

POLYMORPHISM IN SYSTEMATICS AND COMPARATIVE BIOLOGY

John J. Wiens

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

Key Words comparative methods, phylogenetic analysis, phylogeny, intraspecific variation, species-limits

■ **Abstract** Polymorphism, or variation within species, is common in all kinds of data and is the major focus of research on microevolution. However, polymorphism is often ignored by those who study macroevolution: systematists and comparative evolutionary biologists. Polymorphism may have a profound impact on phylogeny reconstruction, species-delimitation, and studies of character evolution. A variety of methods are used to deal with polymorphism in phylogeny reconstruction, and many of these methods have been extremely controversial for more than 20 years. Recent research has attempted to address the accuracy of these methods (their ability to estimate the true phylogeny) and to resolve these issues, using computer simulation, congruence, and statistical analyses. These studies suggest three things: that (a) the exclusion of polymorphic characters (as is commonly done in morphological phylogenetics) is unjustified and may greatly decrease accuracy relative to analyses that include these characters; (b) methods that incorporate frequency information on polymorphic characters tend to perform best, and (c) distance and likelihood methods designed for polymorphic data may often outperform parsimony methods. Although rarely discussed, polymorphism may also have a major impact on comparative studies of character evolution, such as the reconstruction of ancestral character states. Finally, polymorphism is an important issue in the delimitation of species, although this area has been somewhat neglected methodologically. The integration of within-species variation and microevolutionary processes into studies of systematics and comparative evolutionary biology is another example of the benefits of exchange of ideas between the fields of population genetics and systematics.

INTRODUCTION

One of the most important trends in systematics and evolutionary biology in recent years has been an increasing appreciation for the interconnectedness of these fields. For example, phylogenies are used increasingly by evolutionary biologists studying ecology and behavior (e.g. 9, 60, 82), and systematists using DNA

and RNA sequence data are beginning to incorporate more and more details of molecular evolutionary processes into their phylogenetic analyses (e.g. 130). One of the areas in which a phylogenetic approach has had an important impact is the study of within-species variation, particularly in the fields of phylogeography and molecular population genetics (e.g. 4, 45, 51, 62, 126). However, many unresolved questions remain as to what the study of within-species variation and microevolutionary processes might have to offer between-species systematic and comparative evolutionary studies (e.g. 58).

Heritable variation within species is the basic material of evolutionary change and the major subject of research on microevolutionary processes. Intraspecific variation is abundant in all kinds of phenotypic and genotypic traits, including morphology, behavior, allozymes, and DNA sequences. This variation is not really surprising because if characters vary between species, they must also vary within species, at least at some point in their evolution. In many cases, especially among closely related species, this intraspecific variation may persist and may be abundant. For example, among the nine species of the lizard genus *Urosaurus*, 23 of 24 qualitative morphological characters that vary between species were found to vary within one or more species as well (136).

I define polymorphism as variation within species that is (at least partly) independent of ontogeny and sex. I assume that this variation is genetically based and heritable, and for the purposes of this paper I deal primarily with variation in discrete or qualitative characters, rather than continuous variation in quantitative traits.

Despite the prevalence of intraspecific variation, phylogenetic biologists have a long and continuing tradition of ignoring polymorphism. For example, morphological systematists often exclude characters that show any or “too much” variation within species (109a). Both molecular and morphological systematists often “avoid” or minimize polymorphism by sampling only a single individual per species. When polymorphism is dealt with explicitly, as in phylogenetic analyses of allozyme data and some studies of morphology, the appropriateness of different methods for phylogenetic analysis of these data is controversial and has been the subject of heated debate for over 20 years (e.g. 11, 12, 20, 33–35, 39, 40, 43, 75, 90–93, 96, 97, 116, 129, 130, 137–139, 142, 143). The controversy over the efficacy of different methods for analyzing polymorphism is not merely academic because different methods may give very different estimates of phylogeny from the same data (Figure 1; 137).

Different phylogenetic hypotheses may have very different implications for comparative evolutionary studies. But even if the tree is stable, different methods of treating within-species variation in ancestral state reconstructions may lead to radically different hypotheses about how traits evolve (see below). Descriptions of comparative methods designed for discrete traits (e.g. 76, 104, 118) rarely mention that these traits may vary within species or what the potential impact may be of this variation on the methods or results.

Species-level systematics, or alpha taxonomy, also involves analyzing polymorphic characters. Analytically, the main task of species-level systematics is to

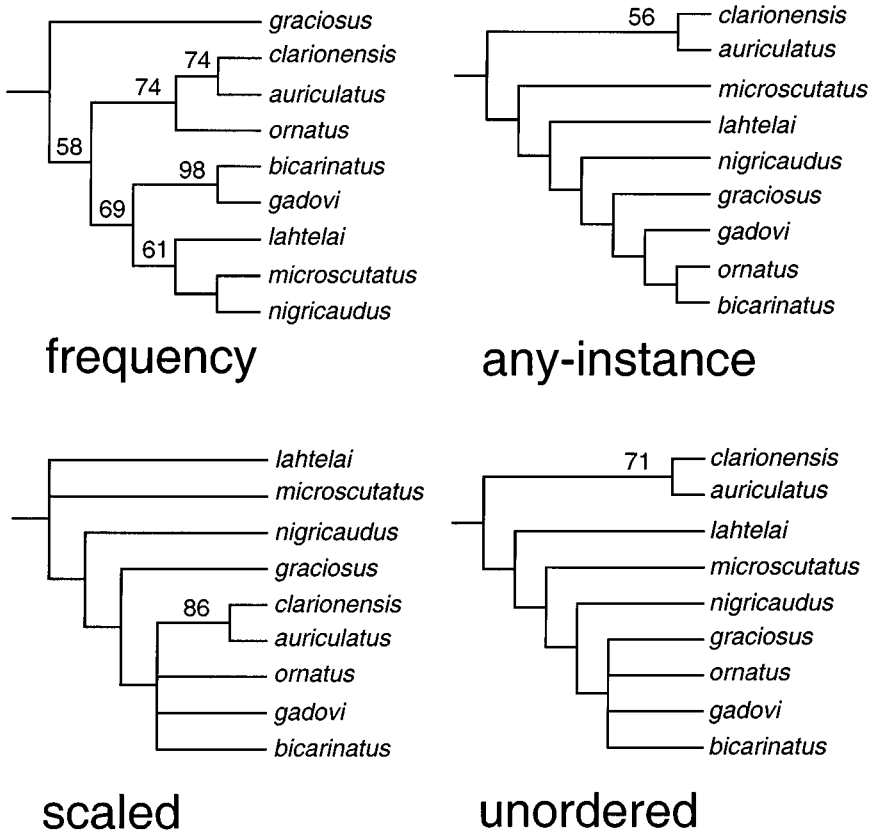


Figure 1 Different methods for coding polymorphic characters for phylogenetic analysis lead to very different hypotheses of evolutionary relationships. Results are based on morphological data for the lizard genus *Urosaurus* (136). Numbers at nodes indicate bootstrap values (42; bootstrap values <50% not shown). Each data set was analyzed with 1000 pseudoreplicates with the branch-and-bound search option.

distinguish between intraspecific and interspecific character variation. The delimitation, diagnosis, and description of species is at least as important an endeavor of systematics as phylogeny reconstruction. Yet, in contrast to phylogeny reconstruction, there has been relatively little methodological improvement in this area, especially as practiced by morphological systematists, who have described and will continue to describe most of the world's species. Alpha taxonomy is a branch of systematics that would benefit tremendously from a more explicit treatment of polymorphism.

In this paper, I review the implications of within-species variation for studies of systematics and comparative biology. I first provide an overview of common methodologies for dealing with polymorphism in phylogeny reconstruction and of

some of the controversies surrounding these methods. I then describe recent studies designed to test the accuracy of these methods and resolve these controversies. I also discuss the impact, although considerably less studied, of within-species variation on comparative studies of the evolution of discrete or qualitative characters. Finally, I review the problem of delimiting species and the operational criteria and methodologies used for delimiting species and distinguishing within and between species variation.

PHYLOGENY RECONSTRUCTION

General Approaches

Polymorphism is important in reconstructing the phylogeny among species for two reasons. First, it is common in data of all types, including morphology, molecules, and behavior. Second, when polymorphism is present, it may have a significant impact on phylogenetic analyses. In particular, various methods for dealing with polymorphism may lead to very different estimates of phylogeny, even when relationships are strongly supported by one or more methods (Figure 1). The abundance and impact of polymorphism are especially clear for closely related species, but different methods for analyzing polymorphic data may affect higher-level relationships as well (e.g. relationships among genera; 138, 139). Yet, surprisingly, the issue of polymorphism is frequently ignored by systematists, particularly those working with morphological and DNA sequence data.

Systematists deal with polymorphism, or avoid dealing with polymorphism, in a number of different ways. These general approaches loosely reflect different types of data. Morphologists often exclude characters in which polymorphism is observed, and in fact this is the most common reason given for excluding characters (109a). This practice may be far more common than is apparent from the literature because morphologists rarely provide criteria for excluding or including characters (109a). The next most common exclusion criterion, excluding characters that show continuous variation, also reflects the desire to avoid characters that vary within and overlap between species. The justification for excluding polymorphic characters is rarely made clear by empirical systematists. Yet, there is a persistent idea in the systematics literature, dating back to Darwin (22), that the more variation characters show within species, the less reliable they will be for inferring the phylogeny among species (32, 86, 123). There have been few empirical tests of this idea.

Systematists working with sequence and restriction-site data typically deal with intraspecific variation by treating each individual organism (or each unique genotype or haplotype) as a separate terminal unit in phylogenetic analyses. Thus, variation within species is effectively treated in the same way as variation between species (134). However, some authors have recently suggested modifications to this general approach, specifically tailored to the problem of analyzing variation

within species (e.g. 18, 19, 127, 132). Of course, one variant of this approach is to sample only a single individual from each species. This sampling regime, although obviously controversial (2, 127, 138, 142, 143), is often employed by both molecular and morphological systematists.

A third general approach involves treating each species (or population) as a terminal unit in the phylogenetic analysis. This approach incorporates intraspecific variation by different methods of coding in a parsimony or discrete character framework or by conversion of trait frequencies to genetic distances (or direct analysis of frequencies using continuous maximum likelihood; 38). This general approach is most frequently applied to allozyme data but is sometimes used for morphological data as well (12, 14, 108). A plethora of methods for dealing directly with polymorphism have been proposed and used, including at least eight parsimony coding methods (described below), two maximum likelihood methods (38, 100), and no less than 36 genetic distance methods (e.g. 114, 115, 130, 148), where each distance method is a combination of tree-building algorithm and genetic distance measure.

These parsimony, likelihood, and distance methods, designed explicitly for polymorphic data, have been the subject of considerable controversy, dating back more than 20 years (20, 39, 90, 91, 96, 97, 129, 130). Two questions have been particularly prominent. First, are frequency data appropriate for phylogenetic analysis? Many authors have argued that the frequencies of traits or alleles within species are not useful for reconstructing phylogenies among species, largely because they are thought to be too variable in space and time within species (e.g. 20, 96, 97) and are not heritable, organismal traits (e.g. 97, 122). Proponents of frequency methods have argued that frequency methods utilize valuable information ignored by other methods (e.g. a trait occurring at a frequency of 1% is different from one occurring at a frequency of 99%), even if frequencies are not stable over a macroevolutionary time scale (129, 130, 137). These authors have also argued that frequency methods downweight rare traits, and therefore they will be less subject to problems of sampling error than methods that merely treat traits as present or absent (i.e. a trait that is rare but present in several related species will be detected only sporadically with finite sample sizes, creating homoplasy, but this homoplasy will have little impact if frequency methods are used).

The second question is whether polymorphic data should be analyzed using parsimony or distance methods (e.g. 33–35, 39, 40, 43, 90, 97, 130). Most of the debate surrounding this topic has not directly involved the accuracy of the methods, but rather issues such as the meaning of branch lengths and negative distances (e.g. 33–35, 39, 40, 43).

The maximum likelihood method most widely applicable to polymorphic data (continuous maximum likelihood or CONTML; 38) has been largely ignored by empirical systematists (but see 120), presumably because it assumes a clearly unrealistic model of evolution (e.g. 71, 129). Namely, it assumes no mutations and no fixations or losses of polymorphic traits (38). However, the sensitivity of

the method to violations of these assumptions has not been thoroughly explored until recently.

