EVOLUTION

INTERNATIONAL JOURNAL OF ORGANIC EVOLUTION

PUBLISHED BY

THE SOCIETY FOR THE STUDY OF EVOLUTION

Vol. 43 August, 1989 No. 5

Evolution, 43(5), 1989, pp. 925–942

DNA DIVERGENCE AMONG HOMINOIDS

ADALGISA CACCONE[1] AND JEFFREY R. POWELL

Department of Biology, Yale University, P.O. Box 6666, New Haven, CT 06511

Abstract.—We have determined the degree of single-copy DNA divergence among the extant members of the Hominoidea employing the technique of DNA-DNA hybridization. The species studied include humans, two species of chimpanzees, gorillas, two subspecies of orangutans, and two species of gibbons; as an outgroup we have used a member of the Old World monkeys (Cercopithecidae), the baboon. Our methods are different from those previously used and allow us to control for two factors other than base-pair mismatch that can affect the thermal stability of DNA duplexes: the base composition and duplex length. In addition, we have studied more than one individual for most species and thus are able to assess the effect of intraspecific variation on phylogenetic conclusions.

The results indicate that the closest extant relatives of humans are the chimpanzees. Gorillas are the next closest, followed by orangutans and gibbons. This result is strongly supported statistically, as there is virtually no overlap in measurements between different taxa. Our conclusions are in agreement with a growing amount of molecular evidence supporting this pattern of relatedness. The data behave as a reasonably good molecular clock, and we do not see an indication of slowdown in molecular evolution in the clade containing humans and African apes, contrary to what has been documented for protein-coding regions. Because of the clocklike nature of the results, we have estimated that the divergence of humans and chimpanzees occurred about 6–8 million years ago. Results from orangutans indicate that the Borneo and Sumatra populations are genetically distinct, about as different as the named species of chimpanzees.

Received November 8, 1988. Accepted March 5, 1989

The higher primates, specifically the Hominoidea, have presented molecular evolutionists and systematists with a difficult challenge. A long-standing goal of primatologists and anthropologists has been to establish a firm phylogeny for the extant members of this group in order to understand the evolutionary significance of fossil remains and thus reach a better understanding of the evolution of our own species. Classical systematic studies led to a number of proposed phylogenies, and it was hoped that molecular studies would shed light on which of the phylogenies was most likely. Data from molecular studies are a very different sort of information, which should serve as independent verification of the true phylogeny.

The main problem began to emerge with the earliest work of Goodman (1963) and Sarich and Wilson (1967). These studies, based on immunological methods, indicated that humans (*Homo sapiens*), common chimpanzees (*Pan troglodytes*), and gorillas (*Gorilla gorilla*) are almost equally distant from one another, and thus it was impossible to establish an unambiguous dichotomous relationship among them. Sarich and Wilson's (1967) introduction of the molecular-clock concept led to the further conclusion that these lineages separated very recently. Thus, the challenging problem presented by this important group is twofold.

[1] On leave from the Dipartimento di Biologia, II Università di Roma "Tor Vergata," Rome, Italy.

First, they split from one another quite recently so relatively little time has elpased for the accumulation of phylogenetically informative changes. Second, assuming that the three lineages did not undergo a trifurcation, the time between dichotomous splits is small. Suffice it to say that all subsequent molecular data have confirmed these original observations. It is not our purpose to review the rather large literature on the subject here; interested readers are referred to several recent reviews: Andrews and Martin (1987), Bishop and Friday (1985), Holmquist et al. (1988*a*), Ruvolo and Pilbeam (1986), and Sibley and Ahlquist (1987).

In addition to the three species just mentioned, Hominoidea includes the pygmy chimpanzee (*Pan paniscus*), the orangutan (*Pongo pygmaeus*), and the gibbons (genus *Hylobates*; sometimes split into a second genus, *Symphalangus*). A consensus has been reached, based on molecular, morphological, and fossil evidence, as to the general phylogeny of the group. Humans and the African apes form one lineage, with the closest relatives being orangutans, followed by gibbons; the cercopithecines (Old World monkeys) are acknowledged to be the extant sister group of Hominoidea (for an alternative view of the placement of the orangutan, see Schwartz [1984]). The bulk of the controversy, therefore, has centered around the phylogeny within the clade containing humans and African apes.

The technique we have used to assess the degree of DNA divergence among hominoids is DNA-DNA hybridization. This technique has been applied to hominoids previously, with the result that a trichotomy among human-chimpanzee-gorilla (combining the common chimpanzee and the pygmy chimpanzee) could not be rejected (Hoyer et al., 1972; Benveniste and Todaro, 1976; O'Brien et al., 1985) or that humans and chimpanzees are more closely related to each other than either is to gorillas (Sibley and Ahlquist, 1984, 1987). This latter result has been widely cited as well as challenged (Lewin, 1988*a*, 1988*b*). Because of the importance of this result, not just for anthropology but for molecular systematics in general, we chose to restudy this group using a different technique to measure the thermal stability of DNA duplexes.

The technique of DNA-DNA hybridization is conceptually straightforward. It is based on the fact that the thermal stability of duplex DNA molecules is dependent upon the fidelity of base-pair matching: C with G, and A with T. The change in median melting temperature, the ΔT_m, is the difference between the T_m of a homoduplex molecule (in which the two strands are from the same individual or species) and the T_m of the heteroduplex (in which each strand comes from a different source). Melting refers to the transition from double-stranded to single-stranded. In theory, the ΔT_m should be linearly related to degree of base-pair mismatch (Britten et al., 1974), and indeed, we have empirically demonstrated this (Caccone et al., 1988*b*). We have used a different method to determine the thermal stability than that used by Sibley and Ahlquist (1984, 1987), who used the hydroxyapatite (HAP) method. We have used the so-called TEACL (tetraethylammonium chloride) method (Britten et al., 1978), which allows us to control for two factors other than base-pair mismatch that determine the thermal stability of DNA duplexes; these are the base composition and the length of the duplexes. Thus, while our results are based on the same principle, the actual technique of measuring the thermal stability of DNA duplexes is different. Therefore, strictly speaking, we did not replicate the Sibley and Ahlquist (1984, 1987) experiments; rather, we have tested whether independent measurements of the same parameter (the thermal stability of DNA molecules) yield the same results.

Materials and Methods

The sources of the DNAs used in our experiments are given in Table 1. Of the 16 samples, nine were prepared by us from blood. Seven samples were total genomic DNA prepared by others and kindly given to us. Of these seven, three were supplied by C. Sibley and were used in the earlier experiments. Thus, we have used mostly independently obtained DNAs, with a few used in the previous studies.

The extraction procedure for blood samples was modified from a procedure provided by J. Rogers (Department of Anthropology, Yale University). Briefly, this consisted of concentrating the white blood cells by centrifugation in saline. The cells were lysed and subjected to prolonged digestion with proteinase-K, followed by repeated phenol-chloroform extractions. After ethanol precipitation, the samples were RNAse-treated, ethanol-precipitated again, and resuspended in water. Concentration and purity were monitored by UV spectrophotometric readings. The sizes of the molecules were monitored by agarose-gel electrophoresis, both before and after sonication. DNA was randomly broken by sonication to a size of 500–800 bp.

