Amino acid substitution of proteins coded for in mitochondrial DNA during mammalian evolution

Jun ADACHI¹ and Masami HASEGAWA^{1,2}

¹*Department of Statistical Science, The Graduate University for Advanced Study, 4-6-7 Minami-Azabu, Minato-ku, Tokyo 106*

²*The Institute of Statistical Mathematics, 4-6-7 Minami-Azabu, Minato-ku, Tokyo 106, Japan*

(Received 14 November 1991)

ABSTRACT

Three Markov models (Dayhoff, Proportional and Poisson models; Hasegawa et al., 1992a) for amino acid substitution during evolution were used for maximum likelihood analyses of proteins coded for in mitochondrial DNA in estimating a phylogenetic tree among human, bovine and murids (mouse and rat) with chicken as an outgroup. It turned out that Dayhoff model is the most appropriate model among the alternatives in approximating the amino acid substitutions of proteins coded for in mitochondrial DNA. In spite of the presence of the complete sequence data of mitochondrial genomes, we could not resolve the trichotomy among human, bovine and murids, probably because the time length separating two branching events among these three lines was short and because chicken is too distant from mammals to be used as an outgroup. It was suggested that the average substitution rate of amino acids coded for in mitochondrial DNA is lower along the bovine line than those along the human or murid lines. Advantages of amino acid sequence analysis over nucleotide sequence analysis in phylogenetic study were discussed.

1. INTRODUCTION

It has been suggested that the evolutionary rate of nuclear DNA (in terms of nucleotide substitution) is higher in murids than in primates and bovine (Kikuno et al., 1985; Wu and Li, 1985; Li et al., 1990). From a statistical analysis of complete sequence data of human, bovine and mouse mitochondrial DNAs (mtDNAs), Hasegawa and Kishino (1989) suggested that the transversion rate is lower in the bovine mtDNA than in the human and mouse ones, but it was equivocal whether the rate in mouse is higher than that in human. Since an appropriate outgroup was lacking at that time, they assumed that the three species diverged almost simultaneously, although this assumption was not necessarily guaranteed. Recently, a complete mtDNA sequence from chicken was published (Desjardins and Morais, 1990), and in this work we used it as an outgroup to estimate the branching order among the three lines of placental mammals and to evaluate the possible heterogeneity of evolutionary rate of mtDNA among them.

2. MATERIALS AND METHODS

Since the selection (negative selection in the framework of the neutral theory; Kimura, 1983) should operate on a codon or an amino acid rather than on a nucleotide in a protein-coding gene, it seems to be better to deal with amino acid sequences than with nucleotide sequences in a phylogenetic study of the relationship among distantly related species such as human, bovine and murids, if we can approximate the process of amino acid substitution during evolution by an appropriate model. Kishino et al. (1990) have developed a maximum likelihood (ML) method for estimating protein phylogeny assuming a Markov model that is based on an empirical transition matrix compiled by Dayhoff et al. (1978) ('Dayhoff model'). The method does not assume constancy of the substitution rate. A PASCAL program implementing the ML method was written by Adachi and Hasegawa (1992), and it was used in this work. To see the robustness of the ML analysis against the violation of the model, we also used two simplified models (Hasegawa et al., 1992a); one ('Proportional model') assumes that, in an infinitesimally short time interval of dt , the probability $P_{ij}(dt)$ of the current amino acid i being replaced by another amino acid j is given by

$$P_{ij}(dt) = udt\pi_j, \quad (1)$$

where u is a parameter that determines the substitution rate and π_j is the frequency of the amino acid j , and the other ('Poisson model') assumes that $P_{ij}(dt)$ is given by

$$P_{ij}(dt) = udt. \quad (2)$$

Proportional model is an extension of Felsenstein's (1981) model for nucleotide sequences into amino acid sequences. Poisson model is the simplest and it has been used in most of the molecular phylogenetic studies even if not stated explicitly.

When the number of parameters differs among models in comparison, the goodness of the approximation of models can be compared by $AIC = -2 \times (\log\text{-likelihood}) + 2 \times (\text{number of free parameters})$. A model that minimizes AIC is considered to be the most appropriate model (Akaike, 1974; Sakamoto et al., 1986; for its application to phylogenetics, see Kishino and Hasegawa, 1990; Hasegawa et al., 1990). The number of parameters of Proportional model is larger than those of Dayhoff and Poisson models by 19, because Proportional model estimates π_j 's from the data.

To evaluate the extent to which the ML tree is a significantly better representation of the true tree, we calculated standard errors of the difference of log-likelihood of alternative trees from that of the ML tree by the formula of Kishino and Hasegawa (1989) and estimated bootstrap probability by the RELL (resam-

pling of estimated log-likelihood) method given by Kishino et al. (1990).

