

# On the Optimization Principle in Phylogenetic Analysis and the Minimum-Evolution Criterion

Olivier Gascuel

Département Informatique Fondamentale et Applications, LIRMM, Montpellier, France

This paper discusses the optimization principle in phylogenetic analysis, in the case of distance data. We argue that the use of this principle cannot be called into question, except for computing time reasons. We show that the minimum-evolution criterion is not perfectly suited for distance data estimated from sequences, and we present another approach, implemented in the BIONJ algorithm, which allows the data features to be taken into account, while being less demanding in computing time. Simulations show that BIONJ significantly outperforms NJ.

## Introduction

Recently, Nei, Kumar, and Takahashi (1998) published a paper entitled “The optimization principle in phylogenetic analysis tends to give incorrect topologies when the number of nucleotides or amino acids used is small.” They simulated the evolution of sequences along a model tree and compared, for each replicate data set, the observed performance of this (true) tree to that of trees inferred by three usual phylogenetic reconstruction approaches, based, respectively, on maximum-parsimony (MP), minimum-evolution (ME), and maximum-likelihood (ML) criteria. The observed performance of a tree refers to its value for a given criterion (MP, ME, or ML) when considering the data set at hand. The optimization of observed performance is the basis of most phylogenetic reconstruction approaches. For example, in parsimony, we search (among all possible trees) for the tree that requires the minimum number of mutational changes to explain the evolutionary change of the studied sequences. Note that the observed performances of trees, and hence the inferred tree, usually vary from one data set to another, especially with short sequences. Moreover, since the number of possible trees is extremely large, even for a moderate number of taxa, it is usually not possible to use an exact algorithm, and we rely on approximate algorithms that find near-optimal trees within a reasonable amount of computation time, e.g., DNAPARS (Felsenstein 1993) in the case of the MP criterion, or NJ (Saitou and Nei 1987) in the case of the ME criterion.

The results presented in the paper of Nei, Kumar, and Takahashi (1998) can be summarized as follows. When using an exact algorithm, the observed performance of the true tree is always worse than or identical to the performance of the optimal tree. This is not surprising, since the true tree is one tree among all possible trees, and therefore it cannot be better than the best tree. Moreover, with short sequences, reconstruction algorithms often fail to discover the true tree, and consequently the true tree appears to be worse than the in-

ferred tree. With approximate algorithms, the situation is not much different. Near-optimal inferred trees are rarely worse than the true tree, and the topological accuracy of these (fast) algorithms is close to that of the exact (time-consuming) algorithms. In light of these results, Nei, Kumar, and Takahashi (1998) suggest that more attention should be given to testing the statistical reliability of inferred trees, e.g., using the bootstrap method (Felsenstein 1985), than to finding optimal trees with excessive computational effort.

Similar results were previously described in Gascuel (1997*a*, 1997*b*) concerning the ME criterion (see also Kumar 1996, figs. 6 and 7). We first summarize these results and then show that, in the case of the ME criterion, considerations on the optimization principle cannot fully explain the observations. We conclude that the ME criterion is not perfectly suited for evolutionary distance data obtained from sequences, and also that the global optimization principle is not the sole way to conceive, describe, and analyze phylogenetic reconstruction algorithms. We then present another approach, implemented in the BIONJ algorithm (Gascuel 1997*a*), which follows the agglomerative scheme, as NJ, and uses a model of distance data estimated from BIOlogical sequences.

## Simulation Schemes and NJ Algorithm Results

In Gascuel (1997*a*), we simulated six 12-taxon model trees, some complying with the molecular-clock hypothesis and the others not. The Kimura two-parameter model was used with a transition/transversion ratio of 2. Sequence lengths were 300 or 600 sites, and four different evolution conditions were considered, corresponding to low, medium, high, and high/low per-site substitution rates. Evolutionary distances were computed using the standard Kimura (1980) estimate, except for the high/low per-site condition, for which we also used the two-parameter gamma estimate ( $\alpha = 1$ ) of Jin and Nei (1990). Five hundred replications were performed for each condition, involving 15,000 data sets per sequence length. The mean results are displayed in table 1.

With a sequence length of 300, we observed that NJ trees were better in the ME sense (i.e., shorter) than the true tree in 61% of the cases, and longer in 11%. With 600 sites, we found 26% and 8%, respectively. This indicates that the optimal ME tree (not computable

Key words: phylogenetic reconstruction, optimal and heuristic algorithms, criterion to be optimized, agglomerative approach, model of the data.

