

# Can Three Incongruence Tests Predict When Data Should be Combined?

Clifford W. Cunningham

Zoology Department, Duke University

Advocates of conditional combination have argued that testing for incongruence between data partitions is an important step in data exploration. Unless the partitions have had distinct histories, as in horizontal gene transfer, incongruence means that one or more data partitions support the wrong phylogeny. This study examines the relationship between incongruence and phylogenetic accuracy using three statistical tests of incongruence. These tests were applied to pairs of mitochondrial DNA data partitions from two well-corroborated vertebrate phylogenies. Of the three tests, the most useful was the incongruence length difference test (ILD, also called the partition homogeneity test). This test distinguished between cases in which combining the data generally improved phylogenetic accuracy ( $P > 0.01$ ) and cases in which accuracy of the combined data suffered relative to the individual partitions ( $P < 0.001$ ). In contrast, in several cases, the Templeton and Rodrigo tests detected highly significant incongruence ( $P < 0.001$ ) even though combining the incongruent partitions actually increased phylogenetic accuracy. All three tests identified cases in which improving the reconstruction model could improve the phylogenetic accuracy of the individual partitions.

## Introduction

One of the outstanding issues in systematics is how to proceed when different data partitions collected from the same taxa support conflicting phylogenies (Kluge 1989; Bull et al. 1993; Rodrigo et al. 1993; de Queiroz, Donoghue, and Kim 1995; Miyamoto and Fitch 1995; Huelsenbeck, Bull, and Cunningham 1996). At the heart of the controversy is whether conflicting data partitions should be analyzed separately (Lanyon 1993; Miyamoto and Fitch 1995) or combined in a simultaneous analysis (Kluge 1989; Barrett, Donoghue, and Sober 1991). Supporters of a third alternative—conditional combination—have suggested that this decision depends on the degree of incongruence (Bull et al. 1993; de Queiroz 1993; Larson 1994; Huelsenbeck, Bull, and Cunningham 1996). While weak incongruence can arise from inadequate sample sizes (Bull et al. 1993; Omland 1994), strong incongruence bears further investigation. Cases of strong incongruence may indicate that partitions have had different histories or that one or more data partitions violate the assumptions of the phylogenetic method.

In some cases, improving the fit between the data and the assumptions of the reconstruction method may reduce the degree of incongruence. For example, most phylogenetic methods assume that characters are homologous. Incorrect DNA sequence alignments can violate this assumption. Kjer (1995) showed that objectively improving amphibian 18S rDNA sequence alignments increased congruence between the DNA sequences and morphologically based taxonomy. Alternatively, the reconstruction method itself can be modified to better fit the data. For example, equally weighted parsimony makes the extreme assumption that all classes of nucleotide substitutions are equally likely. One alternative is to weight nucleotide changes based on their ob-

served frequency in the original data (Williams and Fitch 1990). Marshall (1992) found that applying this method to vertebrate 18S rDNA sequences increased congruence with phylogenies based on the fossil record.

In both of these examples, measuring the degree of congruence before and after adjusting the fit between the data and the reconstruction model was a valuable guide to phylogenetic inference. But what of cases in which the source of incongruence is not identified? Supporters of conditional combination argue that strongly incongruent data partitions should be considered individually until more information is available. Bull et al. (1993) used computer simulations to show that combining strongly incongruent data can reduce phylogenetic accuracy relative to individual partitions, even when those partitions have identical histories. In this study, I test the hypothesis that combining strongly incongruent data partitions can reduce phylogenetic accuracy.

A number of statistical tests have been proposed to investigate incongruence (e.g., Bull et al. 1993; Rodrigo et al. 1993; Farris et al. 1994; Larson 1994; Huelsenbeck and Bull 1996). To date, however, there have been relatively few investigations of the empirical properties of these tests (but see Larson 1994; Lutzoni and Vilgaly 1995; Huelsenbeck, Bull, and Cunningham 1996; Sullivan 1996; Lutzoni 1997). In this paper, I apply three parsimony-based incongruence tests to mitochondrial DNA sequences from two well-corroborated vertebrate phylogenies (Sullivan, Holsinger, and Simon 1995; Graybeal 1994). These tests include two tests of character incongruence (Templeton 1983; Farris et al. 1994) and one test of topological incongruence (Rodrigo et al. 1993). First, I apply the three incongruence tests to each pair of data partitions. Then, I compare phylogenetic accuracy of the individual and combined data partitions. Finally, I ask whether the degree of incongruence detected by a test is sufficient to reduce the accuracy of the combined data.

## Materials and Methods

### Two Well-Corroborated Phylogenies

The first phylogeny was proposed by Sullivan, Holsinger, and Simon (1995) for the rodent genera *Pero-*

Key words: conditional combination, incongruence tests, known phylogenies, phylogenetic accuracy.

Address for correspondence and reprints: Clifford Cunningham, Zoology Department, Duke University, Durham, North Carolina 27705. E-mail: cliff@acpub.duke.edu.

