The Phylogeny of the Hominoid Primates: A Statistical Analysis of the DNA-DNA Hybridization Data[1]

Alan R. Templeton
Department of Biology, Washington University

Sibley and Ahlquist compared the single-copy nuclear DNA sequences of the hominoid primates using DNA-DNA hybridization. From this data set they estimated a phylogeny that clusters man and chimpanzees using a distance Wagner procedure. However, no assessment of statistical confidence in this estimated phylogeny was made, despite the fact that their data set contains internal inconsistencies concerning the correct branching order. This paper presents a modification of Pielou's Q-statistic that allows one to make nonparametric tests of phylogenetic relationship from distance data. The results of this analysis indicate that the estimated phylogeny of Sibley and Ahlquist is without statistical significance owing to the internal inconsistencies of the data set. A survey and additional analyses of other types of molecular data indicate that the phylogeny that clusters chimpanzees and gorillas and has the human lineage splitting off earlier is statistically consistent with all the molecular data (including the DNA-DNA hybridization data), whereas the phylogeny estimated by Sibley and Ahlquist can be rejected at the 5% level using the data on restriction-endonuclease sites in the mitochondrial genome.

Introduction

Sibley and Ahlquist (1984) have recently estimated a branching order for the hominoid primates that clusters chimpanzees with humans (phylogeny 2 in fig. 1) on the basis of DNA-DNA hybridization data. They further conclude that their estimate indicates that all alternative phylogenies for these species are incorrect. In contrast, Templeton (1983a, 1983b) concluded that the branching pattern that clusters chimpanzees and gorillas together (phylogeny 1 of fig. 1) was significantly better (at least at the 5% level of significance) than any alternative branching pattern, including phylogeny 2.

At first glance, then, there appears to be a direct contradiction between the conclusions of Templeton (1983a, 1983b) and of Sibley and Ahlquist (1984). However, only Templeton (1983a, 1983b) presents an actual statistical comparison of alternative branching patterns, so before concluding that a contradiction exists, it is necessary to subject the data of Sibley and Ahlquist to a comparable statistical analysis. Such an analysis is presented in this paper.

Estimation versus Hypothesis Testing

Before analyzing the data set of Sibley and Ahlquist, it will be useful to point out the distinction between estimation and hypothesis testing. The molecular evolution literature contains a very large number of phylogeny estimation algorithms—maximum parsimony, UPGMA, distance Wagner, and compatibility, to name but a few. These estimation algorithms often allow one to generate an estimate of a phylogenetic tree
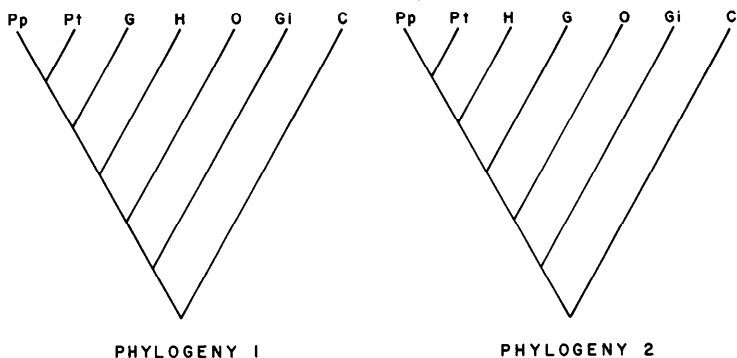

FIG. 1.—Two alternative branching orders for the hominoid primates. Pp = pygmy chimpanzee (*Pan paniscus*), Pt = common chimpanzee (*Pan troglodytes*), G = gorilla (*Gorilla gorilla*), H = man (*Homo sapiens*), O = orangutan (*Pongo pygmaeus*), Gi = gibbon (*Hylobates lar*), and C = cercopithecidae monkeys.

from a particular data set. This tree gives the estimated branching order of the species or populations being studied and often gives the estimated branch lengths as well. These algorithms do *not* allow the investigator to (1) make an assessment of how much confidence to place in the estimate and/or (2) reject alternative hypotheses (Felsenstein 1983). Questions of confidence in the estimate require the development of a hypothesis-testing framework that can ascribe probabilities to possible observable outcomes given a specific hypothesis or set of hypotheses.

Although Sibley and Ahlquist estimated phylogeny 2 by applying a distance Wagner procedure to their data set, they did not present any indication of how well or poorly their data set fits alternative phylogenies. The closest Sibley and Ahlquist (1984) come to testing the branching order is with a *t*-test on the delta $T_{50}H$ values between humans and chimpanzees versus that between humans/chimpanzees and gorillas. The *t*-tests reveal that humans and chimpanzees have a significantly (at the 0.1% level) smaller delta $T_{50}H$ value than either does vis à vis the gorilla. Sibley and Ahlquist state that this result also means that *Homo* and *Pan* are more closely related to one another than either is to *Gorilla;* that is, phylogeny 1 is true, and phylogeny 2 is rejected. However, before this conclusion can be made, two issues must be addressed. First, does the null hypothesis being tested by the *t*-statistic actually correspond to a biological hypothesis about branching order? Second, given that branching order is being tested, are *t*-tests of pairwise comparisons of delta $T_{50}H$ values the appropriate test statistic?

