



## Polymorphic Characters in Phylogenetic Systematics

John J. Wiens

*Systematic Biology*, Vol. 44, No. 4 (Dec., 1995), 482-500.

Stable URL:

<http://links.jstor.org/sici?sici=1063-5157%28199512%2944%3A4%3C482%3APCIPS%3E2.0.CO%3B2-L>

---

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at <http://www.jstor.org/about/terms.html>. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

*Systematic Biology* is published by Society of Systematic Biologists. Please contact the publisher for further permissions regarding the use of this work. Publisher contact information may be obtained at <http://www.jstor.org/journals/ssbiol.html>.

---

*Systematic Biology*

©1995 Society of Systematic Biologists

JSTOR and the JSTOR logo are trademarks of JSTOR, and are Registered in the U.S. Patent and Trademark Office. For more information on JSTOR contact [jstor-info@umich.edu](mailto:jstor-info@umich.edu).

©2003 JSTOR

# POLYMORPHIC CHARACTERS IN PHYLOGENETIC SYSTEMATICS

JOHN J. WIENS<sup>1</sup>

*Department of Zoology and Texas Memorial Museum, University of Texas,  
Austin, Texas 78712-1064, USA*

*Abstract.*—The use of discrete, intraspecifically variable characters in parsimony analysis is reviewed. Seven data sets (two from morphology and five from allozymes) were analyzed to (1) compare different methods for treating polymorphic characters, (2) test for phylogenetic information in polymorphic characters, and (3) determine if there is a relationship between homoplasy and intraspecific variability. The performance of eight methods was compared using five criteria (number of characters treated as informative, number of shortest trees, phylogenetic signal, number of nodes supported by bootstrapping, and sensitivity to reduced sample size). Approaches that incorporate explicit frequency information perform best overall for all the criteria, although the “majority” method ties for best for the bootstrapping criterion. Levels of phylogenetic information in the polymorphic characters differed greatly among data sets and methods. Polymorphic characters in most data sets contained significant phylogenetic structure using most methods, but only one, the frequency method, extracted significant signal from the polymorphic characters in all seven data sets. Fixed characters appear to contain more signal than polymorphic characters, and homoplasy is significantly and positively correlated with intraspecific variability. This study supports the traditional view that polymorphic characters are less reliable in inferring phylogeny but does not necessarily support their exclusion. Systematists working with morphological data often do not report intraspecific variation, the frequencies of different traits, or how polymorphic characters are screened and analyzed; this situation should change. [Polymorphic characters; character coding; character selection; character weighting; homoplasy; allozymes; morphology.]

Polymorphism, used herein as discrete variation within species that is independent of ontogenetic and sexual variation, is a ubiquitous problem in phylogenetic analysis. If characters evolve, they must vary within species, at least at some point in their history. Yet, there is a long tradition in systematics of excluding characters in which polymorphism is observed. This practice is evident from the scarcity of phylogenetic studies that report intraspecific variation but is difficult to document explicitly because systematists seldom mention the characters that are excluded, rarely give explicit criteria for discarding nonfixed characters, and justify this exclusion even less frequently. Presumably, the basis for excluding polymorphic characters is that they are less reliable in phylogenetic inference (e.g., Darwin, 1859; Simpson, 1961; Farris, 1966; Kluge and Farris, 1969; Mayr, 1969) and have traditionally been difficult to deal with analytically (e.g., Wi-

ley, 1981). Despite the general stigma associated with polymorphic characters, a number of methods have been developed for their treatment (mostly in the context of analysis of allozyme data), and the relative merits of these methods have been debated vigorously (e.g., Farris, 1981; Mickevich and Mitter, 1981, 1983; Buth, 1984; Swofford and Berlocher, 1987; Crother, 1990; Campbell and Frost, 1993; Mabee and Humphries, 1993; Murphy, 1993). Few studies have used analyses of real data sets to examine the properties of polymorphic characters and to compare the different methods for using them (Mickevich and Mitter, 1981; Campbell and Frost, 1993).

In this paper, I review the problem of using discrete, intraspecifically variable characters in parsimony analysis. Specifically, I address three questions: (1) What is the best method for phylogenetic analysis of polymorphic characters? (2) Do polymorphic characters contain useful phylogenetic information? (3) Is there a relationship between levels of intraspecific variability and homoplasy in systematic characters? These questions are addressed

<sup>1</sup> Present address: Section of Amphibians and Reptiles, Carnegie Museum of Natural History, Pittsburgh, Pennsylvania 15213-4080, USA. E-mail: wiens@clpgh.org.

TABLE 1. Summary of data sets used for evaluation of polymorphic characters in phylogenetic analysis. The number of characters represents the maximum number that can be considered phylogenetically informative by any method.

Taxon	Data type	No. characters	No. species	Reference
Phrynosomatid lizards	morphology	115	59	Reeder and Wiens (1996)
<i>Urosaurus</i> (lizards)	morphology	24	9	Wiens (1993)
<i>Crotaphytus</i> (lizards)	allozymes	24	7	Montanucci et al. (1975)
<i>Hybognathus</i> (fish)	allozymes	14	7	Cook et al. (1992)
<i>Rana</i> (frogs)	allozymes	24	8	Green (1986)
<i>Rhagoletis</i> (flies)	allozymes	9	6	Swofford and Berlocher (1987)
Sand lizards	allozymes	19	9	de Queiroz (1992)

through analyses of seven data sets that contain large numbers of polymorphic characters and explicit information on frequencies of polymorphic traits (Table 1).

#### MATERIALS AND METHODS

##### *Comparing Methods for Analyzing Polymorphic Characters*

Given that there is a variety of proposed methods for analyzing polymorphic characters, how does one determine which is best? One way is to compare the results obtained using the different methods on the same data set with one or more optimality criteria (e.g., Mickevich and Mitter, 1981; Archie, 1985).

The best performance criterion would be the frequency with which each method recovers the correct phylogeny (i.e., the relative accuracy of the methods). Although this criterion cannot be applied to most real data sets, there are a number of criteria with either an intuitive or a demonstrated relationship to the accuracy of phylogeny reconstruction that can be used to evaluate different methods. These criteria are as follows.

1. *Number of equally parsimonious trees generated.*—Although a single most-parsimonious tree can be obtained for even randomized data sets (Hillis and Huelsenbeck, 1992), large numbers of shortest trees can lead to poorly resolved consensus trees and can make a single correct phylogeny difficult to find. For these comparisons, I consider methods that generate a small number of trees to be superior to those that generate many.

However, this is a relatively weak criterion.

2. *Number of informative characters.*—Whether or not a polymorphic character is phylogenetically informative can depend on the coding scheme used. Computer simulations have shown that the accuracy of parsimony analysis generally increases rapidly with increasing numbers of characters, given that the characters are of similar reliability and the combination of branch lengths is appropriate for parsimony analysis (e.g., Wheeler, 1992; Bull et al., 1993; Huelsenbeck and Hillis, 1993; Hillis et al., 1994). Thus, I consider methods that are able to extract phylogenetic information from a larger number of characters to be superior.
3. *Phylogenetic signal.*—The skewness of the frequency distribution of tree lengths (the  $g_1$  index) has been shown by computer simulations to be a good overall measure of phylogenetic signal (internal congruence and decisiveness) in a data set (Hillis, 1991; Huelsenbeck, 1991). However, the  $g_1$  index is sensitive to the number of informative characters in a data matrix (Huelsenbeck, 1991), which may differ among methods for the same data set. To provide a comparable index among methods, I used the difference between the observed  $g_1$  and the critical value for random data. Critical values for randomized data were generated for each data set for each method by first making 100 new data matrices from each original matrix, with the states of each character ran-

domly shuffled among taxa. I then calculated the  $g_1$  index for each randomized data set and found the lower 95% confidence limit for the 100  $g_1$  values (the minimum  $g_1$  for random data). (Randomizations were implemented using a program written by J. Huelsenbeck, who also suggested the method for standardizing phylogenetic signal among data sets.) Because some of the data sets contained both fixed and polymorphic characters, this measure of phylogenetic signal simultaneously assessed how well the polymorphic characters agree with each other, the congruence of the polymorphic characters with the fixed characters, and the internal congruence of the fixed characters.

