Biological Journal of the Linnean Society (2000), *70:* 613–632. With 6 figures.

doi:10.1006/bijl.1999.0420, available online at http://www.idealibrary.com on IDEAL®

Reconstructing phylogenies from allozyme data: comparing method performance with congruence

JOHN J. WIENS

Section of Amphibians and Reptiles, Carnegie Museum of Natural History, Pittsburgh, Pennsylvania 15213–4080, U.S.A.

Received 6 June 1999; accepted for publication 30 October 1999

Allozyme data are widely used to infer the phylogenies of populations and closely-related species. Numerous parsimony, distance, and likelihood methods have been proposed for phylogenetic analysis of these data; the relative merits of these methods have been debated vigorously, but their accuracy has not been well explored. In this study, I compare the performance of 13 phylogenetic methods (six parsimony, six distance, and continuous maximum likelihood) by applying a congruence approach to eight allozyme data sets from the literature. Clades are identified that are supported by multiple data sets other than allozymes (e.g. morphology, DNA sequences), and the ability of different methods to recover these 'known' clades is compared. The results suggest that (1) distance and likelihood methods generally outperform parsimony methods, (2) methods that utilize frequency data tend to perform well, and (3) continuous maximum likelihood is among the most accurate methods, and appears to be robust to violations of its assumptions. These results are in agreement with those from recent simulation studies, and help provide a basis for empirical workers to choose among the many methods available for analysing allozyme characters.

ADDITIONAL KEY WORDS:—accuracy – distance methods – maximum likelihood – parsimony methods – polymorphic characters – frequency data.

CONTENTS

E-mail: wiensj@carnegiemuseums.org

0024–4066/00/080613+20 $35.00/0

INTRODUCTION

Allozyme data, typically consisting of allele frequencies obtained from starch gel electrophoresis of proteins, are an important source of characters for reconstructing phylogenies among conspecific populations and closely-related species. Despite the increasing use of DNA sequence data in phylogenetics, allozyme data remain widely used in systematic and evolutionary studies (e.g. Brumfield & Capparella, 1996; Ruedi, 1996; Weller, 1996; Mardulyn, Milinkovitch & Pasteels, 1997; Cannatella *et al.*, 1998; Marko, 1998; Nyman, Roininen & Vuorinen, 1998; Klauta *et al.*, 1999) and have many advantages. For example, allozyme data consist of multiple unlinked nuclear loci, with each locus providing an independent estimate of the species phylogeny. Therefore, in contrast to results from mitochondrial DNA sequences, allozyme data are less likely to be systematically misled by mismatches between gene trees and species trees (e.g. Pamilo & Nei, 1988). Furthermore, it is relatively cheap and easy to survey allozyme variation for a large number of individuals (Hillis, Mable & Moritz, 1996).

A problematic aspect of the use of allozyme data, however, is that there is long-standing and continuing controversy as to the preferred method for their phylogenetic analysis (e.g. Mickevich & Johnson, 1976; Mickevich, 1978; Mickevich & Mitter, 1981, 1983; Farris, 1981, 1985, 1986; Buth, 1984; Felsenstein, 1984, 1985a, 1986; Swofford & Berlocher, 1987; Crother, 1990; de Queiroz, 1992; Jones, Kluge & Wolf, 1993; Mabee & Humphries, 1993; Murphy, 1993; Wiens 1995; Swofford *et al.*, 1996; Murphy & Doyle, 1998). Major questions include: (1) is it better to analyze allozyme data with parsimony or genetic distance methods? (2) are allele frequencies sufficiently stable over space and time to be used in phylogenetic analysis directly, or should they be converted to qualitative 'presence/absence' data? and (3) can continuous maximum likelihood be applied to allozyme data (as suggested by Felsenstein, 1981), or are the assumptions of this method too restrictive? Although these issues have been vigorously debated in the literature, there have been few attempts to address these questions quantitatively with empirical data (e.g. Mickevich, 1978; Wiens, 1995). Thus, the question remains: what are the best methods for phylogenetic analysis of allozyme data?

Congruence studies provide an important framework for comparing the relative performance of phylogenetic methods (e.g. Mickevich, 1978; Allard & Miyamoto, 1992; Miyamoto & Fitch, 1995). In congruence studies one typically assumes that clades supported by many different lines of evidence (e.g. morphology, DNA sequences) can effectively be considered to be 'known' (Miyamoto & Fitch, 1995). Given the assumption that these congruent, well-supported relationships are true, subsets of the total data can be analyzed using a variety of methods, to see which methods can yield the 'correct' phylogeny with a finite sample of characters. Congruence studies provide a useful complement to simulation studies (Allard &

Miyamoto, 1992). Simulations allow one to test the accuracy of methods (i.e. how well they recover the known, simulated phylogeny) under a variety of simplified conditions, and it is generally easy to examine the causes of method success and failure by careful manipulation of simulation parameters (e.g. Huelsenbeck & Hillis, 1993). Unfortunately, it is difficult to tell how applicable these results are to real data, because simulations always involve many simplifying assumptions about evolutionary processes (Miyamoto & Fitch, 1995). Congruence studies, on the other hand, use real data generated by natural processes, but they do not necessarily allow one to determine why methods behave as they do, and the correct phylogeny is inferred rather than truly known.

In this study, I use congruence analyses of eight allozyme data sets from a variety of animals (e.g. birds, mammals, reptiles, amphibians, insects) to compare 13 phylogenetic methods, including most of the widely used parsimony, likelihood, and distance methods. A congruence study for allozyme data is particularly important because recent simulation studies most applicable to allozyme data (i.e. those simulating changes in allele frequencies at unlinked loci) used relatively simple models that did not incorporate important features of allozyme data, such as the introduction of new alleles through mutation (Kim & Burgman, 1988; Rohlf & Wooten, 1988; Wiens & Servedio, 1997, 1998). An earlier simulation study (Nei, Tajima & Tateno, 1983), which did incorporate mutation, compared only distance methods. Furthermore, there are many processes at work on allele frequencies in nature that would be difficult to model simultaneously, but which might impact the relative performance of phylogenetic methods, including mutation, selection, drift, and geographic subdivision and migration among populations before, during, and after speciation. The results of congruence analyses provide an important line of evidence for researchers choosing a phylogenetic method to apply to a given empirical problem, and a crucial 'reality check' for results based on simulation studies (Allard & Miyamoto, 1992).

MATERIAL AND METHODS

Eight allozyme data sets were used in this study, representing a diversity of animal taxa. These data sets were chosen largely because, for each group, at least two non-allozyme data sets (e.g. morphology, DNA sequences) were available that yielded congruent relationships. That is, these non-allozyme phylogenies were congruent with each other, but not necessarily congruent with the phylogeny from allozyme data. Although a plethora of allozyme data sets exist in the literature, the necessity of having two other data sets for the same taxa eliminated many from consideration, including any plant groups. I required that two other data sets be available because of the possibility that clades supported by only one data set might be due to random (e.g. due to sampling too few characters) or systematic error (e.g. mismatch between gene and species trees for DNA sequence data). The presence of two independent, congruent data sets supporting a given clade makes these possibilities less likely. Although the 'known' clades were generally strongly supported by one or both of the non-allozyme data sets (i.e. as determined by bootstrapping; Felsenstein, 1985b), it remains an at least theoretical possibility that some of these 'known' clades are not correct. Nevertheless, shared phylogenetic history seems the most likely explanation for the congruence between diverse data sets (Miyamoto & Fitch, 1995).

