IS CONGRUENCE BETWEEN DATA PARTITIONS A RELIABLE PREDICTOR OF PHYLOGENETIC ACCURACY? EMPIRICALLY TESTING AN ITERATIVE PROCEDURE FOR CHOOSING AMONG PHYLOGENETIC METHODS

CLIFFORD W. CUNNINGHAM

Zoology Department, Duke University, Durham, North Carolina 27708-0325, USA; E-mail: cliff@duke.edu

Abstract.—The relationship between phylogenetic accuracy and congruence between data partitions collected from the same taxa was explored for mitochondrial DNA sequences from two well-supported vertebrate phylogenies. An iterative procedure was adopted whereby accuracy, phylogenetic signal, and congruence were measured before and after modifying a simple reconstruction model, equally weighted parsimony. These modifications included transversion parsimony, successive weighting, and six-parameter parsimony. For the data partitions examined, there is a generally positive relationship between congruence and phylogenetic accuracy. If congruence increased without decreasing resolution or phylogenetic signal, this increased congruence was a good predictor of accuracy. If congruence increased as a result of poor resolution, the degree of congruence was not a good predictor of accuracy. For all sets of data partitions, six-parameter parsimony methods show a consistently positive relationship between congruence and accuracy. Unlike successive weighting, six-parameter parsimony methods were not strongly influenced by the starting tree. [Congruence; incongruence tests; maximum likelihood; phylogenetic accuracy; six-parameter parsimony; successive weighting; transversion parsimony; weighted parsimony.]

The relationship between congruence and phylogenetic accuracy is central to systematic biology. For example, the consistency index, a widely used measure of data reliability, is based on the degree of congruence between characters and a tree (Kluge and Farris, 1969). Moreover, the topological congruence between phylogenies supported by independent data partitions is considered some of the strongest support for phylogenetic relationships (Hillis, 1987, 1995; Allard and Miyamoto, 1992; Lanyon, 1993; Miyamoto et al., 1994; Olmstead and Sweere, 1994; Omland, 1994; Miyamoto and Fitch, 1995). Even further, the common assumption that congruence implies phylogenetic accuracy is the principal justification for using consensus methods to compare and summarize information from independent data partitions (reviewed by Swofford, 1991; Lanyon, 1993; but see Barrett et al., 1991). The central importance of congruence has been dramatized by the vigorous debate over whether or not to combine data partitions that support incongruent phylogenies (Kluge, 1989; Bull et al., 1993; de Queiroz, 1993; Rodrigo et al., 1993; Chippindale and Wiens, 1994; de Queiroz et al., 1995; Lutzoni and Vilgalys, 1995; Huelsenbeck et al., 1996; Cunningham, 1997).

Bull et al. (1993) argued that the degree of incongruence between data partitions provides important information about their phylogenetic accuracy. If we can be sure that data partitions have experienced the same history, then any disagreement between them must arise from phylogenetic inaccuracy. Although this inaccuracy might arise from inadequate sampling (see also de Queiroz, 1993; Rodrigo et al., 1993), the source of incongruence might lie in the phylogenetic reconstruction model being employed (e.g., Felsenstein, 1978; Huelsenbeck and Hillis, 1993). When revision of the reconstruction model to better fit the data causes the data partitions to converge on the correct phylogeny, this increased accuracy will be reflected by greater congruence.

If there is a positive relationship between congruence and phylogenetic accuracy, maximizing congruence between partitions might form the basis for choosing
among phylogenetic methods (e.g., Miyamoto et al., 1994). Congruence is especially useful with parsimony, where it is often unclear when to apply weighting methods (e.g., Chippindale and Wiens, 1994; Huls-ensenbeck et al., 1994). For example, vertebrate 18S ribosomal DNA sequences were more congruent with the fossil record when analyzed using six-parameter weighted parsimony (Marshall, 1992) than when using equally weighted parsimony. Similarly, Wheeler (1995) used congruence with a morphological phylogeny as a criterion for choosing appropriate gap weights when aligning DNA sequences.

