



## Further problems with the incongruence length difference test: “hypercongruence” effect and multiple comparisons

Martín J. Ramírez\*

*Museo Argentino de Ciencias Naturales—CONICET, Avenida, Angel Gallardo 470 C1405DJR Buenos Aires, Argentina*

Accepted 18 January 2006

---

### Abstract

The incongruence length difference (ILD) test may produce artificially large significance values with the addition not only of uninformative characters, but also of informative characters not relevant to the groups in conflict. Previously reported problems with the ILD test involved cases of false positives, reporting high incongruence when none is expected. Under certain conditions, the test can suffer with the opposite problem (false negatives), reporting non-significant values in cases of high incongruence. These opposing effects can be combined in a data set, such that a comparison over all partitions appears as congruent, while some of the pair-wise comparisons are reported as significantly incongruent.

© The Willi Hennig Society 2006.

---

The idea of the incongruence length difference (ILD) test is detecting significant incongruence in the phylogenetic trees derived from two or more data partitions (Farris et al., 1995a,b). The ILD test has become a standard procedure in phylogenetic analyses involving more than one gene sequence fragment, or diverse sources of data, as are morphology and nucleotides. Because the core of all phylogenetic methods is vertical transfer, which predicts congruent trees for all data sources, the ILD test rapidly gained a central role in systematics, as a means to corroborate the hypothesis of vertical transfer or alternatively as an indicator of other transfer processes.

Phylogenetic incongruence can be interpreted as indicative that the partitions have different evolutionary history. Following this interpretation, one of the more frequent uses of the test is informing on the decision of combining data from different sources in a simultaneous analysis (Bull et al., 1993; see discussion in Barker and Lutzoni, 2002). The test is also used to detect cases of horizontal transfer (Lecointre et al., 1998; Ballard, 2000; Thornton and DeSalle, 2000; Kroken and Taylor, 2001; Sota and Vogler, 2001; Sota, 2002; Brown et al., 2002,

2003; Escobar-Páramo et al., 2004), and sometimes to evaluate coevolution of associates (Lopez-Vaamonde et al., 2001; Johnson et al., 2001; Downie and Gullan, 2005). Some results are, however, not well explained by the different-history interpretation, as are the cases of incongruence between linked mitochondrial (e.g., Sullivan, 1996; Baker et al., 2001; Vogler et al., 2005) and chloroplast (Yoon et al., 2002) genes, or among first, second and third codon positions of the same gene (Sanderson et al., 2000; Vogler et al., 2005).

An alternative explanation of incongruence is that the partitions have indeed a common history, but certain phylogenetic methods (via their associate models and parameters) perform better than others in finding the common tree. Under this interpretation, the ILD test was used to select among transition/transversion costs (e.g., Allard and Carpenter, 1996; Ah Yong and O’Meally, 2004), or determine the quality of a given source of data (e.g., Quicke and Belshaw, 1999).

There is growing evidence that the ILD test produces unexpected results under certain simple circumstances (see below). So far all known problems with the ILD test involve cases of false positives (Type I error): it reports significant incongruence where the expectation is of congruence, or of a level of incongruence attributable to

---

\*Corresponding author.

E-mail address: ramirez@macn.gov.ar

random variation. I explore here some conditions that can produce the opposite effect, of false negatives (Type II error), reporting non-significant values in cases of high incongruence. A peculiarity of the discussion of the ILD test in the literature is that although the derivation of the test (Farris et al., 1995a,b) is an example of elegance and simplicity, the examples with unexpected results accumulate and there are no explanations as to what is going wrong. As a potential explanation, I suggest here that the problem is the randomization process, which is not able to follow the phylogenetic data when these are highly structured.

## Background

The ILD test was proposed by Farris et al. (1995a,b; also known as “partition homogeneity test”, Swofford et al., 1996) for the detection of a conflicting hierarchical structure between partitions; it is composed of two parts, a test statistic and a null distribution generated by a randomization model. The test statistic is  $D$ , the additional homoplasy that arises when combining different partitions in a simultaneous analysis. For two partitions  $A$  and  $B$ ,

$$D = L_{AB} - (L_A + L_B)$$

Where  $L_A$  and  $L_B$  is the minimum length of partition  $A$  and  $B$ , respectively, and  $L_{AB}$  is the minimum length of the combined data set. Similarly, for multiple partitions,