4. *Bootstrapping*.—Nonparametric bootstrapping (Felsenstein, 1985) uses random resampling of characters to estimate the support for a given clade. Studies of simulated and experimental viral phylogenies have found bootstrapping to be a biased but conservative indicator of the accuracy of individual clades under a wide variety of conditions (Hillis and Bull, 1993). Thus, under conditions appropriate for parsimony analysis, bootstrap values of  $\geq 70\%$  indicate that a given clade has a roughly 95% probability of being correctly reconstructed (Hillis and Bull, 1993). The number of nodes with bootstrap values  $\geq 70\%$  within each data set was used for comparison of methods.
5. *Sensitivity to reduced sample size*.—To evaluate the effects of sampling error on the different methods, a series of subsampling experiments was performed on the morphological data from the lizard genus *Urosaurus* (Wiens, 1993). This data set was the only one for which both information on conditions in individual specimens and large sample sizes ( $> 50$  individuals scored for some characters for some species) were available. For each species, five specimens (or fewer, less than five individuals were available from the full data set in a few species for a few characters) were selected randomly from a list of all in-

dividuals examined, and these specimens were used to make a new matrix of raw frequencies. This resampling procedure was repeated to make 20 new matrices of raw frequencies. These matrices were then coded according to the eight methods for using polymorphic characters. For each method, the tree(s) based on the subsampled matrices was compared with the tree(s) based on the complete sample. The congruence of the subsample trees with the full-sample trees was used as the index of sensitivity to sampling error; high similarity between the full and reduced sample trees indicates robustness to sampling error, whereas less congruence indicates sensitivity to sampling error. Similarity between trees was assessed using the consensus fork index (Colless, 1980): the number of clades in common between the trees from the full and reduced samples divided by the total number of nontrivial clades possible (the number of taxa minus 2). Presumably, methods that are highly sensitive to variation in sample size will be less likely to consistently reconstruct the true phylogeny.

Recent work with computer simulations and experimental phylogenies has upheld the relationship of at least three of these criteria (number of informative characters, phylogenetic signal, and bootstrapping) to accuracy. However, these indices are useful indicators of accuracy only when conditions are amenable to accurate phylogeny reconstruction by parsimony (i.e., when branch lengths are short and relatively similar; Huelsenbeck and Hillis, 1993). This study addresses the merit of different parsimony approaches under conditions where the basic assumptions of parsimony analysis presumably are not violated and does not address the relative sensitivity of these methods to violations of these assumptions. It seems likely that the data sets analyzed do not grossly violate these assumptions because of (1) the presumed independence of each of the allozyme loci and morphological characters and (2) the

TABLE 2. Summary of methods for coding polymorphic characters for parsimony analysis, where 0 is the primitive condition, 1 is derived, and 0/1 indicates polymorphism.

Method	Summary
Any-instance	0/1 or 1 = 1
Frequency	0/1 = frequency of trait 1
Majority	if frequency of 1 $\geq$ 50%, then 0/1 = 1, otherwise 0/1 = 0
Missing	0/1 = ?
Polymorphic	0/1 = (0, 1)
Scaled	0 = 0, 0/1 = 1, 1 = 2; ordered 0 → 1 → 2, change from 0 → 2 is two steps
Unordered	same as scaled, but unordered (0 ↔ 1 ↔ 2 ↔ 0)
Unscaled	same as scaled (ordered), but change from 0 → 2 is one step

close relationship of the species in each data set (mostly congeners; Table 1), making the attraction of long branches seem doubtful.

Values were obtained for each criterion for each method for each data set, and methods were ranked for each criterion within each data set. For each criterion, the relative rankings of the different methods were then summed across data sets to determine which method(s) performs best overall for each criterion. The methods compared are listed below (terminology mostly from Campbell and Frost, 1993) and are summarized briefly in Table 2.

1. *Any-instance*.—A derived condition present at any frequency is coded the same (0/1 [polymorphic] or 1 [fixed] = 1 [fixed]), so that species with the derived condition fixed or polymorphic are treated identically. It is not clear how multistate characters would be coded for this method, however (more than one state in a polymorphic species could be derived), so it was not applied to the allozyme data sets. This approach, referred to as the mutation model, was recommended by Murphy (1993).
2. *Missing*.—A polymorphic species is coded as missing or unknown for the variable character. The polymorphic species is treated as if any state could be assigned to the missing cell, making that cell uninformative in building the tree (Swofford, 1993). The presence of the derived condition as a polymorphism is not used as evidence to group the polymorphic species with other species sharing the derived trait. This method was recommended by Pimentel and Riggins (1987).
3. *Polymorphic*.—A species with two traits at any frequency is coded as polymorphic. In the context of the PAUP program (Swofford, 1993) (this option is unavailable in Hennig86), one of the states observed will be assigned to the polymorphic taxon a posteriori, depending on how the taxon is placed on the tree by other characters. As with the missing method, the presence of the derived condition as a polymorphism is not used as evidence to group the polymorphic species with other species sharing the derived trait. In contrast to the missing method, however, the placement of the polymorphic species is constrained by the observed traits because it is more parsimonious if the placement of the polymorphic species does not require transformation to a trait that is not observed in that species. For characters with only two conditions (binary in the usual sense of the term), the missing and polymorphic methods yield identical results. In my analyses, a single species in which three traits were observed was coded as unknown for this method (because if three states are entered only two can be used by PAUP). Swofford (1993) disavowed the polymorphic option for coding intraspecific variation, although it was used by Campbell and Frost (1993) as their "baseline" method.
4. *Majority*.—The state present in the majority of specimens is coded as being present. For this analysis, I considered frequencies of the derived condition of 0–49% as absence and frequencies of 50–100% as presence. This method was used by Johnson et al. (1988, 1989).
5. *Scaled*.—For characters with only two conditions, polymorphic species are

coded as one state (e.g., 1), and those fixed with the derived state are coded with another (e.g., 2). The character is then ordered from absent to polymorphic to fixed (e.g., 0 → 1 → 2). Characters in which no polymorphism is observed are assumed to have gone through a polymorphic state, so that these characters are weighted by 2 relative to those in which polymorphism is observed. This method is equivalent to that of Mabee and Humphries (1993), which uses step matrices to implement the weighting scheme on characters with multiple derived conditions (i.e., allozyme data).

6. *Unordered*.—This method is the same as the scaled method, except that the character states are unordered. Thus, no special connection is implied between polymorphic and fixed conditions, so that it is as parsimonious to evolve a derived state independently in a fixed and polymorphic species as to have the state go from polymorphism to fixation (Campbell and Frost, 1993; Mabee and Humphries, 1993).
7. *Unscaled*.—This method is almost the same as the scaled method, except that characters in which no polymorphic state is observed are not weighted by 2. Thus, the appearance of the derived condition at any frequency (1–100%) is accorded one step, but the transition from polymorphism to fixation (if a polymorphic taxon is observed) is also considered to be one step (Campbell and Frost, 1993). This method was preferred by Campbell and Frost (1993). I was uncertain as to how to apply this method to characters in which a species may have multiple derived traits, so it was used only on the morphological data sets.
8. *Frequency*.—Frequency methods use the frequencies of the different traits within a species. Swofford and Berlocher (1987) developed an algorithm and corresponding software for direct parsimony analysis of frequency data (FREQPARS). However, FREQPARS has a weak tree-searching routine and is unlikely to find

TABLE 3. Character state designations (from a to y) for frequencies of the derived trait used for the frequency bins method.

Character state	Frequency range (%)
a	0–3
b	4–7
c	8–11
d	12–15
e	16–19
f	20–23
g	24–27
h	28–31
i	32–35
j	36–39
k	40–43
l	44–47
m	48–51
n	52–55
o	56–59
p	60–63
q	64–67
r	68–71
s	72–75
t	76–79
u	80–83
v	84–87
w	88–91
x	92–95
y	96–100

the best tree (by its own criteria) for even small data sets (Swofford and Berlocher, 1987).

To approximate the use of FREQPARS, I used two alternative methods. For the morphological data sets, I used ranges of frequencies of the derived condition as character states (e.g., character state a = frequency of derived state from 0% to 3%; b = 4–7%, etc.; see Table 3). These character states were then ordered, assuming that a derived feature must go through a polymorphic stage between absence and fixation. Unpolarized characters were treated by simply choosing one of the two conditions to code as if it were derived (leaving the hypothetical ancestor as unknown). Coding ranges of frequencies clearly is less precise than using frequencies directly. For example, using this method, there is no cost in a change in the frequency of a derived trait from 0% to 3%, but there is a cost to go from 3% to 4%. However, this method none-

theless incorporates considerable frequency information and allows use of the powerful tree-searching algorithms used in more conventional parsimony analyses. This method was used by Wiens (1993; but with fewer frequency "bins"), and a similar range-coding method was used for the frequency of individual alleles at allozyme loci (treating each allele at a locus as an independent character) by Prober et al. (1990).

