Syst. Biol. 47(3): 397–413, 1998

The Accuracy of Methods for Coding and Sampling Higher-Level Taxa for Phylogenetic Analysis: A Simulation Study

JOHN J. WIENS

Section of Amphibians and Reptiles, Carnegie Museum of Natural History, Pittsburgh, Pennsylvania 15213-4080, USA; E-mail: wiensj@clpgh.org

Abstract.— Many phylogenetic analyses, particularly morphological studies, use higher taxa (e.g., genera, families) rather than species as terminal taxa. This general approach requires dealing with interspecific variation among the species that make up the higher taxon. In this paper, I review different parsimony methods for coding and sampling higher taxa and compare their relative accuracies using computer simulations. Despite their widespread use, methods that involve coding higher taxa as terminals perform poorly in simulations, relative to splitting up the higher taxa and using species as terminals. Among the methods that use higher taxa as terminals, coding a taxon based on the most common condition among the included species (majority or modal coding) is generally more accurate than other coding methods, such as coding taxa as missing or polymorphic. The success of the majority method, and results of further simulations, suggest that in many cases "common equals primitive" within variable taxa, at least for low and intermediate rates of character change. The fixed-only method (excluding variable characters) performs very poorly, a result that is indirectly supported by analyses of published data for squamate reptiles. Sampling only a single species per higher taxon also yields low accuracy under many conditions. Along with recent studies of intraspecific polymorphism, the results of this study show the general importance of (1) including characters despite variation within taxa and (2) using methods that incorporate detailed information on the distribution of states within variable taxa. [Accuracy; coding methods; parsimony; simulations; squamata; taxon sampling.]

Species may be the basic units of evolution and classification, but they are often not the basic units of phylogenetic analysis. Many phylogenetic studies, particularly morphological analyses, deal with the relationships of higher taxa and use supraspecific taxa as their terminal units (e.g., Gauthier et al., 1988; Rowe, 1988; Trueb and Cloutier, 1991; Eernisse et al., 1992; Novacek, 1992; Schultze, 1994; Carlson, 1995; Livezey, 1996, 1997; Smith, 1996; Whiting et al., 1997). Using genera, families, or other higher taxa as terminals may be a useful way to analyze relationships among speciose groups. However, using higher-level terminals often requires dealing with variation among the species that make up the terminal taxa. The question of how to deal with this variation not only involves coding, but also is intimately related to the issue of taxon sampling.

Interspecific variation within a higher taxon is sometimes referred to as polymorphism (e.g., Nixon and Davis, 1991; Donoghue, 1994). However, interspecific variation is fundamentally different from polymorphism within species or populations. Intraspecific polymorphisms evolve via population-genetic processes and can be shared between species through common ancestry, whereas shared interspecific "polymorphisms" are generally due to homoplasy or nonmonophyly of the higher taxa (de Queiroz, 1987).

Systematists use a variety of methods for dealing with interspecific variation within higher-level terminal taxa. These include the practices of: sampling a single species per higher taxon, coding inferred ancestral states, excluding the variable characters, coding variable taxa as polymorphic, or dividing the variable taxa into smaller taxonomic units. The choice among these methods is important, because the application of different methods to the same empirical data set can give different trees (Fig. 1). Because only one phylogeny can be true, this observation suggests that many of these methods must give incorrect estimates of the phylogeny, at least for some data sets.

Computer simulations are an important tool for choosing among phylogenetic methods, because they provide a context in which

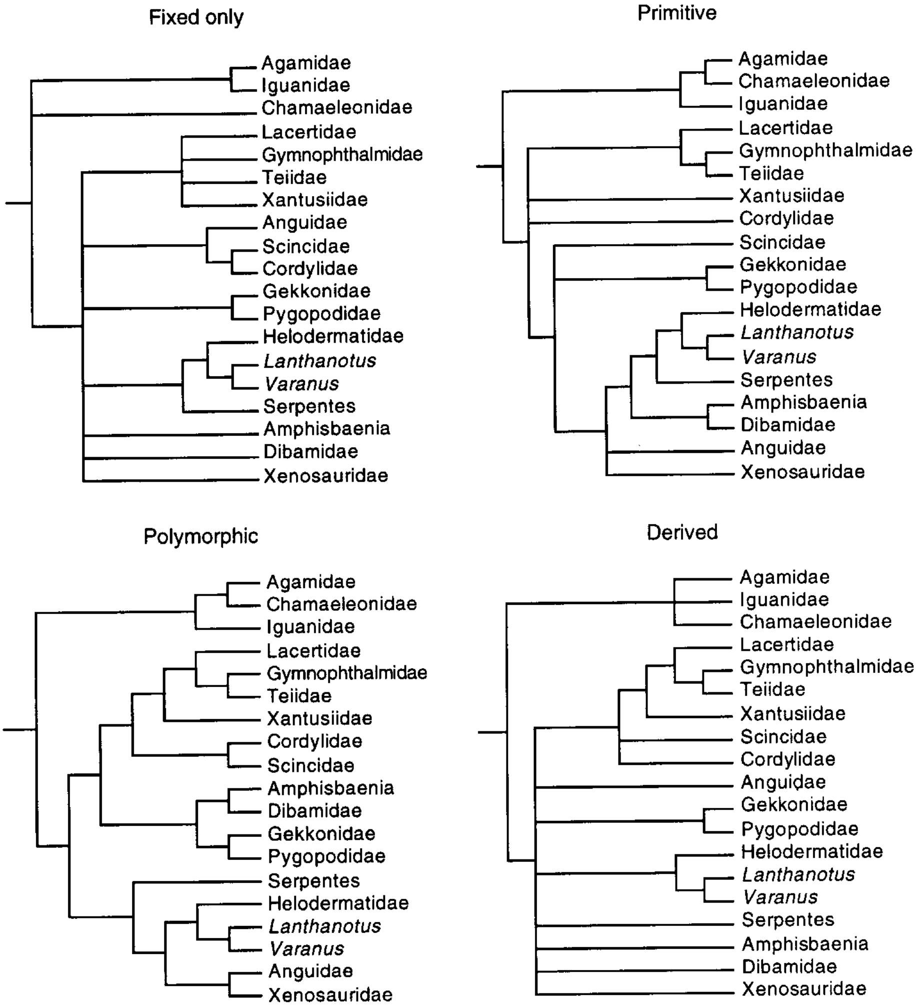

FIGURE 1. Different methods for treating interspecific variation give different trees for the same data. Trees were produced by different parsimony methods for coding interspecific variation with the data of Estes et al. (1988) for families and other higher taxa of squamate reptiles. Trees are either the strict consensus of multiple equally parsimonious trees or a single shortest tree. The missing and polymorphic coding methods give nearly identical results for these data, so only the tree from the polymorphic method is shown. Because Estes et al. (1988) did not provide data for individual species, some methods (majority, type species, species-as-terminals) could not be applied.

the true phylogeny is known. Although simulated data sets never capture the complexity of real data produced by natural processes, the simplicity of simulated conditions allows one to manipulate systematically and understand the parameters that affect the accuracy of phylogenetic methods. Insights gained from simulations can then be used to predict how methods may behave in the real world, where the phylogeny is unknown.

Some previous studies have provided useful discussions of the pros and cons of different methods for dealing with interspecific variation in higher-level taxa (e.g., de Queiroz, 1987; Estes et al., 1988; Nixon and Davis, 1991; Donoghue, 1994; Mishler, 1994; Yeates, 1995; Rice et al., 1997). However, these studies did not address the relative accuracy of these methods (their ability to recover the true phylogeny). In the present paper, I review proposed methods for treating interspecific variation and compare their accuracy using simulations.