Address for correspondence and reprints: Olivier Gascuel, Département Informatique Fondamentale et Applications, LIRMM, 161 rue Ada, 34392, Montpellier cedex, France. E-mail: gascuel@lirmm.fr.

*Mol. Biol. Evol.* 17(3):401–405. 2000

© 2000 by the Society for Molecular Biology and Evolution. ISSN: 0737-4038

**Table 1**  
**Mean Results**

			ME CRITERION			TOPOLOGICAL DISTANCE		
			<	=	>	RF	<	>
Sequence data . . . . .	NJ	300 sites	61	28	11	1.42		
		600 sites	26	66	8	0.47		
	NJ + FTS	300 sites	73	26	1	1.50 (-4)	15	19
		600 sites	37	63	1	0.50 (-9)	8	12
	BIONJ	300 sites	58	33	9	1.33 (8)	19	10
		600 sites	24	70	6	0.42 (10)	9	5
i.i.d. normal data . . . . .	NJ	$\sigma = 0.6$	92	5	4	5.72		
		$\sigma = 0.1$	40	52	8	0.81		
		$\sigma = 0.6$	47	32	11	5.53 (3)	25	14
	UNJ	$\sigma = 0.6$	17	80	3	0.75 (8)	10	5

NOTE.—The sequence data results are derived from Gascuel (1997a), except those of NJ + FTS, which were not previously published. The i.i.d. normal data results are derived from Gascuel (1997b). The ME criterion columns <, =, and > provide the percentages of times the inferred tree  $\hat{T}$  is shorter than, identical to, and longer than the true tree  $T$ , i.e.,  $ME(\hat{T}) < ME(T)$ ,  $\hat{T} = T$ , and  $ME(\hat{T}) > ME(T)$ , respectively. UNJ is compared with NJ and not with the true tree, and in this case, the ME columns correspond to  $ME(\hat{T}_{UNJ}) < ME(\hat{T}_{NJ})$ ,  $\hat{T}_{UNJ} = \hat{T}_{NJ}$ , and  $ME(\hat{T}_{UNJ}) > ME(\hat{T}_{NJ})$ , respectively. RF is half of the Robinson and Foulds distance between  $\hat{T}$  and  $T$ , and the numbers in parentheses are the percentages of relative error reduction regarding NJ, i.e.,  $100(RF(\hat{T}_{NJ}, T) - RF(\hat{T}_X, T))/RF(\hat{T}_{NJ}, T)$ , where  $\hat{T}_X$  denotes the tree inferred by the method  $X$  (e.g., BIONJ). The columns < and > provide the percentages of times the topological accuracy of  $X$  is better and worse than that of NJ, i.e.,  $RF(\hat{T}_X, T) < RF(\hat{T}_{NJ}, T)$  and  $RF(\hat{T}_X, T) > RF(\hat{T}_{NJ}, T)$ , respectively. A simple sign test shows that all differences reported in this table are highly statistically significant ( $\alpha < 0.001$ ).

with 12 taxa) missed the true tree in at least 61% of the cases with 300 sites, and in at least 26% with 600 sites. Moreover, this demonstrates that the ME tree cannot markedly improve the NJ tree (in terms of probability of finding the true tree): at most, 11% improvement with 300 sites and 8% improvement with 600 sites. More generally, this indicates that NJ selects a tree in a region where all trees are more or less equivalent and not discernable from the true tree using the ME criterion (Kumar 1996, fig. 6). Therefore, this basically explains the fact observed by several authors (Saitou and Imanishi 1989; Kumar 1996) that the ME tree is not more accurate than the NJ tree.

In Gascuel (1997b), we simulated another type of data. Let  $D = (d_{ij})$  be a tree distance (i.e., a distance represented by a tree with positive branch lengths), where  $d_{ij}$  is the distance between taxa  $i$  and  $j$ . A noise  $\epsilon_{ij}$  was added to every  $d_{ij}$  to obtain the noisy distance  $(\delta_{ij}) = (d_{ij} + \epsilon_{ij})$ . The noises  $\epsilon_{ij}$  were independently and identically distributed (i.i.d.) and normal with variance  $\sigma^2$  and a null expectation. Such i.i.d. normal data are encountered when  $\delta_{ij}$  estimates are the result of real observations with measurement errors, which is close, for example, to DNA-DNA hybridization data (Felsenstein 1987). We considered high and low noise levels ( $\sigma = 0.6$  or  $\sigma = 0.1$ ) and simulated three 12-taxon and three 24-taxon model trees. Five hundred replications were performed for each condition, which involves 3,000 data sets per noise level. Then, we applied reconstruction algorithms to the noisy matrices in order to recover the true tree corresponding to the original tree distance. The mean results are displayed in table 1.