The TEACL method of DNA-DNA hybridization is described in Britten et al. (1978), Hall et al. (1980), and Hunt et al. (1981). Minor modifications used in our laboratory, as well as our statistical procedures, are described in Caccone and Powell (1987) and Caccone et al. (1987). A few aspects of these procedures are worthy of emphasis here. First, all experiments were done using single-copy DNA (scDNA). This refers to the fact that excess copies of repetitive sequences have been removed by incubating single-stranded DNA to a C_0t of 1,000 M sec/liter and removing the duplexes by two rounds of absorption to hydroxyapatite. The resulting single-stranded fraction was labeled by random-primer ("oligo-labeling") procedures (Feinberg and Vogelstein, 1983) without preheating the DNA. Thus, any double-stranded sequences that may not have been removed by the hydroxyapatite-column step would not be labeled. Tritiated (^{3}H) TTP was used as the radioactive label. Second, the reassociation reactions were carried out under conditions (1 M TEACL, 45°C) that require about 75% or greater base matching to form stable duplexes. Third, the degree of reassociation was determined by nuclease-S1 digestion after the reassociation reaction had come to an equilibrium; we used a C_0t of 12,000 M sec/liter or greater. Fourth, T_m determinations were made in a solution of 2.4 M TEACL; at this concentration, DNA

Table 1. Sources of DNA used in this study; details available upon request.

Abbreviation	Description	Source
Homo sapiens:		
HU1	blood from adult male	—
HU2	blood from adult female	—
Pan paniscus:		
PP1	blood from 8-year-old male	D. Janssen; San Diego Zoo
PP2	blood from 7-year-old female (half sib of individual above)	A. Shima; San Diego Zoo
Pan troglodytes:		
PT1	blood from 11-year-old female	C. Allen; Yerkes
PT2	DNA preparation	O. Ryder; San Diego Zoo
Gorilla gorilla:		
GO1	DNA preparation ("chop" 2565)	C. Sibley
GO2	blood from 26-year-old female	C. Allen; Yerkes
GO3	blood from female	N. Schaffer; Dallas Zoo
GO4	blood from 15-year-old female	C. Allen; Yerkes
Pongo pygmaeus:		
OR1	DNA preparation ("chop" 3123); presumed from Borneo	C. Sibley
OR2	blood from 27-year-old female from Borneo	A. Shima; San Diego Zoo
OR3	DNA preparation from Sumatran source	O. Ryder; San Diego Zoo
Hylobates lar:		
HYL	DNA preparation ("chop" 2566)	C. Sibley
Hylobates syndactylus:		
HYS	DNA preparation	O. Ryder; San Diego Zoo
Papio hamadryas:		
BAB	DNA preparation from wild-caught adult female (central Tanzania)	J. Rogers

TABLE 2. All pairwise ΔT_m's (°C) with associated standard errors given in parentheses.

Driver	Tracers								
	HU1	HU2	PP1	PT1	GO2	OR1	OR2	HYS	BAB
HU1	—	—	1.66 (0.12)	1.62 (0.22)	2.40 (0.28)	3.42 (0.07)	3.27 (0.20)	5.13 (0.13)	6.57 (0.14)
HU2	0.31 (0.10)	—	1.62 (0.09)	1.82 (0.11)	2.53 (0.20)	—	—	—	—
PP1	1.79 (0.14)	1.51 (0.13)	—	0.97 (0.11)	2.75 (0.23)	—	—	—	—
PP2	1.55 (0.14)	1.36 (0.22)	0.37 (0.12)	0.32 (0.21)	2.73 (0.20)	—	—	4.86 (0.15)	—
PT1	1.37 (0.09)	1.92 (0.17)	1.03 (0.13)	—	2.39 (0.26)	3.72 (0.11)	2.96 (0.13)	4.52 (0.13)	—
PT2	1.42 (0.20)	1.39 (0.11)	—	−0.39 (0.19)	2.39 (0.20)	—	—	—	—
GO1	2.48 (0.19)	2.40 (0.10)	2.61 (0.07)	2.24 (0.16)	0.31 (0.23)	3.65 (0.13)	—	—	—
GO2	2.69 (0.20)	2.52 (0.25)	2.97 (0.11)	2.35 (0.16)	—	3.72 (0.10)	3.20 (0.13)	5.04 (0.19)	—
GO3	—	—	—	—	0.14 (0.23)	—	—	—	—
GO4	—	—	—	—	−0.30 (0.22)	—	—	—	—
OR1	3.63 (0.20)	3.63 (0.17)	—	—	3.44 (0.24)	—	0.05 (0.12)	5.11 (0.15)	—
OR2	—	—	3.76 (0.15)	3.17 (0.18)	3.86 (0.23)	0.35 (0.16)	—	—	—
OR3	—	—	—	—	—	1.22 (0.11)	0.84 (0.13)	4.86 (0.17)	—
HYL	4.95 (0.16)	—	4.90 (0.18)	4.48 (0.25)	5.32 (0.23)	—	4.77 (0.23)	1.26 (0.12)	7.21 (0.26)
HYS	—	—	4.76 (0.32)	4.44 (0.23)	4.92 (0.24)	4.80 (0.18)	4.74 (0.14)	—	—
BAB	6.99 (0.14)	—	7.43 (0.09)	6.58 (0.20)	7.12 (0.21)	—	7.33 (0.16)	6.75 (0.19)	—

melting is independent of base composition (Hutton and Wetmur, 1973; Orosz and Wetmur, 1977). Fifth, a sample of the DNA that was melted was reserved for size measurement on alkaline-agarose gels; from the resulting mean duplex size, a correction can be made in the T_m. See Hall et al. (1980) and Hunt et al. (1981) for details and for the theoretical and empirical justification for size correction. Sixth, the melting is done in a temperature gradient set up by placing aliquots in holes in an aluminum block. Thus, the raw data we obtained was in the form of a cumulative melting curve, rather than in terms of the amount of additional DNA melting at each temperature, as is the case with the HAP method.

The PHYLIP computer program package (available from J. Felsenstein, Department of Genetics, University of Washington, Seattle) was used in analyzing the results. Both the KITSCH program (which assumes contemporaneous tips) and the FITCH program (which does not) were used. In the present case, this is tantamount to assuming or not assuming that ΔT_m is behaving as a molecular clock.

RESULTS

Table 2 summarizes the results of all 77 pairwise comparisons we have performed; with replicates these represent a total of 417 median melting determinations. Table 2 presents the corrected ΔT_m's, which we designate with an uppercase "T." More detailed data for each determination of melting temperature are presented in the tables in the Appendix. In the Appendix, we present the uncorrected t_m's (designated by lowercase "t"), the numbers of replicates for each determination, the mean sizes of the duplexes, the corrections made for size, and the resulting T_m's. We also indicate the percentage hybridization for each reassociation reaction. The standard errors presented in Table 2 were calculated as described in Caccone et al. (1987).

