Letter to the Editor

Nonparametric Phylogenetic Inference from Restriction Cleavage Sites

Alan R. Templeton
Department of Biology, Washington University

Nei and Tajima (1985) recently have criticized my algorithms (Templeton 1983a, 1983b) that make phylogenetic inference from restriction-site maps. The purpose of this letter is to address their criticisms and to examine the reasons for our difference of opinion concerning the range of validity of my algorithms. These points will also be illustrated by a worked example using Hawaiian *Drosophila*.

Nei and Tajima (1985) have suggested that my discrimination between two primate phylogenies (Templeton 1983b) might be statistically flawed if the data included a double counting of some restriction sites. Overlap could arise owing to enzymes with multiple recognition sequences that include the recognition sequences of single cutters. However, in generating the scores used to test these hypotheses (table 2 of Templeton 1983b), I used only the novel informative patterns revealed by the multiplic cutters. No site was counted twice. Hence, the concern raised by Nei and Tajima does not apply.

The primary criticism of Nei and Tajima concerns the range of validity of the algorithms. I have concluded (Templeton 1983b) that the algorithms are valid if the average substitution rate (λ) times the time of divergence (t) is \( \lambda t \approx 0.05 \). Nei and Tajima (1985) argue that the much smaller value of 0.01 is necessary. They justify this smaller \( \lambda t \) value on the basis of two theoretical arguments. The first concerns the difficulty, in the absence of outgroup data, of distinguishing a single convergent gain of restriction sites from a double loss. (fig. 3 of Nei and Tajima [1985] and associated text). However, the criticized algorithms were explicitly limited to data with an outgroup (Templeton 1983b). Hence, this theoretical objection lies outside the inference domain delimited by me (Templeton 1983b) and is therefore completely irrelevant to the validity of the 0.05\( \lambda t \) criterion.

The second theoretical objection raised by Nei and Tajima is that the probability of three parallel losses of a restriction site (or of one gain followed by two losses) is not necessarily smaller and can be greater than the probability of two parallel gains (or of one loss followed by a gain). These conclusions are sound. However, Nei and Tajima state: “Templeton (1983a) recognized this problem but concluded that if \( \lambda t < 0.05 \), the error introduced in a parsimony [emphasis mine] tree is small. . . . [A] lower criterion (\( \lambda t < 0.01 \)) is necessary.” However, the 0.05 criterion was specifically recommended for the algorithms of Templeton (1983b), which are not the same as the parsimony criterion used by Nei and Tajima in arriving at their conclusion (see, in particular, the discussion on pp. 227–229 and 239–240 of my article [Templeton 1983b]).

Although I used parsimony (Templeton 1983b), my use of it differed from Nei and Tajima’s use of it in the following two fundamental ways: (1) it used compatibility to restrict the application of parsimony to only a subset of the data, and (2) it used a weighting scheme that can assign different weights to patterns that involve the same number of events and the same weight to patterns that involve different numbers of
events. I will now show how these two differences invalidate the conclusion of Nei and Tajima concerning the limits of validity of my algorithms.

First, consider my algorithm for estimating the topology of a phylogeny. The estimate is based on compatibility, with the data generated by any given restriction enzyme being judged as incompatible if any multiple event has occurred. Hence, it is irrelevant whether a multiple event is regarded as two gains or three losses; both would result in the same qualitative inference of "incompatibility," and parsimony would not be invoked to make any further distinctions. Therefore, it is impossible for my phylogeny-estimation algorithm (Templeton 1983b) to be influenced by the ambiguity between double gains and triple losses. The estimation procedure has the formal mathematical property of being invariant with respect to this ambiguity.

Second, consider my test for discriminating between two hypothesized branching orders. The first step is to construct maximum-parsimony trees for each set of restriction sites cut by a particular enzyme under each of the two alternative branching orders. Although these restriction-enzyme trees are based on maximum parsimony, the actual test is based solely on a score assigned to the trees that assigns different weights to different types of events. A tree is assigned a score of 0 if no multiple events of any sort occur. For each additional event that involves the loss of a restriction site, a score of $-1$ is given, and for each additional event that involves the gain of a site, $-2$ is given. The score of the restriction-enzyme tree is, then, simply the sum of the individual site scores. A convergent gain or a loss-gain (i.e., one additional gain) contributes a $-2$ to the total score. This contribution is identical to the contribution of a triple loss or of a gain, double loss (two additional losses, each with a score of $-1$, which sum to $-2$). Since the test is based exclusively on these scores, the test has the formal mathematical property of being invariant with respect to the ambiguity between triple losses and convergent gains.

The only algorithm that I give (Templeton 1983b) that is not analytically invariant to triple losses versus double gains is the test of the molecular-clock hypothesis. Accordingly, I cautioned (Templeton 1983b) that this test should only be performed on contrasts for which convergences of any sort are rare. Under the recommended conditions, the objections of Nei and Tajima are not critical.

Therefore, neither of the theoretical justifications given by Nei and Tajima (1985) for the 0.01 criterion are valid for my algorithms. Another reason for the discrepancy between my recommendations and those of Nei and Tajima stems from the underlying assumptions of our respective models. Consequently, it is necessary to examine these assumptions.

The genetic-distance approach of Nei and Li (1979) and the probabilistic foundation (Templeton 1983a) of my algorithm (Templeton 1983b) both assume that substitutions at all nucleotides occur independently and with identical distributions that are homogeneous in all lineages over both the molecule and time. Despite some slightly different mathematical approaches and approximations, my results (Templeton 1983a) and those of Nei and Li (1979), Li (1981), and Nei and Tajima (1985) are all in close agreement concerning the probabilities of the various events that determine restriction-site evolution.

