J Mol Evol (1995) 40:622–628

JOURNAL OF MOLECULAR EVOLUTION

Improved Dating of the Human/Chimpanzee Separation in the Mitochondrial DNA Tree: Heterogeneity Among Amino Acid Sites

Jun Adachi,[1] Masami Hasegawa[1,2]

[1] Department of Statistical Science, The Graduate University for Advanced Studies, 4-6-7 Minami-Azabu, Minato-ku, Tokyo 106, Japan
[2] The Institute of Statistical Mathematics, 4-6-7 Minami-Azabu, Minato-ku, Tokyo 106, Japan

Received: 4 May 1994 / Accepted: 20 October 1994

Abstract. The internal branch lengths estimated by distance methods such as neighbor-joining are shown to be biased to be short when the evolutionary rate differs among sites. The variable-invariable model for site heterogeneity fits the amino acid sequence data encoded by the mitochondrial DNA from Hominoidea remarkably well. By assuming the orangutan separation to be 13 or 16 Myr old, a maximum-likelihood analysis estimates a young date of 3.6 ± 0.6 or 4.4 ± 0.7 Myr (±1 SE) for the human/chimpanzee separation, and these estimates turn out to be robust against differences in the assumed model for amino acid substitutions. Although some uncertainties still exist in our estimates, this analysis suggests that humans separated from chimpanzees some 4–5 Myr ago.

Key words: Mitochondrial DNA — Hominoidea — Molecular clock — Maximum likelihood — Site heterogeneity

Introduction: Problems Inherent in the Previous Estimates of Branching Dates

Although molecular phylogenetics has established that the human/chimpanzee separation is younger than 10 Myr, there is still a wide range of variation in the estimate made by researchers, which depends on the data and the method they use (Sarich and Wilson 1967a; Andrews and Cronin 1982; Sibley and Ahlquist 1984, 1987; Hasegawa et al. 1987, 1990; Ueda et al. 1989; Kishino and Hasegawa 1990; Gonzalez et al. 1990; Hasegawa 1991; Bailey et al. 1992).

Recently, Horai et al. (1992) determined 4.8 kbp of mitochondrial DNA (mtDNA) sequences from common chimpanzee (*Pan troglodytes*), pygmy chimpanzee (bonobo; *Pan paniscus*), gorilla (*Gorilla gorilla*), orangutan (*Pongo pygmaeus*), and siamang (*Hylobates syndactylus*). The sequences cover genes coding for ND2, COI, COII, ATPase 8, and 11 tRNAs and partially cover genes for ND1 and ATPase 6. Since mtDNA evolves much more rapidly than nuclear DNA (Brown et al. 1982), these data together with the corresponding sequences of human (*Homo sapiens*) (Anderson et al. 1981) should contain more information than the nuclear DNA data published to date for the purpose of elucidating the phylogenetic place of humans within Hominoidea.

From these sequences, Horai et al. established that the closest relatives of the human are the two chimpanzees rather than the gorilla, in accord with the earlier works (Sibley and Ahlquist 1984, 1987; Miyamoto et al. 1987; Kishino and Hasegawa 1989; Caccone and Powell 1989; Sibley et al. 1990; Ruvolo et al. 1991). By assuming the orangutan separation to be 13 Myr ago, they further estimated the dates of branchings within the African apes/human clade. From the data set that consists of the tRNAs and first and second codon positions (their DATA1), they estimated the human/chimpanzee separation to be 4.3 and 5.6 Myr ago, respectively, by the

Correspondence to: M. Hasegawa

maximum-likelihood (ML) method for DNA phylogeny (Felsenstein 1981; the DNAML program in Felsenstein's package PHYLIP) and the neighbor-joining (NJ) method (Saitou and Nei 1987). They noted that the ML method gave shorter divergence times than the NJ, and they attributed the difference to the problematic synonymous changes in *Leu* codons. It is likely that synonymous changes at the first positions of *Leu* codons have substantial effects on the estimation, as they thought. But this must be a problem not only in the ML but also in the NJ method, and hence this does not explain the difference in the estimates between the two methods.

We think that the difference in the estimates is due to a defect of the NJ in estimating branch lengths as will be shown later. Because of the problem of *Leu* codons, Horai et al. excluded synonymous transition in the first codon positions. They included synonymous transversions in the third codon positions. They applied the NJ method to this data set (their DATA3; the DNAML program cannot be applied to such a data set). Their estimate of the human/chimpanzee separation was 4.7 ± 0.5 Myr (±1 SE). They attributed a younger estimate of 3.9 ± 0.7 Myr for this separation by Hasegawa et al. (1990) to the relatively small region compared (896 bp of Brown et al. 1982).

