Higher-Primate Phylogeny—Why Can't We Decide?[1]

Richard Holmquist, Michael M. Miyamoto,† and Morris Goodman‡*

*Space Sciences Laboratory, University of California, Berkeley; †Department of Zoology, University of Florida; and ‡Department of Anatomy and Cell Biology, Wayne State University

At present, no definitive agreement on either the correct branching order or differential rates of evolution among the higher primates exists, despite the accumulated integration of decades of morphological, immunological, protein and nucleic acid sequence data, and numerous reasonable theoretical models for the analysis, interpretation, and understanding of those data. Of the three distinct unrooted phylogenetic trees, that joining human with chimpanzee and the gorilla with the orangutan is currently favored, but the two alternatives that group humans with either gorillas or the orangutan rather than with chimpanzees also have support. This paper is a synthetic and critical review of the methodological literature and isolates some 20 specific reasons why uncertainty in the evolutionary understanding of our closest living relatives persists. Many of the difficulties are eliminated or ameliorated by Lake's new methods of phylogenetic invariants and operator metrics. In the companion paper these new methods are used to analyze both the nuclear and mitochondrial DNA of the higher primates.

Introduction

The branching order and rates of evolution of human, chimpanzee, gorilla, and orangutan are topics of much current interest. There are three different unrooted phylogenetic trees that are a priori possible: human and chimpanzee joining gorilla and orangutan, human and gorilla joining chimpanzee and orangutan, and human and orangutan joining chimpanzee and gorilla. These will be called *Homo/Pan, Homo/Gorilla,* and *Pan/Gorilla,* respectively.

We begin with a review of published studies—each favoring one or another of the above branching orders—with a deliberate focus on those aspects that prevented final resolution of the question. From this review we cull 21 of those properties of current models that were used to analyze the raw data and that encourage an incorrect—or prevent an unambiguous—interpretation with regard to branching order and branch lengths. We consider parsimony, distance-matrix, and Markov models, with or without maximum-likelihood estimation of parameters. Following this, we look at the manner in which the complexity of the data type can aid or mislead us with regard to branching order and branch lengths. Noting that Lakes's formulation (1987*a,* 1987*b*) of evolutionary parsimony, phylogenetic invariants, and operator metrics is free of the more

1. Key words: primates, evolution, branching order, rate of evolution, complexity, phylogenetic invariants, operator metrics, nuclear DNA, mitochondrial DNA, parsimony, distance matrix methods, Markov models, maximum likelihood.

Address for correspondence: Dr. Richard Holmquist, 760 Mesa Way, Richmond, California 94805. Address for reprints: Professor Morris Goodman, Department of Anatomy and Cell Biology, Wayne State University, 540 East Canfield Avenue, Detroit, Michigan 48201; or Professor Michael M. Miyamoto, Department of Zoology, University of Florida, Gainesville, Florida 32611.

Mol. Biol. Evol. 5(3):201–216. 1988.

0737-4038/88/0503-5301$02.00

serious defects of the above models, in the companion paper (Holmquist et al. 1988) we extend and apply Lake's methods to determine the branching order and rates of evolution among the higher primates.

Evidence for a *Pan/Gorilla* Clade

Morphological data (Delson et al. 1977; Oxnard 1981) have been used in support of this clade. On the basis of the fact that man has many morphological features shared by the orangutan and has few in common with the gorilla or chimpanzee, Schwartz (1984), too, preferred this grouping. Hasagawa et al. (1985, p. 171) offered alternative explanations of the morphological data: "Although brain capacity has increased very much along the human lineage, not only the orangutan but also the human may be living fossils with respect to various morphological features, whereas the chimpanzee and the gorilla may have specialized quickly after they diverged from the human line." They further point out the curious fact that no fossil assigned to be ancestral only to the chimpanzee or gorilla has yet been unearthed: "it is possible that some fossil hominoids that were ancestral to the chimpanzee or gorilla but not the human have been assigned to human ancestors because of some of their residual features."

In contrast to the above explanation of morphological change, humans are actually highly derived in at least certain aspects of their morphology (Gould 1977; Creel 1986). The evolutionary interpretations of many morphological differences among humans and African apes, such as bipedalism versus knuckle-walking, remain open to debate (Washburn 1982; Templeton 1983*b;* Pilbeam 1986).

Immunological (Dene et al. 1976) and chromosomal (Dutrillaux 1979, 1980; Marks 1982; Stanyon and Chiarelli 1982; Bianchi et al. 1985) data have also been cited in support of making chimpanzees and gorillas sister groups (see also Kluge 1983). Bianchi et al.'s (1985) restriction enzyme–induced chromosome-banding studies are particularly forceful: this grouping required a minimum of only 37 events to explain the banding pattern of 30 informative traits. To group humans either with chimpanzees or with gorillas required a quite larger number of events—56 and 57, respectively.

Hixson and Brown (1986), using mitochondrial DNA sequences from the 12S ribosomal RNA genes of man and the great apes, reported that the *Homo/Pan* and *Homo/Gorilla* arrangements were both most parsimonious, but only by one substitution over the third dichotomous possibility grouping *Pan* with *Gorilla.* However—and this may be the telling fact—this less parsimonious grouping was supported by a unique deletion, unlike the two more parsimonious arrangements.

Brown et al. (1982), in examining an 896-bp fragment of mitochondrial DNA, found, using hominoid outgroups, that the most parsimonious tree favored this branching order, but grouping *Homo* and *Pan* together or *Homo* and *Gorilla* together required only two and three more substitutions, respectively. The authors also examined the data by the distance-matrix method of Fitch and Margoliash (1967), with the same result. In a subsequent maximum-parsimony analysis of those data, in which mouse and ox mitochondrial sequences were used as outgroups of the hominoids, the *Pan/Gorilla* clade was again supported, but an equally parsimonious tree grouped *Pan* with *Homo* (Goodman et al. 1985).

Nei et al. (1985) reanalyzed the data of Brown et al. by the unweighted-pair-group method (Sneath and Sokal 1973) and computed a minimum estimate of the SEs of the branch points. This method assumed equal expected rates of evolution among lineages descendant from the same ancestral node. Contrary to the result found

by Brown et al., it grouped human and chimpanzee together, but the difference between their branch point and the earlier branch point of the gorilla was not statistically significant. Hasegawa and Yano (1984*a*, 1984*b*) and Bishop and Friday (1985, 1986) analyzed the Brown et al. fragment by a maximum-likelihood method and also concluded that the chimpanzee/human grouping was most likely. In the studies by Hasegawa and Yano, only transversion-type differences were considered.