Reciprocity.—As noted in Table 2, each DNA preparation can be used as either a "tracer" or "driver." This refers to the fact that in making comparisons between two DNAs, one of the DNAs is labeled with radioactivity, which is used to trace the reassociation. The other source of DNA is not labeled and is in very high concentration in order to drive the reaction. In theory, it should not matter which species is used as tracer or driver; the degree of base-pair mismatch should be the same. Among our comparisons, there are 17 reciprocal combinations, i.e., 17 cases in which each source of DNA was used as tracer and driver in separate experiments. As can be seen in Table 2, reciprocity holds very well. For example, when the tracer prepared from pygmy chimpanzee PP1 is reassociated with the two human DNAs, the ΔT_m's are 1.66 ± 0.12 and 1.62 ± 0.09. When this same PP1 DNA is used as a driver with the two human tracers, the ΔT_m's are 1.79 ± 0.14 and 1.51 ± 0.13. Over all 17 cases, the mean difference between reciprocals is 0.21. The standard errors associated with each ΔT_m are generally between 0.1 and 0.2, with an overall mean of 0.17 (for all the data in Table 2). Therefore, the average difference between reciprocals (0.21) is not significantly different from zero. In fact, of the 17 cases of reciprocals, only one (that involving OR2 and GO2) has ΔT_m's differing by more than two standard errors. This internal consistency of the data is good evidence that no major artifacts have been introduced by experimental procedures.

Tracer Lengths and Temperature Corrections.—Longer duplexes require higher temperatures to dissociate than do shorter duplexes, if the duplexes have the same base composition and degree of base matching. This effect can be taken into consideration if the mean duplex lengths are known. The method to correct the median melting temperature has both a theoretical and an empirical basis (see Hall et al., 1980). We have used the formula presented in Hunt et al. (1981) to correct our t_m's to obtain T_m's.

A test to lend credence to the length correction is to compare reciprocals. If the length correction is legitimate, then the difference between reciprocal Δt_m's should be larger than the differences between ΔT_m's. This is the case. As just mentioned, the mean difference between reciprocal ΔT_m's for the 17 cases in the present data set is 0.21; for these same cases, the mean difference in Δt_m is 1.44. We have previously shown a similar phenomenon in our data for crickets and *Drosophila* (Caccone and Powell, 1987; Caccone et al., 1988*a*).

The mean Δt_m difference between reciprocals (1.44) is higher than desirable. However, this figure is inflated due to the fact that one tracer (HU2) produced hybrids that were consistently smaller than generally observed and, therefore, larger length corrections were required. We emphasize that the tracer lengths reported in the Appendix were determined after a nuclease-S1 digestion of the reassociation reactions; these are not the lengths of the tracers during the reassociation reaction. Any overhanging ends and loops due to insertions or deletions will be removed by the S1 treatment, and this will shorten the tracer lengths. Thus, in the TEACL method, "trimmed up" duplexes are melted. This is not the case with the HAP method.

Generally, we view with suspicion any hybrids with tracer lengths of less than 100 bp after S1 digestion. We only use the results from such hybrids if they are verified by reciprocals or DNA preparations of the same species or individual. This is the case with the present data, and therefore we have not discarded any data. Even if the data based on hybrids of less than 100 bp were disregarded, there would be no significant change in our conclusions.

Percentage Reassociation.—As mentioned in the Materials and Methods section, the reassociation reaction was performed under conditions requiring about 75% or greater base-pair match to form stable duplexes. If there are scDNA sequences that have diverged beyond this stringency requirement or if insertions/deletions have

TABLE 3. Cases for which replicate reassociation reactions were performed. Abbreviations for the species are as in Table 1; other abbreviations: N = number of replicate t_m determinations; t_m = mean uncorrected median melting temperature; SE = standard error of t_m and T_m; T_{corr} = temperature correction; T_m = corrected median melting temperature. Hybrids are indicated as tracer-driver.

Species pair	Replicate	N	t_m	SE	Tracer length (bp)	T_{corr}	T_m	Percentage reassociation
HU1-HU1	A	5	56.44	0.14	167	2.99	59.43	53.3
	B	5	56.65	0.13	172	2.91	59.56	62.2
	C	5	56.67	0.13	177	2.82	59.49	70.6
GO2-OR2	A	4	51.93	0.15	272	1.84	53.77	58.0
	B	4	51.58	0.18	266	1.88	53.46	69.7
PP1-PP1	A	5	56.03	0.07	168	2.98	59.01	65.8
	B	5	56.30	0.08	171	2.92	59.22	64.8
PP1-HU1	A	5	54.78	0.11	175	2.86	57.64	62.2
	B	5	54.62	0.14	188	2.66	57.28	61.0

occurred, parts of the tracer genome will not reassociate. The lowering of the percentage reassociation of heteroduplex reactions compared to homoduplex reactions is called the normalized percentage reassociation (NPR) by us (=normalized percentage hybridization [NPH] of other authors). This is simply the percentage heteroduplex reassociation divided by the corresponding percentage homoduplex reassociation. We determined the percentage of the tracer that had reassociated by a nuclease-S1 treatment after the reaction had come to equilibrium. As we have noted elsewhere (Caccone and Powell, 1987; Caccone et al., 1988*a*), this method of determining percentage reassociation is associated with significant error.

We present the percentage reassociation for each reaction in the Appendix. The problem with this measure is highlighted by the cases in which we made replicate reassociations and determined the T_m's for each. These are presented in Table 3. As can be seen, the t_m's, tracer lengths, and T_m's are all highly replicable, stable parameters. However, the percentage reassociation, as measured in our procedures, is not reliable. Note that some of the data in Table 3 are for homoduplexes and thus would affect all NPR estimates for those tracers. In the Appendix, we present the mean parameters for these separate experiments. Because of the nonreproducibility of NPR, we will only consider ΔT_m in our further analysis. We use T_m to refer to the median melting temperature of that fraction of the genome that hybridized. This is in contrast to T_{50}H (Sibley and Ahlquist, 1984, 1987), T_{median} (Britten, 1986), or T_mR (Benveniste, 1985), as used by other authors to incorporate percentage hybridization into the measures. We will treat the issue of incorporating NPR into the distance measure in the Discussion. Suffice it to say, this issue makes little or no difference for determining phylogenetic relationships.

Intraspecific Variation.—An obvious question in evaluating differences between species concerns the magnitude and effect of intraspecific variation. We have studied DNA prepared from two humans, two common chimpanzees, two pygmy chimpanzees, four gorillas, and three orangutans. There are two ways to assess possible effects of intraspecific variation on interspecific differences. First, one can directly compare individuals of the same species. We made eight intraspecific hybrids; this does not include hybrids between the Sumatran and Borneo orangutan (OR3 with OR1 or OR2). Of these, six have positive ΔT_m's and two have negative ΔT_m's. Of the eight intraspecific ΔT_m's, three are different from zero by more than two standard errors: HU1-HU2, PP1-PP2, and PT1-PT2, this last being a negative value (throughout this paper, when we refer to hybrids, we separate species abbreviations [Table 1] with a hyphen, with the tracer given first [tracer-driver]). In one case, the two individuals are half-sibs (PP1-PP2; see Table 1). Our overall assessment is that DNA-DNA hybridization may be detecting some degree of intraspecific polymorphism; however, it is difficult to conclude this definitively. The mean intraspecific ΔT_m for the eight comparisons is 0.11, with a standard deviation of 0.30.