Methods for Coding Polymorphic Data

In this section, I briefly review some of the methods commonly used for coding polymorphic data for parsimony analysis (see also Figure 2). The terminology for these methods follows Campbell & Frost (12) and Wiens (137).

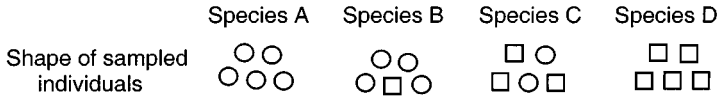
Any Instance Using this coding method, a derived trait is coded as present regardless of the frequency at which it occurs within a species (e.g. 1 to 100%). However, this method is problematic in that it can hide potentially informative reversals (12) such as the reappearance of the primitive trait as a polymorphism (e.g. a transition from 100% to 50% for the derived state). Mutation coding (96, 97) is similar to any-instance coding but potentially allows for characters with multiple derived states to be analyzed. However, its application is “frequently impossible for most loci” (96, p. 32).

Majority Using majority or “modal” coding, a species is coded as having the most common state of the polymorphic character. Potential disadvantages of this method are that it ignores the gain and loss of traits at frequencies less than 50% and that it gives a large weight to small changes in frequency close to 50% (e.g. a change from 49% to 51% has the same weight as a change from 0 to 100%).

Missing When a species that is polymorphic for a given character is coded as missing, the state is treated as unknown in the phylogenetic analysis. Any state is considered a possible assignment to the species, even if the state was not one of the ones observed to be present in the variable species (at least using PAUP). Disadvantages of the missing method are that polymorphic data cells are uninformative in tree reconstruction, and polymorphic states can be treated neither as synapomorphies nor as homoplasies.

Polymorphic Under polymorphic coding, a variable species is coded as having both states (using PAUP or MacClade). When the data are analyzed, the variable species is treated as if either state is present, but the variable cell is largely uninformative in building the tree (although some placements of the variable taxon may be considered more parsimonious than others), and the most parsimonious

Figure 2 Different methods for coding polymorphic characters, illustrated with a hypothetical example. Five individuals are sampled from each of four species, and the circular shape represents the primitive condition and the square shape is derived. The step matrix shows the different costs (in number of steps) for transitions between each of the states; the costs are based on the Manhattan distance between the frequencies of each species for this character. Modified from Wiens (140).



Proportion square	0%	20%	60%	100%
Any-instance	0	1	1	1
Majority	0	0	1	1
Missing	0	?	?	1
Polymorphic	0	(0,1)	(0,1)	1
Scaled	0	1	1	2
Unordered	0	1	1	2
Unscaled	0	1	1	2
Frequency-bin	a	f	p	y
Frequency step matrix	A	B	C	D

Frequency bins

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

Frequency step matrix

	A	B	C	D
A	0	20	60	100
B	20	0	40	80
C	60	40	0	40
D	100	80	40	0

state assignment to the variable taxon is assigned a posteriori. As with the missing method, polymorphic states are not treated as synapomorphies, a serious disadvantage of both methods.

Scaled Using scaled coding, a species is coded as absent (“0”), polymorphic (“1”), or fixed (“2”) for the derived trait. The states are ordered under the assumption that traits pass through a polymorphic stage between absence and fixed presence. If no polymorphic state is observed, it is assumed that the polymorphic stage was present but unobserved (i.e. it costs two steps to go from 0 to 2). The scaled method is equivalent to the step matrix method of Mabee & Humphries (75), but the use of a step matrix allows complex ordering of polymorphic multistate characters when there is no clear relationship among the states (as in the case of different combinations of alleles at an allozyme locus). The scaled method is advantageous in that it allows polymorphisms to act as synapomorphies (unlike the missing and polymorphic methods) and it does not mask reversals (unlike any-instance coding) or the gain and loss of rare traits (as does majority coding). However, it is potentially disadvantageous in that it utilizes no frequency information, and a change from 1% to 100% has the same weight as a change from 99% to 100%.

Unscaled The unscaled method is identical to the scaled method, except that for characters in which no polymorphism is observed it is assumed that the character did not pass through a polymorphic stage between absence and fixation. Therefore, a change from fixed absence to fixed presence has a cost of one step under unscaled coding, but a cost of two steps under scaled coding.

Unordered Unordered coding is identical to scaled and unscaled coding except that all the states are unordered, and there is an equal cost to any transition between any of the character states. As noted by Campbell & Frost (12) and Mabee & Humphries (75), the unordered method is disadvantageous in that it loses any information about the shared presence of traits (i.e. a change from trait absence to fixed presence is no more costly than a change from polymorphic presence to fixed presence).

Confidence Coding The method of Domning (27), which I dub confidence coding, is similar to the majority method but statistically incorporates sample size. For a given species, the 95% confidence interval for the frequency of the commonest trait is found, and if the lower confidence limit is >0.5 , the species is coded as having the majority condition. If not (or if two traits are present at equal frequencies), the taxon is coded “whichever way was more congruent with other characters (i.e. whichever way did not imply a reversal).” The “congruence with other characters” is determined from a preliminary tree.

Frequency Parsimony Methods Frequency methods are a class of methods that use precise information on the frequency of traits within a given species, and weight

changes between states based on differences in frequencies. Genetic distance methods and continuous maximum likelihood are frequency-based approaches, and there are at least three frequency parsimony methods that differ in their precision and versatility.

The most precise method is the FREQPARS program (129), which uses trait frequencies directly. However, FREQPARS has a weak tree-searching algorithm and is unlikely to find the shortest tree unless the data set has only a few taxa.

Wiens (137) used a method (suggested by D Hillis) that approximates the FREQPARS approach while still allowing for thorough tree searching, and this method was described in detail by Berlocher & Swofford (7). The method is implemented by giving each taxon that has a unique set of frequencies a different character state (Figure 2). The cost of a transition between each pair of character states is calculated by finding the Manhattan distance (129) between the frequencies; these transition costs are then entered into a step matrix (Figure 2). The step matrix allows for extremely precise frequency information to be used in character weighting. The main disadvantage of this approach is that step matrices may slow down tree searches prohibitively for large numbers of taxa.

The least precise of the three methods is the frequency-bins method (136, 137; modified from 111). This method is practical for large numbers of taxa (>100; 141) but is designed for binary characters only. With this method, each taxon is assigned one of an array of character-state bins, where each bin corresponds to a small range of frequencies of the putative derived trait (e.g. character state a = frequency of derived trait from 0–3%, b = 4–7%; Figure 2). The bins are then ordered, which forces a large number of steps between large changes in frequency and a small number of steps between small changes in frequency. The choice of bin-size relates to the maximum number of states allowed by the phylogenetic software program; most authors have used 25 bins (Figure 2; 137).

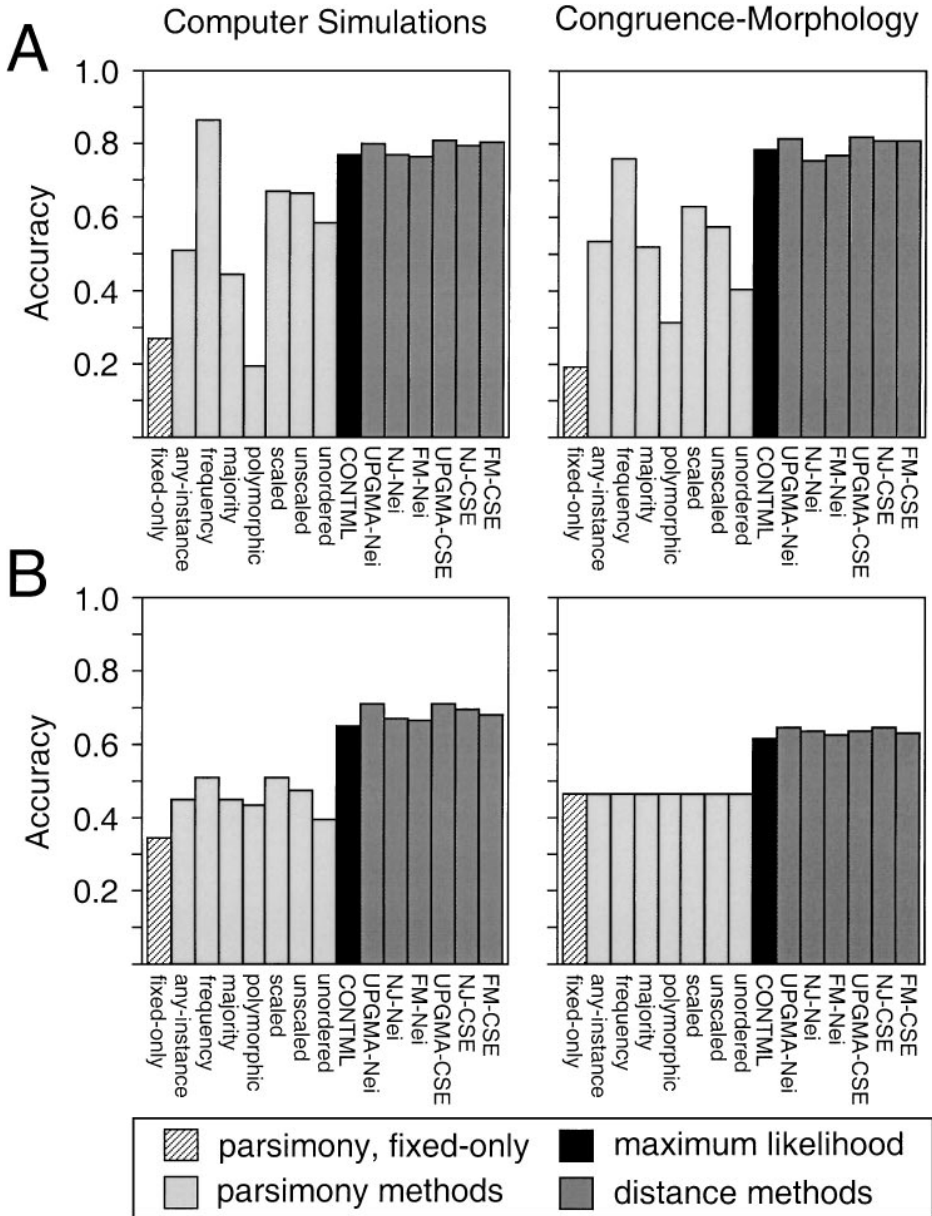
Testing Methods for Phylogenetic Analysis of Polymorphic Data

Recent work has tried to resolve some of the controversies surrounding different approaches for dealing with polymorphic data. In particular, these studies have attempted to address the accuracy of excluding versus including polymorphic characters, sampling single versus multiple individuals per species, and the relative performance of various parsimony, distance, and likelihood methods designed for analyzing polymorphic data. These studies have employed computer simulations (142, 143), congruence analyses of real data (morphology and allozymes; 138, 139), and statistical analyses of empirical data sets (morphology and allozymes; 137). Computer simulation studies involve constructing a known phylogeny, evolving characters on this tree according to some model of evolution, and testing the ability of different methods to estimate this tree given the same data (65). Congruence analyses require finding relationships that are agreed on by multiple data sets, assuming that these well-supported, congruent clades are

effectively “known,” and comparing the frequency with which different methods estimate these clades with the same finite data (95). Congruence studies provide a useful “reality check” on simulation studies, which make many simplifying assumptions about evolutionary processes. Comparing methods according to statistical measures that may relate to accuracy (such as bootstrapping) may be a relatively weak criterion for assessing performance (65). Nevertheless, certain methods do make assumptions that are amenable to statistical testing (e.g. whether or not polymorphic characters or frequency data contain significant nonrandom phylogenetic information; 137).

