Molecular Phylogeny and Man's Place in Hominoidea

Masami HASEGAWA

*The Institute of Statistical Mathematics, and Department of Statistical Science,
The Graduate University for Advanced Studies*

Abstract Branching order among man, chimpanzee and gorilla was inferred with orangutan as an outgroup from the data of the 28S rRNA gene and of the ribosomal internal transcribed spacer recently sequenced by GONZALEZ *et al.* (1990). The analysis was performed by the maximum likelihood method without assuming constancy of the evolutionary rate. It was strongly suggested that chimpanzee is the closest relative of man, consistently with GONZALEZ *et al.*'s analysis. After establishing the branching order, the extent of rate variation among lineages was evaluated, and the branching dates were estimated for the ribosomal internal transcribed spacer by taking into account a possible rate variation. The overall evidence from the DNA sequence data accumulated up to date suggests that man separated from chimpanzee some 4–5 million years ago.

Keywords Man/chimpanzee clade, Molecular clock with variable rate, Branching dates

Introduction

The evolutionary history of organisms is inscribed in the DNA of the living species. Since GOODMAN (1963) and SARICH and WILSON (1967), analyses of proteins and DNAs have been playing an important role in elucidating man's place in the evolution of Hominoidea, and the recent progress of DNA sequencing technique provides us with invaluable data for a profound investigation of this problem (for review, see HASEGAWA, 1990; HASEGAWA *et al.*, 1987, 1989; HASEGAWA and KISHINO, 1991).

It has been well recognized that man is a close relative of apes. Furthermore, all molecular clock analyses with sufficient resolving power con-

sistently group man with chimpanzee and gorilla, and away from orangutan (GOODMAN, 1963; SARICH and WILSON, 1967; ANDREWS and CRONIN, 1982; KOOP *et al.*, 1986; HASEGAWA *et al.*, 1987). However, some morphologists still claim other branching orders. For example, SCHWARTZ (1984) links man and orangutan away from the African apes. He claims that, because the molecular evolutionary rate may differ among different lineages, similarity of DNA or protein sequences does not necessarily mean a phylogenetic closeness.

Although approximate rate constancy is generally a rule in molecular evolution (WILSON *et al.*, 1977; KIMURA, 1983), several instances of deviation from the constancy have been actually

suggested (HASEGAWA *et al.*, 1987, 1990; HASEGAWA and KISHINO, 1989a, and references therein). Such deviation may be due to two causes: one is the change of the mutation rate, and the other is the alteration of the selective constraint of each molecule (KIMURA, 1987). Because of the possibility of the deviation from rate constancy, we should infer the branching order at first without assuming the constancy. A maximum likelihood method developed by FELSENSTEIN (1981) is robust against the violation of the rate constancy, and therefore is suitable for this purpose (HASEGAWA and YANO, 1984a; SAITOU and IMANISHI, 1989).

The method has a merit that we can make an assessment of how much confidence to place in the inferred tree. To place a confidence limit of the inferred tree, bootstrap resampling is applicable (FELSENSTEIN, 1985). We developed a simple method for estimating a bootstrap probability of a tree for being the highest likelihood among alternatives without performing a maximum likelihood estimation for each resampled data sets (KISHINO *et al.*, 1990), and applied the method to the sequence data of mitochondrial DNA (mtDNA) and of several genes in nuclear DNA (HASEGAWA and KISHINO, 1991; HASEGAWA, 1990). It was shown that orangutan is convincingly placed as an outgroup of an African apes/man clade even if evolutionary rate differs among lineages (HASEGAWA and KISHINO, 1989b; KISHINO and HASEGAWA, 1989). Although man/chimpanzee clustering within the African apes/man clade has been preferred over man/gorilla and chimpanzee/gorilla clusterings by the previous analyses (HASEGAWA and YANO, 1984b; HASEGAWA *et al.*, 1985; MIYAMOTO *et al.*, 1987; UEDA *et al.*, 1989), the preference has been only marginal (HASEGAWA 1990; HASEGAWA and KISHINO, 1991). Our analyses of GONZALEZ *et al.*'s (1990) data of the 28S rRNA (LrRNA)

gene and of one of the ribosomal internal transcribed spacer (ITS1) presented in this paper provide strong support of the man/chimpanzee clustering with 99% bootstrap probability.

Another interesting question is when man diverged from the apes. After establishing the branching order, we can estimate whether or not the evolutionary rate is equal among different lineages. Even if the constancy of the evolutionary rate violates, it does not necessarily invalidate the molecular clock analysis, because the variations can be incorporated into the model used in the analysis. We have developed a method for evaluating how much extent the rate differs among lineages and for estimating branching dates that takes into account the variation of the rate among lineages (HASEGAWA *et al.*, 1989; KISHINO and HASEGAWA, 1990). Analysis of the ITS1 data by this method in addition to the overall evidence from the previous analyses of other gene loci (HASEGAWA, 1990) suggests that man diverged from chimpanzee about 4–5 million years before present (MyrBP).

