

Strengths and Limitations of the Minimum Evolution Principle

OLIVIER GASCUEL, DAVID BRYANT, AND FRANÇOIS DENIS

Département Informatique Fondamentale et Applications, L.I.R.M.M., 161 rue Ada, 34392, Montpellier Cedex 5, France;
E-mail: gascuel@lirmm.fr

The idea of inferring phylogenies by selecting trees that minimize the total tree length can be traced back to the 19th century and the mathematician Jakob Steiner. It complies with Occam's principle of scientific inference, which essentially maintains that simpler explanations are preferable to more complicated ones and that ad hoc explanations should be avoided. Parsimony methods, which infer phylogenies directly from character data, are a well-known example of this approach. They search for the tree that requires the minimum number of mutational changes to explain the evolutionary change in the sequences studied. With evolutionary distance data, the definition of simplicity is less obvious. We must first decide how the branch lengths are to be estimated and then how the tree length is to be calculated from these branch lengths. In practice, branch lengths are usually estimated within the least-squares framework. Several different least-squares methods are available to choose from, each using a different model for variances and covariances of the observed distances. Several definitions of tree length have also been proposed, differing from one another by the treatment of negative branch lengths. We shall discuss branch length estimation first (see also Searl, 1971; Bulmer, 1991; Swofford et al., 1996) and then the various definitions of tree length.

Let δ_{ij} be the estimate of the evolutionary distance between taxa i and j , obtained from sequences or any other data, and let $\Delta = (\delta_{(ij)})$ be a column vector containing all the δ_{ij} estimates, with (ij) denoting the index of the pair i, j . Let T be the tree being studied, d_{ij} the distance induced by T between taxa i and j (i.e., $d_{(ij)}$ is equal to the length of the path connecting i to j in T), and $D = (d_{(ij)})$ a column vector containing all ranked d_{ij} distances. Using matrix notation, the branch lengths of T can be represented by a column vector $B = (b_k)$ with b_k denoting the length of branch k , whereas the topology of T can be represented by a 0-1 matrix $A = (a_{(ij)k})$ such that $(a_{(ij)k})$ is

equal to 1 if the branch k lies on the path connecting i and j , but is equal to 0 otherwise. With this notation we have $D = AB$, and the branch lengths are estimated by minimizing the difference between the observation Δ and D . The ordinary least-squares (ordinary-LS) approach involves minimizing the squared Euclidean fit between Δ and D , that is, $(D - \Delta)^T(D - \Delta)$, which yields $B = (A^T A)^{-1} A^T \Delta$. However, this approach implicitly assumes that each δ_{ij} estimate is independent and has the same variance, which is not generally true because of the common evolutionary history of the sequences (or molecules) in question, and because large distances are much more variable than short distances. So, we often use weighted least-squares (weighted-LS), that is, $(D - \Delta)^T V^{-1}(D - \Delta)$, where V is the diagonal matrix containing the variances of the δ_{ij} estimates. This yields $B = (A^T V^{-1} A)^{-1} A^T V^{-1} \Delta$. Weighted-LS accounts for the variable variance of the estimates but not for their dependencies. The minimum variance and hence most reliable branch length estimates are obtained by generalized least-squares (generalized-LS), the formula for which is identical to that of weighted-LS except that V now equals the full variance-covariance matrix of the δ_{ij} estimates. However, generalized-LS is rarely used because the full V matrix is usually poorly known, and because the inversion of V requires a lot of computing time. Ordinary-LS is a special case of weighted-LS, which is obtained when all variances are equal, whereas weighted-LS is a special case of generalized-LS, corresponding to the case in which all covariances are null.

Minimization of these criteria sometimes gives branch lengths with negative values, which does not correspond to any biological process. The general approach for dealing with this problem is non-negative least-squares regression (Lawson and Hanson, 1974), which applies to generalized-LS and thus to weighted-LS and ordinary-LS. Several algorithms (e.g.,

Felsenstein, 1997; Makarenkov and Leclerc, 1999) have been designed for weighted-LS or ordinary-LS on phylogenetic trees. However, incorporating the positivity constraint is not straightforward, and these algorithms tend to be more time consuming than the unconstrained versions derived from the above formulae.