Templeton (1983*a*, 1983*b*, 1987), using Wilcoxon matched-pairs signed-ranks tests, analyzed the data on restriction-endonuclease sites in mitochondria and rejected the *Homo/Pan* grouping at the 5% level of statistical significance in favor of the *Pan/Gorilla* grouping. But Nei and Tajima (1985, 1987) claim that Templeton's methodology forces the chimpanzee to always group with the gorilla, even if that were not the correct phylogeny. Hasegawa et al. (1985) noted that Templeton's data included many silent transitions and that, because the number of such transitions between human and mouse is about the same as that between human and chimpanzee, multiple substitutions at the same base site would have confounded Templeton's analysis. Brown et al. (1982, p. 236) early understood this potential for transitions to make life difficult: "The preponderance of transitions helps us to understand why it has been hard to establish the branching order for hominoid lineages by mtDNA comparisons. A high incidence of transitions inevitably produces parallel and back mutations at the same site among lineages."

Evidence for a *Homo/Pan* Clade

Ciochon (1985), in a review of living and fossil hominoids, favored this grouping, as did the review of fossil and molecular data by Ruvolo and Pilbeam (1986). Because gorillas and chimpanzees—but not humans—share thin tooth enamel, Pilbeam (1986) noted that acceptance of a *Homo/Pan* clade implies morphological convergences in hominoids.

Chromosomal evidence has also been used to support this grouping (Seuánez 1982; Yunis and Prakash 1982).

Goodman et al. (1983) favored this branching order on the basis of the identity of the alpha and beta hemoglobin sequences and on the basis of the fact that the gorilla alpha chain differs from both the chimpanzee and human alpha chains by a glutamic acid at amino acid residue 23. This glutamic acid is shared also with the orangutan, gibbons, and the cebids. As a cautionary note, however, it is worth remembering that the whale (Goldstone and Smith 1966) and camel (Sokolovsky and Moldovan 1972) cytochrome *c* sequences are also identical (Holmquist 1976).

Slightom et al. (1985), in a study of the gamma hemoglobin genes, found that the most parsimonious tree favored the *Homo/Pan* over both the *Pan/Gorilla* and the *Homo/Gorilla* groupings—but only by four and nine fewer substitutions, respectively. However, when the newly completed orangutan and spider monkey γ-hemoglobin gene sequences were added to the parsimony analysis, grouping *Pan* with *Homo* was no longer most parsimonious (Slightom et al. 1987).

The mitochondrial DNA sequence data of Hixson and Brown (1986), cited above in support of a *Pan/Gorilla* clade (on the basis of a unique deletion shared by the chimpanzee and gorilla), would, if the deletion were not interpreted as phylogenetically determinative, otherwise support, with equal force, a *Homo/Pan* or a *Homo/Gorilla* clade.

Sibley and Ahlquist (1984), in a distance-matrix study based on 183 DNA/DNA hybridization values, favored the human/chimpanzee grouping. Though the triangle

inequality was obeyed by all distance values, Templeton (1985) pointed out (1) that the validity of the statistic (Student's *t*-test) that Sibley and Ahlquist used to link *Homo* with *Pan* depended on an assumed rate constancy among all lineages and (2) that, if that assumption were true, there had to be internal inconsistencies in the data with respect to the branching order. Because of the high precision of Sibley and Ahlquist's data—which is now even greater (Sibley and Ahlquist 1987), the data base having been extended to 514 hybrids—these papers together establish that rate nonconstancy exists for single-copy DNA in these lineages.

On analyzing the Sibley and Ahlquist (1984) data with the nonparametric delta-Q statistic, an extension of Pielou's (1979, 1983) Q-statistic, Templeton (1985) found that the *Homo/Pan* grouping was not statistically significantly favored over the *Pan/Gorilla* grouping by the DNA/DNA hybridization data. The virtue of the delta-Q statistic is that it eliminates all shared phylogenetic distances (the "noise" that swamps the informative distances) and utilizes only and all the informative distances that can potentially distinguish between two alternative phylogenies. Saitou (1986) pointed out (1) that Templeton calculated the critical values of this statistic on the invalid assumption that no hierarchical structure exists among these taxa and (2) that the sampling error was sufficiently large for this statistic that it was inadequate for testing topological distances. Ruvolo and Smith (1986) then found that even under the most optimal circumstances the delta-Q test lacked the power to discriminate between the alternative phylogenies. Templeton (1986*a*), in reply to these criticisms, found that incorporating the hierarchical structure into his methodology gave the same result, and by changing his scoring method he was able to answer Ruvolo and Smith's criticism. In that same paper he showed how sensitively the phylogenetic conclusions rested on the choice of the distance Wagner method (Farris 1972 vs. Farris 1985) and even on apparently minor changes in the $\Delta T_{50}H$ DNA/DNA hybridization values. Fitch (1986) used the Mann-Whitney U-test on Sibley and Ahlquist's data and could not reject Templeton's conclusion that the hybridization distance between gorilla and chimpanzee was not statistically greater than that between gorilla and human. However, he also noted that the chimpanzee-human distance was significantly less than *both* the gorilla-chimpanzee and gorilla-human distances, thus supporting the human/chimpanzee clade favored by Sibley and Ahlquist.

Lanave et al. (1986), using a Markov model (Lanave et al. 1984; Preparata and Saccone 1987; Saccone et al. 1987) in which the rate of substitution was constant for a particular type of substitution (e.g., A→C) but was allowed to vary among types, examined the third position within fourfold-degenerate codons in a 720-bp mitochondrial gene fragment containing the ND4 and ND5 genes and found that the date of separation of the human and gorilla lineages was 2.8 ± 3.6 Myr prior to that of the human and chimpanzee. Again the overlap due to the large sampling error from the short sequence precluded a unique answer to the branching pattern. They also found that the ratio of transitions to transversions was different in different lineages. Hasegawa et al. (1985) extended this Markov model to allow for the fact that not all sites are variable (Fitch and Markowitz 1970; Holmquist et al. 1972; Karon 1979; Holmquist 1980; Holmquist and Pearl 1980; Holmquist et al. 1982, 1983) and estimated the model parameters by maximum likelihood (Felsenstein 1981*a*). Their conclusions regarding branching order supported those of Lanave et al. (1986), but again they were inconclusive.