*Mol. Biol. Evol.* 14(7):733–740. 1997

© 1997 by the Society for Molecular Biology and Evolution. ISSN: 0737-4038

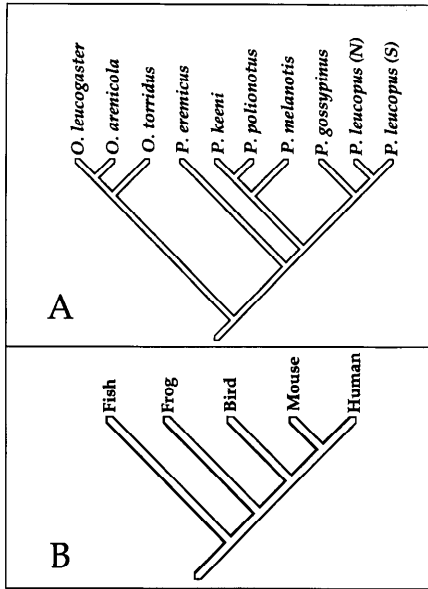


FIG. 1.—Two well-corroborated phylogenies. *A*, Phylogeny of the rodent genera *Peromyscus* and *Onychomys* (from Sullivan, Holsinger, and Simon 1995, fig. 1). Because the relationships within *Onychomys* are not certain, they are not included when determining phylogenetic accuracy. *B*, Higher-level vertebrate phylogeny based on figure 2 of Graybeal (1994).

*myscus* and *Onychomys* (fig. 1A). Because the relationships within the genus *Onychomys* are not well understood (J. Sullivan, personal communication), the node supporting *O. leucogaster* and *O. arenicola* will not be considered when evaluating phylogenetic accuracy. Two mitochondrial genes from this phylogeny have been used to investigate the effects of among-site variation and combination of data on phylogenetic accuracy (Sullivan, Holsinger, and Simon 1995; Sullivan 1996). I used the sequence alignment described in Sullivan, Holsinger, and Simon (1995). Like them, I excluded the ambiguously aligned positions 85–87 and 280–283.

The second well-corroborated phylogeny, shown in figure 1B, was proposed by Graybeal (1994). Five taxa whose entire mitochondrial genomes are available in GenBank were chosen from this phylogeny: fish (*Cyprinus carpio*, GB X61010), frog (*Xenopus laevis*, GB M10217), bird (*Gallus gallus*, GB X52392), mouse (*Mus musculus*, GB J01420), and human (*Homo sapiens*, GB J01415). Five genes were chosen because a preliminary analysis showed varying degrees of support for the expected phylogeny: ATPase 6; cytochrome oxidase I, II, and III; and cytochrome *b*. These will be abbreviated as ATP6, COI, COII, COIII, and *Cytb*. First, the mitochondrial DNA sequences were partitioned by gene. Then all five genes were pooled together and partitioned by codon position. Regions of uncertain homology were removed using a method described by Cunningham (1997). This method is a variant of that proposed by Gatesy, DeSalle, and Wheeler (1993).

#### Measuring Phylogenetic Accuracy

The degree to which each data partition supported the expected tree was calculated using the % clades-

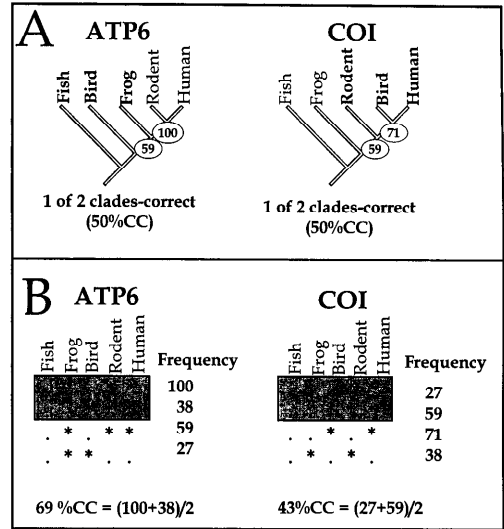


FIG. 2.—Measuring phylogenetic accuracy using the % clades correct (%CC) index for two mitochondrial genes from the higher-level vertebrate phylogeny. The %CC is directly correlated with the symmetric distance (Penny and Hendy 1985) from the expected tree. *A*, Measuring %CC for the most parsimonious trees (MPTs) for each gene. Taxa whose relationships depart from the expected phylogeny are in bold. Numbers in circles are bootstrap proportions for 1,000 pseudoreplications. *B*, Measuring the mean %CC across the MPTs from each of 100 bootstrap pseudoreplications is equivalent to averaging the partition frequencies for the expected nodes (shaded).

correct index (%CC; Hillis, Huelsenbeck, and Cunningham 1994). For any tree, the %CC is simply the percentage of clades in the observed tree that match clades in the expected tree. For example, when “fish” is designated as the outgroup in the five-taxon phylogeny there are two clades that are free to vary. The most parsimonious trees (MPTs) for two mitochondrial genes are shown in figure 2A. The ATP6 tree has one clade that matches the expected relationships (rodent, human) and one that does not (frog, rodent, human). The COI tree also has one clade that supports the expected tree (bird, rodent, human) and one that does not (bird, human). When applied to these MPTs, both genes show an identical degree of support for the expected tree, 50%CC. But a casual inspection of the bootstrap values suggests that the level of support for the expected tree is not the same. In the ATP6 tree, the correct clade found in the MPT is supported by 100% of bootstrap trees. In the COI tree, the correct clade found in the MPT is much more weakly supported (59%).