The common practice of excluding polymorphic characters implicitly assumes one or more of the following: (a) polymorphic characters are more homoplastic than fixed characters (characters that are invariant within species), (b) polymorphic characters do not contain useful phylogenetic information, and (c) inclusion of polymorphic characters will decrease phylogenetic accuracy (relative to excluding them and analyzing only “fixed” characters). Recent studies of empirical data suggest that polymorphic characters are more homoplastic than fixed characters (12, 137). Furthermore, there is a significant positive relationship between levels of homoplasy and intraspecific variability in morphological characters in phrynosomatid lizards (137). These two observations support the long-standing idea that more variable characters may be less useful in phylogeny reconstruction (22, 86, 123) and might be interpreted as supporting their exclusion. Yet, although they are more homoplastic than fixed characters, polymorphic characters nevertheless do contain significant phylogenetic information, as shown by the congruence between trees based on fixed and polymorphic characters (12) and randomization tests of homoplasy levels (137). Furthermore, computer simulations and congruence studies of morphology support the idea that, given a sample of fixed and nonfixed characters of realistic (i.e. limited) size, exclusion of all the polymorphic characters significantly decreases accuracy (Figure 3; 138, 142). In many cases,

Figure 3 Sample of results from simulation and congruence analyses showing the relative performance of methods for analyzing intraspecific variation (with 8 taxa and 25 characters). Data for congruence analyses are from *Sceloporus* (138), and simulations are with branch lengths varied randomly among lineages (from 0.2 to 2.0) and two alleles per locus (142, 143). Each bar represents the average accuracy from 100 replicated matrices, where accuracy is the number of nodes in common between the true and estimated phylogenies. A. Results with $n = 10$ (individuals per species) in simulations and around 10 for many species and characters in the *Sceloporus* data. B. Results with $n = 1$. The parsimony methods give identical results with $n = 1$ for the congruence analyses because heterozygotes are not detectable as such in the morphological data (so there is no polymorphism), whereas heterozygotes can be detected in the simulations. Modified from Wiens (140). CONTML, continuous maximum likelihood (38); NJ, neighbor-joining (117); FM, Fitch-Margoliash (46); Nei, Nei's (98) standard distance; CSE, Cavalli-Sforza & Edwards (13) modified chord distance.



twice as many known clades are recovered when polymorphic characters are included rather than excluded (138). Thus, in this trade-off between having more characters (fixed + polymorphic) and having a smaller number of characters with less homoplasy (fixed-only), it is clearly better to have more characters.

Given the observations that polymorphic characters contain useful phylogenetic information in general but that homoplasy may increase with increasing variability, some authors have proposed downweighting characters based on their level of homoplasy (e.g. using successive approximations; 12) or their degree of intraspecific variability (using a priori weighting; 32, 137). Similarly, many empirical systematists seem to delete the most polymorphic characters from their data sets, excluding characters because of “too much” intraspecific variability as opposed to any variability at all (109a). Simulation and congruence analyses suggest that, while these approaches may improve accuracy in some cases relative to some methods, they rarely improve accuracy relative to the unweighted frequency method including all polymorphic characters (138, 142).

Simulations, congruence studies, and statistical resampling studies also suggest that sample size (individuals per species) may be very important for achieving accurate results, particularly when levels of polymorphism are high (Figure 3; 2, 127, 138, 142, 143). These results argue against the sampling of a single individual per species as a general practice. For example, using congruence analyses of morphological data for spiny lizards (*Sceloporus*), Wiens (138) found that the accuracy of the “best” parsimony method is effectively cut in half by sampling only a single individual per species under some conditions (Figure 3).

An important result of recent congruence and simulation analyses is that the methods that generally perform best are those that make direct use of frequency information, whether they be parsimony, distance, or likelihood. That is, the frequency parsimony method, the genetic distance methods, and continuous maximum likelihood tend to recover more of the well-supported or known clades than do any of the nonfrequency parsimony methods. This same result is obtained for a variety of simulated branch lengths, sample sizes, numbers of taxa, and numbers of characters, and in congruence studies of both morphological (Figure 3) and allozyme (Figure 4) data. Furthermore, statistical analyses of two morphological and five allozyme data sets show that frequency-coded polymorphic characters do contain significant, nonrandom phylogenetic information (137), and that frequency methods perform best among the parsimony coding methods (or are tied for best) for a number of statistical performance criteria. These results contradict the idea that frequency data are too unstable to be used in phylogenetic analysis and that they are misleading.

Recent simulation and congruence studies also show that distance and likelihood methods may outperform all parsimony methods (both frequency and nonfrequency) in many cases. One such situation is equivalent to the “Felsenstein Zone” effect described for fixed characters (e.g. 37, 68, 69), which occurs when there are two unrelated terminal lineages with long branches separated by

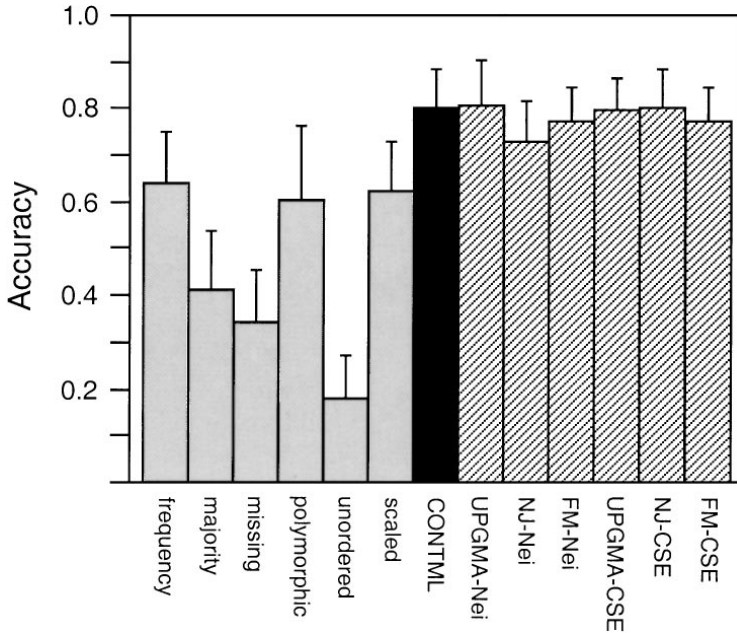
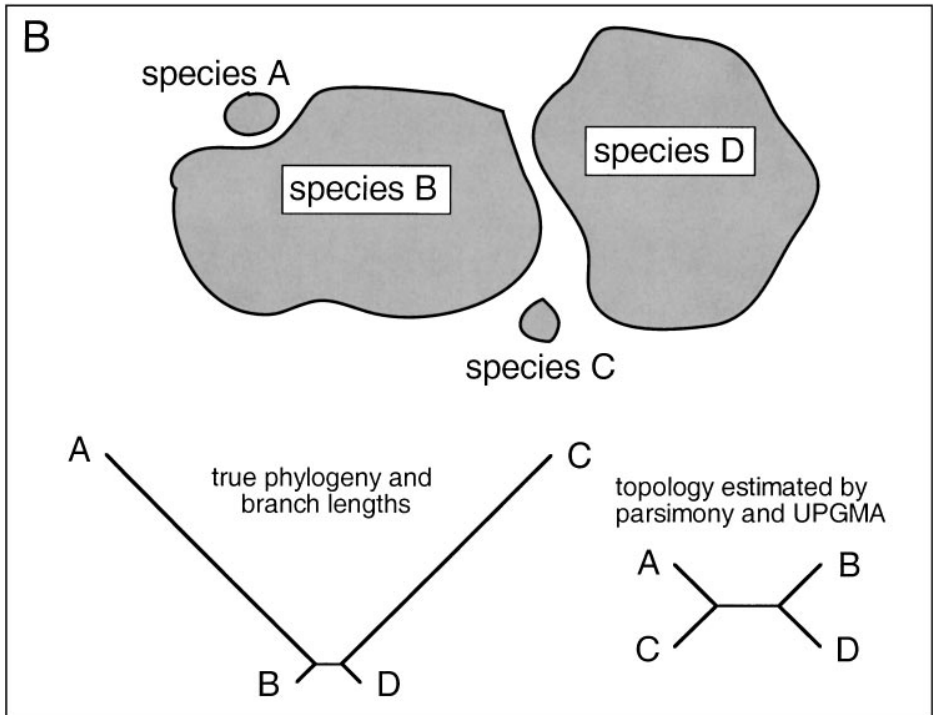
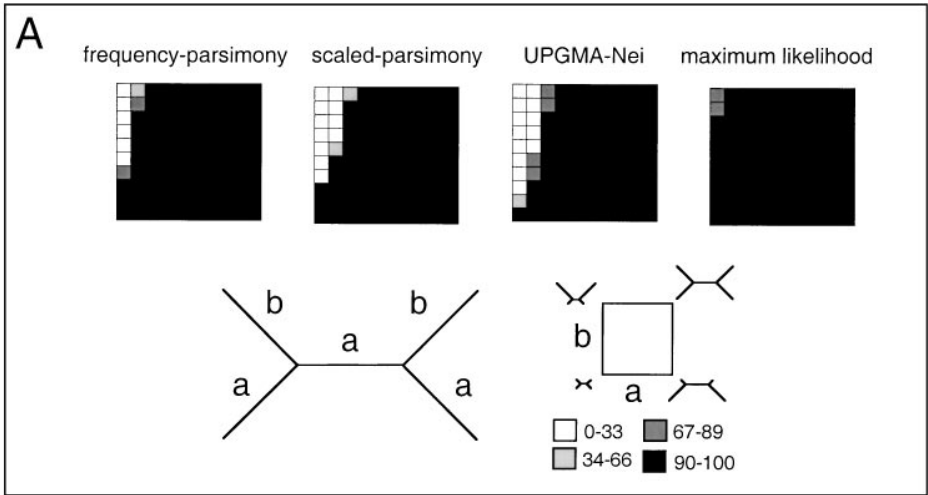


Figure 4 Accuracy of 13 phylogenetic methods averaged across eight allozyme data sets. The line above each bar indicates the standard error of each mean. Accuracy is the proportion of well-supported clades that are correctly resolved by each method. Modified from Wiens (139). See Figure 3 for abbreviations.

a short internal branch. In simulations of polymorphic data using a model in which allele frequencies evolve along branches by random genetic drift (72, 143), this Felsenstein Zone effect might occur when there are two unrelated species with small population sizes (long branches, with high probability of fixation, loss, and/or large changes in trait frequency) that are separately derived from an ancestor with a very large population size (short branch). This might correspond to a peripheral isolate model of speciation in the long-branch species. Under these conditions, parsimony and UPGMA tend to place the taxa with long branches together as sister taxa (incorrectly), even if a large number of characters are sampled (Figure 5). In contrast, continuous maximum likelihood and the additive distance methods (neighbor-joining and Fitch-Margoliash) will tend to estimate the correct tree, especially when given a large sample of characters (72, 143). The Felsenstein Zone effect for polymorphic data is interesting for several reasons: (a) it is very similar to the effect described for fixed characters, even though the simulated models of evolution are extremely different (i.e. the fixed character model has change as mutation only, whereas there is no mutation in the pure drift model), (b) increased taxon sampling to subdivide the long branches is not a potential solution



to the problem, and (c) the effect for polymorphic data has a simple biological explanation and may therefore occur commonly (143).

Of course, the Felsenstein Zone scenario described above may represent a very special case. Furthermore, the possibility of long branch repulsion (i.e. the failure of maximum likelihood to place together two long branches that are actually sister taxa; 70, 121, 150) has not been explored for polymorphic data. Yet, distance and likelihood methods outperform parsimony methods under many other conditions apart from the Felsenstein Zone. In simulations, distance and likelihood methods generally performed as well as or better than any of the parsimony methods under a variety of branch lengths, sample sizes, and numbers of characters and taxa, and the nonparsimony methods consistently outperformed parsimony when sample sizes were very small ($n = 1$ or 2 ; Figure 3; 142). In congruence analyses of morphology, at least some distance methods consistently performed better than any parsimony methods (138), and in congruence analyses of allozyme data sets (139), each of the distance and likelihood methods outperformed (on average) all the parsimony methods (Figure 4). The fact that continuous maximum likelihood performs as well as it does on real and simulated data sets (especially the allozyme data) is particularly interesting, given that the assumptions of this method were almost certainly not met in these data sets (e.g. the method assumes no mutation and no fixation or loss of traits). These results strongly suggest that continuous maximum likelihood will perform well even when its assumptions are violated.