The frequency bins method was used for the morphological data set because for most of these characters there is only a single derived condition (binary characters in the usual sense of the term). For allozyme data sets, each species was given a unique character state, and Manhattan distances between species (based on differences in allele frequencies) were used to weight changes between the states in a step matrix. This method will be described in detail and compared further elsewhere (Hillis, Chippindale, and Wiens, unpubl.). In summary, both methods make use of frequency information but still allow powerful tree searches.

Parsimony analyses were performed using PAUP 3.1+6 (Swofford, 1993). For the morphological data sets, characters not based on my observations (i.e., myological, molecular, and behavioral characters), characters with more than one derived condition (multistate characters in the usual sense of the term), and characters that were uninformative under any coding scheme were excluded. Because the any-in-state method requires the use of polarized characters, the unpolarized characters in the phrynosomatid data set were polarized based largely on optimizing the ancestral condition on the combined tree of Reeder and Wiens (1996), following Campbell and Frost (1993). For the allozyme data sets, I used only species as terminal taxa and lumped frequency information presented for different populations of the same species because I did not wish to confound my comparisons of methods with the problem of reconstructing intra-

specific phylogenies. For the analysis of de Queiroz's (1992) sand lizard data, I did not include his hypothetical reconstructed outgroup taxa in my analyses. Otherwise, all species (both ingroup and outgroup) were included in the allozyme analyses.

In coding the allozyme data sets, the locus, rather than the presence or absence of individual alleles, was treated as the character because the presence, absence, and frequency of different alleles at a locus are not independent of each other (e.g., Buth, 1984).

The two morphological data sets were chosen because they contained explicit information on frequencies of traits and because polymorphic characters were not excluded solely due to intraspecific variability. The allozyme data sets were chosen because of their relatively large sample sizes per species (generally >10 individuals/species/locus) and concomitant high proportion of polymorphic characters. I would not expect the different methods to yield very different results if many of the characters were not polymorphic. Other allozyme data sets were excluded because they included too few species to be informative (less than four) or did not include raw frequency information. (All the data matrices analyzed are available in annotated electronic format from the author.)

#### *Phylogenetic Information in Polymorphic Characters*

To test for significant phylogenetic signal in the polymorphic characters, I compared the observed  $g_1$  to the critical  $g_1$  for randomized data (calculated as described above) for the polymorphic characters in the seven data sets. Because polymorphic characters were analyzed alone, this index measured only the extent to which the polymorphic characters are congruent with each other. Among the seven data sets, the phrynosomatid lizard data set is the only one with large numbers of both fixed ( $n = 33$ ) and polymorphic ( $n = 79$ ) characters; the other data sets have no fixed characters (*Crotaphytus*, *Hybognathus*, sand lizards), one fixed character (*Rhagoletis*, *Urosaurus*), or two fixed characters

(*Rana*). The phrynosomatid data set was used to compare the levels of phylogenetic signal ([observed  $g_1$ ] – [critical  $g_1$  for random data]) in the two classes of characters.

#### *Intraspecific Variability and Homoplasy*

To explicitly examine the relationship between homoplasy and intraspecific variability, I quantified the levels of homoplasy and variability for each character in the *Urosaurus* and phrynosomatid data sets and tested the relationship between them using Spearman's rank correlation (implemented using Statview<sup>®</sup>). The measure of homoplasy used was the homoplasy index (1 – consistency index) (Swofford, 1993). The consistency index is the minimum change possible in a character (i.e., number of steps) divided by the observed change (Kluge and Farris, 1969). Although the consistency index has been criticized as a comparative measure of homoplasy because of its sensitivity to the number of taxa and characters in a data set (Archie, 1989), it was used only to compare levels of homoplasy among characters within a data set. Changes hypothesized to occur within polymorphic terminal taxa were not included in calculations of tree length and consistency index when the polymorphic coding method was used. Because all characters (both fixed and polymorphic) were included in these analyses, the homoplasy index measured incongruence both between and within characters of the two classes. The allozyme data sets were not used in these analyses because the scaled and frequency methods require the use of step matrices for the allozyme data and it currently is not possible to calculate the expected change for step matrix characters.

The homoplasy index for each character was calculated based on the trees generated by each of the eight coding methods. When a method generated multiple equally parsimonious trees for a data set, the homoplasy index for each character was averaged across the trees. Some methods generated large numbers of shortest trees (i.e., >20,000). In cases where there were  $\geq 300$  equally parsimonious trees, I randomly sampled 200 of them and then used

these trees to estimate the levels of homoplasy in individual characters. A HyperCard<sup>™</sup> program (written by D. Cannatella) was used to read consistency indices from a PAUP output file, average them, and convert them to homoplasy indices.

Although there are a number of measures to assess the intraspecific variability of quantitative traits, the best measure for discrete characters is not obvious. I considered a species to be the most variable when the alternative conditions for a character were present at nearly equal frequencies (e.g., 50%). The index used was the mean intraspecific variability (MIV), which for a given character is the sum of the frequencies of the rarer condition in each species, multiplied by 200 (to scale the index to 100), and divided by the number of species:

$$\text{MIV} = \left[ \frac{\sum \text{frequency of rarest trait in each species}}{\text{number of species}} \right] (200)$$

This value provides an index for a character with two conditions (binary in the usual sense of the word), which has a score of 100 when all the species have both traits at a frequency of 50% and a score of 0 when none of the species are intraspecifically variable.

To assess sensitivity to sample size, I also examined the relationship between homoplasy and variability using 10 of the subsampled *Urosaurus* data matrices (with maximum sample size of five individuals per species) with the eight different coding methods.

## RESULTS

### *Comparison of Methods*

The relative performances (the rankings for the five criteria summed across the data sets) of six of the methods are shown in Table 4. The any-instance and unscaled methods are excluded from these rankings because they could not be assessed for the allozyme data. The rankings of all eight methods based on the morphological data sets only are shown in Table 5. In general, there was a great deal of variation among data sets in the performances of the differ-



TABLE 4. Relative performance (based on five criteria) of six methods for parsimony analysis of polymorphic characters summed across the seven morphological and allozyme data sets. Methods are listed in order of best to worst, with brackets showing ties. Raw scores for all criteria are given in Appendices 1-5.

No. characters	No. trees	Phylogenetic signal	Bootstrapping	Sampling error ( <i>Urosaurus</i> only)
frequency	frequency	frequency	[ frequency ]	frequency
scaled	scaled	polymorphic		majority
polymorphic	polymorphic	majority	polymorphic	scaled
majority	majority	[ missing ]	scaled	unordered
unordered	unordered		unordered	missing
missing	missing	scaled	unordered	

ent methods for the different criteria. However, the frequency method always ranked best (or was tied for best) for number of trees, number of informative characters, and robustness to reduced sample size and generally performed well for phylogenetic signal and proportion of nodes with bootstrap values  $\geq 70\%$ . Although the frequency method is not the best method for every criterion for every data set, it ranks as best overall for all the criteria when relative performance ranks are summed across all the data sets (but the majority method tied for best for the bootstrapping criterion). The raw scores for all criteria, methods, and data sets are given in Appendices 1-5.

#### *Phylogenetic Information in Polymorphic Characters*

The levels of phylogenetic signal observed in the polymorphic characters in the seven data sets using the eight methods are presented in Table 6. These levels differ greatly among data sets and among methods, but there is significant phylogenetic information in the polymorphic characters in most of the data sets using most of the methods. However, the frequency method is the only method that recovered significant phylogenetic signal from the polymorphic characters in all the data sets examined; each of the other methods completely obscured the phylogenetic information in the polymorphic characters in at least one of the data sets. These results suggest that polymorphic characters do contain significant phylogenetic information and that the frequency method is the best way to extract it.

Comparison of the levels of phylogenetic signal in the fixed and polymorphic characters in the phrynosomatid lizard data suggests that fixed characters contain much more signal (observed  $g_1$  - critical random  $g_1 = -1.850$ ) than do polymorphic characters, regardless of how the polymorphic characters are coded (range =  $-0.294$  to  $-0.490$ ; Table 6).

#### *Homoplasy and Intraspecific Variability*

For the phrynosomatid data set, there is a strong positive relationship between homoplasy and intraspecific variability, regardless of the coding method used (Table 7). A sample graph indicating this relationship is shown in Figure 1, with the data coded using the frequency method.

The *Urosaurus* data also suggest that there is a positive relationship between homoplasy and intraspecific variability but that this relationship is sensitive to the method used and to sample size (Table 8). For full sample sizes, the correlation is significant ( $P < 0.01$ ) only when the frequency and majority methods are used (see Fig. 1), although the correlation is nearly significant for the unordered method. This relationship is much stronger when assessed on the subsampled data matrices for the any-instance, scaled, unordered, and unscaled methods.