A more sensitive technique for determining the degree of support for the expected tree involves calculating the %CC for each set of MPTs for each of 1000 pseudoreplicates. Taking the mean %CC across all bootstraps gives the bootstrapped %CC. Conveniently, this value is the same as that derived from simply averaging the bootstrap support for each clade in the expected tree. These values can be taken from a table of bootstrap partition frequencies (fig. 2B). All %CCs reported in this paper are bootstrapped %CCs.

#### Incongruence Tests

##### Templeton Test

Goodness-of-fit tests can be used to measure incongruence between data partitions (Bull et al. 1993;

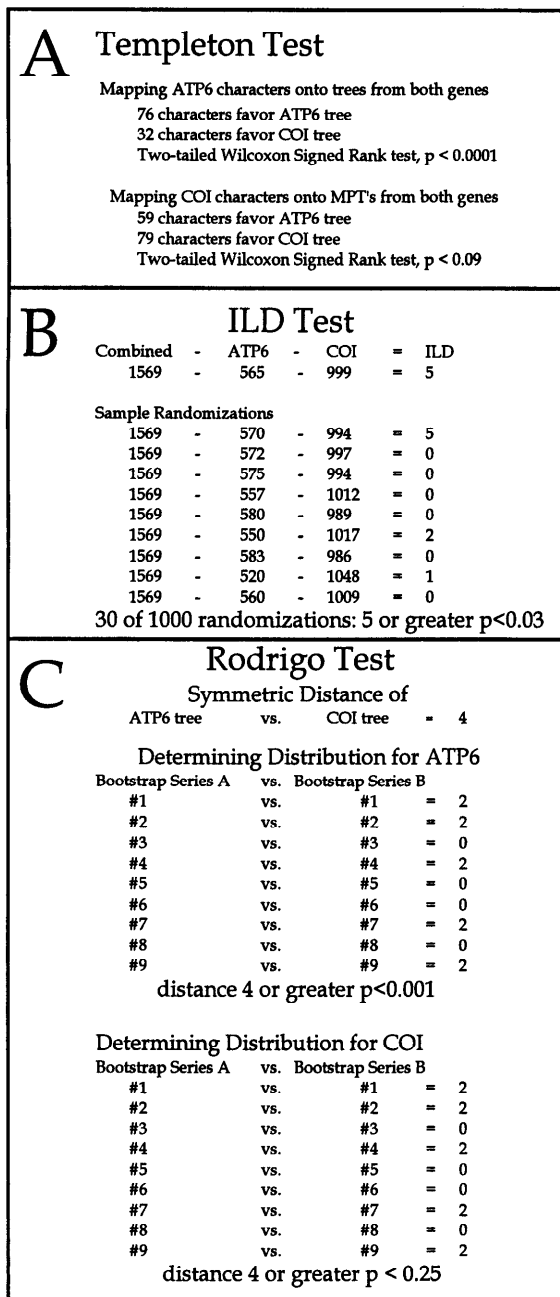


FIG. 3.—Applying three incongruence tests to the pair of genes from fig. 2. Note that, unlike in table 1, the  $P$  values presented here are not corrected for multiple comparisons.

Larson 1994). These tests determine whether alternative topologies differ significantly in how well they fit a data partition (Templeton 1983; Prager and Wilson 1988; Kishino and Hasegawa 1989). In the context of conditional combination, each data partition in turn is mapped onto the MPTs from both data partitions (Bull et al. 1993; Larson 1994). Then, for the Templeton test, the numbers of steps required for each character on each topology are compared using a two-tailed Wilcoxon signed-rank test (fig. 3A; Templeton 1983). The results presented here were obtained using the method described by Larson (1994). For the data partitions considered here, the Templeton test gives results identical

to Kishino and Hasegawa's (1989) modification of the Templeton test as implemented in the DNAPARS program in PHYLIP 3.5c (Felsenstein 1995).

#### The Incongruence Length Difference (ILD) Test

The difference between the numbers of steps required by individual and combined analyses is the incongruence length difference (ILD, fig. 3B; Mickevich and Farris 1981; Farris et al. 1994). The distribution of the ILD statistic can be estimated by calculating the ILD first for the original partition and then for a series of randomized partitions (fig. 3B; Farris et al. 1994). These randomized partitions are the same sizes as the original partitions, but each represents a mixture of characters from each partition. This randomization procedure was applied for each pair of partitions using PAUP\* 4d52 (Swofford 1997, in which the ILD test is called the partition homogeneity test). Invariant characters were always removed before applying the ILD test. This step is especially important when the original data partitions differ in the number of variable characters, as is usually the case when comparing morphological and molecular data.

#### Topological Incongruence Test

Rodrigo et al. (1993) proposed three tests for evaluating incongruence between data partitions. The second is the most appropriate for determining whether partitions are significantly incongruent (discussed in detail by Lutzoni and Vilgalys 1995; Lutzoni 1997). This second test will be referred to henceforth as the Rodrigo test. There are two steps for the Rodrigo test. First, the symmetric distance (SD) is calculated between the MPTs from each data partition (Penny and Hendy 1985). Then, the distribution of this statistic is determined separately for each partition by calculating the mean SD between MPTs from bootstrap pseudoreplicates taken from the same data partition. For example, if every bootstrap pseudoreplicate for a data partition always supports the same tree, the SD between the bootstraps will always be zero. If there is a large amount of variation between the trees supported by each bootstrap, there will be wide distribution of SDs.