In summary, estimating the branch lengths by generalized-LS is theoretically superior to using weighted-LS, which is in turn superior to using ordinary-LS, and incorporating the positivity constraint is desirable. Note that all these criteria can be used to infer whole phylogenies (and not only the branch lengths corresponding to a given tree topology) by selecting the tree for which they are minimum, as introduced in phylogenetics by Fitch and Margoliash (1967) and Cavalli-Sforza and Edwards (1967). In this case, several authors (e.g., Kuhner and Felsenstein, 1994) have shown that the positivity constraint must be accounted for; otherwise, highly suboptimal trees can use negative branch lengths to produce a low apparent error.

Kid and Sgaramella-Zonta (1971) as well as Rzhetsky and Nei (1993) proposed that ordinary-LS without the positivity constraint be used to estimate branch lengths. They gave two different definitions of tree length. Kidd and Sgaramella-Zonta (1971) defined the tree length to be the sum of the absolute values of the branch lengths, whereas Rzhetsky and Nei (1993) defined the tree length to be the sum of all branch lengths, regardless of whether they were positive or negative. In a third variant, suggested by Swofford et al. (1996), one simply adds those branch lengths that are positive and neglects the negative ones. A fourth variant is to account for the positivity constraint during estimation, then sum the non-negative branch lengths to obtain the tree length. We denote these four variants of the minimum evolution principle as absolute-BL, all-BL, positive-BL, and non-negative-BL. Note that these four variants are not equivalent (see the first example below), and that accounting for positivity during estimation is very different from selecting the positive branch lengths or using absolute values. For any of these four definitions, the minimum evolution principle involves selecting the shortest tree as being the correct phylogeny.

Statistical consistency is a central issue in phylogenetic inference. In the case of

distance-based methods, it is defined as follows. Let T be the correct tree, D the associated tree distance matrix, and Δ the matrix of estimated distances. Assuming that Δ is a consistent estimate of D , the more data we have (e.g., the longer the sequences used to estimate the pairwise distances), the closer Δ is to D . Statistical consistency of tree inference then means that T is obtained with certainty as soon as Δ is sufficiently close to D . In other words, assuming that the model used to estimate the pairwise distance matrix is satisfied, the more data we have, the higher is the probability of recovering the correct tree. This property is essential and has been discussed at length (e.g., Felsenstein, 1978). Consistent methods are in contrast with inconsistent ones (e.g., parsimony in some cases), which may converge towards a wrong tree when the amount of data increases. Numerous phylogenetic inference methods have been proved consistent. Using ordinary-LS, weighted-LS, and generalized-LS (with or without positivity constraint) to infer phylogenies directly is consistent, simply because these criteria have value 0 for T when $\Delta = D$ and are strictly positive for any other tree topology. The agglomerative, distance-based algorithms are also consistent, but the proof is more complex (Atteson, 1997). Finally, Rzhetsky and Nei (1993) demonstrated that the minimum evolution principle is consistent when combined with ordinary-LS and their own definition of tree length (all-BL).

This paper investigates the possible generalizations and extensions of the consistency result of Rzhetsky and Nei (1993). Ordinary-LS fits poorly the features of evolutionary distance data, as explained above. Moreover, we have shown with computer simulations (Gascuel, 2000a) that Rzhetsky and Nei's (1993) version of the minimum evolution principle is not very efficient in terms of topological accuracy and it is improved by other simple approaches that better take into account sequence data characteristics. A similar conclusion was reached by Kuhner and Felsenstein (1994), when comparing neighbor-joining (NJ; Saitou and Nei, 1987) and FITCH (Felsenstein, 1997). NJ is based on Rzhetsky and Nei's (1993) version of the minimum evolution principle, whereas FITCH is based on weighted-LS with the positivity constraint. Numerous authors (e.g., Swofford et al., 1996; Bryant