To investigate the relationship between congruence and phylogenetic accuracy, data partitions must have shared the same history. Because I used linked mitochondrial DNA sequences, any incongruence between partitions must be caused by phylogenetic inaccuracy (see Allard and Miyamoto, 1992; Allard and Carpenter, 1996). To evaluate the accuracy of individual partitions, it is necessary to have an a priori expectation of the correct phylogeny. This expectation ensures that increased congruence is not caused by data partitions converging on the wrong phylogeny. To this end, I used linked mitochondrial DNA sequences from two well-corroborated vertebrate phylogenies (Graybeal, 1994; Sullivan et al., 1995).

I evaluated an iterative procedure for investigating the effect of parsimony weighting methods on congruence and phylogenetic accuracy (Bull et al., 1993). First, the degree of phylogenetic signal and character incongruence between data partitions was measured under a simple reconstruction model. Second, the phylogenetic signal and character incongruence was measured again after the reconstruction model had been modified. Finally, I determined whether the degree of character incongruence under the modified reconstruction model was associated with a corresponding change in phylogenetic accuracy of the weighted partitions.

**MATERIALS AND METHODS**

**The Data**

Well-corroborated phylogenies are widely used to evaluate phylogenetic methods (Hillis and Dixon, 1991; Friedlander et al., 1992, 1996; Marshall, 1992; Kumazawa and Nishida, 1993; Saccone et al., 1993; Graybeal, 1994; Miyamoto and Fitch, 1995; Russo et al., 1996; Cunningham, 1997). The first well-corroborated phylogeny I considered includes the rodent genera *Peromyscus* and *Onychomys* (Fig. 1a). As noted by Sullivan et al. (1995), there is a remarkable concordance among phylogenies inferred for these taxa from a variety of morphological and molecular data partitions (reviewed by Carleton, 1989; Hogan et al., 1993). Sullivan et al. (1995) obtained partial sequences of cytochrome *b* (Cyt *b*, 319 bp) and of the small subunit of mitochondrial ribosomal DNA (12S, 770 bp, not including 7 ambiguously aligned positions). The sequence alignments used in the original study were provided by J. Sullivan. Because intrageneric relationships within *Onychomys* are not well established (J. Sullivan, R. Honeycutt, pers. comm.), the in-
trageneric node was not considered when calculating phylogenetic accuracy.

The second well-corroborated phylogeny was used by Graybeal (1994) as a model system to evaluate the phylogenetic utility of genes (Fig. 1b). I chose representative species from this phylogeny whose entire mitochondrial genomes are available in GenBank: fish (Cyprinus carpio, GB X61010), frog (Xenopus laevis, GB M10217), bird (Gallus gallus, GB X52392), rodent (Mus musculus, GB J01420), and human (Homo sapiens, GB J01415). Five protein-coding genes were chosen because they are widely used for phylogenetic analysis and because they showed various degrees of support for the expected phylogeny under equally weighted parsimony (ATPase6 [ATP6], cytochrome oxidase I-III [COI-III], Cyt b). The mitochondrial DNA sequences were first analyzed as individual genes, and then the five genes were combined and partitioned by codon position.

DNA Sequence Alignment

When phylogenetic methods are evaluated, it is important to use unambiguous alignments (Lake, 1991). The five genes from the higher level vertebrate phylogeny were aligned using a procedure designed to remove regions that differed when aligned under different weighting parameters (e.g., Gatesy et al., 1993; Bridge et al., 1995). The procedure produced unambiguously aligned sequences for each gene (ATP6, 603 bp; COI, 1,506 bp; COII, 648 bp; COIII, 786 bp; Cyt b, 1,095 bp). These aligned sequences are available from the EBI FTP server under accession code DS29741 either by anonymous FTP from FTP/EMBL.AT.UK in directory /pub/databases/embl/align or by sending an e-mail message to netserv@ebi.ac.uk including the line GET ALIGN;DS29741.DAT.