The mtDNA sequences provided for the analyses are from human (Anderson et al., 1981), bovine (Anderson et al., 1982), mouse (Bibb et al., 1981), rat (Gadeleta et al., 1989) and chicken (Desjardins and Morais, 1990). Amino acid sequences of all the protein-coding genes transcribed from the H-strand of mtDNA were used for the analyses, but the following sites were excluded; 1) initiation codons, 2) overlapping regions between two genes, 3) regions where alignments are ambiguous, and 4) sites that experienced deletion or insertion in any of the five species under consideration. The segments used for the analyses by excluding the above regions are as follows (by Anderson et al.'s numbering for the human mtDNA); ND1 (3316–4257), ND2 (4617–4670, 4758–5117, 5211–5387), COI (5907–6347, 6351–7442), COII (7589–7738, 7742–8245), ATPase 8 (8369–8392), ATPase 6 (8644–9072, 9097–9204), COIII (9210–9989), ND3 (10095–10154, 10158–10298, 10347–10403), ND4L (10500–10592, 10665–10760), ND4 (10763–10852, 10934–11242, 11348–12028, 12044–12127), ND5 (12571–12900, 12970–13611, 13663–13752, 13915–13941), COB (14750–15439, 15470–15883). Thus, the number of amino acid sites used in this work was 2955.

Since it has been suggested that COII has undergone exceptionally rapid evolution in the human line compared to the bovine and murid lines (Brown and Simpson, 1982; Cann et al., 1984), we analyzed also the data excluding COII (the number of sites is 2737).

3. RESULTS AND DISCUSSIONS

Table 1 shows numbers of differences between the species. The chicken sequence differs from bovine, human, mouse and rat at 791, 848, 835 and 840 among 2955 sites, and 723, 779, 766 and 771 among 2737 sites excluding COII. The numbers for bovine are smaller than those for human and for murids, suggesting that the evolutionary rate is lower in the bovine line than in the human and murid lines as was suggested by Hasegawa and Kishino (1989), while the rate is nearly the same between human and murids.

Table 1. Number of amino acid differences between species in the 2955 sites data (upper right half) and in the 2737 sites data excluding COII (lower left half)

	human	bovine	mouse	rat	chicken
human		525	600	612	848
bovine	470		477	493	791
mouse	543	458		213	835
rat	555	474	210		840
chicken	779	723	766	771	

The significance of the rate difference in the relative rate test can be evaluated by bootstrap resampling (Hasegawa et al., 1987). The test compares D_{bovine} , number of amino acid differences of bovine from chicken, with D_{man} , D_{mouse} and D_{rat} , that represent numbers of differences of human, mouse and rat, respectively, from chicken. Fig. 1 represents bootstrap histograms of $D_{man}-D_{bovine}$, $D_{mouse}-D_{bovine}$ and $D_{rat}-D_{bovine}$ (the 2736 sites data). All of the lower bounds of 99% confidence intervals are positive in this figure, and therefore we conclude that the amino acid substitution rate of proteins coded for in mtDNA is significantly lower in bovine than in human and murids.

The results of ML analysis are summarized in Table 2. Three trees are

Fig. 1. Bootstrap histograms of $D_{man}-D_{bovine}$, $D_{mouse}-D_{bovine}$ and $D_{rat}-D_{bovine}$ (the 2736 sites data; sample size of 10^4), where D_{bovine} , D_{man} , D_{mouse} and D_{rat} refer to numbers of amino acid differences of bovine, human, mouse and rat, respectively, from chicken.

Table 2. Phylogenetic relationships among human, bovine and murids (mouse and rat) with chicken as an outgroup

Tree topology	Dayhoff Model		Prop. Model		PoissonModel	
	Δl_i	P_i	l_i	P_i	l_i	P_i
2955 sites						
1. (man,(bovine,murid))	(-17750.24)	.8535	(-17844.36)	.6783	(-18986.24)	.7048
2. (murid,(man,bovine))	-22.39 ± 21.41	.1434	-12.65 ± 23.18	.2805	-14.60 ± 26.78	.2720
3. (bovine,(man,murid))	-44.27 ± 19.28	.0031	-30.10 ± 21.62	.0412	-38.83 ± 24.64	.0232
2737 sites excluding COII						
1. (man,(bovine,murid))	(-16455.44)	.5943	-11.52 ± 21.02	.2423	-12.00 ± 24.71	.2802
2. (murid,(man,bovine))	-5.06 ± 20.33	.3917	(-16482.23)	.6212	(-17584.27)	.6371
3. (bovine,(man,murid))	-28.61 ± 17.91	.0140	-17.09 ± 20.85	.1365	-24.43 ± 23.90	.0827