In summary, the results of simulation, congruence, and statistical analyses suggest that (a) polymorphic characters should not be excluded, (b) methods that use frequency data may perform best, and there is no evidence that frequency data are misleading, and (c) distance and likelihood methods may be

Figure 5 The effects of branch lengths and the Felsenstein Zone effect for polymorphic data. A. The effects of branch lengths on the accuracy of four phylogenetic methods, where darker shading represents higher accuracy. The data consist of 500 loci (characters), with two alleles per locus and complete sampling of individuals within each species. Modified from Wiens & Servedio (143). B. Hypothetical example illustrating the Felsenstein Zone effect for polymorphic data. The shaded areas represent the geographic distributions (and relative population sizes) of four species. Species A and C have small geographic distributions, small population sizes, and long branch lengths under a genetic drift model. Species B and D (and the ancestors of all four species) have large population sizes and short branch lengths. Traits will tend to remain polymorphic in B and D but become fixed or lost in A and C. Parsimony methods and UPGMA will tend to put A and C together based on shared fixations, losses, and changes in trait frequency. In contrast, continuous maximum likelihood and the additive distance methods (neighbor-joining and Fitch-Margoliash) can give accurate results under these conditions, given enough characters. Modified from Wiens (140). Nei, Nei's (98) standard distance.

superior to parsimony methods for analyzing polymorphic data under many conditions.

Objections to Frequency Methods

The results of these recent studies provide some resolution to the controversies surrounding the analysis of polymorphic data. These studies show that frequency methods use more informative variation than do any of the other methods (137), are less subject to errors caused by limited sample sizes (129, 137) and unequal branch lengths (142, 143) than are nonfrequency parsimony methods, and generally give more accurate estimates of phylogeny in simulation and congruence analyses than do other methods (138, 139, 142, 143). A number of recent empirical studies have used frequency methods to include and code polymorphic characters, including studies of allozymes (e.g. 10, 31, 88), behavior (e.g. 57), and morphology (e.g. 14, 16, 55, 56, 59, 67, 87, 108, 113, 136, 141). Nevertheless, the use of frequency information in phylogenetic analysis remains controversial (e.g. 94, 97, 122).

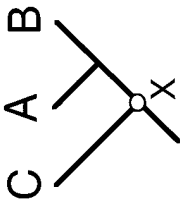
The most common objection to the use of frequency data in phylogenetic analysis appears to be the idea that frequencies are too variable over space and time within species to be used in reconstructing relationships between species. Several authors have cited the study of Crother (20) as evidence that frequency data are unstable and therefore unusable (e.g. 12, 75, 94, 97). This example does not withstand closer scrutiny. Crother analyzed allele frequency data from four populations of *Microtus ochrogaster* and found that phylogeny estimates for these populations based on the same locus differed from year to year (using data from 50). Crother (20) concluded from this example that frequencies vary too much over time and space to be phylogenetically informative for reconstructing relationships among species. However, it should be noted that the "populations" were not natural populations from different localities but were individuals drawn from the same locality confined in four enclosures (50). Thus, there was no true phylogenetic history to be estimated for these populations, and the absence of stable phylogenetic signal in the frequency data is hardly surprising. The fact that there are different estimates of phylogeny from year to year does demonstrate that frequency methods may resolve clades that have little or no support (137). Yet, the weak support for these phylogenies is obvious from low bootstrap values (<50%) and g_1 analysis (i.e. the data contain no significant phylogenetic structure; JJ Wiens, unpublished data). Extrapolating this rather artificial example to the interspecific case and generalizing the results to all applications of frequency data clearly is unjustified, especially in the face of growing evidence that frequency-coded polymorphic data do contain significant, nonrandom phylogenetic structure at the between-species level (87, 108, 137–139).

How can frequencies be highly variable within species but still informative between species? Population genetics theory (e.g. 73) suggests that traits with

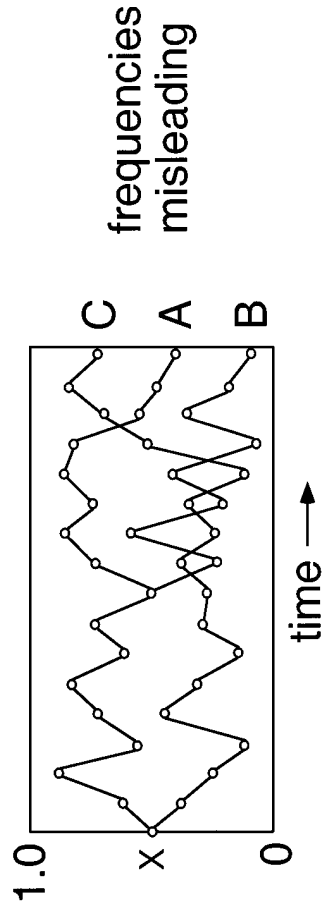
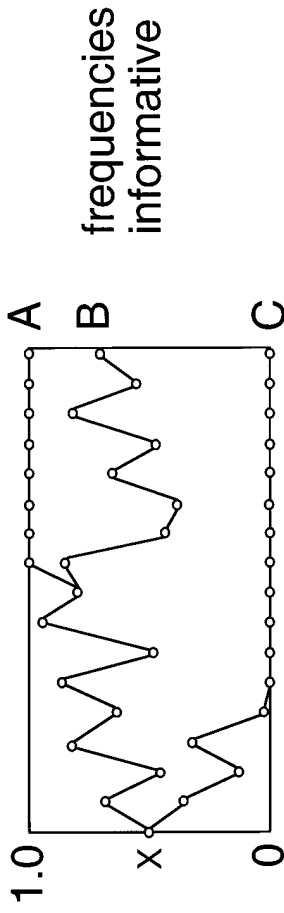
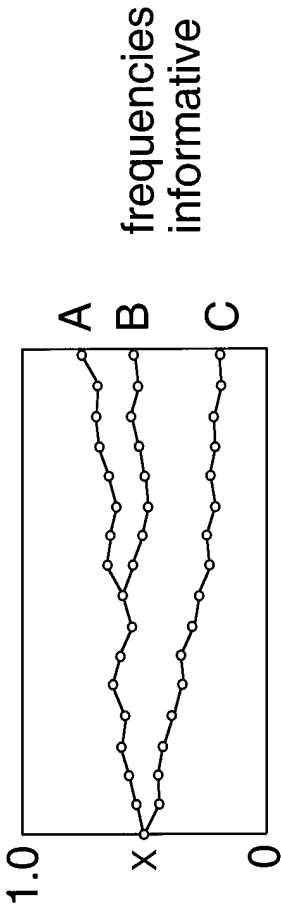
frequencies that are highly variable over time within species are more likely to become fixed or lost over longer evolutionary time scales. Thus, a model in which frequencies change rapidly without fixation or loss for thousands or millions of years seems unrealistic for the vast majority of characters (Figure 6). When fixations and losses do occur, theory predicts that traits at high frequencies are more likely to be fixed than lost, and vice versa. This relationship is mirrored in the weighting scheme of frequency methods. In contrast, nonfrequency methods (except majority) assume that it is just as easy to go to fixation from a frequency of 1% as it is from a frequency of 99%. Thus, in simulations, frequency methods may be superior estimators of phylogeny even when the frequencies of nonfixed traits are nearly randomized between splitting events (142). Furthermore, results of statistical and congruence analyses (137–139) imply that trait frequencies are conserved enough to contain at least some historical information.

Another objection to the use of frequency-based methods is that frequencies are not heritable and/or organismal traits (e.g. 97, 122; note that “heritable” refers to transmission from ancestor to descendant, and not to the quantitative genetic meaning of the term). Although the idea that frequencies are never heritable is not strictly accurate (because frequencies *are* heritable if populations are at Hardy-Weinberg equilibrium), it is likely that nonfixed frequencies between 0 and 100% are rarely passed from an ancestor species to a descendant species without at least some change. But, from a practical perspective, it is clear that these changes in trait frequencies do not prevent frequency methods from accurately estimating phylogenies (e.g. 138, 139, 142, 143). Obviously, if frequencies never changed there would be no variation with which to reconstruct trees. The objection to nonorganismal traits appears to be questionable as well. It is true that frequencies are features of populations and species, and not of individual organisms. However, this is true for polymorphic and intraspecifically-variable quantitative data (133), not just frequency data. The fact that Kluge has argued against inclusion of frequency data is ironic because the exclusion of potentially informative data is contrary to the maxim of total evidence (74).

An objection to frequency methods sometimes raised in specific cases is that sample sizes (individuals per species) may be insufficient (e.g. 12), with the implicit assumption that frequency-based methods will be less accurate with small sample sizes than qualitative coding methods. In fact, simulations suggest that as sample sizes decrease, the performance of all methods becomes increasingly similar (i.e. if there were no polymorphism, all the polymorphism coding methods would be identical). But even with small sample sizes (e.g. $n = 1$ or 2 individuals per species), there are still noticeable differences in accuracy among methods, with frequency methods generally outperforming other coding methods (Figure 3B; 142, 143). An objection to frequency methods based on finite sample sizes is surprising because the putative robustness of frequency methods to finite sample sizes has traditionally been the major argument to justify their use (129, 137).



frequency of trait "a"



Areas for Future Research

Morphology Further research is needed on the phylogenetic analysis of polymorphic data for a variety of reasons. Recent studies on polymorphic morphological characters suggest that two common practices of morphological systematists, excluding polymorphic characters and ignoring frequency data, may lead to relatively poor estimates of phylogeny. But the congruence results, and those showing the information content of polymorphic characters, are so far based on morphological data from only one clade of lizards (the Phrynosomatidae). On the positive side, these conclusions are also upheld by simulations. Furthermore, recent studies in other groups of vertebrates show significant phylogenetic information in frequency-coded data, or at least frequency-based trees that are congruent with previous taxonomy or other data sets (e.g. 14, 16, 56, 67, 87, 108). Nevertheless, the generality of these conclusions should be tested in morphological data sets from other groups of organisms, especially plants and invertebrates. These conclusions will be difficult to test until more morphological systematists publish data on polymorphic characters and trait frequencies.

Previous work on intraspecific variation in morphology has focused largely on discrete, or at least qualitatively described traits. The analysis of quantitative characters with ranges of trait values overlap between species (such as meristic and morphometric variables) is also in need of study. Overlapping quantitative data, like qualitative polymorphic data, are also frequently excluded from morphological phylogenetic studies (109a), and yet numerous methods for their analysis have been proposed and debated (1, 36, 44, 52, 107, 128, 133). Using data from plants, Thiele (133) has shown that overlapping quantitative data do contain significant nonrandom covariation and produce trees that are significantly congruent with trees based on qualitative data, at least when using his gap-weighting method (which is very similar to the frequency-bins method; 137). Like the results from qualitative polymorphic characters, Thiele's results from quantitative characters support the inclusion of characters despite within-species variation, and the use of methods that treat continuously valued data (e.g. frequencies, means of quantitative traits) as continuous (e.g. frequency methods, gap weighting). However, more studies are needed, in plants and other organisms, to test the generality of these conclusions

←

Figure 6 Hypothetical example showing the changes in a trait “a” over time among three species (A, B, C). If frequencies change only slightly over time, trait frequencies should track the phylogeny and frequency methods should be effective (*top*). If frequencies change rapidly over time and traits go to fixation or loss then frequencies can still be informative (*middle*), because fixations and losses can be synapomorphies and will prevent further oscillations in trait frequencies. Frequencies are most likely to be misleading when frequencies change rapidly over time without becoming lost or fixed (*bottom*), but this seems unlikely without some unusual mechanism to simultaneously drive change and prevent fixation and loss (e.g. frequency dependent selection). The starting frequency at time “X” is 0.5 in all three cases.

and to compare the performance of different methods for analyzing quantitative morphological data.

Allozymes There are also many unanswered questions in the phylogenetic analysis of polymorphic allozyme data. Using congruence analyses of eight allozyme data sets, Wiens (139) found that the relative performance of different methods varies greatly from data set to data set. For example, for some data sets, the polymorphic coding method is the best of all the parsimony, distance, and likelihood methods (i.e. recovers more “known” clades than any other method), whereas for other data sets it is one of the worst. A comparable situation exists for UPGMA with Nei’s distance, which on most data sets is the most accurate method (or is tied for most accurate) but on other data sets performs relatively poorly. Although the failure of these methods can be easy to explain (e.g. UPGMA is sensitive to unequal branch lengths), their strong performance on certain data sets is perplexing. This variability is particularly vexing because it makes it difficult to choose a single method that will be “the best” for every data set, or to understand which method may be preferred in a particular case. Simulation studies, with data explicitly designed to mimic allozymes, may be necessary to better understand why certain methods behave so well on some data sets but not others.