and Waddell, 1998; Makarenkov and Leclerc, 1999; Gascuel, 2000a) have suggested that the minimum evolution principle could be combined with a more reliable estimation of branch lengths, using weighted-LS or generalized-LS and eventually the positivity constraint. The basic question we address is determining whether such combinations are statistically consistent. We also examine the influence of the various definitions of tree length. In the first section, we provide positive results indicating that the proof of Rzhetsky and Nei (1993), for example, extends to other tree length definitions. In the second section, we show that, unfortunately, the minimum evolution principle has serious limitations when combined with weighted-LS or generalized-LS. A discussion concludes the paper.

STRENGTHS: CONSISTENCY RESULTS

Rzhetsky and Nei's (1993) result is easily seen to extend to the positive-BL and absolute-BL tree length definitions. A direct consequence of statistical consistency is that when $\Delta = D$, the correct tree T has the shortest length among all possible tree topologies. This shortest length property is necessary for consistency. It is also sufficient, as we shall see. The length associated with a tree topology relative to a distance matrix is a continuous function of this matrix. Therefore, when Δ is sufficiently close to D , the estimated tree lengths relative to Δ and to D become close, and T becomes the shortest tree for Δ as it already is for D ; T is then inferred with certainty from Δ . To prove consistency, Rzhetsky and Nei (1993) thus demonstrated that when $\Delta = D$ then $\text{all-BL}(T) < \text{all-BL}(X)$ holds for any tree topology X different from T (using ordinary-LS branch length estimates). Moreover, for any X we have $\text{all-BL}(X) \leq \text{positive-BL}(X) \leq \text{absolute-BL}(X)$, and $\text{all-BL}(T) = \text{positive-BL}(T) = \text{absolute-BL}(T)$ as soon as $\Delta = D$. Combining these elements, we obtain: $\text{positive-BL}(T) < \text{positive-BL}(X)$ and $\text{absolute-BL}(T) < \text{absolute-BL}(X)$ when $X \neq T$ and $\Delta = D$, and the result follows. Note that this proof does not directly apply to the non-negative-BL tree length definition. To the best of our knowledge, the consistency status in that setting is still unknown.

Moreover, we recently demonstrated (Denis and Gascuel, 2000) that the minimum

evolution principle (combined with all-BL, positive-BL, and absolute-BL tree length definitions) remains statistically consistent in a classical model, which generalizes ordinary-LS but is contained in weighted-LS. This model assumes that every taxon i is associated with a strictly positive weight w_i , such that the variance of the δ_{ij} estimate is equal to $1/w_i w_j$. The covariances of the distance estimates are supposed to be null. This model, combined with least-squares estimation of branch lengths, is denoted as taxon-weighted-LS. Ordinary-LS is obtained when w_i is the same for all taxa i . Taxon-weighted-LS is clearly a special case of weighted-LS. A usual application of this model concerns the case where every taxon represents a collection of individuals. For example, in population genetics, suppose we are studying the relationships between groups of people and we take a representative sample of size w_i for each group i . The distance between the groups i and j is then the average of the pairwise individual distances, and, assuming that individual distances have the same variance, the variance of the δ_{ij} estimate is proportional to $1/w_i w_j$. This model can also be used to express that distances associated with some taxa (typically belonging to the outgroup) are less reliable than other (ingroup) distances.

LIMITATIONS: INCONSISTENCY RESULTS

That the minimum evolution principle is not statistically consistent when combined with weighted-LS and generalized-LS is demonstrated here on the basis of the counter-example described in Figure 1 and Table 1. Figure 1a represents the correct tree T ; in this tree every branch has length 0.1, so T has length 0.9 and (e.g.) $\delta_{12} = d_{12} = 0.2$ and $\delta_{35} = d_{35} = 0.4$. Any reasonable method should infer T from Δ , because T is the unique tree that exactly represents Δ . The estimated length of T is 0.9, for any tree length definition and least-squares criterion. Table 1 provides a set of hypothetical variances of the δ_{ij} estimates, e.g., $\text{Var}(\delta_{12}) = 0.01$ and $\text{Var}(\delta_{35}) = 0.50$. The variances are different one from the other, and the covariances are equal to zero, so we are in the weighted-LS framework.