Hasegawa et al.'s (1990) estimate was done by classifying sites into two classes—third codon positions and the remainder—and suffers from the problem of synonymous changes at the first positions of *Leu* codons. Therefore, we admit that Hasegawa et al.'s estimate should be reexamined by an improved method with more abundant data. This does not necessarily mean that Horai et al.'s (1992) estimate is the most reliable one to be made from their data.

In Horai et al.'s data set DATA3, they included nonsynonymous differences and synonymous transversion differences in protein-encoding genes, and all differences in tRNA genes. They considered that the differences between species under consideration were small enough to be far from the saturation level, and hence they did not take account of multiple substitutions in a site in their NJ analysis. Since the number of differences between even the most distant pair is only a small fraction of the total number of sites, the multiple-hit correction should be negligibly small by conventional formulas such as of Jukes and Cantor (1969) and of Kimura (1980), and therefore their procedure might seem to be justified at a first glance.

Actually, however, variability differs among sites (even among nonsynonymous sites), and all the sites under consideration are not equally variable (Fitch and Markowitz 1970; Hasegawa et al. 1985; Reeves 1992; Sidow et al. 1992). Although the human/chimpanzee clade has been firmly established for the 4.8-kbp data of Horai et al. (1992), there are still many sites in DATA3 that support other branchings by the parsimony principle, indicating multiple substitutions in these sites. Such a multiple-hit effect has not been taken into account in their NJ analysis, while it can be taken into account to some extent by the ML analysis, as will be shown later. Since the multiple-hit effect is more serious in a longer branch than in a shorter one, their dating of the human/chimpanzee branching could be biased to be older. We attribute this effect to the cause of the difference of the estimates between the NJ and ML methods for DATA1.

Since a more realistic model is available for amino acid substitutions than for nucleotide substitutions in protein-encoding genes (Kishino et al. 1990; Adachi and Hasegawa 1992), we now reexamine their data by the ML method at the amino acid sequence level, taking account of the heterogeneity of rate among amino acid sites.

Modeling of Amino Acid Sequence Evolution

Comparison Between the ML and NJ Methods in Estimating Branch Lengths

All phylogenetic inferences depend on their underlying models. To have confidence in inferences, it is necessary to have confidence in the models (Goldman 1993). Adachi and Hasegawa (1992) published the PROTML program for the ML inference of protein phylogeny based on the Dayhoff model (Dayhoff et al. 1978), and it has been used widely. Subsequently, it has turned out that this model is far more appropriate than the Proportional and Poisson models (Hasegawa et al. 1992) for approximating the evolution of the diverse protein data (Hasegawa et al. 1993a; Adachi et al. 1993; Hashimoto et al. 1993, 1994). Recently, Jones et al. (1992) updated the amino acid substitution matrix by using about 40 times more abundant substitution data than those of Dayhoff et al. (1978). The new version of PROTML (version 2.2) allows us to use this model (called the JTT model) as well as the Dayhoff, Proportional, and Poisson models, and it has turned out that the JTT model better approximates the evolution of diverse proteins than the Dayhoff model, except for globins (Cao et al. 1994a).

Both the Dayhoff and the JTT models assume the averaged amino acid frequencies of the proteins that were used in estimating the respective substitution matrices as the equilibrium frequencies. However, the amino acid frequencies of the individual protein species under analysis generally differ from those of the average one, and hence it might be better to use the actual amino acid frequencies of the protein under analysis as the equilibrium frequencies. The new version of PROTML (version 2.2) allows us to use this option for the JTT, Dayhoff, and Poisson models (the "F" option; the Proportional model corresponds to the F option of the Poisson model). When it was applied to mtDNA-encoded proteins of tetrapods, it turned out that, among the alter-