METHODS FOR TREATING INTERSPECIFIC VARIATION IN HIGHER TAXA

Fixed only.—Given the scarcity with which interspecific variation in higher taxa is reported in systematic studies, the practice of excluding variable characters appears to be common. A survey of morphological phylogenetic studies published in 12 journals from 1986 to 1995 confirms that variability within higher taxa is one of the most common criteria for excluding characters (Wiens, unpubl. data). This practice may have its basis in the idea that characters that vary within a terminal taxon are likely to be homoplastic between terminal taxa as well (Kluge and Farris, 1969). A relationship between homoplasy and variability has been found in studies of intraspecific polymorphism (Wiens, 1995), but has not been tested for interspecifically variable characters in higher taxa. De Queiroz (1987) and Estes et al. (1988) rejected the practice of discarding variable characters as a general solution because it requires ignoring potentially informative data.

Splitting taxa.—The practice of splitting up a variable higher taxon into smaller, presumably monomorphic units (e.g., species) has a number of advantages. The first is that it avoids making assumptions about the monophyly of the variable taxon (Nixon and Davis, 1991). Monophyly of the higher taxon is a serious concern, because interspecific variation is a priori evidence that the variable taxon may not be monophyletic, if the derived state occurs in other higher taxa besides the variable one (de Queiroz, 1987). The practice of splitting up higher taxa also avoids arbitrary coding of variable terminals (Nixon and Davis, 1991; Yeates, 1995). Some authors have argued that this approach may be impractical in some cases, especially if it leads to analyzing hundreds of terminal taxa simultaneously (e.g., Donoghue, 1994; Mishler, 1994; Rice et al., 1997). On the other hand, a higher taxon can be represented by a more limited sample of species or exemplars (Yeates, 1995).

Kluge and Farris (1969) recommended splitting up a variable taxon into two or more dummy taxa, each invariant for one of the states for a given character. This method has been criticized because it may require a new taxon for each instance of interspecific variation in each character (de Queiroz, 1987; Estes et al., 1988). Furthermore, these dummy taxa may not correspond to monophyletic groups, especially given that at least one of the dummy taxa is likely to be defined based on a primitive state (Donoghue, 1994).

The next six methods involve assigning a single character state to a variable higher taxon. All of these methods assume that the higher taxon is monophyletic.

Inferring the ancestral state from the phylogeny within the higher taxon (IAS).—In theory, the goal of coding higher-level terminal taxa is to represent the character states present in the ancestral species of the higher taxon. Many authors have used information on the phylogeny within the higher taxon (if available) to estimate and code the ancestral state (e.g., Doyle and Donoghue, 1986; Carpenter, 1987; de Queiroz, 1987; Estes et al., 1988; Gauthier et al., 1988; Rowe, 1988; Frost and Etheridge, 1989; Trueb and Cloutier, 1991; Wiens, 1993; Schultze, 1994; Livezey, 1996). The phylogenetic information used comes from previous or independent studies within the higher taxon. This traditional approach (Yeates, 1995) has recently received a number of different names, including the placeholder approach (Donoghue, 1994), compartmentalization (Mishler, 1994), intuitive groundplan method (Yeates, 1995), and inferred ancestral states (IAS; Rice et al., 1997).

In many cases, information on the phylogeny within the variable terminal taxon is questionable or unavailable, or optimization of the ancestral state yields ambiguous results. At least five methods have been proposed for these cases.

Primitive state.—One method (primitive state) involves coding the variable taxon with the plesiomorphic state determined in a higher-level outgroup analysis (e.g., given that genera make up the ingroup, then the state present in the outgroups of these genera). It should be understood that the primitive state method is distinct from the IAS method. The former is based on outgroup analysis, whereas the latter is based on the phylogeny within the variable terminal taxon. In justifying their use of the primitive state method, Estes et al. (1988) argued that it would avoid circularity, because the derived state determined by outgroup analysis would be plesiomorphic for the terminal taxon only if certain relationships were obtained among these terminals. However, this method assumes that the derived state within a variable taxon is convergent with the derived condition found in other terminal taxa, and conversely, that the presence of the plesiomorphic state is not due to reversal (de Queiroz, 1987; Estes et al., 1988). Kluge (1989b) criticized this method because there is no empirical evidence to suggest such a predominance of one type of homoplasy over another, and cited data from a study (Kluge, 1989a) in which the observed ratio of convergences to reversals was nearly 50:50.

Derived state.— This method is merely the opposite of the preceding method; rather than coding the variable taxon with the primitive state determined from the higher-level outgroup analysis, one codes it with the derived state. This method is included for the sake of completeness, though I am unaware of empirical studies in which it has been used.

Type species/single species.— Walker et al. (1990) advocated coding a higher taxon based on the taxon's name-bearing type. Although there is no phylogenetic rationale for this method, it may have heuristic value if the monophyly of the terminal taxon is in doubt and the assignment of species to the higher taxon is likely to change (Yeates, 1995). Many phylogenetic studies (particularly molecular ones) deal with interspecific variation in a comparable way, by sampling only a single species or exemplar from each higher taxon.

Majority.— Some authors have coded variable higher taxa based on the modal condition among the species (e.g., Livezey, 1986) or a "consensus" (Trueb and Cloutier, 1991). Although these authors did not provide a justification, the method rests implicitly on an assumption of "common equals primitive," which has been criticized (e.g., Watrous and Wheeler, 1981; Wiley, 1981). Despite the criticism, the majority method does incorporate at least some information on the distribution of variation within the higher taxa.

Polymorphic/missing.— A frequently used set of methods for dealing with interspecific variation is to code the variable higher taxon as being polymorphic (having both states) or unknown (missing), particularly when information on the phylogeny within the variable taxon is absent or gives ambiguous reconstructions (e.g., Doyle and Donoghue, 1986; Frost and Etheridge, 1989; Wiens, 1993; Livezey, 1996). The main disadvantage of these methods is that missing or polymorphic cells in the data matrix are largely uninformative in reconstructing the tree. A taxon coded as missing or polymorphic is treated as having the state that is most parsimonious, given the position of the taxon on the tree, as determined by other characters. The two methods differ only in that polymorphic cells are treated as if either of the observed states is a possible assignment to the polymorphic taxon, whereas missing cells are treated as if any state is possible. Thus, the placement of the variable taxon is constrained somewhat by the observed states when using the polymorphic method, at least for multistate characters. The missing and polymorphic methods give identical results for binary characters. Nixon and Davis (1991) used a hypothetical data matrix to show that coding variable higher taxa as missing led to trees that were inconsistent with those based on scoring species as terminals. Although these authors considered the differences in tree topology to be errors on the part of the missing method, their study did not address which of the trees was correct.

MATERIALS AND METHODS

Simulations

Computer simulations were used to compare the accuracy of the proposed methods. Two sets of model trees were used (Fig. 2). For the first, a 42-species tree was simulated, consisting of six higher taxa with seven species each. The ability of different methods to recover the correct unrooted tree of the six higher taxa was tested. An asymmetric tree was chosen for the

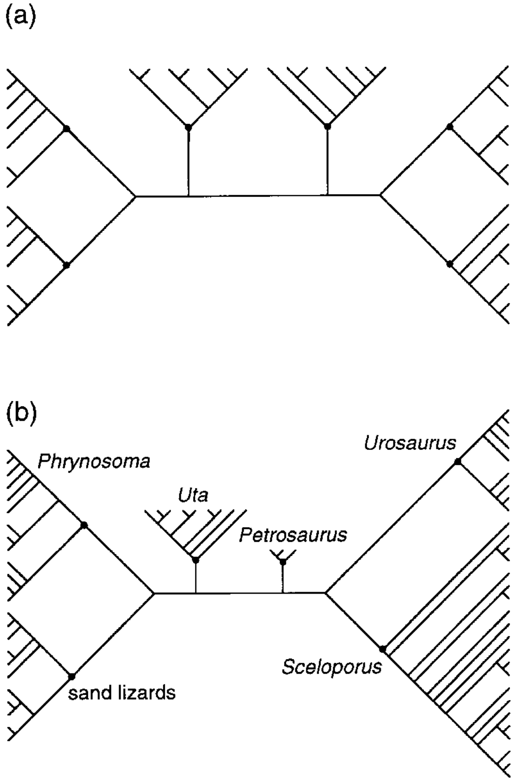

FIGURE 2. Model trees used in simulations. (a) Sample 42-species tree. Relationships among the seven species within each of the six higher taxa are chosen randomly in each replicate; some of the most common tree shapes are illustrated here. (b) The 66-species tree based on hypothesized phylogenies of the lizard family Phrynosomatidae.

relationships among the higher taxa; an asymmetric tree shape is more likely than the more-symmetric tree, given a model where speciation is equally likely to occur on any part of a growing phylogeny (Harding, 1971; Slowinski and Guyer, 1989). Furthermore, tree shape should have little influence on the performance of methods, given the small, unrooted tree (the symmetric and asymmetric topologies are nearly identical). Rooted topologies relating the seven species within each higher taxon were chosen randomly for each higher taxon in each replicate. The probability of selection for each of the possible topologies (or tree shapes) for seven taxa was based on a Markovian model with an equal probability of speciation along any branch (Harding, 1971; Slowinski and Guyer, 1989). The number of species and higher taxa was chosen somewhat arbitrarily, but was intended as a compromise between numbers that are realistically large and computationally tractable.