Figure 1b represents an incorrect tree topology with the associated weighted-LS branch length estimates, under both

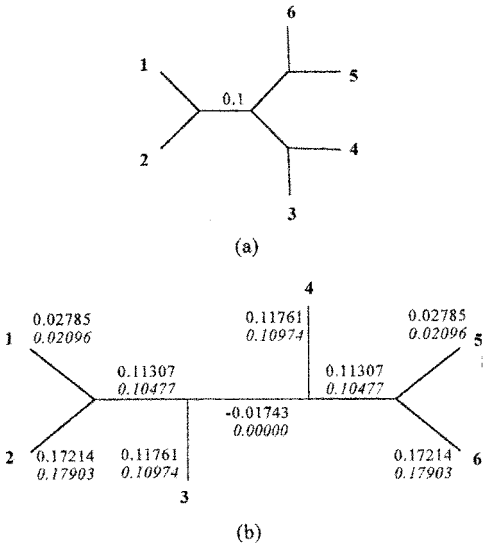


FIGURE 1. (a) The correct tree T , every branch has length 0.1; (b) A shortest, incorrect tree; the unconstrained weighted-LS branch length estimates are indicated by ordinary characters, whereas italic characters refer to those constrained by positivity.

the constrained and unconstrained settings. These estimates were obtained by using the above formulae and non-negative-LS regression, respectively. Computing the length of this incorrect tree, we find the all-BL length is 0.8439, positive-BL is 0.8614, absolute-BL is 0.8788, and non-negative-BL is 0.8290. In other words, the four variants of the minimum evolution principle would not choose the correct tree because an incorrect tree exists that is shorter. This demonstrates that none of these four variants of the minimum evolution principle is statistically consistent when extended to weighted-LS (and thus to generalized-LS) estimation of branch lengths.

However, this counter-example can be seen as artificial, because the variances (Table 1) are highly contrasted, some having value 0.01 and the others 0.50. Moreover,

TABLE 1. Variances of the δ_{ij} estimates.

| i | 1 | 2 | 3 | 4 | 5 | 6 |
|-----|------|------|------|------|------|---|
| 1 | — | | | | | |
| 2 | 0.01 | — | | | | |
| 3 | 0.50 | 0.01 | — | | | |
| 4 | 0.50 | 0.01 | 0.01 | — | | |
| 5 | 0.50 | 0.01 | 0.50 | 0.50 | — | |
| 6 | 0.01 | 0.50 | 0.01 | 0.01 | 0.50 | — |

these variances are quite different from the variances expected with sequence data. In this case, the variance of a distance estimate is an increasing function of the real evolutionary distance that is estimated (see below), and, because the evolutionary distances are supposed to be tree-like, the variances must be close to a tree distance. This is clearly not the case with the variances in Table 1. Because of the four-point condition (Zaretskii, 1965), one would expect, for example, that $Var(\delta_{15}) + Var(\delta_{26}) \approx Var(\delta_{16}) + Var(\delta_{25})$, whereas we have $Var(\delta_{15}) + Var(\delta_{26}) = 1.00$ and $Var(\delta_{16}) + Var(\delta_{25}) = 0.02$ (Table 1). In fact, finding a counterexample for absolute-BL is not easy, at least with 6 taxa, which explains why the variance matrix (Table 1) is special. Finding a counter-example for all-BL is easier, because all-BL is generally strictly smaller than absolute-BL. Thus, minimum evolution combined with weighted-LS may conceivably provide good practical results for realistic variance matrices, especially in the case of absolute-BL. However, depending on the number of taxa and the tree length definition chosen, this combination might also lead to inconsistency in some realistic configurations. An interesting direction for further research would be to characterize the inconsistency zones of the variants of the minimum evolution principle when it is combined with weighted-LS.