Nei and Li (1979) use the parametric form of these probabilities to generate expectations that serve as the basis of a genetic-distance measure. In contrast, I used these probability models only as guidelines in developing nonparametric ranking criteria (Templeton 1983a, 1983b). To achieve robustness to the underlying assumptions, I based my ranking criteria only on events whose parametric probabilities differed consistently by at least an order of magnitude in the relevant range of $\lambda t$ values ($\lambda t \leq 0.05$). In using this same criterion, the probability models of Nei and Tajima (1985) implied that convergent gains, loss-gains, triple losses, and gain-double losses should all be treated as an equivalence class in terms of the ranking procedure. As has already
been demonstrated, my ranking procedure (Templeton 1983b) does indeed treat these different events as equivalent.

My nonparametric approach (Templeton 1983b) was explicitly motivated by doubts concerning the validity of the homogeneity assumptions. The disagreement between me and Nei and Tajima over the 0.05 criterion stems in part from the latter's different views concerning the homogeneity assumption. In the Nei and Tajima (1985) model, $\lambda t$ is a set parametric value; for me (Templeton 1983b), it is an empirical average, with some portion of the data being above this average and some below. As I have shown, (Templeton 1983b, pp. 227-228), the compatibility approach effectively gives much more weight to that subset of the data with the lower realized $\lambda t$ values. Hence, when the average $\lambda t$ value is $\leq 0.05$, the actual statistical inference in my algorithm is primarily drawn from a subset of the data with realized $\lambda t$ values considerably $< 0.05$. Consequently, in practice, there is less of a discrepancy between my recommendations and those of Nei and Tajima than is implied by the latter.

One can make an empirical evaluation of the 0.05 criterion by studying a group of organisms for which much is already known about their phylogenetic relationships and that have $\lambda t$ values in the 0.05 range. This is the approach taken by DeSalle (1984), who examined mtDNA restriction-site evolution in Hawaiian picture-wing Drosophila. Since there is probably no other group of living organisms for which we have as much nonmolecular data available for phylogenetic inference (Carson 1983, and references therein), Hawaiian Drosophila constitute a good empirical test case of the validity of various algorithms of phylogenetic inference. Figure 1 shows a phylogeny for eight species of Hawaiian Drosophila, estimated by my algorithm (Templeton 1983b) on the basis of data on 23 restriction enzymes (DeSalle 1984). This phylogeny is completely consistent with the outside information available on these species. DeSalle (1984) also used my algorithm to test the ability of these data to discriminate between alternative phylogenies. With the exception of the internal branching order of the alpha lineage, the estimated branching order is statistically significant at the 5% level against alternative branching orders.

Figure 2 shows the results of applying the genetic-distance, unweighted pair-group method of analysis (UPGMA) algorithm of Nei et al. (1985) to these data. As is immediately evident from the broadly overlapping SD bars surrounding the nodes shown in figure 2, not a single node is cleanly resolved. Moreover, the UPGMA tree

---

**FIG. 1.**—The estimated phylogeny, as determined on the basis of my algorithm (Templeton 1983b), for eight Hawaiian Drosophila species. The nonmolecular data indicate that D. neopicta is in a different taxonomic section than the remaining species, which, in turn, are split into two lineages; the alpha lineage (D. cyrtoloma, D. melanocephala, and D. neoperkinsi) and the beta lineage (D. hemipeza, D. differens, D. planitibia, and D. silvestris). Asterisks indicate those branch orderings that, as determined on the basis of my testing algorithm (Templeton 1983b), are significant at the 5% level against alternatives.
clearly clusters *D. neopicta* within the alpha lineage despite the fact that the alpha and beta lineages are in the same section and *neopicta* is in a different section. This represents a gross phylogenetic error.

This example illustrates two additional points. First, from figure 2 it is evident that the depth of this phylogeny is $\sim 0.06 \lambda t$ units, with only one $\lambda t$ depth $< 0.01$. Moreover, the $\lambda t$ depth of the beta lineage is $0.03 < 0.05$. My estimation and hypothesis-testing algorithms (Templeton 1983b) gave a 100% accurate and statistically significant resolution within this range. Recall that Nei and Tajima (1985) claimed that my approach would yield erroneous inferences unless all branch lengths are $< 0.01$ and would display lower statistical power than the genetic-distance approach. Hence, this example provides empirical support for the conclusion that my algorithms are valid and can have greater statistical power than genetic-distance approaches when $\lambda t$ is $\sim 0.05$.

Second, this example shows that my algorithm is more robust to deviations in the evolutionary assumption of rate homogeneity than is the genetic-distance approach, which is known to be violated in this case (DeSalle 1984). The issue of the validity of assumptions is the critical point in deciding which algorithm to use. If the homogeneity assumptions of Nei et al. (1985) hold, their algorithm represents an excellent technique for analyzing restriction-site data, even when $\lambda t$ values are $> 0.05$. If these assumptions are not valid, my algorithms are superior, but only if $\lambda t$ is $\sim < 0.05$. Consequently, the choice of algorithms depends completely on the degree of confidence that the investigator has in the evolutionary assumptions, and this degree of confidence will undoubtedly vary from data set to data set. What is needed are more empirical examples of the sort given by DeSalle (1984), to illuminate just how frequently or infrequently the evolutionary assumptions commonly made are valid.
Acknowledgments

This work was supported by National Institutes of Health grant R01 GM31571. I would like to thank Drs. Peter Smouse and Walter Fitch for their critical reading of an earlier draft of this letter.

LITERATURE CITED