The second, more crucial means for assessing the effect of intraspecific variation on phylogenetic conclusions is to determine whether using different individuals in interspecific comparisons produces any difference. Even if there is significant intraspecific polymorphism, most of it may have accumulated after the splitting of lineages and, therefore, may not affect the interspecific comparisons. An examination of the results in Table 2 indicates little or no effect of intraspecific differences on interspecific comparisons. Where the sex of the individual from which DNA was prepared is known (Table 1), there are two males and eight females represented. The sex of the individual makes no detectable difference, implying that Y-chromosomal DNA is not greatly affecting the results.

Phylogenetic Trees.—Various tree-building algorithms are applicable to distance data generated by DNA-DNA hybridization. We present here (Fig. 1) three of the more commonly used methods; UPGMA (Unweighted Pair-Group Method with Arithmetic Mean; Sokal and Michener, 1958) and the Fitch and Margoliash (1967) (or Cavalli-Sforza and Edwards [1967]) method, with and without the assumption of contemporaneous tips. These latter trees were determined by the KITSCH and FITCH programs of PHYLIP. These trees were based on the simple mean of the independent measures between species as given in Table 4. By "independent" we mean that different individuals were compared or that reciprocals of the same individuals were studied; i.e., each entry in Table 2 was considered to be independent. The important point is that all three methods give identical topology with virtually identical branch length.

A natural question arises as to whether all the nodes depicted in the trees in Figure 1 differ significantly from one another. Table 5 lists the independent measures (as defined above) through each node. Note that the measures comparing nodes overlap in only a single case: two values through nodes C and D, which overlap by 0.01 ΔT_m. Therefore, even conservative tests such as the nonparametric Mann-Whitney U test indicate with very high statistical confidence that these nodes are significantly different from one another. Thus, our results clearly support the previous conclusion reached by Sibley and Ahlquist (1984, 1987) that humans and chimpanzees are more closely related to one another than either species is to gorillas. The rest of the trees are in perfect agreement with the general consensus: orangutan is the next closest rel-

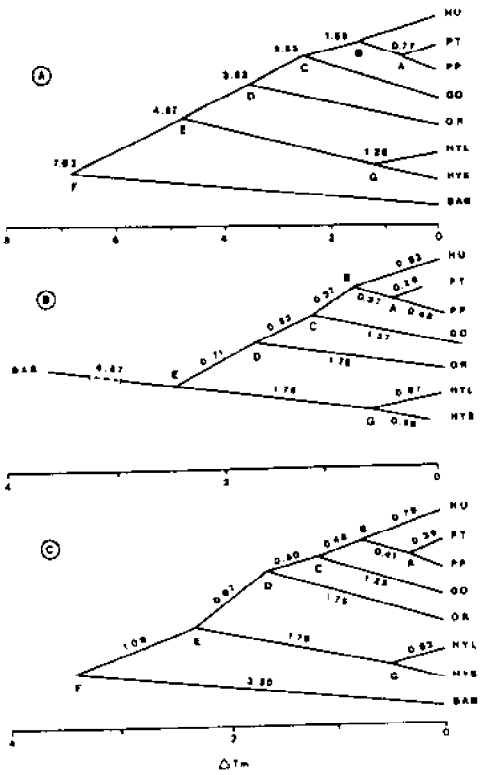

FIG. 1. Dendrograms based on three different algorithms: A) UPGMA (Sokal and Michener, 1958) tree, B) Fitch and Margoliash (1967) tree, not assuming contemporaneous tips; C) Fitch and Margoliash (1967) tree, assuming contemporaneous tips. In tree B, HU (human) was placed at zero, with all branch lengths proportional to the horizontal scale; vertical distances are arbitrary. This tree was "rooted" using the baboon (BAB) as the outgroup. Tree A has the ΔT_m's associated with each node indicated, while trees B and C have branch lengths indicated. The scale is in ΔT_m units, with the tree A spanning a scale twice as long as that for the other two trees. The abbreviations of the species are as indicated in Table 4, with the addition of BAB = *Papio hamadryas* (baboon). The numbers used to generate these trees were the means in Table 4.

TABLE 4. Mean ΔT_m's between species. These are the values used in the construction of the trees in Figure 1.

Species	Species						
	PP	PT	GO	OR	HYL	HYS	BAB
Homo sapiens (HU)	1.58	1.59	2.50	3.49	4.95	5.13	6.78
Pan paniscus (PP)	—	0.77	2.76	3.76	4.90	4.76	7.43
P. troglodytes (PT)		—	2.34	3.28	4.48	4.48	6.58
Gorilla gorilla (GO)			—	3.57	5.32	4.98	7.12
Pongo pygmaeus (OR)				—	4.77	4.88	7.33
Hylobates lar (HYL)					—	1.26	7.21
H. syndactylus (HYS)						—	6.74

ative, followed by gibbons, with the cercopithecines being the sister taxa of the Hominoidea.

Other Taxa.—There are two points not involving the relationships among humans, chimpanzees, and gorillas that are worth mentioning. First is the degree of genetic differentiation between the siamang, *Hylobates syndactylus*, and the common gibbon, *H. lar* (HYS vs. HYL in the tables). As can be seen, these species are more genetically distinct than the two named species of chimpanzees but somewhat less genetically different than are humans and chimpanzees.

A second question concerns the genetic distinctness of the populations or subspecies of orangutans that are native to Sumatra and Borneo. We do not know the origin of OR1, although it is thought to have come from Borneo (C. Sibley, pers. comm.). We do know that OR2 is from Borneo and that OR3 is from Sumatra. Assuming that OR1 is from Borneo, we have measured differences between the populations at ΔT_m's of 1.22 ± 0.11 and 0.84 ± 0.13 (Table 2). This indicates that the differentiation of these populations is about that measured between the different named species of chimpanzees. Certainly these populations are more genetically distinct than any of the other intraspecific pairs we have measured in the Hominoidea.

TABLE 5. Independent measures from Table 2 through each node indicated in Figure 1. With the exception of the values in bold type, there is no overlap among nodes.