$$D = L_{ABC\dots N} - (L_A + L_B + L_C + \dots + L_N)$$

The greater the incongruence between partitions, the larger the homoplasy that arises when the partitions are combined to accommodate to a unique tree, which is reflected in a larger value of  $D$ . For simplicity, I will refer most of the time to the two-partitions case. The observed value of the statistic  $D$  ( $D_{\text{obs}}$ ) is compared with a null distribution of randomized values of  $D$  ( $D_{\text{rand}}$ ) generated by randomly distributing the characters in two data sets  $P$  and  $Q$  of the same sizes as  $A$  and  $B$ , that is, comparing  $L_{AB} - (L_A + L_B)$  versus  $L_{PQ} - (L_P + L_Q)$ . Because  $L_{AB}$  and  $L_{PQ}$  are the same entire data set, this is equivalent as comparing  $(L_A + L_B)$  versus  $(L_P + L_Q)$ . The null hypothesis is that  $(L_A + L_B) \geq (L_P + L_Q)$ , which is identified with the partitions being congruent. The alternative hypothesis  $(L_A + L_B) < (L_P + L_Q)$ , is identified with the partitions being incongruent.

The tree lengths can be expressed as minimum steps plus homoplasy ( $H$ ), and because the minimum number of steps for a character under fixed homology is constant, the comparison is equivalent to  $(H_A + H_B)$  versus  $(H_P + H_Q)$ . The ILD test was analogized with the Mann–Whitney  $U$ -test for difference in location of an ordinal variable (Farris et al., 1995a; Dolphin et al.,

2000). The Mann–Whitney and other simple tests require independence of the observations, that is, when observation 1 is sampled, it does not tell anything about observation 2. Even if characters are independent (e.g., Felsenstein, 2004, p. 338), the requisite of independence is not on the characters itself (i.e., the distribution of character states), but on the variable that is being measured, in this case number of homoplasious steps. Because the steps are realized over a common tree, the homoplasy score for one character depends on the conflict with other characters. Homoplasy scores are not attributes of characters, but interactions among multiple characters. This situation should not necessarily be a problem for the ILD test, because randomization tests are used for cases where independence is violated, as is the case here. The ILD test has been under heavy scrutiny during the last years, and I summarize below the main problems reported so far.

### *Random data*

Graham et al. (1998) reported high levels of significance of the ILD test when comparing real and random data, which is unexpected as random data do not have a structured hierarchy. Dolphin et al. (2000) explored this effect in more detail, and compared the effect of shuffling cells of characters to produce noisy data. They compared one partition fully resolved and without homoplasy, with a copy of the same partition, where some of the characters have the cells randomly shuffled. The more shuffled characters, the more significant the ILD test. The expectation in this case is that randomly shuffled characters should not produce significant incongruence values.

### *Uninformative characters*

Lee (2001) found that moderately incongruent partitions are highly significant for the ILD test when many uninformative characters are added to one of the partitions. As uninformative characters do not contain hierarchical structure, the expectation is that incongruence should be the same, or at most “diluted” by the uninformative characters. Lee found this effect in real data sets, for example in comparisons of morphology versus DNA sequences, where the sequence data has many uninformative sites. His simple hypothetical example will be discussed below in more detail. Cunningham (1997) already pointed out that it was important to delete invariant characters before running the test, without providing further explanation.

### *Global versus localized incongruence*

The test statistic  $D$  is a summary measure for the entire tree. Several authors raised the question whether a

global measure of incongruence can reflect cases appropriately when only a small portion of the tree is involved in the conflict (Baker et al., 2001; Damgaard and Cognato, 2003; see References in Lambkin, 2004). This distinction is most relevant in studies deciding on the combinability of data partitions, because a small area of conflict may lead to a partitioned analysis with the consequent loss of support throughout the entire tree (Gatesy et al., 1999). Localizing areas of conflict is, however, not trivial. Baker et al. (2001) used the variation of ILD scores on constrained trees, and Yoder et al. (2001) and Hipp et al. (2004) observed the variation of *P*-values from the ILD test after the removal of terminals. These strategies are only feasible for small data sets, or when there are previous hypotheses about specific taxa as potential sources of incongruence; otherwise the number of clades that can be constrained or excluded grows exponentially with the number of terminals. The partitioned Bremer support and its derivatives are often used to detect local incongruence (e.g., Baker and DeSalle, 1997; Baker et al., 2001; Damgaard and Cognato, 2003; Lambkin, 2004). Lavoué et al. (2003) reported conflicting results between the ILD test and partitioned Bremer supports.