Furthermore, for both allozyme and morphological data, it is unclear whether the Felsenstein Zone effect described in simulations under a genetic drift model applies to many real data sets. Perhaps just as importantly, it is unclear whether the methods that appear to be robust to this problem in simulations (e.g. continuous maximum likelihood) will also be robust in real data sets. Simulations that utilize more complex models than those employed by Kim & Burgman (72) and Wiens & Servedio (143) may be particularly useful for addressing this question.

DNA Data Sequence data and restriction-site data are becoming widely used for inferring relationships among closely related species. This is a level where polymorphism may often have a significant impact on phylogenetic studies, but the simulation and congruence studies mentioned above may not be applicable to DNA data. These studies focused on trait frequencies at multiple, unlinked loci, whereas DNA data typically consist of linked characters at a single locus (i.e. a single nuclear gene or one or more mitochondrial or chloroplast genes). Theoretical work on the impact of within-species variation on interspecific phylogenetic inference using DNA sequence data has dealt primarily with the problem of incomplete lineage sorting of ancestral polymorphisms (e.g. 78, 99, 105, 131, 149). In this situation, the phylogeny of the gene(s) may not be congruent with the phylogeny of the species (especially when population sizes are large and/or divergence times are recent), and theoreticians have explored the effects of sampling multiple individuals and loci as possible solutions. There seems to be general agreement that sampling enough unlinked loci will resolve the problem. When only one locus is available (e.g. data from the mitochondrial or chloroplast genome), sampling multiple individuals from each species may also be helpful (131).

But how does one infer species phylogeny when multiple individuals are sampled? Surprisingly, this question has hardly been explored. In general, empirical systematists treat each haplotype (unique genotype) as a separate terminal taxon in the phylogenetic analysis, and these individuals may or may not cluster with their putatively conspecific haplotypes. The within-species phylogeny is inferred simultaneously with the among-species phylogeny, and no distinction is made between the two. However, two modifications to the haplotype-as-terminal-taxon approach have been suggested.

Using principles from population genetics and coalescent theory, Templeton et al (132) have recently developed a method (dubbed TCS) specifically tailored for within-species phylogeny reconstruction, and they have suggested its application, in combination with more traditional methods, to better infer between-species phylogeny. The algorithm was designed to overcome two major problems of within-species phylogenetics: (a) the scarcity of informative characters and (b) the problem of rooting the relatively similar within-species haplotypes with relatively divergent haplotype(s) from a different species. Crandall & Fitzpatrick (18) have combined the TCS method with more traditional among-species methods (see also 6). Using this combined approach, the most likely connections between intraspecific haplotypes are inferred using the TCS method, and these relationships are then constrained in a global parsimony or maximum likelihood search that includes all the haplotypes from all the species [although Hedin (61) found that unconstrained parsimony searches recover the same clades that are connected by the TCS method]. The combined approach seems very promising for dealing with polymorphism in DNA data, but whether it actually improves the accuracy of estimated interspecific trees has yet to be shown. Using data from a known bacteriophage phylogeny, Crandall (17) has shown that the TCS method by itself may outperform parsimony in some cases. Simulation and congruence studies are needed to further test the accuracy of the TCS and combined approaches at the interspecific level.

Smouse et al (127) proposed a method for estimating species phylogenies from restriction-site data when multiple individuals are sampled from each species. Their method involves estimating multiple phylogenies for each data set, each derived using a single individual to represent each species (i.e. the first phylogeny based on the first individual sampled from each species, the second phylogeny based on the second individual, etc). The species phylogeny is considered to be the "average" topology from among these trees, in accord with the idea of a species phylogeny as the "central tendency" of a diverse cloud of gene histories (78). The general approach of Smouse et al (127) seems readily applicable to DNA sequence data as well, but its accuracy relative to other methods has not been tested.

Nucleic acid sequence and restriction site data are not the only DNA data used in phylogenetic analysis, and recent years have seen increasing use of data from microsatellites and other hypervariable loci to estimate relationships among populations and closely-related species (e.g. 8, 89). Because of their rapid rate of evolution, the appropriateness of these markers for any but the most closely related taxa is questionable (66). Microsatellites are among the most slowly evolving

of these markers and may be the most appropriate for interspecific phylogeny reconstruction. Microsatellite data are similar in many ways to allozyme data (i.e., multiple unlinked loci with high levels of polymorphism and many alleles per locus). These data have been analyzed using both an *individuals as terminal taxa* approach (using, for example, the proportion of alleles shared between individuals as a measure of distance), as commonly applied to DNA sequence data, and by analyzing the allele frequencies of populations using genetic distance methods (8). Genetic distances designed for allozymes have been used on microsatellite data, but a number of allele frequency-based distances specifically designed for phylogenetic analysis of microsatellite data have recently been developed, and the accuracy of all of these distance methods have now been tested extensively using simulated microsatellite data (e.g. 52ab, 109a, 131a).

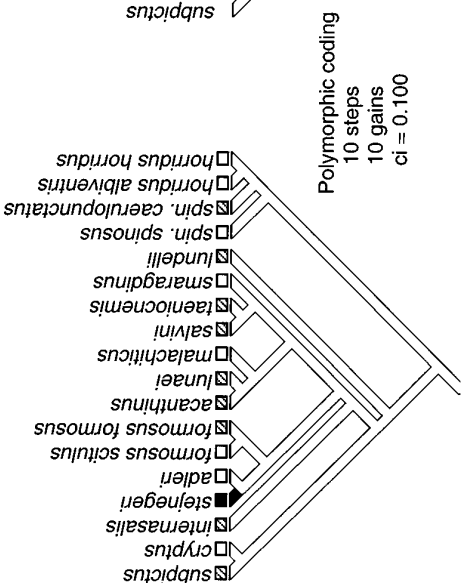
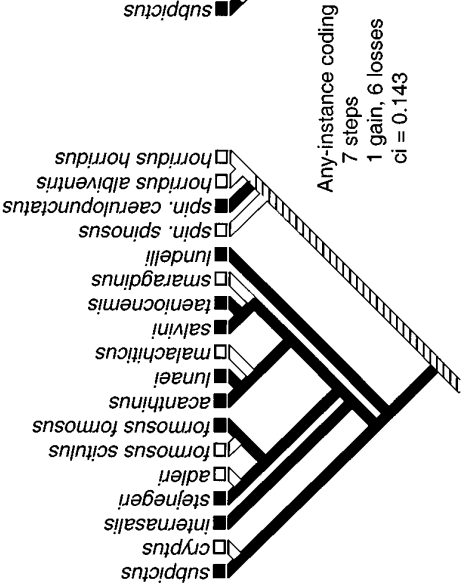
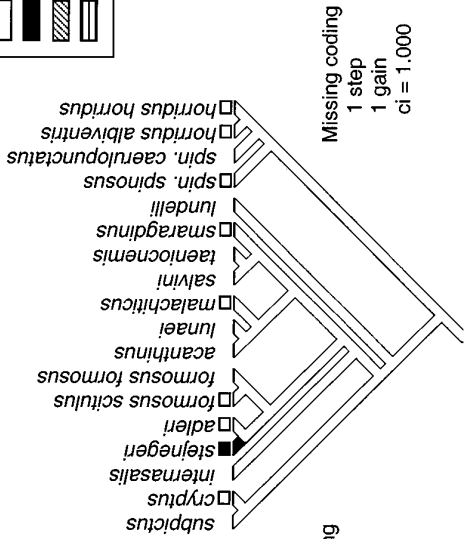
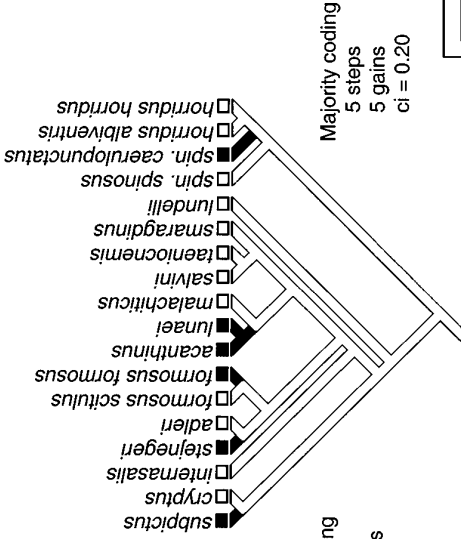
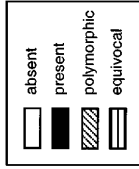
COMPARATIVE EVOLUTIONARY STUDIES

Recent years have seen burgeoning use of phylogenies in studying patterns and processes of character evolution (9, 28, 41), and there is growing interest in testing and refining the methodologies used in comparative evolutionary studies, for both continuous (e.g. 26, 83, 84) and discrete data (e.g. 21, 77, 103, 118, 119). However, there has been little discussion of the impact of intraspecific variation in characters that are the focus of comparative studies (but see 29 and 85). Whereas studies of continuous traits typically use mean values for species, intraspecific variation in discrete traits is rarely mentioned in comparative studies. Yet, different ways of treating this variation may have a profound impact on evolutionary reconstructions and inferences.

For example, Figure 7 shows the effects of different ways of coding a single character—presence of colored female belly patches in spiny lizards (*Sceloporus*)—on evolutionary inferences. Depending on how polymorphism in this one character is coded, the trait may exhibit: (a) a preponderance of losses relative to gains (6 to 1), (b) a preponderance of gains (10 to 0), (c) a high degree of homoplasy (10 changes among 18 species), or (d) no homoplasy at all (1 change). Furthermore, the trait may be inferred to have evolved in the common ancestor of the entire clade (any-instance coding) or within a single, relatively derived species (missing coding). Clearly, the choice of methodology for dealing with polymorphism in comparative studies can be extremely important.

Given that many different methodologies are available for coding polymorphism (Figure 2), which one might be the best to apply in comparative studies?

Figure 7 Different methods of coding polymorphism can produce radically different hypotheses of character evolution for the same data and tree. The presence or absence of female belly patches is mapped among 18 closely related species of spiny lizard (*Sceloporus*) using MacClade (79). Data and tree from Wiens & Reeder (141). *Sceloporus tanneri* is excluded from this example because the state for this character is unknown.



The answer may depend on the specific question being asked. Many comparative studies are concerned with the timing, number, and pattern of gains and losses of qualitative traits, which require reconstructing ancestral states and (presumably) coding polymorphism in some way, regardless of whether the reconstruction method is parsimony or likelihood.

Given its strong performance in phylogeny reconstruction, frequency coding may seem an obvious choice for ancestral state reconstructions as well. However, frequency methods work by differentially weighting changes based on trait frequencies, and differential weighting is largely meaningless for reconstructing the ancestral state of a linearly ordered trait (80). Therefore, I briefly review and evaluate some of the candidate methods.

The any-instance method, which codes a derived trait as present regardless of its frequency, is problematic (12) in that it can potentially hide reversals to the primitive state (for example, if the primitive trait was regained and present at a frequency of 99%). This severely limits the effectiveness of the method for tracking the gains and losses of polymorphic traits.

The polymorphic coding method treats a variable species as having either of the two traits, but not both (during the reconstruction, taxa are treated as having whichever state is most parsimonious). Thus, ancestral nodes are never reconstructed as being polymorphic. Instead, each instance of polymorphism within a species is treated as an independent evolutionary event (as implemented by MacClade; 79), rather than allowing the possibility that polymorphisms are inherited. This approach therefore maximizes homoplasy rather than homology. This problem also applies to coding individuals as terminals when reconstructing ancestral states, or when variable species are broken up into monomorphic units (as recommended in 29, 103).

Similarly, the missing method (coding polymorphic species as unknown) also does not allow polymorphic ancestors. But in contrast to the polymorphic method, the missing method never treats polymorphisms as homoplasies, even when they seem clearly to be the result of reversal or parallelism (as in *Sceloporus spinosus caerulopunctatus* in Figure 7). Instead, polymorphic taxa coded as missing are treated as having whatever state is most parsimonious given the reconstruction of the trait based on other taxa. This property seriously compromises the ability of this method to estimate levels of homoplasy and the pattern of gains and losses in a character of interest.

The majority method (coding the commonest state as the only state present) does allow polymorphic ancestors but hides gains and losses of rare traits. In some ways, this insensitivity to rare traits may be an advantage, because the apparent gain and loss of rare traits may be due only to sampling error (129). Yet, the majority method is disadvantageous, in part because small changes in frequency close to 50% might be due to sampling error as well.