First, consider the biological meaning of the null hypothesis being tested by the *t*-statistic. Sarich and Wilson (1967) long ago proposed the relative-rate test of the hypothesis that all lineages evolve at the same rate in absolute time. If phylogeny 1 were true, one potential relative-rate test would compare the delta $T_{50}H$ values of the human-chimpanzee contrast to the corresponding value between either of these two species and gorillas. This is *exactly* the same *t*-test as performed by Sibley and Ahlquist (1984). Hence, given the topology, the *t*-test of Sibley and Ahlquist is a relative-rate test. Given rate constancy in all the lineages, the *t*-test of Sibley and Ahlquist is a test of branching order. Obviously, these two biological hypotheses are confounded in the null hypothesis of the *t*-statistic. Thus, quite apart from the issue of statistical significance of this test, there is considerable ambiguity as to its biological significance.

As shown above, the *t*-test is potentially a test of branching order only if there is rate constancy in *all* the lineages being compared. However, even given rate constancy,

are pairwise t-tests an appropriate test of branching order? The first problem concerns what data should be regarded as a single sample for the t-tests. Sibley and Ahlquist (1984) report $N = 19$ for their "*Homo* × *Pan*" comparisons by pooling the distance data of the *P. paniscus* vs. *Homo* ($N = 7$) with that of the *P. troglodytes* vs. *Homo* comparisons ($N = 12$). Because there is no dispute that the two chimpanzee species share a more recent common ancestor than either does with *Homo*, any evolution that occurs along the segment leading to this ancestor will be reflected in the distances from humans to both chimpanzees. Obviously, these two sets of comparisons are not independent, making it difficult to calculate the degrees of freedom. Even worse, Sibley and Ahlquist (1984) report that $N = 26$ for their "*Gorilla* × *Pan, Homo*" comparisons. Obviously, this sample size was obtained by pooling the *Gorilla–P. paniscus* ($N = 6$), *Gorilla–P. troglodytes* ($N = 10$), and *Gorilla–Homo* ($N = 10$) contrasts. Not only does the issue of nonindependence arise, but treating *Homo* and *Pan* as homogeneous entities for comparison with *Gorilla* presupposes that phylogeny 2 is true; yet, this is precisely the phylogeny that is supposed to be tested by this t-statistic.

A second basic problem with this testing procedure is that altering the position of one species in the phylogeny simultaneously alters the expected values for several pairwise comparisons rather than just that for one. Ignoring for the sake of argument the problems of the molecular clock and pooling, we note that Sibley and Ahlquist (1984) report a t-test that implies that phylogeny 2 is correct and that phylogeny 1 is incorrect. However, another pairwise comparison that tests these phylogenies under the rate-constancy hypothesis is that between the gorilla-chimpanzee and the gorilla-human delta $T_{50}H$ values. If phylogeny 1 is correct, gorillas and chimpanzees should be significantly closer than gorillas and humans, but if phylogeny 2 is true, there should be no significant difference. This expectation was tested using the data from Sibley and Ahlquist's table 1 on the *Gorilla–Pan troglodytes* and *Gorilla–Homo* comparisons. The resulting t-value is -2.78 with 18 degrees of freedom (significant at the 1% level), thereby indicating that phylogeny 1 is correct and that phylogeny 2 is incorrect.

The above test reveals that the data set of Sibley and Ahlquist (1984) is internally inconsistent with respect to the branching order of these primates. The existence of such inconsistencies raises the question of how one should deal with them. The simple-minded solution of doing all pairwise comparisons or a standard ANOVA to determine the general pattern will unfortunately not yield statistically justifiable conclusions. The reason is that these t-tests are not independent of one another, so there is no straight-forward way of combining them into a single probabilistic statement. Moreover, such pairwise comparisons still suffer from the confoundment between branching-order hypotheses and the rate-constancy hypothesis.

Clearly, a test statistic is needed that is robust to deviations in rate constancy, so that its null hypothesis corresponds primarily to a branching-order hypothesis. More-over, this test statistic should take into account all the informative distance comparisons in the data set of Sibley and Ahlquist (1984) and not just a biased portion. Such a test is given in the next section.

The Q-Test

Pielou (1979, 1983) has considered the problem of interpretation of paleoecological similarity matrices from a statistical point of view. The basic data set of Sibley and Ahlquist (1984) consists of a distance matrix giving the delta $T_{50}H$ values for the various primate comparisons (table 1). It is quite trivial to extend Pielou's basic testing procedure from similarity matrices to distance matrices.

Table 1
Matrix of Average Delta $T_{50}H$ Distance Values for Each Pairwise Set of Comparisons (from Sibley and Ahlquist 1984)

	Pan paniscus	Homo sapiens	Gorilla gorilla	Pongo pygmaeus	Hylobates lar	Cerco- pithecids
Pan troglodytes	0.7	1.8	2.1	3.7	5.1	7.7
P. paniscus		1.9	2.3	3.7	5.6	8.0
H. sapiens			2.4	3.6	5.2	7.7
G. gorilla				3.8	5.4	7.5
P. pygmaeus					5.1	7.6
H. lar						7.4

Pielou's testing procedure works best for null hypotheses that result in a strictly increasing (or decreasing) order of the distance (similarity) values contained in the matrix. This represents a severe limitation to the use of her statistic for phylogeny testing because many phylogenies do not yield a strictly increasing ordering of the species' expected distances. Fortunately, however, the two phylogenies under consideration here do satisfy this restriction. For example, consider the order *Pan paniscus* (hereinafter abbreviated by Pp), *Pan troglodytes* (Pt), *Gorilla gorilla* (G), *Homo sapiens* (H), *Pongo pygmaeus* (O), *Hylobates lar* (Gi), and Cercopithecidae (C). Under phylogeny 1, this order should ideally define a strictly increasing series of distance relationships; that is, if one chooses any species in this ordering and then progressively looks at the species to the right of the chosen species, the evolutionary distances would all be expected to increase. Similarly, phylogeny 2 defines the order: Pp, Pt, H, G, O, Gi, and C. Hence, the first step in testing these hypotheses is to arrange the delta $T_{50}H$ matrix according to these orders. For example, table 1 reproduces the data of Sibley and Ahlquist in the form of an upper-triangular distance matrix that is ordered under phylogeny 2.