TABLE 1. Basic descriptions of the allozyme data sets used in this study (OTU = operational taxonomic unit, or terminal taxon). The number of alleles per locus and the number of individuals per locus are averaged across OTUs and loci, respectively. The proportion of polymorphic loci refers to the number of loci that vary within one or more species. Sample sizes were unavailable for the *Geomydoecus* data set. Additional statistics (e.g. mean heterozygosity) would be difficult to calculate with the data provided in these papers. The number of OTUs is sometimes greater than suggested by Figs 1–3 because some species are represented by multiple populations

Taxon	No. OTUs	No. loci	Alleles per locus	Individuals per OTU	Polymorphic loci (%)
Aneides	6	28	4.75	5.00	60.7
Physalemus	10	25	5.08	9.00	72.0
Sand lizards	9	26	5.35	13.44	96.2
Ammodramus	5	18	3.10	4.50	72.2
Peromyscus	12	19	2.42	14.08	78.9
Geomydoecus	6	12	3.08	–	8.3
Gonioctena	22	17	13.41	35.0	100.0
Ophraella	14	19	10.16	57.68	89.5

Within a given group, there were some cases in which the two data sets were not fully congruent or in which one of the data sets was incomplete or did not give a fully resolved tree. In these cases, only a subset of the clades within the group were used to compare the accuracy of methods. These details are provided in the account for each data set in the results, and basic descriptions of each data set are summarized in Table 1. Unless otherwise noted, the trees for the non-allozyme data sets are based on the parsimony analyses of the original authors. Some may argue that using parsimony-based trees might bias the results of this study to favour parsimony methods for allozyme data. However, parsimony was the method used by most of the authors and is the only method readily applicable to most morphological data sets. Furthermore, the results of this study show distance and likelihood methods generally outperforming parsimony, which suggests that the expected bias is either absent or not overwhelming.

For a given data set, the success (= accuracy or performance) of a given method was measured as the proportion of clades that were correctly resolved as monophyletic by that method, from among the total number of clades that were considered to be well-supported or 'known'. Although other approaches for scoring accuracy could have been used in theory (e.g. tree-to-tree distances), a tallying of correct clades is most appropriate for this study given that, in many cases, only parts of trees (not entire trees) were considered to be known. When multiple equally parsimonious trees were generated by a given parsimony search, a given clade was considered to be successfully resolved only if it was present in the strict consensus tree, following standard practice in empirical studies (and following Wiens & Servedio, 1997, 1998; Wiens, 1998). In the Results, I distinguish between clades that are unresolved (in the consensus tree) and those that are resolved incorrectly. To examine the robustness of the results to an alternate method of dealing with multiple shortest trees, the success of a method for a given data set was also summarized by taking the average accuracy from among the equally parsimonious solutions (as recommended by Hillis, Huelsenbeck & Cunningham, 1994). In cases where thousands of equally shortest trees were generated from a search, a sample of 50 trees was used to estimate the

average accuracy. I did not perform extensive statistical testing of average differences in method accuracy because of problems of multiple tests (Rice, 1989).

A total of 13 phylogenetic methods was examined on all eight data sets, representing most of the common methods for analyzing allozyme data. These included six parsimony coding methods, six genetic distance methods (where each 'method' is a combination of tree-building approach and genetic distance measure), and continuous maximum likelihood ([CONTML]; Felsenstein, 1981). The parsimony coding methods were as follows:

(1) *Frequency*: implemented using step matrices to weight changes between taxa based on the Manhattan (or Prevosti) distance (Wright, 1978) between the allele frequencies for each locus (Wiens, 1995; Berlocher & Swofford, 1997).

(2) *Majority*: coding a given species as having the most common allele for that locus (when a species had two alleles present at equal frequencies, the species was coded as 'polymorphic' for that locus, see below).

(3) *Missing*: coding a species with the allele observed for that locus, but coding species as unknown if more than one allele is present.

(4) *Polymorphic*: coding a species with the observed allele if invariant, but coding the species as having both states if the species was polymorphic (analytically, the polymorphic species is treated as having either of the alleles, but not both; species in which more than two alleles were present for a given locus were coded as having the two most common alleles, or as missing if it was not clear which were the two most frequent).

(5) *Unordered*: coding each unique combination of alleles within a species as a different character state, and unordering the states so that all transitions between states have equal cost.

(6) *Scaled*: coding each unique combination of alleles within a species as a different character state, but using step matrices to weight the cost of transition between states based on the shared presence of alleles (see Mabee & Humphries, 1993; Mardulyn & Pasteels, 1994).

These methods are described in more detail by Wiens (1995); however, the majority, missing, polymorphic, and unordered methods are no longer widely used on allozyme data, and are included largely for the sake of completeness. An additional parsimony method proposed for allozyme data, the 'mutation model' of coding (Murphy, 1993), was not included because its practical application is "frequently impossible for most loci" (p. 32). The related "quadraphenic evaluation procedure" (Murphy, 1993) is not clearly defined, and was also not included. The coding of individual alleles as characters (as opposed to coding the locus as the character) has been refuted repeatedly (e.g. Buth, 1984; Murphy, 1993; Mabee & Humphries, 1993) and never strongly defended theoretically, and was also not evaluated.

Six distance methods were applied to all eight data sets. These consisted of three tree-reconstruction methods, UPGMA (Sokal & Michener, 1958), neighbor-joining (Saitou & Nei, 1987), and Fitch–Margoliash ([FM]; Fitch & Margoliash, 1967), each applied to the standard genetic distance of Nei (1972) and the modified Cavalli-Sforza & Edwards (71967; [CSE]) chord distance. Nei's (1972) distance is very widely used in empirical studies, and the CSE distance has been convincingly argued for on theoretical grounds (Felsenstein, 1985a; Rogers, 1986). These two distance measures generally gave very similar results. Four additional distance measures were also analysed using the three clustering methods on a more limited sample of data sets (see below), and gave similar results. These distances were the unbiased Nei's

(1978), standard Rogers (1972), modified Rogers (Wright, 1978), and Prevosti (Wright, 1978). Thus, a total of 18 distance methods were examined on at least five of the data sets.

Parsimony methods were implemented using a test version of Swofford's (1998) PAUP* (4.0d64). Distance and likelihood methods were implemented using PHYLIP 3.57c (Felsenstein, 1995), but BIOSYS I (Swofford & Selander, 1981) was used to calculate the distances of Rogers (standard and modified), Prevosti, and the unbiased Nei's distance. To standardize the methods as much as possible, all methods were treated as estimating unrooted trees, including UPGMA. Although populations were used as terminal taxa in some of the data sets (as opposed to species), none of the well supported clades involved within-species relationships, and the accuracy of methods designed explicitly for within-species phylogenies was not addressed (e.g. Reynolds, Weir & Cockerham, 1983; Nielsen *et al.*, 1998).

RESULTS

Salamanders (Aneides)

The relationships of species in the plethodontid salamander genus *Aneides* (and the outgroup *Plethodon neomexicanus*) are supported by parsimony analysis of morphological data and by Fitch-Margoliash analysis of albumin immunological data (Larson *et al.*, 1981). The only disagreement between these two data sets concerns the relationships among the species *A. ferreus*, *A. flavipunctatus*, and *A. lugubris* (Fig. 1A). There are two well-supported clades among the six species for testing the performance of methods on the allozyme data of Larson *et al.* (1981).

None of the parsimony methods recover the two well-supported clades. The majority, missing, and unordered parsimony methods all give unresolved consensus trees for *Aneides* relationships. The frequency, polymorphic, and scaled parsimony methods all resolve the same, incorrect tree (*P. neomexicanus*, *lugubris (hardii (ferreus (aeneus + flavipunctatus*))))). This tree is not only rejected by the morphological and immunological data, but also makes little sense biogeographically (i.e. *ferreus*, *lugubris*, and *flavipunctatus* all occur in the extreme western U.S., whereas *aeneus* occurs in the Appalachians). UPGMA (both distances) recovers the two well-supported clades correctly, whereas CONTML, neighbor-joining, and FM group the West Coast species (*ferreus*, *lugubris*, *flavipunctatus*) together but incorrectly place *A. aeneus* as the sister taxon to this clade rather than *A. hardii*.