The situation was basically the same as that for sequence data. With high noise (corresponding to short sequences), we observed that NJ trees were shorter than the true tree in 92% of the cases, and longer in 4%. With low noise, we found values of 40% and 8%, respectively. This indicates that there is little room for improvement by optimizing the ME criterion: 4% im-

provement with high noise and 8% improvement with low noise (in terms of the probability of finding the true tree).

**Performance of Near-Optimal ME Trees**

However, as we shall see in this section, the results differ with both types of data when the optimization of the ME criterion is better than that achieved by NJ.

The i.i.d. normal data were dealt with by the UNJ algorithm (Gascuel 1997b). This algorithm is close to NJ but is unweighted, just as UPGMA (Sokal and Michener 1958). The unweighted approach involves giving the same importance to each original taxon, which can be seen as the fundamental principle of the ordinary least-squares (OLS) criterion, and UNJ uses formulae that are exact in the OLS sense for estimating the branch lengths and reducing the distance matrix at each step, while formulae used by NJ are approximate (Saitou and Nei 1987). Moreover, the OLS criterion is critical in the ME criterion definition, since the ME value of a tree is its OLS length estimate (Rzhetsky and Nei 1993). As expected, we observed that UNJ optimized the ME criterion better than NJ did: with high noise, UNJ trees were shorter than NJ trees in 47% of the cases and longer in 11%, while with low noise, we found values of 17% and 3%, respectively (table 1). Moreover, UNJ had better topological accuracy than NJ. For every inferred tree, we measured the number of incorrect branches, which is equal to half of the usual Robinson and Foulds 1981 topological distance between this tree and the true tree. We observed that UNJ trees were closer to the true tree than were NJ trees by 3% with high noise and by 8% with low noise (table 1).

When applying UNJ to sequence data, we observed, as expected, that it optimized the ME criterion better than NJ did. However, it had lower topological accuracy (data not shown). The results obtained with tree swapping were even more surprising. To every NJ

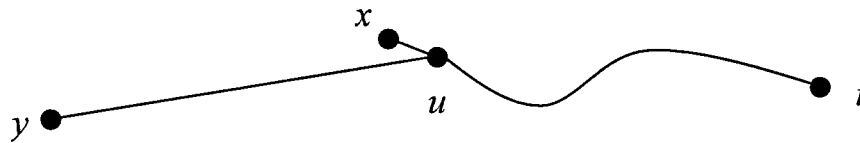


FIG. 1.— $x$ ,  $y$  is the selected pair,  $u$  represents the closest common ancestor of  $x$  and  $y$ , and  $i$  is any taxon (original, or intermediary in the tree building process).

tree, we applied our fast tree-swapping (FTS) algorithm (Gascuel 1996), which is an optimized implementation of Rzhetsky and Nei's (1993) close-neighbor interchange (CNI) algorithm, but restricted to nearest-neighbor interchanges. Given a tree, FTS searches for the pair of nonintersecting clusters separated by one interior branch, which provides the greatest reduction in the ME criterion when exchanged. This process is repeated until no cluster exchange diminishes the ME criterion. Starting with NJ trees, FTS necessarily finds trees that are identical or better in the ME sense. Indeed, we found that these "NJ+FTS" trees were longer than the true tree in only 1% of the cases, with 300 sites as well as with 600 sites, as compared with 11% and 8%, respectively, for NJ alone. However, the topological accuracy was reduced by 4% with 300 sites and by 9% with 600 sites (results not previously published; see table 1). This means that FTS, which is fully justified from the ME principle point of view, tends to degrade the good properties of NJ trees.