The second model tree was taken from empirical studies (Fig. 2). An empirically derived tree is advantageous in that it has different numbers of species within each higher taxon (as do most real data sets) and complex trees at the species level. The estimated phylogeny of phrynosomatid lizards was used as if it represented the true phylogeny. The accuracy of this estimated tree is not critical to the simulation results, however, because in the simulations the true phylogeny is known. Six higher taxa were used (*Phrynosoma*, the sand lizard clade, *Uta*, *Petrosaurus*, *Urosaurus*, *Sceloporus*). The asymmetric unrooted topology was based on Reeder and Wiens (1996, combined analysis). Species-level relationships within these clades were based on the following sources: *Phrynosoma* (Montanucci, 1987), sand lizards (*Uma*, *Callisaurus*, *Cophosaurus*, *Holbrookia*; de Queiroz, 1989, combined analysis), *Uta* (Ballinger and Tinkle, 1972; their fig. 13), *Urosaurus* (Reeder and Wiens, 1996; combined analysis), and *Sceloporus* (Wiens and Reeder, 1997, combined analysis with all taxa). Because an odd number of taxa in each clade was desired (to eliminate ties when coding the majority method), some minor modifications to these phylogenies were made (the polytypic taxa *Phrynosoma douglasi*, *Cophosaurus texanus*, and *Petrosaurus thalassinus* were each split into two species). For computational simplicity, the tree for *Sceloporus* included only one species from each species group (so that only 22 species were represented rather than 80), and one species group was removed to ensure an odd number of taxa. Thus, the second model tree consisted of six higher taxa and 66 species, with different (odd) numbers of species within each higher taxon. The number of species ranged from 3 to 21 per higher taxon.

The model of character evolution used was extremely simple. All characters were binary, both for simplicity and because the majority of morphological characters in empirical studies are described with only two states. Each character began its evolution with the state 0, and the branch length was considered to be the probability of a change occurring by the end of the branch (to state 1). Gains and losses were assumed to be equally likely.

Branch lengths were varied in three ways. First, all lengths were held constant across all characters and all branches of the tree. This assumes an extreme punctuated model of change (as opposed to having divergence increase linearly with time), but allows the effects of a given branch length to be tested. Six different lengths were tested (0.005, 0.01, 0.05, 0.10, 0.15, 0.20). At the longest branch length (0.20), all methods perform very poorly because of high levels of homoplasy. With lengths shorter than 0.005, performance decreases because many characters are invariant. Thus, the six lengths tested include a broad range of lengths over which phylogenies can be reconstructed accurately under this model, given a finite number of characters. In a second set of analyses, branch lengths were varied randomly among all lineages (species and higher taxa) but held constant among characters. In the third set of analyses, rates of change (i.e., branch lengths) were varied randomly among characters but held constant across lineages. For computational simplicity, the third set of analyses (rates of change varying among characters) was applied only to the tree with an invariant topology (the 66-taxon phrynosomatid tree). Randomly selected branch lengths ranged from 0 to 0.10 and from 0 to 0.20. The results from these lengths are very similar to those obtained when using equal branch lengths at the midpoint of these ranges (0.05, 0.10). Although other ranges could have been explored, using a longer maximum length would likely cause all methods to perform very poorly, and using a lower maximum length would certainly produce results similar to those for low-equal branch lengths. For each set of branch length conditions examined, three different numbers of characters were used (100, 200, and 400).

The accuracy of almost all of the proposed methods was tested, including fixed-only, splitting taxa (henceforth referred to as the species-as-terminals method), and the majority, missing, primitive state, and derived state coding methods. The effects of scoring a higher taxon based on a single randomly chosen species were also tested; this is analytically equivalent to the "type species" method. For the fixed-only method, a given character was excluded if there was any variation among the species within any of the six higher taxa. Thus, the fixed-only method used a smaller number of characters than other methods. For the species-as-terminals method, trees were constrained so that higher taxa would be monophyletic; this allowed direct comparison of the accuracy of this method to the other coding methods, which also constrain monophyly. For the primitive and derived methods, polymorphic higher taxa were coded with the known primitive (0) or derived (1) state.

The IAS coding method, which uses an ancestral state inferred for a higher taxon based on a priori information on the phylogeny within that taxon, was not included. This approach would be difficult to model realistically, because in the real world this a priori phylogenetic information varies in quality and quantity from taxon to taxon. The species-as-terminals method might be considered similar to this approach, but the only phylogenetic information assumed is the monophyly of the terminal taxon, and the ancestral state for the higher taxon is inferred through a global, simultaneous analysis rather than an a priori analysis. Also, the "dummy taxa" approach (Kluge and Farris, 1969) was excluded because it is not widely used and is not practical to apply to real or simulated data sets unless levels of variation are extremely low (see above).

For each set of conditions (i.e., model tree, branch length, number of characters), 100 replicated matrices were simulated. The accuracy of a method was the similarity between the estimated phylogeny (or the strict consensus of multiple equally parsimonious estimates) and the true phylogeny, averaged across the 100 replicates. Similarity was measured as the proportion of nodes in common between the true and estimated trees, using the consensus fork index of Colless (1980). Given that results were very similar for closely matched simulated conditions (and there is little random variation in method performance), 100 replicates appears to be adequate. Results obtained using an alternative measure of accuracy are discussed later (see "Robustness of Results to Changes in the Model and Methods").

Trees were estimated with parsimony, using PAUP* (provided by David Swofford), versions 4.0 d53–4.0 d55. All methods except the species-as-terminals method used only six taxa. The small number of taxa made it possible to use the branch-and-bound search option, which guarantees finding the shortest tree. For the species-as-terminals method, which used either 42 or 66 taxa, the heuristic search option was used, with TBR branch swapping and 20 random-addition sequences. The programs for simulating and coding the data were written in C by the author.

Limited Taxon Sampling

It was assumed in the preceding simulations that all species within a higher taxon were sampled (except for the type species method). In the real world, sampling all the species within a given higher taxon may not be possible, and a few exemplar species are often used to represent higher groups. To address whether incomplete sampling of species within the higher taxa might affect the relative performance of methods, I ran a limited set of analyses in which only three species were sampled (randomly) from each of the higher taxa.

Common Equals Primitive

The majority method implicitly assumes that within a variable higher taxon the common state will be primitive. To test the assumption more explicitly, I simulated phylogenies with seven species each and with 200 binary characters evolving at various rates (branch lengths), and recorded how often for each variable character the commonest state was the known primitive condition. This was done for each of the 11 possible rooted tree shapes for seven taxa, and was then repeated for the same lengths for a 65-species tree for phrynosomatids (one species was deleted to eliminate ties).