Frogs (Physalaemus)

Cannatella *et al.* (1998) examined the phylogeny of the *pustulosus* group of *Physalaemus* using characters from DNA sequences (COI and 12S genes), morphology, allozymes, and vocalizations. Relationships among the 10 species are largely congruent between the DNA sequence data sets, morphology, and combined data (the vocalization data set conflicts with all other data sets and was not considered further). The position of *P. pustulosus* is in conflict between the COI gene and the 12S gene and morphology, and I have chosen the placement for *P. pustulosus* supported by 12S, morphology,

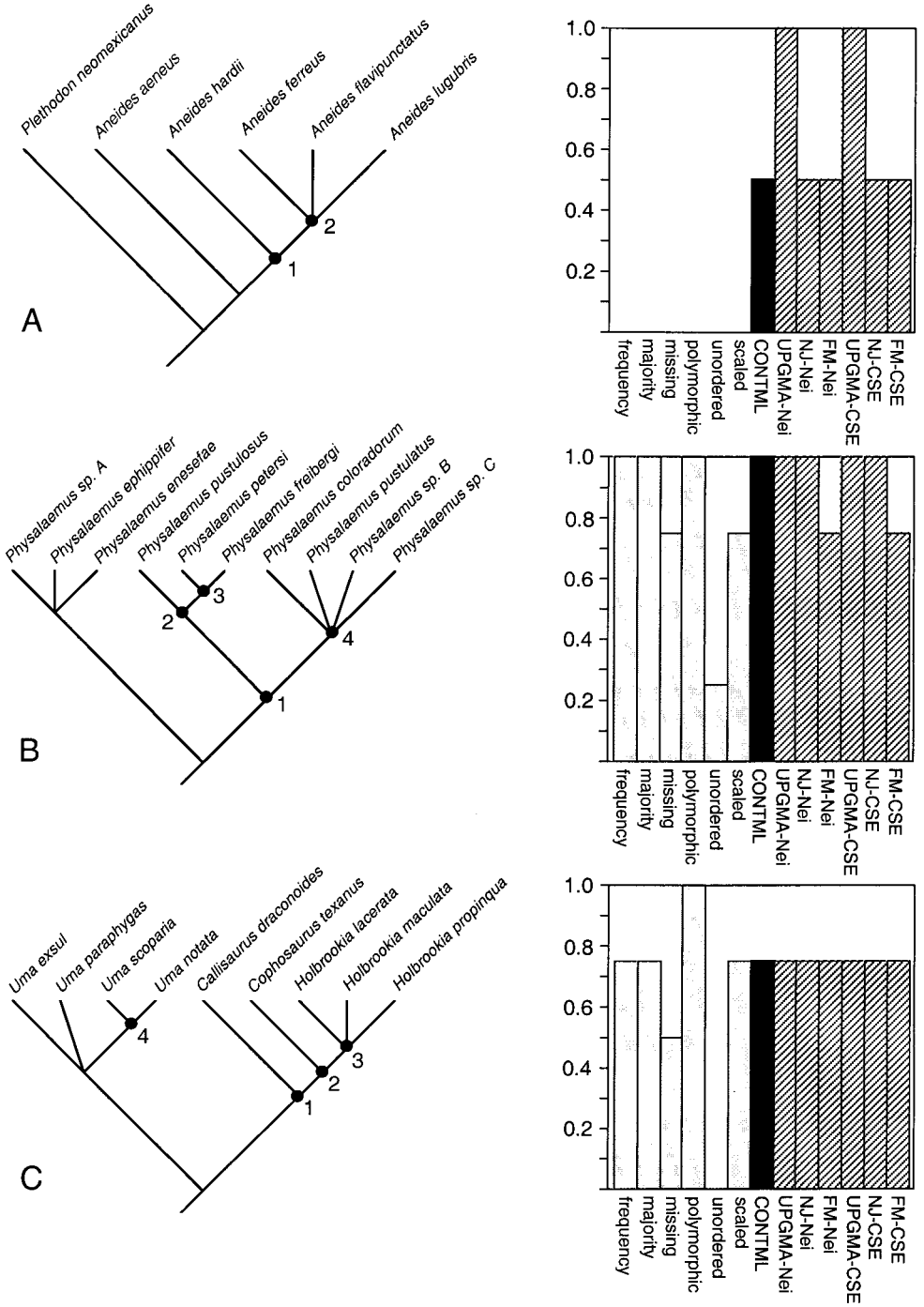

Figure 1. Well-supported phylogenies and results of congruence analyses for (A) salamanders, (B) frogs, and (C) lizards. Bar graphs indicate the proportion of the 'known' clades (numbered) that are correctly resolved by each method. NJ = neighbor-joining. Nei = Nei's standard (1972) distance.

and the combined data. There is also some incongruence and/or lack of resolution concerning the relationships among some of the other species as well (Fig. 1B), leaving four well supported ('known') clades among the 10 species.

Most methods were able to recover all four of these clades correctly. The only exceptions were the missing, unordered, and scaled parsimony methods, which yielded multiple equally parsimonious trees (which included both correct and incorrect resolutions for one or more of these clades), and FM method with Nei's standard distance (which incorrectly resolved one clade).

Sand lizard clade

The phrynosomatid lizard genera *Uma*, *Callisaurus*, *Cophosaurus*, and *Holbrookia* form a clade known as the 'sand lizards' (Etheridge & de Queiroz, 1988). Relationships among these species are supported by morphology (de Queiroz, 1989) and DNA sequence data (Wilgenbusch & de Queiroz, in press), although some intrageneric relationships are not fully resolved by both types of data (Fig. 1C). The DNA sequence tree of Reeder (1995) is not congruent with these relationships, but the incongruent clades are very poorly supported (bootstrap <50%) and based on limited taxon sampling, and a combined analysis of these data with non-DNA data produces the conventional generic-level phylogeny (Reeder & Wiens, 1996). There are four well supported clades among the nine species, and the allozyme data of de Queiroz (1992) are available for these species.

Of all the methods tested, only the polymorphic coding method resolves all the relationships correctly. The frequency, majority, and scaled methods correctly resolve the well supported clades except for the *Uma notata* and *Uma scoparia* clade (relationships incorrectly resolved by the frequency method, unresolved by majority and scaled). This clade (and the monophyly of *Uma*) is supported by the missing method, but the missing method leaves the other 'known' clades unresolved. The unordered method yields a completely unresolved strict consensus tree. CONTML, neighbor-joining, and FM resolve most relationships correctly, but incorrectly place *Cophosaurus* as the sister taxon of *Callisaurus* rather than *Holbrookia*. UPGMA places *Cophosaurus* inside of *Holbrookia*, but otherwise recovers the correct clades.

Birds (Ammodramus)

Zink & Avise (1990) analysed relationships within the sparrow genus *Ammodramus* using mtDNA restriction site data and allozymes, and discussed their results in light of a UPGMA tree based on morphometric data by Robins & Schnell (1971). The mtDNA phylogeny and morphometric tree are fully congruent with regard to the relationships of *A. maritimus*, *A. caudacutus*, *A. lecontei*, and *A. henslowii*, but there is some disagreement concerning the placement of *A. bairdii* with these four species or with three others (*A. aurifrons*, *A. savannarum*, *Passerculus sandwichensis*). Deleting these latter three species produces an unrooted tree of five closely-related species (Fig. 2A), with two clades supported by both DNA and morphological data.

All methods correctly place *A. lecontei*, *A. maritimus*, and *A. caudacutus* together. However, within this group, most methods place *A. caudacutus* and *A. lecontei* as sister