Node	ΔT_m's
A	1.05, 0.97, 0.32
B	1.65, 1.62, 1.62, 1.82, 1.79, 1.55, 1.51, 1.36, 1.37, 1.92, 1.42, 1.39
C	2.40, 2.53, 2.75, 2.73, 2.39, 2.39, 2.69, 2.48, 2.52, 2.40, **2.97**, 2.61, 2.35, 2.24
D	3.42, 3.27, 3.72, 3.65, **2.96**, 3.20, 3.63, 3.63, 3.76, 3.17, 3.44, 3.86
E	5.13, 4.86, 4.52, 5.04, 5.11, 4.86, 4.95, 4.90, 4.76, 4.48, 4.44, 5.32, 4.92, 4.80, 4.77, 4.74
F	6.57, 7.21, 6.99, 7.43, 6.58, 7.12, 7.33, 6.74

DISCUSSION

Technical Considerations.—Before proceeding to a more general discussion of these results, we discuss some technical aspects. The first point to be considered is the accuracy with which the thermal stability of DNA molecules reflects the actual degree of base-pair mismatch. We have studied this problem in a direct manner by DNA-DNA hybridization of cloned DNA fragments of known sequence and, thus, with known degrees of base-pair mismatch (Caccone et al., 1988*b*). Our results showed a remarkable correspondence between ΔT_m as measured by our TEACL technique and the percentage base-pair mismatch: the correlation between ΔT_m and base-pair mismatch is 0.985 or greater. The second important result from these studies is that this relationship is linear at least over the range we studied, which was up to 7.2% mismatch. This is strong empirical evidence that the technique measures sequence divergence, that it is accurate, and that it is linear. This empirical result is in accord with a large body of theory reviewed in Britten et al. (1974).

Given that the technique works, what is the best statistic to describe the results? Three measures have been presented in the literature. One is the change in median melting temperature of the portion of the genomes that hybridized (the ΔT_m presented here). Other workers have incorporated the portion of the tracer genome that did not hybridize and have calculated a statistic to describe the difference in median melting temperature of the whole genome (that which hybridized and that which did not); this is the ΔT_{median} (Britten, 1986), the ΔT_{50}H (Sibley and Ahlquist, 1984, 1987), or the ΔT_mR (Benveniste, 1985). Finally, some authors use the modal temperature of melting, the ΔT_{mode} (Brownell, 1983). For many purposes, in particular for phylogenetic reconstructions, we argue that ΔT_m is the preferred statistic. First, as discussed above, we have empirically shown that this measure is extremely well correlated with the degree of base-pair mismatch, which is the thing we want to measure. Second, as discussed below, the measurement of the percentage reassociation is both difficult and somewhat arbitrary; it is this measurement, along with the measurement of ΔT_m, that provides the ΔT_{50}H. Thus, ΔT_{50}H has two sources of experimental error. Third, as we have shown for insects and birds (Caccone et al., 1988*a*) and as Benveniste (1985) has shown for vertebrates, there is a linear relationship between ΔT_m and ΔT_{50}H. Thus, for reconstruction of phylogenetic trees, the relative distance of the two measures will be nearly identical. This is true up to a certain distance, certainly up to a ΔT_m of 10, which is within the range of importance for the Hominidae. As for ΔT_{mode}, we simply point out that it has neither empirical nor theoretical justification. It is a less accurate statistic to measure, unless one resorts to curve-fitting. This would be necessary with the TEACL technique, because the primary data generate a cumulative curve. Thus, we conclude that ΔT_m is the best measure of the genetic differences measured by DNA-DNA hybridization when considering phylogenetic relationships over reasonably short distances. Knowing the percentage reassociation is certainly important and is added information; but incorporating this into the distance measure in order to reconstruct phylogeny may add more noise than information.

The problem of the measurement and arbitrariness of the percentage reassociation stems from two sources. First, the HAP and TEACL techniques measure reassociation in different ways. With HAP, the criterion is nonbinding to HAP, while with TEACL it is nuclease-S1 sensitivity. We have shown (Caccone et al., 1988*b* fig. 3) that for the same taxa, these two methods give somewhat different results (see also Hall et al. [1980]). The percentage reassociation is consistently smaller with the nuclease-S1 treatment, as this trims up overhanging ends and cuts out insertion/deletion differences. The somewhat arbitrary nature of determining the normalized percentage reassociation is due to the fact that it can vary, depending upon the stringency of the hybridization conditions and the conditions of digestion with nuclease-S1. This is shown in Benveniste (1985 table 6) for some of the same primates used in the present study. An important point in Benveniste's results is that, while the NPR varies as the conditions of hybridization vary, the ΔT_m does not. We will return to this point later. For now, it is sufficient to point out that, unless the percentage reassociation is determined under highly controlled conditions, using the same reassociation criteria, and using the same method to quantify reassociation, it is impossible to compare results from study to study. Thus, we are unable to convert our ΔT_m's into T_{50}H's in order to compare them with the ΔT_{50}H's of Sibley and Ahlquist (1984, 1987). The remarkable congruence of our ΔT_m's and Sibley and Ahlquist's (1984, 1987) ΔT_{50}H's (Table 6) is probably due to the fact that, in Sibley and Ahlquist's study, the NPR contributed relatively little to increasing ΔT_m's to generate ΔT_{50}H's. From a partial data set, Roy Britten (pers. comm.) has determined that the T_m's and T_{50}H's in the Sibley and Ahlquist (1984, 1987) data differ on average by only 0.2°C.

Despite all of the above, we need to examine the percentages of reassociation in our present data in more detail. In our previous work on insects (Caccone and Powell, 1987; Caccone et al., 1988*a*) we also found that the normalized percentage of reassociation was not a reliable measure for any single experiment. However, when averaged over many experiments the normalized percentages of reassociation became stable for a given range of ΔT_m, and the average ΔT_m and average normalized percentage reassociation showed a remarkably linear relationship (Caccone et al., 1988*a* fig. 3). Our hominoid data do not present as clear a pattern. We suspect that there are

TABLE 6. Comparison of results from the present study with the results of Sibley and Ahlquist (1984, 1987) and those of Goodman and colleagues (Miyamoto et al., 1987; Goodman et al., 1989) for sequence data on the $\psi\eta$-globin gene. Species abbreviations are as in Table 4, except 1) the two species of chimpanzees are combined (CH) as are the two species of gibbons (GIB) and 2) CER = various species of cercopithecines used in the different studies. The 1984 data were reported to one decimal place; when averaged over more than one species, a second decimal place may appear.