Branching Order among Man, Chimpanzee and Gorilla

The maximum likelihood analyses were carried out by using DNAML in FELSENSTEIN's (1988) program package PHYLIP (Version 3.1). Three tree topologies shown in Fig. 1 were examined. The trees analyzed by the maximum likelihood method that does not assume rate constancy are unrooted, and hence a proper outgroup is necessary to place a root. Since it has been convincingly shown that man and the African apes form a clade excluding orangutan as an outgroup (bootstrap probabilities of other relationships are about 10^{-87} ; HASEGAWA, 1990; HASEGAWA and KISHINO, 1991), orangutan was used as an outgroup to place the root, and the three possible relations among man, chimpanzee and gorilla shown in Fig. 1 were

Fig. 1. Candidate trees of Hominoidea.

compared.

Table 1 summarizes the previous analyses of various genes by the maximum likelihood (from HASEGAWA, 1990, but the analysis of the *Ig- $\phi\epsilon$* was revised excluding 3'-flanking sequence. See the following section). The data used are a fragment of 896 nucleotides of mtDNA (BROWN *et al.*, 1982), small rRNA gene (SrRNA) (HIXSON and BROWN, 1986), η -globin pseudogene ($\phi\eta$ -globin) (MIYAMOTO *et al.*, 1987), spacer between $\phi\eta$ and δ -globin genes (η - δ spacer) (MAEDA *et al.*, 1988), and immunoglobulin ϵ pseudogene (*Ig- $\phi\epsilon$*) (UEDA *et al.*, 1989). The maximum likelihood method has an advantage that allows

us to evaluate the confidence limits of the inferred trees by taking account of all the data from diverse genes (KISHINO and HASEGAWA, 1989; HASEGAWA, 1990; HASEGAWA and KISHINO, 1991). This can be achieved by summing up log-likelihood values for diverse genes and by comparing the total log-likelihoods among alternative trees.

The overall evidence by the maximum likelihood analyses suggests that the man/chimpanzee clustering has the bootstrap probability of 76%, while the man/gorilla and chimpanzee/gorilla clusterings have bootstrap probabilities of 12% and 13%, respectively. Therefore, although the man/chimpanzee clustering is the most likely, the two alternative trees cannot be excluded by the previous analyses.

Additional sequence data used here in inferring the branching order in Hominoidea are the 28S rRNA (LrRNA) gene and the ribosomal internal transcribed spacer (ITS1) from man, chimpanzee, gorilla and orangutan sequenced by GONZALEZ *et al.* (1990). GONZALEZ and coworkers themselves analyzed their data by various methods for inferring trees, and all the methods consistently suggested the man/chimpanzee clustering. To place the confidence limit of the inferred tree, they tested whether or not the internal branch length of the inferred tree is significantly positive. This is the test proposed by LI (1989); that is, a given branching would be supported if the internal branch length is significantly greater than zero, would be rejected if it is significantly less than zero, or would be unsupported if it is approximately zero. The internal branch length of the man/chimpanzee tree is negative at the 2% level of significance for LrRNA and at the 9% level for ITS1, and they concluded that the LrRNA data clearly supports the man/chimpanzee tree. However, since this test is performed under the assumption that the inferred tree is a correct one, and since an in-

Table 1. Bootstrap probabilities of hominoid trees for being the highest likelihood among three candidate trees in Fig. 1, and the difference of log-likelihood from that of Tree-1 (in parentheses) (revised from Table 1 in HASEGAWA, 1990). The sites that experienced insertion or deletion were excluded from the analysis except for the $\phi\eta$ -globin. The insertion/deletion events were taken into account for the $\phi\eta$ -globin data by the method of KISHINO *et al.* (1990). n denotes the number of nucleotide sites (number of insertion/deletion positions in parentheses). The 1st, 2nd and 3rd refer to the positions in a codon. Bootstrap probabilities were estimated by the multivariate normal distribution (MND) method developed by KISHINO *et al.* (1990) (sample size of 10^4). The bootstrap probability of the highest likelihood tree for each data is denoted by boldface. $\pm$ refers to 1SE estimated by the formula of KISHINO and HASEGAWA (1989).

	n	Tree-1	Tree-2	Tree-3
<i>mitochondrial DNA</i>				
3rd	232	0.71	0.28 (-3.29 ± 5.92)	0.02 (-5.45 ± 4.74)
1st, 2nd & tRNA	662	0.26	0.07 (-1.98 ± 4.18)	0.67 ($+3.29 \pm 6.15$)
SrRNA	939	0.19	0.68 ($+3.97 \pm 6.22$)	0.12 (-0.84 ± 3.75)
<i>nuclear DNA</i>				
$\phi\eta$ -globin	6,166 (74)	0.71	0.01 (-12.76 ± 8.02)	0.29 (-5.59 ± 9.81)
η - δ spacer	2,242	0.33	0.63 ($+2.09 \pm 5.38$)	0.04 (-1.82 ± 2.58)
Ig- $\phi\epsilon$	2,250	0.74	0.19 (-2.24 ± 3.26)	0.07 (-2.52 ± 3.12)
Total		0.76	0.12 (-14.22 ± 13.97)	0.13 (-13.93 ± 13.67)

correct tree can have significantly positive internal branch length, an alternative test seems to be desirable to confirm their results.