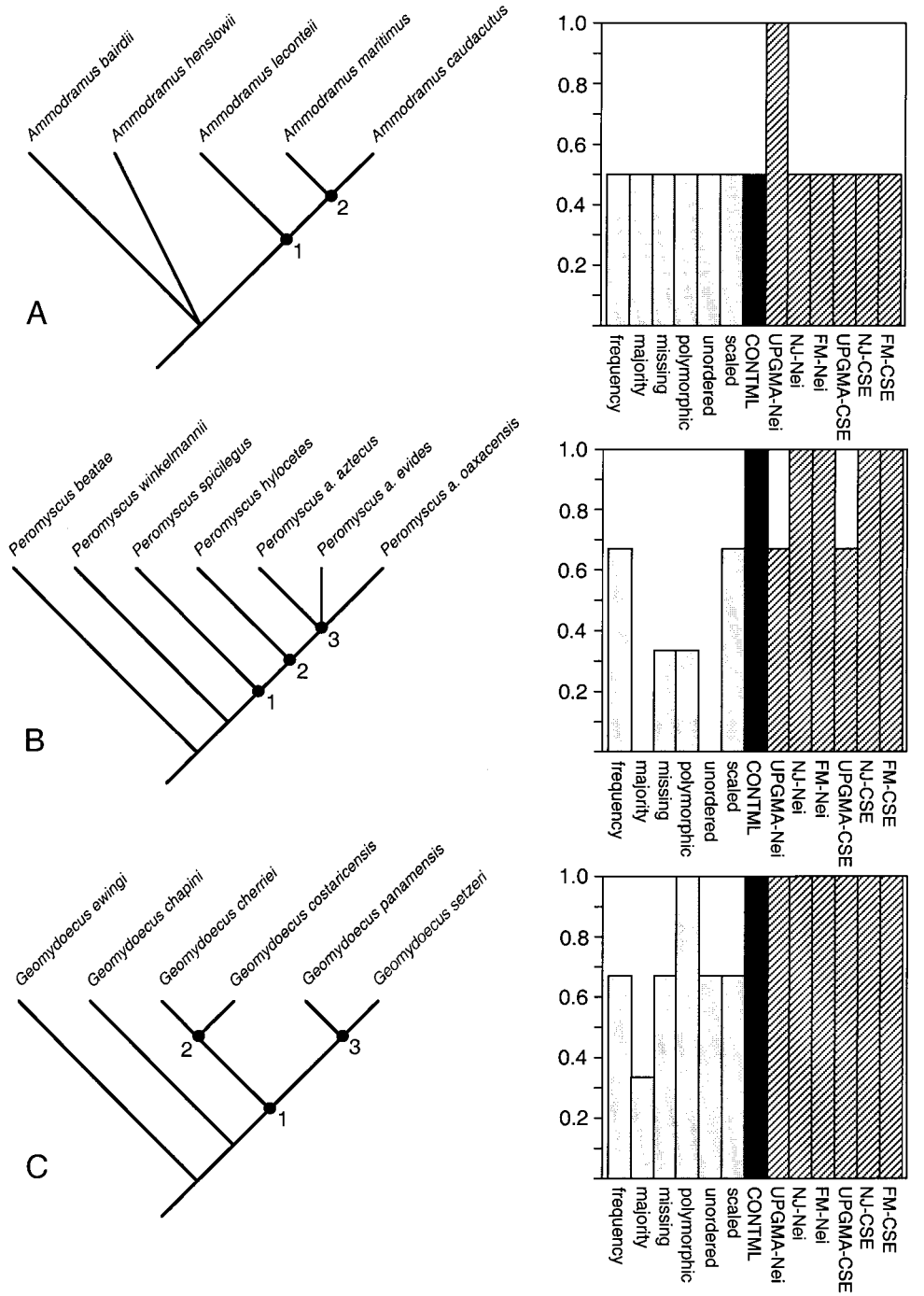

Figure 2. Well-supported phylogenies and results of congruence analyses for (A) birds, (B) rodents, and (C) chewing lice. Bar graphs indicate the proportion of the 'known' clades (numbered) that are correctly resolved by each method. NJ = neighbor-joining. Nei = Nei's standard (1972) distance.

taxa rather than *A. maritimus* and *A. caudacutus*. The only exceptions are UPGMA with Nei's standard distance (which recovers the well-supported clade), and the missing and unordered methods (which do not resolve the relationships among these three species).

Mammals (Peromyscus)

Relationships among the murid rodents of the *Peromyscus aztecus* species group have been examined using data from allozymes (Sullivan & Kilpatrick, 1991), morphology (Bradley & Schmidly, 1987), chromosomes (Smith, 1990), and mitochondrial DNA sequences (Sullivan, Market & Kilpatrick, 1997). Combined analysis of the morphology and chromosomes (Wiens, unpubl. results) gives a tree that is largely congruent with the tree based on cytochrome *b* sequences (Sullivan *et al.*, 1997), except for the relationships among *P. a. aztecus*, *P. a. evides* and *P. a. oaxacensis*, which are unresolved by the non-molecular data (Fig. 2B). There are three well-supported, 'known' clades. I used the populations sampled by Sullivan & Kilpatrick (1991) as terminal units in the analysis, but populations with small sample sizes (n<4) were excluded.

None of the parsimony methods correctly resolve all of the well supported relationships. The scaled and frequency methods have difficulty in correctly placing *P. spicilegus*; the frequency method places *P. spicilegus* with *P. beatae*, and the scaled method places *P. spicilegus* with *P. winkelmani* in one of the shortest trees. The missing and polymorphic methods are unable to resolve any relationships except for the monophyly of *P. aztecus*, and the majority and unordered methods do not resolve any of the well supported relationships. CONTML, neighbor-joining, and FM resolve all the well supported clades correctly, whereas UPGMA (with both distances) incorrectly places *P. spicilegus* and *P. hylocetes* as sister taxa. The failure of UPGMA in this instance may be due to the relatively short terminal branch for *P. spicilegus* (branch length estimated by CONTML), as UPGMA is known to be sensitive to unequal branch lengths (e.g. Huelsenbeck & Hillis, 1993; Wiens & Servedio, 1998).

Chewing lice (Geomydoecus)

Hafner & Nadler (1990) provided allozyme data for eight species of the chewing lice genus *Geomydoecus*. Page, Price & Hellenthal (1995) generated a morphological phylogeny that includes all eight of these species, and the mtDNA sequence data set of Hafner *et al.* (1994) includes six of these species. Parsimony and maximum likelihood analysis of the DNA sequence data (Hafner *et al.*, 1994) yields a molecular phylogeny for these six species that is fully congruent with the morphology tree (Fig. 2C), with three well-supported clades.

The polymorphic coding approach is the only parsimony method that recovers all three well-supported clades. The frequency, missing, scaled, and unordered methods recover the *G. costaricensis–G. cherriei* clade and the *G. panamensis–G. setzeri* clade, but fail to resolve the *G. ewingi–G. chapini* clade. The majority method recovers only the *G. panamensis–G. setzeri* clade. CONTML and the distance methods recover all three of the well supported clades.

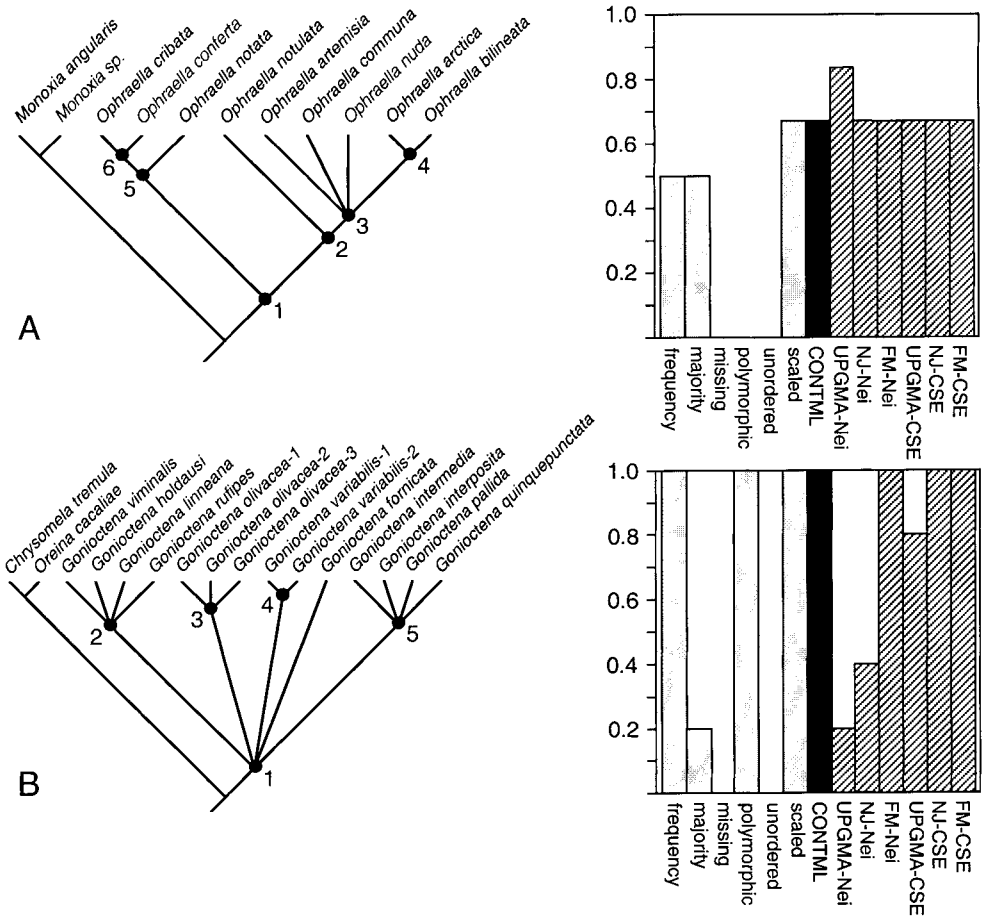

Figure 3. Well-supported phylogenies and results of congruence analyses for two groups of leaf beetles (Chrysomelidae). (A) *Ophraella* and outgroups. (B) *Gonioctena* and outgroups. Bar graphs indicate the proportion of the 'known' clades (numbered) that are correctly resolved by each method. NJ = neighbor-joining. Nei = Nei's standard (1972) distance.