Koop et al. (1986) examined 2,251 nucleotide positions of the η-globin gene and found two most parsimonious trees, one favoring the chimpanzee/human linkage and

the other favoring the chimpanzee/gorilla linkage; but only the former was supported by unique insertion-deletion events. They concluded that "within the Hominidae the relationships remain unclear" (Koop et al. 1986, p. 237).

In the most extensive molecular study to date, on a 7.1-kb region including the 5′ and 3′ noncoding flanking regions of the ψη-globin gene, Miyamoto et al. (1987) favored the human/chimpanzee grouping on the basis of the fact that humans and chimpanzee share eight derived features while the gorilla/human or gorilla/chimpanzee groupings share but two and three derived characters, respectively.

Evidence for a *Homo/Gorilla* Clade

As noted above, the mitochondria DNA sequences from the 12S ribosomal RNA genes of the great apes (Hixson and Brown 1986) could be used in support of this grouping, but not unambiguously so.

The fact that satellite DNA fractions II and IV, though present in humans, have not been found in chimpanzees or the orangutan (Jones et al. 1972; Godsen et al. 1977) could be interpreted as evidence for a *Homo/Gorilla* clade, though the absence of data for the gorilla makes such an interpretation tenuous.

Parsimony Methods

Reconstructions of phylogenetic history by the parsimony principle are based on variants of two simple rules: (1) global minimization of the total number of character state changes over the evolutionary history and (2) the "two-thirds" rule—i.e., when comparing three taxa, if at least two of these taxa share the same character state, those taxa are grouped together and that state is chosen to be the ancestral condition.

Rule (1) has the consequence that the evolutionary history in one branch of a tree can influence the reconstruction in another branch in a manner that violates temporal causality (Holmquist 1976). For example, a base substitution in a mammalian lineage would be introduced if it would reduce by two or more the number of substitutions needed to explain the data in a more ancient branch. The selective pressures during these two evolutionary epochs are clearly independent of this minimization process and are not related in a causal manner. Regardless of whether temporal causality is violated, Lake (1987*a*; p. 59) observed that "substitutions in one branch of the tree can alter the measured length of another."

Rule (2) has the consequence that spurious identities among homologous residue positions in proteins—or among homologous nucleotide sites in nucleic acids—are not rare. In proteins, for three taxa, King (1980) has shown that even for unrelated taxa, at least two of the taxa will share the same amino acid 16% of the time, purely by chance. For DNA we find that, for three taxa, such spurious identities occur ~62% of the time ($\Sigma[3p_i^2 - 2p_i^3]$, the four-element set $\{p_i\}$ being the base composition). Spurious identities are more common in DNA than in proteins because there are only four bases in DNA, whereas in proteins there are 20 amino acids. The magnitude of this sequence noise is large. It should alert those who assume that nucleic acids are automatically more informative than proteins, and it indicates that the interpretation of phylogenetic reconstructions made by using the parsimony method must adequately isolate these spurious identities.

This noise leads to errors in the ancestral sequences inferred by parsimony, the error increasing in a determinable manner with increasing distance from the present (Holmquist 1979), even when the correct topology is known beforehand. Because branch lengths are calculated from the number of differences between ancestral nodes—

or between an ancestral node and an extant sequence—these lengths (also called distances) too can be in error.

The phylogenetic branching order (topology) found by parsimony can err either when there are different rates of evolution in different lineages (Cavender 1978; Felsenstein 1978) or when there is heterogeneity in the rates of evolution of different characters (Felsenstein 1981*b*). For four taxa, when the probability of a nucleotide substitution is large for two of the lineages and small for the other two and for the central branch of the unrooted tree, the number of substitutions by which one tree must exceed another before the less parsimonious tree can be rejected is $\geq 3/16$ the total number of sites examined (Cavender 1981; Felsenstein 1983). For a nucleotide sequence 7,000 bases long, such as the ψη-globin gene region, this critical value is thus at least 1,312 substitutions (the exact number is 1,366 at the 5% significance level). Felsenstein comments that the size of the critical numbers is striking. Under some conditions, a stronger result is possible. For example, for a rooted three-species tree, the critical value can be much smaller: if a molecular clock holds—i.e., if the probability for a base substitution per unit time is the same along all lineages—then the critical value is at least one-third the number of those sites considered phylogenetically informative by the parsimony method (Felsenstein 1985). This results in a lower critical value because the sites considered phylogenetically informative by the parsimony method normally constitute a small subset of the total sites. In any case, when the most parsimonious tree is only slightly more parsimonious than other trees found in the search procedure, one is unjustified to conclude that any particular one of these trees is any more likely than another to be historically correct. As a referee of this paper pointed out, "the fact that one finds several not particularly closely 'related' trees with almost the same numbers of 'minimal steps' suggests that we are making incorrect inferences about nodal states with considerable regularity. We are simply going to have to face up to the fact that we have to allow for different possibilities for the unobservable nodal states. Lake's (1987*a*, 1987*b*) methods do that, as do the standard likelihood procedures."

In examining enzyme restriction-site data, Templeton (1983*b*) correctly concludes that "the topology of the tree inferred from maximum parsimony can be other than the topology of the true phylogenetic tree even as an infinite amount of data is accumulated." Penny and Hendy (1986, p. 227) brought home the difficulty of the problem when they estimated that, with (1) weighting for the quality of the data and (2) 11 mammalian taxa (for which there are 34,459,425 unrooted binary trees), more than 300 phylogenetically informative sites would be needed to find the historically true tree by parsimony methods. As we shall see in the accompanying paper (Holmquist et al. 1988), the frequency of occurrence of nucleotide patterns that allow one to distinguish between branching topologies is considerably rarer than one might at first believe.

Nei and Tajima (1985, p. 199) estimated that when the average number of nucleotide substitutions per site is >0.01 "the parsimony method can give an erroneous inference about the evolutionary change of restriction sites and that the estimate of the number of mutational changes for a given branch could be larger than the true number."

More disturbing is the fact that, even if absolute substitution rates are low, the relative lengths of individual branches can be over- or underestimated by an order of magnitude or more, provided only that relative substitution rates on adjacent branches differ significantly (Lake 1987*a*), the effect being greater when substitutions are irre-

versible. Further, for four taxa, parsimony can fail to converge to the true tree for essentially all choices for the lengths of the peripheral branches, provided that the central branch of the tree is sufficiently short (Lake 1987*b*).

