Letter to the Editor

Problems Arising in Phylogenetic Inference from Restriction-Site Data

Masatoshi Nei* and Fumio Tajima†

*Center for Demographic and Population Genetics, The University of Texas Health Science Center at Houston; and †Department of Biology, Kyushu University, Fukuoka, Japan

One of the major problems we were concerned with (Nei and Tajima 1985) is the extent of errors that would occur in estimates of the number of restriction-site changes for the various branches of the evolutionary tree when Templeton’s (1983) method is used. Templeton (1987) addresses this problem by stating that tests of the molecular-clock hypothesis should only be performed by means of his method if “convergences of any sort are rare.” Since such tests depend on the accuracy of the estimates of restriction-site changes, we take his statement to indicate agreement with our conclusion that his method does not provide reliable estimates of the numbers of restriction-site changes unless convergences of any sort are rare.

Templeton’s (1987) letter principally concerns itself with the conditions under which his method is appropriate for testing alternative tree topologies. He concludes: “If the homogeneity assumptions of Nei et al. (1985) hold, their algorithm represents an excellent technique for analyzing restriction-site data, even when λt values are >0.05. If these assumptions are not valid, my algorithms are superior, but only if λt is ~ ≤0.05.” Despite the moderateness of Templeton’s conclusion, we remain unconvinced.

At the present time, the theoretical basis for testing the difference in the maximum-parsimony or compatibility score between two topologies is not well established. We have shown (Nei and Tajima 1985) that if we use Templeton’s test, his phylogeny 1 (which places the chimpanzee closer to the gorilla than to the human—and is denoted as phylogeny A in our [Nei and Tajima 1985] paper) for the human, chimpanzee, gorilla, orangutan, and gibbon is never inferior to phylogeny 4 (which represents tri- furcation of the human, chimpanzee, and gorilla—and is denoted as phylogeny B in our paper) even if phylogeny 4 is the correct one. This clearly indicates that Templeton’s test is not valid for this case. This conclusion remains the same regardless of whether the ambiguity between double gains and triple losses affects Templeton’s scoring system.

Let us explain this problem in a little more detail. Let s be Templeton’s score for an enzyme for the comparison of phylogenies 1 and 4. In the Wilcoxon signed-rank test that Templeton used, s must be a random variable with a mean of zero under the null hypothesis. In other words, the maximum-parsimony estimate of mutational changes for phylogeny 1 must have the same distribution as that for phylogeny 4. Actually, this is not true, and the mean or expectation of s is positive, as we showed. Therefore, the Wilcoxon test is not applicable. In the other phylogeny comparisons that Templeton considered, s is not strictly nonnegative, but its expectation (E[s]) may not be zero under the null hypothesis. Without proof that E(s) = 0 under the null hypothesis, one cannot use the Wilcoxon test. Since s is computed from the

1. Key words: phylogenetic trees, hominoid evolution, restriction-site data.

Address for correspondence and reprints: Dr. Masatoshi Nei, Center for Demographic and Population Genetics, The University of Texas Health Science Center at Houston, P.O. Box 20334, Houston, Texas 77225.
maximum-parsimony estimate for each enzyme, E(s) is unlikely to be zero in most cases. Furthermore, it is not clear what kind of tree corresponds to the null hypothesis in each case.

When there are three species involved, it is possible to conduct a statistical test of topological differences, regarding a trifurcating tree as the null-hypothesis tree (Felsenstein 1985). Under this null hypothesis, the three species are related to one another with equal distance as long as the molecular clock applies. Rejection of this null hypothesis leads to the conclusion that a particular pair of species are more closely related to each other than to the remaining one. Li (1986) applied this method to Ferris et al.'s (1981) data for mitochondrial DNA in order to distinguish between the alternative phylogenies of the human, chimpanzee, and gorilla but could not find any significant difference between them. This indicates that Templeton's (1983) conclusion—that is, that his phylogeny 1 is significantly better than phylogeny 2—is not warranted. Li also showed that the number of phylogenetically informative sites (31) in Ferris et al.'s data is far less than the number (~400) required to obtain, with a probability of 95%, the correct tree (see Li 1986, table 9).