Leaf beetles (Ophraella)

The phylogeny of the chrysomelid beetle genus *Ophraella* has been analyzed with morphology and allozymes (Futuyma & McCafferty, 1990) and with mitochondrial DNA sequences (Funk *et al.*, 1995). After removal of a few problematic taxa (scored for only one data type or of conflicting placement; e.g. *O. pilosa*, *O. slobodkini*) a strict consensus of the morphological phylogeny (Futuyma & McCafferty 1990, Fig. 4, including all characters) and the tree based on mtDNA sequences (Funk *et al.*, 1995; Fig. 3A) reveals six clades supported by both types of data (Fig. 3A).

The scaled parsimony method recovers two thirds of the well supported clades, the frequency and majority methods recover half, and the missing, polymorphic, and unordered methods give very poorly resolved consensus trees that contain none of the 'known' clades. Most of the distance and likelihood methods recover two

thirds of the well supported clades, whereas UPGMA with Nei's standard distance recovers all but one.

Leaf beetles (Gonioctena)

Mardulyn *et al.* (1997) analysed the relationships of *Gonioctena* beetles using allozymes and mitochondrial DNA sequences. Their DNA tree supported many clades recognized as subgenera by morphologists (see Mardulyn *et al.*, 1997 for review), and these five clades are treated as well supported nodes for comparing methods with the allozyme data (Fig. 3B). The populations sampled by Mardulyn *et al.* (1997) for allozyme data were used as terminal units in the phylogenetic analyses.

Most of the parsimony, distance, and likelihood methods recover all five of the 'known' clades. Among the parsimony methods, however, the majority and missing methods give poorly resolved trees, and the unordered method incorrectly resolves all five clades. Among the distance methods, UPGMA and neighbor-joining with Nei's distance incorrectly resolve many of the well supported groups, and UPGMA with CSE incorrectly resolves one of them. Examining the branch lengths estimated by CONTML suggests that there are many large differences in branch lengths among lineages, which may be at least partly responsible for the failure of UPGMA.

Overall performance of the methods

The overall success of the methods (Fig. 4) was examined by averaging method accuracy across the eight data sets (causing all data sets to contribute equally to overall performance, regardless of the number of 'known' clades) and by comparing the proportion of correctly resolved clades summed across all data sets (allowing data sets with more 'known' clades to have a greater contribution to overall method success). The distance and likelihood methods have higher overall accuracy than the parsimony methods when method success was both averaged and summed across data sets. Among parsimony methods, the frequency method and scaled method outperformed the other coding methods (frequency slightly outperformed scaled when accuracies are averaged), and the polymorphic method was a very close third. Curiously, despite its unexceptional performance on average, the polymorphic method is the most accurate of all 13 methods for the sand lizard data set, and the most accurate parsimony method for the *Geomydoecus* data set. The majority, missing, and unordered parsimony methods performed relatively poorly. The overall accuracies of the seven distance and likelihood methods were very similar, but CONTML and neighbor-joining with the CSE distance performed consistently well among these methods. Analysis of 12 additional distance methods on the five smaller data sets (*Aneides*, *Physalaemus*, *Ammodramus*, *Peromyscus*, *Geomydoecus*) further suggests that choice among the different distance measures has relatively little impact on the overall performance of methods (Fig. 5). Nevertheless, the Rogers and modified Rogers distances (with UPGMA) have slightly higher accuracies on average than the Prevosti and unbiased Nei distances.

Using an alternate method for scoring accuracy when there are multiple equally parsimonious trees (averaging among trees) gives generally similar results (Fig. 6),

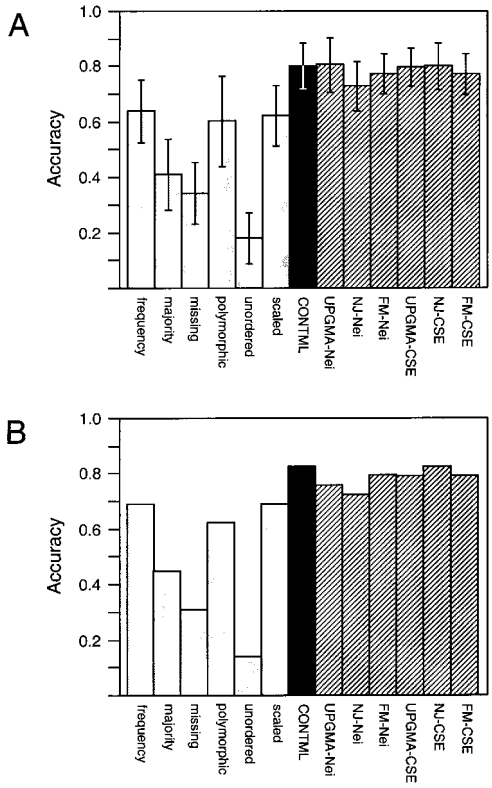

Figure 4. Overall accuracy of 13 phylogenetic methods for eight allozyme data sets. (A) Accuracy of each method from each data set averaged across the eight data sets, such that each data set contributes equally to the measure of overall method success. The line above each bar indicates the standard error of each mean. (B) Accuracy based on the overall proportion of correctly resolved clades summed across all eight data sets, such that data sets with more 'known' clades have a greater contribution to overall method success. NJ = neighbor-joining. Nei = Nei's standard (1972) distance.

although the parsimony methods have somewhat higher accuracy. The scaled and polymorphic methods slightly outperform the frequency parsimony method, and these two methods are slightly more accurate than one of the distance methods (neighbor-joining with Nei's distance) when accuracies are summed (rather than averaged) across data sets (Fig. 6B).

DISCUSSION

The results of this study shed light on three long standing controversies concerning the phylogenetic analysis of allozyme data: (1) the use of parsimony versus distance methods, (2) the utility of frequency data, and (3) the success of continuous maximum likelihood (CONTML). First, the results suggest that distance and likelihood methods generally recover more of the 'known' clades than do any of the parsimony methods. This is not simply a result of the distance and likelihood methods producing more well-resolved trees; in many cases (e.g. *Aneides*) parsimony methods incorrectly resolve

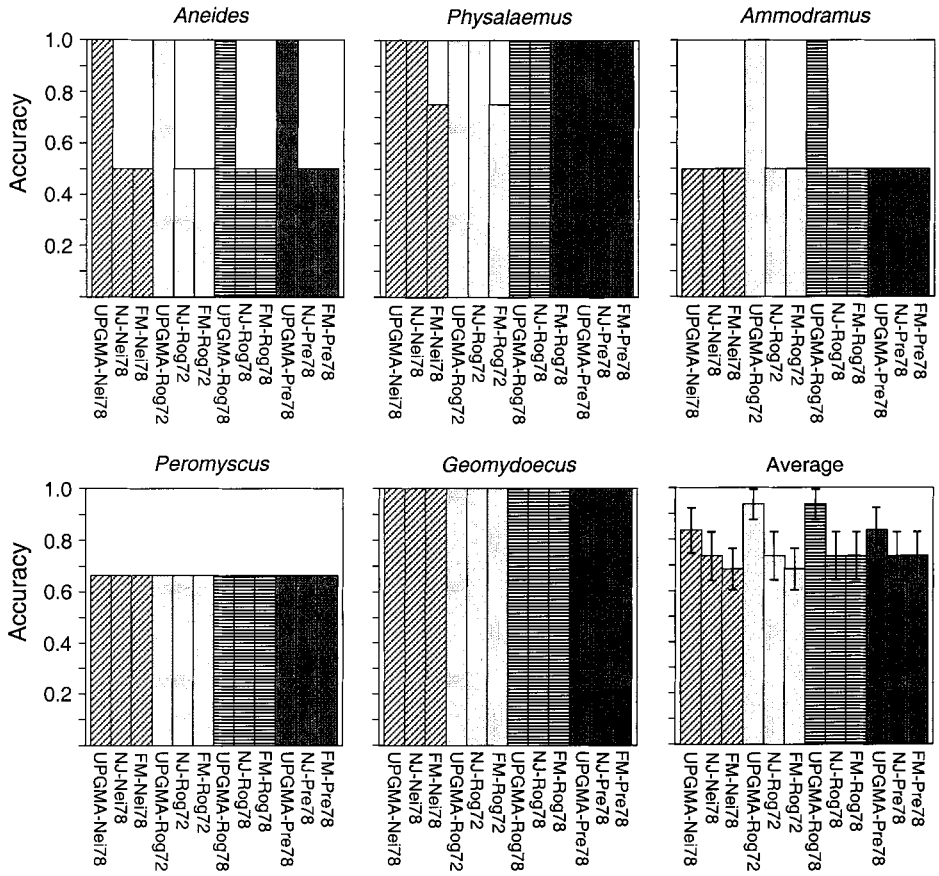

Figure 5. Results of congruence analyses (proportion of well-supported clades resolved correctly) for 12 additional distance methods. On the graph showing average accuracy across the five data sets, the line above each bar indicates the standard error for each mean. NJ = neighbor-joining. Nei78 = Nei's (1978) unbiased distance. Rog72 = Rogers (1972) standard distance. Rog78 = Rogers modified distance (Wright, 1978). Pre78 = Prevosti distance (Wright, 1978).