Distance-Matrix Methods

Reconstruction of phylogenetic history by distance-matrix methods begins with a matrix, each element of which represents the pairwise distance between the two taxa identified by the row and column labels. For taxa A, B, and C and distances D_{AB}, D_{AC}, and D_{BC}, taxa A and B are grouped together if $D_{AB} < D_{AC}$ and $D_{AB} < D_{BC}$. For the unrooted tree the branch lengths to the ancestral node can be found from simple linear algebra: $L_A = \frac{1}{2}(D_{AB} + D_{AC} - D_{BC})$, for example. Although parsimony methods are superior to matrix methods when the number of substitutions per site is very small or the data set is reasonably dense (Goodman et al. 1978; Holmquist 1978), distance methods can, for nondense data sets, be superior to parsimony as the substitutions per site increase—and *provided* that the rate of nucleotide substitution remains more or less the same for all evolutionary lineages (Schwartz and Dayhoff 1978; Klotz and Blanken 1981).

There are many ways by which the same data set of nucleic acid or protein sequences can be analyzed to define pairwise distance values (Fitch and Margoliash 1967; Schwartz and Dayhoff 1978; Klotz and Blanken 1981; Wilbur 1985), and the quality of the phylogenetic reconstruction, with respect to both topology and branch lengths, depends on the reasonableness of these values (Sneath and Sokal 1973). Regardless of method, there is a large loss of information in converting sequence data to distance matrices (Penny 1982). The idealized goal has been to obtain a distance that is metrical in the mathematical sense, that is logically and biologically consistent with known evolutionary mechanisms and experimental constraints, and that is maximally useful for reconstructing the branching times, order of species divergence, and internodal distances of the tree.

Lake (1987*a*) points out that the above algebraic expressions for branch lengths (L_A, etc.) can distort those lengths by distributing substitutions that are of uncertain origins (as reflected in the pairwise distances) equally among all three branches, even though historically they may have been unequally distributed among the branches. Given this fact, which has been known for 2 decades, it is somewhat surprising that with the important exceptions of the work of Moore et al. (1976), Moore (1977), Fitch and Bruschi (1987), and Lake's own recent efforts cited above, all work has gone into finding more adequate distance values, and none at all into finding more realistic expressions for the branch lengths. Thus, though the relative rate test (Wilson et al. 1977) is useful for detecting rate inequalities among lineages, it does not always suffice to estimate these in an accurate quantitative manner and can, depending on the outgroup chosen, fail to detect such inequalities in rate entirely.

Heterogeneity in evolutionary rates among lineages—or among different characters—can obviously perturb the above inequalities so that distance-matrix methods will return an incorrect branching order, as well as incorrect branch lengths (see preceding paragraph).

The problem of noise due to spurious identities is also present in distance-matrix methods. Even were noise absent, Andrews (1987) noted that, for any distance statistic, the small number of derived characters uniquely shared between two closely related species can be swamped by the much larger number of shared primitive characters, unless the latter can be eliminated from consideration.

Rates of Evolution

One of the more insightful conclusions to come out of Koop et al.'s (1986) recent study of the η-globin gene region was: "Our analysis of the primate η-globin genealogy also raises questions concerning the validity of using global clocks to estimate divergence times. Because of the variability in rates of divergence in different lineages, we suggest that local clocks in some cases yield more accurate divergence times."

This admonition is underscored by Gingerich's (1986) finding that, for DNA hybridization data, immunological distances, amino acid sequence data, and nucleotide sequence data, none bore a linear relationship between time and molecular change. Somewhat surprisingly, the nucleotide sequence data were the most nonlinear of all.

Britten (1986) has noted that even for selectively neutral DNA evolutionary rates vary by as much as a factor of five.

Complexity of Data Type

In simplest form, the complexity of a data set can be defined as the amount of detail (information) it can provide about the evolutionary history of a set of organisms. It is inefficient and obscuring to include in an analysis more detail from the data than is necessary to answer the evolutionary questions posed. On the other hand, if an analysis uses a data set of insufficient complexity, it will be impossible to obtain unambiguous answers to those questions.

Sibley and Ahlquist (1987, p. 104) state that "the notion that protein sequences are more informative than the data of DNA hybridization is erroneous. . . . DNA hybridization data are enormously complex, indexing essentially the entire information content of the genome." We believe the truth or falsity of this statment depends on circumstances. Most of the information in the DNA of living organisms reflects quite distant evolutionary events and is irrelevant to the most recent speciation event. In proteins and nucleotide sequences, sites can be grouped together to form a hybrid character useful in drawing accurate phylogenetic conclusions. An illustration occurs in the mRNAs coding for myoglobin in the animal kingdom: consider the hybrid character defined by the nucleotide sextet {A, A, C, A, U, U}, where the letters represent, respectively, the bases occupying nucleotide sites 220, 221, 304, 305, 307, and 308. The presence or absence of this character separates the families Macropodidae and Didelphidae from the other mammalian families (Romero-Herrera et al. 1978). DNA/DNA hybridization does not reveal this type of detail, though complete DNA sequences do. A further problem with DNA hybridization data is that the melting temperature is raised by the spurious identities discussed earlier, even though these are phylogenetically uninformative and even misleading in some cases. Also, multiply hit sites that result in a base difference between the two DNAs being hybridized, though phylogenetically less informative than singly hit sites, lower the melting temperature just as much as (from the point of view that sites that distort the phylogenetic history should ideally be transparent to the measurement procedure) or no more than (from the point of view that DNA/DNA hybridization is intended to estimate total substitutions) the latter. Andrews (1987) notes that "the problem with DNA/DNA hybridization data is the same as for any distance statistic, that it is not possible to isolate the changes that are being measured." On the other hand, DNA/DNA hybridization is the only technique currently available that is able to detect rate differences among many weakly constrained regions, such as synonymous sites and introns, in a

cost-effective manner. Hybridization can do this because of the huge total number of sites involved, but protein and single-gene DNA cannot (Wu and Li 1985).

The matter of complexity also relates to the question as to whether morphological, chromosomal, or molecular data are more informative with regard to branching order. Until the genetic basis of morphological change and of chromosomal banding patterns is better understood, the extent to which each may be an indicator of speciation events is simply unknown. And only recently (Hutter and Ashburner 1987) has the manner in which gene sequences may relate to speciation come into focus.