#### DISCUSSION

##### *Do Different Methods Yield Different Trees?*

An important consideration when evaluating methods is whether different methods actually make a difference in the topologies obtained. When there is little

TABLE 5. Relative performance (based on five criteria) of the eight methods for parsimony analysis of polymorphic characters summed across the two morphological data sets. Methods are listed in order of best to worst, with brackets showing ties. Raw scores for all criteria are given in Appendices 1-5.

No. characters	No. trees	Phylogenetic signal	Bootstrapping ( <i>Urosaurus</i> only)	Sampling error ( <i>Urosaurus</i> only)
[ frequency scaled unscaled ]	[ frequency any-instance scaled ]	[ frequency missing polymorphic ]	majority frequency	frequency majority
[ any-instance unordered ]	[ unscaled majority unordered ]	[ majority scaled ]	[ missing polymorphic scaled ]	scaled unordered
[ majority ]	[ missing ]	[ unordered unscaled ]	[ unscaled ]	any-instance
[ missing polymorphic ]	[ polymorphic ]	any-instance	[ any-instance unordered ]	

polymorphism in a data set, there is no reason to expect substantial differences in the trees generated by the different methods. But when the number of polymorphic characters is large, the differences can be striking. As an example, the trees (or consensus trees) produced by the different methods for the *Urosaurus* data are shown in Figure 2. For this data set, there is only one clade (*U. auriculatus* + *U. clarionensis*) shared among all the trees produced by the different methods, and many of the trees are very different from one another (compare the frequency and any-instance trees).

Patterns of congruence among the trees produced by the different methods are shown in Table 9. This table summarizes the proportion of data sets in which there are one or more trees in common among the sets of trees produced by a given method. The most obvious patterns are the con-

sistent agreement between the frequency and majority trees (one or more trees in common for every data set) and the consistent disagreement between the frequency and scaled trees (no trees in common for any data set).

#### *Superiority of the Frequency Method*

This study shows that the frequency method generally performs better than the other methods for all the criteria examined (although the majority method performed as well overall for the bootstrapping criterion). Why is the frequency method superior? The frequency method has two important advantages: (1) it uses the most phylogenetic information, and (2) it reduces noise and sampling error by giving minimal weight to conditions occurring at low frequencies (Swofford and Berlocher, 1987).

Swofford and Berlocher (1987) provided

TABLE 6. Phylogenetic signal in polymorphic characters. Phylogenetic signal was measured as the observed  $g$ , minus the expected critical value for the data set for the method when randomized. Negative values indicate significant phylogenetic signal; values  $\geq 0$  indicate that the phylogenetic information is indistinguishable from random data. The any-instance method could not be applied to the allozyme data sets.

Method	Data sets						
	Phrynosomatidae	<i>Urosaurus</i>	<i>Crotaphytus</i>	<i>Hybognathus</i>	<i>Rana</i>	<i>Rhagoletis</i>	Sand lizards
Any-instance	-0.370	+0.028					
Frequency	-0.413	-0.503	-0.070	-1.165	-0.214	-0.151	-0.579
Majority	-0.368	-0.448	-0.208	-1.056	-0.144	+0.157	-0.626
Missing	-0.490	-0.391	-0.033	-0.482	-0.264	0.000	-0.455
Polymorphic	-0.490	-0.391	-0.101	-0.964	-0.321	0.000	-0.444
Scaled/unscaled <sup>a</sup>	-0.447	-0.100	+0.006	-0.884	-0.257	+0.113	-0.490
Unordered	-0.294	+0.055	-0.114	-0.369	-0.355	0.000	-0.539

<sup>a</sup> Results from the scaled and unscaled methods are identical because they differ only in how fixed characters are weighted.

TABLE 7. The results of Spearman's rank correlation ( $\rho$ ) of homoplasy and mean intraspecific variability for the phrynosomatid lizard data using the eight methods for coding polymorphic characters.

Method	$\rho$	$P$
Any-instance	0.823	0.0001
Frequency	0.779	0.0001
Majority	0.671	0.0001
Missing/polymorphic <sup>a</sup>	0.326	0.0006
Scaled	0.799	0.0001
Unordered	0.797	0.0001
Unscaled	0.789	0.0001

<sup>a</sup> Missing and polymorphic methods yield identical results for binary characters.

a convincing argument to explain the problem of rare conditions. They showed statistically that conditions occurring at low frequencies will often be missed, even when sample sizes are very large. By chance, these rare traits will be missed in some species and detected in others but not necessarily in taxa that are closely related. Thus, a rare trait is likely to appear to be a synapomorphy for a group of species that are not closely related. Such cases will be problematic for the any-instance, scaled, unordered, and unscaled methods, which give rare and frequent variants equal weight. Using a frequency method, these rare polymorphisms are given very little weight (almost as if they were not de-

tected at all), and their potentially misleading impact is minimized.

The problem of rare conditions can be demonstrated empirically with the *Urosaurus* subsampling experiments. When sample sizes are reduced, conditions occurring at low frequencies are more likely to go undetected (Swofford and Berlocher, 1987). When the methods that weight rare and common variants equally, i.e., any-instance, scaled, unordered, and unscaled, are applied to the 20 subsampled matrices ( $n \leq 5$ /species), we can predict that (1) trees based on the subsampled matrices will be very different from the trees based on the full sample and (2) phylogenetic signal should increase with the decreased sample size.

These four methods never retrieved the tree(s) based on the full sample and consistently recovered trees that differed greatly from the full-sample trees (Fig. 3). The different methods seem to converge somewhat on the frequency tree upon subsampling. The scaled, unscaled, and any-instance methods yielded trees that were identical or consistent with the frequency tree in 3 of 20 replicates, and the unordered method did so in 6 of 20 replicates. Subsampling effectively downweights the rare polymorphisms by causing them to go

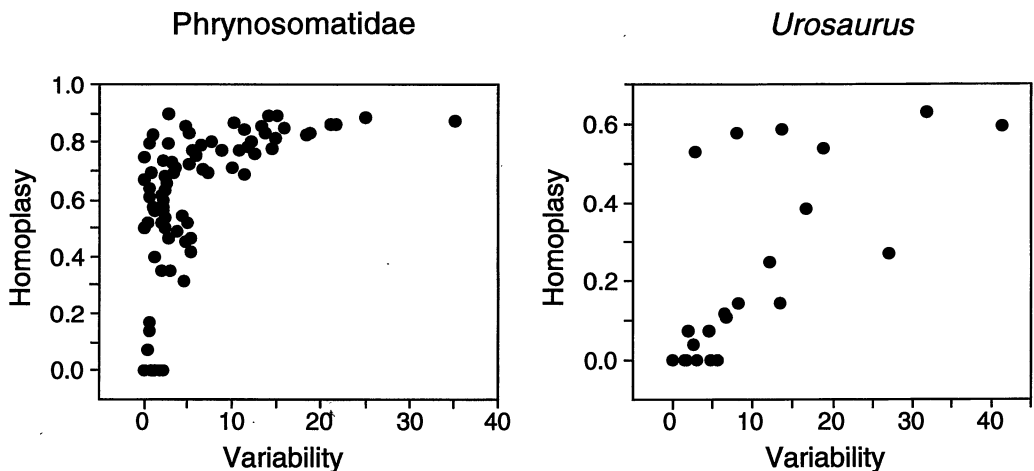


FIGURE 1. The relationship between homoplasy and intraspecific variability for the phrynosomatid lizard and *Urosaurus* data sets using the frequency coding method.

TABLE 8. The results of Spearman's rank correlation ( $\rho$ ) of homoplasy and mean intraspecific variability for the *Urosaurus* data using the eight methods for coding polymorphic characters. To assess the sensitivity of this relationship to sample size, the relationship also was examined for 10 subsampled data matrices ( $n \leq 5$  individuals/species).

Method	Full sample		Subsample	
	$\rho$	$P$	$\rho$	$P$
Any-instance	0.099	0.635	0.529	0.021
Frequency	0.831	0.0001	0.756	0.001
Majority	0.518	0.013	0.487	0.029
Missing/polymorphic <sup>a</sup>	-0.166	0.507	0.193	0.469
Scaled	0.145	0.486	0.681	0.002
Unordered	0.472	0.024	0.621	0.006
Unscaled	0.145	0.486	0.571	0.030

<sup>a</sup>Missing and polymorphic methods yield identical results for binary characters.

undetected, resulting in trees that are similar to the ones in which rare polymorphisms are given less weight.