Excluding Interspecifically Variable Characters

The fixed-only method implicitly assumes that characters that vary within higher taxa will be less reliable for inferring the relationships between these taxa. Regardless of the simulation results, this approach might be justified if, in real data, (1) characters that vary within terminal taxa are so noisy that they do not contain any phylogenetic information, and/or (2) there is a consistent, positive relationship between levels of interspecific variability in higher taxa and homoplasy. To test these hypotheses, I analyzed the morphological data of Estes et al. (1988) for families (and other higher taxa) of squamate reptiles. This data set is unusual in that the authors explicitly avoided excluding characters due to interspecific variation, and 114 of the 148 characters that they analyzed vary within one or more of their terminal taxa. Using these data, I compared the levels of phylogenetic signal in the fixed and variable characters (relative to randomized data) according to the g_1 index (Hillis, 1991) and examined the relationship between homoplasy and variability through use of the Spearman rank correlation (following Wiens, 1995). I also assessed qualitatively if the fixed-only method was able to recover traditionally recognized clades of squamates (Estes et al., 1988), with the idea that failure to recover these groups might suggest failure of this method (although these clades are obviously not known to be correct). The clades were Iguania (Iguanidae, Agamidae, and Chamaeleonidae), Acrodonta (Agamidae and Chamaeleonidae), Gekkota (Gekkonidae and Pygopodidae), Anguimorpha (Anguidae, Xenosauridae, Helodermatidae, *Lanthanotus*, *Varanus*, and possibly Serpentes), Scincomorpha (Scincidae, Cordylidae, Xantusiidae, Lacertidae, Teiidae, and Gymnophthalmidae), and Scincoidea (Scincidae and Cordylidae).

RESULTS AND DISCUSSION

The simulation results (Figs. 3, 4) are very consistent across the different conditions examined. The general conclusions are as follows.

1. The fixed-only method (excluding variable characters) performs very poorly.
2. The species-as-terminals method is the most accurate under almost all conditions, and is often superior to the other methods by a large margin.
3. Among the methods that code higher taxa as terminals, the majority method generally performs best.

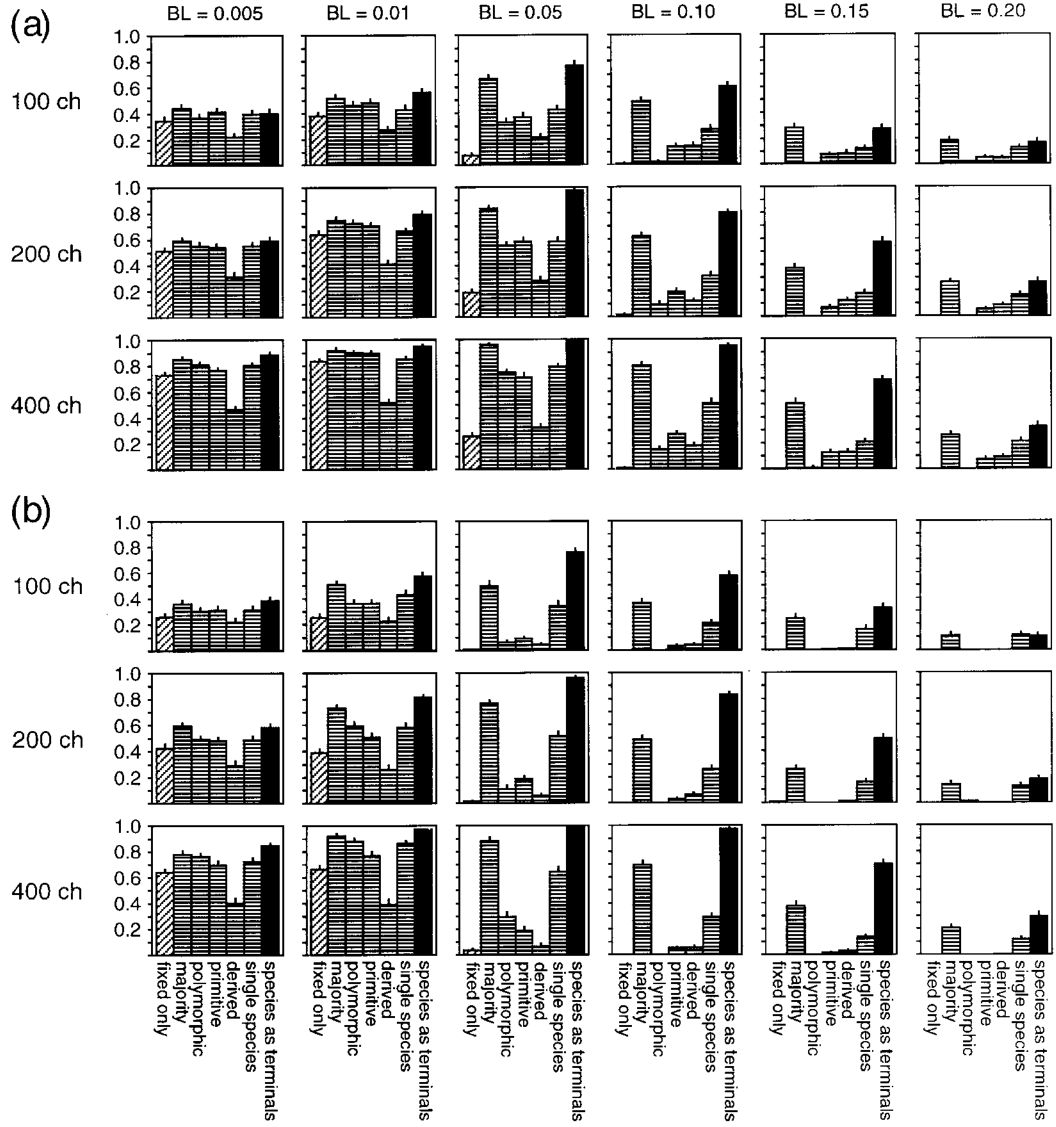

FIGURE 3. Accuracy of parsimony methods for analyzing interspecific variation, when branch lengths (BL) are invariant among lineages and characters (ch). The polymorphic and missing methods give identical results for binary characters. Each bar is the average accuracy from 100 replicated matrices, the line above each bar is the standard error. (a) 42 species. (b) 66 species.

In general, the methods show similar accuracy at the lowest branch lengths (when there is relatively little variation within higher taxa) and highest branch lengths (when all methods perform poorly), and the greatest differentiation is seen at intermediate branch lengths. Under conditions where levels of interspecific variation are high (Table 1), methods that exclude variable characters (fixed-only), render variable data cells uninformative (missing, polymorphic), or arbitrarily fill in variable data cells with either all 0's or all 1's (primitive state, derived state) perform very poorly. These five methods do not utilize any information on the distribution of character states within the higher taxa, and they treat all