The scaled method avoids the problems of the preceding methods, although it does not downweight rare traits. The unordered method is similar to the scaled method but is problematic in that it implies no special relationship among states. Finally, the frequency method is similar to the scaled method for the purposes of

ancestral state reconstruction, because, although it downweights rare traits, this differential weighting is meaningless in this context (80). The scaled or frequency method is recommended for reconstructing ancestral states in polymorphic characters. However, it should be noted that these methods are potentially sensitive to sampling error in this context, and very large sample sizes may be needed to confidently distinguish traits that are truly absent from those that are present as polymorphisms at low frequencies (112, 129, 144).

Instead of focusing specifically on the reconstruction of ancestral states, comparative evolutionary studies may also address questions about correlations between pairs of characters (e.g. 41, 76), or about differences in rates of change between characters and/or lineages (e.g. 47, 102). For these questions, the weighting of changes is important, and frequency methods may therefore be advantageous. For example, most nonfrequency methods (all but majority coding) would treat a change from 0 to 1% the same as a change from 0 to 99%, yet clearly more change has occurred in the latter case than in the former. Frequency methods are therefore recommended for studies of this kind. Treating polymorphic discrete traits as continuous frequencies may also facilitate the use of the many continuous data comparative methods (e.g. independent contrasts; 41), and the combination of discrete (polymorphic) and continuous traits. However, the performance of these methods using frequency data needs to be tested. There is also the need to modify comparative methods for discrete characters (e.g. 76, 104, 118) to accommodate polymorphism and frequency data.

POLYMORPHISM AND SPECIES DELIMITATION

Although it is tempting to equate systematics with phylogenetic analysis, the delimitation and description of species-level diversity is a major endeavor of systematics that is at least as important as reconstructing phylogenies. Species-delimitation is linked to the issue of polymorphism in that, for most alpha-level systematists, the main analytical task is distinguishing fixed (or nearly fixed) diagnostic features from those that are polymorphic. Curiously, while a voluminous literature has accumulated and continues to accumulate on the philosophical question of species concepts (what species are), the practical, methodological aspects of how we differentiate one species from another have received relatively little attention in the systematics literature (but see 23, 124), although species concepts and criteria for species recognition are often confounded (49). Recent authors have distinguished between character-based and tree-based approaches for delimiting species (e.g. 5), and this division is followed here. Character-based approaches delimit species based on character state distributions among geographic samples, whereas tree-based approaches generally use a phylogeny of haplotypes, individuals, or populations to infer species limits.

In this review of approaches to species-delimitation, I purposely avoid discussing the pros and cons of particular species concepts. However, my personal bias favors the evolutionary species concept (48, 123, 146), and I therefore follow

the view that species are real entities that exist in nature, regardless of whether or not there is character or tree-based evidence that allows us to detect them (49).

Character-Based Species Delimitation

Newly discovered species are generally delimited, diagnosed, and described using a character-based approach, and this approach is widely used to test the validity of described taxa. For example, when a potentially new species is discovered, it is compared to similar species already described, and “diagnostic” (generally meaning intraspecifically invariant, or non-overlapping) character states are sought to distinguish them. This describes the basic task of most practicing alpha-level systematists for the past few hundred years. There has been surprisingly little methodological advancement in this area, especially relative to the burgeoning methods of phylogenetic analysis.

Davis & Nixon (23) described an explicit character-based methodology for delimiting species, given a set of populations with unknown species boundaries, which they called population aggregation analysis (PAA). PAA involves systematically comparing character distributions among populations, aggregating sets of populations that differ only in polymorphic traits, and considering sets of populations that differ from others by at least one fixed difference (or which share no states for a given character) to be different species. As pointed out by the authors, PAA is problematic in that: (a) unless many characters are sampled, the number of species present may be underestimated and (b) unless sample sizes are large, the number of species may be overestimated (by considering traits to be fixed that are actually polymorphic). PAA has never been modified to account for these problems, or at least to detect when the data are inadequate to make a decisive resolution. However, Wiens & Servedio (144) recently proposed a statistical test to determine whether or not sufficient characters and individuals have been sampled to argue that one or more seemingly fixed characters are truly diagnostic for a given species (i.e. the state of the other species is either absent or below a given frequency).

The methodology of PAA raises an important question: Why should fixed traits be the only ones that can delimit species? Why not differences in trait frequencies? One potential reason is that fixed differences may indicate an absence of gene flow, in a way that differences in trait frequencies may not. For example, suppose we compare two putative species and find that for a given character (e.g. flower color), all the individuals of each putative species differ from the other in this character (e.g. red versus white). The most likely explanation for this consistent difference would be that no individuals of one of the putative species are breeding with the other. Conversely, if all potentially diagnostic traits are polymorphic in one or both species, the explanation may be that there is gene flow between the species, or that the species have split too recently for any differences to become fixed.

Of course, in empirical studies, “fixed” differences are inferred from a finite sample of individuals, and very large sample sizes are necessary to be confident that fixed traits are not actually polymorphic at a low frequency. For example, if

an investigator seeks a 95% probability of being able to detect a polymorphism in a putatively fixed trait occurring at a frequency of 1% in a given species, about 150 individuals would need to be sampled from that species (based on equations in 129). This estimated sample size is based on simplified assumptions about population subdivision, but more realistic conditions would likely require more sampling rather than less (112). Given the difference between the large sample sizes needed and those typically available in empirical studies, it seems unlikely that a systematist could ever say with any confidence that a diagnostic character state was actually fixed. In fact, to be truly certain that a trait was not polymorphic and present at a very low frequency would require sampling the entire population.

An alternative approach is to use frequency differences in character-based species delimitation. For example, it seems reasonable to consider a 95% frequency of red flower color in one putative species and 95% white in the other to be evidence that these taxa are distinct, although this character would not be considered informative using PAA. Surely, such large differences in frequencies must indicate that gene flow between these putative species is rare if not absent. By considering differences in frequencies, characters that would be considered uninformative for species delimitation by a strict “fixed-only” criterion could be incorporated, thus increasing the power of any test of species boundaries. Unfortunately, it is not clear what constitutes a sufficient difference in trait frequencies to be adequate for recognizing putative species as distinct, given that some discontinuity in gene flow is to be expected among conspecific populations.

Some authors have used measures of genetic distance between populations to make species-level decisions, an approach that does incorporate frequency differences between putative species. Perhaps the simplest way this can be done is to find a standard level of distance between “good” species and then apply this value to cases that are less clear (e.g. 63, 64). However, this approach has met with considerable resistance (e.g. 48, 135), partly on the grounds that it relies on an “arbitrary measure of similarity.”

A more sophisticated usage of genetic distance data (obtained from multiple populations of two or more putative species) in species delimitation involves various techniques that address the relationship between genetic and geographic distance (see review in 25). The general idea is that for conspecific populations, genetic distance should increase with geographic distance (98), but that this relationship should not hold for heterospecific populations. de Queiroz & Good (25) reviewed a number of techniques that could be used to apply the expected relationships between geographic and genetic distances to species delimitation, including the Mantel test (81) and spatial autocorrelation (15).

Porter (110) has suggested using estimates of gene flow between populations derived from population genetics (e.g. 125, 147, 148) to help determine species boundaries (i.e. certain values of F_{GT} [110] indicate that gene flow is absent or negligible between groups of populations). This approach appears to be promising for determining how much of the similarity between putative species is due

to ongoing gene flow. However, all three of these frequency-based approaches are designed primarily for allozyme data, and it is not clear how successfully they could be applied or adapted to morphological data (the data used to describe and delimit most species).

Finally, Doyle (30) proposed a nonfrequency, character-based approach for species delimitation from DNA sequence data from nuclear genes, based on the sharing of alleles in heterozygous individuals. Conspecific individuals share a set of alleles not found in other species, and the combinations of alleles that define a species are seen in the heterozygous individuals. This approach seems likely to be highly sensitive to the failure to detect heterozygous individuals with finite sample sizes.

Tree-Based Species Delimitation

Much of the recent literature on “the species problem” has focused on the implications of intraspecific phylogenies (particularly of individual organisms or haplotypes) for delimiting species, and on the conceptual nature of species (3, 5, 24, 54, 101). Most authors seem to agree that when all the individuals sampled of a putative species appear as each others’ closest relatives on a gene tree or trees (relative to other putative species), this is support for the presence of a distinct species. However, when the individuals of a species-level taxon fail to cluster together, the results are more difficult to interpret. Possible explanations include: (a) interbreeding between the putative species and other taxa (i.e. possibly suggesting the putative species is invalid); (b) incomplete lineage sorting of ancestral polymorphisms [i.e. possibly meaning the putative species is valid but is very recently diverged (e.g. 99)]; (c) the presence of multiple, unrelated species hidden by previous taxonomy (e.g. 145); and (d) insufficient data, such that the estimated phylogeny fails to match the gene tree.

Another issue in the tree-based approach is that it may be difficult to delimit species without reference to some extrinsic character data (e.g. 30). For example, given only a phylogeny of haplotypes, how do we determine which lineages are species and which are merely clades within species? One approach is to assume that within-species phylogenies will not be concordant between genes (because of gene flow and lineage sorting), sample multiple unlinked genes, and consider species boundaries to be the points that are congruent between gene trees (3, 5). However, the theory behind this approach has not been well explored [for example, how many genes need to be sampled before we can say that relationships are truly congruent or discordant? (3)]. A similar approach assumes that haplotype phylogenies between species will be concordant with geography, but that haplotype phylogenies within species will show discordance with geography (i.e. individuals from the same locality or population will not cluster together, suggesting gene flow between populations). Discordance between geography and gene phylogeny forms the basis for certain measures of gene flow (e.g. 126), but these methods have not been widely applied (if at all) to making species-level decisions.

Finally, it should be noted that tree-based species-delimitation is not restricted to gene trees from DNA sequence data; many authors have applied a tree-based approach to allozyme data to help infer species boundaries; using populations as terminal units and testing whether or not putative conspecifics cluster together (e.g. 53). Morphological data can be used this way as well (67), and the same morphological and/or allozyme data can be analyzed from both a tree and character-based perspective. The congruence and incongruence of population-level trees from diverse types of data (e.g. morphology, allozymes, and DNA sequences) with geography and with groups recognized by tree and character-based perspectives may be particularly revealing about species limits (3, 106, 124).

CONCLUDING REMARKS

The question of how we analyze intraspecific variation in between-species evolutionary studies lies at the intersection of the fields of population genetics, comparative biology, and systematics. In this review, I have discussed three fundamental areas of phylogenetic biology—phylogeny reconstruction, species-delimitation, and comparative studies of character evolution—where polymorphism can have a major impact. I have argued that treating polymorphic traits directly as frequencies may improve analyses in all three areas, although the methodological treatment of polymorphism in the latter two areas is very poorly explored. The frequency of a given trait within a species or population is the most basic parameter of population genetics, but one that is ignored by many systematic and comparative biologists. Future progress in this area may come not only from applying frequency information to additional questions, but also by incorporating additional information on within-species evolutionary processes into systematic and comparative analyses (e.g. 58, 110, 132).

ACKNOWLEDGMENTS

I thank Paul Chippindale, Keith Crandall, Kevin de Queiroz, Zhexi Luo, David Posada, Steve Poe, Maria Servedio, H Bradley Shafer, Andrew Simons, and Jack Sites, Jr. for helpful comments on the manuscript.