*Unequal transformation costs*

Dowton and Austin (2002) explored the behavior of the ILD test to changes in costs of transitions, transversions, and different costs matrices for each partition. Their purpose was to explore the ILD measure (scaled as WILD;  $ILD_w$  in Aagesen et al., 2005) and the ILD test as criteria of meta-optimality. That is, selecting the set of costs that maximize congruence between partitions, to produce the final trees (see also Wheeler, 1995; Aagesen et al., 2005). Their results on simulations show that both the WILD and the ILD test often select two extremes of cost relations as the less incongruent. Hipp et al. (2004) found a similar effect in a reanalysis of Yoder et al.’s (2001) real data. What is most disturbing is that the cost

matrices that would be optimal according to the ILD test are different from those used to generate the simulations.

Aagesen et al. (2005) explored the effect of extreme cost ratios on incongruence measures. Their aim was selecting the set of transformation costs that minimized some measure of incongruence, without any significance test involved. They explored the WILD, a rescaled ILD (RILD), and the meta-retention index for fragments within partitions (MRI). They found that all selected indices may slide towards trivial optima when the cost ratios are extreme; the circumstances under which the trivial optimum was reached differed for each index. Nevertheless, in several cases all indices did converge on the same cost set as being optimal, in agreement with a topological index. It would be interesting to test these new indices on simulated data to see whether the cost set used to generate the data will be selected as the optimal one. The above-mentioned problems all arise when comparing congruence values based on different transformation costs. These findings are interesting because they report problems in the measures of incongruence, independent of any randomization test. I will, therefore, concentrate on the randomization test only, and to avoid these problems only use examples based on uniform and equal costs.

**Materials and methods**

ILD significance values were calculated with a custom script in TNT 1.0 (Goloboff et al., 2003–2005), with 10 000 replicates, using as a default three random addition sequences plus TBR swapping. The script and data sets are available at <http://www.cladistics.org/journal/data>.

*Not only uninformative characters*

Figure 1 is a modification of the hypothetical example case presented by Lee (2001). Partitions A

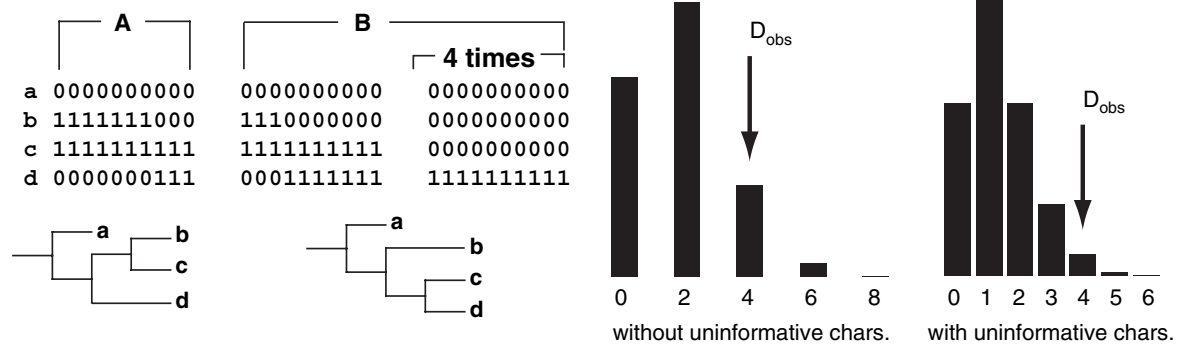


Fig. 1. A moderately incongruent data set with two partitions with 10 characters each ( $P = 0.177$ ) appears highly incongruent when 40 uninformative characters are added to partition B ( $P = 0.034$ ) (modified after Lee, 2001). Histograms of randomized values of  $D$ ,  $D_{obs} = D$  observed.