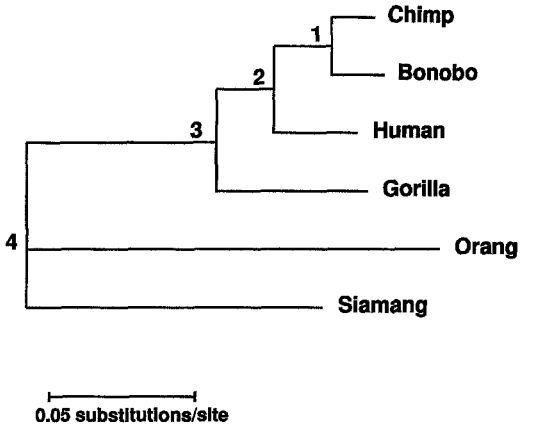

Fig. 1. The ML tree of the mtDNA-encoded proteins based on the JTT-F model. The horizontal length of each branch is proportional to the estimated number of substitutions. The root of this tree is located somewhere within the 4-siamang branch. Among several models implemented in the PROTML program (version 2.2), which assume homogeneity among sites, the JTT-F model best approximates the data.

native models, the JTT-F model best approximates the evolution of all the 13 proteins encoded by mtDNA (Cao et al. 1994b).

In this work, the JTT-F, JTT, and Poisson models were used. The Akaike Information Criterion defined by AIC = $-2 \times$ (log-likelihood) $+ 2 \times$ (number of parameters) is one of a number of model selection criteria used in statistics. A model that minimizes AIC is considered the most appropriate model (Akaike 1974). For the purpose of comparisons with the ML, we also used the NJ method. The distances estimated by the PROTML for two-species trees based on the respective models were used in the NJ analyses.

The following protein-encoding regions in Horai et al.'s (1992) and Anderson et al.'s (1981) data were used in this work: ND1 (4123–4260 in the numbering of Anderson et al.), ND2 (4470–5510), COI (5904–7442), COII (7586–8266), ATPase 8 (8366–8524), ATPase 6 (8575–9024, overlapping region with ATPase8 region 8525–8574 was excluded). The total number of deduced amino acid sites was 1344.

Figure 1 shows the ML tree estimated from the JTT-F model assuming homogeneity across sites. The left-hand side of Table 1 gives the branch lengths estimated by the NJ and ML methods based on the JTT-F, JTT, and Poisson models that assume site homogeneity. It is apparent that, although the terminal branch lengths do not differ systematically between the NJ and ML methods, the internal branch lengths estimated by the NJ are consistently shorter than those by the ML. This is particularly true for the two most internal branches, 4–3 and 3–2, for which the ratios of NJ to ML estimates are nearly 0.7–0.8. This discrepancy between the two methods can be attributed to the fact that the multiple substitutions are underestimated in the NJ method because it does not take account of the states of the internal nodes.

Table 2 gives numbers of differences in the 1,344 amino acid sites. The difference between siamang and orangutan is significantly larger than those between siamang and the members of the African apes/human clade. Furthermore, the differences between orangutan and the African apes/human are even larger than those between siamang and the African apes/human. Since the siamang is highly likely to be the outgroup to all the other species used in this analysis (Hayasaka et al. 1988; Hasegawa et al. 1990), these differences indicate that the evolutionary rate in the orangutan lineage accelerated relative to the African apes and human lineages, as suggested by Horai et al. (1992). Except for this violation of the molecular clock, the relative rate tests (Sarich and Wilson 1967b; Hasegawa et al. 1987) at the amino acid level do not suggest any rate variation among chimpanzee, bonobo, human, and gorilla, which allows molecular clock analyses of these data.

From the estimates of branch lengths, we estimated branching dates by the following procedure, similar to Horai et al.'s. The depth of a node (numbered 1–4 as in Fig. 1) from tips was estimated as follows from branch lengths represented as l_{XY} between X and Y (either nodes or tips):

$$d_1 = (l_{1C} + l_{1B})/2 \quad (1)$$

$$d_2 = (l_{2H} + l_{21} + d_1)/2 \quad (2)$$

$$d_3 = (l_{3G} + l_{32} + d_2)/2 \quad (3)$$

$$d_4 = l_{43} + d_3 \quad (4)$$

Since the rate in the orangutan lineage is higher than in other lineages, l_{4O} was not used in estimating d_4. Assuming 13 Myr for node 4 (Pilbeam 1988; Andrews 1992; McCrossin and Benefit 1993), dates of the other nodes are estimated by

$$t_i = (d_i/d_4) \times 13 \ (i = 1, 2, \text{and } 3) \quad (5)$$

The human/chimpanzee separation is estimated to be 4.4 and 3.7 (or 3.6 for the JTT-F model) Myr old, respectively, by the NJ and ML methods, when rate homogeneity among sites is assumed. The older estimate by NJ than that by ML is due to the underestimate of the internal branch lengths by NJ. The JTT-F model has turned out to be the best among the alternative models in approximating the data, but the estimated branch lengths and the divergence dates are almost the same among different models as long as the site homogeneity is assumed. We shall examine the Poisson model in further detail because of its simplicity.

