On the Delta Q-Test of Templeton[1]

Naruya Saitou
Center for Demographic and Population Genetics, The University of Texas
Health Science Center at Houston

Using a modified form (delta Q [ΔQ]-test) of Pielou's (1979) Q-test, Templeton (1985) reanalyzed Sibley and Ahlquist's (1984) DNA-DNA hybridization data for hominoid species. He concluded that the phylogeny (fig. 1A) supported by Sibley and Ahlquist is not significantly better than the phylogeny (fig. 1B) favored by Templeton (1983). However, Templeton's (1985) ΔQ-test has several statistical problems, and his conclusion does not seem to be justified.

In brief, Pielou's Q-test and Templeton's ΔQ-test are as follows: Consider the distance matrix in table 1A, where $d_{ij}(i > j)$ denotes the evolutionary distance between species i and j. In the Q-test, d_{ij} is compared with d_{kl} ($k > l$) with the restriction $i < k$, and a random variable x, which takes 1 when $d_{ij} < d_{kl}$, 0 when $d_{ij} > d_{kl}$, and 0.5 when $d_{ij} = d_{kl}$, is considered. The Q-statistic is the sum of x for all possible comparisons of d_{ij} and d_{kl}. If we assume that phylogeny A in figure 1 is correct and that d_{ij} increases in proportion to evolutionary time without error, we will have the inequality $d_{2j} < d_{3j} < d_{4j} < d_{5j}$ ($j = 1, \cdots 4$). In this case Q = 9 + (2 × 7) + (3 × 4) = 35 for distance matrix A of table 1. This is the maximum value Q can take for the five-species case. In practice, d_{ij} may be subject to random errors, and, furthermore, the phylogeny A of figure 1 may not be true. Pielou's (1979) original test is for examining whether or not there is a hierarchical structure of species (strata in her original problem), and the null hypothesis is that there is no hierarchical structure, i.e., that all species diverged at the same time. To test this null hypothesis, Pielou's Q-test is justified. In evolutionary studies, however, there is almost always some hierarchical structure. For this reason, Templeton (1985) introduced the ΔQ-test. In this test the Q-statistic is computed for two different phylogenies (e.g., A and B in fig. 1), and the difference in Q between the two phylogenies is called ΔQ. Templeton obtained the probability distribution of ΔQ and used this distribution for testing the statistical significance of ΔQ. However, this distribution is based on the assumption that there is no hierarchical structure.

As mentioned earlier, the Q-value for matrix A of table 1 is 35. By contrast, the Q-value for matrix B, which corresponds to phylogeny B of figure 1 becomes 29–31, depending on the differences among d_{41}, d_{42}, and d_{43} (see table 1B). The distribution of ΔQ (Templeton's table 2) indicates that ΔQ = 6 is required to achieve statistical significance at the 5% level, but other, smaller values are nonsignificant. If we apply the above test to Sibley and Ahlquist's (1984) data, ΔQ becomes 4. Therefore, this test suggests that phylogeny A is not favored against phylogeny B.

However, we note that ΔQ depends only on the ranks among d_{4j}'s in the present case. With the inequality $d_{2j} < d_{3j} < d_{4j} < d_{5j}$ ($j = 1, \cdots 4$) maintained, ΔQ becomes 6 only when d_{43} is smaller than both d_{41} and d_{42}. Even when the strict rate constancy is assumed, where $d_{41} = d_{42} = d_{43}$, ΔQ = 5 (see table 1). Obviously, d_{41}, d_{42}, and d_{43} may vary by chance effects even if species 4 is remotely related to species 1, 2, and 3, and ΔQ = 4 may be obtained with a high probability. It is therefore clear that the ΔQ-test is inadequate for testing topological differences.
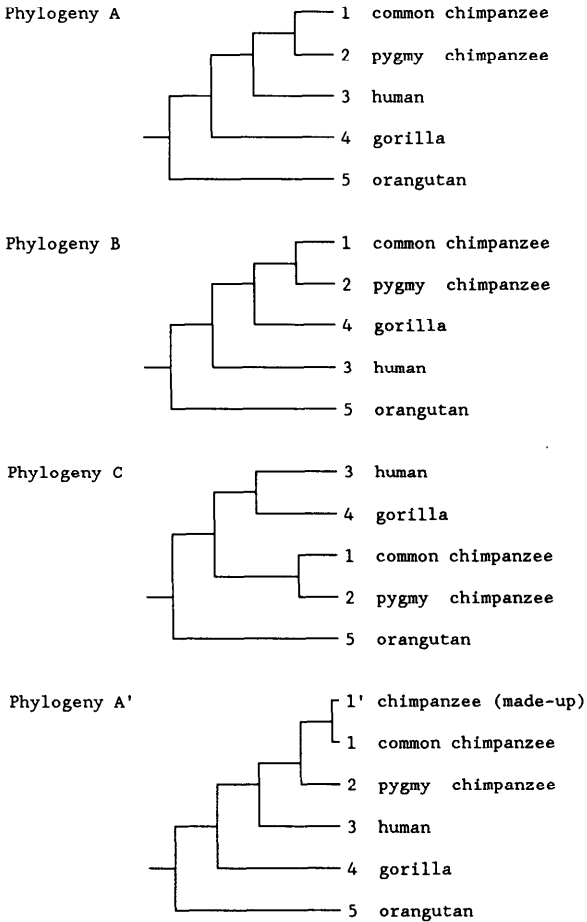
FIG. 1.—Three probable phylogenies (A, B, and C) for five hominoid species and a derivative (A') of phylogeny A in which one chimpanzee species (made-up) is included.