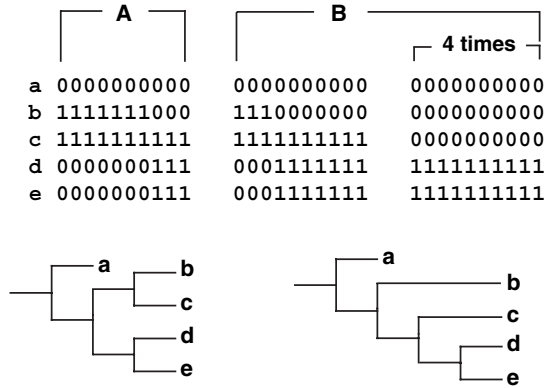


Fig. 2. The same effect as in Fig. 1 is now obtained with the addition of informative characters ( $P = 0.034$ ). In this case taxon D is a copy of taxon C, similarly as occurs when two closely related species are together in the same data set.

and B are moderately incongruent ( $P = 0.177$ ), but when a series of uninformative characters are added to partition B, the test reports high significance ( $P = 0.034$ ). This effect occurs because the vast majority of the randomizations will assign most of the informative characters to the larger partition Q, while the smaller partition P will contain mostly uninformative characters. The incongruence that originally occurred between partitions is now combined within partition Q, thus expressing most of the potential homoplasy; the most frequent value of  $H_P + H_Q$  is very close to  $H_{AB}$  (Fig. 1).

Lee proposed the elimination of uninformative characters as a way of dealing with this problem. However, the same effect may occur with informative characters, as long as they are not relevant to the groups in conflict, as occurs when the last terminal in the example of Fig. 1 is duplicated (Fig. 2). In practice, many uninformative characters can become informative just by adding a species closely related to one already represented in the data set.

The “hypercongruence” effect

The ILD test was designed as a one-tailed test, presumably because at that time the only known or interesting departures from random distributions were negative (i.e.,  $H_P + H_Q > H_A + H_B$ ), and caused by incongruence. It is, however, possible to obtain a similar effect as shown above, but in the opposite direction, that is, a partition of data such that all or most ( $H_P + H_Q$ ) will be smaller than ( $H_A + H_B$ ). This is produced simply by having some homoplasy in one of the partitions, and characters that are irrelevant to the groups in conflict in the other (Fig. 3). Because all the potential homoplasy is already realized in the first partition, the distribution of ( $H_P + H_Q$ ) has an upper value in ( $H_A + H_B$ ), hence in all randomizations ( $H_A + H_B$ )  $\geq$  ( $H_P + H_Q$ ), and all  $D_{rand}$  are greater or equal to  $D_{obs}$ ; the result is as non-significant as it can be ( $P = 1$ ). Of course the partitions are not especially congruent—just indifferent. In this example any random partitioning in P and Q will be more incongruent than the originals A and B, thus producing an illusion of “hypercongruence”. This situation is not especially interesting in terms of incongruence, but real DNA sequence data may, nevertheless, be of such a structure because phylogenetic information (and hence homoplasy) is often concentrated in informative regions alternated with uninformative ones.

In the pure examples of Figs 1–3, the cases where additional homoplasy arises when partitions are combined, as in incongruence, have a  $D_{obs}$  on the right tail of the distribution of  $D_{rand}$ . Those cases where the random partitions are more incongruent than the originals have a  $D_{obs}$  in the opposed tail. If the second situation is not of special interest, it may however, determine the outcome of the test. By combining incongruent and “hypercongruent” data set sectors, an incongruent data set may appear as congruent to the ILD test (Fig. 4), producing a false negative. As in the

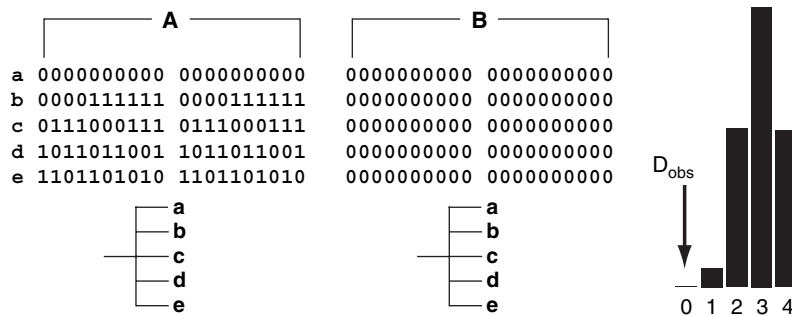


Fig. 3. An “hypercongruent” data set ( $P = 1$ ). Partition A has some homoplasy, and partition B is uninformative. Any random partition has less or at most the same homoplasy. Histogram of randomized values of D,  $D_{obs} = D$  observed.