Next, a Q-statistic is calculated that represents a nonparametric measurement of how well the hypothesized ordering of species actually orders the distance values within the matrix. Pielou's original Q-statistic was designed to test hypotheses about community structure, and the matrices were ordered by diagonals. However, phylogenetic hypotheses should order distance matrices across columns rather than across diagonals (J. Felsenstein, personal communication). For example, in table 1 the distances in a column are very similar, whereas the distances between columns increase as one goes from left to right. This is the expected pattern in a distance matrix if the hypothesized species ordering represents the true phylogeny.

The degree of order in the distance matrix is measured by the following modification of Pielou's Q-statistic. Suppose that n species or taxa are included in the study. All pairwise distances between these species can be represented by an upper-triangular matrix with $n - 1$ columns (e.g., table 1). Let the columns be numbered 1 through $n - 1$ going from left to right, with the first column having one element, the second two, and the rth column having r elements. Consider the tth element in the rth column. Suppose its numerical value is x. Define q_{rt} as the number of elements in columns $(r + 1)$, $(r + 2)$, ... $(n - 1)$ with values greater than x. Define Q as the sum

$$Q = \sum_{r=1}^{n-2} \sum_{t=1}^{r} q_{rt}.$$

Tied Pairs Each Contribute One-Half to the Sum Q

Q measures the degree to which the distance matrix is ordered by a particular series of species or taxa. If the series of species corresponds exactly to one of increasing phylogenetic divergence, all the elements in columns $r + 1$ through $n - 1$ should ideally be larger than any of the elements in column r, and Q will take on its maximal value. If the hypothesized series does not correspond to one of increasing phylogenetic divergence and/or if the data set contains little or no phylogenetic information, many of the elements in columns $r + 1$ through $n - 1$ could be smaller than some of the elements in column r. Hence, the larger the value of Q, the better the fit of the data to a particular ordering of the species.

It is also important to note that the value of Q is determined by the ranks of the distance values and not by their absolute magnitudes. Hence, the Q-statistic will measure phylogenetic relationship as long as the distance measures tend to increase with time, such that more recently diverged species have smaller distances than more remotely diverged species. On the other hand, to test phylogenetic relationship, the previously mentioned t-statistics not only require the distance measures to increase with time but also to increase at the same rate and with the same functional relationship in all lineages. Extreme deviations from rate constancy can obviously undermine the expected rank relationships measured by the Q-statistic, but the Q-statistic is much more robust to deviations from rate constancy than are the previously discussed t-tests. This focuses more of the Q-statistic's power on the hypothesized branching order than on branch lengths. In addition, the Q-statistic measures how a proposed branching order influences the ordering of the entire data matrix. Therefore, the Q-statistic satisfies the requirements for a test statistic that were mentioned in the previous section.

There are two sorts of stochastic error that can potentially affect the Q-value. The first is simply measurement error. This refers to the experimental error inherent in measuring the delta $T_{50}H$ values. This type of error can be controlled by rigorously controlling the measurement conditions and by doing sufficient replication. Sibley and Ahlquist (1984) argue that they have greatly minimized this error by their measurement conditions and by averaging a large number of replicates. As will be mentioned in the Discussion, serious questions still remain concerning the nature of their replications, but for now Sibley and Ahlquist's argument that measurement error has been virtually eliminated through averaging will be accepted. Hence, the Q-statistic will only be calculated using the average delta $T_{50}H$ values. It is important to note that because the Q-statistic depends only on ranks, the effective elimination of measurement error affecting ranks is far more readily accomplished than that affecting distance magnitudes.

However, the second source of stochastic variation cannot be eliminated or even reduced by the experimenter. Evolution itself is a stochastic process, and, accordingly, variations from expected distance values and rankings can be expected even if the distances are measured with 100% accuracy (see Ewens [1983] for an excellent discussion of these two types of stochastic error as they affect statistics relating to molecular evolution). Even if the relative rankings of the delta $T_{50}H$ values averaged over several replicates are regarded as accurately estimated, the evolutionary error still persists. Hence, statistical evaluation of the average distance values is still needed even if one feels that measurement error has been completely eliminated by means of both careful measurement conditions and the averaging of numerous replications.

In order to make probability statements about the value of Q, Pielou (1979) obtained the distribution of Q through computer simulation of matrices with random

distance elements and produced a table of significance values. Noting that large values rather than small values are in the significance region because the data matrix is a distance rather than a similarity matrix, we see that the simulation results of Pielou (1979) indicate that phylogeny 2 results in a Q-value of 175 that is significant at the 1% level. In other words, significant phylogenetic information is imbedded in the order favored by Sibley and Ahlquist (1984).