Δl_i is the difference of log-likelihood of tree- i from that of the ML tree, and $\pm$ is 1SE. The log-likelihood of the ML tree is shown in parentheses. Bootstrap probability, P_i , of tree- i being the maximum likelihood tree among three alternatives was estimated by the REL method (Kishino et al. 1990) with sample size of 10^4 .

possible among human, bovine and murids by using chicken as an outgroup. When the 2955 sites data are used, tree-1, ((bovine, murids), human), is invariably the ML tree among the alternatives irrespective of the model for amino acid substitutions. However, tree-2 of (murids, (bovine, human)) has log-likelihood lower than that of tree-1 only by 22.39 ± 21.41 , 12.65 ± 23.18 , and 14.60 ± 26.78 ($\pm$: 1SE), respectively, for Dayhoff, Proportional, and Poisson models, and cannot be excluded. On the other hand, tree-3 of (bovine, (murids, human)) is highly unlikely.

When COII is excluded (the 2737 sites data), Dayhoff model selects tree-1 again as the ML tree, while tree-2 has log-likelihood lower only by 5.06 ± 20.33 . The bootstrap probabilities of trees-1 and 2 are 0.5943 and 0.3917, and these trees are almost indistinguishable by the likelihood criterion. Proportional and Poisson models, on the other hand, select tree-2 as the ML tree, but the preference is again of no statistical significance. In this way, the ML tree sometimes differs among the different models for amino acid substitutions. The dependence of the ML tree on the assumed model for amino acid substitution does not necessarily mean that the ML method is sensitive against the violation of the assumed underlying model in estimating a tree topology, but merely indicates that the data is not sufficient enough to discriminate among the candidate trees, because any of the preferences are with no statistical significance.

By using AIC, we can compare the adequacy of the approximation of the three models for amino acid substitution. Dayhoff model is the most appropriate among the alternatives, while Proportional and Poisson models have higher AIC values by 226.24 ± 124.55 (91.58 ± 125.72) and 2471.96 ± 124.20 (2257.66 ± 125.76)

Table 3. Differences of AIC between models for amino acid substitutions. Δ AIC is the difference of AIC from that of the minimum AIC model, and $\pm$ is 1SE estimated by Kishino and Hasegawa's (1989) formula

Model	ML tree	Δ AIC
2955 sites		
Dayhoff	tree-1	0
Proportional	tree-1	+226.24 $\pm$ 124.55
Poisson	tree-1	+2471.96 $\pm$ 124.20
2737 sites excluding COII		
Dayhoff	tree-1	0
Proportional	tree-2	+91.58 $\pm$ 125.72
Poisson	tree-2	+2257.66 $\pm$ 125.76

Table 4. Amino acid frequencies of Dayhoff model and of proteins coded for in mtDNA analyzed in this work (average of human, bovine, mouse, rat and chicken). Amino acid frequencies of human proteins are shown in parentheses

Amino acid	Dayhoff model	2955 sites	2737 sites
Ala	0.087	0.074 (0.075)	0.076
Arg	0.041	0.021 (0.020)	0.020
Asn	0.040	0.036 (0.037)	0.037
Asp	0.047	0.020 (0.020)	0.018
Cys	0.033	0.007 (0.005)	0.007
Gln	0.038	0.023 (0.024)	0.023
Glu	0.050	0.026 (0.024)	0.023
Gly	0.089	0.060 (0.059)	0.062
His	0.034	0.029 (0.029)	0.029
Ile	0.037	0.088 (0.086)	0.088
Leu	0.085	0.162 (0.167)	0.162
Lys	0.081	0.023 (0.022)	0.023
Met	0.015	0.054 (0.050)	0.054
Phe	0.040	0.062 (0.059)	0.064
Pro	0.051	0.055 (0.055)	0.054
Ser	0.070	0.072 (0.071)	0.072
Thr	0.058	0.081 (0.089)	0.082
Trp	0.010	0.028 (0.029)	0.029
Tyr	0.030	0.036 (0.037)	0.035
Val	0.065	0.043 (0.041)	0.041

than Dayhoff model for the 2955 (2737) sites data (Table 3).