Heterogeneity Among Sites in the Evolution of Amino Acid Sequences

The left-hand side of Table 3 shows a comparison of the observed distribution of configurations of amino acid

Table 1. Branch lengths (numbers of subtitutions per 100 amino acids) and branching dates estimated from the amino acid sequences of mtDNA encoded proteins by the NJ and ML methods[a]

	Homogeneous							Heterogeneous Poisson		
	JTT-F		JTT		Poisson					
	NJ	ML	NJ	ML	NJ	ML	NJ/ML	NJ	ML	NJ/ML
Terminal branch										
l_{1C} (1–chimp)	0.78	0.72	0.79	0.72	0.79	0.71	1.12	0.83	0.72 ± 0.24 (0.24)	1.15
l_{1B} (1–bonobo)	0.85	0.91	0.86	0.93	0.85	0.93	0.91	0.85	0.94 ± 0.28 (0.28)	0.91
l_{2H} (2–human)	1.43	1.43	1.45	1.51	1.43	1.38	1.03	1.46	1.41 ± 0.37 (0.36)	1.04
l_{3G} (3–gorilla)	2.36	2.58	2.37	2.57	2.34	2.38	0.98	2.50	2.48 ± 0.48 (0.49)	1.01
l_{4O} (4–orang)	6.41	6.96	6.40	6.91	6.26	6.82	0.92	7.70	7.75 ± 0.88 (0.87)	0.99
l_{4S} (4–siamang)	4.96	4.92	4.97	5.00	4.88	4.89	1.00	5.74	5.35 ± 0.73 (0.72)	1.07
Internal branch										
l_{21} (2–1)	0.82	0.94	0.83	0.93	0.84	1.00	0.84	0.92	1.02 ± 0.32 (0.31)	0.91
l_{32} (3–2)	0.80	0.97	0.81	0.97	0.81	1.14	0.71	0.91	1.15 ± 0.37 (0.36)	0.79
l_{43} (4–3)	2.20	3.14	2.22	3.16	2.20	3.06	0.72	2.73	3.23 ± 0.59 (0.59)	0.85
Branching date (Myr)										
t_1 (chimp/bonobo)	2.33	1.85	2.34	1.86	2.35	1.90	1.24	2.08	1.84 ± 0.43 (0.46)	1.13
t_2 (human/chimp)	**4.38**	**3.63**	**4.40**	**3.69**	**4.41**	**3.70**	1.19	**4.00**	**3.60** ± 0.58 (0.70)	1.11
t_3 (human/gorilla)	6.70	5.86	6.71	5.85	6.70	5.92	1.13	6.22	5.83 ± 0.72 (0.99)	1.07
t_4 (human/orang)	13	13	13	13	13	13		13	13	
ln L		−5,510.6		−5,741.7		−6,144.9			−5,747.7	
df		28		9		9			10	
AIC		11,077.2		11,501.4		12,307.8			11,515.4	
ΔAIC		0		424.2		1,230.6			438.2	

[a] The homogeneous model assumes that all 1,344 amino acid sites are equally variable. The heterogeneous model assumes that some portion of the sites are invariable and the remainings are equally variable. Branch lengths are represented as the averages of all sites irrespective of variable or invariable. ± refers to 1 SE estimated by replicating bootstrap resampling (Felsenstein 1985) (1,000 replications). The SEs estimated from the curvature of likelihood surface (given by PROTML) are shown in parentheses. Log-likelihood for the heterogeneous Poisson model is given by $\ln L = \ln L_{var} - (\text{number of invariable sites}) \times \ln 20$, where $\ln L_{var}$ is the total log-likelihood for the variable sites. *df* refers to a degree of freedom of the model for the ML method. For the JTT and Poisson models, nine branch lengths are estimated; for the JTT-F model, 20 amino acid frequencies are additionally estimated under the constraint that the summation is 1 (additional 19 *df*); and for the heterogeneous model, the fraction of variable sites is estimated (additional 1 *df*).