The situation is quite different with generalized-LS. Consider the example of Figure 2 and Table 2. Figure 2a represents the correct tree T ; in this tree every branch has length 0.1, so T has length 0.5 and (e.g.) $\delta_{12} = d_{12} = 0.2$ and $\delta_{23} = d_{23} = 0.3$. Once again, any reasonable method should infer T from Δ , because T is the unique tree that exactly represents Δ . The estimated length of T is 0.5, for any tree length definition and least-squares criterion. Table 2 provides a hypothetical variance-covariance matrix. The elements in this matrix are obtained from the evolutionary distances represented by T , assuming the Jukes and Cantor (1969) model and using the following formula (Nei and Jin, 1989; Bulmer, 1991):

$$Covar(\delta_{ij}, \delta_{kl}) = f(c_{ij,kl}) = 3(e^{8c_{ij,kl}/3} + 2e^{4c_{ij,kl}/3} - 3)/16$$

where $c_{ij,kl}$ is the length of the intersection of the paths i, j and k, l in T . For

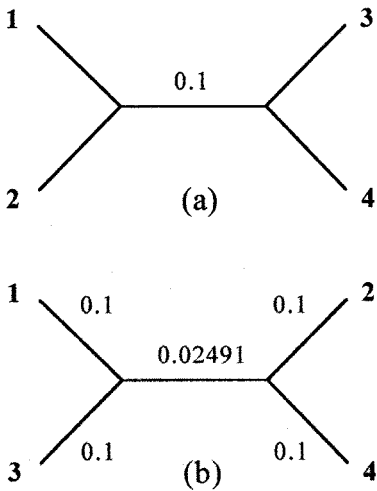


FIGURE 2. (a) The correct tree T , every branch has length 0.1. (b) A shortest, incorrect tree, with the associated generalized-LS branch length estimates.

example, $c_{12,12} = 0.2$, $c_{23,34} = 0.1$, and $c_{12,34} = 0.0$, and we have $Var(\delta_{12}) = f(c_{12,12}) \approx 1.315$, $Covar(\delta_{23,34}) = f(c_{23,34}) \approx 0.590$, and $Covar(\delta_{12}, \delta_{34}) = f(0.0) = 0.0$. In the original formula, the function f is divided by the sequence length. This term is removed from the values given in Table 2, because it does not influence branch length estimation, which is easily seen from the formula $B = (A^T V^{-1} A)^{-1} A^T V^{-1} \Delta$. The use of another model of sequence evolution, for example, Kimura's (1980) or any other, would yield a similar variance-covariance matrix; accordingly, the variance-covariance matrix given in Table 2 can be seen as fully representative of sequence data (within the sequence length factor, which is unimportant).

Figure 2b represents an incorrect tree topology with the associated generalized-LS branch length estimates. Because all branches are positive, the estimated length of this tree is equal to 0.4249 for any tree length definition. As a result, this incorrect

tree is preferred to the correct tree by any of the four variants of the minimum evolution principle. The third possible tree topology (separating $\{1,4\}$ from $\{2,3\}$) has the same estimated length (0.4249) and is also preferred to the correct tree by the minimum evolution principle. Moreover, we replaced the five branch lengths in T (Fig. 2a) with values independently and uniformly drawn from the interval $[0.01, 1.0]$ and modified the variance-covariance matrix accordingly, using the same formula as above. We then observed that the minimum evolution principle is consistent in only $\sim 25\%$ of the trials, whenever the variant that is used. This means that the minimum evolution principle is profoundly inconsistent when combined with generalized-LS and when dealing with sequence data.

DISCUSSION

Table 3 summarizes the various results presented in this paper. To the best of our knowledge, only the consistency of all-BL combined with ordinary-LS was previously published (Rzhetsky and Nei, 1993). The practical impact of the inconsistency of the combination of the minimum evolution principle and weighted-LS remains to be studied, and for generalized-LS, our results seem to preclude any possible use of the minimum evolution principle, at least with sequence data.