Species (tracer-driver)	$\Delta T_{50}H$ (Sibley and Ahlquist, 1984, 1987) 1984	1987	ΔT_m (this study)	Percentage divergence (Miyamoto et al., 1987; Goodman et al., 1989)
PT-PP	0.7	0.69	0.77	—
HU-CH	1.85	1.64	1.59	1.61
HU-GO	2.4	2.27	2.50	1.72
CH-GO	2.2	2.28	2.55	1.84
HU-OR	3.6	3.60	3.49	3.39
CH-OR	3.7	3.57	3.52	3.52
GO-OR	3.8	3.55	3.57	3.47
HU-GIB	5.2	4.76	5.04	—
CH-GIB	5.35	4.85	4.66	—
GO-GIB	5.4	4.69	5.15	—
OR-GIB	5.1	4.83	4.83	—
HU-CER	7.7	7.34	6.78	7.43
CH-CER	7.85	7.17	7.01	7.59
GO-CER	7.5	7.18	7.12	7.38
OR-CER	7.6	7.43	7.33	7.43
GIB-CER	7.4	7.05	6.98	—

two problems. First, mammalian genomes are about 30 times the size of insect genomes, so that reassociation kinetics are much more complex. Thus, accurate measurement of percentage reassociation is more difficult. A second problem, which we can actually detect in the data, is that our DNA preparations came from various sources using different purification procedures; this was not the case in our insect studies. For example, the samples from C. Sibley were stored in 0.48 M phosphate buffer; phosphate is a strong inhibitor of nuclease-S1. We attempted to remove the phosphate by dialysis and ethanol precipitation but may have been variously successful from sample to sample. Also, because the DNAs were in different concentrations, different hybrids required different lengths of incubation for the reassociation reaction; this varied from three to ten days. If various preparations contained different contaminants that slowly degraded the DNA during reassociation, this is another source of error.

Despite these problems, we can detect in our data patterns of reassociation that are in good agreement with previous studies. In our reassociation reactions, most of the DNA was driver. By the time tracer DNA was put into the reaction, it had undergone several additional steps which homogenize any differences among the initial sources from which the tracers were prepared; also, only a small fraction of the volume of the reassociation reaction was contributed by tracer. Thus, if DNA preparations affect percentage reassociation, we should see the effect most strongly for driver DNA. If we consider only reassociation reactions in which the DNA was prepared in the same manner in about the same concentrations (therefore requiring similar incubation times), much of the variation in normalized percentage reassociation should disappear. In our data set, ten of the DNA preparations were prepared the same way and were not stored for extended periods. These are the nine preparations prepared by us and the baboon DNA supplied by J. Rogers (Table 1). If we consider only the reassociation reactions in which these DNAs were drivers, we obtain the results in Figure 2. Benveniste (1985 fig. 2) presents these same data for primates and carnivores, and we include his linear relationship in our Figure 2. As can be seen, this set of "cleaner" data is in good agreement with the previous work on mammals. It is important to note that the Benveniste (1985) results are for nuclease-S1 resistance, as are ours.

While different sources of driver DNA may affect the normalized percentage reassociation, this factor does not seem to affect ΔT_m's. As noted previously, Table 2 indicates that, whether the driver DNA was prepared by us relatively recently or was obtained from various sources and had been stored for considerable time (in some cases, years), the resulting ΔT_m's are quite homogeneous. This is another argument in favor of using ΔT_m as the measurement of distance. We suspect that the stability of T_m in our system is due to two factors. 1) In determining the percentage reassociation, the sample is treated with nuclease-S1 and then passed over a sephadex G-100 column

(Caccone and Powell, 1987). This step removes degraded products and possibly some contaminants. 2) This reassociated fraction is then transfered to 2.4 M TEACL, further diluting any contaminants and homogenizing the final preparation on which the T_m is determined.

The next point we consider is the temperature correction due to differences in tracer lengths. Previous experiments and theoretical considerations have established that the relationship between the melting temperature and the length of the duplex can be expressed as observed $T_m + B/L$ where L is the length of the duplexes and B is a constant for the conditions of melting (Crothers et al., 1965; Hayes et al., 1970). Hall et al. (1980) determined empirically that in 2.4 M TEACL, $B = 500$. This was done by melting restriction fragments of known length. In addition to this justification for correction for fragment length, our own data indicate further justification. First, as pointed out in the Results and in our previous studies (Caccone and Powell, 1987; Caccone et al., 1988*a*) reciprocals are in much better agreement after temperature correction than before. Second, in our study of known base-pair mismatch and ΔT_m (Caccone et al., 1988*b*) our corrected ΔT_m's give a much better correlation with base-pair mismatch than do uncorrected Δt_m's. Thus, keeping in mind that what we want to measure is base-pair mismatch, this last fact should be ample justification for using the tracer-length temperature correction; in association with the previous empirical and theoretical studies, the evidence in favor of the correction is overwhelming.

Next, we consider the possibility of complications due to paralogous sequences, sequences which are homologous but which exist in the same genome, presumably due to gene duplication. We doubt that this is a serious problem with DNA-DNA hybridization. First, in dealing with reasonably closely related taxa, such as the hominoids, one would expect that most gene duplications occured before the separation of the taxa. Thus, if there is a complication due to paralogous hybridization, the effect should be the same for homoduplex and heteroduplex reactions; the ΔT_m would be unaffected. Another consideration is that, if the

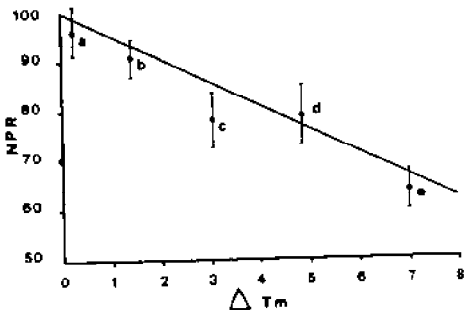

FIG. 2. Relationship between normalized percentage reassociation (NPR) and ΔT_m for selected data; see text for explanation. Definitions of points: a = the mean of all intraspecific hybrids; b = the mean for interspecific hybrids up to a ΔT_m of 2; c = the mean over the range of 2–4; d = the mean over the range of 4–6; and e = the mean over the range of 6–8. Bars indicate two standard errors of the mean. The line is that obtained by Benveniste (1985), using the nuclease-S1 criterion for primates and other mammals.

duplications predate the splitting, even though paralogous pairing may occur in theory, orthologous pairing would be thermodynamically favored. If, on the other hand, the paralogous sequences were highly similar to one another, then they would behave like true repetitive sequences and be removed in the tracer preparation. Finally in this context, we note again the results of Benveniste (1985), showing that one can increase the stringency of reassociation and lower the percentage reassociation but not affect the ΔT_m. It is possible that some of the lowering of the percentage reassociation is due to relatively weakly pairing paralogous sequences; but, if this fraction is removed from homoduplex and heteroduplex reactions to the same extent, the ΔT_m is unaffected.

There are a few other technical aspects concerning DNA-DNA hybridization that have been questioned (e.g., Marks et al., 1988). First, we used a low-energy radioisotope (tritium), so damage due to high-energy disintegration is not a problem. Second, we used a labeling procedure that does not harm the DNA and that specifically labels only single-stranded DNA. Thus, any double-stranded DNA that might have passed through the hydroxyapatite column would not be labeled. Third, we have taken into consideration both the tracer length and

the effect of base composition. Thus, our procedures have taken into account virtually all of the criticisms aimed at the HAP technique, specifically as applied by Sibley and Ahlquist (1984, 1987). Since we obtained results nearly identical to those of Sibley and Ahlquist when we studied the same taxa, it would seem that these possible technical problems do not significantly affect the results. Furthermore, we argue that the similarilty of our results to those previously obtained not only reinforces the conclusions concerning the phylogeny of the particular group studied, but also increases confidence in the technique itself: independent studies of the same taxa yield nearly identical results.