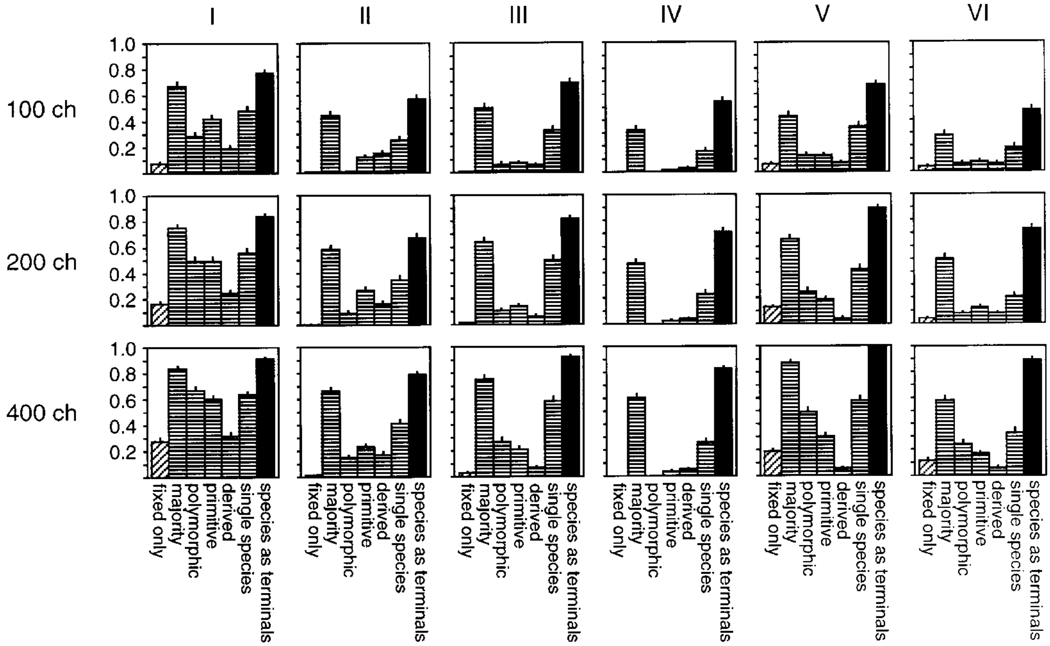

FIGURE 4. Accuracy of parsimony methods for analyzing interspecific variation when branch lengths vary among lineages and characters (ch). The polymorphic and missing methods give identical results for binary characters. Each bar is the average accuracy from 100 replicated matrices, the line above each bar is the standard error. I: 42 species, branch lengths vary randomly among lineages (range 0–0.10). II: 42 species, branch lengths vary randomly among lineages (0–0.20). III: 66 species, branch lengths vary randomly among lineages (0–0.10). IV: 66 species, branch lengths vary randomly among lineages (0–0.20). V: 66 species, rates of change (branch lengths) vary randomly among characters (0–0.10). VI: 66 species, rates of change vary randomly among characters (0–0.20).

instances of interspecific variation identically. Given this, it seems likely these methods will perform poorly under a much wider range of conditions than those simulated here, including cases where there are many more species within the higher taxa.

The strong performance of the species-as-terminals method is not entirely surprising (e.g., Nixon and Davis, 1991; Yeates, 1995). Within each variable higher taxon, the ancestral state is estimated in the course of a global parsimony analysis, and may be determined by both the estimated phylogeny within the higher taxon and the relationship of the higher taxon to others. Presumably, this provides a more consistently accurate estimate of the

TABLE 1. Levels of interspecific variability within higher taxa at different branch lengths in the simulated data matrices (based on a sample of 10 replicated matrices with 10 characters each for each set of conditions). A character was considered variable if there was any variation among the species within any of the six higher taxa. A data cell was considered variable if (for a given higher taxon and character) there was any variation among the species.

	42 species		66 species	
Branch length	Variable characters (%)	Variable data cells (%)	Variable characters (%)	Variable data cells (%)
0.005	24	4.3	46	9.2
0.01	53	11.0	53	14.2
0.05	98	43.5	100	57.3
0.10	100	75.2	100	79.6
0.15	100	83.8	100	86.7
0.20	100	91.5	100	90.8

ancestral state than does the "common equals primitive" assumption of the majority method. Unlike the species-as-terminals method, the majority method does not incorporate information on the state outside of the higher taxon in estimating the taxon's ancestral state, and uses only information from a single character at a time in coding the higher taxon.

Some authors have advocated coding higher taxa rather than species because including all the species as separate terminals will lead to huge matrices that are difficult or impossible to analyze effectively (e.g., Donoghue, 1994; Mishler, 1994; Rice et al., 1997). However, the results of the present study suggest that the loss of information inherent in coding higher taxa as terminals may greatly outweigh the benefits of being able to use more effective search strategies. Of course, this may depend somewhat on the number of taxa being analyzed. In this study, relatively superficial heuristic searches (20 addition sequence replicates) consistently estimated more accurate higher-level trees with 66 species than did branch-and-bound searches with only six taxa. Perhaps the problems of effectively searching for and accurately estimating trees with very large numbers of species are not as great as anticipated, especially with increasing computing power and search algorithm speed. For example, Hillis (1996) found that heuristic searches of simulated data sets with > 200 taxa could consistently achieve 100% accuracy when given a large sample of characters.

Robustness to Incomplete Taxon Sampling

The results are generally similar when only three species are sampled from each higher taxon (Fig. 5). Under these conditions, the species-as-terminals method still outperforms the other methods, albeit by a smaller margin. There are two reasons for this: (1) subsampling species decreases the accuracy of the species-as-terminals method, and (2) subsampling increases the accuracy of the other coding methods (except the majority method, which may be slightly outperformed by other methods at low branch lengths when few species are sampled). Thus, the latter methods (missing, primitive, derived) seem to give worse results as more data (species sampled) are added. This disturbing phenomenon is presumably caused by the fact that the observed interspecific variability increases as more species are sampled; and as this variability increases, these methods either exclude more characters (fixed-only), render more data cells uninformative (missing, polymorphic) or arbitrarily fill more data cells with 0's or 1's (primitive, derived).

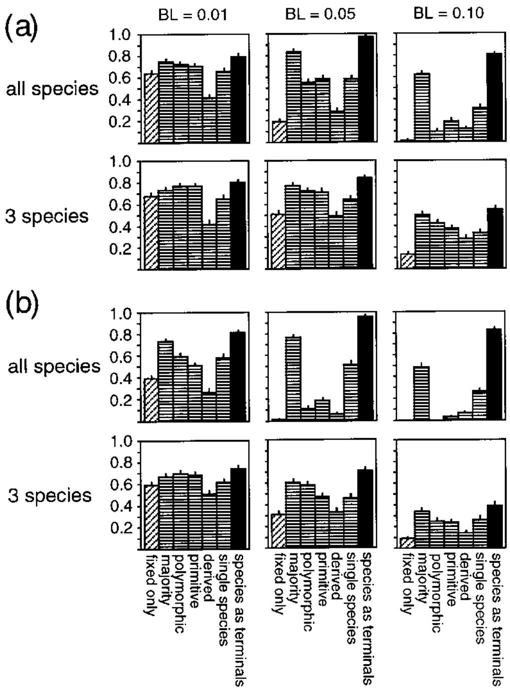

FIGURE 5. The effects of incomplete sampling of species on methods for analyzing interspecific variation. The polymorphic and missing methods give identical results for binary characters. Each bar is the average accuracy from 100 replicated matrices, lines above each bar represent the standard error. (a) 42 species, 200 characters. (b) 66 species, 200 characters.

The results from the subsampling analyses also suggest that subsampling taxa has its greatest negative impact on accuracy when branches are relatively long (0.05, 0.10). Conversely, at a relatively low branch length (0.01), sampling only three species gives a similar level of accuracy to sampling all the species. However, at all branch lengths examined, sampling three species per higher taxon consistently gives more accurate results than does sampling only one.

Sampling a Single Species Decreases Accuracy

The poor performance of coding based on a single (randomly chosen) species relative to treating species as terminals demonstrates the deleterious effects of incomplete taxon sampling (see also Wheeler, 1992). Under almost all conditions examined, sampling a single species from each higher taxon decreases accuracy (relative to including all species), often by a very large margin. Furthermore, when a single species is sampled and branches are moderately long ($\geq$0.10) or variable in length, increasing the number of characters increases accuracy only slightly. There is no evidence from these simulations that increasing the number of taxa decreases accuracy when using species-as-terminals (contra Charleston et al., 1994; Wiens and Reeder, 1995), even when the ratio of characters to taxa is decreased. However, unlike those studies, the analyses in this paper considered only the relationships among the higher taxa (as did Wheeler, 1992); this may actually be a more realistic way to assess the effects of taxon sampling. In summary, these results show the importance of sampling multiple species when inferring relationships among higher taxa.

Does Common Equal Primitive?

The relatively strong performance of the majority method may be surprising to some. This coding method rests on the assumption that, within a variable higher taxon, "common equals primitive." This assumption has been criticized as being unsupported, and has been widely rejected as a criterion for determining character polarity (e.g., Watrous and Wheeler, 1981; Wiley, 1981). However, the simulation results suggest that this assumption may have some predictive value, at least as applied in this study. As a generalization (rather than a strict rule), "common equals primitive" makes intuitive sense. For example, consider a rooted three-taxon tree (A (B, C)). Given a binary character where 0 is primitive and 1 is derived, and assuming that the character changes once within the group, there are four possible outcomes in which the character is variable: (0(1, 1)), (1(0, 0)), (0(1, 0)), and (0(0, 1)). Although the first case would mislead the "common equals primitive" assumption, it is the only one of the four that would, and assuming that all cases are equally probable, would be relatively unlikely. Thus, in this very simple scenario, the "common equals primitive" assumption holds true roughly 75% of the time.