Estimating the length of any branch of an incorrect topology has no clear mathematical meaning, which probably explains the inconsistency results. However, this does not explain why the minimum evolution principle becomes consistent when combined with ordinary-LS or taxon-weighted-LS, which can be seen as "natural" from a biological point of view but remains intriguing from a mathematical standpoint.

In Figure 1 the length of the erroneous branch separating $\{1,2,3\}$ and $\{4,5,6\}$ is either negative or null, depending on the estimation procedure. One might thus consider modifying the minimum evolution principle by rejecting any tree that requires negative or null optimal value for any branch, as suggested by Kidd and Sgaramella-Zonta (1971) and others. However, we have observed (Gascuel, 1997a) during computer simulations that the correct tree very often contains negative branch estimates, and Swofford

TABLE 2. Variances (on the diagonal) and covariances of the δ_{ij} estimates.

| ij | 12 | 13 | 14 | 23 | 24 | 34 |
|------|-------|-------|-------|-------|-------|-------|
| 12 | 1.315 | | | | | |
| 13 | 0.590 | 2.209 | | | | |
| 14 | 0.590 | 1.315 | 2.209 | | | |
| 23 | 0.590 | 1.315 | 0.590 | 2.209 | | |
| 24 | 0.590 | 0.590 | 1.315 | 1.315 | 2.209 | |
| 34 | 0.000 | 0.590 | 0.590 | 0.590 | 0.590 | 1.315 |

TABLE 3. Consistency status of the variants of the minimum evolution principle.

| | All-BL | Positive-BL | Absolute-BL | Non-negative-BL |
|-------------------|--------------|--------------|--------------|-----------------|
| Ordinary-LS | Consistent | Consistent | Consistent | Unknown |
| Taxon-weighted-LS | Consistent | Consistent | Consistent | Unknown |
| Weighted-LS | Inconsistent | Inconsistent | Inconsistent | Inconsistent |
| Generalized-LS | Inconsistent | Inconsistent | Inconsistent | Inconsistent |

et al. (1996) suggested that "this extreme approach runs the risk of rejecting the correct tree in certain realistic situations." Such an approach would therefore perhaps be statistically consistent but most likely inefficient in practice.

As explained above, ordinary-LS is not perfectly suited to deal with sequence data, and weighted-LS or generalized-LS, when possible, is preferable. However, as we have shown, combining the minimum evolution principle with weighted-LS or generalized-LS presents serious limitations. A well-founded solution for tree inference is to directly minimize these criteria, using the positivity constraint. Several efficient algorithms exist for this purpose in weighted-LS (De Soete, 1983; Felsenstein, 1997; Makarenkov and Leclerc, 1999), but none in generalized-LS. Alternatively, we can use agglomerative algorithms that perform local weighted-LS or generalized-LS estimations, require little computing time, and are statistically consistent (Gascuel, 1997b, 2000b; Bruno et al., 2000).

ACKNOWLEDGMENTS

Many thanks to Mike Steel and to two anonymous referees for their helpful remarks on the preliminary versions of this paper.