Finally in this section we discuss statistical procedures. The analysis of data generated by DNA-DNA hybridization is far from straightforward, primarily because there are inevitable correlations in the data matrix. This is due to the fact that a single homoduplex T_m determination is used in a number of ΔT_m determinations. Thus, not all ΔT_m's in a matrix such as in Table 2 are truly independent. Felsenstein (1987) has dealt with this problem in some detail for Sibley and Ahlquist's (1987) results and concluded that the topology claimed (the same as in our Fig. 1) was statistically defensible. Because our data are so similar to Sibley and Ahlquist's, they would almost certainly be at least as statistically robust. We have used more individuals, which would represent a greater number of truly independent measures. Because our data are so clear, with virtually no overlap in measurements between adjacent nodes (Table 5), we have simply applied conservative tests such as the Mann-Whitney U test and obtained results indicating that the distributions in Table 5 are highly significantly different from one another. We have studied the distribution of the errors associated with our T_m determinations and found them not to deviate from a normal distribution, indicating that parametric tests would probably be valid for our data. These would only be more powerful and would support the results of the nonparametric tests.

General Considerations.—Virtually all molecular phylogenetic studies, including these, have a major underlying assumption: the genetic similarity or difference among taxa is an indication of phylogenetic relatedness. Lineages that diverged more recently will be genetically more similar to one another than will be lineages with more ancient splits. This assumption cannot be tested with absolute rigor, as we can never be absolutely certain about the correct phylogeny of any group, just as we cannot be certain about any historical event. However, two facts lend powerful credence to the assumption. First, virtually all molecular studies are in perfect accord with phylogenies deduced by other methodologies. For example, the molecular data for the hominoids are in excellent agreement with all other evidence indicating that humans and African apes are each other's closest relatives, followed by orangutans, gibbons, and the Old World monkeys. It is only in the details of particularly difficult clades, such as the human-African ape clade, that molecular and more traditional data may appear to conflict.

In addition to the empirical observations in support of the assumption stated above, there is a large body of theory to support it as well. This is population genetics. Over a wide range of models incorporating a large number of parameters and assumptions, there is one common result: isolated populations (=species) will become genetically differentiated directly in proportion to time. This is true whether mutation, selection, genetic drift, or, most likely, some combination of these forces is driving evolution. One need not be a neutralist to accept the assumption of molecular phylogenetics. One could in theory contrive a population-genetics model wherein isolated populations will not diverge with time, but it would likely be either trivial or totally unrealistic.

If one accepts this assumption, the next consideration is to assess methodologies to measure genetic differentiation of lineages. Diverse methods are available, which produce data more or less useful in different situations. For most purposes, we argue that measurements of DNA divergence are the best indication of phylogenetic affinities. First, DNA is a phylogenetically continuous character; all (eukaryotic) organisms have DNA, and it is directly inherited from parent to offspring. Second, DNA is probably the phylogenetically continuous trait that is least affected by selection. This is not to imply DNA is not subject to any selective forces, which would be an absurd view. Rather, compared to proteins, morphology, physiology, etc., DNA is less exposed to selection on average, because it is furthest removed from the phenotype. Characters more exposed to selection can be misleading in phylogenetic reconstruction, due to convergence and parallelism. Third, because of the large genome sizes of eukaryotes, DNA can have virtually an infinite number of character states. With four possible bases at each of millions or billions of positions, the complexity and information content of DNA are enormous. We recognize that DNA also has some problems, such as the facts that it undergoes insertions and deletions and that back mutations are not unlikely, since there are only four bases. However, while these complicating factors must be considered, they do not negate the favorable properties of DNA. (For an argument that amino-acid sequences may be more accurate if one wishes to invoke a molecular clock, see Zuckerkandl [1987].)

Given the problematic situation posed by the human and African ape lineage, what approaches to DNA studies are most useful for clarifying the relationships? As pointed out in the Introduction, the problems are that the lineages split fairly closely in time and that they did so quite recently. In order to increase the phylogenetically informative changes, two strategies are possible. First, one could study a large portion of the genome to detect as many differences as possible. Second, one could study in detail a part of the genome that is evolving unusually fast. Both approaches have been used in the hominoids, and they arrive at the same conclusion.

The first approach, to study a large portion of the genome, is the strategy of DNA-DNA hybridization, which measures DNA divergence over virtually the total single-copy genome. This method has been applied in an extensive manner to hominoids using two different techniques, the HAP technique used by Sibley and Ahlquist (1984, 1987) and the TEACL technique reported here. Table 6 compares the outcome of these studies. Clearly, there is remarkable agreement between the two sets of HAP data and the TEACL measurements. This shows that when the technique of DNA-DNA hybridization is applied to the same taxa, even when using somewhat different methods, similar results are obtained. We discussed above the matter of the different measures reported in Table 6 and concluded that this is a moot point. These two measures are linearly related, and the two sets of data display identical relative distances. All phylogenetic reconstructions based on either set yield identical branching patterns and virtually identical relative branch lengths.

The approach used to study a particularly rapidly evolving section of the genome has been Goodman and colleagues' study of the pseudogene for the η-globin (abbreviated $\psi\eta$-globin) of the β-globin cluster (Miyamoto et al., 1987, 1988; Goodman et al., 1989). This region is rapidly evolving because it is probably under no selective constraints and, thus, is an ideal candidate for the problem presented by the hominoids. Table 6 includes a summary of the results. Again, good agreement is evident as to the relative divergence of the taxa; good absolute agreement is obtained if one assumes that ΔT_m is converted to percentage base-pair mismatch in a 1:1 manner. The one aspect of the $\psi\eta$-globin data that differs somewhat from the hybridization results concerns the difference between humans and gorillas and that between chimpanzees and gorillas, crucial measurements for the splitting of the trichotomy. Fortunately, because sequence data are richer in information than are distance data generated by hybridization studies, they can be further analyzed using a variety of techniques. When this was done with these data, the phylogeny most favored is the same as that favored by results from DNA-DNA hybridization (Miyamoto et al., 1987, 1988; Goodman et al., 1989). In a parsimony analysis, breaking the association of humans and chimpanzees requires eight additional mutations. A cladistic analysis indicates that there are 11 synapomorphies favoring human-chimpanzee, while only three favor either human-gorilla or chimpanzee-gorilla as the most closely related pair. Felsenstein's (1981) likelihood method indicates that the human-chimpanzee association is nine times more likely

than the alternative two associations. Similar analyses of the sequence data for the other taxa strongly support the consensus phylogeny.

Another source of rapidly evolving DNA sequences is mitochondrial DNA (mtDNA). Several initial analyses of mtDNA data were unable to reject the trichotomy for human-chimpanzee-gorilla (reviewed in Holmquist et al. [1988*a*]). However, more recent analyses using a variety of statistical models have tended to support the human-chimpanzee association. These include the Markov model of Lanave et al. (1986), the Markov-maximum-likelihood model of Hasegawa et al. (1985), and the "Lake method" (Holmquist et al., 1988*b*).