Table 2. Numbers of amino acid differences in the 1,344 sites of mtDNA-encoded proteins of Hominoidea

	Orang	Gorilla	Human	Chimp	Bonobo
Siamang	142	121	116	127	123
Orang		138	139	141	136
Gorilla			61	61	64
Human				39	43
Chimp					22
Bonobo					

sites with that expected from the homogeneous Poisson model. The fitting of the model to the data is terribly bad ($\chi^2 = 116.27$ with 10 *df*) as was pointed out by Reeves (1992) for the mtDNA-encoded proteins of tetrapods. This may be attributed to the facts that not all sites are equally variable and that some of the sites are invariable due to functional constraints. Therefore, we assume that some portion of the sites is invariable and that the remaining sites are equally variable (Hasegawa et al. 1985; Hasegawa and Horai 1991). When this heterogeneous Poisson model is applied, the fraction of variable sites turns out to be 372/1,344 = 0.277, and the fitting to the data improves drastically ($\chi^2 = 3.59$ with 9 *df*) (Table 3). Consequently, the AIC of the heterogeneous Poisson model improves over that of the homogeneous Poisson model (Table 1). The estimates of branching dates by ML remain almost unchanged by this improvement of the model, while those estimated by NJ become nearer those estimated by ML (Table 1).

A combination of the heterogeneous model and the JTT-F model should further improve the fit to the data, but we did not take this approach because of the ambiguity in removing sites with this model. The variable-invariable classification is only an approximation, and the rate variation among sites must be more continuous (Kocher and Wilson 1991; Yang 1993; Tamura and Nei 1993). Nevertheless, it is clear that the ML estimates of the branching dates would remain almost unchanged by these further improvements of the model. It is noteworthy in Table 1 that branch lengths estimated by ML are

Table 3. Distribution of configurations of amino acid sites for the homogeneous and heterogeneous Poisson models (ML estimates)[a]

Configuration	Number of changes	Homogeneous model			Heterogeneous model		
		Observed	Expected	$\frac{(\text{Obs-Exp})^2}{\text{Exp}}$	Observed	Expected	$\frac{(\text{Obs-Exp})^2}{\text{Exp}}$
(C,B,H,G,O,S)	0	1128	1074.4	2.67	156	156.0	—
(C,B,H,G,S)(O)	1	53	76.1	7.01	53	50.5	0.12
(C,B,H,G,O)(S)	1	39	54.2	4.26	39	33.6	0.86
(C,B,H,G)(O,S)	1	20	33.7	5.57	20	20.0	0.00
(C,B,H,O,S)(G)	1	11	26.0	8.65	11	14.7	0.94
(C,B,G,O,S)(H)	1	5	15.0	6.67	5	8.2	1.24
(C,B,H)(G,O,S)	1	7	12.4	2.35	7	6.8	0.01
(C,B)(H,G,O,S)	1	5	10.9	3.19	5	5.9	0.13
(C,H,G,O,S)(B)	1	6	10.1	1.66	6	5.4	0.06
(B,H,G,O,S)(C)	1	5	7.7	0.95	5	4.1	0.19
Others	≥2	65	23.5	73.29	65	66.7	0.04
Total		1,344	1,344.0	$\chi^2 = 116.27$ $df = 10$ $P \ll 0.00001$	372	372.0	$\chi^2 = 3.59$ $df = 9$ $P = 0.94$

[a] C, B, H, G, O, and S refer to common chimpanzee, bonobo, human, gorilla, orangutan, and siamang. In the specification of a configuration of a site, the amino acids of the species within common parentheses are the same, while those in different parentheses are different. For the heterogeneous model, zero-change sites were deleted one by one until the expected number of the zero-change sites coincided with the observed number for the remainder that were assumed to evolve homogeneously across sites. When 972 sites were deleted from the 1,128 sites of zero change, the coincidence was attained

affected only slightly by taking account of the site heterogeneity. Those estimated by NJ are affected more greatly, particularly for the deepest internal branch 4–3. This indicates that, while the multiple-hit effect is taken into account automatically to some extent in ML even under the homogeneity assumption because the method takes account of the states of the internal nodes, it is underestimated by distance methods such as the NJ.

For the ML analysis of the heterogeneous Poisson model, SEs of branch lengths and branching dates were estimated by replicating bootstrap resampling (Felsenstein 1985) and from the curvature of likelihood surface (given by PROTML) as well (Table 1). The SEs of each branch length are nearly identical between the two methods of estimation, suggesting that the SEs estimated in the PROTML are good approximations. However, since the covariances between different branches are neglected in the estimation from the curvature (PROTML does not estimate covariances), the SEs of the branching dates turned out to be overestimated.