Noise clearly is decreased by the reduced sample sizes when the any-instance, scaled, unordered, and unscaled methods are used (Fig. 4). Thus, these four methods seem to have a disturbing tendency to give

worse results as more data are included, as predicted by Swofford and Berlocher (1987). Because the frequency method buffers against rare conditions (Swofford and Berlocher, 1987), there is relatively little change in topology and phylogenetic signal when the method is applied to the subsampled matrices (Figs. 3, 4).

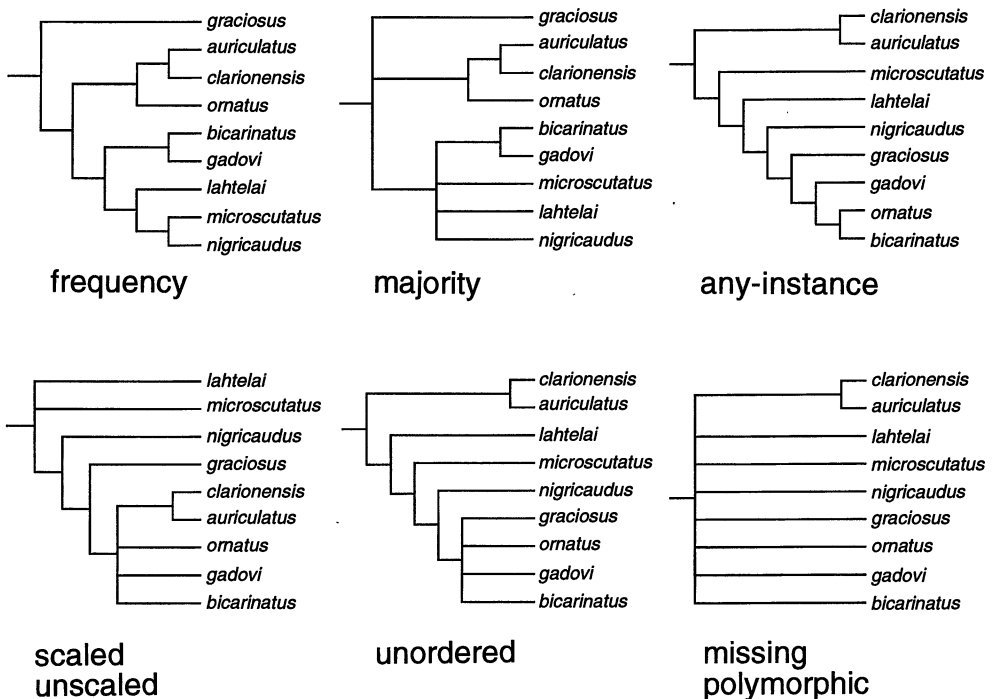


FIGURE 2. Trees produced by the eight different methods for parsimony analysis of polymorphic characters for the *Urosaurus* data set. The trees shown represent either the single shortest tree (any-instance, frequency) or the strict consensus tree derived from several equally parsimonious trees.

TABLE 9. Patterns of congruence among the results from different methods for analyzing polymorphic characters, measured as the percentage of data sets in which there is one or more trees in common among the sets of trees produced by two methods. Because of the unmanageable number of trees, results from the phrynosomatid lizard data set were not used for this comparison. The any-instance and unscaled methods could not be applied to the allozyme data sets.

	Fre- quency	Majority	Missing	Scaled	Un- ordered
Majority	100				
Missing	67	83			
Scaled	0	17	33		
Unordered	17	50	67	33	
Polymorphic	50	67	83	17	50

The majority method also gives little weight to rare variants (they are effectively ignored) and generally gives results similar to those of the frequency method (Table 9). However, this method discards much potentially useful information about frequencies (e.g., a frequency of 0% is treated

the same as a frequency of 49%) and so does not perform as well overall as the frequency method (Table 4). The polymorphic and missing methods also discard considerable information, although the polymorphic method retains more than does the missing method.

#### *Objections to the Frequency Method*

The use of explicit frequency information has been a particularly controversial aspect of the phylogenetic analysis of polymorphic data. Swofford and Berlocher (1987) argued that direct use of frequencies is the best method for using polymorphic data because it is the least sensitive to sampling error. Their conclusions are strongly upheld by empirical data in this study. Mickevich and Johnson (1976), Farris (1981), Mickevich and Mitter (1981, 1983), Crother (1990), Carpenter et al. (1993), Mabee and Humphries (1993), and others have questioned whether frequencies con-

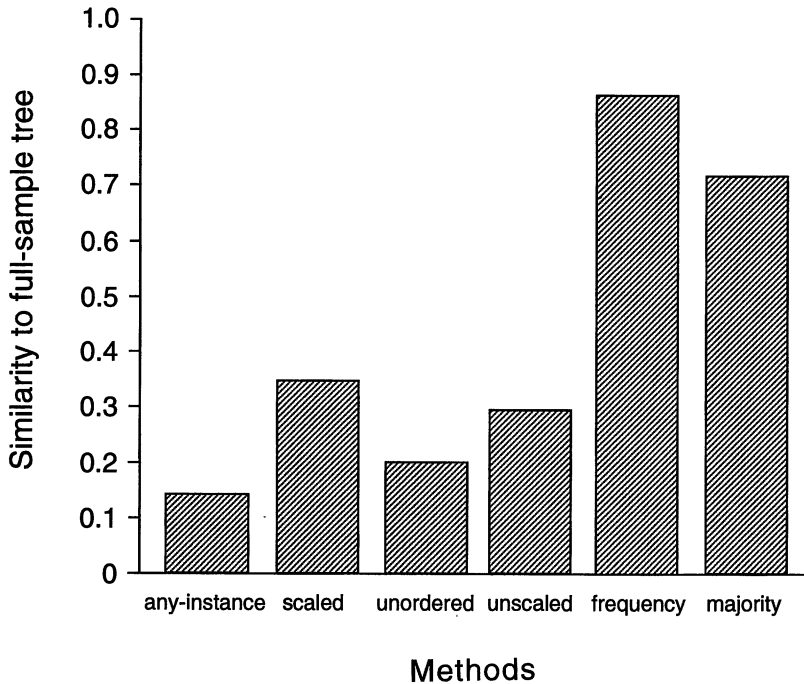


FIGURE 3. Mean similarity between trees from the original *Urosaurus* data set and trees based on 20 replicated matrices with reduced sample sizes ( $n \leq 5$ ) for six methods of coding polymorphic data. The missing and polymorphic methods are not shown because of the large number of shortest trees produced by the original matrix (489).

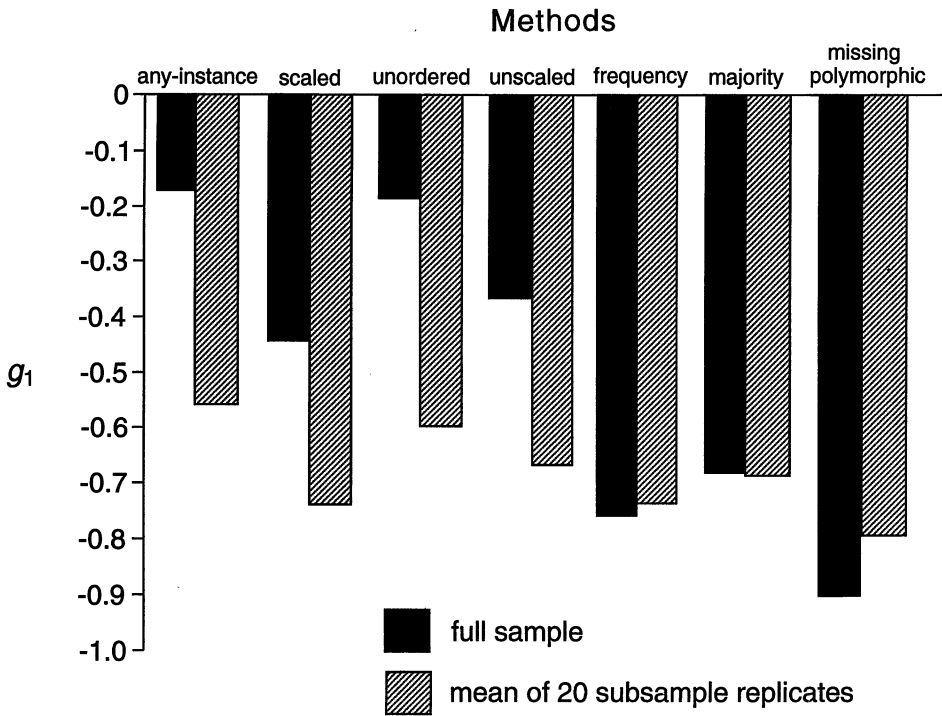


FIGURE 4. Effect of sample size on levels of phylogenetic signal (measured using the  $g_1$  index) using different methods of coding polymorphic characters for the *Urosaurus* morphological data set. Larger sample sizes decrease phylogenetic signal using the any-instance, scaled, unordered, and unscaled methods.

tain phylogenetically useful information at all, largely because of their variability in space and time. Empirical studies have shown that frequencies of certain traits can vary considerably over space and time within certain species (e.g., Gaines et al. [1978] as used by Crother [1990]). However, the relevance of these cases to the problem of reconstructing relationships among several species using several characters has simply been assumed. The analyses of phylogenetic signal in the polymorphic characters alone using the frequency method suggest that frequencies do contain significant phylogenetic information on interspecific relationships (Table 6).