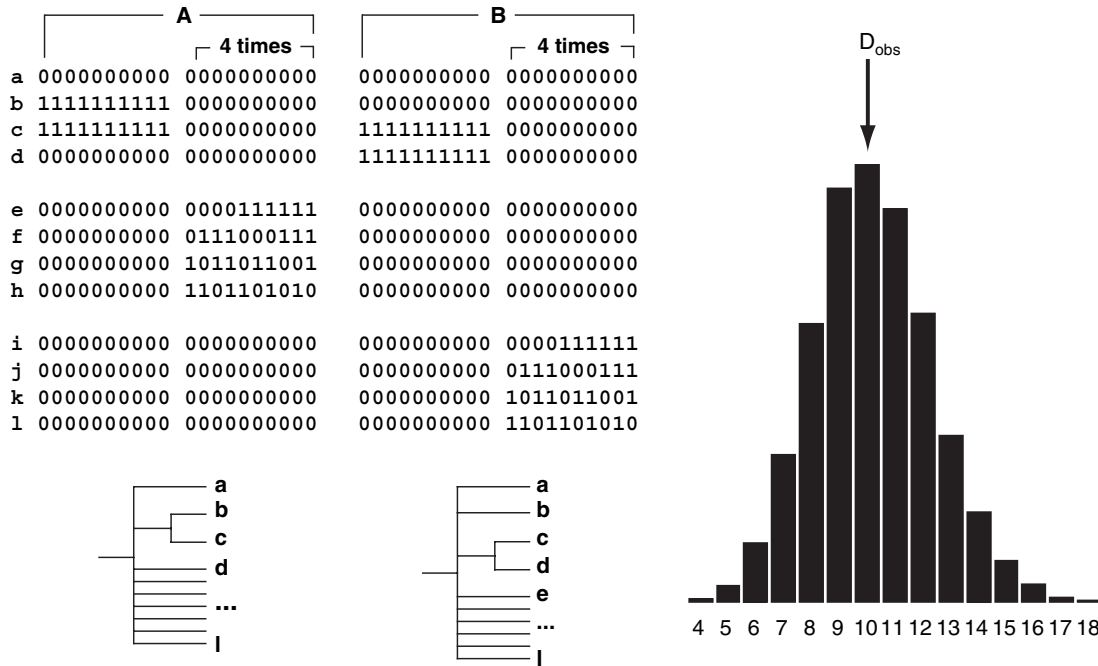


Fig. 4. Adding a “hypercongruent” sector (for taxa e–k) to incongruent partitions produces a net appearance of congruence ( $P = 0.601$ ). Histogram of randomized values of D,  $D_{obs} = D$  observed.

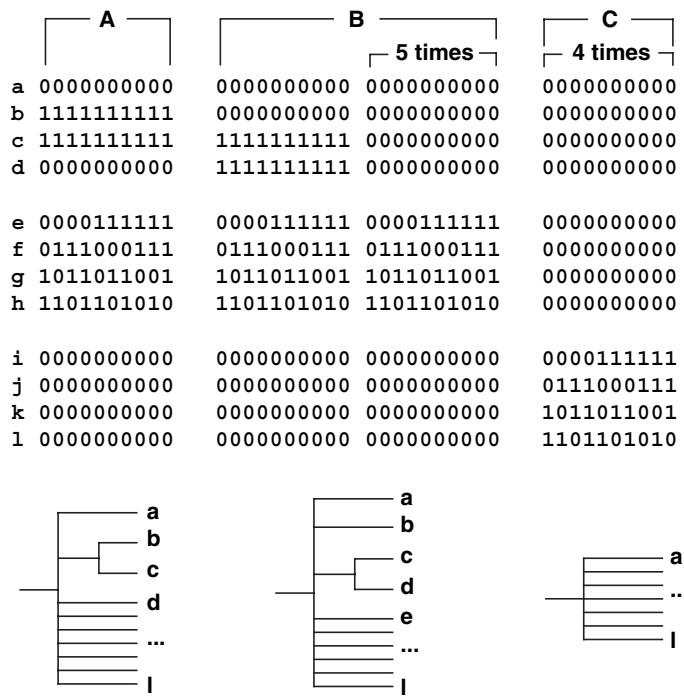


Fig. 5. Three partitions A–C, appear as congruent in the simultaneous comparison ( $P_{ABC} = 0.856$ ), but partitions A and B are highly incongruent ( $P_{AB} < 0.0001$ ). This occurs because of a “hypercongruent” sector involving the comparisons of C with A or B ( $P_{AC} = 1$ ,  $P_{BC} = 1$ ).

previous examples, the variation in  $P$ -values is obtained just by adding characters and taxa that are irrelevant to the groups in conflict.