From the ML analysis of the heterogeneous Poisson model, we estimate 1.84 ± 0.43 Myr for the chimpanzee/bonobo separation, 3.60 ± 0.58 Myr for the human/chimpanzee, and 5.83 ± 0.72 Myr for the human/gorilla (Table 1). The latter two estimates are in accord with the previous estimates of 3.9 ± 0.7 and 5.1 ± 0.8 Myr from shorter mtDNA sequences (Hasegawa et al. 1990). The remarkable fit of the heterogeneous model to the data and the robustness of the ML estimates of branching dates to changes in model assumptions raise the possibility that the human/chimpanzee separation was more recent than has been generally thought even by molecular evolutionists. However, there are two factors that may cause our estimate to be too young. First, we assumed the orangutan separation to be 13 Myr old. If it was 16 Myr, which is probably the oldest limit (Pilbeam 1988; Andrews 1992; McCrossin and Benefit 1993), the human/chimpanzee separation is estimated to be 4.43 ± 0.71 Myr old. Second, there may have been variation of the evolutionary rate which cannot be detected by the relative rate test. If the rate along the 4–3 branch was as high as that along the orangutan (4–0) branch, the human/chimpanzee separation is estimated to be 4.70 ± 0.99 Myr old. These possibilities cannot be excluded, and therefore some uncertainties exist in our estimates.

Discussion

In spite of the uncertainties discussed above, it seems unlikely from our analysis that the human/chimpanzee separation in the mtDNA tree was much older than 5 Myr, and the most likely date would be 4–5 Myr. Our dating of the human/chimpanzee separation is closely relevant to the dating of the deepest root of the human mtDNA tree, and is in favor of the recent-origin hypothesis of modern humans (Cann et al. 1987; Kocher and Wilson 1991; Hasegawa et al. 1993b; Ruvolo et al. 1993) rather than the more-ancient-origin hypothesis (Thorne and Wolpoff 1992; Pesole et al. 1992).

Molecular clock analyses that take account of the rate heterogeneity among lineages (Kishino and Hasegawa 1990; Hasegawa 1991) gave 4.0 ± 1.1 and 4.7 ± 0.8 Myr dates for the human/chimpanzee separation from the ribosomal internal transcribed spacers (ITS1) (Gonzalez et

al. 1990) and the immunoglobulin ϵ pseudogene (Ueda et al. 1989), and 6.3 ± 0.9 and 7.4 ± 0.8 Myr from the intergenic spacer between η and δ-globin genes (Maeda et al. 1988) and the η-globin pseudogene (Miyamoto et al. 1987) for the trifurcation among human, chimpanzee, and gorilla (the tricotomy could not be resolved by these data) with the same reference of 13 Myr for the orang-utan separation. Although the estimate from ITS1 is consistent with that from mtDNA, the estimates from the other nuclear genes are older. It should be noted that such gene trees do not necessarily agree with the species tree mainly because of ancestral polymorphism (e.g., Nei 1987). Older coalescence is expected for some nuclear genes.

The expected duration time of polymorphism is proportional to the effective population size under neutrality (Kimura 1983). Since the effective population size of mtDNA is about one-fourth that of nuclear genes, because of its maternal inheritance and of the haploid nature (Takahata 1985), polymorphism is likely to be maintained for a longer time in the nuclear genes than in the mtDNA. The discrepancy among the dates of human/chimpanzee separation estimated from different genes is thus likely to be due to polymorphism particularly of the η-globin pseudogene and of the globin spacer in the common ancestral species of human and the African apes (Hasegawa et al. 1987; Hasegawa 1991). If this is the case, it would be reasonable to consider that humans and chimpanzees diverged 4–5 Myr ago as suggested by the mtDNA and ITS1 clocks.

Acknowledgments. We thank H. Kishino and J. Reeves for helpful discussions. The helpful comments of anonymous reviewers are also appreciated. This work was carried out under the Institute of Statistical Mathematics Cooperative Research Program (93-ISM · CRP-C2) and was supported by grants from the Ministry of Education, Science, and Culture of Japan.