The mean SDs between bootstraps were obtained by bootstrapping the same data partition twice with different random number seeds and saving the trees to separate treefiles. The trees being compared were pasted into a new file and imported into PAUP\* 4d52 (Swofford 1997), which calculated the SD between trees from the corresponding pseudoreplicate for each bootstrap run (fig. 3C; Rodrigo et al. 1993; Lutzoni and Vilgalys 1995).

#### Multiple Most-Parsimonious Trees

The Templeton and Rodrigo tests can be difficult to interpret in cases in which there is more than one MPT. This was resolved by considering the most congruent trees from each data partition. If two data partitions share one MPT, the tests report no incongruence ( $P = 1.0$ ). This convention allows the Templeton and Rodrigo tests to be directly compared with the ILD test,

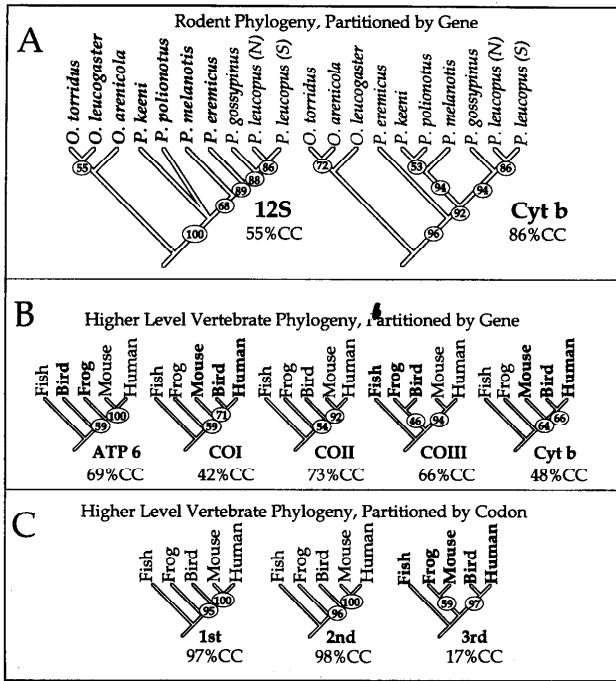


FIG. 4.—Parsimony analysis for mtDNA sequences from a rodent phylogeny (A) and for mtDNA sequences from a higher-level vertebrate phylogeny, partitioned by gene (B) and by codon position (C). Taxa whose relationships depart from the expected phylogeny are in bold. Numbers in circles are bootstrap values from 1,000 pseudoreplications.

which also shows no incongruence in this circumstance (Swofford 1991). In general, this convention will cause the tests to underestimate incongruence (Swofford 1991).

## Results

### Phylogenetic Analysis of Individual Data Partitions

As reported by Sullivan, Holsinger, and Simon (1995), the *Cytb* gene strongly supports the expected rodent phylogeny, while the 12S gene supports an alternative tree (fig. 4A). The major conflict between the genes lies in the strong support by the 12S gene for the placement of *P. eremicus* as the sister group to *P. gossypinus* and *P. leucopus* (89% bootstrap support; fig. 4A).

When mtDNA sequence data from the higher-level vertebrate phylogeny are partitioned by gene, only one of five individual partitions (COII) supports the expected tree (fig. 4B). In fact, the only tree supported by more than one partition places birds and humans as sister taxa relative to rodents (COI, *Cytb*). When the mtDNA sequences from this phylogeny are partitioned by codon position, first and second positions strongly support the expected tree, while third positions strongly support a bird/human sister taxon relationship (97% bootstrap support).

### Incongruence and Phylogenetic Accuracy of Combined Data

The incongruence tests were applied to a total of 14 pairs of data partitions from the two phylogenies (table 1). For these data, the ILD test was the least sensitive

**Table 1**  
**Three Incongruence Tests**

	ILD	CALCULATED WITH RESPECT TO FIRST PARTITION		CALCULATED WITH RESPECT TO SECOND PARTITION	
		Rodrigo	Templeton	Rodrigo	Templeton
<b>Rodent phylogeny</b>					
12S vs. <i>Cytb</i> . . . . .	<0.02	NS	NS	NS	<0.001
<b>Higher-level vertebrate phylogeny (partitioned by gene)</b>					
ATP6 vs. COI . . . . .	NS	<0.001	<0.001	NS	NS
ATP6 vs. COII . . . . .	NS	NS	NS	NS	NS
ATP6 vs. COIII . . . . .	NS	NS	NS	NS	NS
ATP6 vs. <i>Cytb</i> . . . . .	NS	<0.001	<0.001	NS	NS
COI vs. COII . . . . .	NS	NS	NS	NS	NS
COI vs. COIII . . . . .	NS	NS	NS	NS	NS
COI vs. <i>Cytb</i> . . . . .	NS	NS	NS	NS	NS
COII vs. COIII . . . . .	NS	NS	NS	NS	NS
COII vs. <i>Cytb</i> . . . . .	NS	NS	NS	NS	NS
COIII vs. <i>Cytb</i> . . . . .	NS	NS	NS	NS	NS
<b>Higher-level vertebrate phylogeny (partitioned by codon position)</b>					
first vs. second . . . . .	NS	NS	NS	NS	NS
first vs. third . . . . .	<0.001	<0.001	<0.001	NS	NS
second vs. third . . . . .	<0.001	<0.001	<0.001	NS	NS

NOTE.—All *P* values were multiplied by the number of multiple comparisons. *P* values higher than 0.05 are labeled "NS." The significance of the Templeton and Rodrigo tests is determined with respect to each data partition in turn, whereas the ILD test is applied simultaneously to each pair of data partitions. Randomizations for the ILD test were performed 10,000 times. The Rodrigo tests were based on 1,000 bootstraps in most cases. When a significant result was found, the number of bootstraps was increased to 10,000. When the Rodrigo test was applied to the rodent data, only 100 bootstraps were necessary to show a nonsignificant result.