Simulations designed to explicitly test this assumption (Fig. 6) suggest that at low rates of change (0.10 or below) the plesiomorphic state within a group can be deduced correctly without knowledge of the phylogeny based on its commonality among the species about 80% of the time or more. The frequency with which common equals primitive is generally higher in the 65-taxon case (except at high rates of change), which suggests that this conclusion will hold true for even larger numbers of taxa. The assumption that the common state is primitive becomes less tenable as branch lengths increase (approaching 50%, or randomly picking one of the two states). How-ever, under those conditions in which the common state is not usually primitive, branch lengths are so

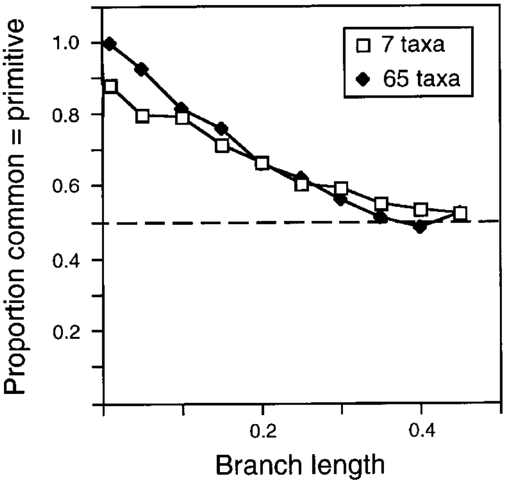

FIGURE 6. A test of the assumption that common equals primitive, with use of simulations of binary characters evolving at different rates of change (branch lengths). Results for the seven-taxon case are averaged across the 11 possible rooted topologies for seven taxa. For each topology, the results are the proportion of variable characters in which the known primitive state is the most common state among the species, based on 200 potentially variable characters. The 65-taxon tree is based on the phylogeny of phrynosomatid lizards (Fig. 1), and each symbol represents the average result from 200 potentially variable characters.

long that all parsimony methods have extremely low accuracy (< 33%). In summary, the common-equals-primitive assumption does seem to hold true with enough frequency to allow the majority method to outperform the other coding methods under many conditions. This observation may have applications to other situations where the ancestral state of a higher taxon with unresolved internal relationships is sought, such as studies of character evolution. Frolich (1987) provided a somewhat different quantitative analysis of the common-equals-primitive assumption, found some support for its reliability, and suggested its application to outgroup analysis (i.e., when outgroup relationships are unresolved).

Excluding Interspecifically Variable Characters

The fixed-only method, which excludes characters that vary within higher taxa, performed very poorly in this study. This suggests that characters that vary within higher taxa can be reliable for inferring relationships between them. How likely is it that this result holds for empirical data sets? Three results from the analyses of the Estes et al. (1988) data support the inclusion of interspecifically variable characters. First, the variable characters contain significant phylogenetic information relative to randomized data, although they contain somewhat less phylogenetic signal than fixed characters alone (Table 2). Second, there is no significant correlation between levels of homoplasy and interspecific variability in these characters (Fig. 7), unless the data are coded by using the derived state method (which performs very poorly in simulations). Third, the trees based on the fixed characters alone contradict several groupings that are traditionally recognized by squamate systematists (e.g., Acrodonta, Anguimorpha, Scincomorpha), whereas one or more of the analyses that include variable characters recovers these clades (Fig. 1). Thus, both empirical analysis and simulations support the inclusion of interspecifically variable characters in higher-level phylogenetic analyses.

TABLE 2. Levels of phylogenetic signal in fixed and variable characters for the data of Estes et al. (1988) for squamate reptiles using different methods for coding interspecific variation. Phylogenetic signal was measured as the difference between the observed g_1 index (Hillis, 1991) and the critical g_1 value for the data when randomized (the lower 95% confidence interval for 100 data sets with states randomly shuffled among taxa; randomization program written by J. Huelsenbeck). Because Estes et al. (1988) did not provide data for individual species, some methods (majority, type species, species-as-terminals) could not be applied.

Methods	Phylogenetic signal: With fixed characters	Phylogenetic signal: Polymorphic only
Fixed only	−0.675	
Derived state	−0.469	−0.330
Primitive state	−0.576	−0.474
Missing	−0.657	−0.506
Polymorphic	−0.631	−0.492

The exclusion criterion for variable characters in this study was extreme (a character was excluded if there was any variation within any taxon), whereas more-forgiving criteria may be used by practicing systematists. These criteria would probably give more accurate results than excluding all variable characters. However, given that little evidence from the present analysis suggests that more-interspecifically variable characters are generally more homoplastic (Fig. 7), using less extreme exclusion criteria seems unlikely to give better estimates than does including all variable characters.

Using Species as Terminal Taxa

The species-as-terminals method performed extremely well in this study. Despite this strong suggestion that this method may be preferable to using higher-level taxa as terminals, some caveats should be mentioned. Accuracy of the species-as-terminals method was not compared to the method in which inferred ancestral states are coded (IAS), and it is theoretically possible that the IAS method might be superior. In the real world, the effectiveness of the IAS method will presumably depend on the quality of the a priori estimated phylogenies and of the ancestral state reconstructions within the higher taxa, factors that are difficult to realistically model in simulations. However, the species-as-terminals method has a clear advantage in that the relationships that are simply assumed with the IAS method can be tested directly using the species-as-terminals