Table 2 gives the results of the maximum likelihood analyses. Consistently with GONZALEZ *et al.*'s analysis, the LrRNA data strongly favours Tree-1 of the man/chimpanzee clade with the bootstrap probability of 97%, while the ITS1 data also favours it at a lower significance level (90%). In the lowest row in Table 2, the total of LrRNA and ITS1 is shown. From the data analyzed here, the man/

chimpanzee tree is the most likely with the bootstrap probability of 99%, and the two alternatives are highly unlikely. This conclusion holds even if the evolutionary rate differs among different lineages (HASEGAWA and YANO, 1984a).

Date of Man/Ape Separation

Once the branching order is known, we can evaluate whether the molecular evolutionary rate is equal among different lineages. One of the advantages of using DNA sequences in estimating

Table 2. Bootstrap probabilities of hominoid trees of LrRNA gene and ITS1 for being the highest likelihood among three candidate trees in Fig. 1, and the difference of log-likelihood from that of Tree-1 (in parentheses). Sequence data presented in Fig. 2a and 2b of GONZALEZ *et al.* (1990) were used. The sites that experienced insertion or deletion among man, chimpanzee, gorilla and orangutan were excluded from the analysis. Bootstrap probabilities were estimated by the MND method developed by KISHINO *et al.* (1990) (sample size of 10^4). $\pm$ refers to 1SE estimated by the formula of KISHINO and HASEGAWA (1989).

	<i>n</i>	Tree-1	Tree-2	Tree-3
LrRNA	826	0.97	0.00 (-21.01 $\pm$ 9.88)	0.02 (-19.22 $\pm$ 10.26)
ITS1	893	0.90	0.02 (-8.36 $\pm$ 5.54)	0.08 (-7.73 $\pm$ 5.75)
Total		0.99	0.00 (-29.37 $\pm$ 11.32)	0.01 (-27.65 $\pm$ 11.76)

the branching dates is the approximate constancy of nucleotide substitution rate (WILSON *et al.*, 1977; KIMURA, 1983). The constancy, however, is not absolute, and the rate sometimes differs among different lineages. We developed a statistical method for inferring the branching dates taking into account the possibility of the violation of the rate constancy (KISHINO and HASEGAWA, 1990).

We assume that each site evolves according to a MARKOV process in which a nucleotide (*T*, *C*, *A* or *G*) is replaced by another nucleotide *y* in an infinitesimally short interval of time, *dt*, with a probability of $P_{xy}(dt)$, as follows,

$$P_{xy}(dt) = \begin{cases} \alpha\pi_y dt & \text{for transition} \\ & (T \leftrightarrow C, A \leftrightarrow G) \\ \beta\pi_y dt & \text{for transversion} \\ & (T, C \leftrightarrow A, G), \end{cases} \quad (1)$$

where π_y is the composition of nucleotide *y* (HASEGAWA *et al.*, 1984, 1985). The parameters α and β can be different among different

branches in a tree. We assume that all of the nucleotide sites are independently and homogeneously variable with the substitution probability represented by Equation (1). This assumption seems to be justified as an approximation when we deal with pseudogenes or spacers that have no known functions. However, when we deal with functional genes, it must be taken into account that not all the sites are equally variable. Hence we introduced an additional parameter *f* that represents the fraction of variable sites and assumed that the remaining sites are invariable in the time scale under consideration (HASEGAWA *et al.*, 1984, 1985; HASEGAWA and HORAI, 1991).

The α 's, β 's, and branching dates are parameters of the model. In a model where evolutionary rates are not equal among branches, different sets of rate parameters, α_k and β_k , are assigned to branches along which rate differs. An extreme case is that every branch has its own parameters, α_k and β_k ($k = 1, 2, \dots$, number of branches), and the other extreme is rate constancy throughout the tree. Between these two extremes, there are many intermediate models in which several, but not all, branches share a common set of rate parameters. As we introduce many parameters, fitting of the model of the data improves, and accordingly the likelihood increases. However, we cannot accept a model with an unnecessarily large number of parameters as an appropriate model. A penalty should be imposed on increasing the number of parameters. Based on an information theoretical consideration, AKAIKE (1974) showed that, in comparing the goodness of the approximation of models with different numbers of parameters, AIC (AKAIKE Information Criterion) defined by

$$\text{AIC} = -2 \times (\log\text{-likelihood}) + 2 \times (\text{number of parameters}), \quad (2)$$

provides a good criterion of model selection (for

review, see SAKAMOTO *et al.*, 1986). The log-likelihood and parameters of the model are estimated by the maximum likelihood. The better the fitting of the model to the data, the lower is the first term. On the other hand, the more complex is the model, the higher is the second term. A model which minimize the AIC is considered to be the most appropriate model. This implies that, when there are several models whose values of log-likelihood are nearly the same, we should choose the one with the smallest number of parameters.

To calibrate a molecular clock that takes account of the possibility of rate variation, the branching of orangutan from the African apes-man clade is taken as a reference, and it is assumed to be 13 MyrBP (PILBEAM 1985; ANDREWS 1986). This assumption is based on some interpretations of fossil evidence that ramapithecines of 7–12 MyrBP in Asia have shared derived characters with orangutan and that older ones than 13 Myr in Africa (called *Kenyapithecus*) are common ancestors of the great apes clade before the separation of orangutan.