**Table 2**  
Phylogenetic Accuracy of Data Partitions Before and After Combining

	Individual Analyses (Best of Both)	Combined Analysis	Effect of Combining
<b>Rodent Phylogeny</b>			
12S, <i>Cytb</i> .....	86	92/	6
<b>Higher-level vertebrate phylogeny (partitioned by gene)</b>			
ATP6, COI .....	69	97/	28
ATP6, COII .....	73/	82/	9
ATP6, COIII .....	69	74/	5
ATP6, <i>Cytb</i> .....	69	77/	8
COI, COII .....	73/	72/	-1
COI, COIII .....	66	78/	12
COI, <i>Cytb</i> .....	48	46	-2
COII, COIII .....	73/	74/	1
COII, <i>Cytb</i> .....	73/	74/	1
COIII, <i>Cytb</i> .....	66	73/	7
<b>Higher-level vertebrate phylogeny (partitioned by codon position)</b>			
first, second .....	98	99/	1
first, third .....	97	87/	-10
second, third .....	98	80/	-18

NOTE.—Phylogenetic accuracy for each pair of genes is given in terms of % clades-correct measured across 1,000 bootstraps (see fig. 2). A check mark (/) indicates that the correct tree was recovered by the indicated data.

to incongruence (3 of 14 pairs,  $P < 0.05$ ). The Templeton test was the most sensitive, with 5 of 14 pairs significantly incongruent with respect to one partition or the other ( $P < 0.05$ ). To evaluate the effects of different partitions on phylogenetic accuracy, both phylogenies were partitioned by gene, and the higher-level phylogeny was also partitioned by codon position.

#### Partitioning by Gene

Despite the incongruence detected by some of the tests, combining genes generally improved phylogenetic accuracy. This improvement was often dramatic. Although only one gene from each phylogeny supported the expected phylogeny (fig. 4A and B), combining pairs of genes supported the expected phylogeny in all but one case (table 2). Interestingly, the case in which combining the genes gave the wrong phylogeny was also the only case of perfect congruence between genes—both genes supported the same wrong phylogeny to begin with (table 2). The improvement in accuracy was measured by comparing the accuracy of the combined data with the best of the individual partitions. In no case did the accuracy of the combined data go down by more than 2%CC, and for 7 of the 11 pairs of gene partitions the improvement from combining genes was  $\geq 5\%$ CC (table 2).

Clearly, the highly significant incongruence ( $P < 0.001$ ) detected between some pairs of genes by the Templeton and Rodrigo tests was not sufficient to reduce the accuracy of the combined data (table 1). By contrast, the ILD test found only one pair of genes to be significantly incongruent after correcting for multiple comparisons ( $P < 0.02$ ).

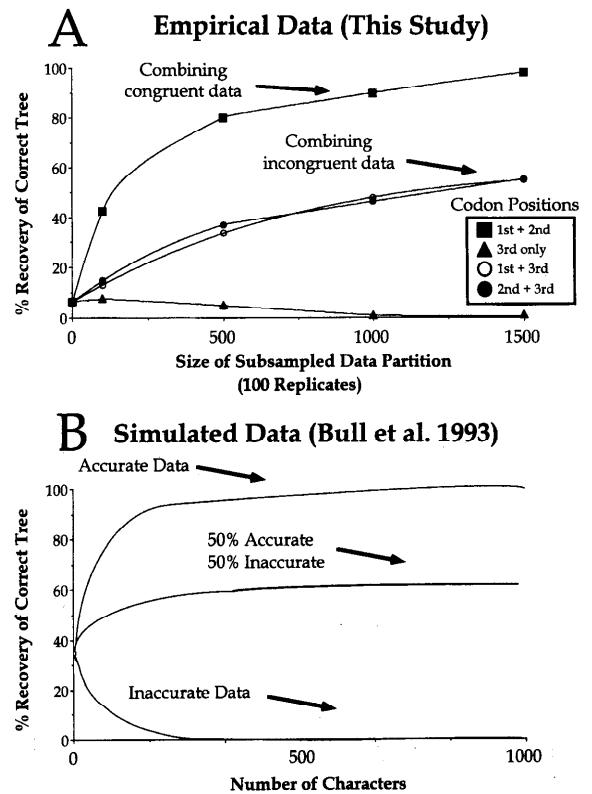


FIG. 5.—A, Power graphs showing the effect of combining congruent and incongruent codon positions. To generate the graphs, the appropriate codon positions from all five genes were pooled, and subsamples of a particular size (e.g., 100 bp) were drawn from the pool without replacement. Each subsample was analyzed with equally weighted parsimony and evaluated to determine whether it supported the correct tree, as described by Cummings, Otto, and Wakeley (1995). B, The effect of combining accurate and inaccurate data generated by computer simulations for a four-taxon phylogeny is shown for comparison (drawn after fig. 4A of Bull et al. 1993).