clades that are correctly resolved by distance and likelihood methods. Furthermore, the frequency parsimony method gives well-resolved estimates but did not perform as well as the distance and likelihood methods. For example, the accuracy of CONTML (the most generally accurate distance/likelihood method) is significantly higher than that of the frequency parsimony method (the most accurate parsimony method from Fig. 4), based on a paired *t*-test ($P = 0.0498$) of the results summarized in Figure 4A. Nevertheless, the greater resolution of the distance and likelihood methods (and the frequency parsimony method) may contribute somewhat to their higher accuracy, as suggested by the higher accuracies of parsimony methods using an alternate method for treating multiple shortest trees (e.g. compare Figs 4 and 6).

Second, the results suggest that frequency-based methods perform well. Many authors have argued that the frequencies of traits (e.g. alleles) vary too much over space and time within species to be useful in reconstructing phylogeny between species (e.g. Mickevich & Johnson, 1976; Crother, 1990; Jones *et al.*, 1993; Mabee

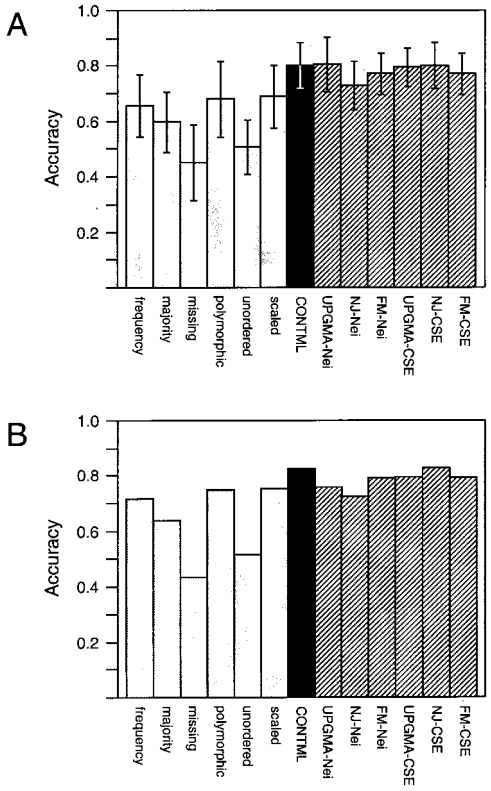

Figure 6. Overall accuracy of 13 phylogenetic methods for eight allozyme data sets, using the average accuracy of multiple equally parsimonious trees to score the success of parsimony methods. (A) Accuracy of each method from each data set averaged across the eight data sets, such that each data set contributes equally to the measure of overall method success. The line above each bar indicates the standard error of each mean. (B) Accuracy based on the overall proportion of correctly resolved clades summed across all eight data sets, such that data sets with more 'known' clades have a greater contribution to overall method success. NJ = neighbor-joining. Nei = Nei's standard (1972) distance.

& Humphries, 1993; Murphy, 1993; Murphy & Doyle, 1998). However, the methods that utilized frequency information (the distance methods, continuous maximum likelihood, and frequency parsimony) were generally the most accurate methods in this study (Fig. 4). The overall accuracy of the frequency parsimony method was very similar to that of the scaled (Mabee & Humphries) parsimony method, and was slightly inferior using one method for scoring accuracy (Fig. 6). This observation may indicate that the Manhattan/Prevosti distance step matrix method does not utilize frequency information as well as the non-parsimony methods, and/or that the superiority of the distance and likelihood methods is partly independent of their use of frequency information. Regardless, it is clear that this study rejects the idea that frequency information is misleading in phylogenetic analyses of allozyme data. This result echoes those of Wiens (1995), who found that many allozyme data sets do contain significant phylogenetic information when analysed using frequency methods.

Third, the results of this study strongly support the use of continuous maximum

likelihood (CONTML) to analyse allozyme data. Although maximum likelihood methods have many desirable properties in phylogenetic inference in general (e.g. Felsenstein, 1978; Swofford *et al.*, 1996), it seems likely that many researchers have hesitated to use CONTML because of the unrealistic assumptions it makes about the evolutionary process (e.g. Swofford & Berlocher, 1987). CONTML explicitly assumes a Brownian motion model of evolution, in which there is no fixation or loss of alleles nor any introduction of new alleles through mutation (Felsenstein, 1981). Yet, CONTML was among the most accurate methods in this study, despite the fact that all of the data sets used exhibit: (1) loci that appear to be fixed within species, and (2) loci with many alleles within and/or between species (which suggests the introduction of new alleles through mutation). Thus, the results suggest that CONTML may perform well even when its assumptions are seemingly violated. In support of this idea, simulations show that CONTML is robust to violations of the assumption of no fixation or loss (Wiens & Servedio, 1998). The robustness of CONTML is particularly important because simulation results predict that CONTML should give accurate results in the 'Felsenstein Zone' (when there are long, unrelated terminal branches united by a short internal branch; Huelsenbeck & Hillis, 1993), whereas parsimony methods and UPGMA will be misled (Wiens & Servedio, 1998). However, the accuracy of CONTML under branch length conditions where likelihood analyses of simulated DNA sequence data may perform poorly (Yang, 1996; Huelsenbeck, 1998; Siddall, 1998) has yet to be tested.

The overall results of this study generally agree with those from recent simulation and congruence studies of polymorphic data (i.e. characters that vary within as well as between species). Wiens & Servedio (1998) examined the performance of 16 methods (including the 13 methods analysed herein) using simulated allele frequency data evolving by random genetic drift. Wiens (1998) compared the same methods in a congruence study of polymorphic morphological characters in phrynosomatid lizards. These two studies and the present one agree on the following points: (1) distance and likelihood methods are as or more accurate (on average) than parsimony methods under most conditions, (2) frequency-based methods (whether parsimony, distance, or likelihood) typically outperform non-frequency parsimony methods (although the superiority of frequency parsimony is slight in this study and depends on how accuracy is scored), and (3) among non-frequency parsimony methods, the scaled (Mabee & Humphries) method tends to perform best. The concordance between the simulation, morphological, and allozyme results suggest that these conclusions may apply to many data sets.

Given the results of this study, what is the best method to apply to phylogenetic analyses of allozyme data? In this study, the distance and likelihood methods generally outperformed the parsimony methods, but the accuracy of the distance and likelihood methods were relatively similar overall. CONTML and neighbor-joining with the CSE distance performed relatively well among the distance and likelihood methods using both averaged and summed accuracies. Furthermore, parsimony methods and UPGMA are known to be sensitive to certain types of unequal branch lengths, whereas CONTML, neighbor-joining, and Fitch-Margoliash appear to be robust (Wiens & Servedio, 1998). Unlike neighbor-joining and UPGMA, CONTML employs an optimality criterion, which allows one to evaluate the success of tree searches (Swofford *et al.*, 1996). Given these results and considerations, I recommend CONTML for phylogenetic analyses of allozyme data, but acknowledge the need for further study of its performance relative to other methods.