Table 3 represents the numbers of transition and transversion differences among the ITS1's from Hominoidea. Since the number of transition differences between gorilla and orangutan, 28, is larger than those between man and orangutan, 23, and between chimpanzee and orangutan, 24, a model (Model 2) where rate

change in gorilla is allowed (by assigning to the gorilla line a set of rate parameters, α_2 and β_2 , different from α_1 and β_1 assigned to other lines) is compared in Table 4 with a constant rate model (Model 1; a single set of rate parameters is assigned to all branches). Although AIC of Model 2 is higher than that of the constant rate Model 1, estimates of β_1 and β_2 from Model 2 do not differ significantly. Therefore, Model 3 allowing only transition rate change in gorilla is examined, and it turns out that this model with lesser number of parameters improves AIC and is more appropriate than Model 1 of constant rate.

More complicated models that takes account of minor differences of branch lengths, for example between man and chimpanzee, are not preferred. Introduction of an additional parameter f that represents the fraction of variable sites (HASEGAWA *et al.*, 1984, 1985; HASEGAWA and HORAI, 1991) does not improve AIC, and therefore it seems to be justified to assume as an approximation that all the sites in ITS1 are equally variable. Furthermore, Model 4 that allows rate change in the man/chimpanzee clade rather than in gorilla is examined, but it turns out not to be appropriate. The minimum AIC model (Model 3) indicates that gorilla and chimpanzee separated from man 5.6 ± 1.3 and 4.0 ± 1.1 MyrBP ($\pm$: 1SE).

In a previous paper (HASEGAWA, 1990), I analyzed the immunoglobulin ϵ pseudogene (Ig- $\phi\epsilon$) sequenced by UEDA *et al.* (1989). However, I missed to take into account the remark of UEDA and his coworkers that the gorilla sequence has an unusually high number of substitutions in the 3'-flanking region downstream from position 2,393 in their numbering. Therefore, the acceleration of the transition rate in gorilla Ig- $\phi\epsilon$ suggested in HASEGAWA (1990) may be an artifact caused by gene conversion or unequal crossing over. In Table 5, the data was

Table 3. Numbers of transition (upper right half) and transversion (lower left half) differences among ITS1 from Hominoidea (893 sites)

	Man	Chimp	Gorilla	Orang
Man		12	19	23
Chimp	7		17	24
Gorilla	12	11		28
Orang	32	31	30	

Table 4. Evolutionary rates and divergence dates estimated by the molecular clock of ITS1 (893 sites). Model 1 is a constant rate model. Model 2 allows rate change of transition and transversion in gorilla (α_2 , β_2). Model 3 allows only transition rate change in gorilla (α_2). Model 4 allows rate change of transition and transversion in the man/chimpanzee clade (α_2 , β_2). $\pi_T = 0.133$, $\pi_C = 0.398$, $\pi_A = 0.063$, and $\pi_G = 0.405$.

	Model 1	Model 2	Model 3	Model 4
	constant rate	rate change in gorilla		rate change in man/chimp
Rates (10^{-3} /Myr)				
$\hat{\alpha}_1$	8.45 ± 1.36	7.50 ± 1.53	7.51 ± 1.53	8.31 ± 1.71
$\hat{\alpha}_2$		13.85 ± 4.79	14.83 ± 5.11	8.26 ± 3.04
$\hat{\beta}_1$	2.50 ± 0.47	2.69 ± 0.49	2.61 ± 0.47	2.69 ± 0.51
$\hat{\beta}_2$		1.91 ± 1.13		1.82 ± 0.94
Branching dates (Myr)				
t_1 (orangutan/man)	13	13	13	13
t_2 (gorilla/man)	6.5 ± 1.2	6.0 ± 1.4	5.6 ± 1.3	6.9 ± 1.6
t_3 (chimpanzee/man)	4.4 ± 1.0	4.2 ± 1.1	4.0 ± 1.1	5.0 ± 1.6
AIC	42.51	43.51	41.90	45.74

 minimum AIC

reanalyzed excluding the 3'-flanking region.

Model 2 that allows rate change in gorilla has a higher AIC than Model 1 of the constant rate. Another model that allows only transition rate change in gorilla ($\hat{\beta}_1$ and $\hat{\beta}_2$ are almost identical in Model 2) has AIC of 39.51, that is still higher than that of Model 1. Any models that allows rate change in man or chimpanzee do not reduce AIC. Therefore, the constant rate model is the most appropriate model for the Ig- ψ s data, in agreement with UEDA *et al.*'s (1989) analysis by the relative rate test. Model 1' that assumes trifurcation among man, chimpanzee and gorilla with rate constancy has a slightly higher AIC than Model 1, and the tricotomy among man and the African apes is resolved under the most appropriate model of rate

constancy.

From Model 1, the branching dates of gorilla and chimpanzee from man are estimated to be 5.4 ± 0.8 and 4.7 ± 0.8 MyrBP. By assuming 14 MyrBP for the orangutan separation, UEDA *et al.* (1989) estimated that gorilla and chimpanzee separated from man 5.9 ± 0.9 and 4.9 ± 0.9 MyrBP. Our estimates are consistent with theirs.