One can repeat this procedure for the species ordering implied by phylogeny 1, and the resulting Q-value is 171, a value also significant at the 1% level. The significance of other orderings can be evaluated as well. For example, consider exchanging the position of man with orangutans in phylogeny 1 to obtain a phylogeny in which the human lineage splits off before orangutans and the African apes diverge from one another (this is phylogeny 5 of Templeton [1983a]). This third ordering of the species results in a Q-value of 159, which is significant at the 5% level. As can be seen from these examples, the Q-test does not tell one that a particular species ordering corresponds to the true phylogenetic ordering but merely that a particular ordering is better than a random ordering. The three specific orderings considered here all share the relative position of the two chimpanzee species to one another and the relative positions of gibbons and monkeys to the other hominoid primates. The fact that one can permute the relative placement of humans, gorillas, and orangutans and still get significant Q-values implies that considerable phylogenetic information resides within the nonpermuted subset of the original data matrix.

The question of whether phylogeny 2 is better than phylogeny 1 remains to be addressed. Both phylogenies produced ordered distance matrices that are significantly better than random ordering. However, these two phylogenies only differ in the placement of two species, so much of the phylogenetic information contained in these two orderings is potentially contained in the shared portions—and is thereby irrelevant to the relative merits of phylogeny 1 versus phylogeny 2. What is needed is a test statistic that ignores the shared phylogenetic expectations and focuses only on the distance information that can distinguish between phylogenies 1 and 2 (or any other alternative). Such a test is proposed in the next section.

The Delta Q-Test

In order to compare two alternative orderings, the Q-statistic can still be used, because, as pointed out by Pielou (1983), the magnitude of Q measures how well the hypothesized branching pattern actually orders the elements of the data matrix. Basically, the larger the Q-value, the better the ordering. To test two alternative orderings, say a and b, the first step is to calculate Q under these orderings, yielding Q(a) and Q(b). Let the test statistic that directly compares the two alternative orderings be delta Q(a, b) = Q(a) − Q(b). For example, letting a = phylogeny 2 and b = phylogeny 1, delta Q(2, 1) = 175 − 171 = 4. This implies that the DNA-DNA hybridization data of Sibley and Ahlquist (1984) do indeed favor phylogeny 2 over phylogeny 1. Moreover, it is important to note that the delta Q-value depends only on the q_{rt} values from the two columns that were permuted in going from phylogeny 2 to phylogeny 1. All the q_{rt} values from the shared portions of the two phylogenies are canceled out during the subtraction of the two Q-values. Hence, the value of delta Q in this case is determined only by six distance contrasts involving five distance values: the distances between man and the two chimpanzee species, the distances between gorillas and the two chimpanzee species, and the distance between man and gorillas. All other distance

values are irrelevant, as they ought to be since they have identical evolutionary expectations regardless of which phylogeny is true. Hence, the delta Q-statistic eliminates all the shared phylogenetic distances and utilizes only the distances that can potentially distinguish between the two alternative phylogenies. However, unlike the t-test used by Sibley and Ahlquist (1984), the delta Q-statistic is a function of *all* the potentially informative distance contrasts. Hence, the delta Q-statistic has the desired statistical properties of being a function of all of the informative distance ranks in the data matrix and of being invariant to the ranks relating to the shared portions of the phylogenies being contrasted. All that remains is to evaluate the statistical significance of the delta Q-value.

As mentioned above, the delta Q-value for the contrast between phylogenies 1 and 2 depends on five informative distances. Under the null hypothesis in which it is assumed that there is no phylogenetic information in these five distances, all delta Q-values associated with the 120 (5!) permutations of the five distances should be equally probable. For any given permutation, a Q-value was first calculated and then the distances associated with the third and fourth taxa were switched (the same type of difference that exists between phylogenies 1 and 2) and a second Q-value was calculated. Delta Q is simply the difference between the first and second Q-values. This procedure was repeated over all possible permutations, and table 2 gives the exact probability distribution of the delta Q-statistic for switching taxa 3 and 4 in an ordered distance matrix. Since the Q-value of phylogeny 2 is maximal, it seems appropriate to treat this as a one-tailed test, which will strengthen the rationale for rejecting phylogeny 1. The probability of obtaining a delta Q-value of 4 or larger in a one-tailed test is 0.16 (from table 2). Hence, the data of Sibley and Ahlquist (1984) do not support the rejection of phylogeny 1 in favor of phylogeny 2 at the 5% level of significance.

Table 2 also shows that if a delta Q-value of 6 had been obtained, phylogeny 1 could have been rejected at the 5% level in favor of phylogeny 2. Consequently, the failure of Sibley and Ahlquist's data set to resolve the problem of the relative merit of phylogenies 1 and 2 can be attributed to the data set's internal inconsistencies that were noted earlier, and it is *not* due to the fact that the delta Q-test lacks sufficient power to ever make such a resolution.

As another example, consider testing phylogeny 1 against phylogeny 5 of Templeton (1983a), the phylogeny that switches the positions of man and orangutans relative to phylogeny 1. In this case delta $Q(1, 5) = 171 - 159 = 12$. Here there are 5,040 (7!) permutations of seven informative distances when taxa 4 and 5 are switched in an ordered distance matrix. Table 2 also gives the exact distribution in this case, and the one-tailed probability of obtaining a value of 12 (the maximum possible) is 0.0045. Hence, one can safely conclude that the data set of Sibley and Ahlquist (1984) allows one to reject phylogeny 5 of Templeton (1983a) in favor of phylogeny 1 (as is also true for phylogeny 2).