Visit the Annual Reviews home page at <http://www.AnnualReviews.org>

LITERATURE CITED

1. Archie JW. 1985. Methods for coding variable morphological features for numerical taxonomic analysis. *Syst. Zool.* 34:326–345
2. Archie JW, Simon C, Martin A. 1989. Small sample size does decrease the stability of dendrograms calculated from allozyme-frequency data. *Evolution* 43:678–83
3. Avise JC, Ball RM. 1990. Principles of genealogical concordance in species concepts and biological taxonomy. *Oxford Surv. Evol. Biol.* 7:45–67

4. Avise JC, Arnold J, Ball RM, Bermingham E, Lamb T, et al. 1987. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annu. Rev. Ecol. Syst.* 18:489–522
5. Baum DA, Donoghue MJ. 1995. Choosing among alternative “phylogenetic” species concepts. *Syst. Bot.* 20:560–73
6. Benabib M, Kjer KM, Sites JW Jr. 1997. Mitochondrial DNA sequence-based phylogeny and the evolution of viviparity in the *Sceloporus scalaris* group (Reptilia: Squamata). *Evolution* 51:1262–75
7. Berlocher SH, Swofford DL. 1997. Searching for phylogenetic trees under the frequency parsimony criterion: an approximation using generalized parsimony. *Syst. Biol.* 46:211–15
8. Bowcock AM, Ruiz-Linares A, Tomfohrde J, Minch E, Kidd JR, Cavalli-Sforza LL. 1994. High resolution of human evolutionary trees with polymorphic microsatellites. *Nature* 368:455–57
9. Brooks DR, McLennan DA. 1991. *Phylogeny, Ecology, and Behavior: A Research Program in Comparative Biology*. Chicago, IL: Univ. Chicago Press
10. Brumfield RT, Capparella AP. 1996. Historical diversification of birds in northwestern South America: a molecular perspective on the role of vicariant events. *Evolution* 50:1607–24
11. Buth DG. 1984. The application of electrophoretic data in systematic studies. *Annu. Rev. Ecol. Syst.* 15:501–22
12. Campbell JA, Frost DR. 1993. Anguillid lizards of the genus *Abronia*: revisionary notes, description of four new species, a phylogenetic analysis, and key. *Bull. Am. Mus. Nat. Hist.* 216:1–121
13. Cavalli-Sforza LL, Edwards AWF. 1967. Phylogenetic analysis: models and estimation procedures. *Am. J. Hum. Genet.* 19:233–57
14. Chu PC. 1998. A phylogeny of the gulls (Aves: Larinae) inferred from osteological and integumentary characters. *Cladistics* 14:1–43
15. Cliff AD, Ord JK. 1973. *Spatial Autocorrelation*. London: Pion
16. Coloma LA. 1997. *Morphology, systematics and phylogenetic relationships among frogs of the genus Atelopus (Anura: Bufonidae)*. PhD thesis, Univ. Kansas. 287 pp
17. Crandall KA. 1994. Intraspecific cladogram estimation: accuracy at higher levels of divergence. *Syst. Biol.* 43:222–35
18. Crandall KA, Fitzpatrick JF. 1996. Crayfish molecular systematics: using a combination of procedures to estimate phylogeny. *Syst. Biol.* 45:1–26
19. Crandall KA, Templeton AR, Sing CF. 1994. Intraspecific phylogenetics: problems and solutions. In *Models in Phylogeny Reconstruction*, ed. RW Scotland, DJ Siebert, DM Williams, pp. 273–97. Oxford, UK: Clarendon
20. Crother BI. 1990. Is “some better than none” or do allele frequencies contain phylogenetically useful information? *Cladistics* 6:277–81
21. Cunningham CW, Omland KE, Oakley TH. 1998. Reconstructing ancestral character states: a critical reappraisal. *Trends Ecol. Evol.* 13:361–66
22. Darwin C. 1859. *On the Origin of Species*. Cambridge: Harvard Univ. Press
23. Davis JI, Nixon KC. 1992. Populations, genetic variation, and the delimitation of phylogenetic species. *Syst. Biol.* 41:421–35
24. de Queiroz K, Donoghue MJ. 1988. Phylogenetic systematics and the species problem. *Cladistics* 4:317–38
25. de Queiroz K, Good DA. 1997. Phenetic clustering in biology: a critique. *Q. Rev. Biol.* 72:3–30
26. Diaz-Uriarte R, Garland T. 1996. Testing hypotheses of correlated evolution using phylogenetically independent contrasts: sensitivity to deviations from Brownian motion. *Syst. Biol.* 45:27–47
27. Domning DP. 1994. A phylogenetic anal-

- ysis of the Sirenia. *Proc. San Diego Soc. Nat. Hist.* 29:177–89
28. Donoghue MJ. 1989. Phylogenies and the analysis of evolutionary sequences, with examples from seed plants. *Evolution* 43:1137–56
29. Donoghue MJ, Ackerly DD. 1996. Phylogenetic uncertainties and sensitivity analysis in comparative biology. *Philos. Trans. R. Soc. Lond. B* 351:1241–49
30. Doyle JJ. 1995. The irrelevance of allele tree topology for species delimitation, and a non-topological alternative. *Syst. Bot.* 20:574–588
31. Echelle AA, Echelle AS. 1998. Evolutionary relationships of pupfishes in the *Cyprinodon eximius* complex (Atherinomorpha: Cyprinodontiformes). *Copeia* 1998:852–65
32. Farris JS. 1966. Estimation of conservatism of characters by constancy within biological populations. *Evolution* 20:587–91
33. Farris JS. 1981. Distance data in phylogenetic analysis. In *Advances in Cladistics*, Volume 1. *Proceedings of the First Meeting of the Willi Hennig Society*, ed. VA Funk, DR Brooks, pp. 3–23. New York: New York Bot. Gard.
34. Farris JS. 1985. Distance data revisited. *Cladistics* 1:67–85
35. Farris JS. 1986. Distances and cladistics. *Cladistics* 2:144–57
36. Farris JS. 1990. Phenetics in camouflage. *Cladistics* 6:91–100
37. Felsenstein J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. *Syst. Zool.* 27:401–10
38. Felsenstein J. 1981. Evolutionary trees from gene frequencies and quantitative characters: finding maximum likelihood estimates. *Evolution* 35:1229–42
39. Felsenstein J. 1984. Distance methods for inferring phylogenies: a justification. *Evolution* 38:16–24
40. Felsenstein J. 1985. Phylogenies from gene frequencies: a statistical problem. *Syst. Zool.* 34:300–11
41. Felsenstein J. 1985. Phylogenies and the comparative method. *Am. Nat.* 125:1–15
42. Felsenstein J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39:783–91
43. Felsenstein J. 1986. Distance methods: a reply to Farris. *Cladistics* 2:130–43
44. Felsenstein J. 1988. Phylogenies and quantitative characters. *Annu. Rev. Ecol. Syst.* 19:445–71
45. Felsenstein J. 1992. Estimating effective population sizes from samples of sequences: inefficiency of pairwise and segregation sites as compared to phylogenetic estimates. *Genet. Res. Camb.* 56:139–57
46. Fitch WM, Margoliash E. 1967. Construction of phylogenetic trees. *Science* 155:279–84
47. Foster SA, Cresko WA, Johnson KP, Tlusty MU, Willmott HE. 1996. Patterns of homoplasy in behavioral evolution. In *Homoplasy. The Recurrence of Similarity in Evolution*, ed. MJ Sanderson, L Hufford, pp. 245–69. San Diego, CA: Academic
48. Frost DR, Hillis DM. 1990. Species in concept and practice: herpetological applications. *Herpetologica* 46:87–104
49. Frost DR, Kluge AG. 1994. A consideration of epistemology in systematic biology, with special reference to species. *Cladistics* 10:259–94
50. Gaines MS, McLenaghan LR, Rose RK. 1978. Temporal patterns of allozyme variation in fluctuating populations of *Microtus ochrogaster*. *Evolution* 32:723–39
51. Golding B, Felsenstein J. 1990. A maximum likelihood approach to the detection of selection from a phylogeny. *J. Mol. Evol.* 31:511–23
52. Goldman N. 1988. Methods for discrete coding of variable morphological features for numerical analysis. *Cladistics* 4:59–71
- 52a. Goldstein DB, Linares AR, Cavalli-Sforza LL, Feldman MW. 1995. An evaluation of

- genetic distances for use with microsatellite data. *Genetics* 139:463–71
- 52b. Golstein DB, Pollock DD. 1997. Launching microsatellites: a review of mutation processes and methods of phylogenetic inference. *J. Hered.* 88:335–42
53. Good DA. 1989. Hybridization and cryptic species in *Dicamptodon* (Caudata: Dicamptodontidae). *Evolution* 43:728–44
54. Graybeal A. 1995. Naming species. *Syst. Biol.* 44:237–50
55. Grismer LL. 1999. Phylogeny, taxonomy, and biogeography of *Cnemidophorus hyperythrus* and *C. ceralbensis* (Squamata: Teiidae) in Baja California, Mexico. *Herpetologica* 55: In press
56. Gutberlet RL Jr. 1998. The phylogenetic position of the Mexican black-tailed pitviper (Squamata: Viperidae: Crotalinae). *Herpetologica* 54:184–206
57. Halloy M, Etheridge RE, Burghardt GM. 1998. To bury in sand: phylogenetic relationships among lizard species of the *boulengeri* group, *Liolaemus* (Reptilia: Squamata: Tropiduridae), based on behavioral characters. *Herpetol. Monogr.* 12:1–37
58. Hansen TF, Martins EP. 1996. Translating between microevolutionary process and macroevolutionary patterns: the correlation structure of interspecific data. *Evolution* 50:1404–17
59. Harvey M, Gutberlet RL Jr. 1999. A phylogenetic analysis of the Tropidurini (Squamata: Tropiduridae) using new characters of squamation and epidermal microstructure. *Zool. J. Linn. Soc.* In press
60. Harvey PH, Pagel MD. 1991. *The Comparative Method in Evolutionary Biology*. Oxford, UK: Oxford Univ. Press
61. Hedin MC. 1997. Speciation history in a diverse clade of habitat specialized spiders (Araneae: Nesticidae: *Nesticus*): inferences from geographic based sampling. *Evolution* 51:1929–45
62. Hey J. 1993. Using phylogenetic trees to study speciation and extinction. *Evolution* 46:627–40
63. Highton R. 1989. Biochemical evolution in the slimy salamanders of the *Plethodon glutinosus* complex in the eastern United States. Part I. Geographic protein variation. *Univ. Ill. Biol. Monogr.* 57:1–78
64. Highton R. 1998. Is *Ensatina escholtzii* a ring-species? *Herpetologica* 54:254–78
65. Hillis DM. 1995. Approaches for assessing phylogenetic accuracy. *Syst. Biol.* 44:3–16
66. Hillis DM, Mable BK, Moritz C. 1996. Applications of molecular systematics: the state of the field and a look to the future. In *Molecular Systematics*, ed. DM Hillis, C Moritz, B Mable, pp. 515–43. Sunderland, MA: Sinauer. 2nd ed.
67. Hollingsworth BD. 1998. The systematics of chuckwallas (*Sauromalus*) with a phylogenetic analysis of other iguanid lizards. *Herpetol. Monogr.* 12:38–191
68. Huelsenbeck JP, Hillis DM. 1993. Success of phylogenetic methods in the four-taxon case. *Syst. Biol.* 42:247–64
69. Huelsenbeck JP. 1995. The performance of phylogenetic methods in simulation. *Syst. Biol.* 44:17–48.
70. Huelsenbeck JP. 1998. Systematic bias in phylogenetic analysis: is the Strepsiptera problem solved? *Syst. Biol.* 47:519–37
71. Jones TR, Kluge AG, Wolf AJ. 1993. When theories and methodologies clash: a phylogenetic reanalysis of the North American ambystomatid salamanders (Caudata: Ambystomatidae). *Syst. Biol.* 42:92–102
72. Kim J, Burgman MA. 1988. Accuracy of phylogenetic-estimation methods under unequal evolutionary rates. *Evolution* 42:596–602
73. Kimura M. 1955. Random genetic drift in multi-allelic locus. *Evolution* 9:419–35
74. Kluge AG. 1989. A concern for evidence and a phylogenetic hypothesis among