Table 6 summarizes branching dates in Hominoidea estimated by the molecular clocks of diverse genes taking account of the possibility of rate variation among lineages. The two data sets of mitochondrial DNA are the 896 nucleotides of ND4, ND5 and rRNAs (BROWN *et al.*, 1982) and the SrRNA gene (HIXSON and BROWN, 1986). The four data sets of nuclear DNA are the $\phi\eta$ -globin (MIYAMOTO *et al.*,

Table 5. Evolutionary rates and divergence dates estimated by the molecular clock of immunoglobulin $\psi\epsilon$ gene excluding the 3'-flanking region downstream from position 2393 in UEDA *et al.*'s (1989) numbering (2250 sites). Model 1 is a constant rate model. Model 1' is also the constant rate model, and assumes trifurcation among man, chimpanzee and gorilla. Model 2 allows rate change of transition and transversion in gorilla (α_2, β_2). $\pi_T=0.199$, $\pi_C=0.282$, $\pi_A=0.291$, and $\pi_G=0.228$.

	Model 1	Model 1'	Model 2
	constant rate		rate change in gorilla
Rates (10^{-3} /Myr)			
$\hat{\alpha}_1$	4.16 ± 0.50	4.10 ± 0.50	4.04 ± 0.53
$\hat{\alpha}_2$			5.20 ± 1.61
$\hat{\beta}_1$	1.07 ± 0.17	1.05 ± 0.17	1.09 ± 0.18
$\hat{\beta}_2$			1.05 ± 0.53
Branching dates (Myr)			
t_1 (orangutan/man)	13	13	13
$\hat{t}_2$ (gorilla/man)	5.4 ± 0.8	5.0 ± 0.7	5.1 ± 0.9
$\hat{t}_3$ (chimpanzee/man)	4.7 ± 0.8		4.6 ± 0.8
AIC	38.12	38.21	41.50

	minimum AIC		

Table 6. Branching dates (Myr) in Hominoidea estimated by molecular clocks of DNA that take account of the possibility of rate variation (from the minimum AIC model for the respective data; from this work and from HASEGAWA 1990). The clocks are calibrated by assuming the separation of orangutan from man to be 13 Myr ago. $\pm$ refers to 1SE.

	mtDNA		nuclear DNA			
	ND4&5 + tRNAs	SrRNA	$\psi\eta$ -globin	η - δ spacer	Ig- $\psi\epsilon$	ITS1
Orang/Man	13	13	13	13	13	13
Gorilla/Man	5.1 ± 0.8	4.2 ± 1.3	7.4 ± 0.8	6.3 ± 0.9	5.4 ± 0.8	5.6 ± 1.3
Chimp/Man	3.9 ± 0.7				4.7 ± 0.8	4.0 ± 1.1
Common/Pygmy chimp	—	1.2 ± 0.7	—	—	—	—

1987), the η - δ spacer (MAEDA *et al.*, 1988), the Ig- $\phi\epsilon$ (UEDA *et al.*, 1989), and the ITS1 (GONZALEZ *et al.*, 1990). The SrRNA of mtDNA, the $\phi\eta$ -globin and the η - δ spacer cannot resolve the tricotomy among man, chimpanzee and gorilla, and the date of trifurcation was estimated for these genes. The 896 nucleotides of mtDNA gave estimates of 3.9 ± 0.7 and 5.1 ± 0.8 MyrBP for the chimpanzee and gorilla separations from man, and the SrRNA gene of mtDNA gave an estimate (4.2 ± 1.3 MyrBP for the trifurcation) congruently with them if the standard errors are taken into account. Furthermore, the ITS1 and Ig- $\phi\epsilon$ also gave congruent estimates with those from the mtDNA, that is, 4.0 ± 1.1 and 4.7 ± 0.8 MyrBP for the chimpanzee separation and 5.6 ± 1.3 and 5.4 ± 0.8 MyrBP for the gorilla separation. However, the dates of 7.4 ± 0.8 and 6.3 ± 0.9 MyrBP for the trifurcation among man, chimpanzee and gorilla estimated from the $\phi\eta$ -globin and the η - δ spacer seem to be older than the estimates from other gene loci.

It should be noted that a gene tree does not necessarily agree with a species tree, in which a branching refers to the reproductive isolation between the two descendent species, because of the possibilities of ancestral polymorphism (*e.g.* NEI, 1987) and of introgression (gene transfer across the species boundary through hybridization) (FERRIS *et al.* 1983). Since the three independent data of mtDNA, ITS1 and Ig- $\phi\epsilon$ gave congruent estimates for the dates of man/chimpanzee and man/gorilla separations, the older dates estimated from the $\phi\eta$ -globin and the η - δ spacer are likely to be due to polymorphism of these gene loci in the common ancestral species of man and the African apes (HASEGAWA, 1990; HASEGAWA and KISHINO, 1991). If this is the case and if the possibility of introgression can be ignored, it might be reasonable to consider that man and chimpanzee diverged at the species level

some 4 MyrBP that the mtDNA and ITS1 indicate. This estimate depends on the assumption of 13 MyrBP for the orangutan separation. If this separation occurred 16 MyrBP as some morphologists think likely (PICKFORD, personal communication), the man/chimpanzee separation is estimated to be some 5 MyrBP. Therefore, it might be safe at present to suggest that man separated from chimpanzee some 4–5 MyrBP after gorilla separated from these two species.