This strong rejection of phylogeny 5 illustrates another statistical attribute of the delta Q-statistic; namely, that the statistical power increases with an increase in the number of informative distances. For an upper-triangular matrix, switching the fourth and fifth taxa involves many more elements than switching the third and fourth, and accordingly the range of possible delta Q-values in much larger. Moreover, a larger percentage of the possible delta Q-values could yield statistically significant results at the 5% level. This increase in power is a natural consequence of the number of potentially informative contrasts. Thus, in trying to resolve the relative phylogenetic

Table 2
Exact Probabilities of the Delta Q-Statistic under the Null Hypothesis
That the Distance Ranks Contain No Phylogenetic Information

	SPECIES INTERCHANGED		
DELTA Q-VALUE	Second and Third	Third and Fourth	Fourth and Fifth
−12	. . .	. . .	0.0071
−11	. . .	. . .	0.0143
−10	. . .	. . .	0.0143
−9	. . .	. . .	0.0286
−8	. . .	. . .	0.0214
−7	. . .	. . .	0.0286
−6	. . .	0.0333	0.0429
−5	. . .	0.0667	0.0571
−4	. . .	0.0667	0.0500
−3	. . .	0.0667	0.0714
−2	0.1667	0.1000	0.0571
−1	0.3333	0.0667	0.0857
0	0.0000	0.2000	0.0429
1	0.3333	0.0667	0.0857
2	0.1667	0.1000	0.0571
3	. . .	0.0667	0.0714
4	. . .	0.0667	0.0500
5	. . .	0.0667	0.0571
6	. . .	0.0333	0.0429
7	. . .	. . .	0.0286
8	. . .	. . .	0.0214
9	. . .	. . .	0.0286
10	. . .	. . .	0.0143
11	. . .	. . .	0.0143
12	. . .	. . .	0.0071

NOTE.—The probability distribution is given for discriminating between two phylogenies in which the phylogenetic positions of either the second and third, third and fourth, or fourth and fifth species are interchanged in an ordered distance matrix.

relationships of man to the African apes, only a small subset of the total data matrix (five distances in this case) contains any relevant information. In contrast, much more of the data matrix is informative about the relative placement of the orangutan. It is therefore natural that the power of the delta Q-test should increase as increasing amounts of the data set become informative.

This statistical analysis of the DNA-DNA hybridization data set is completely compatible with the previous statistical analyses of Templeton (1983a, 1983b) on the mtDNA restriction maps of Ferris et al. (1981), the globin restriction maps of Zimmer (1981), and the partial mtDNA nucleotide-sequence data of Brown et al. (1982). In no case do the analyses of any of these molecular data sets yield statistically significant rejections of phylogeny 1. Hence, the rejection of phylogeny 1 on the basis of DNA-DNA hybridization data has yet to be demonstrated, despite the claims of Sibley and Ahlquist (1984).

The Implications of Rate Inconstancy

Under phylogeny 1, the *t*-tests of Sibley and Ahlquist (1984) imply that the human lineage has a slower rate of molecular evolution than that of the African apes, and especially that of the gorilla. This result is consistent with tests of the molecular clock performed with the mtDNA restriction-site data (Templeton 1983*a*) and with the nuclear DNA sequence data on globin pseudogenes (Templeton 1985). In addition, Marks (1982) concluded that man has a slower rate of karyotypic evolution than the African apes, and Stanyon and Chiarelli (1983) concluded that man has a slower rate of karyotypic evolution than gibbons. This pattern raises an interesting question— What factor or factors could cause such a slowdown in the human lineage, one that would simultaneously affect both nuclear and mitochondrial genome evolution?

The neutral theories of Ohta (1976) and Kimura (1979) provide a straightforward answer to this question. Ohta (1976) and Kimura (1979) have shown that under a model of neutrality that incorporates some small selective differences, the rate of evolution is a decreasing function of population size and of generation time. One of the important features of human evolution has been a dramatic slowdown in rate of development and an attendant increase in generation time over the past 3 Myr (Cutler 1975; Lovejoy 1981). In addition, the human lineage greatly expanded in size, starting at the very least with *Homo erectus,* which had a fossil distribution covering the entire Old World and not just Africa (Lovejoy 1981). The simultaneous increase in both generation time and population size would result in a dramatic slowdown in human evolutionary rates under Ohta's (1976) or Kimura's (1979) theories.

This slowdown, when coupled with the branching order given by phylogeny 1, offers a simple explanation for the discrepancies observed in *distance* data sets such as Sibley and Ahlquist's. Under phylogeny 1, the actual time separating man from the African apes is greater than that separating the gorilla and chimpanzee, but with a slowdown in the human lineage, the genetic distance between man and the African apes would be more similar to the distance between chimpanzees and gorillas than would be expected under a uniform-rate hypothesis. If the expectations for all these distances were roughly comparable as a result of the slowdown in the human evolutionary rate, the rankings of the observed distances would be greatly influenced by the stochastic variation inherent in the evolutionary process. Therefore, there is a high probability that distance data sets will show internal inconsistencies of the sort seen in Sibley and Ahlquist's data. On the other hand, suppose phylogeny 2 represents the correct branching order. In that case, the internal inconsistencies present in the data set imply that chimpanzees experienced the slowdown in the rate of molecular evolution. Since chimpanzees have a shorter generation time than man and much smaller population sizes, there is no straightforward explanation for such a slowdown. Indeed, under phylogeny 2, the theories of Ohta (1976) or Kimura (1979) would predict that the gorilla-chimpanzee distances should be greater than the gorilla-human distances owing to the expected slowdown in the one lineage showing an expanding population size and an increasing generation time. This is just the opposite of what is observed.