Amino acid sequences were first aligned with CLUSTAL V (Higgins et al., 1992). The alignments were performed with assigned gap weights of 10 under three weighting methods: equal weighting, the Dayhoff 100 matrix, and the Dayhoff 250 matrix (Dayhoff et al., 1978). After ambiguous regions were removed, the amino acid alignments were used to align the original DNA sequences, which were then aligned with MALIGN 1.91 (Wheeler and Gladstein, 1994) given a nucleotide substitution cost of 2 and gap weights of 10, 5, and 3 (additional gaps were weighted the same as initial gaps). As before, regions that differed among weighting schemes were eliminated. In a few cases, complementary single base-pair gaps that disrupted the reading frame over a very short region (3-6 base pairs) were ignored. In general, the truncated sequence alignments produced with CLUSTAL V did not change greatly when the DNA sequences were compared using MALIGN 1.91.

Incongruence Length Difference Test

The incongruence length difference (ILD) test compares the number of steps required for minimum-length trees in separate and combined analyses. In this test, between-partition incongruence is measured by the additional steps required when the data are combined in a simultaneous analysis (Mickey and Farris, 1981; Farris et al., 1994). The statistical distribution of the ILD is determined by randomization (Farris et al., 1994). The ILD test (also called the partition homogeneity test in PAUP*; Swofford, 1997) can be simultaneously applied to any number of data partitions and can be applied under any character weighting method. In the following analyses, invariant characters were always removed before applying the ILD test (see Cunningham, 1997). Removing invariant characters is especially important when the original data partitions differ in the percentage of variable characters, as is often the case when comparing morphological and molecular data.

Weighting Methods

A total of nine weighting methods were applied to each data partition. These methods were chosen to represent a range of possible approaches and include equally weighted parsimony, transversion parsimony, successive weighting, and six variants of six-parameter parsimony. Equally weighted parsimony was implemented by
invoking the unordered states option in PAUP* 4.0, and transversion parsimony was implemented by constructing a step matrix in which all transitions were given a weight of zero.

In successive weighting, each character is weighted according to the degree of homoplasy it shows on the most-parsimonious tree (Farris, 1969). Characters with high levels of homoplasy are given relatively lower weights. Successive weighting was applied using MacClade 3.05 (Maddison and Maddison, 1992). First, each data partition was mapped onto its own set of minimum-length trees obtained using equally weighted parsimony. Then, weights were calculated according to the mean rescaled consistency index across all trees for each character, using integral weights on a scale of 1 to 10 (identical results were obtained if the weights were based on the consistency index or on the retention index; results not shown).

Six-parameter parsimony is a special case of generalized parsimony, where a cost is assigned to the transformation from any character state to any other (Sankoff and Cederberg, 1983; Swoford and Olsen, 1990). For nucleotide data, six substitution classes were considered in this study (A ↔ C, A ↔ G, A ↔ T, C ↔ G, C ↔ T, G ↔ T). Each of the six classes is given a weight based on the observed frequency in the data (Williams and Fitch (1989, 1990). The object of this weighting is that frequent changes are considered more likely than rare changes to have experienced homoplasy. Six-parameter parsimony has the advantage of designing a reconstruction model based on the data partition being analyzed. Unfortunately, there are any number of ways to calculate substitution frequencies from the observed number of substitutions or to calculate weights from those substitution frequencies. The six methods I chose represent only a few of these possibilities.

For all methods, each data partition was mapped onto its own set of minimum-length trees obtained with equally weighted parsimony. The estimated number of substitutions in each nucleotide class was taken as the average across all possible character state reconstructions (MacClade 3.05). The weights were calculated as follows:

\[ K_{ij} = -\ln(X_{ij}/X) \] (LN weighting) \hspace{1cm} (1)

\[ K_{ij} = 1/(X_{ij}/X) \] (INV weighting) \hspace{1cm} (2)

\[ K_{ij} = -\ln[X_{ij}/(X_i + X_j)] \]
(based on Wheeler, 1990) \hspace{1cm} (3)

\[ K_{ij} = 1/[X_{ij}/(X_i + X_j)] \] \hspace{1cm} (4)

\[ K_{ij} = -\ln[X_{ij}/(N_i + N_j)] \]
(based on Rodrigo, 1992) \hspace{1cm} (5)

\[ K_{ij} = 1/[X_{ij}/(N_i + N_j)] \]
(6)