When there are four or more species involved, a statistical test of topological differences becomes exceedingly difficult and there seems to be no existing method that is mathematically sound (Felsenstein 1985). (For an unrooted tree of four species, Cavender [1981] developed a nonparametric test.) This is particularly so when one drops the assumption of the molecular clock. One of the major problems for this case is the difficulty in setting up a proper null hypothesis or null-hypothesis tree.

Templeton argues that the effect of rate heterogeneity on his algorithm is unimportant. We believe that his argument should be documented by mathematical proof or computer simulation, since a verbal argument is unreliable in this type of problem.

At the present time, phylogenetic trees are often constructed using parsimony or compatibility methods. As we mentioned in our 1985 paper (p. 201), these methods seem to be useful for the case of small \( \lambda t \) values. However, parsimony or compatibility methods do not utilize data on so-called singular sites (Fitch 1977), so that it is not clear whether they are superior to distance methods. As noted in our previous paper, there are many distance methods that are applicable to the case of rate heterogeneity (see Nei 1987); and Saitou and Nei's (1986) computer simulation indicates that some of these methods are better than parsimony or compatibility methods, at least under certain circumstances.

The final problem that Templeton discusses is an empirical test of tree-making methods. He states that the evolutionary relationship of some Hawaiian Drosophila species has been well established by means of nonmolecular studies and that this can be used as a test case of the validity of various algorithms of phylogenetic inference. According to him, the topology of the tree constructed by his algorithm when DeSalle's (1984) restriction-site data for mtDNA is used is in complete agreement with the known phylogeny but the topology obtained by Nei et al.'s (1985) unweighted pair-group method of analysis (UPGMA) is not.

First, we do not believe that the phylogeny of the eight Hawaiian species in question is firmly established. There are no fossil records for this group of organisms, and the evolutionary relationship that has been derived from morphological, behavioral, and chromosomal data is only hypothetical (Carson 1983).

Second, the difference in topology between the tree obtained with Templeton's method and that obtained with ours is due to the fact that Templeton used D. neopicta as an outgroup species for his method but not for ours. If D. neopicta is known to be an outgroup, one should use it as such and do the UPGMA analysis for the rest of the data. The results obtained would then be the same for both methods.
FIG. 1.—Eight different topologies considered by DeSalle (1984) for the seven species of Hawaiian Drosophila: D. neopicta (N1, N2), D. cyrtoloma (CY), D. melanocephala (MC), D. neoperkinsi (NK), D. hemipeza (HZ), D. differs (DI), and D. planitibia (PL). N1 and N2 represent different individuals from D. neopicta. Drosophila silvestris was not considered by DeSalle. The α lineage consists of D. cyrtoloma, D. melanocephala, and D. neoperkinsi, whereas the β lineage is composed of D. hemipeza, D. differs, D. planitibia, and D. silvestris.

Citing DeSalle (1984), Templeton states that when his algorithm is used, “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.” This statement seems to be incorrect, even if one accepts Templeton’s test of topological differences. DeSalle (1984) considered the eight different topologies given in figure 1 and found that tree II is significantly different only from tree V. Of course, this does not mean that each of the branching orders with asterisks in Templeton’s figure 1 (equivalent to tree II) is statistically significant against alternatives. (Strictly speaking, DeSalle’s statistical conclusion should not be applied to Templeton’s fig. 1, because the species used are not identical.) In fact, the application of Li’s (1986) three-species method to DeSalle’s data indicates that the branching orders within the β-lineage group are not statistically significant against alternatives.

It should also be noted that nonparametric tests such as Templeton’s should be less powerful than parametric tests such as Nei et al.’s (1985). Since Nei et al.’s method does not discriminate between different branching orders, it is hard to imagine how Templeton’s test can.

Finally, where polymorphism for orthologous genes persists between two speciation events, Nei (1987, p. 288–289) has shown that the correct evolutionary relationship for one allele from each of several species may not be the same as that for the species themselves. This noncongruity is more likely when the species are more closely related and when fewer orthologous loci are examined. In the case of DeSalle’s Hawaiian Drosophila, the species are closely related and only one independently evolving unit of DNA (mtDNA) was studied.

Acknowledgments

We thank R. DeSalle for sending us a copy of his Ph.D. thesis. We also thank Wen-Hsiung Li, Paul Sharp, and Clay Stephens for their comments on an earlier draft of this paper.
LITERATURE CITED