Although parsimony methods did not perform as well as distance and likelihood methods in this study, parsimony methods have distinct advantages that are not directly related to accuracy, such as ease of including characters with missing data (impossible with recent versions of CONTML in PHYLIP), combining data sets evolving under different models (e.g. allozymes, DNA sequences), and evaluating the contribution of individual characters to the recovery of specific clades. Because of these considerations, parsimony methods may be desirable in some situations. The results of this study suggest that the frequency and scaled methods may be the most accurate parsimony methods to use (the polymorphic method is relatively inconsistent between data sets), and the results of simulations, statistical analyses, and other congruence studies suggest that the frequency method may often be advantageous (Wiens, 1995, 1998; Wiens & Servedio, 1997, 1998).

A surprising result of this study is the extent to which method success varies, both within and between data sets. For half of the data sets examined in this study, one or more of the methods recovered all of the known clades, whereas one or more of the other methods recovered none of the known clades for the same data set. Similarly, some methods vary greatly in their relative accuracy between data sets. This is especially true of UPGMA (with Nei's standard distance) and the polymorphic parsimony coding method. For example, for the *Ophraella* (leaf beetle) data set the polymorphic coding method has an accuracy of 0 whereas UPGMA has an accuracy of 0.833. In contrast, for the *Gonioctena* (leaf beetle) data set (which is extremely similar; Table 1), the polymorphic method has an accuracy of 1.000 and UPGMA has an accuracy of 0.200. The poor performance of UPGMA on some data sets may be explained by obvious differences in estimated branch lengths among lineages, to which UPGMA is known to be sensitive (e.g. Huelsenbeck & Hillis, 1993; Wiens & Servedio, 1998). However, it is unclear why UPGMA and the polymorphic method perform so well relative to other methods on certain data sets. Unfortunately, the bases for differences in method success can be difficult to determine in congruence studies (because the relevant evolutionary parameters are unknown and/or not amenable to systematic manipulation), a disadvantage relative to simulation studies (Wiens, 1998). Although the results of this study reveal trends that should be important for empirical researchers seeking to choose among the available methods, the results also show the need for further simulation and congruence studies testing the behavior and accuracy of methods for analysing allozyme data. Such studies will be particularly important because the present results suggest that the choice of phylogenetic method may have a considerable impact on the accuracy of phylogenies estimated from these data.

ACKNOWLEDGEMENTS

I thank Doug Futuyma for suggesting that I pursue this study, and for sending me his *Ophraella* data. I am grateful to Kevin de Queiroz and Jim Wilgenbusch for permission to cite their unpublished work on sand lizards. I thank Tonya Penkrot for help with data input, Mahmood Sasa for valuable assistance with BIOSYS, David Swofford for allowing me to use a test version of his PAUP* program, and Brad Livezey, Patrick Mardulyn, Maria Servedio, and an anonymous reviewer for helpful comments on the manuscript.

REFERENCES

Allard MW, Miyamoto MM. 1992. Testing phylogenetic approaches with empirical data, as illustrated with the parsimony method. *Molecular Biology and Evolution* **9:** 778–786.

Berlocher SH, Swofford DL. 1997. Searching for phylogenetic trees under the frequency parsimony criterion: an approximation using generalized parsimony. *Systematic Biology* **46:** 211–215.

Bradley RD, Schmidly DJ. 1987. The glans penes and bacula of Latin American taxa of the *Peromyscus boylii* group. *Journal of Mammalogy* **68:** 595–616.

Brumfield RT, Capparella AP. 1996. Historical diversification of birds in northwestern South America: a molecular perspective on the role of vicariant events. *Evolution* **50:** 1607–1624.

Buth DG. 1984. The application of electrophoretic data in systematic studies. *Annual Review of Ecology and Systematics* **15:** 501–522.

Cannatella DC, Hillis DM, Chippindale PT, Weight L, Rand AS, Ryan MJ. 1998. Phylogeny of frogs of the *Physalaemus pustulosus* species group, with an examination of data set incongruence. *Systematic Biology* **47:** 311–335.

Cavalli-Sforza LL, Edwards AWF. 1967. Phylogenetic analysis: models and estimation procedures. *American Journal of Human Genetics* **19:** 233–257.

Crother BI. 1990. Is "some better than none" or do allele frequencies contain phylogenetically useful information? *Cladistics* **6:** 277–281.

de Queiroz K. 1989. Morphological and biochemical evolution in the sand lizards. Unpublished Ph.D. Thesis, University of California, Berkeley.

de Queiroz K.1992. Phylogenetic relationships and rates of allozyme evolution among the lineages of sceloporine sand lizards. *Biological Journal of the Linnean Society* **45:** 333–362.

Etheridge R, de Queiroz K. 1988. A phylogeny of Iguanidae. In: Estes R, Pregill GK, eds. *Phylogenetic relationships of the lizard families: essays commemorating Charles L. Camp*. Stanford, California: Stanford Univ. Press, 283–368.

Farris JS. 1981. Distance data in phylogenetic analysis. In: Funk VA, Brooks DR, eds. *Advances in cladistics, Volume 1. Proceeding of the first meeting of the Willi Hennig Society*. New York: New York Botanical Garden, 3–23.

Farris JS. 1985. Distance data revisited. *Cladistics* **1:** 67–85.

Farris JS. 1986. Distances and cladistics. *Cladistics* **2:** 144–157.

Felsenstein J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. *Systematic Zoology* **27:** 401–410.

Felsenstein J. 1981. Evolutionary trees from gene frequencies and quantitative characters: finding maximum likelihood estimates. *Evolution* **35:** 1229–1242.

Felsenstein J. 1984. Distance methods for inferring phylogenies: a justification. *Evolution* **38:** 16–24.

Felsenstein J. 1985a. Phylogenies from gene frequencies: a statistical problem. *Systematic Zoology* **34:** 300–311.

Felsenstein J. 1985b. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39:** 783–791.

Felsenstein J. 1986. Distance methods: a reply to Farris. *Cladistics* **2:** 130–143.

Felsenstein J. 1995.*PHYLIP: phylogeny inference package. Version 3.57c*. Department of Genetics, University of Washington, Seattle, Washington.

Fitch WM, Margoliash E. 1967. Construction of phylogenetic trees. *Science* **155:** 279–284.

Funk DJ, Futuyma DJ, Ortí G, Meyer A. 1995. Mitochondrial DNA sequences and multiple data sets: a phylogenetic study of phytophagous beetles (Chrysomelidae: *Ophraella)*. *Molecular Biology and Evolution* **12:** 627–640.

Futuyma DJ, McCafferty S. 1990. Phylogeny and the evolution of host plant associations in the leaf beetle genus *Ophraella* (Coleoptera: Chrysomelidae). *Evolution* **44:** 1885–1913.

Hafner MS, Nadler SA. 1990. Cospeciation in host-parasite assemblages: comparative analysis of rates of evolution and timing of cospeciation. *Systematic Zoology* **39:** 192–204.

Hafner MS, Sudman PD, Villablanca FX, Spradling TA, Demastes JW, Nadler SA. 1994. Disparate rates of molecular evolution in cospeciating hosts and parasites. *Science* **265:** 1087–1090.

Hillis DM, Huelsenbeck JP, Cunningham CW. 1994. Application and accuracy of molecular phylogenies. *Science* **264:** 671–677.

Hillis DM, Mable BK, Moritz C. 1996. Applications of molecular systematics: the state of the field and a look to the future. In: Hillis DM, Moritz C, Mable BK, eds. *Molecular systematics. 2nd Edition*. Sunderland, Massachusetts: Sinauer Associates, 515–543.

Huelsenbeck JP. 1998. Systematic bias in phylogenetic analysis: is the Strepsiptera problem solved? *Systematic Biology* **47:** 519–537.

Huelsenbeck JP, Hillis DM. 1993. Success of phylogenetic methods in the four-taxon case. *Systematic Biology* **42:** 247–264.

Jones TR, Kluge AG, Wolf AJ. 1993. When theories and methodologies clash: a phylogenetic reanalysis of the North American ambystomatid salamanders (Caudata: Ambystomatidae). *Systematic Biology* **42:** 92–102.

Kim J, Burgman MA. 1988. Accuracy of phylogenetic-estimation methods under unequal evolutionary rates. *Evolution* **42:** 596–602.

Klauta M, Russo CAM, Lazoski C, Boury-Esnault N, Thorpe JP, Sole-Cava. 1999. Does cosmopolitanism result from overconservative systematics? A case study using the sponge *Chondrilla nucula*. *Evolution* **53:** 1414–1422.