Discussion

As pointed out earlier, all the molecular evidence analyzed to date is consistent with phylogeny 1 and none is inconsistent at a statistically significant level. However, the data of Sibley and Ahlquist (1984) do favor phylogeny 2 over phylogeny 1 (although

not strongly), so a closer examination of the molecular and nonmolecular evidence that could potentially distinguish between these phylogenies seems warranted.

First, as pointed out by Sibley and Ahlquist (1984), phylogeny 1 is the pattern most often suggested by morphological studies (e.g., see Oxnard 1981; Martin 1985). Moreover, recent studies using comparative electromyography on muscle usage during locomotion (Stern and Susman 1981; Vangor and Wells 1983), comparative joint-motion studies (Prost 1980), allometric studies on both recent and fossil hominoids (Aiello 1981), the fossil finds that bipedality is very old in the hominids (Johanson and White 1979; Lovejoy 1981; Charteris et al. 1982), and an analysis of the type of vertebral pathologies found in the oldest hominid fossils (Cook et al. 1983) all suggest that knuckle-walking was never a primitive condition in the hominids. It is very difficult to maintain that knuckle-walking is not primitive under phylogeny 2 because to do so implies that knuckle-walking, which represents a complex suite of traits, must have evolved independently in gorillas and chimpanzees. However, this interpretation fits well with phylogeny 1, since knuckle-walking could have evolved in the common ancestor of chimpanzee and gorillas after the hominid lineage had split off (Templeton 1983a).

The strongest evidence supporting phylogeny 1 with statistical backing is the analysis of the mitochondrial DNA restriction maps (Templeton 1983a, 1985). Because this is a character-state data set, it has several advantages over a distance data set. First, going from character states to distances can result in a net loss of phylogenetic information (Penny 1982). Second, the distance Wagner procedure (the algorithm used by Sibley and Ahlquist) can have the undesirable property of yielding an estimated phylogeny that does not allow sufficient evolution to minimally explain the evolution of the character states (Fitch and Smith 1982). Third, with restriction-site or nucleotide-sequence data, the type of mutation (point substitution, deletion, insertion, etc.) can usually be inferred. As a result, detailed models can be constructed for the way character states evolve that allow phylogenetic information to be extracted not only from the number of mutational events that occurred but also from what types of events occurred (Templeton 1983b). In addition, one can examine the DNA molecule for mutational "hot spots" that can greatly contribute to distance differences but that can also have the highest probability of yielding phylogenetically misinformative results (Templeton 1985). When the mtDNA restriction-site data are corrected for mutational hot spots (Templeton 1985), phylogeny 1 is favored even more strongly than that indicated by the original statistical analysis given in Templeton (1983a). In contrast, with DNA-DNA hybridization data, one has no handle on what types of mutational events are contributing to the observed distances. This makes phylogenetic inference from such data sets precarious, since it is known that mutational events can create extremely large and phylogenetically misleading distances as detected by single-copy DNA-DNA hybridization over short time intervals (Hake and Walbot 1980; Zwiebel et al. 1982).

The support for phylogeny 1 is not limited to the mitochondrial genome but comes from some nuclear genetic studies as well. S. O'Brien (personal communication) surveyed man and the African apes for 46 isozyme loci. This data set represents a substantial increase in the number of loci sampled in previously published isozyme surveys (Bruce and Ayala 1979; Lucotte and Lefebvre 1981; Nozawa et al. 1982). Unfortunately, O'Brien (personal communication) only examined one chimpanzee species, so there are only three informative distance contrasts. Contrasting phylogenies 2 and 1 is equivalent to switching taxa 2 and 3 in an ordered distance matrix, and the

range of delta Q-values is only -2 to $+2$ (table 2). The delta $Q(2, 1)$-value in this case is -2, indicating a perfect fit to phylogeny 1 with no internal inconsistencies. However, because of the small number of informative distances in this case, even this most extreme value of the delta Q-statistic is not significant at the 5% level (table 2). It is also possible to analyze isozyme data as character-state data and perform an analogue to the nonparametric test performed on restriction sites (Templeton 1985). This procedure requires an outgroup species, which unfortunately is not provided by O'Brien et al. (personal communication). Hence, although the isozyme data supports phylogeny 1 as strongly as possible, the result is not significant with the delta Q-test, and additional data are needed for more powerful statistical analyses.

There are several other data sets that are almost equally probable under phylogenies 1 and 2. First are the chromosomal data. Although Sibley and Ahlquist (1984) cite Yunis and Prakash (1982) to claim that the chromosomal data favor phylogeny 2, two other recent karyotypic analyses favor phylogeny 1 (Marks 1982; Stanyon and Chiarelli 1982). Templeton (1985) performed a statistical analysis of the pooled (and partially overlapping) karyotypic data from all three studies and concluded that six of 10 phylogenetically informative karyotypic changes favor phylogeny 1 over phylogeny 2. A second nearly neutral data set is the partial mitochondrial DNA nucleotide-sequence data of Brown et al. (1982). As shown by Templeton (1983b), this data set results in almost equally likely probability values under either phylogeny, with 12 nucleotide positions favoring phylogeny 1 over phylogeny 2 and 10 positions favoring phylogeny 2 over phylogeny 1. As a third example of a nearly neutral data set, the hemoglobin–amino acid sequence data favor phylogeny 2, but the human-chimpanzee clustering is defined by only one amino acid substitution (Goodman et al. 1983). Such a result is without statistical significance (Templeton 1985). Slightly stronger support stems from restriction mapping of nuclear genes. In Zimmer's (1981) globin-DNA data, one structural change (insertions and deletions) favors phylogeny 2 over phylogeny 1 and one informative restriction site favors phylogeny 1 over phylogeny 2 (Templeton 1983a). Wilson et al. (1984) published some restriction maps of primate ribosomal DNA that had two informative sites (PVU II a and b in their maps), one of which favored phylogeny 1 over phylogeny 2 and the other of which favored phylogeny 2 over phylogeny 1. Wilson et al. (1984) also report one structural change that favors phylogeny 2 over phylogeny 1. None of these data sets is even close to having statistical significance (Templeton 1985), and even taken together these data sets are virtually neutral concerning the relative merits of these two phylogenies; in toto, 20 character-state changes favor phylogeny 1, and 18 favor phylogeny 2.