Other objections concern the nature of the evidence that frequency data provide. Authors have rejected the use of frequencies because they are continuous, because changes in frequency do not represent evolutionary 'novelties', and/or because frequencies purportedly are not heritable nor

are they organismal traits (e.g., Crother, 1990; Campbell and Frost, 1993; Murphy, 1993). The basic disagreement lies in what can legitimately be considered a character. Thiele (1993) argued convincingly that a character state is merely a frequency distribution of traits for a given sample of a given taxon.

A potential problem of the frequency method is that it usually yields a fully resolved tree, regardless of the level of support. Thus, clades might be supported by only a minor (e.g., 1%) change in frequencies. For example, when the *Urosaurus* data are randomized, the frequency method usually gives only one ( $\bar{x} = 1.1$ ;  $n = 100$ ) most-parsimonious tree for each randomized data matrix, although these trees are clearly not well supported. However, none of the coding schemes distinguish between well and poorly supported clades, and there are many methods available to address the stability of specific nodes after an

initial analysis (e.g., Faith's T-PTP [1991], decay index [Donoghue et al., 1992], non-parametric bootstrapping [Felsenstein, 1985]). The other methods are also likely to resolve nodes supported by only minor changes in frequencies (e.g., 0–1%, 99–100%, 49–50%).

#### *Polymorphism in Asexual Species and Characters from Nonrecombining Organelles*

This discussion has dealt implicitly with bisexual species. Should polymorphism within asexual species be treated differently? If a polymorphism in a unisexual species can be maintained through a speciation event, then the presence of a polymorphic condition in two unisexual species can be evidence that they are closely related, just as in sexual species. Thus, the same methods are applicable for treating polymorphism in sexual and asexual species, and the frequency method is most appropriate for both. The same applies for coding characters from nonrecombining organelles, such as DNA sequences from chloroplasts and mitochondria.

#### *Phylogenetic Information in Polymorphic Characters*

The results of this study indicate that polymorphic characters can contain significant phylogenetic information. However, this phylogenetic information can be completely obscured by many of the methods for coding polymorphic characters, at least for some data sets. The frequency method always extracted significant signal from the polymorphic characters in all the data sets, which further supports the use of this method. These results also confirm the idea that fixed characters contain more phylogenetic signal than do polymorphic characters.

Campbell and Frost (1993) recently addressed the issue of phylogenetic information content of polymorphic characters using data on external morphology from the anguid lizard genus *Abronia*. They also concluded that polymorphic characters contain phylogenetic signal and that they show more noise than fixed characters, but their conclusions were based on criteria

different from those used in the present study. Campbell and Frost (1993) concluded that polymorphic characters contain phylogenetic signal because many of the variable characters were "coherent to varying degrees" when optimized onto the tree based on fixed characters only, although it is not clear how this criterion could have been used to show that the polymorphic characters did not contain phylogenetic signal. Campbell and Frost (1993) considered the variable characters to contain more homoplasy than the fixed characters because more equally parsimonious trees were found when polymorphic characters were included (despite the larger number of characters), but the general relationship between number of shortest trees and data set quality is questionable (Hillis and Huelsenbeck, 1992). Although I prefer other criteria, the ones used by Campbell and Frost (1993) do have intuitive appeal, and I believe the agreement between our conclusions is not a coincidence.

#### *Homoplasy and Intraspecific Variability*

The *Urosaurus* and phrynosomatid data support the idea that there is a strong, positive relationship between intraspecific variability and homoplasy, confirming the long-standing conventional wisdom that more intraspecifically variable characters are inherently less reliable in inferring phylogeny (e.g., Darwin, 1859; Simpson, 1961; Mayr, 1969). Although the present study may be the first to specifically address the relationship between homoplasy and intraspecific variability, several studies have examined the relationship between variability and evolutionary rate within and between populations (e.g., Kluge and Kerfoot, 1973; Sokal, 1976; Baker and Moeed, 1979; Pierce and Mitton, 1979; Baker, 1980; Wiens and Rottenberry, 1980). These studies purportedly found a correlation between levels of variability within and between populations (often called the Kluge-Kerfoot phenomenon) but were later shown (by Rohlf et al., 1983) to be plagued with statistical artifacts.

Several factors may account for the relationship between homoplasy and intra-

specific variability. Many of the polymorphisms in morphology may not track the phylogeny because they do not have a strictly genetic basis. These polymorphisms could be the result of environmental or epigenetic effects in ontogeny or even ambiguities or errors in scoring characters. Although the problem of nonheritable variation alone might be taken as a rationale for excluding polymorphic morphological characters from phylogenetic analyses (or morphological characters in general), it is clear that the polymorphic morphological characters examined do have significant phylogenetic signal as a whole (Table 6). The greater homoplasy in more intraspecifically variable characters may also be due to sampling error; estimates of the presence, absence, or frequency of a trait will generally be less accurate for species that are polymorphic than for those that are fixed, given finite sampling of individuals and populations. Thus, there could appear to be homoplasy in a polymorphic character even if the actual distribution of frequencies among species matched the phylogeny perfectly. Higher levels of homoplasy in polymorphic characters might also result if the frequencies of polymorphic traits in ancestral species are extensively modified when passed on to their descendant species. For example, if there were geographic variation in the frequency of a polymorphic allele in an ancestral species (and speciation occurred by geographic subdivision), a derived allele present in only part of the range of the ancestral species would be rare or absent in one of the daughter species. If this process occurred repeatedly, the distribution of the polymorphic allele would show little congruence with the phylogeny. Farris (1966) suggested that characters that are highly variable within populations may be more variable between populations because they are less constrained by natural selection.

If there is a general relationship between homoplasy and variability, how could this information be used to improve phylogenetic inference? The most important question may be whether polymorphic characters should be included in phylogenetic

analysis at all. Although they can contain significant phylogenetic information, they also can contain more homoplasy than fixed characters. Campbell and Frost (1993) argued that polymorphic characters must be included in phylogenetic analyses because the tree that excludes them may not be the shortest tree for all the data (e.g., fixed and nonfixed characters). Yet, the change in tree topology caused by their inclusion could make the phylogenetic estimate less similar to the correct phylogeny. Intuitively, including more characters with significant phylogenetic information should improve the accuracy of phylogeny estimation (the probability of inferring the true phylogeny), whereas adding characters that are more homoplastic should decrease it. These intuitions are supported by computer simulation studies (e.g., Bull et al., 1993; Hillis et al., 1994). Computer simulations (Chippindale and Wiens, 1994) also have shown that downweighting more rapidly evolving characters can improve the accuracy of phylogeny reconstruction relative to excluding these more homoplastic characters or weighting the rapidly and slowly evolving characters equally. Weighting characters by the inverse of their intraspecific variability (as suggested by Farris, 1966) offers a possible approach to improving accuracy with polymorphic characters; phylogenetic information from all the characters is included, yet the characters that are expected to be more homoplastic are weighted accordingly.

Campbell and Frost (1993) recommended successive approximations (i.e., reweighting characters based on their levels of homoplasy observed in an initial unweighted analysis [Farris, 1969]), coupled with the unscaled method, as the best way to increase the information extracted from polymorphic characters. Given that there is a relationship between homoplasy and variability, this approach might yield similar benefits (and similar results) to the a priori weighting scheme suggested above. However, the relationship between homoplasy and variability can be highly sensitive to the method used, and the unscaled method seems particularly unlikely to re-



flect this relationship (Table 8). Thus, high weights may be given to the most variable characters, and vice versa. Furthermore, successive approximations itself remains controversial (e.g., Cannatella and de Queiroz, 1989; Swofford and Olsen, 1990; Guyer and Savage, 1992; Campbell and Frost, 1993).