If, as argued by Zuckerkandl (1987), amino-acid replacements provide a better "clock" than DNA base changes, it is of interest to inquire what evidence may be obtained from protein sequences. Goodman et al. (1983, 1989) have reviewed the evidence from nine protein chains that have been sequenced in both humans and chimpanzees and noted 0.4% divergence for the 1,271 amino-acid positions. Somewhat fewer data are available for gorillas, but the average difference between gorillas and either humans or chimpanzees is 0.7%. It is interesting to note that the relative distances indicated by these amino-acid data and the DNA-DNA hybridization data are nearly the same. The ratios of the distance between human and gorilla to the distance between human and chimpanzee are 1.7 and 1.6, respectively.

On technical or statistical grounds, one may question the strength of any single source of molecular data bearing on the hominoid problem. However, there is a clear consensus emerging as more data accumulate. The data from DNA-DNA hybridization clearly support the human-chimpanzee association to a high degree of statistical certainty and also strongly support the rest of the consensus tree. As more DNA sequence data for nuclear genes (such as the $\psi\eta$-globin gene) have been obtained, the case has become stronger that this type of information is in total agreement with the hybridization results. The mtDNA data are now shown to be compatible with the other two data sets. Finally, amino-acid replacement data support the DNA-based data.

We submit that the problem of the relative genetic relatedness of the hominoids has been largely solved. Humans and chimpanzees are genetically most close, followed in order by gorillas, orangutans, gibbons, and the Old World monkeys. If one accepts the assumption of molecular phylogenetics, we argue that the problem of the phylogeny of the extant species has also been solved. This is not to imply that further molecular studies of these species will not lead to important insights. More information on rates and patterns of molecular evolution are important for their potential to provide further confidence in the absolute genetic differences among hominoids and their primate relatives. However, the evidence is strong enough to predict with reasonable confidence that the topology of the phylogenetic tree of the hominoids will not prove to be different from those in Figure 1.

What then of the morphological evidence? Generally morphological evidence has been invoked to support a chimpanzee-gorilla clade with humans as the sister group. However, the evidence is not unambiguous, and a human-chimpanzee group could also be supported (Groves, 1986). Two of the major traits primatologists have relied on in support of a chimpanzee-gorilla lineage are thin tooth enamel and knuckle-walking, which are apparently synapomorphic for the gorilla-chimpanzee clade. However, humans may be considered to be a highly derived, or autapomorphic, species, and it is certainly plausible that humans have simply lost these characters (D. Pilbeam, pers. comm.). This leads to the hypothesis of the presence of thin tooth enamel and knuckle-walking in the ancestors of the lineage containing humans and African apes, a testable hypothesis. If one discounts these two traits, the remaining morphological evidence is certainly compatible with the molecular data and may actually support it.

The molecular-clock hypothesis can be invoked to use molecular data to put absolute times on lineage splits. We do not intend to discuss this in detail, as an enormous literature exists. We wish, however, to make two points. First, given that our

results are so similar to those previously obtained by Sibley and Ahlquist (1984, 1987) and that Felsenstein's (1987) statistical analysis could not reject the molecular-clock hypothesis for their data, the same is likely to be true for our results. This can be seen in Figure 1B. This tree was built placing *Homo* at zero, without assuming the molecular clock. If the clock were perfect and our data were perfect, all tips would line up on zero. This is indeed the case, with the possible exception of tips representing the gorilla and the pygmy chimpanzee. However, the distance between these species is small and would not present a serious challenge to the overall clocklike nature of the data set. This result stands in contrast to that obtained from protein-coding regions which appear to have undergone a slowdown in the lineage containing humans and African apes (Goodman et al., 1983, 1989; Li and Tanimura, 1987). Results from DNA-DNA hybridization experiments pertain to the total single-copy genome, most of which is not coding for protein sequences.

The second point is that the paleontological data are compatible with the clocklike nature of our results. The two best established dates within the hominoids are for their split from the cercopithecines, placed at 25–30 million years ago, and the divergence of orangutans, 12–16 million years ago (Gingerich, 1984; Pilbeam, 1986). Our data for the distance from baboons to the hominoids ($\Delta T_m = 6.78$) is almost exactly double the distance at which orangutans diverged ($\Delta T_m = 3.53$). Humans split from chimpanzees at just under half the distance, as did orangutans; thus, if the clock remains linear, our data indicate that humans diverged from chimpanzees about 6–8 million years ago.

Initial attempts to apply a molecular clock to the emergence of the human lineage led to what was considered to be a radical conclusion: the human lineage diverged from the African apes five million years ago (Sarich and Wilson, 1967) or possibly as recently as three million years ago (Sarich and Cronin, 1976). This was at a time when anthropologists interpreted the fossil record as indicating that the minimum age of the human lineage was 12–15 million years. Since then, a reassessment of the fossil record has led anthropologists to estimate a more recent date, and as more molecular data have become available, the molecular-clock estimates have led to an older date, such that the two sets of data are now compatible (Ruvolo and Pilbeam, 1986).

Acknowledgments

We thank the individuals and institutions listed in Table 1 for supplying blood samples or DNA preparations. J. Marks was also helpful in obtaining samples and in encouraging us to undertake this project. G. Amato and K. Goddard assisted in the laboratory work. We thank M. Goodman, J. Marks, M. E. Ruvolo, and C. Sibley for comments and helpful suggestions concerning the manuscript. Financial support was provided by the National Science Foundation (BSR 8500207) and the National Institutes of Health (GM 36648). This manuscript was written while J.R.P. was on leave at the II Università di Roma; we thank the Dipartimento di Biologia, especially V. Sbordoni, for hospitality.

Literature Cited

Andrews, P., and L. Martin. 1987. Cladistic relationships of extant and fossil hominoids. J. Hum. Evol. 16:101–118.

Benveniste, R. E. 1985. The contributions of retroviruses to the study of mammalian evolution, pp. 359–417. *In* R. J. MacIntyre (ed.), Molecular Evolutionary Genetics. Plenum, N.Y.

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.

Bishop, M. J., and A. E. Friday. 1985. Evolutionary trees from nucleic acid and protein sequences. Proc. Roy. Soc. Lond. B 226:271–302.

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

Britten, R. J., A. Cetta, and E. H. Davidson. 1978. The single-copy DNA sequence polymorphism of the sea urchin *Strongylocentrotus purpuratus*. Cell 15:1175–1186.

Britten, R. J., D. E. Graham, and B. R. Neufeld. 1974. Analysis of repeating DNA sequences by reassociation, pp. 363–418. *In* L. Grossman and K. Moldave (eds.), Methods in Enzymology, Vol. 29e. Academic Press, N.Y.

Brownell, E. 1983. DNA/DNA hybridization studies of muroid rodents: Symmetry and rates of molecular evolution. Evolution 37:1034–1051.

Caccone, A., G. D. Amato, and J. R. Powell. 1987. Intraspecific DNA divergence in *Drosophila*: A study