Previously published DNA-DNA hybridization analyses are also neutral regarding the relative merits of phylogenies 1 and 2. Both Hoyer et al. (1972) and Benveniste and Todaro (1976) concluded that their respective data sets could not resolve the phylogenetic relationships of man to African apes. Sibley and Ahlquist (1984) use "the advantage of hindsight" (i.e., the phylogeny estimated by Sibley and Ahlquist) to "trim" discrepant data from Hoyer et al. (1972) in order to claim that the Hoyer et al. (1972) data set supports phylogeny 2. Without performing any statistical analysis as a justification, Sibley and Ahlquist rationalized this trimming by claiming that the discrepant data were "probably caused by greater experimental error." Such trimming

of a data set—by using an estimate in order to claim that the data set supports the estimate—is statistically indefensible.

Recently, O'Brien et al. (1985) measured the distances between several primate species through DNA-DNA hybridization of single-copy DNA. Unlike Sibley and Ahlquist, O'Brien et al. (personal communication) obtained the smallest distances between gorillas and man rather than between chimpanzees and man. The delta Q-value is only 1 for the O'Brien et al. data set, indicating virtual neutrality on the merits of phylogeny 1 vs. phylogeny 2. There are three potential reasons why O'Brien et al. (personal communication) obtained different rankings than Sibley and Ahlquist (1984) did. First, there may be differences in laboratory techniques. Second, the differences may be simply due to measurement error. Third, the differences may be biologically real owing to intraspecific variation. None of the studies using DNA-DNA hybridization has attempted to document the amount of intraspecific variation found in the relevant primate species. Conclusions about the evolutionary significance of the interspecific differences are not valid unless these differences are clearly larger than the intraspecific differences. That intraspecific differences may be of a comparable order of magnitude are suggested by recent results of J. R. Powell (personal communication) that the delta $T_{50}H$ value (the same measurement used by Sibley and Ahlquist) between two strains of the subspecies *Drosophila mercatorum mercatorum* is 1.2. This value is greater than any of the delta $T_{50}H$ values reported by Sibley and Ahlquist between the two *Pan* species and is close to their lowest *Homo* $\times$ *Pan* measurement of 1.3. Hence, the results of Sibley and Ahlquist must be viewed as suspect until the issue of intraspecific variation has been resolved.

The strongest discrimination in favor of phylogeny 2 is provided by the DNA sequence data on globin pseudogenes (Goodman et al. 1984). Six of seven informative nucleotide sites favor phylogeny 2 over phylogeny 1, but the result is not significant at the 5% level (Templeton 1985).

At present, there is no analysis of any data set that rejects phylogeny 1 at the 5% level in favor of phylogeny 2, but phylogeny 2 can be rejected at the 5% level in favor of phylogeny 1 on the basis of the mitochondrial restriction-site data. Hence, at present, all the molecular data are statistically compatible with phylogeny 1 and only with phylogeny 1.

Acknowledgments

LITERATURE CITED

AIELLO, L. C. 1981. The allometry of primate body proportions. Symp. Zool. Soc. Lond. **48**: 331–358.

BENVENISTE, R. E., and G. J. TODARO. 1976. Evolution of type C viral genes: evidence for an Asian origin of man. Nature **261**:101–108.

BROWN, W. M., E. M. PRAGER, A. WANG, and A. C. WILSON. 1982. Mitochondrial DNA sequences of primates: tempo and mode of evolution. J. Mol. Evol. **18**:225–239.

BRUCE, E. J., and F. J. AYALA. 1979. Phylogenetic relationships between man and the apes: electrophoretic evidence. Evolution **33**:1040–1056.

CHARTERIS, J., J. C. WALL, and J. W. NOTTRODT. 1982. Pliocene hominid gait: new interpretations based on available footprint data from Laetoli. Am. J. Phys. Anthropol. **58**:133–144.

COOK, D. C., J. E. BUIKSTRA, C. J. DEROUSSEAU, and D. C. JOHANSON. 1983. Vertebral pathology in the Afar Australopithecines. Am. J. Phys. Anthropol. **60**:83–101.

CUTLER, R. G. 1975. Evolution of human longevity and genetic complexity governing aging rate. Proc. Natl. Acad. Sci. USA **72**:4664–4668.

EWENS, W. J. 1983. The role of models in the analysis of molecular genetic data, with particular reference to restriction fragment data. Pp. 45–73 in B. S. WEIR, ed. Statistical analysis of DNA sequence data. Marcel Dekker, New York.