Given the demonstrated phylogenetic information in polymorphic characters (Table 6), that there is no real distinction between polymorphic and fixed characters (because all characters must be polymorphic at some point in their evolution), and that apparent absence of polymorphism in seemingly fixed characters may only be an artifact of sampling error (e.g., Campbell and Frost, 1993), it seems difficult to justify the exclusion of characters in which polymorphism is observed. Yet, the fact that homoplasy may increase with increasing variability should not be ignored. Developing and evaluating different ways to incorporate this information should be a rewarding subject for future research.

#### *Traditional Systematics*

Many published phylogenetic analyses from the past 30 years give the impression that there is no intraspecific variation in morphology. A typical data matrix contains only rows of 0's, 1's, and 2's, and the character descriptions simply list the different morphological traits. There often is no mention or discussion of intraspecific variation. Yet, it is difficult to believe that there was no intraspecific variation observed in any of the characters examined. Were characters excluded if found to be intraspecifically variable, or was variation observed within species simply ignored, or both? In some studies, the authors have included polymorphic characters and/or mentioned characters that were excluded because of intraspecific variability, but rarely is there a general discussion of how polymorphism was coded, a quantitative assessment of how much variability there was in a character, or a discussion of how much variation was too much for a character to be included and why. This largely unwritten policy of "don't ask, don't tell"

regarding polymorphism in morphology is simply unscientific. Instead, morphologists should follow the example of systematists using allozyme data and should report the observed frequencies of the different traits for all characters that are adequately characterized, as well as the method used for coding polymorphism and the criteria for character selection.

#### ACKNOWLEDGMENTS

For many valuable discussions and for advice, I thank D. Cannatella, P. Chippindale, D. Frost, D. Hillis, J. Huelsenbeck, P. Mabee, J. McGuire, T. Reeder, M. Servedio, A. Simons, and T. Titus, although my gratitude should not imply their agreement with my views. I am especially grateful to D. Cannatella for writing the program for calculating homoplasy and to J. Huelsenbeck for use of his program for randomizing data matrices and for random numbers. D. Cannatella, P. Chippindale, K. de Queiroz, D. Hillis, and J. McGuire made useful comments on the manuscript. Financial support for some of the research was provided by an NSF Graduate Fellowship.

#### REFERENCES

- ARCHIE, J. W. 1985. Methods for coding variable morphological features for numerical taxonomic analysis. *Syst. Zool.* 34:326-345.
- ARCHIE, J. W. 1989. Homoplasy excess ratios: New indices for measuring levels of homoplasy in phylogenetic systematics and a critique of the consistency index. *Syst. Zool.* 38:253-269.
- BAKER, A. J. 1980. Morphometric differentiation in New Zealand populations of the house sparrow (*Passer domesticus*). *Evolution* 34:638-653.
- BAKER, A. J., AND A. MOEED. 1979. Evolution in the introduced New Zealand population of the common myna, *Acridotheres tristis* (Aves: Sturnidae). *Can. J. Zool.* 57:570-584.
- BULL, J. J., J. P. HUELSENBECK, C. W. CUNNINGHAM, D. L. SWOFFORD, AND P. J. WADDELL. 1993. Partitioning and combining data in phylogenetic analysis. *Syst. Biol.* 42:384-397.
- BUTH, D. G. 1984. The application of electrophoretic data in systematic studies. *Annu. Rev. Ecol. Syst.* 15: 501-522.
- CAMPBELL, J. A., AND D. R. FROST. 1993. Anguid lizards of the genus *Abronja*: Revisionary notes, description of four new species, a phylogenetic analysis, and key. *Bull. Am. Mus. Nat. Hist.* 216:1-121.
- CANNATELLA, D. C., AND K. DE QUEIROZ. 1989. Phylogenetic systematics of the anoles: Is a new taxonomy warranted? *Syst. Zool.* 38:57-69.
- CARPENTER, J. M., J. E. STRASSMAN, S. TURILLAZZI, C. R. HUGHES, C. R. SOLÍS, AND R. CERVO. 1993. Phylogenetic relationships among paper wasp social parasites and their hosts (Hymenoptera: Vespidae: Polistinae). *Cladistics* 9:129-146.
- CHIPPINDALE, P. T., AND J. J. WIENS. 1994. Weighting,

- partitioning, and combining characters in phylogenetic analysis. *Syst. Biol.* 43:278–287.
- COLLESS, D. H. 1980. Congruence between morphometric and allozyme data for *Menidia* species: A reappraisal. *Syst. Zool.* 29:288–299.
- COOK, J. A., K. R. BESTGEN, D. L. PROPST, AND T. L. YATES. 1992. Allozyme divergence and systematics of the Rio Grande Silvery Minnow, *Hybognathus amarus* (Teleostei: Cyprinidae). *Copeia* 1992:36–44.
- CROTHER, B. I. 1990. Is "some better than none" or do allele frequencies contain phylogenetically useful information? *Cladistics* 6:277–281.
- DARWIN, C. 1859. On the origin of species. Harvard Univ. Press, Cambridge, Massachusetts.
- DE QUEIROZ, K. 1992. Phylogenetic relationships and rates of allozyme evolution among the lineages of sceloporine sand lizards. *Biol. J. Linn. Soc.* 45:333–362.
- DONOGHUE, M. J., R. G. OLMSTEAD, J. F. SMITH, AND J. D. PALMER. 1992. Phylogenetic relationships of Dipsacales based on *rbcL* sequences. *Ann. Mo. Bot. Gard.* 79:333–345.
- FAITH, D. P. 1991. Cladistic permutation tests for monophyly and nonmonophyly. *Syst. Zool.* 40:366–375.
- FARRIS, J. S. 1966. Estimation of conservatism of characters by constancy within biological populations. *Evolution* 20:587–591.
- FARRIS, J. S. 1969. A successive approximations approach to character weighting. *Syst. Zool.* 18:374–385.
- FARRIS, J. S. 1981. Distance data in phylogenetic analysis. Pages 3–23 in *Advances in cladistics*, Volume 1 (V. A. Funk and D. R. Brooks, eds.). New York Botanical Garden, New York.
- FELSENSTEIN, J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* 39:783–791.
- GAINES, M. S., L. R. MCCLENAGHAN, JR., AND R. K. ROSE. 1978. Temporal patterns of allozyme variation in fluctuating populations of *Microtus ochragaster*. *Evolution* 32:723–739.
- GREEN, D. M. 1986. Systematics and evolution of western North American frogs allied to *Rana aurora* and *Rana boylei*: Electrophoretic evidence. *Syst. Zool.* 35:283–296.
- GUYER, C., AND J. M. SAVAGE. 1992. Anole systematics revisited. *Syst. Biol.* 41:89–110.
- HILLIS, D. M. 1991. Discriminating between phylogenetic signal and random noise in DNA sequences. Pages 278–294 in *Phylogenetic analysis of DNA sequences* (M. M. Miyamoto and J. Cracraft, eds.). Oxford Univ. Press, New York.
- HILLIS, D. M., AND J. J. BULL. 1993. An empirical test of bootstrapping as a method for assessing confidence in phylogenetic analysis. *Syst. Biol.* 42:182–192.
- HILLIS, D. M., AND J. P. HUELSENBECK. 1992. Signal, noise, and reliability in molecular phylogenetic analyses. *J. Hered.* 83:189–195.
- HILLIS, D. M., J. P. HUELSENBECK, AND C. W. CUNNINGHAM. 1994. Application and accuracy of molecular phylogenies. *Science* 264:671–677.
- HUELSENBECK, J. P. 1991. Tree-length distribution skewness: An indicator of phylogenetic information. *Syst. Zool.* 40:257–270.
- HUELSENBECK, J. P., AND D. M. HILLIS. 1993. Success of phylogenetic methods in the four-taxon case. *Syst. Biol.* 42:247–264.
- JOHNSON, N. K., J. A. MARTEN, AND C. J. RALPH. 1989. Genetic evidence for the origin and relationships of Hawaiian honeycreepers (Aves: Fringillidae). *Condor* 91:379–396.
- JOHNSON, N. K., R. M. ZINK, AND J. A. MARTEN. 1988. Genetic evidence for relationships in the avian family Vireonidae. *Condor* 90:428–445.
- KLUGE, A. G., AND J. S. FARRIS. 1969. Quantitative phyletics and the evolution of anurans. *Syst. Zool.* 18:1–32.
- KLUGE, A. G., AND W. C. KERFOOT. 1973. The predictability and regularity of character divergence. *Am. Nat.* 107:426–442.
- MABEE, P. M., AND J. HUMPHRIES. 1993. Coding polymorphic data: Examples from allozymes and ontogeny. *Syst. Biol.* 42:166–181.
- MAYR, E. 1969. Principles of systematic zoology. McGraw-Hill, New York.
- MICKEVICH, M. F., AND M. F. JOHNSON. 1976. Congruence between morphological and allozyme data in evolutionary inference and character evolution. *Syst. Zool.* 25:260–270.
- MICKEVICH, M. F., AND C. MITTER. 1981. Treating polymorphic characters in systematics: A phylogenetic treatment of electrophoretic data. Pages 45–58 in *Advances in cladistics*, Volume 1 (V. A. Funk and D. R. Brooks, eds.). New York Botanical Garden, New York.
- MICKEVICH, M. F., AND C. MITTER. 1983. Evolutionary patterns in allozyme data: A systematic approach. Pages 169–176 in *Advances in cladistics*, Volume 2 (N. Platnick and V. A. Funk, eds.). Columbia Univ. Press, New York.
- MONTANUCCI, R. R., R. W. AXTELL, AND H. C. DESAUER. 1975. Evolutionary divergence among colored lizards (*Crotaphytus*), with comments on the status of *Gambelia*. *Herpetologica* 31:336–347.
- MURPHY, R. W. 1993. The phylogenetic analysis of allozyme data: Invalidity of coding alleles by presence/absence and recommended procedures. *Biochem. Syst. Ecol.* 21:25–38.
- PIERCE, B. A., AND J. B. MITTON. 1979. A relationship of genetic variation within and among populations: An extension of the Kluge-Kerfoot phenomenon. *Syst. Zool.* 28:63–70.
- PIMENTEL, R. A., AND R. RIGGINS. 1987. The nature of cladistic data. *Cladistics* 3:201–209.
- PROBER, S., J. C. BELL, AND G. MORAN. 1990. A phylogenetic and allozyme approach to understanding rarity in the "green ash" eucalypts (Myrtaceae). *Plant Syst. Evol.* 172:99–118.
- REEDER, T. W., AND J. J. WIENS. 1996. Evolution of the lizard family Phrynosomatidae as inferred from diverse types of data. *Herpetol. Monogr.* (in press).
- ROHLF, F. J., A. J. GILMARTIN, AND G. HART. 1983. The Kluge-Kerfoot phenomenon—A statistical artifact. *Evolution* 37:180–202.

