When Does the Incongruence Length Difference Test Fail?

Pierre Darlu* and Guillaume Lecointre†

*INSERM, U535 Génétique épidémiologique et Structure des populations humaines, Bâtiment Gregory Pincus, 80 rue du Général Leclerc, 94276 Le Kremlin Bicêtre Cedex; and †Laboratoire d'Ichtyologie, Service de systématique moléculaire, IFR-CNRS 1541, Muséum National d'Histoire Naturelle, 43 rue Cuvier, 75231 Paris Cedex 05

This paper examines the efficiency of the incongruence length difference test (ILD) proposed by Farris et al. (1994) for assessing the incongruence between sets of characters. DNA sequences were simulated under various evolutionary conditions: (1) following symmetric or asymmetric trees, (2) with various mutation rates, (3) with constant or variable evolutionary rates along the branches, and (4) with different among-site substitution rates. We first compared two sets of sequences generated along the same tree and under the same evolutionary conditions. The probability of a Type-I error (wrongly rejecting the true hypothesis of congruence) was substantially below the standard 5% level of significance given by the ILD test; this finding indicates that the choice of the 5% level is rather conservative in this case. We then compared two data sets, still generated along the same tree, but under different evolutionary conditions (constant vs. variable evolutionary rate, homogeneity vs. heterogeneity rate of substitution). Under these conditions, the probability of rejecting the true hypothesis of congruence was greater than the 5% given by the ILD test and increased with the number of sites and the degree to which the tree was asymmetric. Finally, the comparison of the two data sets, simulated under contrasting tree structures (symmetric vs. asymmetric) but under the same evolutionary conditions, led us to reject the hypothesis of congruence, albeit weakly, particularly when the number of informative sites was low and among-site substitution rate heterogeneous. We conclude that the ILD test has only limited power to detect incongruence caused by differences in the evolutionary conditions or in the tree topology, except when numerous characters are present and the substitution rate is homogeneous from site to site.

Introduction

Farris et al. (1994) first proposed the incongruence length difference test (ILD) to quantify the conflicts that can occur between sets of characters from different data sources, such as nuclear or mitochondrial DNA sequences, protein sequences, RFLP or RAPD characters, isoenzymes, or even morphological traits. Each of these various data types provide phylogenetic information that can either converge toward the same phylogenetic tree or show discrepancies leading to conflicting conclusions. The conditional combination approach (Bull et al. 1993; Huelsenbeck, Bull, and Cunningham 1996; Baker, Yu, and DeSalle 1998) uses the ILD test as a preliminary step before choosing either to combine congruent data and thus increase the accuracy of the phylogenetic reconstruction or to analyze the data separately and attempt to discover the reasons for the incongruence.

Several statistical tests to measure character incongruence between partitions have already been proposed (Rodrigo et al. 1993; Farris et al. 1994; Huelsenbeck and Bull 1996) and their respective performances compared (see Huelsenbeck, Bull, and Cunningham 1996; Cunningham 1997a). To date, in the context of parsimony, ILD appears to be the most useful test, as Cunningham (1997a) pointed out, and it is widely used as a tool for studying various phylogenetic problems (Sullivan 1996; Lecointre et al. 1998; O'Grady, Clark, and Kidwell 1998; Vidal and Lecointre 1998; Allard, Farris, and Carpenter 1999; Denamur et al. 2000). Although

other methods have also been developed to test congruence in the maximum likelihood context (Waddell, Kishino, and Ota 2000), the parsimony approach remains the least inappropriate method for handling data that include different kinds of characters, which are difficult to integrate in a probabilistic model (such as morphology and DNA sequences).

The purpose of this work is to explore more deeply some of the evolutionary conditions that may influence the confidence we may have using the ILD test to reject or accept the hypothesis of congruence (Darlu and Lecointre 1999).

Methods and Simulations

The Incongruence Length Difference Test

The ILD test compares the numbers of steps of the most parsimonious trees built from the separate data partitions, the combined data, and random partitions of the data.

We denote as X_i, $i = 1 \ldots s$, the s different sets of data, X_i including n_i and as L_i the length of the most parsimonious tree obtained from the X_i data set. With all sets composed of the same taxa, we can express the observed value of the ILD test (ILD_o) as:

$$ILD_o = L - \sum L_i,$$

where L is the length of the most parsimonious tree obtained by combining the s data sets, and the sum is taken over all data sets (Mickevich and Farris 1981).