where \( K_{ij} \) is the cost from going from nucleotide \( i \) to nucleotide \( j \) or from \( j \) to \( i \), \( X_{ij} \) is the observed number of changes between \( i \) and \( j \) in either direction, \( X \) is the number of changes from \( i \) to any other nucleotide, \( N \) is the total number of changes on the tree, and \( N_i \) is the estimated proportion of \( i \) nucleotides in the data matrix multiplied by the total length of the sequence. These calculations were performed by importing a chart calculated by MacClade 3.05 into a Microsoft Excel 4.0 spreadsheet (the spreadsheet used to convert these charts into step matrices is available on request). The symmetric step matrices based on these weights were then corrected with MacClade 3.05 for violations of the triangle inequality. Correcting for the triangle inequality ensures that changing directly from one nucleotide to another is never more expensive than passing through an intermediate nucleotide state.

Although several of these equations (Eqs. 1–3) are similar to those calculated by MacClade 3.05, the weights used in this study were not generated using MacClade's "chart to type" command. That command requires the user to specify the range of weights (e.g., 1–50). The method used here, following Williams and Fitch (1990), allows the data themselves to determine the range between the highest and lowest weights.
Phylogenetic Analysis

Branch-and-bound searches were performed using PAUP* 4.0d54 (Swofford, 1997). All bootstrap analyses included 1,000 pseudoreplicates. For bootstrapping, characters were sampled with equal probability and weights were applied. For ILD testing with six-parameter parsimony, each character was assigned to a step matrix derived from its own substitution frequencies. When the characters were randomized, each character retained its original assignment regardless of which partition it was assigned to.

Testing for Phylogenetic Signal

Some weighting methods sharply decreased phylogenetic resolution. To quantify this decrease, it was necessary to use a measure of phylogenetic signal that can be applied under a variety of weighting methods. I used the permutation tail probability (PTP) test of Faith and Cranston (1991) as implemented in PAUP* 4.0d54. This test has been criticized because it can incorrectly detect phylogenetic signal where none exists (Alroy, 1994). However, in the present study the PTP test was used strictly to identify cases where phylogenetic signal has been lost, not gained. In all cases, only ingroup taxa were randomized because the node defined by an outgroup is often very strongly supported and can overshadow loss of resolution in less strongly supported ingroup nodes.

Percentage of Clades Correct Index

The degree of support of each data partition for the expected tree was calculated using the index of percentage of clades correct (%CC; Hillis et al., 1994; Cunningham, 1997). For each bootstrap replicate, the percentage of clades that match the expected phylogeny is determined for the set of most-parsimonious trees. This statistic is then averaged across all bootstrap replications. For any tree, %CC is strongly correlated with the symmetric distance from the expected tree (Penny and Hendy, 1985). A major advantage of the %CC is that it can be rapidly calculated from the table of partition frequencies for each bootstrap run by simply averaging the partition frequencies for each clade that corresponds to the expected tree (see Cunningham, 1997).

RESULTS

For each set of data partitions, I examined the relationship between phylogenetic accuracy of each data partition and the degree of congruence as measured by the ILD test. This relationship was examined in detail for five of the weighting methods and summarized graphically for all nine methods. For each set of partitions, equally weighted parsimony was used as the standard for comparison among methods.

Rodent Phylogeny: Partitioned by Gene

As reported by Sullivan et al. (1995), the Cyt b gene strongly supported the expected phylogeny under equally weighted parsimony (Fig. 2a). With the 12S gene, there was strong bootstrap support for the incorrect placement of Peromyscus eremicus (89% bootstrap). Both genes showed highly significant phylogenetic signal, and incongruence between the two genes was significant ($P < 0.02$ ILD).

Transversion parsimony not only decreased the phylogenetic accuracy of the 12S gene relative to equally weighted parsimony, but it dramatically reduced phylogenetic resolution (Fig. 2b). This loss of resolution was indicated by 135 most-parsimonious trees and a sharp reduction in phylogenetic signal ($P > 0.44$ PTP). Under transversion parsimony, the ILD test found the two genes to be entirely congruent ($P = 1.0$). This congruence resulted because one of the 135 most-parsimonious trees supported by the 12S gene matched a tree also supported by Cyt b. In this case, the congruence measured by the ILD appears to be attributable more to a loss of resolution than to any great improvement in phylogenetic accuracy.