REFERENCES

- ATTESON, K. 1997. The performance of the NJ method of phylogeny reconstruction. Pages 133–148 in *Mathematical hierarchies and biology* (B. Mirkin, F. R. McMorris, F. S. Roberts, and A. Rzhetsky, eds.). American Mathematical Society, Providence, Rhode Island.
- BRUNO, W. J., N. D. SOCCI, AND A. L. HALPERN. 2000. Weighted neighbor joining: A likelihood-based approach to distance-based phylogeny reconstruction. *Mol. Biol. Evol.* 17:189–197.
- BRYANT, D., AND P. WADDELL. 1998. Rapid evaluation of least-squares and minimum-evolution criteria on phylogenetic trees. *Mol. Biol. Evol.* 7:1346–1359.
- BULMER, M. 1991. Use of the method of generalized least squares in reconstructing phylogenies from sequence data. *Mol. Biol. Evol.* 8:868–883.
- CAVALLI-SFORZA, L. L., AND A. W. F. EDWARDS. 1967. Phylogenetic analysis: Models and estimation procedures. *Am. J. Hum. Genet.* 19:223–257.
- DENIS, F., AND O. GASCUEL. 2000. On the consistency of the minimum-evolution principle of phylogenetic inference. LIRMM Research Report 2000-219, available on request from the authors.
- DE SOETE, G. 1983. A least-squares algorithm for fitting additive trees to proximity data. *Psychometrika* 48:621–626.
- FELSENSTEIN, J. 1978. Cases in which parsimony and compatibility methods will be positively misleading. *Syst. Zool.* 27:401–410.
- FELSENSTEIN, J. 1997. An alternating least-squares approach to inferring phylogenies from pairwise distances. *Syst. Biol.* 46:101–111.
- FITCH, W. M., AND E. MARGOLIASH. 1967. Construction of phylogenetic trees. *Science* 155:279–284.
- GASCUEL, O. 1997a. Concerning the NJ algorithm and its unweighted version, UNJ. Pages 149–170 in *Mathematical hierarchies and biology* (B. Mirkin, F. R. McMorris, F. S. Roberts, and A. Rzhetsky, eds.). American Mathematical Society, Providence, Rhode Island.
- GASCUEL, O. 1997b. BIONJ: An improved version of the NJ algorithm based on a simple model of sequence data. *Mol. Biol. Evol.* 14:685–695.
- GASCUEL, O. 2000a. On the optimization principle in phylogenetic analysis and the minimum evolution criterion. *Mol. Biol. Evol.* 17:401–405.
- GASCUEL, O. 2000b. Data model and classification by trees: The minimum variance reduction (MVR) method. *J. Classif.* 17:67–99.
- JUKES, T. H., AND C. R. CANTOR. 1969. Evolution of protein molecules. Pages 21–132 in *Mammalian protein metabolism* (H. N. Munro, ed.). Academic Press, New York.
- KIDD, K. K., AND L. A. SGARAMELLA-ZONTA. 1971. Phylogenetic analysis: Concepts and methods. *Am. J. Hum. Genet.* 23:235–252.
- KIMURA, M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* 16:111–120.
- KUHNER, M. K., AND J. FELSENSTEIN. 1994. A simulation comparison of phylogeny algorithms under equal and unequal evolutionary rates. *Mol. Biol. Evol.* 11:459–468.
- LAWSON, C. M., AND R. J. HANSON. 1974. *Solving least squares problems*. Prentice Hall, Englewood Cliffs, New Jersey.
- MAKARENKOV, V., AND B. LECLERC. 1999. An algorithm for the fitting of a tree metric according to a weighted least-squares criterion. *J. Classif.* 16:3–26.
- NEI, M., AND L. JIN. 1989. Variances of the average numbers of nucleotide substitutions within and between populations. *Mol. Biol. Evol.* 6:290–300.

- RZHETSKY, A., AND M. NEI. 1993. Theoretical foundation of the minimum-evolution method of phylogenetic inference. *Mol. Biol. Evol.* 10:1073–1095.
- SAITOU, N., AND M. NEI. 1987. The neighbor-joining method: A new method for reconstruction of phylogenetic trees. *Mol. Biol. Evol.* 4:406–425.
- SEARL, S. R. 1971. *Linear models*. Wiley, New York.
- SWOFFORD, D. L., G. L. OLSEN, P. J. WADDELL, AND D. M. HILLIS. 1996. Phylogenetic inference. Pages 407–514 *in* *Molecular systematics* (D. M. Hillis, C. Moritz, and B. K. Mable, eds.). Sinauer, Sunderland, Massachusetts.
- ZARESTKII, K. 1965. Reconstructing a tree from the distances between its leaves. *Usp. Math. Nauk* 20:90–92 (in Russian).

Received 12 July 2000; accepted 6 October 2000

Associate Editor: R. Olmstead