Discussions

Although the branching order among man, chimpanzee and gorilla has been equivocal by the previous analyses of the maximum likelihood without assuming the constancy of evolutionary rate, the analyses of the recently sequenced data of LrRNA and of ITS1 strongly support the man/chimpanzee tree. Furthermore, the 896 nucleotide segment of mtDNA favours significantly the man/chimpanzee tree under the minimum AIC model, and the Ig- $\phi\epsilon$ also favours it although marginally. The closer relation of chimpanzee with man rather than with gorilla is supported also by DNA hybridization data (SIBLEY and AHLQUIST, 1984; CACCONI and POWELL, 1989; SIBLEY *et al.*, 1990).

Inconsistently with the data mentioned above, shared repeats and shared marker nucleotides in the segment of repeats of the involucrin gene suggest the chimpanzee/gorilla tree (DJIAN and GREEN, 1989). Although we cannot estimate the branching dates in the involucrin gene tree, this discrepancy is likely to be due to polymorphism of the involucrin gene in the common ancestral species of man, chimpanzee and gorilla. Further analyses of many independent gene loci are desirable to settle the issue.

It might be noteworthy that the transition and transversion parameters, α and β , for the ITS1 are about twice of those for the Ig- $\phi\epsilon$. Table 7 summarizes the transition rate $\nu_S =$

Table 7. Evolutionary rates of various parts of nuclear DNA in Hominoidea (10^{-9} /year/site). ν_S , ν_V and ν_T are the transition rate, transversion rate and total nucleotide substitution rate, respectively.

Gene loci	Lineage	ν_S	ν_V	ν_T
ITS1	except gorilla	1.18 ± 0.24	1.30 ± 0.24	2.48 ± 0.32
	gorilla	2.33 ± 0.80	1.30 ± 0.24	3.63 ± 0.89
Ig- $\phi\epsilon$		1.02 ± 0.12	0.54 ± 0.08	1.55 ± 0.16
$\phi\eta$ -globin	except man	0.83 ± 0.09	0.39 ± 0.05	1.21 ± 0.10
	man	0.62 ± 0.16	0.16 ± 0.08	0.78 ± 0.19
η - δ spacer		0.67 ± 0.12	0.26 ± 0.06	0.93 ± 0.13

$2(\pi_T \pi_C + \pi_A \pi_G) \alpha$, transversion rate $\nu_V = 2(\pi_T + \pi_C)(\pi_A + \pi_G) \beta$, and the total substitution rate $\nu_T = \nu_S + \nu_V$, in Hominoidea for the four gene loci of nuclear DNA estimated under the minimum AIC model for the respective gene locus. Although $\hat{\alpha}$ of ITS1 is about twice of that of Ig- $\phi\epsilon$, the transition rate ν_S does not differ between the two gene loci (except in the gorilla lineage) because of the extreme bias in base composition in ITS1. Extremely higher transversion rate in ITS1 than in other pseudogenes and spacer is remarkable. In principle this can be an artifact due to paralogous comparison between orangutan and the African apes/man clade. Even if this is the case, however, it is remarkable that the transversion rate ν_V is nearly equal to the transition rate ν_S , while transversion is rare in other gene loci. If the Ig- $\phi\epsilon$ and $\phi\eta$ -globin are really pseudogenes, these discrepancies cannot be due to the difference of the functional constraint, but must be due to the difference of mutation rate among these gene loci. WOLFE *et al.* (1989) suggested that mutation rate differs among regions of the mammalian genome. Our analysis suggests that not only the mutation rate as a whole but also the transition/transversion ratio may differ among regions of the genome.

Apparent acceleration of the evolutionary rate

in a particular lineage can be an artifact caused by gene conversion or unequal crossing over as was shown for the Ig- $\phi\epsilon$. Therefore, when we encounter a case of rate acceleration, we must carefully evaluate whether the rate variation is real. Acceleration of transition rate of ITS1 in gorilla relative to those in man and chimpanzee was suggested in Table 4. Since we observe no cluster of nucleotide substitutions in the gorilla ITS1, this suggestion seems to be real. Slowdown of transition and transversion rates of $\phi\eta$ -globin in man relative to those in chimpanzee and gorilla was suggested by MIYAMOTO *et al.* (1987) and by KISHINO and HASEGAWA (1990). On the other hand, no significant rate variation among man, chimpanzee and gorilla is observed for the η - δ spacer (possible acceleration in man, but not significant; HASEGAWA *et al.*, 1989) and for the Ig- $\phi\epsilon$. In this way, the pattern of rate variation among lineages differ from different gene loci even in nuclear DNA with no apparent functions. Although several unknown factors seem to be operating in determining mutation rate and, therefore, in determining evolutionary rate, we can estimate branching dates taking account of the possibility of rate variation as shown above.

Overall evidence from the DNA sequence data accumulated up to date suggests that man separated from chimpanzee some 4–5 MyrBP.

A new and relatively complete fossil hominoid skull has recently been described by DE BONIS *et al.* (1990). The fossil was found at Macedonia in Northern Greece and was dated to be 9–10 Myr old. DE BONIS *et al.* considers that this Miocene hominoid might be a direct human ancestor. However, our analyses of DNA sequence data suggest that this fossil is too old to be our direct ancestor after we separated from chimpanzee.