ILD_o is equal to 0 when there is at least one most parsimonious tree that is shared by each of the s data sets. Conversely, minimizing the number of homoplasies in some data sets produces more homoplasies in other sets, ILD_o is large. To test the null hypothesis of congruence, Farris et al. (1994) proposed that data sets be drawn at random from the combined data set, with each
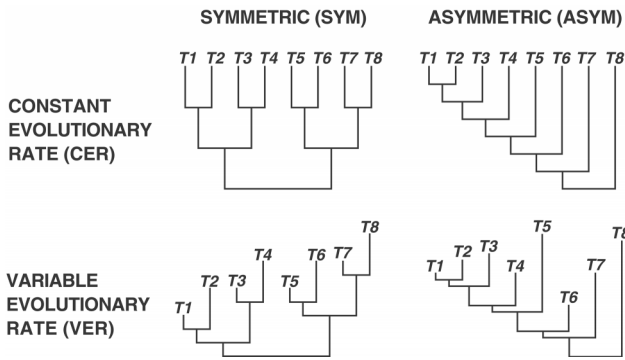
Key words: incongruence, ILD test, phylogeny.

FIG. 1.—Symmetric (SYM) and asymmetric (ASYM) simulated trees according to constant (CER) or variable (VER) evolutionary rate. Long and short branch lengths are in the ratio of three when the evolutionary rate is variable (VER).

random set having the same size as the original set. Thus, each represents a random mixture of characters extracted from the overall data. The ILD value thus obtained, ILD_r, is then calculated for a large number (e.g., $k = 1,000$) of such random partitions. We then count the number of times the observed ILD_o value is larger than the k random ILD_r values. Finally, the null hypothesis of congruence can be rejected at $P > 0.05$, that is, when ILD_o is larger than 95% of the random ILD_r.

The ILD test does not specifically compare the topologies of the trees obtained from different data sets, unlike, for instance, Robinson and Foulds's metric (Robinson and Foulds 1981; Makarenkov and Leclerc 1999), which gives the topological distance between trees. In the parsimony context, the ILD test is simply intended to evaluate whether the combined data produce a parsimonious tree with a length statistically comparable to the sum of the lengths of the most parsimonious trees obtained from the separate data. This is *Ho,* the null hypothesis of congruence. When the null hypothesis is accepted, the ILD test cannot help us to decide whether the trees obtained from the separate data are the correct

trees (i.e., in our case, the simulated trees). Moreover, the causes for a statistical rejection of *Ho,* the hypothesis of congruence, cannot be straightforwardly and easily ascertained. We can only conclude that the data sets compared do not share at least one identical parsimonious tree. Several reasons may be put forward: the tree structures may really be incongruent (e.g., horizontal transfers undergone within one data set), or the evolutionary conditions may be such that, although the two data sets stem from the same phylogeny, the parsimony method is inconsistent and leads to infer different topologies instead of reconstructing the correct one, or some combination of both the explanations. Our purpose here is to clarify some of these issues through simulations. Our scope does not, however, cover the question of whether combining the data is the most efficient method of reconstructing the correct tree, after the ILD test has not rejected the hypothesis of congruence. We focus chiefly on the power of the ILD test.

Simulations

To test the ILD test, we used the PAML generator (version 1.4) written by Z. Yang (1997) to simulate several sets of eight DNA sequences under conditions that varied as follows:

1. The structure of the tree was either symmetric (SYM) or asymmetric (ASYM) (fig. 1).
2. The evolutionary rate was either constant (CER), based on a molecular clock, or variable (VER), leading to two different branch lengths (alternating short and long branches in a length ratio of 3).
3. The length of the sequence, L, was either 100 or 1,000 sites.
4. The mutation rate was scaled at different values (s = 0.02, 0.10, 0.20, 0.40). The lowest (s = 0.02) and highest values represent, respectively, about 2 or 40 mutations along each branch, per 100 sites.
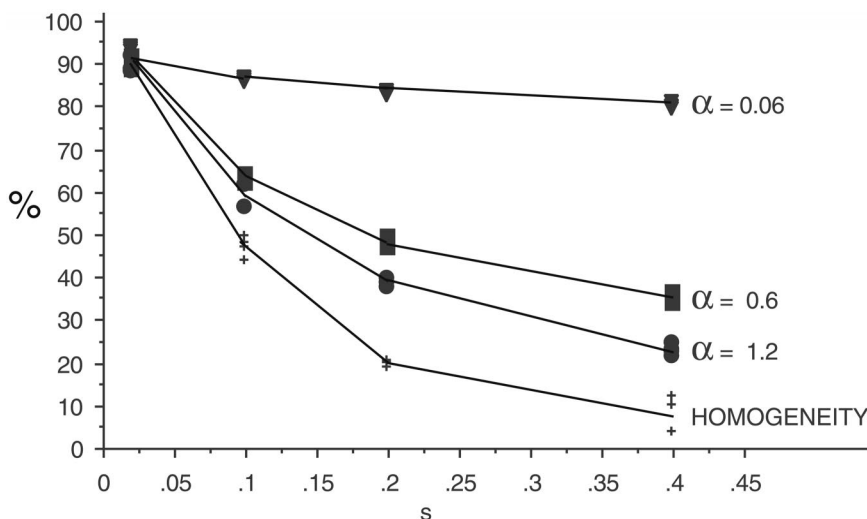

FIG. 2.—Relationships between the proportion (%) of uninformative sites, the mutation rate (s), and the α parameter estimating the among-site heterogeneity. The values obtained under CER and VER, with L = 100 and 1,000, are plotted although they are not significantly different.

Table 1
Results of the ILD Tests Comparing Data Simulated with Identical Tree and Parameters

| | HOM | | | | HET α = 0.06 | | | | HET α = 0.6 | | | |
| | SYM*SYM | | ASYM*ASYM | | SYM*SYM | | ASYM*ASYM | | SYM*SYM | | ASYM*ASYM | |
	L = 100	L = 1,000	L = 100	L = 1,000	L = 100	L = 1,000	L = 100	L = 1,000	L = 100	L = 1,000	L = 100	L = 1,000
CER... s = 0.02	0.6	0.0	1.6	0.4	1.2	1.4	1.0	4.0	2.2	0.0	1.6	3.0
s = 0.10	1.2	0.0	2.2	0.0	2.0	3.6	3.2	4.6	1.6	0.0	2.8	3.0
s = 0.20	2.4	0.0	3.2	1.4	3.2	2.6	3.2	3.8	1.8	0.0	1.8	3.8
s = 0.40	3.0	0.0	3.0	4.0	1.8	3.4	3.0	3.0	2.8	0.0	3.8	2.6
VER... s = 0.02	0.8	0.2	0.6	2.4	1.2	4.0	1.0	3.2	1.4	0.8	1.2	2.2
s = 0.10	1.2	0.0	2.2	3.0	3.2	4.6	2.8	3.8	2.0	0.2	2.8	3.4
s = 0.20	2.6	0.0	3.6	4.2	2.2	3.8	3.0	3.8	3.2	0.0	3.6	4.6
s = 0.40	3.2	4.4	2.2	4.6	3.4	3.0	3.2	3.2	3.6	3.2	3.2	3.2

NOTE.—Proportion (%) of the simulations ($n = 500$) leading to rejection of *Ho* (at $P = 0.05$), the true hypothesis of congruence between two data sets simulated with identical parameters and onto identical tree structures. Results are given according to the mutation rate s, the evolutionary rate (constant, CER, or variable, VER), the sequence length (L = 100 or 1,000 nucleotides), the among-site homogeneity (HOM) or heterogeneity (HET; α = 0.06 and 0.6) rate of substitution, and the tree structure (symmetric, SYM, or asymmetric, ASYM).

5. The heterogeneity of the substitution rate among sites was simulated by a gamma distribution of parameter α equal to 0.06, 0.6, and 1.2. The first value corresponds to the most heterogeneous values observed in the literature and the second to a median value (Yang 1996). The third value (α = 1.2) leads to a symmetric distribution of the among-site substitution rate, with moderate heterogeneity. Homogeneity was obtained by setting α to a high value: all sites then changed at the same rate.