A Summary of Factors Contributing to Uncertain Inference of Branching Orders and Branch Lengths

From the above review, the only honest statement about our present knowledge concerning the branching order and rates of evolution among higher-primate lineages is that we do not really *know* either. Even a single investigator examining the same data in different ways can reach different conclusions: the chimpanzee/human branching order was favored by Andrews (1986); but the chimpanzee/gorilla order is favored in Andrews (1987).

The reasons for the current inability to pinpoint the correct branching order, branch lengths, and rates of evolution among higher primates can be summarized into four categories:

Category I: Factors Inherent to the Evolutionary Process

These include (1) the relative rareness in closely related species of phylogenetically informative sites, (2) the existence of polymorphisms (Nei 1986), and (3) sampling error due to finite sequence length causing overlap in the estimates of the positions of branch points.

Category II: Factors More or Less Common to All Current Methodologies

These factors are (4) the incorrect apportioning of evolutionary events among the various branches of a phylogenetic tree because current models permit (5) events along more recent lineages to influence estimates of the number of events along more ancient branches or (6) events along more ancient lineages to influence estimates of the number of events along more recent, but independently, evolving lineages. (7) Some models ignore or do not sufficiently allow for heterogeneity of evolutionary rates among different characters or among different lineages descendant from the same or, more generally, different ancestral nodes. This includes (8) using global clocks to estimate local behavior, and (9) inadequately recognizing that molecular change is not always proportional to time. (10) The models do not correct for the fact that two distantly related species can share a pattern expected for two more closely related species. (11) The statistics for distinguishing between topologies do not have the power to do so. (12) The models fail to distinguish phylogenetically informative from non-informative sites. This can take various forms from including in the analysis (13) shared—but not informative—positions, such as unvaried sites; (14) noise caused by multiple hits, reversals, and parallelisms; (15) spurious chance identities arising from the assumptions of the evolutionary model adopted for interpreting the data; and (16) failing to distinguish between the nucleotide differences resulting from multiple hits and single hits and between genetic transitions and transversions.

Category III: Factors Unique to Particular Methodologies

With regard to parsimony, (17) the number of extra substitutions necessary to make it possible to reject statistically trees other than the most parsimonious one is much larger than the number of extra substitutions usually found in "nearby," not maximally parsimonious, trees. (18) Inaccuracies in the ancestral sequences cause errors in branch-length and branching-order estimations. With regard to distance-matrix methods, (19) there is a large loss of information in converting sequence data to pairwise distance methods. (20) These methods are sensitive to small changes in pairwise distance values. This is particularly so when these distance values are all small.

Category IV: Biological Factors

(21) Estimates of branching order and branch lengths may be confounded by differential generation times, age at first breeding, and number of cell cycles per unit time in the germ line (see Sibley and Ahlquist 1987 for review).

With regard to the factors in category I, we can improve our inferences by collecting more data to increase the number of phylogenetically informative sites and to reduce the sampling error. Estimates of how much more data are necessary to estimate unequivocally a particular evolutionary branching order or branch length can be helpful here (Ferris et al. 1981; Penny and Hendy 1986; Maeda et al. 1988). This does not mean rare events should be ignored. On the contrary, they can be important in guiding us to the historically correct inference (Miyamoto et al. 1987). However, the above historical review shows that gathering more data does not in itself suffice to resolve the problems of determining accurate branching orders and branch lengths. The number of base pairs sampled is not a reliable indicator of how informative the data are concerning the relative merits of two alternative phylogenies, particularly when the variable sites are unique to a single species (Templeton 1986*b*) and hence cladistically uninformative about branching order.

Most of the remaining problems arise from inadequate theory and not fully realistic evolutionary models. Lake's recent work (1987*a*, 1987*b*) has jogged the field out of its complacency with regard to the factors in categories II and III. He has demonstrated that these factors are not necessary evils with which we have to live but that, by a combination of looking at experimental macromolecular patterns different than those to which we have become accustomed and better theory, many of these factors can be ameliorated or eliminated. Lake's work suggests that if one concentrates on the proper patterns in nucleic acid sequence data, it is not necessary to know the ancestral sequence in order to deduce either the correct branching order (Lake 1987*b*) or the correct branch lengths (Lake 1987*a*). He reduces the problems of multiple hits by preferentially examining transversions. In summary, his methods significantly ameliorate or eliminate all the problems in categories II and III above. His methods work in cases in which both parsimony and distance-matrix methods do not.

In category III, Penny and Hendy (1986) have made progress by reposing the question. Instead of trying to decide between the most parsimonious and nearby trees on the basis of insufficient data, they ask: For a given number of taxa, what is the number of (informative) characters necessary to insure, at a given statistical significance level, that the most parsimonious tree is the historically correct tree?

Current understanding (Li et al. 1987; Sibley and Ahlquist 1987; Maeda et al.

1988) of the factors in category IV is very primitive, and they will not be discussed further here.

In the companion paper (Holmquist et al. 1987) we extend and apply Lake's (1987*a*, 1987*b*) methods to determine the branching order and rates of evolution among the higher primates.

Note added in proof: Ellen M. Prager (Department of Biochemistry, University of California, Berkeley; personal communication), in collaboration with Allan C. Wilson, independently obtained the same evolutionary parsimony results reported in tables 2 and 3 for the branching order based on the mtDNA sequence data of Brown et al. (1982). Using the same mtDNA data set, Prager also did the other four four-taxon comparisons possible among human (H), chimpanzee (C), gorilla (G), orangutan (Or), and gibbon (Gb) as well as standard parsimony analyses for all five comparisons. Letting the phylogenetic invariants X, Y, and Z, respectively, be associated with the networks joining the first and second taxa listed in each comparison, the first and third taxa, and the first and fourth taxa, she derived the following results: (*a*) HCGGb: X = 3, $P = 0.25$; Y = 2, P = 0.50; Z = −1, $P = 1.00$. (*b*) HCOrGb: X = 10, $P = 0.0063$; Y = 1 and Z = −1, $P = 1.00$. (*c*) HGOrGb: X = 2, $P = 0.75$; Y = 0 and Z = −1, $P = 1.00$. (*d*) CGOrGb: X = 3, $P = 0.51$; Y = 0 and Z = 1, $P = 1.00$.