Acknowledgements

I thank Dr. H. KISHINO and Dr. T. YANO for continuous help and discussions, Dr. N. SAITOU and Dr. M. PICKFORD for suggestions, and Miss M. FUJIWARA and Mr. J. ADACHI for computer programming. This work was supported by grants from the Ministry of Education, Science, and Culture of Japan, and by the Joint Studies Program of The Graduate University for Advanced Studies.

抄 録

分子系統学とヒト上科における人類の系統的位置

長谷川政美

DNA塩基配列データからヒト、チンパンジー、ゴリラの間の系統関係を、進化速度の一定性を仮定せずに、最尤法によって推定した。これまでの知見から、オランウータンがこれら3種からみて遠い関係にあることは確かであるから、これをアウトグループと仮定して解析した。GONZALEZら(1990)によって塩基配列が決定された28SリボソームRNA遺伝子とそれに隣接したスペーサーからは、ヒト/チンパンジー近縁説が99%のブートストラップ確率で支持され、ヒト/ゴリラあるいはチンパンジー/ゴリラ近縁説は棄却された。このことは、GONZALEZらがほかの分子系統樹推定法を用いて得た結論を支持するものである。さらに、オランウータンがヒトとアフリカ類人猿のグループから分かれたのが今から1300万年前と仮定し、リボソームRNAのスペーサーについて進化速度が系統ごとに異なる可能性を考慮に

いれて分岐年代を推定すると、ヒトがチンパンジーから分かれたのがおよそ400万年前ということになった。これまでのわれわれの研究で、ミトコンドリアDNAからも同じような分岐年代が推定されている。オランウータンの分岐がこれよりも古くて1600万年前である可能性もあるので、ヒトとチンパンジーの分岐はおよそ400ないし500万年前であったことが示唆される。

References

- AKAIKE, H., 1974: A new look at the statistical model identification. *IEEE Trans. Autom. Contr.* AC19: 716-723.
- ANDREWS, P., 1986: Fossil evidence on human origins and dispersal. *Cold Spring Harbor Symp. Quant. Biol.*, 52: 419-428.
- DE BONIS, L., G. BOUVRAIN, D. GERAADS and G. KOUFOS, 1990: New hominoid skull material from the late Miocene of Macedonia in Northern Greece. *Nature*, 345: 712-714.
- ANDREWS, P. and J.E. CRONIN, 1982: The relationship of *Sivapithecus* and *Ramapithecus* and the evolution of the orang-utan. *Nature*, 297: 541-546.
- BROWN, W.M., E.M. PRAGER, A. WANG and A.C. WILSON, 1982: Mitochondrial DNA sequence of primates: tempo and mode of evolution. *J. Mol. Evol.*, 18: 225-239.
- CACCONE, A. and J.R. POWELL, 1989: DNA divergence among hominoids. *Evolution*, 43: 925-942.
- DJIAN, P. and H. GREEN, 1989: Vectorial expansion of the involucrin gene and the relatedness of the hominoids. *Proc. Natl. Acad. Sci. USA*, 86: 8447-8451.
- 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.
- FELSENSTEIN, J., 1988: PHYLIP, Version 3.1. Univ. of Washington, Seattle.
- FERRIS, S.D., R.D. SAGE, C.-M. HUANG, J.T. NIELSEN, U. RITTE and A.C. WILSON, 1983: Flow of mitochondrial DNA across a species boundary. *Proc. Natl. Acad. Sci. USA*, 80: 2290-2294.
- GONZALEZ, I.L., J.E. SYLVESTER, T.F. SMITH, D. STAMBOLIAN and R.D. SCHMICKEL, 1990: Ribosomal RNA gene sequences and hominoid phylogeny. *Mol. Biol. Evol.*, 7: 203-219.
- GOODMAN, M., 1963: Serological analysis of the systematics of recent hominoids. *Hum. Biol.*, 35: 377-436.