Using both s and α parameters enabled us to generate data sets with various proportions of uninformative sites (invariant sites and sites showing autapomorphies): these proportions increase when α and s decrease, as shown in figure 2. We chose to simulate DNA sequences because it was an easy way to obtain several data sets with known evolutionary properties. For the moment, our purpose is simply to test the ILD method, now recognized as an appropriate method for evaluating congruence (Cunningham 1997*a*) with parsimony when the data sets are heterogeneous in nature.

The ILD test compared two sets of data, simulated either under the same tree structure or under contrasting tree structures, depending on the various evolutionary parameters. The test was performed with Farris's XRN program (Farris et al. 1994; Allard, Farris, and Carpenter 1999). The lengths of the parsimonious trees were obtained after branch swapping by testing five different randomly selected addition sequences. We performed 1,000 randomizations of data sets to determine the distribution of the null hypothesis and confirmed the results with the PAUP4b program (Swofford 1998).

Several comparisons are possible:

1. Comparing two sets of data generated along the same tree under the same evolutionary conditions makes it possible to evaluate the probability of wrongly rejecting the true hypothesis of congruence (type-I error).

2. Comparing two sets of data generated along the same tree but with contrasting evolutionary conditions (CER vs. VER, homogeneity vs. heterogeneity among-site substitution rate) enables us to evaluate the effect of these various conditions on type-II error (accepting *Ho,* the null hypothesis of congruence, when it is false because of incongruence caused by different evolutionary conditions for the two data sets along the same tree). When, however, our interest is the topological congruence between trees inferred from the data, and not the character congruence between the two data sets, these simulations allow us to determine whether the ILD test is robust enough to accept the null hypothesis of topological congruence in various evolutionary conditions.

3. Comparing two sets of data generated along two different trees (symmetric and asymmetric) with the same evolutionary conditions allows us to evaluate the probability of accepting the false hypothesis of congruence between the two data sets (type-II error) because the incongruence is caused in this case only by the topological differences. Moreover, we tested

Table 2
Results of the ILD Tests Comparing Data Simulated with Identical Tree and Different Rates of Evolution

		HOM CER*VER		HET $\alpha = 0.6$ CER*VER		HET $\alpha = 0.06$ CER*VER	
		L = 100	L = 1,000	L = 100	L = 1,000	L = 100	L = 1,000
SYM	s = 0.02	0.2	0.2	2.0	0.4	2.0	5.2
	s = 0.10	4.0	0.0	2.6	0.0	1.8	8.2
	s = 0.20	3.2	0.0	5.6	0.4	2.0	5.8
	s = 0.40	5.2	6.2	5.8	17.2	3.6	11.0
ASYM. . . .	s = 0.02	0.8	2.0	1.4	5.8	1.0	4.0
	s = 0.10	4.2	15.8	4.2	17.6	3.2	6.6
	s = 0.20	4.6	12.8	5.6	7.6	2.6	5.4
	s = 0.40	5.2	27.6	4.6	9.2	2.8	4.4

NOTE.—Proportion (%) of the simulations ($n = 500$) leading to rejection, at $P = 0.05$, of the hypothesis Ho of congruence between two data sets simulated with contrasted parameters (constant, CER, vs. variable, VER, evolutionary rate), the identical tree structure (SYM, symmetric, or ASYM, asymmetric). Results are given according to the mutation rate s, the sequence length (L = 100 or 1,000 nucleotides), and the among-site homogeneity (HOM) or heterogeneity (HET, $\alpha = 0.06$ and 0.6) rate of substitution.

the effect of withdrawing uninformative sites before performing the ILD test, as suggested by Cunningham (1997a, 1997b) and, more recently, by Lee (2001), who used a theoretical approach and molecular and morphological examples to demonstrate that a slight bias can occur when the proportions of informative characters differ too greatly between the data sets.

Results and Discussion

Table 1 gives the results of the ILD test that compared data simulated under the same evolutionary conditions and onto the same tree topology. The number of simulations that led to the erroneous conclusion that the data sets were incongruent (at $P < 0.05$) was always less than 5%. Thus, this significance level for the ILD test seems to be quite conservative, i.e., the risk of rejecting $Ho,$ the true hypothesis of congruence between the two data sets, is far less than the standard 5% level, particularly for the long sequences (1,000 sites), for which the ILD test never led to the wrong conclusion,

even when informative sites accounted for only about 10% of all the sites (see fig. 2).