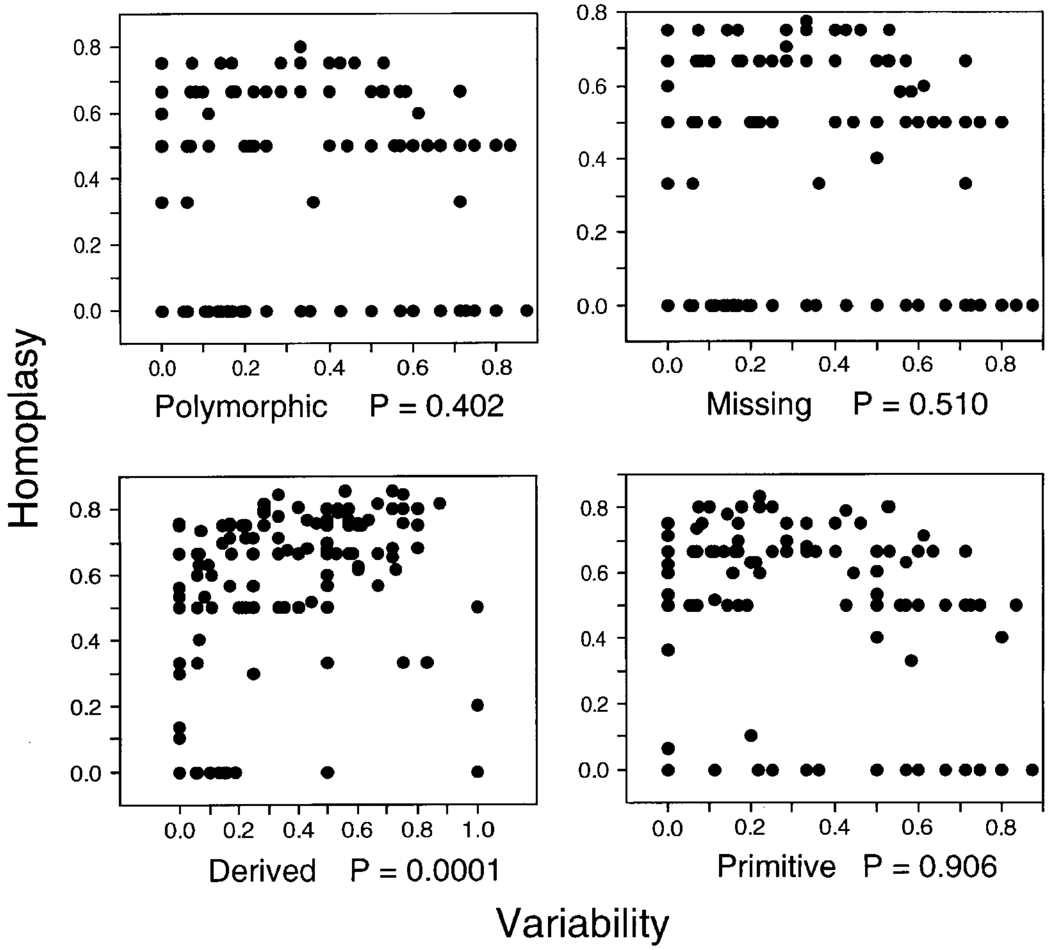

FIGURE 7. Relationship between homoplasy and interspecific variability for the data of Estes et al. (1988) for families of squamate reptiles, using different coding methods. Significance levels are for Spearman's rank correlation of the variables (Statview^{TM} software package). The measure of homoplasy for each character was the homoplasy index (1 − consistency index). Consistency indices for each character were obtained from the trees generated by each of four coding schemes for variable characters. When multiple equally parsimonious trees were obtained from an analysis, consistency indices for each character were averaged across trees. Changes occurring within variable taxa are not included in calculations of homoplasy because these changes do not affect tree reconstruction. Variability was determined as the proportion of taxa that were variable for the derived character divided by the total number of taxa having the derived condition (whether variable or monomorphic within a terminal taxon). Thus, a character with a derived state present only as interspecific variation would have a score of 1, and a character in which no taxa were variable would have a score of 0. The missing and polymorphic methods give slightly different results because some characters are multistate. Because Estes et al. (1988) did not provide data for individual species, some methods (majority, single species, species-as-terminals) could not be applied.

method by including the relevant characters and taxa.

A major strength of the species-as-terminals method is its potential to resolve relationships within each higher-level taxon (in the course of a global analysis), and thereby better resolve the ancestral states of the higher taxa. In the real world, resolving the species relationships within the higher taxa may require scoring additional characters that were uninformative for the higher-level analysis. Thus, using this method may involve extra effort. In the simulations it was assumed that all methods used the same sample of characters. Although in the real world more effort may be involved with the species-as-terminals method, the simulations suggest that this method gives a much better return than the other methods (in terms of accuracy) for the same amount of data.

In these simulations the species-as-terminals method was implemented with the constraint that the higher taxa be monophyletic. This was necessary to make the results comparable to the other methods (which also assume monophyly) and to make the results more easily interpretable. This constraint negates an important advantage of the species-as-terminals method because it does not allow the analysis to test the monophyly of the higher taxon. It should also be noted that the species-as-terminals method might not perform as well if the monophyly constraint is lifted.

In empirical studies, the coding of a higher taxon may be based on different species for different characters. In such cases, using the species-as-terminals method may be problematic, because it may require including many taxa with many missing data entries for some characters. However, resampling studies suggest that the effects of including incomplete taxa may often be only slightly worse than including taxa that are complete (Wiens and Reeder, 1995).

Given the relatively strong performance of the method of splitting up the variable taxon for analyzing interspecific variation, one might ask whether or not this method would perform as well for intraspecific polymorphism (e.g., using individuals as terminal taxa). In fact, this general approach is widely applied to the analysis of intraspecific variation in DNA sequence and restriction-site data, where each genotype is treated as a separate terminal taxon in the analysis (e.g., Crandall and Fitzpatrick, 1996; Shaw, 1996). Nevertheless, this approach has an important disadvantage in the intraspecific case, in that it does not allow intraspecific polymorphisms shared between species to act as synapomorphies. Instead, intraspecific polymorphisms must be treated as homoplasies or as evidence that the species are not "monophyletic." In the intraspecific case, methods designed to group species based on shared polymorphisms may be preferable (e.g., Wiens, 1995), and simulation results suggest that some of these methods may be more accurate than the individuals-as-terminals method when there is an abundance of shared intraspecific polymorphisms (Wiens, M. Servedio, and R. Servedio, unpubl. data). This remains an area in need of further study.

Robustness of Results to Changes in the Model and Methods

As in all simulations, a number of simplifying assumptions were made in this study, including: (1) only six higher taxa; (2) only two states per character; (3) a finite number of species within terminal taxa (maximum of 21), and a limited number of tree shapes relating them; and (4) no intraspecific polymorphism, no missing data, and complete independence of characters. These assumptions are probably not met in most real data sets, and violations of any or all of these assumptions may affect the accuracy

TABLE 3. Accuracy of parsimony methods for analyzing interspecific variation when the estimate for a given method for a given data matrix is based on a randomly selected, fully resolved shortest tree (so that all methods have the same level of resolution). Results are similar to those obtained when accuracy is measured using the strict consensus of the shortest trees as the estimate for a given method (Figs. 3–5), particularly in terms of the relative success of the methods. The number of characters is 200, and branch lengths (BL) are invariant among characters and lineages.

	42 species		66 species	
Methods	BL = 0.01	BL = 0.10	BL = 0.01	BL = 0.10
Fixed only	0.753	0.150	0.497	0.133
Majority	0.817	0.663	0.787	0.550
Missing	0.823	0.300	0.720	0.120
Primitive state	0.820	0.310	0.637	0.130
Derived state	0.490	0.157	0.270	0.100
Single species	0.727	0.327	0.700	0.297
Species-as-terminals	0.857	0.863	0.907	0.863

of the methods. However, it is unclear how realistic violations of any of these assumptions could overturn the major results of this study in terms of the relative success of the methods. Furthermore, the results of this study appear to be robust to a number of changes in: (1) number of characters, (2) number of species within higher taxa, (3) tree shape within higher taxa, (4) branch lengths, (5) differences in probabilities of change among lineages and among characters, and (6) incomplete sampling of species within higher taxa.

In this study the accuracy of a given method for a given data matrix was measured using the strict consensus of the shortest trees as the estimate of phylogeny. This approach was chosen to reflect common practice in empirical studies; however, the results may be biased against methods that give poorly resolved trees. Nonetheless, using instead a randomly-selected, fully-resolved shortest tree as the estimate (so that all methods have the same level of resolution) gives similar results to those obtained by using consensus trees (Table 3).

RECOMMENDATIONS AND CONCLUSIONS

Morphological systematists commonly code higher taxa as terminal units for phylogenetic analysis, and this traditional approach has recently been advocated for analyzing large molecular data sets as well. However, the simulation results of this study suggest that splitting up higher taxa and using species as terminals gives consistently more accurate estimates than do the other coding methods, even when only a few species are sampled from each higher taxon. This approach is strongly recommended for empirical studies. If it is not possible to split up the higher taxa, then the goal of coding is to represent the ancestral state within each taxon. Under these circumstances, the IAS approach (using a priori information on the phylogeny within the higher taxon) is recommended, although its accuracy was not directly addressed in this study. If such phylogenetic information is unavailable, the commonality of character states among species may be useful for inferring the ancestral state, and the majority method generally performs better in simulations than methods that disregard the distribution of states within the higher taxon (missing, polymorphic, primitive, derived). The simulation results also demonstrate that sampling multiple species within each higher taxon is crucial for recovering accurate trees.

Despite the widespread practice among morphologists of excluding characters that vary within higher taxa, the results of this study strongly support their inclusion. Simulation results suggest that their exclusion may greatly decrease phylogenetic accuracy under many conditions, and empirical data from squamate reptiles confirm that they contain useful phylogenetic information.