Although successive weighting was strongly influenced by the starting trees, the mean accuracy of partitions increased (Fig. 2c). This increased accuracy was ac-
Figure 2. Phylogenetic analysis of mitochondrial DNA from a strongly supported rodent phylogeny under a variety of weighting schemes. The phylogenies shown represent the strict consensus of the most-parsimonious trees. Taxon labels are only included for those taxa whose relationships differ from the expected phylogeny in Figure 1a. Numbers at nodes refer to the results of 1,000 bootstrap pseudoreplicates.
companied by strongly decreased congruence ($P < 0.001$). Using six-parameter step matrices, the accuracy of both genes improved slightly (Figs. 2d, 2e), enough so that congruence increased (LN, $P < 0.05$; INV, $P < 0.09$). For both forms of step matrix weighting, the phylogenetic signal of the genes remained highly significant ($P < 0.001$ PTP).

Higher Level Vertebrate Phylogeny: Partitioned by Gene

Under equally weighted parsimony, only one gene supported the expected vertebrate phylogeny (COII; Fig. 3a). Although the other genes supported different relationships, none of the incorrect nodes were strongly supported (all incorrect nodes $<71\%$). This weakly supported incongruence was also reflected by a nonsignificant ILD value ($P < 0.11$).

Under transversion parsimony, congruence between genes increased sharply (ILD $P > 0.93$, Fig. 3b). Although this greatly increased congruence was accompanied by a modest increase in phylogenetic accuracy of the individual data partitions, the major reason for the increased congruence appears to be the loss of resolution. This decreased resolution was reflected in the greater number of most-parsimonious trees for two of the genes (three trees for ATP6, two trees for COII); a third gene (Cyt b) showed a decrease in phylogenetic signal, as measured by the PTP test.

In all cases, successive weighting simply reinforced the starting tree, whether or not it was correct (Fig. 3c). Because four of the five starting trees were incorrect, this circularity sharply decreased congruence without increasing accuracy. In all cases, the existing phylogenetic signal, as measured by the PTP test, was amplified for each gene, even when that signal was misleading.

When either INV or LN six-parameter step matrices were applied to the data, phylogenetic accuracy ($\%CC$) increased for all five genes relative to equally weighted parsimony, and the overall congruence also increased (Figs. 3d, 3e). The improved accuracy was also reflected by the observation that under both INV and LN four of the five genes supported the expected tree (Figs. 3d, 3e).

Higher Level Vertebrate Phylogeny: Partitioned by Codon Position

In one group of analyses, data from all five genes were pooled and then partitioned by codon position. In all cases, first and second codon positions strongly supported the correct tree (Figs. 4a–e). Thus, the weighting methods had the most effect on third positions.

Under equally weighted parsimony, third positions strongly supported the incorrect bird/human clade (97% bootstrap, Fig. 4a). The incongruence between codon positions was highly significant ($P < 0.00001$ ILD), and this misleading phylogenetic signal was significant ($P < 0.02$, PTP). The misleading signal in third positions appears to be a result of saturation by multiple substitutions, as suggested by the observation that only 15% of third positions are invariant compared with 66% and 84% of first and second positions, respectively. Furthermore, the transition/transversion ratio of third positions (0.73) is sharply lower than that of first and second positions (1.0 and 1.4, respectively; calculated by mapping substitutions onto the expected tree using MacClade 3.05). A drop in this ratio is expected as data approach saturation (Holmquist, 1983; DeSalle et al., 1987; Larson, 1994). For third positions, the drop is greater than predicted by differences in base composition (Holmquist, 1983; analysis not shown).