#### Partitioning by Codon Position

When the data from the higher-level phylogeny were partitioned by codon position, all three tests found highly significant incongruence between third codon positions and both first and second positions (table 1). Unlike the gene partitions considered above, combining incongruent codon positions did reduce the accuracy of the combined data (table 2). Despite this reduced accuracy, the sample size of the combined data was large enough to recover the correct tree.

Two conclusions about sample size can be drawn from the power graphs in figure 5A. First, third codon positions by themselves cause parsimony to be inconsistent, so increasing sample size actually reduces accuracy. This is a classic example of inappropriate data for phylogenetic analysis (Felsenstein 1978; Bull et al. 1993). Second, the combined congruent codon positions recover the correct phylogeny at a rate of 50% just under 150 bp (fig. 5A). By contrast, over 1,100 bp are needed for the same level of accuracy when the incongruent third positions are included. This represents a difference of nearly an order of magnitude. The empirical data from the present study are remarkably similar to those from an earlier computer simulation study of incongru-

**Table 3**  
**Overcoming the Effect of Third Codon Positions in the Higher-Level Vertebrate Phylogeny**

	Equally Weighted Parsimony, All Positions	Equally Weighted Parsimony, No Third Positions	Six-parameter Parsimony, All Positions	Mean Effect of Adding One Gene
ATP6 .....	69	83/	73	83/
COI .....	42	94/	93/	73/
COII .....	73/	73/	89/	76/
COIII .....	66	91/	75/	72/
<i>Cytb</i> .....	48	80/	71/	68/
Mean accuracy .....	60	83	80	74

NOTE.—A check mark (✓) indicates that the correct tree was recovered by the indicated data. Numbers refer to phylogenetic accuracy as measured in %CC. The mean effect of adding one gene was calculated by considering each case in table 2 where the indicated gene was combined with each of the other four genes. The six-parameter stepmatrix was constructed by mapping each gene onto the most parsimonious tree found with equally weighted parsimony (MacClade 3.06, Maddison and Maddison 1992). The substitution frequencies were determined by each of six classes of nucleotide substitution (AC, AG, AT, CG, CT, GT). Each substitution frequency was converted to a weight by taking the absolute value of its natural log and then correcting for the triangle inequality. See Cunningham (1997) for a full description.

ent data from a four-taxon phylogeny (fig. 5B; Bull et al. 1993). For both the simulated and empirical data, the effect of misleading data partitions can be overcome, but only with thousands of characters (fig. 5A and B).

## Discussion

This study compares the performance of three parsimony-based incongruence tests under two criteria. First, how well were the tests able to predict when combining data would increase phylogenetic accuracy? Second, did investigating significant incongruence lead to improved accuracy?

### Incongruence and Phylogenetic Accuracy

Of the three tests, the ILD test was best able to distinguish between the degree of incongruence between genes ( $P > 0.01$ , table 1) and the degree of incongruence between codon positions ( $P < 0.001$ , table 1). In keeping with these differing degrees of incongruence, combining genes generally improved phylogenetic accuracy, whereas combining the highly incongruent codon positions did not (table 2). By contrast, the Templeton and Rodrigo tests found the same degree of incongruence between some genes as between codon positions. For example, both detected highly significant incongruence between the ATP6 and COI genes ( $P < 0.001$ , table 1). Yet combining these genes yielded the largest improvement in phylogenetic accuracy of any pair of data partitions (+28%CC).

### Overcoming Incongruence Between Codon Positions: More Data or Better Analysis?

As shown above, the misleading third codon positions cause only four out of the five genes to support the correct phylogeny (fig. 4B). This is true even though there are only half as many third positions as first and second positions. The misleading signal in third positions appears to be a result of saturation by multiple substitutions. This is suggested by the observation that only 15% of third positions are invariant, as compared

to 66% and 84% for first and second positions, respectively. Furthermore, the transition/transversion ratio of third positions (0.73) is sharply lower than those of first and second positions (1.0 and 1.4, respectively; calculated by mapping substitutions onto the expected tree using MacClade 3.06 [Maddison and Maddison 1992]). A drop in this ratio is expected as data approach saturation (Holmquist 1983; DeSalle et al. 1987; Larson 1994), and for third positions, the drop is greater than predicted by differences in base composition (Holmquist 1983; analysis not shown).

Three approaches were taken to overcome the misleading effect of third positions. First, because of strong evidence for saturation, third positions were excluded, causing all five genes to recover the correct phylogeny under equally weighted parsimony (table 3). Second, third positions were included, but the parsimony reconstruction model was improved by the a posteriori construction of a six-parameter stepmatrix (Williams and Fitch 1990; described in table 3 legend). Objectively improving the reconstruction model in this way caused the %CC to increase for all five genes, and the correct tree was recovered for four of five genes (table 3).

Finally, the effect of misleading data could be overcome by simply adding more data. The power of additional data is dramatized by the observation that in 9 of 10 cases, combining two genes caused equally weighted parsimony to recover the correct tree (table 2). The mean improvement in phylogenetic accuracy from combining each gene with the four remaining genes is shown in table 3. All three approaches improved phylogeny reconstruction. Interestingly, however, less appears to be more. Adding an additional gene improved accuracy less than did simply removing third positions (table 3). Furthermore, adding another gene was not markedly more effective than improving the parsimony reconstruction model with a six-parameter stepmatrix.