Tables 2 and 3 summarize the ILD results for congruence between data simulated in identical tree structures but under different evolutionary conditions. They allow us to estimate the type-II error—accepting a false hypothesis of congruence. When we focus only on the topological congruence, these tables can also be interpreted as an evaluation of the type-I error of rejecting the true hypothesis of topological congruence when the evolutionary conditions in the two data sets vary. Table 2, for instance, shows that when the two tree structures were identical, either symmetric or asymmetric, with one tree simulated under a constant evolutionary rate and the other under a variable rate, the null hypothesis of congruence between data sets with L = 100 sites was rejected in no more than 5.8% of the simulations. This finding underlines the low power for rejecting Ho in these conditions, or, conversely, the robustness of the test, which does not reject the true hypothesis of topological congruence when the branch lengths vary between the two data sets. A large number of sites, specific

Table 3
Results of the ILD Tests Comparing Data Simulated with Identical Tree and Different Substitution Rates Among Sites

		HOM*HET ($\alpha = 0.6$)				HOM*HET ($\alpha = 0.06$)			
		SYM		ASYM		SYM		ASYM	
		L = 100	L = 1,000	L = 100	L = 1,000	L = 100	L = 1,000	L = 100	L = 1,000
CER.	s = 0.02	1.2 (1.4)	0.0 (0.0)	1.4 (1.2)	2.6 (2.6)	9.8 (8.6)	6.8 (6.2)	5.4 (5.8)	28.0 (27.2)
	s = 0.10	4.4 (3.8)	0.0 (0.0)	3.8 (4.2)	4.6 (4.2)	31.2 (23.4)	13.6 (14.0)	16.8 (17.2)	50.0 (26.8)
	s = 0.20	4.8 (3.8)	0.0 (0.0)	3.4 (3.0)	4.2 (3.2)	24.8 (19.4)	18.8 (19.2)	8.6 (10.8)	40.6 (16.6)
	s = 0.40	3.8 (3.6)	0.0 (0.0)	4.2 (4.0)	4.8 (3.8)	9.2 (7.2)	31.6 (7.6)	3.4 (5.6)	9.0 (3.4)
VER	s = 0.02	0.4 (0.6)	0.4 (0.4)	1.0 (0.4)	3.2 (3.2)	9.0 (7.4)	10.2 (10.2)	3.8 (3.6)	22.2 (22.2)
	s = 0.10	2.8 (3.0)	0.0 (0.0)	3.4 (4.2)	7.0 (6.4)	25.0 (20.8)	27.8 (21.0)	9.2 (12.8)	21.0 (13.4)
	s = 0.20	3.2 (3.2)	0.4 (0.4)	3.2 (2.8)	4.6 (3.0)	17.8 (14.8)	34.6 (17.2)	7.4 (12.2)	18.8 (11.4)
	s = 0.40	3.0 (2.0)	6.6 (0.4)	3.6 (4.0)	8.6 (8.6)	13.6 (11.0)	21.6 (8.4)	2.2 (5.6)	9.6 (8.0)

NOTE.—Proportion (%) of the simulations ($n = 500$) leading to rejection, at $P = 0.05$, of the hypothesis Ho of congruence between two data sets simulated with different among-site rate of substitution (HOM vs. HET, $\alpha = 0.6$ and 0.06), onto identical tree structures (SYM, symmetric, or ASYM, asymmetric). Results are given according to the mutation rate s, the sequence length (L = 100 or L = 1,000 nucleotides), and the constant (CER) or variable (VER) evolutionary rate. Numbers between parentheses are proportions obtained after removing uninformative sites.