Under transversion parsimony (Fig. 4b), the phylogenetic accuracy of third positions increased substantially, and congruence among codon positions also increased (ILD $P < 0.32$). As before, successive weighting was entirely circular, always increasing support for the starting tree (Fig. 4c). Of the six-parameter parsimony methods, only INV increased accuracy of third positions (Fig. 4e). This improvement was reflected in an increase in congruence of over two orders of magnitude (Fig. 4e).
FIGURE 3. Phylogenetic analysis of mitochondrial DNA, partitioned by gene, from a strongly supported vertebrate phylogeny. The phylogenies shown represent the strict consensus of the most-parsimonious trees. Taxon labels are only included for those taxa whose relationships differ from the expected phylogeny in Figure 1b. Numbers at nodes refer to the results of 1,000 bootstrap pseudoreplicates.
Under LN, there was little change in the accuracy or the degree of incongruence.

Relationship between Congruence and Phylogenetic Accuracy

The relationship between congruence and accuracy was investigated for the five methods considered in Figures 2–4 and for four additional six-parameter parsimony methods. For these nine weighting methods, the log of the mean ILD value was plotted against the mean %CC of the individual data partitions. For heuristic purposes, a regression line was drawn through the points corresponding to the six-parameter methods; this line is not in-
tended to imply that these points satisfy the assumptions of regression analysis.

When the data from each phylogeny were partitioned by gene, equally weighted parsimony falls near the line drawn through the points for the six-parameter methods (Figs. 5a, 5b). For both phylogenies, transversion parsimony falls well below the line, suggesting that transversion parsimony has a relatively high ILD value for its degree of phylogenetic accuracy. In both phylogenies, successive weighting falls well above the line, with an ILD value lower than that expected from its degree of phylogenetic accuracy. When the higher level vertebrate phylogeny was partitioned by codon position, all methods examined show a generally log-linear relationship between ILD and %CC (Fig. 5c).

**DISCUSSION**

Congruence between data partitions does not necessarily imply phylogenetic accuracy. Because congruence depends in part on the phylogenetic method being used, an inappropriate reconstruction method can introduce systematic error. For example, a reconstruction method that simply joins taxa according to their order in the data matrix guarantees perfect topological congruence among data partitions, but this congruence bears no relation to phylogenetic accuracy.

In this study, I examined the relationship between congruence and accuracy for a specific set of character weighting methods. These modifications of the parsimony reconstruction model are based on the principle that character changes most likely to result in homoplasy should be given lower weights. Although these methods are based on similar principles, they vary not only in their accuracy but also in how this accuracy relates to congruence. These methods were evaluated in an iterative framework (Bull et al., 1993), whereby congruence and accuracy are measured before and after modifying a simple phylogenetic reconstruction model: equally weighted parsimony.

**FIGURE 5.** The relationship between congruence and phylogenetic accuracy for three sets of data partitions. For all comparisons, there were $10^9-10^6$ replications of the ILD test. The phylogenetic accuracy is measured as the mean percentage of correct clades (%CC) across all partitions for that method. The regression lines shown are drawn through the triangles corresponding to the six-parameter parsimony methods. Open triangles refer to six-parameter methods whose weights were based on the negative natural log of the substitution frequency, and the closed triangles refer to methods whose weights were based on the inverse of the frequency. The largest open and closed triangles refer to the LN and INV methods, respectively (Figs. 2-4). (a) Rodent phylogeny; $R^2 = 0.82$. (b) Higher level vertebrate phylogeny partitioned by gene; $R^2 = 0.61$. (c) Higher level vertebrate phylogeny partitioned by codon position; $R^2 = 0.81$. 
Phylogenetic Accuracy, Congruence, and Character Weighting

Transversions are generally less common than transitions (Brown et al., 1982; Swofford and Olsen, 1990), and transversion parsimony accordingly gives all transitions a weight of zero. For the data partitions considered here, transversion parsimony had the greatest effect on accuracy when it was applied to codon positions (Fig. 4b). In most other partitions, the effect on accuracy was modest (Figs. 2–4). Whenever it was applied, however, transversion parsimony sharply increased character congruence among partitions. This increased congruence was usually associated with a decrease in phylogenetic signal, as measured by the PTP test, and by a general increase in the overall number of most-parsimonious trees (Figs. 2–4). For transversion parsimony, it appears that the increased congruence between partitions had less to do with improved accuracy than with the loss of resolution caused by ignoring transitions. In a similar study of vertebrate mitochondrial DNA sequences, Allard and Carpenter (1996) also found that transversion parsimony increased congruence, according to the ILD test, without a corresponding improvement in accuracy.