When these effects occur in more than two partitions, the ILD test may produce seemingly self-contradictory results. The data set of Fig. 5, with three partitions A, B

and C, appears congruent for the simultaneous ILD test ( $P = 0.856$ ), while the comparison of A and B appears as highly incongruent ( $P < 0.0001$ ). Both comparisons A with C, and B with C are “hypercongruent” ( $P = 1$ ).

This happens because the null distribution is built from net deviation of homoplasy scores, adding over positive and negative deviations. Negative scores from incongruent partitions can be cancelled by positive scores from unequally concentrated (“hypercongruent”) partitions. It is possible to build a distribution by computing only the absolute (unsigned) deviation of randomized scores from observed scores, as in a regular analysis of variance, but then partitions with unequal density of information may produce significant results, even in the absence of incongruence.

### *The null distribution*

A closer scrutiny of the generation of the null distribution for the ILD may help understand why the test, under certain circumstances, produces significance values not related with incongruence levels. In a data set there are multiple possible instances of homoplasious steps, but not every character is relevant to the conflicts producing each instance. In cases of structured data, the experimental design and the randomizations should capture this structure, so that the randomized data sets are plausible versions of the original data set, reflecting the null hypothesis. In the simple case of Fig. 1, most randomized data sets will contain a small partition with just a few informative characters, which is definitely not a plausible version of the original partition. As an exercise, we can try to develop a structured, custom randomization model to produce plausible versions of the original partitions. Take for example the steps that arise in the conflict of (b (c d)) versus (d (b c)) in Fig. 4. For those steps, it seems better to limit the randomization within the first 10 characters of each partition, because all other characters are irrelevant to choose between these groupings, hence they are not interchangeable candidates. This is analogous to skipping oranges and dust mottles when testing the distribution of small and large apples in two bags, and is in line with the solution proposed by Lee (2001) of leaving uninformative characters out of the randomization. If there is a second area of conflict in the tree (e.g., terminals e–h in Fig. 4), we might isolate the characters relevant to that area (the last 40 characters of partition A), and randomize them separately, as if they were small and large oranges (in this case, all random partitions are the same as in the original). The same idea can be applied to the remaining area of conflict (terminals i–l). In more general cases, however, each character is relevant to many areas of the tree at the same time, as occurs with the entire partition A in Fig. 5, and a custom randomization cannot be implemented in this simple way. In

real data sets, with multiple, nested areas of conflict, and characters relevant to many of them at the same time, structuring approximate custom randomizations of this kind seems unfeasible. What is worst, Goloboff et al. (2003, Fig. 2) have shown that it is often not possible to classify characters as supporting, irrelevant or opposing a given group. In summary, the equiprobable model of character randomization may produce a majority of non-plausible versions of the original partitions. Phylogenetic data can be intricately structured, and maintaining homogeneous partition sizes is not a sufficient control to generate plausible randomized versions. Skipping uninformative characters is a step towards reflecting some of the structure of the data in the randomization process. In principle, if characters could be segregated as relevant or irrelevant for given groupings, this idea might be extended to produce more structured randomizations, but this is not the case in real data.

### **Conclusions**

The results of the ILD test are nowadays interpreted with caution, as common evolutionary conditions like heterogeneous rates of change between partitions may cause the test to report higher significance of incongruence than expected. The results presented here suggest that heterogeneous rates of change between clades may also affect the outcome of the test: while computing the null distribution of the ILD, some tree sectors with high within-partition homoplasy may produce an “hypercongruent” effect that cancels the contribution of sectors with high incongruence between partitions. In this situation, highly incongruent sectors may pass undetected by the test.

Congruence is perhaps the most general prediction of phylogenetics, hence a powerful test of incongruence would be a tool for the evaluation of reconstruction methods, quality of data, transformation parameters, and exceptions to vertical transfer. It seems, however, that there is a long list of issues to be fixed in the ILD test before it can be used in such ambitious tasks.

### **Acknowledgments**

Lone Aagesen, Norberto Giannini, Camilo Mattoni, Cecilia Kopuchian, Julián Faivovich, and an anonymous reviewer provided helpful criticism and ideas. This work was initiated from discussions at the Seminars of Theory and Methods of Phylogenetic Reconstruction at the Museo Argentino de Ciencias Naturales, Buenos Aires. Support was provided by PEI 6558 from CONICET, PICT03/14092 from ANPCyT, and ATOL NSF

Grant “Assembling the Tree of Life: Spiders” 03-011-1612-216.