- SIMPSON, G. G. 1961. Principles of animal taxonomy. Columbia Univ. Press, New York.
- SOKAL, R. R. 1976. The Kluge-Kerfoot phenomenon reexamined. *Am. Nat.* 110:1077-1091.
- SWOFFORD, D. L. 1993. PAUP: Phylogenetic analysis using parsimony, version 3.1. Illinois Natural History Survey, Champaign.
- SWOFFORD, D. L., AND S. H. BERLOCHER. 1987. Inferring evolutionary trees from gene frequency data under the principle of maximum parsimony. *Syst. Zool.* 36:293-325.
- SWOFFORD, D. L., AND G. J. OLSEN. 1990. Phylogeny reconstruction. Pages 411-501 in *Molecular systematics* (D. M. Hillis and C. Moritz, eds.). Sinauer, Sunderland, Massachusetts.
- THIELE, K. 1993. The holy grail of the perfect character: The cladistic treatment of morphometric data. *Cladistics* 9:275-304.
- WHEELER, W. C. 1992. Extinction, sampling and molecular phylogenetics. Pages 205-215 in *Extinction and phylogeny* (M. Novacek and Q. Wheeler, eds.). Columbia Univ. Press, New York.
- WIENS, J. A., AND J. T. ROTTENBERRY. 1980. Patterns of morphology and ecology in grassland and shrub-steppe bird populations. *Ecol. Monogr.* 50:287-308.
- WIENS, J. J. 1993. Phylogenetic systematics of the tree lizards (genus *Urosaurus*). *Herpetologica* 49:399-420.
- WILEY, E. O. 1981. *Phylogenetics. The theory and practice of phylogenetic systematics.* John Wiley and Sons, New York.

Received 22 August 1994; accepted 9 March 1995

APPENDIX 1. Raw scores for the number of characters treated as phylogenetically informative by the eight methods for phylogenetic analysis of polymorphic characters. The any-instance and unscaled methods could not be applied to the allozyme data sets.

Method	Data sets						
	Phrynosomatidae	<i>Urosaurus</i>	<i>Crotaphytus</i>	<i>Hybognathus</i>	<i>Rana</i>	<i>Rhagoletis</i>	Sand lizards
Any-instance	115	20					
Frequency	115	24	10	14	24	9	19
Majority	112	17	5	9	17	4	15
Missing	102	9	2	4	9	1	10
Polymorphic	102	9	8	10	20	5	15
Scaled	115	24	9	14	24	8	19
Unordered	114	24	6	7	16	1	14
Unscaled	115	24					

APPENDIX 2. Raw scores for the number of equally parsimonious trees generated by the eight methods for phylogenetic analysis of polymorphic characters. The any-instance and unscaled methods could not be applied to the allozyme data sets.

Method	Data sets						
	Phrynosomatidae	<i>Urosaurus</i>	<i>Crotaphytus</i>	<i>Hybognathus</i>	<i>Rana</i>	<i>Rhagoletis</i>	Sand lizards
Any-instance	130	1					
Frequency	3	1	1	1	1	1	1
Majority	>20,000	5	10	3	4	6	5
Missing	>20,000	489	24	21	3	15	21
Polymorphic	>20,000	489	3	5	1	12	1
Scaled	80	4	2	2	1	1	2
Unordered	4,065	6	4	1	9	15	57
Unscaled	256	4					

APPENDIX 3. Raw scores for phylogenetic signal used to compare eight methods for phylogenetic analysis of polymorphic characters. Phylogenetic signal was measured as the observed  $g_1$  minus the expected critical value for the data set for the method when randomized. Negative values indicate significant phylogenetic signal, and values of  $\geq 0$  indicate that the phylogenetic information is indistinguishable from random data. The any-instance and unscaled methods could not be applied to the allozyme data sets.

Method	Data sets						
	Phrynosomatidae	<i>Urosaurus</i>	<i>Crotaphytus</i>	<i>Hybognathus</i>	<i>Rana</i>	<i>Rhagoletis</i>	Sand lizards
Any-instance	-0.720	+0.015					
Frequency	-0.917	-0.512	-0.070	-1.165	-0.232	+0.165	-0.579
Majority	-0.829	-0.424	-0.208	-1.056	-0.202	+0.330	-0.626
Missing	-1.088	-0.391	-0.033	-0.482	-0.289	+0.007	-0.455
Polymorphic	-1.088	-0.391	-0.101	-0.964	-0.334	+0.286	-0.444
Scaled	-1.030	-0.254	+0.006	-0.884	-0.291	+0.038	-0.490
Unordered	-0.787	+0.009	-0.114	-0.369	-0.379	+0.067	-0.539
Unscaled	-0.708	-0.181					

APPENDIX 4. Raw scores\* for the number of clades supported by bootstrapping (with frequency of the clade  $\geq 70\%$  among 100 replicates) for six data sets using eight methods for phylogenetic analysis of polymorphic characters. Bootstrapping could not be applied to the phrynosomatid data set because of the large number of taxa. The any-instance and unscaled methods could not be applied to the allozyme data sets.

Method	Data sets					
	<i>Urosaurus</i>	<i>Crotaphytus</i>	<i>Hybognathus</i>	<i>Rana</i>	<i>Rhagoletis</i>	Sand lizards
Any-instance	0/7					
Frequency	2/7	1/5	1/5	0/6	0/4	6/6
Majority	4/7	1/5	3/5	0/6	0/4	3/6
Missing	1/7	0/5	0/5	2/6	1/4	1/6
Polymorphic	1/7	0/5	1/5	2/6	0/4	4/6
Scaled	1/7	0/5	1/5	3/6	1/4	1/6
Unordered	0/7	0/5	0/5	2/6	0/4	0/6
Unscaled	1/7					

\* Number of clades supported/maximum number clades possible (number of taxa minus 2).

APPENDIX 5. Similarity between trees from the original *Urosaurus* data set and trees based on 20 replicated matrices with reduced sample sizes ( $n \leq 5$ ) for six methods for phylogenetic analysis of polymorphic characters (see also Fig. 3). The missing and polymorphic methods were not included because of the large number of shortest trees produced by the original matrix (489).

Method	Similarity	
	$\bar{x} \pm SD$	Range
Any-instance	0.134 $\pm$ 0.086	0-0.286
Frequency	0.861 $\pm$ 0.124	0.571-1.000
Majority	0.709 $\pm$ 0.042	0.642-0.771
Scaled	0.334 $\pm$ 0.100	0.143-0.536
Unordered	0.196 $\pm$ 0.068	0.072-0.372
Unscaled	0.291 $\pm$ 0.087	0.143-0.476