Based on these mitochondrial data alone, (*a*) shows that if gibbon is used in an evolutionary parsimony analysis as the outgroup to human, chimpanzee, and gorilla, the association of human with chimpanzee is nonsignificantly favored but may be weaker than if orangutan (table 3) is used as the outgroup. In both instances standard parsimony analyses supported the same association, but they did not produce statistically significant resolution among the three possible topologies. For (*b*), evolutionary parsimony grouped human with chimpanzee at a high level of significance, as did standard parsimony: 42 phylogenetically informative sites favored the tree joining human and chimpanzee, and six sites favored each of the alternatives. If the null hypothesis is that the expected number of informative sites favoring one tree equals the number favoring a second tree, then the probability of obtaining this result by chance is 10^{-7}. In (*c*) and (*d*), evolutionary parsimony showed a marginal but statistically nonsignificant association of the gorilla with either human or chimpanzee. Standard parsimony, in contrast, strongly favors this association in both comparisons: 34 or 39 informative sites support joining gorilla with human or chimpanzee, respectively, and only 10 or 11 sites support the alternatives. Under the above null hypothesis, the probability that these results would be obtained by chance is less than one in a thousand. The contrasting results between standard and evolutionary parsimony in (*c*) and (*d*) are in part due to the fact that standard parsimony considers as phylogenetically informative both sites with transition differences and sites with transversion differences. See Prager and Wilson (accepted) for additional insights into the strengths and weaknesses of the standard and evolutionary parsimony methods, with attention to diverse types of sequences, base compositions, evolutionary distances, and modes and tempos of evolution.

The results for the mitochondrial data can be summarized as follows: for all five four-taxon comparisons possible, both standard and evolutionary parsimony favored the same topology and support the hypothesis that humans, chimpanzees, and gorillas form a monophyletic group. For some of the cases in which evolutionary parsimony supported a particular topology weakly, standard parsimony did so strongly. In other cases, both methods provided only weak support. These observations suggest caution

in the use of both standard and evolutionary parsimony to elucidate branching order until the relative merits of all factors affecting each are better understood.

Acknowledgments

This work was supported by grants BSR-8304140 (to R.H.) and BSR-8307336 (to M.G.) from the National Science Foundation.

LITERATURE CITED

ANDREWS, P. 1986. Molecular evidence for catarrhine evolution. Pp. 107–129 *in* B. WOOD, L. MARTIN, and P. ANDREWS, eds. Major topics in primate and human evolution. Cambridge University Press, Cambridge.

———. 1987. Aspects of hominoid phylogeny. Pp. 23–53 *in* C. PATTERSON, ed. Molecules and morphology in evolution: conflict or compromise? Cambridge University Press, Cambridge.

BIANCHI, N. O., M. S. BIANCHI, J. E. CLEAVER, and S. WOLFF. 1985. The pattern of restriction enzyme–induced banding in the chromosomes of chimpanzee, gorilla, and orangutan and evolutionary significance. J. Mol. Evol. **22**:323–333.

BISHOP, M. J., and A. E. FRIDAY. 1985. Evolutionary trees from nucleic acid and protein sequences. Proc. R. Soc. Lond. [Biol.] **226**:271–302.

———. 1986. Molecular sequences and hominoid phylogeny. Pp. 150–156 *in* B. WOOD, L. MARTIN, and P. ANDREWS, eds. Major topics in primate and human evolution. Cambridge University Press, Cambridge.

BRITTEN, R. J. 1986. Rates of DNA sequence evolution differ between different taxonomic groups. Science **231**:1393–1398.

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.

CAVENDER, J. A. 1978. Taxonomy with confidence. Math. Biosci. **40**:271–280 (also see Erratum, **44**:309 [1979].)

———. 1981. Tests of phylogenetic hypotheses under generalized models. Math. Biosci. **54**: 217–229.

CIOCHON, R. L. 1985. Hominoid cladistics and the ancestry of modern apes and humans. Pp. 345–363 *in* R. L. CIOCHON and J. G. FLEAGLE, eds. Primate evolution and human origins. Benjamin-Cummings, Menlo Park, Calif.

CREEL, N. 1986. Size and phylogeny in hominoid primates. Syst. Zool. **35**:81–99.

DELSON, E., N. ELDREDGE, and I. TATTERSALL. 1977. Reconstruction of hominid phylogeny: a testable framework based on cladistic analysis. J. Hum. Evol. **6**:263–278.

DENE, H. T., M. GOODMAN, and W. PRYCHODKO. 1976. Immunodiffusion evidence on the phylogeny of the primates. Pp. 171–196 *in* M. GOODMAN and R. E. TASHIAN, eds. Molecular anthropology. Plenum, New York.

DUTRILLAUX, B. 1979. Chromosomal evolution in primates: tentative phylogeny from *Microcebus murinus* (prosimian) to man. Hum. Genet. **48**:251–314.

———. 1980. Chromosomal evolution of the great apes and man. J. Reprod. Fertil. [Suppl.] **28**:105–111.

FARRIS, J. S. 1972. Estimating phylogenetic trees from distance matrices. Am. Nat. **106**:645–668.

———. 1985. Distance data revisited. Cladistics **1**:67–85.

FELSENSTEIN, J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. Syst. Zool. **27**:401–409.

———. 1981*a*. Evolutionary trees from DNA sequences: a maximum likelihood approach. J. Mol. Evol. **17**:368–376.

———. 1981*b*. A likelihood approach to character weighting and what it tells us about parsimony and compatibility. Biol. J. Linn. Soc. **16**:183–196.

———. 1983. Inferring evolutionary trees from DNA sequences. Pp. 133–150 *in* B. S. WEIR, ed. Statistical analysis of DNA sequence data. Marcell Dekker, New York.

———. 1985. Confidence limits on phylogenies with a molecular clock. Syst. Zool. **34**:152–161.

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. 1986. Commentary. Mol. Biol. Evol. **3**:296–298.

FITCH, W. M., and M. BRUSCHI. 1987. The evolution of prokaryotic ferredoxins—with a general method correcting for unobserved substitutions in less branched lineages. Mol. Biol. Evol. **4**:381–394.

FITCH, W., and E. MARGOLIASH. 1967. Construction of phylogenetic trees. Science **155**:279–284.

FITCH, W., and E. MARKOWITZ. 1970. An improved method for determining codon variability in a gene and its application to the rate of fixation of mutations in evolution. Biochem. Genet. **4**:579–593.

GINGERICH, P. D. 1986. Temporal scaling of molecular evolution in primates and other mammals. Mol. Biol. Evol. **3**:205–221.

GODSON, J. R., A. R. MITCHELL, H. N. SEUÁNEZ, and C. M. GODSEN. 1977. The distribution of sequences complementary to human satellite DNAs I, II and IV in the chromosomes of chimpanzee (*Pan troglodytes*), gorilla (*Gorilla gorilla*), and orang utan (*Pongo pygmaeus*). Chromosoma **63**:253–271.