The sharp increase in congruence associated with loss of resolution under transversion parsimony is caused by a property of the measure of congruence used in this study. In the ILD test, partitions are perfectly congruent if they share even one most-parsimonious tree (Swofford, 1991). For the rodent phylogeny, transversion parsimony yielded 135 shortest trees for the 12S partition and 3 for the Cyt b partition. Because these partitions had one tree in common, the ILD test indicated perfect congruence even though the accuracy of both partitions decreased when compared with the results of equally weighted parsimony (Fig. 2).

Successive weighting uses the set of trees from an equally weighted parsimony analysis as a basis for weighting individual characters (Farris, 1969). The degree of homoplasy shown by each character is calculated, and the character is weighted accordingly. For the data partitions considered here, successive weighting was entirely circular and always greatly increased support for the starting trees. Hence, if the partitions supported different trees under equally weighted parsimony, this incongruence was greatly amplified under successive weighting (Figs. 2–4). Because correct nodes were strengthened along with incorrect nodes, this method does not show a consistent relationship between congruence and phylogenetic accuracy (Fig. 5). The influence of the starting tree in successive weighting is so great that when codon positions were reweighted according to the tree for third positions (Fig. 4a), the incorrect starting tree was strongly supported by all three positions (results not shown).

Like successive weighting, weights for six-parameter parsimony are based on the set of trees from an equally weighted parsimony analysis (Williams and Fitch, 1990). Instead of considering each character individually, as in successive weighting, six-parameter parsimony is based on the observed frequency of each of the six nucleotide substitution classes. This method allows for heterogeneity of substitution frequencies within as well as between transitions and transversions. For most of the data partitions examined in this study, six-parameter parsimony improved accuracy relative to equally weighted parsimony. Of the weighting methods examined, only six-parameter parsimony showed a consistent positive relationship between congruence and phylogenetic accuracy (Fig. 5). This positive relationship is more consistent than the success of any of the individual six-parameter weighting methods. Although no individual method was always the most successful, the six-parameter method with the highest degree of congruence was always very close to being the most accurate of the methods examined (Fig. 5).

Unlike successive weighting, six-parameter parsimony does not appear to be strongly influenced by the starting trees. When the higher level vertebrate phylange-
ny was partitioned by gene, only one gene supported the correct tree under equally weighted parsimony (Fig. 3a). When six-parameter parsimony was applied to the genes, four of five genes supported the correct tree (Figs. 3d, 3e). This increase in accuracy was observed even when the six-parameter weights were based on incorrect starting trees.

The success of six-parameter parsimony methods is striking considering their simplicity. Because they are inferred by parsimony, these weights do not consider the effects of substitutional saturation on the frequencies of nucleotide substitution classes and thus are subject to biases in character state reconstruction caused by base composition (Collins et al., 1994b). Like any parsimony method, six-parameter parsimony suffers from the difficulty of applying a single model to all parts of the tree, even though substitution frequencies may change among lineages. Other proposed methods for determining six-parameter weights need to be evaluated with these and other data sets (e.g., Williams and Fitch, 1989, 1990; Knight and Mindell, 1993; Collins et al., 1994a).

**Congruence, Accuracy, and Likelihood**

Unlike the parsimony methods considered here, maximum likelihood provides an explicit framework for choosing among nested models of DNA sequence evolution (e.g., Goldman, 1993; Yang et al., 1994). In some cases, however, sophisticated models of DNA sequence evolution greatly improve likelihood but make little difference in phylogenetic accuracy. For the 12S gene from the rodent phylogeny considered in this study, Sullivan et al. (1995) found that a model allowing rate heterogeneity among sites greatly improved likelihood, but they noted no improvement in accuracy. In a study of multiple genes from a four-taxon amniote phylogeny, Huelsenbeck and Bull (1996) compared two models of DNA sequence. For all five genes they considered, a simple model (Jukes and Cantor, 1969) fit the data significantly worse than did a more sophisticated model of DNA sequence evolution that allowed different rates of transitions and transversions and rate heterogeneity among sites (Hasegawa et al., 1985; Huelsenbeck and Bull, 1996). Despite a great improvement in likelihood, the more sophisticated model did not improve accuracy; under both models, only two of the five genes recovered the expected phylogeny.