## References

- Aagesen, L., Petersen, G., Seberg, O., 2005. Sequence length variation, indel costs, and congruence in sensitivity analysis. *Cladistics*, 21, 15–30.
- Ahyong, S.T., O’Meally, D., 2004. Phylogeny of the Decapoda Reptantia: resolution using three molecular loci and morphology. *Raffles Bull. Zool.* 52, 673–693.
- Allard, M., Carpenter, J., 1996. On weighting and congruence. *Cladistics* 12, 183–198.
- Baker, R., DeSalle, R., 1997. Multiple sources of character information and the phylogeny of Hawaiian drosophilids. *Syst. Biol.* 46, 654–673.
- Baker, R.H., Wilkinson, G.E.S., DeSalle, R., 2001. Phylogenetic utility of different types at molecular data used infer evolutionary relationships among stalk-eyed flies (Diptera: Diopsidae). *Syst. Biol.* 50, 87–105.
- Ballard, J.W.O., 2000. When one is not enough: introgression of mitochondrial DNA in *Drosophila*. *Mol. Biol. Evol.* 17, 1126–1130.
- Barker, F.K., Lutzoni, F.M., 2002. The utility of the incongruence length difference test. *Syst. Biol.* 51, 625–637.
- Brown, E.W., Kotewicz, M.L., Cebula, T.A., 2002. Detection of recombination among *Salmonella enterica* strains using the incongruence length difference test. *Mol. Phylogenet. Evol.* 24, 102–120.
- Brown, M.K., LeClerc, M.J.E., Cebula, T.A., 2003. Limited boundaries for extensive horizontal gene transfer among *Salmonella* pathogens. *Proc. Natl Acad. Sci. USA*, 100, 15676–15681.
- Bull, J., Huelsenbeck, J., Cunningham, C., Swofford, D., Waddell, P., 1993. Partitioning and combining data in phylogenetic analysis. *Syst. Biol.* 42, 384–397.
- Cunningham, C., 1997. Can three incongruence tests predict when data should be combined? *Mol. Biol. Evol.* 14, 733–740.
- Damgaard, J., Cognato, A.I., 2003. Sources of character conflict in a clade of water striders (Heteroptera: Gerridae). *Cladistics* 19, 512–526.
- Dolphin, K., Belshaw, R., Orme, C.D.L., Quicke, D.L.J., 2000. Noise and incongruence: interpreting results of the incongruence length difference test. *Mol. Phyl. Evol.* 17, 401–406.
- Downie, D.A., Gullan, P.J., 2005. Phylogenetic congruence of mealybugs and their primary endosymbionts. *J. Evol. Biol.* 18, 315–324.
- Dowton, M., Austin, A.D., 2002. Increased congruence does not necessarily indicate increased phylogenetic accuracy – the behavior of the incongruence length difference test in mixed-model analyses. *Syst. Biol.* 51, 19–31.
- Escobar-Páramo, P., Sabbagh, A., Darlu, P., Pradillon, O., Vauray, C., Denamur, E., Lecointre, G., 2004. Decreasing the effects of horizontal gene transfer on bacterial phylogeny: the *Escherichia coli* case study. *Mol. Phyl. Evol.* 30, 243–250.
- Farris, J.S., Källersjö, M., Kluge, A.G., Bult, C., 1995a. Testing significance of incongruence. *Cladistics* 10, 315–319.
- Farris, J.S., Källersjö, M., Kluge, A.G., Bult, C., 1995b. Constructing a significance test for incongruence. *Syst. Biol.* 44, 570–572.
- Felsenstein, J., 2004. *Inferring Phylogenies*. Sinauer Associates, Sunderland, MA.
- Gatesy, J., O’Grady, P., Baker, R.H., 1999. Corroboration among data sets in simultaneous analysis: hidden support for phylogenetic relationships among higher level artiodactyl taxa. *Cladistics*, 15, 271–313.
- Goloboff, P.A., Farris, J.S., Källersjö, M., Oxelman, B., Ramírez, M.J., Szumik, C.A., 2003. Improvements to resampling measures of group support. *Cladistics* 19, 324–332.
- Goloboff, P.A., Farris, J.S., Nixon, K., 2003–2005. TNT: Tree Analysis Using New Technology Version 1.0. Program and documentation, available from the authors, and at <http://www.zmuc.dk/public/phylogeny/>.