GOLDSTONE, A., and E. SMITH. 1966. Amino acid sequence of whale heart cytochrome *c*. J. Biol. Chem. **241**:4480–4486.

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

GOODMAN, M., J. CZELUSNIAK, and J. BEEBER. 1985. Phylogeny of primates and other eutherian orders: a cladistic analysis using amino acid and nucleotide sequence data. Cladistics **1**:171–185.

GOODMAN, M., J. CZELUSNIAK, G. W. MOORE, A. E. ROMERO-HERRERA, and G. MATSUDA. 1978. Fitting the gene lineage into its species lineage, a parsimony strategy illustrated by cladograms constructed from globin sequences. Syst. Zool. **28**:132–163.

GOULD, S. J. 1977. Ontogeny and phylogeny. Harvard University Press, Cambridge, Mass.

HASEGAWA, M., H. KISHINO, and T. YANO. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J. Mol. Evol. **22**:160–174.

HASEGAWA, M., and T. YANO. 1984*a*. Maximum likelihood method of phylogenetic inference from DNA sequence data. Bull. Biometric Soc. Japan **5**:1–7.

———. 1984*b*. Phylogeny and classification of Hominoidea as inferred from DNA sequence data. Proc. Jpn. Acad. **60B**:389–392.

HIXSON, J. E., and W. M. BROWN. 1986. A comparison of the small ribosomal RNA genes from the mitochondrial DNA of the great apes and humans: sequence, structure, evolution, and phylogenetic implications. Mol. Biol. Evol. **3**:1–18.

HOLMQUIST, R. 1976. Random and nonrandom processes in the molecular evolution of higher organisms. Pp. 89–116 *in* M. GOODMAN and R. E. TASHIAN, eds. Molecular anthropology. Plenum, New York.

———. 1978. A measure of the denseness of a phylogenetic network. J. Mol. Evol. **11**:225–231.

———. 1979. The method of parsimony: an experimental test and theoretical analysis of the adequacy of molecular restoration studies. J. Mol. Biol. **135**:939–958.

———. 1980. Evolutionary analysis of α and β hemoglobin genes by REH theory under the assumption of the equiprobability of genetic events. J. Mol. Evol. **15**:149–159.

HOLMQUIST R., C. CANTOR, and T. H. JUKES. 1972. Improved procedures for comparing homologous proteins and nucleic acids. J. Mol. Biol. **64**:145–161.

HOLMQUIST, R., M. GOODMAN, T. CONROY, and J. CZELUSNIAK. 1983. The spatial distribution of fixed mutations within genes coding for proteins. J. Mol. Evol. **19**:437–448.

HOLMQUIST, R., M. M. MIYAMOTO, and M. GOODMAN. 1988. Analysis of higher-primate phylogeny from transversion differences in nuclear and mitochondrial DNA by Lake's methods of evolutionary parsimony and operator metrics. Mol. Biol. Evol. **5**:217–236.

HOLMQUIST, R., and D. PEARL. 1980. Theoretical foundations for quantitative paleogenetics. Part III. The molecular divergence of nucleic acids and proteins for the case of genetic events of unequal probability. J. Mol. Evol. **16**:211–267.

HOLMQUIST, R., D. PEARL, and T. H. JUKES. 1982. Nonuniform molecular divergence: the quantitative analysis of genes and messenger RNAs under selective structural constraints. Pp. 281–315 *in* M. GOODMAN, ed. Macromolecular sequences in systematic and evolutionary biology. Plenum, New York.

HUTTER, P., and M. ASHBURNER, 1987. Genetic rescue of inviable hybrids between *Drosophila melanogaster* and its sibling species. Nature **327**:331–333.

JONES, K. W., J. PROSSER, G. CORNEO, E. GINELLI, and M. BOBROW. 1972. Satellite DNA, constitutive heterochromatin, and human evolution. Pp. 45–74 *in* R. A. PFEIFFER, ed. Modern aspects of cytogenetics: constitutive heterochromatin in man. Schattauer, Stuttgart.

KARON, J. M. 1979. The covarion model for the evolution of proteins: parameter estimates and comparison with Holmquist, Cantor, and Jukes' stochastic model. J. Mol. Evol. **12**:197–218.

KING, J. L. 1980. Does the information density of amino acid composition increase? J. Mol. Evol. **15**:73–75.

KLOTZ, L. C., and R. L. BLANKEN. 1981. A practical method for calculating evolutionary trees from sequence data. J. Theor. Biol. **91**:261–272.

KLUGE, A. G. 1983. Cladistics and the classification of the great apes. Pp. 151–177 *in* R. L. CIOCHON and R. S. CORRUCCINI, eds. New interpretations of ape and human ancestry. Plenum, New York.

KOOP, B., M. GOODMAN, P. XU, K. CHAN, and J. L. SLIGHTOM. 1986. Primate η-globin DNA sequences and man's place among the great apes. Nature **319**:234–238.

LAKE, J. A. 1987*a*. Determining evolutionary distances from highly diverged nucleic acid sequences: operator metrics. J. Mol. Evol. **26**:59–73.

———. 1987*b*. A rate-independent technique for analysis of nucleic acid sequences: evolutionary parsimony. Mol. Biol. Evol. **4**:167–191.

LANAVE, C., G. PREPARATA, C. SACCONE, and G. SERIO. 1984. A new method for calculating evolutionary substitution rates. J. Mol. Evol. **20**:86–93.

LANAVE, C., S. TOMMASI, G. PREPARATA, and C. SACCONE. 1986. Transition and transversion rate in the evolution of animal mitochondrial DNA. BioSystems **19**:273–283.

LI, W.-H., M. TANIMURA, and P. SHARP. 1987. An evaluation of the molecular clock hypothesis using mammalian DNA sequences. J. Mol. Evol. **25**:330–342.

MAEDA, N., C.-I. WU, J. BLISKA, and J. RENEKE. 1988. Molecular evolution of intergenic DNA in higher primates: pattern of DNA changes, molecular clock, and evolution of repetitive sequences. Mol. Biol. Evol. **5**:1–20.

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

MIYAMOTO, M. M., J. L. SLIGHTOM, and M. GOODMAN. 1987. Phylogenetic relationships of humans and African apes as ascertained from DNA sequences (7.1 kilobase pairs) of the ψη-globin region. Science **238**:369–373.