In one case (i.i.d. normal data), further optimization of the ME criterion tends to improve the topological accuracy of NJ, while in the other (sequence data), this yields a degradation. The explanation is related not to the optimization principle, but to the ME criterion and its use for sequence data. In fact, the UNJ algorithm, the OLS, and the ME criterion are perfectly suited for i.i.d. normal data (Gascuel 1997*b*). A similar result was found by Degens (1983) concerning UPGMA and the reconstruction of ultrametric trees (rooted with a molecular clock). On the other hand, evolutionary distance data estimated from sequences are very far from the i.i.d. normal model. The variance of the distance estimates is not constant and is much higher for long than for short distances. Moreover, the covariance of these estimates is often high (Nei and Jin 1989; Bulmer 1991). In other words, the variance-covariance matrix of the distance estimates obtained from sequences differs markedly from the identity matrix that corresponds to the i.i.d. model. Also, as we shall explain in the next section, NJ trees are more accurate than ME and UNJ trees simply because NJ takes sequence data features into better account.

### The BIONJ Algorithm: Principle and Performance

We designed BIONJ in response to the above observations (Gascuel 1997*a*). Agglomerative algorithms, such as ADDTREE (Sattath and Tversky 1977) or NJ, progressively build a tree by picking a pair of taxa, creating a new node that represents the cluster of these taxa, and computing a new reduced distance matrix where both taxa are replaced by this node. This scheme admits

of variations, especially concerning the reduction of the distance matrix. As shown in figure 1, assume that the pair  $x$ ,  $y$  is selected and that we have to estimate the distance between the new node  $u$  and any remaining taxon  $i$ . Then it is easily shown that regardless of the value of  $\lambda$ , the following formula is suitable:

$$\delta_{ui} = \lambda(\delta_{xi} - \delta_{xu}) + (1 - \lambda)(\delta_{yi} - \delta_{yu}),$$

where  $\delta_{xu}$  and  $\delta_{yu}$  are any unbiased estimates of the branch lengths  $d_{xu}$  and  $d_{yu}$ . NJ corresponds to  $\lambda = 1/2$  and UNJ to  $\lambda = n_x/(n_x + n_y)$ , where  $n_z$  is the number of original taxa contained in the cluster  $z$ . UNJ is unweighted because  $\lambda$  is proportional to the number of original taxa contained in the clusters and hence gives the same weight to each of these taxa. On the other hand, NJ is weighted, just as WPGMA (McQuitty 1966), and should be denominated as WNJ for the sake of accuracy. The weighted approach considers that two or more close taxa belonging to the same cluster do not yield more information than an isolated taxon, which can be seen as consistent in the phylogenetic-inference context (Sneath and Sokal 1973; see also below).

BIONJ exploits the above property to obtain reliable minimum-variance  $\delta_{ui}$  estimates. To achieve this, it uses a simple first-order model of the variances and covariances of evolutionary distance estimates obtained from sequences. This yields the formula

$$\lambda = 1/2 + \varphi$$

where  $\varphi$  is a correction term that depends on  $\delta_{xu}$  and  $\delta_{yu}$  (at least when  $x$  and  $y$  are original taxa). When  $\delta_{xu}$  and  $\delta_{yu}$  are equal, then  $\varphi = 0$ ,  $\lambda = 1/2$ , and BIONJ is equivalent to NJ. When both differ, i.e., when the substitution rates vary among lineages,  $\varphi$  becomes not null and places more confidence on the shorter and hence more reliable distance. For example, in figure 1, we have  $\varphi \approx 1/2$ ,  $\lambda \approx 1$ , and  $\delta_{ui} \approx (\delta_{xi} - \delta_{xu})$ , which is consistent with the fact that  $x$  is close to  $u$ , which is an ancestor of  $y$ . Moreover, the solution used by NJ ( $\lambda = 1/2$ ) appears to be the first approximation of the above equation. This explains why NJ is more accurate than UNJ and NJ+FTS with sequence data: it is closer to the underlying statistical model and hence performs better.

BIONJ is equivalent to or more accurate than NJ for all tested conditions. The mean results are displayed in table 1. The topological accuracy (measured using the Robinson and Foulds distance) is improved by 8% with 300 sites and by 10% with 600 sites, while the probability of finding the true tree is augmented by 5% and 4%, respectively. The gain is much higher for certain conditions. With highly varying rate model trees and with high substitution rates (maximum pairwise divergence  $\approx 1.0$  substitutions per site), the topological ac-

curacy may be improved by as much as 50%, and the probability of finding the true tree may be improved by 15%. Note, however, that BIONJ is neither better nor worse than NJ in optimizing the ME criterion.