## Conclusions

As a "litmus test" for predicting the accuracy of combined data, the ILD test performed the best of the

three tests. Whenever the ILD test found a *P* value greater than 0.01, combining the data improved or did not reduce phylogenetic accuracy. On the other hand, when the ILD detected *P* values lower than 0.001, the combined data suffered relative to the individual partitions. While it is premature to identify the appropriate threshold of incongruence from these results, this study agrees with Sullivan (1996) that a significance threshold of 0.05 may be too conservative for the ILD test. This study also agrees with Lutzoni and Vilgalys (1995) that goodness-of-fit tests such as the Templeton test are too conservative.

As a tool for detecting cases in which investigating incongruence can improve phylogenetic accuracy, all three tests identified highly significant incongruence between codon positions. This incongruence could be successfully overcome by adding more data, deleting third codon positions, or improving the phylogenetic reconstruction model. But even when there is no incongruence, modifying the phylogenetic reconstruction models is prudent. For example, only two genes were perfectly congruent under equally weighted parsimony, but this was because they both supported the wrong tree (COI and *Cytb*). When six-parameter parsimony was applied, both genes were still congruent, but now the reason was because they both supported the correct tree.

The ILD test has other advantages. Like the Rodrigo test, but unlike the Templeton test, the ILD test can be reapplied after the reconstruction model has been adjusted by weighted parsimony to determine whether congruence has increased. Unlike the Rodrigo test, the ILD test is easily implemented in widely available software packages (PAUP\* 4d52 [Swofford 1997], Random Cladistics 4.0 [Siddall 1996]). Finally, the ILD test is the only one of the three tests that can be applied to multiple data partitions simultaneously (Farris et al. 1994). Although the ILD test is presently only implemented with parsimony, there is no reason why it cannot be applied under any minimization criterion, such as maximum likelihood or minimum evolution.

## Acknowledgments

I thank P. Fritsch, J. Thorne, P. Manos, J. Huelsenbeck, M. Leslie, J. Marcus, T. Oakley, K. Omland, D. Swofford, R. Vilgalys, B. Wiegmann, and R. Zechmann for their comments on earlier versions of this manuscript. The thoughtful reviews of Rodney Honecutt and an anonymous reviewer greatly improved the final manuscript. Thanks to J. Huelsenbeck for writing the program which generated the power graphs.

## LITERATURE CITED

- BARRETT, M., M. J. DONOGHUE, and E. SOBER. 1991. Against consensus. *Syst. Zool.* **40**:486–493.
- BULL, J. J., J. P. HUELSENBECK, C. W. CUNNINGHAM, D. L. SWOFFORD, and P. J. WADELL. 1993. Partitioning and combining data in phylogenetic analysis. *Syst. Biol.* **42**:384–397.
- CUMMINGS, M. P., S. P. OTTO, and J. R. WAKELEY. 1995. Sampling properties of DNA sequence data in phylogenetic analysis. *Mol. Biol. Evol.* **12**:814–823.
- CUNNINGHAM, C. W. 1997. Is congruence between data partitions a reliable predictor of phylogenetic accuracy? Empirically testing an iterative procedure for choosing among phylogenetic methods. *Syst. Biol.* (in press).
- DE QUEIROZ, A. D. 1993. For consensus (sometimes). *Syst. Biol.* **42**:368–372.
- DE QUEIROZ, A. D., M. DONOGHUE, and J. KIM. 1995. Separate versus combined analysis of phylogenetic evidence. *Annu. Rev. Ecol. Syst.* **26**:657–681.
- DESALLE, R., T. FREEDMAN, E. M. PRAGER, and A. C. WILSON. 1987. Tempo and mode of sequence evolution in mitochondrial DNA of Hawaiian *Drosophila*. *J. Mol. Evol.* **26**:157–164.
- FAITH, D. P. 1991. Cladistic permutation tests for monophyly and nonmonophyly. *Syst. Zool.* **40**:366–375.
- FARRIS, J. S., M. KÄLLERSJÖ, A. G. KLUGE, and C. BULT. 1994. Testing significance of incongruence. *Cladistics* **10**:315–319.
- FELSENSTEIN, J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. *Syst. Zool.* **27**:401–410.
- . 1981. A likelihood approach to character weighting and what it tells us about parsimony and compatibility. *Biol. J. Linn. Soc.* **16**:183–196.
- . 1995. PHYLIP (phylogeny inference package). Version 3.5c. Distributed by the author, Department of Genetics, University of Washington, Seattle.
- GATESY, J., R. DESALLE, and W. C. WHEELER. 1993. Alignment-ambiguous nucleotide sites and the exclusion of systematic data. *Mol. Phylogenet. Evol.* **2**:152–157.
- GRAYBEAL, A. 1994. Evaluating the phylogenetic utility of genes: a search for genes informative about deep divergences among vertebrates. *Syst. Biol.* **43**:174–193.
- HILLIS, D. M., J. HUELSENBECK, and C. W. CUNNINGHAM. 1994. Application and accuracy of molecular phylogenetics. *Science* **264**:671–676.
- HOLMQUIST, R. 1983. Transitions and transversions in evolutionary descent. *J. Mol. Evol.* **19**:134–144.
- HUELSENBECK, J. P., and J. J. BULL. 1996. A likelihood ratio test to detect conflicting phylogenetic signal. *Syst. Biol.* **45**:92–98.
- HUELSENBECK, J. P., J. J. BULL, and C. W. CUNNINGHAM. 1990. Combining data in phylogenetic analysis. *TREE* **11**:152–158.
- KISHINO, H., and M. HASEGAWA. 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. *J. Mol. Evol.* **29**:170–179.
- KJER, K. M. 1995. Use of rRNA secondary structure in phylogenetic studies to identify homologous positions: an example of alignment and data presentation from the frogs. *Mol. Phylogenet. Evol.* **4**:314–330.
- KLUGE, A. J. 1989. A concern for evidence and a phylogenetic hypothesis of relationships among *Epicrates* (Boidae, Serpentes). *Syst. Zool.* **38**:7–25.
- LANYON, S. M. 1993. Phylogenetic frameworks: towards a firmer foundation for the comparative approach. *Biol. J. Linn. Soc.* **49**:45–61.
- LARSON, A. 1994. The comparison of morphological and molecular data in phylogenetic systematics. Pp. 371–390 in B. SCHIERWATER, B. STREIT, G. P. WAGNER, and R. DESALLE, eds. *Molecular ecology and evolution: approaches and applications*. Birkhauser Verlag, Berlin.
- LUTZONI, F. 1997. Phylogeny of lichen- and non lichen-forming omphalinoid mushrooms and the utility of testing for combinability among multiple data sets. *Syst. Biol.* (in press).