- Graham, S.W., Kohn, J.R., Morton, B.R., Eckenwalder, E., Barrett, C.H., 1998. Phylogenetic congruence and discordance among one morphological and three molecular data sets from Pontederiaceae. *Syst. Biol.* 47, 545–567.
- Hipp, A.L., Hall, J.C., Sytsma, K.J., 2004. Congruence versus phylogenetic accuracy: revisiting the incongruence length difference test. *Syst. Biol.* 53, 81–89.
- Johnson, K.P., Drown, M.D., Clayton, D.H., 2001. A data based parsimony method of cophylogenetic analysis. *Zool. Scripta* 30, 79–87.
- Kroken, S., Taylor, J.W., 2001. Outcrossing and recombination in the lichenized fungus *Letharia*. *Fungal Genet. Ecol.* 34, 83–92.
- Lambkin, C.L., 2004. Partitioned Bremer support localizes significant conflict in bee flies (Diptera: Bombyliidae: Anthracinae). *Invert. Syst.* 18, 351–360.
- Lavoué, S., Sullivan, J.P., Hopkins, C.D., 2003. Phylogenetic utility of the first two introns of the S7 ribosomal protein gene in African electric fishes (Mormyroidea: Teleostei) and congruence with other molecular markers. *Biol. J. Linnean Soc.* 78, 273–292.
- Lecointre, G., Rachdi, L., Darlu, P., Denamur, E., 1998. *Escherichia coli* molecular phylogeny using the incongruence length difference test. *Mol. Biol. Evol.* 15, 1685–1695.
- Lee, M.S.Y., 2001. Uninformative characters and apparent conflict between molecules and morphology. *Mol. Biol. Evol.* 18, 676–680.
- Lopez-Vaamonde, C., Rasplus, J.Y., Weiblen, G.W., Cook, J.M., 2001. Molecular phylogenies of fig wasps: partial coeladogenesis of pollinators and parasites. *Mol. Phyl. Evol.* 21, 55–71.
- Quicke, D.L., Belshaw, R., 1999. Incongruence between morphological data sets: an example from the evolution of endoparasitism among parasitic wasps (Hymenoptera: Braconidae). *Syst. Biol.* 48, 436–454.
- Sanderson, M.J., Wojciechowski, M.F., Hu, J.-M., Sher Khan, T., Brady, S.G., 2000. Error, bias, and long-branch attraction in data for two chloroplast photosystem genes in seed plants. *Mol. Biol. Evol.* 17, 782–797.
- Sota, T., 2002. Radiation and reticulation: extensive introgressive hybridization in the carabid beetles *Ohomopterus* inferred from mitochondrial gene genealogy. *Popul. Ecol.* 44, 145–156.
- Sota, T., Vogler, A.P., 2001. Incongruence of mitochondrial and nuclear gene trees in the carabid beetles *Ohomopterus*. *Syst. Biol.* 50, 39–59.
- Sullivan, J., 1996. Combining data with different distributions of among-site rate variation. *Syst. Biol.* 45, 375–380.
- Swofford, D.L., Olsen, G.J., Waddell, P.J., Hillis, D.M., 1996. Phylogenetic inference. In: Hillis, D.M., Moritz, C., Mable, B.K. (Eds.), *Molecular Systematics*, 2nd edn. Sinauer Associates, Sunderland, MA, pp. 407–514.
- Thornton, J.W., DeSalle, R., 2000. A new method to localize and test the significance of incongruence: detecting domain shuffling in the nuclear receptor superfamily. *Syst. Biol.* 49, 183–201.
- Vogler, A.P., Cardoso, A., Barraclough, T.G., 2005. Exploring rate variation among and within sites in a densely sampled tree: species level phylogenetics of North American tiger beetles (genus *Cicindela*). *Syst. Biol.* 54, 4–20.
- Wheeler, W., 1995. Sequence alignment, parameter sensitivity, and the phylogenetic analysis of molecular data. *Syst. Biol.* 44, 321–331.
- Yoder, A.D., Irwin, J.A., Payseur, B.A., 2001. Failure of the ILD to determine data combinability for slow loris phylogeny. *Syst. Biol.* 50, 408–424.
- Yoon, H.S., Hackett, J.D., Pinto, G., Bhattacharya, D., 2002. The single, ancient origin of chromist plastids. *Proc. Natl Acad. Sci. USA*, 99, 15507–15512.