BIONJ actually does not optimize any global criterion. At each step, it selects a pair of taxa by optimizing a partial (or local) criterion, denoted as  $Q$ , which was proposed by Studier and Keppler (1988) and is equivalent to the NJ criterion (Saitou and Nei 1987; Gascuel 1994). When applied to any tree distance  $D$ ,  $Q$  designates with certainty a pair of neighbors of the tree representing  $D$ . In practice, we only have an estimate  $\hat{D}$  of  $D$ , and hence an estimate  $\hat{Q}$  of the true value of  $Q$ . Since we optimize  $\hat{Q}$  and not  $Q$  itself, we do not systematically select a true pair of neighbors. However, the more reliable  $\hat{D}$  is in estimating  $D$ , the more reliable  $\hat{Q}$  (computed from  $\hat{D}$ ) is in selecting neighbors. The BIONJ approach involves iteratively (1) selecting a pair of neighbors by optimizing  $\hat{Q}$  and (2) reducing  $\hat{D}$  such that the new estimates in the reduced matrix are as reliable as possible. This clearly strays from the global optimization paradigm.

BIONJ is a special case of what we have called the minimum variance reduction (MVR) method (Gascuel 2000). This general approach involves achieving step 2 above owing to knowledge of the variance-covariance matrix  $V$  of estimates in  $\hat{D}$ . The computing time needed by MVR depends on  $V$ . In the general case, MVR requires  $O(n^4)$ , where  $n$  is the number of taxa. However, MVR only requires  $O(n^3)$ , just as NJ does, in several special cases, e.g., UNJ ( $V$  is equal to the identity matrix), WLSNJ ( $V$  is diagonal), or BIONJ ( $V$  is nondiagonal and defined by a simple first-order model of sequence data). The MVR approach thus allows the data features to be taken into account and is much less demanding in computing time than current approximate algorithms used for the global optimization of least-squares criteria. For example, FITCH (Felsenstein 1997) requires  $O(n^4)$  to deal with OLS or weighted least-squares (WLS;  $V$  is diagonal), while generalized least-squares (GLS; no constraint on  $V$ ) necessitates (among other things) inverting  $V$ , which requires  $O(n^6)$  (Bulmer 1991). Note also that the above-mentioned agglomerative algorithms are statistically consistent, which means that they converge toward the true tree when the number of sites increases (and when  $\hat{D}$  is a consistent estimate of  $D$ ). This property is essential in phylogenetic inference (Felsenstein 1984) and is actually more important than the capacity to find the optimal tree according to a given criterion. Moreover, Atteson (1997) has demonstrated that the theoretical convergence rate of agglomerative algorithms such as ADDTREE, NJ, UNJ, and BIONJ is in some sense optimal for all possible distance-based methods (at least when the distances in  $D$  are bounded within a reasonable range; see Erdős et al. 1997).

## Conclusions

The optimization principle and its use in phylogenetic inference cannot be called into question. It serves

for model identification in many domains and possesses solid mathematical foundations, especially concerning the least-squares and maximum-likelihood approaches. When there are few available sites, optimization-based methods often fail to recover the correct phylogeny, but any conceivable method would behave similarly in this case, because the information is not sufficient and/or too noisy, and systematically recovering the correct phylogeny is simply impossible.

Actually, the main question that arises is that of which criterion to optimize. The results presented in this paper indicate that the ME criterion is not perfectly suited for distance data estimated from sequences due to its closeness to OLS. Improvement could probably be obtained by combining the ME principle with the WLS (or GLS) estimation of tree length, but optimizing this refined criterion would then become harder and likely not easier than directly optimizing WLS (or GLS).

However, global optimization is not the sole way to conceive phylogenetic reconstruction algorithms. In the case of distance data, the agglomerative approach also possesses solid foundations. Combined with MVR, it allows the data features to be taken into account and leads to simple, fast, and accurate algorithms such as BIONJ. However, the simplicity and speed of these algorithms is perhaps somewhat counterbalanced by a lower ability to recover the true tree, which should clearly be further studied and evaluated.