Table 4
Results of the ILD Tests Comparing Data Simulated with Different Trees and Identical Parameters

		HOM SYM*ASYM		HET $\alpha = 1.2$ SYM*ASYM		HET $\alpha = 0.6$ SYM*ASYM		HET $\alpha = 0.06$ SYM*ASYM	
		L=100	L=1,000	L=100	L=1,000	L=100	L=1,000	L=100	L=1,000
CER	s = 0.02	20.4	100.0	20.0	100.0	16.4	100.0	3.8	86.1
	s = 0.10	83.6	100.0	65.0	100.0	50.0	100.0	8.0	83.2
	s = 0.20	75.4	100.0	59.8	100.0	46.8	100.0	6.2	73.6
	s = 0.40	53.0	100.0	48.6	100.0	36.8	100.0	6.6	70.0
VER	s = 0.02	35.2	100.0	35.0	100.0	27.4	100.0	5.4	93.6
	s = 0.10	94.4	100.0	80.2	100.0	68.4	100.0	10.6	93.6
	s = 0.20	90.6	100.0	77.0	100.0	59.4	100.0	8.4	90.0
	s = 0.40	74.6	100.0	67.2	100.0	57.2	100.0	8.2	83.2

NOTE.—Proportion (%) of the simulations ($n = 500$) leading to rejection of, at $P = 0.05$, the hypothesis Ho of congruence between two different data sets simulated onto contrasted tree structures (SYM, symmetric, and ASYM, asymmetric) with identical evolutionary parameters for both trees. Results are given according to the mutation rate s, the among-site rate of substitutions (HET or HOM, $\alpha = 0.06$, 0.6, and 1.2), the sequence length (L = 100 or L = 1,000 nucleotides), and the constant (CER) or variable (VER) evolutionary rate.

tree topology conditions (such as asymmetry), and a high mutation rate are needed to reject the null hypothesis of congruence between the two data sets. Even then, the power remains low because the highest proportion of simulations that rejected the false hypothesis of congruence was 27.6% (table 2).

Differences in the among-site substitution rate seemed to affect the ILD test more (table 3). Nonetheless, only when the contrast was large enough (HOM vs. HET, $\alpha = 0.06$) Ho was rejected more often than 5% of the time. For example, with symmetric trees, L = 100 and s = 0.2, we found that 24.8% of the simulations rejected the false hypothesis of congruence between data sets at $P < 0.05$. The ILD test can obviously reject, albeit weakly, the false hypothesis of congruence between data sets only when the among-site substitution rate varies. Sullivan (1996) illustrated this by showing that the null hypothesis of character congruence between two genes (*cyt b* and 12S in mice) was wrongly rejected because of their different among-site rate variation. Our results may, however, be explained by the substantial difference in the proportion of informative sites generated in the two data sets with HOM and HET options, with $\alpha = 0.06$ (see fig. 2). Therefore, we also performed ILD tests after removing both invariant sites and autapomorphies from the data sets. Table 3 summarizes these results, which indicate a slight bias, as observed by Lee (2001). The proportion of simulations rejecting the null hypothesis of congruence was usually slightly lower than when all sites were kept. This bias remains weak, however, and does not modify our conclusions.

Table 4 shows results of the ILD test comparing two data sets simulated onto different tree topologies, one symmetric and the other asymmetric, with identical evolutionary parameters. The only situation where the null hypothesis of congruence was clearly rejected, in all the simulations, involved a large number of sites (L = 1,000) and intermediate or low among-site substitution rates ($\alpha = 0.6$ and 1.2). In these situations, the test power is high. In other situations, a far lower proportion of cases rejected the false null hypothesis. For example,

when $\alpha = 0.6$, L = 100, CER, and s = 0.1, only half the simulations rejected the false hypothesis of congruence. Even when the substitution rate was homogeneous among sites, the proportion of simulations rejecting the false hypothesis of congruence was low—between 20.4% and 94.4% when L = 100. When the number of sites was low (L = 100) and the among-site substitution rate heterogeneous ($\alpha = 0.06$), less than 10% of the simulations rejected the null hypothesis of congruence, even though the tree structures were strongly divergent. We conclude that the power of the ILD test to detect incongruence between data sets generated under different topologies is highly sensitive to the number of sites investigated, to the number of informative sites, and to the among-site substitution rate heterogeneity, or to all of these (as fig. 2 shows), even when these parameters are identical in the data sets being compared.

Finally, we conclude that the ILD test is quite conservative (table 1), at least when the hypothesis of congruence is correct, i.e., when both topology and evolutionary parameters are congruent between the compared data sets. Moreover, the incongruence caused by unequal branch lengths does not appear to be detected easily by the ILD test, a finding that suggests that its efficiency for them, at least in the situations we investigated. Lastly, its power to detect incongruence is extremely low when the incongruence is caused by different topologies, when the number of informative sites is small, and the heterogeneity of among-site substitution rate is large.