- LUTZONI, F., and R. VILGALYS. 1995. Integration of morphological and molecular data sets in estimating fungal phylogenies. *Can. J. Bot.* **73**:S649–S659.
- MADDISON, W. P., and D. R. MADDISON. 1992. *MacClade: analysis of phylogeny and character evolution*. Version 3.0. Sinauer, Sunderland, Mass.
- MARSHALL, C. R. 1992. Substitution bias, weighted parsimony, and amniote phylogeny as inferred from 18S rRNA sequences. *Mol. Biol. Evol.* **9**:370–373.
- MICKEVICH, M. F., and J. S. FARRIS. 1981. The implications of congruence in *Menidia*. *Syst. Zool.* **30**:351–370.
- MIYAMOTO, M. M., and W. M. FITCH. 1995. Testing species phylogenies and phylogenetic methods with congruence. *Syst. Biol.* **44**:64–76.
- OLMSTEAD, R. G., and J. A. SWEERE. 1994. Combining data in phylogenetic systematics: an empirical approach using three molecular data sets in the Solanaceae. *Syst. Biol.* **43**:467–481.
- OMLAND, K. E. 1994. Character congruence between a molecular and a morphological phylogeny for dabbling ducks (*Anas*). *Syst. Biol.* **43**:369–386.
- PENNY, D., and M. D. HENDY. 1985. The use of tree comparison metrics. *Syst. Zool.* **34**:75–82.
- PRAGER, E. M., and A. C. WILSON. 1988. Ancient origin of lactalbumin from lysozyme: analysis of DNA and amino acid sequences. *J. Mol. Evol.* **27**:326–335.
- RODRIGO, A. G., M. KELLY-BORGES, P. R. BERGQUIST, and P. L. BERGQUIST. 1993. A randomisation test of the null hypothesis that two cladograms are sample estimates of a parametric phylogenetic tree. *N. Z. J. Bot.* **31**:257–268.
- RUSSO, C., N. TAKEZAKI, and M. NEI. 1996. Efficiencies of different genes and different tree-building methods in recovering a known phylogeny. *Mol. Biol. Evol.* **13**:525–536.
- SIDDALL, M. 1996. *Random cladistics*. Version 4.0. Virginia Institute of Marine Science, School of Marine Science (ftp.vims.edu), College of William and Mary, Gloucester Point, Va.
- SULLIVAN, J. 1996. Combining data with different distributions of among-site variation. *Syst. Biol.* **45**:375–380.
- SULLIVAN, J., K. E. HOLSINGER, and C. SIMON. 1995. Among-site variation and phylogenetic analysis of 12S rRNA in signomodontine rodents. *Mol. Biol. Evol.* **12**:988–1001.
- SWOFFORD, D. L. 1997. *Phylogenetic analysis using parsimony (PAUP\* 4.0\*)*. Sinauer, Sunderland, Mass. (in press).
- . 1991. When are phylogeny estimates from molecular and morphological data incongruent? Pp. 295–333 in M. M. MIYAMOTO and J. CRACRAFT, eds. *Phylogenetic analysis of DNA sequences*. Oxford University Press, New York.
- TEMPLETON, A. R. 1983. Phylogenetic inference from restriction endonuclease cleavage site maps with particular reference to the humans and apes. *Evolution* **37**:221–244.
- WILLIAMS, P. L., and W. M. FITCH. 1990. Phylogeny determination using dynamically weighted parsimony method. *Methods Enzymol.* **183**:615–626.
- YANG, Z. 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximate methods. *J. Mol. Evol.* **39**:306–314.

RODNEY L. HONEYCUTT, reviewing editor

Accepted March 24, 1997