MOORE, G. W. 1977. Proof of the populous path algorithm for missing mutations in parsimony trees. J. Theor. Biol. **91**:621–636.

MOORE, G. W., M. GOODMAN, C. CALLAHAN, R. HOLMQUIST, and H. MOISE. 1976. Stochastic *versus* augmented maximum parsimony method for estimating superimposed mutations in the divergent evolution of protein sequences: methods tested on cytochrome *c* amino acid sequences. J. Mol. Biol. **105**:15–37.

NEI, M. 1986. Stochastic errors in DNA evolution and molecular phylogeny. Pp. 133–147 *in* H. GERSHOWITZ, D. L. RUCKNAGEL, and R. E. TASHIAN, eds. Evolutionary perspectives and the new genetics. Alan R. Liss, New York.

NEI, M., J. C. STEPHENS, and N. SAITOU. 1985. Methods of computing the standard errors of branching points in an evolutionary tree and their application to molecular data from humans and apes. Mol. Biol. Evol. **2**:66–85.

NEI, M., and F. TAJIMA. 1985. Evolutionary change of restriction cleavage sites and phylogenetic inference for man and apes. Mol. Biol. Evol. **2**:189–205.

———. 1987. Problems arising in phylogenetic inference from restriction-site data. Mol. Biol. Evol. **4**:320–323.

OXNARD, C. 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.

PENNY, D., and M. HENDY. 1986. Estimating the reliability of evolutionary trees. Mol. Biol. Evol. **3**:403–417.

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

———. 1983. Spatial and temporal change in biogeography: gradual or abrupt? Pp. 25–56 *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.

PILBEAM, D. 1986. Hominoid evolution and hominoid origins. Am. Anthropol. **88**:295–312.

PRAGER, E. M., and A. C. WILSON. Ancient origin of lactalbumin from lysozyme: analysis of DNA and amino acid sequences. J. Mol. Evol. (accepted).

PREPARATA, G., and C. SACCONE. 1987. A simple quantitative model of the molecular clock. J. Mol. Evol. **26**:7–15.

ROMERO-HERRERA, A. E., H. LEHMANN, K. A. JOYSEY, and A. E. FRIDAY. 1978. On the evolution of myoglobin. Philos. Trans. R. Soc. Lond. [Biol.] **283**:61–163.

RUVOLO, M., and D. PILBEAM. 1986. Hominoid evolution: molecular and palaeontological pattern. Pp. 157–160 *in* B. WOOD, L. MARTIN, and P. ANDREWS, eds. Major topics in primate and human evolution. Cambridge University Press, Cambridge.

RUVOLO, M., and T. F. SMITH. 1986. Phylogeny and DNA-DNA hybridization. Mol. Biol. Evol. **3**:285–289.

SACCONE, C., G. PREPARATA, C. LANAVE, and S. TOMASSI. 1987. Building up molecular clocks from mitochondrial messenger RNA coding sequences. Pp. 39–48 *in* V. PESCE DELFINO, ed. International Symposium on Biological Evolution. Adriatica Editrice, Bari, Italy.

SAITOU, N. 1986. On the delta-Q test of Templeton. Mol. Biol. Evol. **3**:282–284.

SCHWARTZ, J. H. 1984. The evolutionary relationships of man and the orang-utan. Nature **308**:501–505.

SCHWARTZ, R., and M. O. DAYHOFF. 1978. Origins of prokaryotes, eukaryotes, mitochondria, and chloroplasts. Science **199**:395–403.

SEUÁNEZ, H. N. 1982. Chromosome banding and primate phylogeny. Pp. 224–235 *in* A. B. CHIARELLI and R. S. CORRUCCINI, eds. Advanced views in primate biology. Springer, Berlin.

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.

———. 1987. DNA hybridization evidence of hominoid phylogeny: results from an expanded data set. J. Mol. Evol. **26**:99–121.

SLIGHTOM, J. L., L.-Y. E. CHANG, B. F. KOOP, and M. GOODMAN. 1985. Chimpanzee fetal $^{G}\gamma$ and $^{A}\gamma$ globin gene nucleotide sequences provide further evidence of gene conversions in hominine evolution. Mol. Biol. Evol. **2**:370–389.

SLIGHTOM, J. L., T. W. THIESEN, B. F. KOOP, and M. GOODMAN. 1987. Orangutan fetal globin genes: nucleotide sequences reveal multiple gene conversions during hominid phylogeny. J. Biol. Chem. **262**:7427–7483.

SNEATH, P. H. A., and R. SOKAL. 1973. Numerical taxonomy. W. H. Freeman, San Francisco.
SOKOLOVSKY, M., and M. MOLDOVAN. 1972. Primary structure of cytochrome *c* from the camel, *Camelus dromedarius.* Biochemistry **11**:145–149.
STANYON, R., and B. CHIARELLI. 1982. Phylogeny of the Hominoidea: the chromosome evidence. J. Hum. Evol. **11**:493–504.
TEMPLETON, A. R. 1983*a*. 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.
———. 1983*b*. Phylogenetic inference from restriction endonuclease cleavage site maps 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.
———. 1986*a*. Further comments on the statistical analysis of DNA-DNA hybridization data. Mol. Biol. Evol. **3**:290–295.
———. 1986*b*. Relation of humans to African apes: a statistical appraisal of diverse types of data. Pp. 365–388 *in* S. KARLIN and E. NEVO, eds. Evolutionary processes and theory. Academic Press, New York.
———. 1987. Nonparametric phylogenetic inference from restriction cleavage sites. Mol. Biol. Evol. **4**:315–319.
WASHBURN, S. L. 1982. Human evolution. Perspect. Biol. Med. **25**:583–602.
WILBUR, W. J. 1985. On the PAM matrix model of protein evolution. 1985. Mol. Biol. Evol. **2**:434–447.
WILSON, A. C., S. S. CARLSON, and T. J. WHITE. 1977. Biochemical evolution. Annu. Rev. Biochem. **46**:573–639.
WU, E.-I., and W.-H. LI. 1985. Evidence for higher rates of nucleotide substitution in rodents than in man. Proc. Natl. Acad. Sci. USA **82**:1741–1745.
YUNIS, J. J., and O. PRAKASH. 1982. The origin of man: a chromosomal pictorial legacy. Science **215**:1525–1530.

WALTER M. FITCH, reviewing editor

Received July 21, 1987; revision received November 17, 1987