References

Adachi J, Hasegawa M (1992) Computer science monographs, No. 27. MOLPHY: programs for molecular phylogenetics, I.—PROTML: maximum likelihood inference of protein phylogeny. Institute of Statistical Mathematics, Tokyo

Adachi J, Cao Y, Hasegawa M (1993) Tempo and mode of mitochondrial DNA evolution in vertebrates at the amino acid sequence level: rapid evolution in warm-blooded vertebrates. J Mol Evol 36:270–281

Akaike H (1974) A new look at the statistical model identification. IEEE Trans Autom Contr AC-19:716–723

Anderson S, Bankier AT, Barrell BG, de Bruijn MHL, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith ALH, Staden R, Young IG (1981) Sequence and organization of the human mitochondrial genome. Nature 290:457–464

Andrews P, Cronin JE (1982) The relationships of *Sivapithecus* and *Ramapithecus* and the evolution of the orang-utan. Nature 297:541–546

Andrews P (1992) Evolution and environment in the Hominoidea. Nature 360:641–646

Bailey WJ, Hayasaka K, Skinner CG, Kehoe S, Sieu LC, Slightom JL, Goodman M (1992) Reexamination of the African hominoid trichotomy with additional sequences from the primate β-globin gene cluster. Mol Phyl Evol 1:97–135

Brown WM, Prager EM, Wang A, Wilson AC (1982) Mitochondrial DNA sequences of primates: tempo and mode of evolution. J Mol Evol 18:225–239

Caccone A, Powell JR (1989) DNA divergence among hominoids. Evolution 43:925–942

Cann RL, Stoneking M, Wilson AC (1987) Mitochondrial DNA and human evolution. Nature 325:31–36

Cao Y, Adachi J, Yano T, Hasegawa M (1994a) Phylogenetic place of guinea pigs: no support of the rodent polyphyly hypothesis from maximum likelihood analyses of multiple protein sequences. Mol Biol Evol 11:593–604

Cao Y, Adachi J, Janke A, Pääbo S, Hasegawa M (1994b) Phylogenetic relationships among eutherian orders estimated from inferred sequences of mitochondrial proteins: instability of a tree based on a single gene. J Mol Evol 39:519–527

Dayhoff MO, Schwartz RM, Orcutt BC (1978) A model of evolutionary change in proteins. In: Dayhoff MO (ed) Atlas of protein sequence and structure, vol 5, suppl 3. National Biomedical Research Foundation, Washington DC, pp 345–352

Felsenstein J (1981) Evolutionary trees from DNA sequences: a maximum likelihood approach. J Mol Evol 17:368–376

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791

Fitch WM, Markowitz E (1970) An improved method for determining codon variability in a gene and its application to the rate of fixations of mutations in evolution. Biochem Genet 4:579–593

Goldman N (1993) Statistical tests of models of DNA substitution. J Mol Evol 36:182–198

Gonzalez IL, Sylvester JE, Smith TF, Stambolian D, Schmickel RD (1990) Ribosomal RNA gene sequences and hominoid phylogeny. Mol Biol Evol 7:203–219

Hasegawa M, Kishino H, Yano T (1985) Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J Mol Evol 22:160–174

Hasegawa M, Kishino H, Yano T (1987) Man's place in Hominoidea as inferred from molecular clocks of DNA. J Mol Evol 26:132–147

Hasegawa M, Kishino H, Hayasaka K, Horai S (1990) Mitochondrial DNA evolution in primates: transition rate has been extremely low in lemur. J Mol Evol 31:113–121

Hasegawa M (1991) Molecular phylogeny and man's place in Hominoidea. J Anthrop Soc Nippon 99:49–61

Hasegawa M, Horai S (1991) Time of the deepest root for polymorphism in human mitochondrial DNA. J Mol Evol 32:37–42

Hasegawa M, Cao Y, Adachi J, Yano T (1992) Rodent polyphyly? Nature 355:595–595

Hasegawa M, Hashimoto T, Adachi J, Iwabe N, Miyata T (1993a) Early divergences in the evolution of eukaryotes: ancient divergence of *Entamoeba* that lacks mitochondria revealed by protein sequence data. J Mol Evol 36:380–388

Hasegawa M, Di Rienzo A, Kocher TD, Wilson AC (1993b) Toward a more accurate time scale for the human mitochondrial DNA tree. J Mol Evol 37:347–354

Hashimoto T, Otaka E, Adachi J, Mizuta K, Hasegawa M (1993) The giant panda is most close to a bear, judged by α- and β-hemoglobin sequences. J Mol Evol 36:282–289

Hashimoto T, Nakamura Y, Nakamura F, Shirakura T, Adachi J, Goto N, Okamoto K, Hasegawa M (1994) Protein phylogeny gives a robust estimation for early divergences of eukaryotes: phylogenetic place of a mitochondria-lacking protozoan, *Giardia lamblia*. Mol Biol Evol 11:65–71