- Epicrates* (Boidae, Serpentes). *Syst. Zool.* 38:7–25
75. Mabee PM, Humphries J. 1993. Coding polymorphic data: examples from allozymes and ontogeny. *Syst. Biol.* 42:166–81
76. Maddison WP. 1990. A method for testing the correlated evolution of two binary characters: are gains or losses concentrated on certain branches of a phylogenetic tree? *Evolution* 44:539–57
77. Maddison WP. 1995. Calculating the probability distributions of ancestral states reconstructed by parsimony on phylogenetic trees. *Syst. Biol.* 44:474–81
78. Maddison WP. 1997. Gene trees in species trees. *Syst. Biol.* 46:523–36
79. Maddison WP, Maddison DR. 1992. MacClade Ver. 3.0. *Analysis of Phylogeny and Character Evolution*. Sunderland, MA: Sinauer
80. Maddison WP, Slatkin M. 1990. Parsimony reconstructions of ancestral states do not depend on the relative distances between linearly-ordered states. *Syst. Zool.* 39:175–78
81. Mantel N. 1967. The detection of disease clustering and a generalized regression approach. *Cancer Res.* 27:209–20
82. Martins EP (ed.). 1996. *Phylogenies and the Comparative Method in Animal Behavior*. Oxford, UK: Oxford Univ. Press
83. Martins EP. 1996. Phylogenies, spatial autoregression, and the comparative method: a computer simulation test. *Evolution* 1750–65
84. Martins EP, Garland T Jr. 1991. Phylogenetic analyses of the correlated evolution of continuous characters: a simulation study. *Evolution* 45:534–57
85. Martins EP, Hansen TF. 1997. Phylogenies and the comparative method: a general approach to incorporating phylogenetic information into the analysis of interspecific data. *Am. Nat.* 149:646–67
86. Mayr E. 1969. *Principles of Systematic Zoology*. New York: McGraw-Hill
87. McGuire JA. 1996. Phylogenetic systematics of crotaphytid lizards (Reptilia: Iguania: Crotaphytidae). *Bull. Carnegie Mus. Nat. Hist.* 32:1–143
88. Mendoza-Quijano F, Flores-Villela O, Sites JW Jr. 1998. Genetic variation, species status, and phylogenetic relationships in rose-bellied lizards (*variabilis* group) of the genus *Sceloporus* (Squamata: Phrynosomatidae). *Copeia* 1998:354–66
89. Meyer E, Wiegand P, Rand SP, Kuhlmann D, Brack M, Brinkman B. 1995. Microsatellite polymorphisms reveal phylogenetic relationships in primates. *J. Mol. Evol.* 41:10–14
90. Mickevich MF. 1978. Taxonomic congruence. *Syst. Zool.* 27:143–58
91. Mickevich MF, Johnson MS. 1976. Congruence between morphological and allozyme data in evolutionary inference and character evolution. *Syst. Zool.* 25:260–70
92. Mickevich MF, Mitter C. 1981. Treating polymorphic characters in systematics: a phylogenetic treatment of electrophoretic data. In *Advances in Cladistics*. Volume 1. *Proceeding of the first meeting of the Willi Hennig Society*, ed. V. A. Funk and D. R. Brooks, pp. 45–58. New York: New York Bot. Gard.
93. Mickevich MF, Mitter C. 1983. Evolutionary patterns in allozyme data: a systematic approach. In *Advances in Cladistics*. Vol. 2. *Proceeding of the Second Meeting of the Willi Hennig Society*, ed. NI Platnick, VA Funk, pp. 169–76. New York: Columbia Univ. Press
94. Mink DG, Sites JW Jr. 1996. Species-limits, phylogenetic relationships, and origins of viviparity in the *scalaris* complex of the lizard genus *Sceloporus* (Phrynosomatidae: Sauria). *Herpetologica* 52:551–71
95. Miyamoto MM, Fitch WM. 1995. Testing species phylogenies and phylogenetic methods with congruence. *Syst. Biol.* 44:64–76

96. Murphy RW. 1993. The phylogenetic analysis of allozyme data: invalidity of coding alleles by presence/absence and recommended procedures. *Biochem. Syst. Ecol.* 21:25–38
97. Murphy RW, Doyle KD. 1998. Phylogenetics: frequencies and polymorphic characters in genealogical estimation. *Syst. Biol.* 47:737–61
98. Nei M. 1972. Genetic distance between populations. *Am. Nat.* 106:238–92
99. Neigel JE, Avise JC. 1986. Phylogenetic relationships of mitochondrial DNA under various demographic models of speciation. In *Evolutionary Processes and Theory*, ed. E Nevo, S Karlin, pp. 515–34. New York: Academic
100. Nielsen R, Mountain JL, Huelsenbeck JP, Slatkin M. 1998. Maximum likelihood estimation of population divergence times and population phylogeny in models without mutation. *Evolution* 52:669–77
101. Olmstead RG. 1995. Species concepts and plesiomorphic species. *Syst. Bot.* 20: 623–30
102. Omland KE. 1997a. Correlated rates of molecular and morphological evolution. *Evolution* 5:1381–93
103. Omland KE. 1997b. Examining two standard assumptions of ancestral reconstructions: repeated loss of dichromatism in dabbling ducks (Anatini). *Evolution* 5:1636–46
104. Pagel MD. 1994. Detecting correlated evolution on phylogenies: a general method for the comparative analysis of discrete characters. *Proc. R. Soc. Lond. B Biol. Sci.* 255:37–45
105. Pamilo P, Nei M. 1988. Relationships between gene trees and species trees. *Mol. Biol. Evol.* 5:568–83
106. Patton JL, Smith MF. 1994. Paraphyly, polyphyly, and the nature of species boundaries in pocket gophers (genus *Thomomys*). *Syst. Biol.* 43:11–26
107. Pimentel RA, Riggins, R. 1987. The nature of cladistic data. *Cladistics* 3:201–9
108. Poe S. 1998. Skull characters and the cladistic relationships of the Hispaniolan dwarf ig *Anolis*. *Herpetol. Mon.* 12:192–236
109. Poe S, Wiens JJ. 2000. Character selection and the methodology of morphological phylogenetics. In *Phylogenetic Analysis of Morphological Data*, ed. JJ Wiens, Washington, DC: Smithsonian Press. In press
- 109a. Pollock DD, Bergman A, Feldman MW, Goldstein DB. 1998. Microsatellite behavior with range constraints: parameter estimation and improved distances for use in phylogenetic reconstruction. *Theor. Popul. Biol.* 53:256–71
110. Porter AH. 1990. Testing nominal species boundaries using gene flow statistics: the taxonomy of two hybridizing admiral butterflies (*Limenitis*: Nymphalidae). *Syst. Zool.* 39:148–61
111. Prober S, Bell JC, Moran G. 1990. A phylogenetic and allozyme approach to understanding rarity in the “green ash” eucalypts (Myrtaceae). *Plant Syst. Evol.* 172:99–118
112. Rannala B. 1995. Polymorphic characters and phylogenetic analysis: a statistical perspective. *Syst. Biol.* 44:421–29
113. Reeder TW, Wiens JJ. 1996. Evolution of the lizard family Phrynosomatidae as inferred from diverse types of data. *Herpetol. Mon.* 10:43–84
114. Reynolds J, Weir BS, Cockerham CC. 1983. Estimation of coancestry coefficient: basis for a short-term genetic distance. *Genetics* 105:767–79
115. Rogers JS. 1986. Deriving phylogenetic trees from allele frequencies: a comparison of nine genetic distances. *Syst. Zool.* 35:297–310
116. Rohlf FJ, Wooten MC. 1988. Evaluation of the restricted maximum-likelihood method for estimating phylogenetic trees using simulated allele-frequency data. *Evolution* 42:581–95

117. Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 4:406–25
118. Schluter D, Price T, Mooers AØ, Ludwig D. 1997. Likelihood of ancestor states in adaptive radiation. *Evolution* 51:1699–1711
119. Schultz TR, Crocrot RB, Churchill GA. 1996. The reconstruction of ancestral character states. *Evolution* 50:504–11
120. Shaffer HB, Clark JM, Kraus F. 1991. When molecules and morphology clash: a phylogenetic analysis of the North American ambystomatid salamanders (Caudata: Ambystomatidae). *Syst. Zool.* 40:284–303
121. Siddall ME. 1998. Success of parsimony in the four–taxon case: long–branch repulsion by likelihood in the Farris Zone. *Cladistics* 14:209–220
122. Siddall ME, Kluge AG. 1997. Probabilism and phylogenetic inference. *Cladistics* 13:313–36
123. Simpson GG. 1961. *Principles of Animal Taxonomy*. New York: Columbia Univ. Press
124. Sites JW Jr, Crandall KA. 1997. Testing species boundaries in biodiversity studies. *Conserv. Biol.* 11:1289–197
125. Slatkin M, Barton NH. 1989. A comparison of three indirect methods for estimating average levels of gene flow. *Evolution* 43:1349–68
126. Slatkin M, Maddison WP. 1989. A cladistic measure of gene flow inferred from the phylogenies of alleles. *Genetics* 123:603–13
127. Smouse PE, Dowling TE, Tworek JA, Hoeh WR, Brown WM. 1991. Effects of intraspecific variation on phylogenetic inference: a likelihood analysis of mtDNA restriction site data in cyprinid fishes. *Syst. Zool.* 40:393–409
128. Strait D, Moniz M, Strait P. 1996. Finite mixture coding: a new approach to coding continuous characters. *Syst. Biol.* 45:67–78
129. Swofford DL, Berlocher SH. 1987. Inferring evolutionary trees from gene frequency data under the principle of maximum parsimony. *Syst. Zool.* 36:293–325
130. Swofford DL, Olsen GJ, Waddell PJ, Hillis DM. 1996. Phylogeny reconstruction. In *Molecular Systematics*, ed. DM Hillis, C Moritz, B Mable, pp. 407–514. Sunderland, MA: Sinauer. 2nd ed.
131. Takahata N. 1989. Gene genealogy in three related populations: consistency probability between gene and population trees. *Genetics* 122:957–66
- 131a. Takezaki N, Nei M. 1996. Genetic distances and reconstruction of phylogenetic trees from microsatellite data. *Genetics* 144:389–99
132. Templeton AR, Crandall KA, Sing CF. 1992. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics* 132:619–33
133. Thiele K. 1993. The holy grail of the perfect character: the cladistic treatment of morphometric data. *Cladistics* 9:275–304
134. Vrana P, Wheeler W. 1992. Individual organisms as terminal entities: laying the species problem to rest. *Cladistics* 8:67–72
135. Wake DB, Schneider CJ. 1998. Taxonomy of the plethodontid salamander genus *Ensatina*. *Herpetologica* 54:279–98
136. Wiens JJ. 1993. Phylogenetic systematics of the tree lizards (genus *Urosaurus*). *Herpetologica* 44:399–420
137. Wiens JJ. 1995. Polymorphic characters in phylogenetic systematics. *Syst. Biol.* 44:482–500
138. Wiens JJ. 1998. Testing phylogenetic methods with tree congruence: phylogenetic analysis of polymorphic morphological characters in phrynosomatid lizards. *Syst. Biol.* 47:411–28

139. Wiens JJ. Reconstructing phylogenies from allozyme data: comparing method performance with congruence. *Biol. J. Linn. Soc.* Submitted
140. Wiens JJ. 2000. Coding morphological variation for phylogenetic analysis: analyzing polymorphism and interspecific variation in higher taxa. In *Phylogenetic Analysis of Morphological Data*, ed. JJ Wiens. Washington, DC: Smithsonian. In press
141. Wiens JJ, Reeder TW. 1997. Phylogeny of the spiny lizards (*Sceloporus*) based on molecular and morphological evidence. *Herpetol. Mon.* 11:1–101
142. Wiens JJ, Servedio MR. 1997. Accuracy of phylogenetic analysis including and excluding polymorphic characters. *Syst. Biol.* 46:332–45
143. Wiens JJ, Servedio MR. 1998. Phylogenetic analysis and intraspecific variation: performance of parsimony, distance, and likelihood methods. *Syst. Biol.* 47:228–53
144. Wiens JJ, Servedio MR. Species delimitation in systematics: inferring “fixed” diagnostic differences between species. *Proc. R. Soc. Lond. B Biol. Sci.* Submitted
145. Wiens JJ, Reeder TW, Nieto A. 1999. Molecular phylogenetics and evolution of sexual dichromatism among populations of the Yarrow’s Spiny Lizard (*Sceloporus jarrovii*). *Evolution*. In press
146. Wiley EO. 1978. The evolutionary species concept reconsidered. *Syst. Zool.* 27:17–26
147. Wright S. 1931. Evolution in Mendelian populations. *Genetics* 16:97–159
148. Wright S. 1978. *Evolution and the Genetics of Populations*. Vol. IV. *Variation within and among Natural Populations*. Chicago: Univ. Chicago Press
149. Wu C–I. 1991. Inferences of species phylogeny in relation to segregation of ancient polymorphisms. *Genetics* 127:429–35
150. Yang Z. 1996. Phylogenetic analysis using parsimony and likelihood methods. *J. Mol. Evol.* 42:294–307