Although the lack of improvement in phylogenetic accuracy for the amniote phylogeny is not reflected by the likelihood value of each model, a likelihood ratio test of incongruence resulted in identical P values for incongruence under both models (P < 0.03; Huelsenbeck and Bull, 1996). These preliminary results suggest that congruence among data partitions may be useful for evaluating models of DNA sequence evolution.

**The Importance of an Iterative Framework**

The iterative framework used here measures congruence and phylogenetic signal before and after revising a parsimony reconstruction model. Because there was strong corroboration for the relationships among the taxa considered in this study, it was possible to ask how these measures related to phylogenetic accuracy. In most other studies, however, relationships among taxa are not known. For the taxa included in this study, the iterative procedure would have been successful even if the phylogeny were not known, so long as congruence and phylogenetic signal were considered together.

- If congruence increased relative to equally weighted parsimony and phylogenetic signal did not decrease, the weighting method always increased the mean accuracy of the partitions involved.
- If congruence increased relative to equally weighted parsimony but was accompanied by a drop in phylogenetic signal and/or an increase in the number of shortest trees, the degree of congruence was a poor predictor of phylogenetic accuracy. This was the case for transversion parsimony, which always increased congruence.
and generally lost phylogenetic signal or resolution relative to equally weighted parsimony but varied greatly in its effect on accuracy. Despite this increased congruence, overall accuracy of transversion parsimony only increased when applied to codon positions (Fig. 4b).

Considering phylogenetic signal in an iterative framework had especially interesting results when applied to codon positions. Under equally weighted parsimony, third positions produced a significant but misleading phylogenetic signal. Under both transversion and INV parsimony, the signal for these data was no longer significant. An absence of significant phylogenetic signal has been proposed as a criterion for excluding data from further analyses (Faith and Cranston, 1991; Hillis, 1991). By this criterion, the misleading third positions would have been excluded after modifying the reconstruction model. However, the criterion of phylogenetic signal suffers from a logical difficulty. If data are divided into too many partitions, the sample size in any partition might drop so low that no significant result is obtained (J. Thorne, pers. comm.), thus leading to exclusion of potentially important data. In partitions with a sufficiently large sample size, however, phylogenetic signal may indeed be an appropriate criterion for data inclusion.

Recommendations
1. Tests for congruence and phylogenetic signal should be carried out iteratively, before and after objectively modifying the reconstruction model to better fit the data. This recommendation also holds for cases where the data themselves are being modified to better meet the assumptions of the reconstruction method, as when objectively improving DNA sequence alignment (Kjer, 1995).
2. If the degree of congruence decreases or remains the same, the modifications to the reconstruction model or to the data should be discarded.
3. If the degree of congruence increases and resolving power of the data set goes down dramatically, caution is in order. The loss of resolution may be due to the phylogenetic method (e.g., transversion parsimony) or to reduction of misleading signal (e.g., third codon positions). Poorly resolved data sets may yield a high degree of congruence, but this congruence is not necessarily desirable.
4. If the degree of congruence increases and the resolving power increases or stays the same, the modification to the reconstruction model should be retained with appropriate caution.
5. Exclude invariant characters before applying the ILD test.

Acknowledgments
This study was inspired by the work of A. Graybeal (1994). Versions of this paper have benefitted from the comments of P. Fritsch, J. Thorne, P. Manos, J. Hulsenbeck, M. Leslie, T. Oakley, K. Omland, D. Swofford, R. Vilgalys, B. Wiegmann, and R. Zechman. Careful reviews by T. Collins, C. Simon, and J. Sullivan were especially valuable. I thank D. Swofford for allowing me to use a test version of PAUP*.

References


Received 12 August 1996; accepted 17 January 1997

Associate Editor: Chris Simon