Larson A, Wake DB, Maxson LR, Highton R. 1981. A molecular phylogenetic perspective on the origins of morphological novelties in the salamanders of the tribe Plethodontini (Amphibia, Plethodontidae). *Evolution* **35:** 405–422.

Mabee PM, Humphries J. 1993. Coding polymorphic data: examples from allozymes and ontogeny. *Systematic Biology* **42:** 166–181.

Mardulyn P, Pasteels JM. 1994. Coding allozyme data using step matrices: defining new original states for the ancestral taxa. *Systematic Biology* **43:** 567–572.

Mardulyn P, Milinkovitch MC, Pasteels JM. 1997. Phylogenetic analyses of DNA and allozyme data suggest that *Gonioctena* leaf beetles (Coleoptera: Chrysomelidae) experienced convergent evolution in their history of host-plant family shifts. *Systematic Biology* **46:** 722–747.

Marko PB. 1998. Historical allopatry and the biogeography of speciation in the prosobranch snail genus *Nucella*.*Evolution* **52:** 757–774.

Mickevich MF. 1978. Taxonomic congruence. *Systematic Zoology* **27:** 143–158.

Mickevich MF, Johnson MS. 1976. Congruence between morphological and allozyme data in evolutionary inference and character evolution. *Systematic Zoology* **25:** 260–270.

Mickevich MF, Mitter C. 1981. Treating polymorphic characters in systematics: a phylogenetic treatment of electrophoretic data. In: Funk VA, Brooks DR, eds. *Advances in cladistics, Volume 1. Proceeding of the first meeting of the Willi Hennig Society*. New York: New York Botanical Garden, 45–58.

Mickevich MF, Mitter C. 1983. Evolutionary patterns in allozyme data: a systematic approach. In: Funk VA, Brooks DR, eds. *Advances in cladistics, Volume 2. Proceeding of the second meeting of the Willi Hennig Society*. New York: Columbia University Press, 169–176.

Miyamoto MM, Fitch WM. 1995. Testing species phylogenies and phylogenetic methods with congruence. *Systematic Biology* **44:** 64–76.

Murphy RW. 1993. The phylogenetic analysis of allozyme data: invalidity of coding alleles by presence/absence and recommended procedures. *Biochemical Systematics and Ecology* **21:** 25–38.

Murphy RW, Doyle KD. 1998. Phylophenetics: frequencies and polymorphic characters in genealogical estimation. *Systematic Biology* **47:** 737–761.

Nei M. 1972. Genetic distance between populations. *American Naturalist* **106:** 238–292.

Nei M. 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics* **89:** 583–590.

Nei M, Tajima F, Tateno Y. 1983. Accuracy of estimated phylogenetic trees from molecular data. II. Gene frequency data. *Journal of Molecular Evolution* **19:** 153–170.

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:** 669–677.

Nyman T, Roininen H, Vuorinen JA. 1998. Evolution of different gall types in willow-feeding sawflies (Hymenoptera: Tenthredinidae). *Evolution* **52:** 465–474.

Page RD, Price RDP, Hellenthal RA. 1995. Phylogeny of *Geomydoecus* and *Thomomydoecus* pocket gopher lice (Phthiraptera: Trichodectridae) inferred from cladistic analysis of adult and first-instar morphology. *Systematic Entomology* **20:** 129–143.

Pamilo P, Nei M. 1988. Relationships between gene trees and species trees. *Molecular Biology and Evolution* **5:** 568–583.

Reeder TW. 1995. Phylogenetic relationships among phrynosomatid lizards as inferred from mitochondrial ribosomal DNA sequences: substitutional bias and information content of transitions relative to transversions. *Molecular Phylogenetics and Evolution* **4:** 203–222.

Reeder TW, Wiens JJ. 1996. Evolution of the lizard family Phrynosomatidae as inferred from diverse types of data. *Herpetological Monographs* **10:** 43–84.

Reynolds J, Weir BS, Cockerham CC. 1983. Estimation of coancestry coefficient: basis for a short-term genetic distance. *Genetics* **105:** 767–779.

Rice WR. 1989. Analyzing tables of statistical tests. *Evolution* **43:** 223–225.

Robins JD, Schnell GD. 1971. Skeletal analysis of the *Ammodramus-Ammospiza* grassland sparrow complex: a numerical taxonomic study. *Auk* **88:** 567–590.

Rogers JS. 1972. Measures of genetic similarity and genetic distance. *Studies in Genetics VII. University of Texas Publications* **7213:** 145–153.

Rogers JS. 1986. Deriving phylogenetic trees from allele frequencies: a comparison of nine genetic distances. *Systematic Zoology* **35:** 297–310.

Rohlf FJ, Wooten MC. 1988. Evaluation of the restricted maximum-likelihood method for estimating phylogenetic trees using simulated allele-frequency data. *Evolution* **42:** 581–595.

Ruedi M. 1996. Phylogenetic evolution and biogeography of Southeast Asian shrews (genus *Crocidura*: Soricidae). *Biological Journal of the Linnean Society* **58:** 197–219.

Saitou N, Nei M. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4:** 406–425.

Siddall ME. 1998. Success of parsimony in the four-taxon case: long-branch repulsion by likelihood in the Farris Zone. *Cladistics* **14:** 209–220.

Smith SA. 1990. Cytosystematic evidence against the monophyly of the *Peromyscus boylii* species group (Rodentia: Cricetidae). *Journal of Mammalogy* **71:** 654–667.

Sokal RR, Michener CD. 1958. A statistical method for evaluating systematic relationships. *University of Kansas Scientific Bulletin* **28:** 1409–1438.

Sullivan JM, Kilpatrick CW. 1991. Biochemical systematics of the *Peromyscus aztecus* assemblage. *Journal of Mammalogy* **72:** 681–696.

Sullivan J, Markert JA, Kilpatrick CW. 1997. Phylogeography and molecular systematics of the *Peromyscus aztecus* species group (Rodentia: Muridae) inferred using parsimony and likelihood. *Systematic Biology* **46:** 426–440.

Swofford DL. 1998. *PAUP*: Phylogenetic analysis using parsimony (*and other methods), version 4.0*. Sunderland, Massachusetts: Sinauer Associates.

Swofford DL, Berlocher SH. 1987. Inferring evolutionary trees from gene frequency data under the principle of maximum parsimony. *Systematic Zoology* **36:** 293–325.

Swofford DL, Selander RB. 1981. *BIOSYS-1: a computer program for the analysis of genetic variation*. Department of Development and Genetics, University of Illinois, Urbana, Illinois.

Swofford DL, Olsen GJ, Waddell PJ, Hillis DM. 1996. Phylogeny reconstruction. In: Hillis DM, Moritz C, Mable BK, eds. *Molecular systematics. 2nd Edition*. Sunderland, Massachusetts: Sinauer Associates, 407–514.

Weller SG. 1996. Allozyme diversity and genetic identity in *Schiedea* and *Alsinidendron* (Caryophyllaceae: Alsinoidae) in the Hawaiian islands. *Evolution* **50:** 23–34.

Wiens JJ. 1995. Polymorphic characters in phylogenetic systematics. *Systematic Biology* **44:** 482–500.

Wiens JJ. 1998. Testing phylogenetic methods with tree congruence: phylogenetic analysis of polymorphic morphological characters in phrynosomatid lizards. *Systematic Biology* **47:** 411–428.

Wiens JJ, Servedio MR. 1997. Accuracy of phylogenetic analysis including and excluding polymorphic characters. *Systematic Biology* **46:** 332–345.

Wiens JJ, Servedio MR. 1998. Phylogenetic analysis and intraspecific variation: performance of parsimony, distance, and likelihood methods. *Systematic Biology* **47:** 228–253.

Wilgenbusch J, de Queiroz K. Phylogenetic relationships among the phrynosomatid sand lizards inferred from mitochondrial DNA sequences. *Systematic Biology* in press.

Wright S. 1978. *Evolution and the genetics of populations. Vol. IV. Variation within and among natural populations*. Chicago, Illinois: University of Chicago Press.

Yang Z.1996. Phylogenetic analysis using parsimony and likelihood methods. *Journal of Molecular Evolution* **42:** 294–307.

Zink RM, Avise JC. 1990. Patterns of mitochondrial DNA and allozyme evolution in the avian genus *Ammodramus*. *Systematic Zoology* **39:** 148–161.