Of course, one can still argue that phylogeny A is better than phylogeny B when ΔQ becomes 6. When we consider the branching orders of humans, chimpanzees, and gorillas, however, there is another possible phylogeny (fig. 1C). This phylogeny cannot be compared with phylogeny A or B in terms of ΔQ (Templeton 1985). Therefore, even if phylogeny A is judged to be superior to phylogeny B with $\Delta Q = 6$, it is not necessarily the best tree; the true tree could be phylogeny C.

Another problem of the ΔQ-test is that the power of the test depends on the number of OTUs (operational taxonomic units) compared. Templeton (1985) acknowledged this point but considered it as a good support for the ΔQ-test. Suppose that we include the third species of chimpanzee (made-up) and obtain phylogeny A' of figure 1. In this case, if we switch species 3 and 4, ΔQ becomes 9–12, again depending on the values of d_{41} ($\approx d_{41'}$), d_{42}, and d_{43}. The smallest ΔQ (=9) is still not significant at the 5% level, but the probability of $\Delta Q \geq 9$ is now 0.0643. As the number of OTUs increases from the chimpanzee lineage gradually, even the smallest ΔQ becomes statistically significant. This peculiar property has occurred because the probability distribution of ΔQ is obtained under the assumption that no hierarchical structure of OTUs exists. This assumption is certainly unrealistic in most statistical studies of phylogenetic trees.

Table 1
Various ΔQ-Values for Two Distance Matrices

	A. Correctly Ordered Matrix					B. Reordered Matrix (Species 3 and 4 Exchanged)			
2	d_{21}				2	d_{21}			
3	d_{31}	d_{32}			4	d_{41}	d_{42}		
4	d_{41}	d_{42}	d_{43}		3	d_{31}	d_{32}	d_{43}	
5	d_{51}	d_{52}	d_{53}	d_{54}	5	d_{51}	d_{52}	d_{54}	d_{53}
Species	1	2	3	4	Species	1	2	4	3
	($Q = 35$)					($Q' = 29$–31)			

Q'	ΔQ	Conditions		
31	4	$d_{43} > d_{41}$ and $d_{43} > d_{42}$		
30.5	4.5	$d_{41} < d_{42} = d_{43}$ or $d_{42} < d_{41} = d_{43}$		
30	5	$d_{41} < d_{43} < d_{42}$ or $d_{42} < d_{43} < d_{41}$ or $d_{41} = d_{42} = d_{43}$
29.5	5.5	$d_{41} > d_{42} = d_{43}$ or $d_{42} > d_{41} = d_{43}$		
29	6	$d_{43} < d_{41}$ and $d_{43} < d_{42}$		

In my view, Templeton's (1985) criticism of Sibley and Ahlquist's (1984) use of the t-test is not really justified. Since Sibley and Ahlquist considered only experimental errors, all the $T_{50}H$ values are independent. Therefore, I do not think that there is anything wrong with the t-test. The pooling of "*Gorilla* $\times$ *Pan, Homo*" comparisons as done by Sibley and Ahlquist (1984) also seems to be valid, since they wanted to show that the human-chimpanzee clustering obtained by the distance Wagner method was significant. Templeton (1985) used the t-test for *Gorilla-Homo* vs. *Gorilla-P. troglodytes* comparisons and showed that the former is significantly larger than the latter. This may suggest that Sibley and Ahlquist's (1984) DNA-DNA hybridization data are not strictly rate constant. However, this result alone should not, despite Templeton's (1985) suggestion to the contrary, be taken as the evidence of the existence of inconsistency in Sibley and Ahlquist's data.

At any rate, if we can apply this t-test to DNA-DNA hybridization data, why is it necessary to use Templeton's ΔQ-test, which has almost no statistical power?

LITERATURE CITED

PIELOU, E. C. 1979. Interpretation of paleoecological similarity matrices. Paleobiology 5:435–443.

SIBLEY, C. G., and J. E. AHLQUIST. 1984. The phylogeny of the hominoid primates. J. Mol. Evol. 20:2–15.

TEMPLETON, A. R. 1983. Phylogenetic inference from restriction endonuclease cleavage site map with particular reference to the evolution of humans and the apes. Evolution 37:221–244.

————. 1985. The phylogeny of the hominoid primates: a statistical analysis of the DNA-DNA hybridization data. Mol. Biol. Evol. 2:420–433.