Hayasaka K, Gojobori T, Horai S (1988) Molecular phylogeny and evolution of primate mitochondrial DNA. Mol Biol Evol 5:626–644

Horai S, Satta Y, Hayasaka K, Kondo R, Inoue T, Ishida T, Hayashi S, Takahata N (1992) Man's place in Hominoidea revealed by mito-

chondrial DNA genealogy. J Mol Evol 35:32–43; Erratum 37:89–89 (1993)

Jones DT, Taylor WR, Thornton JM (1992) The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci 8:275–282

Jukes TH, Cantor CR (1969) Evolution of protein molecules. In: Munro HN (ed) Mammalian protein metabolism, vol III. Academic Press, New York, pp 21–132

Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16:111–120

Kimura M (1983) The neutral theory of molecular evolution. Cambridge Univ Press, Cambridge

Kishino H, Hasegawa M (1989) Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in Hominoidea. J Mol Evol 29:170–179

Kishino H, Hasegawa M (1990) Converting distance to time: an application to human evolution. Methods Enzymol 183:550–570

Kishino H, Miyata T, Hasegawa M (1990) Maximum likelihood inference of protein phylogeny and the origin of chloroplasts. J Mol Evol 30:151–160

Kocher TD, Wilson AC (1991) Sequence evolution of mitochondrial DNA in humans and chimpanzees: control region and a protein-coding region. In: Osawa S, Honjo T (eds) Evolution of life: fossils, molecules, and culture. Springer-Verlag, Tokyo, pp 391–413

Maeda N, Wu C-I, Bliska J, Reneke J (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

McCrossin ML, Benefit BR (1993) Recently recovered *Kenyapithecus* mandible and its implications for great ape and human origins. Proc Natl Acad Sci USA 90:1962–1966

Miyamoto MM, Slightom JL, Goodman M (1987) Phylogenetic relations of humans and African apes from DNA sequences in the ψη-globin region. Science 238:369–373

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York

Pesole G, Ebisá E, Preparata G, Saccone C (1992) The evolution of the mitochondrial D-loop region and the origin of modern man. Mol Biol Evol 9:587–598

Pilbeam D (1988) Human origins and evolution. In: Fabian AC (ed) Origins. Cambridge University Press, Cambridge, pp 89–114

Reeves JH (1992) Heterogeneity in the substitution process of amino acid sites of proteins coded for by mitochondrial DNA. J Mol Evol 35:17–31

Ruvolo M, Disotell TR, Allard MW, Brown WM, Honeycutt RL (1991) Resolution of the African hominoid trichotomy by use of a mitochondrial gene sequence. Proc Natl Acad Sci USA 88:1570–1574

Ruvolo M, Zehr S, von Dornum M, Pan D, Chang B, Lin J (1993) Mitochondrial COII sequences and modern human origins. Mol Biol Evol 10:1115–1135

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Sarich VM, Wilson AC (1967a) Immunological time scale for hominid evolution. Science 158:1200–1203

Sarich VM, Wilson AC (1967b). Rates of albumin evolution in primates. Proc Natl Acad Sci USA 58:142–148

Sibley CG, Ahlquist JE (1984) The phylogeny of the hominoid primates, as indicated by DNA-DNA hybridization. J Mol Evol 20: 2–15

Sibley CG, Ahlquist JE (1987) DNA hybridization evidence of hominoid phylogeny: results from an expanded data set. J Mol Evol 26:99–121

Sibley CG, Comstock JA, Ahlquist JE (1990) DNA hybridization evidence of hominoid phylogeny: a reanalysis of the data. J Mol Evol 30:202–236

Sidow A, Nguyen T, Speed TP (1992) Estimating the fraction of invariable codons with a capture-recapture method. J Mol Evol 35: 253–260

Takahata N (1985) Population genetics of extranuclear genomes: a model and review. In: Ohta T, Aoki K (eds) Population genetics and molecular evolution. Japan Sci Soc Press, Tokyo, pp 195–212

Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol 10:512–526

Thorne AG, Wolpoff MH (1992) The multiregional evolution of humans. Sci Am 266(4):76–83

Ueda S, Watanabe Y, Saitou N, Omoto K, Hayashida H, Miyata T, Hisajima H, Honjo T (1989) Nucleotide sequences of immunoglobulin-epsilon pseudogenes in man and apes and their phylogenetic relationships. J Mol Biol 205:85–90

Yang Z (1993) Maximum-likelihood estimation of phylogeny from DNA sequences when substitution rates differ over sites. Mol Biol Evol 10:1396–1401