Acknowledgments

LITERATURE CITED

ALLARD, M. W., J. S. FARRIS, and J. M. CARPENTER. 1999. Congruence among mammalian mitochondrial genes. Cladistics **15**:75–84.

BAKER, R. H., X. YU, and R. DESALLE. 1998. Assessing the relative contribution of molecular and morphological characters in simultaneous analysis trees. Mol. Phylogenet. Evol. **9**(3):427–436.

BULL, J. J., J. P. HUELSENBECK, C. W. CUNNINGHAM, D. L. SWOFFORD, and P. J. WADDELL. 1993. Partitioning and combining data in phylogenetic analysis. Syst. Biol. **42**(3):384–397.

CUNNINGHAM, C. W. 1997a. Can three incongruence tests predict when data should be combined? Mol. Biol. Evol. **14**(7): 733–740.

———. 1997b. Is congruence between data partitions a reliable predictor of phylogenetic accuracy? Empirically testing an iterative procedure for choosing among phylogenetic methods. Syst. Biol. **46**:464–478.

DARLU, P., and G. LECOINTRE. 1999. How powerful and robust are the incongruence length differences tests? Ann. Hum. Genet. **63**:356–357.

DENAMUR, E., G. LECOINTRE, P. DARLU et al. (12 co-authors). 2000. Evolutionary implications of the frequent horizontal transfer of mismatch repair genes. Cell **103**:711–721.

FARRIS, J. S., M. KÄLLERSJÖ, A. G. KLUGE, and C. BULT. 1994. Testing significance of congruence. Cladistics **10**:315–319.

HUELSENBECK, J. P., and J. J. BULL. 1996. A likelihood ratio test to detect conflicting phylogenetic signal. Syst. Biol. **45**(1):92–98.

HUELSENBECK, J. P., J. J. BULL, and C. W. CUNNINGHAM. 1996. Combining data in phylogenetic analysis. Trends Ecol. Evol. **11**(4):152–157.

LECOINTRE, G., L. RACHDI, P. DARLU, and E. DENAMUR. 1998. Escherichia coli molecular phylogeny using the incongruence length difference test. Mol. Biol. Evol. **15**(12):1685–1695.

LEE, M. S. Y. 2001. Uninformative characters and apparent conflict between molecules and morphology. Mol. Biol. Evol. **18**(4):676–680.

MAKARENKOV, V., and B. LECLERC. 1999. The fitting of a tree to a given dissimilarity with the weighted least squares criterion. J. Classif. **16**:3–26.

MICKEVICH, M. F., and J. S. FARRIS. 1981. The implications of congruence in *Menidia*. Syst. Zool. **30**:351–370.

O'GRADY, P. M., J. B. CLARK, and M. G. KIDWELL. 1998. Phylogeny of the Drosophila saltans species group based on combined analysis of nuclear mitochondrial DNA sequences. Mol. Biol. Evol. **15**(6):656–664.

ROBINSON, D. R., and L. R. FOULDS. 1981. Comparison of phylogenetic trees. Math. Biosci. **53**:131–147.

RODRIGO, A. G., M. KELLY-BORGES, P. R. BERGQUIST, and P. L. BERGQUIST. 1993. A randomization test of the null hypothesis that two cladograms are sample estimates of a parametric phylogenetic tree. N Z J. Bot. **31**:257–258.

SULLIVAN, J. 1996. Combining data with different distributions of among-site rate variation. Syst. Biol. **45**(3):375–380.

SWOFFORD, D. L. 1998. PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4. Sinauer Associates, Sunderland, Mass.

VIDAL, N., and G. LECOINTRE. 1998. Weighting and congruence: a case study based on three mitochondrial genes in pitvipers. Mol. Phylogenet. Evol. **9**(3):366–374.

WADDELL, P. J., H. KISHINO, and R. OTA. 2000. Rapid evaluation of the phylogenetic congruence of sequence data using likelihood ratio tests. Mol. Biol. Evol. **17**(12):1988–1992.

YANG, Z. 1996. Among-site rate variation and its impact on phylogenetic analyses. Trends Ecol. Evol. **11**(9):367–372.

———. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. CABIOS **13**:555–556.