## LITERATURE CITED

- ATTESON, K. 1997. The performance of the NJ method of phylogeny reconstruction. Pp. 133–148 in B. MIRKIN, F. R. McMORRIS, F. S. ROBERTS, and A. RZHETSKY, eds. *Mathematical hierarchies and biology*. American Mathematical Society, Providence, R.I.
- BULMER, M. 1991. Use of the method of generalized least squares in reconstructing phylogenies from sequence data. *Mol. Biol. Evol.* **8**:868–883.
- DEGENS, P. O. 1983. Hierarchical cluster methods as maximum likelihood estimators. Pp. 249–253 in J. FELSENSTEIN, ed. *Numerical taxonomy*. Springer, Berlin.
- ERDÖS, P. L., M. A. STEEL, L. A. SZÉKELY, and T. WARNOW. 1997. Constructing big trees from short sequences. *Automata Lang. Programming 24th Int. Colloquium Lect. Notes Comput. Sci.* **1256**:827–837.
- FELSENSTEIN, J. 1984. Distance methods for inferring phylogenies: a justification. *Evolution* **38**:16–24.
- . 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**:783–791.
- . 1987. Estimation of hominoid phylogeny from a DNA hybridization data set. *J. Mol. Evol.* **26**:123–131.
- . 1993. PHYLIP (phylogeny inference package). Distributed by the author, Department of Genetics, University of Washington, Seattle.
- . 1997. An alternating least-squares approach to inferring phylogenies from pairwise distances. *Syst. Biol.* **46**:101–111.
- GASCUEL, O. 1994. A note on Sattath and Tversky's, Saitou and Nei's and Studier and Keppler's algorithms for inferring phylogenies from evolutionary distances. *Mol. Biol. Evol.* **11**:961–963.

- . 1996. Algorithmes efficaces pour la construction d'arbres d'évolution minimum. Rapport de recherche LIRMM-96019.
- . 1997a. BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. *Mol. Biol. Evol.* **14**:685–695.
- . 1997b. Concerning the NJ algorithm and its unweighted version, UNJ. Pp. 149–170 in B. MIRKIN, F. R. MC-MORRIS, F. S. ROBERTS, and A. RZHETSKY, eds. *Mathematical hierarchies and biology*. American Mathematical Society, Providence, R.I.
- . 2000. Data model and classification by trees: the minimum variance reduction (MVR) method, *J. Classif.* (in press).
- JIN, L., and M. NEI. 1990. Limitations of the evolutionary parsimony method of phylogenetic analysis. *Mol. Biol. Evol.* **7**:82–102.
- 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.
- KUMAR, S. 1996. A stepwise algorithm for finding minimum evolution trees. *Mol. Biol. Evol.* **13**:584–593.
- MCQUITTY, L. L. 1966. Similarity analysis by reciprocal pairs for discrete and continuous data. *Educ. Psychol. Meas.* **26**:825–831.
- NEI, M., and L. JIN. 1989. Variances of the average numbers of nucleotide substitutions within and between populations. *Mol. Biol. Evol.* **6**:290–300.
- NEI, M., S. KUMAR, and K. TAKAHASHI. 1998. The optimization principle in phylogenetic analysis tends to give incorrect topologies when the number of nucleotides or amino acids used is small. *Proc. Natl. Acad. Sci. USA* **95**:12390–12397.
- ROBINSON, D. F., and L. R. FOULDS. 1981. Comparison of phylogenetic trees. *Math. Biosci.* **53**:131–147.
- 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. IMANISHI. 1989. Relative efficiencies of the Fitch-Margoliash, maximum-parsimony, maximum-likelihood, minimum-evolution, and neighbor-joining methods of phylogenetic reconstructions in obtaining the correct tree. *Mol. Biol. Evol.* **6**:514–525.
- SAITOU, N., and M. NEI. 1987. The neighbor-joining method: a new method for reconstruction of phylogenetic trees. *Mol. Biol. Evol.* **4**:406–425.
- SATTATH, S., and A. TVERSKY. 1977. Additive similarity trees. *Psychometrika* **42**:319–345.
- SNEATH, P. H. A., and R. R. SOKAL. 1973. *Numerical taxonomy*. Freeman, San Francisco.
- SOKAL, R. R., and C. D. MICHENER. 1958. A statistical method for evaluating systematic relationships. *Univ. Kans. Sci. Bull.* **38**:1409–1438.
- STUDIER, J. A., and K. J. KEPPLER. 1988. A note on the neighbor-joining method of Saitou and Nei. *Mol. Biol. Evol.* **5**:729–731.

MANOLO GOUY, reviewing editor

Accepted November 8, 1999