- HASEGAWA, M., 1990: Phylogeny and molecular evolution in primates. *Jpn. J. Genet.*, 65: 243-266.
- HASEGAWA, M. and S. HORAI, 1991: Time of the deepest root for polymorphism in human mitochondrial DNA. *J. Mol. Evol.*, 32: 37-42.
- HASEGAWA, M. and H. KISHINO, 1989a: Heterogeneity of tempo and mode of mitochondrial DNA evolution among mammalian orders. *Jpn. J. Genet.*, 64: 243-258.
- HASEGAWA, M. and H. KISHINO, 1989b: Confidence limits on the maximum-likelihood estimate of the hominoid tree from mitochondrial-DNA sequences. *Evolution*, 43: 672-677.
- HASEGAWA, M. and H. KISHINO, 1991: DNA sequence analysis and evolution of Hominoidea. In: M. KIMURA and N. TAKAHATA (eds.) *New Aspects of the Genetics of Molecular Evolution*, Japan Sci. Soc. Press, Tokyo, pp.303-317.
- HASEGAWA, M., H. KISHINO, K. HAYASAKA and S. HORAI, 1990: Mitochondrial DNA evolution in primates: transition rate has been extremely low in lemur. *J. Mol. Evol.*, 31: 113-121.
- HASEGAWA, M., H. KISHINO and T. YANO, 1985: Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.*, 22: 160-174.
- HASEGAWA, M., H. KISHINO and T. YANO, 1987: Man's place in Hominoidea as inferred from molecular clocks of DNA. *J. Mol. Evol.*, 26: 132-147.
- HASEGAWA, M., H. KISHINO and T. YANO, 1989: Estimation of branching dates among primates by molecular clocks of nuclear DNA which slowed down in Hominoidea. *J. Hum. Evol.*, 18: 461-476.
- HASEGAWA, M. and T. YANO, 1984a: Maximum likelihood method of phylogenetic inference from DNA sequence data. *Bull. Biomet. Soc. Japan*, 5: 1-7.
- HASEGAWA, M. and T. YANO, 1984b: Phylogeny and classification of Hominoidea as inferred from DNA sequence data. *Proc. Japan Acad.*, B60: 389-392.
- HASEGAWA, M., T. YANO and H. KISHINO, 1984: A new molecular clock of mitochondrial DNA and the evolution of hominoids. *Proc. Japan Acad.*, B60, 95-98.
- HIKSON, J.E. and W.M. BROWN, 1986: A comparison of the small ribosomal RNA genes from the mitochondrial DNA of the great apes and humans: sequence, structure, evolution, and phylogenetic implications. *Mol. Biol. Evol.*, 3: 1-18.
- KIMURA, M., 1983: *The Neutral Theory of Molecular Evolution*. Cambridge Univ. Press, Cambridge.
- KIMURA, M., 1987: Molecular evolutionary clock and the neutral theory. *J. Mol. Evol.*, 26: 24-33.
- KISHINO, H. and M. HASEGAWA, 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. and M. HASEGAWA, 1990: Converting distance to time: an application to human evolution. *Methods Enzymol.*, 183: 550-570.
- KISHINO, H., T. MIYATA and M. HASEGAWA, 1990: Maximum likelihood inference of protein phylogeny and the origin of chloroplasts. *J. Mol. Evol.*, 31: 151-160.
- KOOP, B.F., M. GOODMAN, P. XU, K. CHAN and J.L. SLIGHTOM, 1986: Primate η -globin DNA sequences and man's place among the great apes. *Nature*, 319: 234-238.
- LI, W.-H., 1989: A statistical test of phylogenies estimated from sequence data. *Mol. Biol. Evol.*, 6: 424-435.
- MAEDA, N., C.-I. WU, J. BLISKA and J. RENEKE, 1988: Molecular evolution of intergenic DNA in higher primates: pattern of DNA changes, molecular clock, and evolution of repetitive sequences. *Mol. Biol. Evol.*, 5: 1-20.
- MIYAMOTO, M.M., J.L. SLIGHTOM and M. GOODMAN, 1987: Phylogenetic relations of humans and African apes from DNA sequences in the $\phi\eta$ -globin region. *Science*, 238: 369-373.
- NEI, M., 1987: *Molecular Evolutionary Genetics*. Columbia Univ. Press, New York.
- PILBEAM, D., 1985: Patterns of hominoid evolution. In: E. DELSON (ed.), *Ancestors: The Hard Evidence*. Alan R. Liss., New York, pp.51-59.
- SAITOU, N. and T. IMANISHI, 1989: Relative efficiencies of the FITCH-MARGOLISH, maximum-parsimony, maximum-likelihood, minimum-evolution, and neighbor-joining methods of phylogenetic tree construction in obtaining the correct tree. *Mol. Biol. Evol.*, 6: 514-525.
- SAKAMOTO, Y., M. ISHIGURO and G. KITAGAWA, 1986: *AKAIKE Information Criterion Statistics*. D. Reidel Publ. Comp., Dordrecht.
- SARICH, V.M. and A.C. WILSON, 1967: Immunological time scale for hominid evolution. *Science*, 158: 1200-1203.
- SCHWARTZ, J.H., 1984: The evolutionary relationships of man and orang-utans. *Nature*, 308: 501-505.
- SIBLEY, C.G. and J.E. AHLQUIST, 1984: The phylogeny of the hominoid primates, as indicated by DNA-DNA hybridization. *J. Mol. Evol.*, 20: 2-15.
- SIBLEY, C.G., J.A. COMSTOCK, and J.E. AHLQUIST, 1990: DNA hybridization evidence of hominoid phylogeny: a reanalysis of the data. *J. Mol. Evol.*, 30: 202-236.
- UEDA, S., Y. WATANABE, N. SAITOU, K. OMOTO, H. HAYASHIDA, T. MIYATA, H. HISAJIMA and T. HONJO, 1989: Nucleotide sequences of immunoglobulin-epsilon pseudogenes in man and apes and their phylogenetic relationships. *J. Mol. Biol.*,

- 205: 85-90.
- WILSON, A.C., S.S. CARLSON and T.J. WHITE, 1977: Biochemical evolution. *Annu. Rev. Biochem.*, 46: 573-639.
- WOLFE, K.H., P.M. SHARP and W.-H. LI, 1989: Mutation rates differ among regions of the mammalian genome. *Nature*, 337: 283-285.
-

長谷川 政美 統計数理研究所
〒106 東京都港区南麻布 4 - 6 - 7

Masami HASEGAWA The Institute of Statistical Mathematics
4-6-7 Minami-Azabu, Minato-ku,
Tokyo 106, Japan