FELSENSTEIN, J. 1983. Statistical inference of phylogenies. J. R. Stat. Soc. A **146**:246–272.

FERRIS, S. D., A. C. WILSON, and W. M. BROWN. 1981. Evolutionary tree for apes and humans based on cleavage maps of mitochondrial DNA. Proc. Natl. Acad. Sci. USA **78**:2432–2436.

FITCH, W. M., and T. F. SMITH. 1982. Implications of minimal length trees. Syst. Zool. **31**:68–75.

GOODMAN, M., G. BRAUNITZER, A. STAUGL, and B. SCHRANK. 1983. Evidence on human origins from haemoglobins of African apes. Nature **303**:546–548.

GOODMAN, M., B. F. KOOP, J. CZELUSNIAK, M. L. WEISS, and J. L. SLIGHTOM. 1984. The -globin gene: its long evolutionary history in the -globin gene family of mammals. J. Mol. Biol. **180**:803–823.

HAKE, S., and V. WALBOT. 1980. The genome of *Zea mays:* its organization and homology to related grasses. Chromosoma **79**:251–270.

HOYER, B. H., N. W. VAN DE VELDE, M. GOODMAN, and R. B. ROBERTS. 1972. Examination of hominoid evolution by DNA sequence homology. J. Hum. Evol. **1**:645–649.

JOHANSON, D. C., and T. D. WHITE. 1979. A systematic assessment of early African hominids. Science **203**:321–330.

KIMURA, M. 1979. Model of effectively neutral mutations in which selective constraint is incorporated. Proc. Natl. Acad. Sci. USA **76**:3440–3444.

LOVEJOY, O. 1981. The origin of man. Science **211**:341–350.

LUCOTTE, G., and J. LEFEBVRE. 1981. Distances électrophorètiques entre l'homme, le chimpanzé (*Pan troglodytes*) et le gorille (*Gorilla gorilla*) basées sur la mobilité des enzymes érythrocytaires. Hum. Genet. **57**:180–184.

MARKS, J. 1982. Evolutionary tempo and phylogenetic inference based on primate karyotypes. Cytogenet. Cell Genet. **34**:261–264.

MARTIN, L. 1985. Significance of enamel thickness in hominoid evolution. Nature **314**:260–263.

NOZAWA, K., T. SHOTAKE, Y. KAWAMOTO, and Y. TANABE. 1982. Electrophoretically estimated genetic distance and divergence time between chimpanzee and man. Primates **23**:432–443.

OHTA, T. 1976. Rate of very slightly deleterious mutations in molecular evolution and polymorphism. Theor. Popul. Biol. **10**:254–275.

OXNARD, O. E. 1981. The place of man among the primates: anatomical, molecular and morphometric evidence. Homo **32**:149–176.

PENNY, D. 1982. Towards a basis for classification: the incompleteness of distance measures, incompatibility analysis and phenetic classification. J. Theor. Biol. **96**:129–142.

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

———. 1983. Spatial and temporal change in biogeography: gradual or abrupt? Pp. 29–56 in R. W. SIMS, J. H. PRICE, and P. E. S. WHALLEY, eds. Evolution, time and space: the emergence of the biosphere. Academic Press, New York.

PROST, J. H. 1980. Origin of bipedalism. Am. J. Phys. Anthropol. **52**:175–189.

SARICH, V. M., and A. C. WILSON. 1967. Immunological time scale for hominid evolution. Science **158**:1200–1204.

SIBLEY, C. G., and J. E. AHLQUIST. 1984. The phylogeny of the hominoid primates, as indicated by DNA-DNA hybridization. J. Mol. Evol. **20**:2–15.

STANYON, R., and B. CHIARELLI. 1982. Phylogeny of the Hominoidea: the chromosome evidence. J. Hum. Evol. **11**:493–504.

———. 1983. Mode and tempo in primate chromosome evolution: implications for hylobatid phylogeny. J. Hum. Evol. **12**:305–315.

STERN, J. T., JR., and R. L. SUSMAN. 1981. Electromyography of the gluteal muscles in *Hylobates, Pongo,* and *Pan:* implications for the evolution of hominid bipedality. Am. J. Phys. Anthropol. **55**:153–166.

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

———. 1983*b*. Convergent evolution and nonparametric inferences from restriction data and DNA sequences. Pp. 151–179 *in* B. S. WEIR, ed. Statistical analysis of DNA sequence data. Marcel Dekker, New York.

———. 1985. Relation of humans to African apes: a statistical appraisal of diverse types of data. Accepted *in* S. KARLIN and E. NEVO, eds. Evolutionary processes and theory. Academic Press, New York.

VANGOR, A. K., and J. P. WELLS. 1983. Muscle recruitment and the evolution of bipedality: evidence from telemetered electromyography of spider, wooly and patas monkey. Ann. Sci. Nat. Zool. **5**:125–135.

WILSON, G. N., M. KNOLLER, L. L. SZURA, and R. D. SCHMICKEL. 1984. Individual and evolutionary variation of primate ribosomal DNA transcription initiation regions. Mol. Biol. Evol. **1**:221–237.

YUNIS, J. J., and O. PRAKASH. 1982. The origin of man: a chromosomal pictorial legacy. Science **215**:1525–1530.

ZIMMER, E. A. 1981. Evolution of globin genes. Ph.D. diss. University of California, Berkeley.

ZWIEBEL, L. J., V. H. COHN, D. R. WRIGHT, and G. P. MOORE. 1982. J. Mol. Evol. **19**:62–71.