Intraspecific and interspecific variation are fundamentally different. However, there are some interesting similarities between the results of this study and recent analyses of intraspecifically polymorphic characters (Wiens, 1995; Wiens and Servedio, 1997). Both suggest that, despite the frequent exclusion of polymorphic characters, these characters contain useful phylogenetic information, and their exclusion decreases phylogenetic accuracy. Many of the methods commonly used to code inter- and intraspecific variation are the same (e.g., majority, missing, polymorphic). Although the methods that seem to be most accurate for each type of variation differ (intraspecific = frequency; interspecific = species-as-terminals), both methods share an important feature; they both use detailed information on the distribution of states within the variable taxa. It makes considerable intuitive sense that methods that utilize the most information should perform the best, even if these methods are not the ones currently most widely used.

ACKNOWLEDGMENTS

I thank David Swofford for permission to use and publish results from test versions of his PAUP* software package. I am grateful to John Huelsenbeck for use of his program for randomizing data matrices and to David Cannatella for writing the program for calculating homoplasy levels (programs used in the squamate analyses). I thank Chris Beard, Harold Bryant, David Cannatella, Phil Chu, Kevin de Queiroz, Richard Leschen, Brad Livezey, John Rawlins, Maria Servedio, and Joe Slowinski for useful comments on various versions of this manuscript.

REFERENCES

BALLINGER, R. E., AND D. W. TINKLE. 1972. Systematics and evolution of the genus *Uta* (Sauria: Iguanidae). Misc. Publ. Mus. Zool. Univ. Mich. 145:1–83.

CARLSON, S. J. 1995. Phylogenetic relationships among extant brachiopods. Cladistics 11:131–197.

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

CHARLESTON, M. A., M. D. HENDY, AND D. PENNY. 1994. The effects of sequence length, tree topology, and number of taxa on the performance of phylogenetic methods. J. Comput. Biol. 1:133–151.

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

CRANDALL, K. A., AND J. F. FITZPATRICK, JR. 1996. Crayfish molecular systematics: Using a combination of procedures to estimate phylogeny. Syst. Biol. 45:1–26.

DE QUEIROZ, K. 1987. Phylogenetic systematics of iguanine lizards. A comparative osteological study. Univ. Calif. Publ. Zool. 118:1–203.

DE QUEIROZ, K. 1989. Morphological and biochemical evolution in the sand lizards. Ph.D. Dissertation, Univ. of California, Berkeley.

DONOGHUE, M. J. 1994. Progress and prospects in reconstructing plant phylogeny. Ann. Missouri Bot. Gard. 81:405–418.

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

EERNISSE, D. J., J. S. ALBERT, AND F. E. ANDERSON. 1992. Annelida and Arthropoda are not sister taxa: A phylogenetic analysis of spiralian metazoan morphology. Syst. Biol. 41:305–330.

ESTES, R., K. DE QUEIROZ, AND J. A. GAUTHIER. 1988. Phylogenetic relationships within Squamata. Pages 119–281 *in* Phylogenetic relationships of the lizard families (R. Estes and G. K. Pregill, eds.). Stanford Univ. Press, Stanford, California.

FROLICH, M. W. 1987. Common-is-primitive: A partial validation by tree counting. Syst. Bot. 12:217–237.

FROST, D. R., AND R. ETHERIDGE. 1989. A phylogenetic analysis and taxonomy of iguanian lizards (Reptilia: Squamata). Misc. Publ. Univ. Kans. 81:1–65.

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

HARDING, E. F. 1971. The probabilities of rooted tree-shapes generated by random bifurcation. Adv. Appl. Probab. 3:44–77.

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. 1996. Inferring complex phylogenies. Nature 383:130–131.

KLUGE, A. G. 1989a. A concern for evidence and a phylogenetic hypothesis of relationships among *Epicrates* (Boidae, Serpentes). Syst. Zool. 38:7–25.

KLUGE, A. G. 1989b. Progress in squamate classification. Herpetologica 45:368–379.

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

LIVEZEY, B. C. 1986. A phylogenetic analysis of recent anseriform genera using morphological characters. Auk 103:737–754.

LIVEZEY, B. C. 1996. A phylogenetic reassessment of the tadornine–anatine divergence (Aves: Anseriformes: Anatidae). Ann. Carnegie Mus. 65:27–88.

LIVEZEY, B. C. 1997. A phylogenetic analysis of basal Anseriformes, the fossil *Presbyornis*, and the interordinal relationships of waterfowl. Zool. J. Linn. Soc. 121:361–428.

MISHLER, B. D. 1994. Cladistic analysis of molecular and morphological data. Am. J. Phys. Anthropol. 94:143–156.

MONTANUCCI, R. R. 1987. A phylogenetic study of the horned lizards, genus *Phrynosoma*, based on skeletal and external morphology. Contrib. Sci. Natl. Hist. Mus. Los Angel. Cty. 113:1–26.

NIXON, K. C., AND J. I. DAVIS. 1991. Polymorphic taxa, missing values, and cladistic analysis. Cladistics 7: 233–241.

NOVACEK, M. J. 1992. Fossils, topologies, missing data, and the higher level phylogeny of eutherian mammals. Syst. Biol. 41:58–73.

REEDER, T. W., AND J. J. WIENS. 1996. Evolution of the lizard family Phrynosomatidae as inferred from diverse types of data. Herpetol. Monogr. 10:43–84.

RICE, K. A., M. J. DONOGHUE, AND R. G. OLMSTEAD. 1997. Analyzing large data sets: *rbc*L revisited. Syst. Biol. 46:554–563.

ROWE, T. 1988. Definition, diagnosis, and origin of Mammalia. J. Vert. Paleontol. 8:241–264.

SCHULTZE, H.-P. 1994. Comparison of hypotheses on the relationships of sarcopterygians. Syst. Biol. 43:155–173.

SHAW, K. L. 1996. Sequential radiations and patterns of speciation in the Hawaiian cricket genus *Laupala* inferred from DNA sequences. Evolution 50:237–255.

SLOWINSKI, J. B., AND C. GUYER. 1989. Testing the stochasticity of patterns of organismal diversity: An improved null model. Am. Nat. 134:907–921.

SMITH, J. F. 1996. Tribal relationships within Gesneriaceae: A cladistic analysis of morphological data. Syst. Bot. 21:497–513.

TRUEB, L., AND R. CLOUTIER. 1991. A phylogenetic investigation of the inter- and intrarelationships of the Lissamphibia (Amphibia: Temnospondyli). Pages 223–313 *in* Origins of the higher groups of tetrapods: Controversy and consensus (H.-P. Schultze and L. Trueb, eds.). Cornell Univ. Press, Ithaca, New York.

WALKER, A. K., I. J. KITCHING, AND A. D. AUSTIN. 1990. A reassessment of the phylogenetic relationships within the Microgastrinae (Hymenoptera: Braconidae). Cladistics 6:291–306.

WATROUS, L. E., AND Q. D. WHEELER. 1981. The outgroup comparison method of character analysis. Syst. Zool. 30:1–11.

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.

WHITING, M. F., J. C. CARPENTER, Q. D. WHEELER, AND W. C. WHEELER. 1997. The Strepsiptera problem: Phylogeny of the holometabolous insect orders inferred from 18S and 28S ribosomal DNA sequences and morphology. Syst. Biol. 46:1–67.

WIENS, J. J. 1993. Phylogenetic relationships of phrynosomatid lizards and monophyly of the *Sceloporus* group. Copeia 1993:287–299.

WIENS, J. J. 1995. Polymorphic characters in phylogenetic systematics. Syst. Biol. 44:482–500.

WIENS, J. J., AND T. W. REEDER. 1995. Combining data sets with different numbers of taxa for phylogenetic analysis. Syst. Biol. 44:548–558.

WIENS, J. J., AND T. W. REEDER. 1997. Phylogeny of the spiny lizards (*Sceloporus*) based on molecular and morphological evidence. Herpetol. Monogr. 11:1–101.

WIENS, J. J., AND M. R. SERVEDIO. 1997. Accuracy of phylogenetic analysis including and excluding polymorphic characters. Syst. Biol. 46:332–345.

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

YEATES, D. K. 1995. Groundplans and exemplars: Paths to the tree of life. Cladistics 11:343–357.

Received: 20 June 1997; accepted 1 February 1998
Associate Editor: K. de Queiroz