Dayhoff model assumes a stationary state in terms of amino acid frequencies, and given the transition matrix, the frequencies are uniquely given as represented in Table 4. However, as is obvious in the table, the frequencies of the amino acids coded for in mtDNA deviate from the stationary frequencies of Dayhoff model. For example, methionine, isoleucine, leucine and tryptophan are more abundant in our data than in Dayhoff model, and cysteine, arginine, aspartic acid, glutamic acid and lysine are less abundant. This is because the proteins analyzed in this work are of different kinds from those used by Dayhoff et al. (1978) in constructing the transition matrix. Proportional model assumes that stationary frequencies are the same as those of the data under analysis, and therefore is expected to give lower AIC value than Dayhoff model when the amino acid frequencies of the data differ drastically from those of Dayhoff model. Nevertheless, the mitochondrial data prefer Dayhoff model than Proportional model.

This may indicate that Dayhoff et al.'s transition matrix is not necessarily specific only to the proteins they used in constructing it, and that it is applicable also to the evolution of other proteins. The frequency of substitutions in the matrix correlate well with the physico-chemical similarity between amino acids. This suggests that amino acid substitutions with small physico-chemical changes occur more frequently than those with large ones, which is consistent with the neutral theory (Kimura, 1983). Therefore, it seems to be a good approximation to assume that the evolution of proteins under consideration follows Dayhoff model. Deviation of the actual amino acid frequencies from Dayhoff's ones is probably a minor problem in estimating tree topologies. Computer simulation suggests that the ML method is robust against such a deviation in estimating tree topologies (M.H. and M. Fujiwara, in preparation).

Hasegawa and Kishino (1989) have shown that the L-strand of human mtDNA has significantly higher C frequency and lower T and A frequencies than those of bovine and murids. They attributed this tendency to be due to a directional mutation pressure operating in the human line. This sort of directional change of base frequency could obscure the phylogenetic relationship among species. Ribosomal RNAs (rRNAs) widely used in molecular phylogenetics sometimes lead to an erroneous tree because of a biased G+C content. Although the cellular slime mold *Dictyostelium discoideum* has been consistently placed as an outgroup to Plantae-Fungi-Animalia from the rRNA sequences, this placing is likely to be erroneous because of the unusually low G+C content of *D. dictyostelium* rRNA, and amino acid sequence data of proteins are more reliable in such a case (Loomis and Smith, 1990; Hasegawa et al., 1992b). Table 4 shows that, in spite of the difference of base frequencies of the human mtDNA from the bovine and murid ones, amino acid frequencies does not differ significantly. This observation reinforces our argument favoring protein sequences in phylogenetic analyses.

Table 5 represents estimated numbers of amino acid substitutions per site along

Table 5. Estimated number of amino acid substitutions per site along each branch of the tree in Fig. 1. $\pm$ refers to 1SE

Tree	Branch	Estimated number of substitutions	
		2955 sites	2737 sites
Tree-1	1-human	0.1132 ± 0.0071	0.1074 ± 0.0072
	1-2	0.0355 ± 0.0046	0.0305 ± 0.0045
	2-3	0.0868 ± 0.0061	0.0898 ± 0.0065
	3-mouse	0.0336 ± 0.0037	0.0360 ± 0.0040
	2-3-mouse	0.1203 ± 0.0072	0.1258 ± 0.0076
	3-rat	0.0423 ± 0.0041	0.0452 ± 0.0044
	2-3-rat	0.1271 ± 0.0074	0.1350 ± 0.0078
	2-bovine	0.0678 ± 0.0055	0.0711 ± 0.0059
	1-chicken	0.2585 ± 0.0107	0.2579 ± 0.0111
Tree-2	1-2	0.0282 ± 0.0042	0.0304 ± 0.0044
	2-human	0.1366 ± 0.0075	0.1263 ± 0.0075
	2-bovine	0.0669 ± 0.0055	0.0697 ± 0.0058
	1-3	0.0699 ± 0.0058	0.0713 ± 0.0061
	3-mouse	0.0343 ± 0.0038	0.0368 ± 0.0041
	3-rat	0.0415 ± 0.0041	0.0444 ± 0.0044
	1-chicken	0.2653 ± 0.0108	0.2598 ± 0.0112
Tree-3	1-bovine	0.0617 ± 0.0054	0.0637 ± 0.0057
	1-2	0.0198 ± 0.0037	0.0211 ± 0.0039
	2-human	0.1347 ± 0.0075	0.1244 ± 0.0075
	2-3	0.0835 ± 0.0060	0.0865 ± 0.0064
	3-mouse	0.0338 ± 0.0038	0.0363 ± 0.0041
	3-rat	0.0419 ± 0.0041	0.0448 ± 0.0044
	1-chicken	0.2738 ± 0.0110	0.2690 ± 0.0113

branches in the trees shown in Fig. 2. If tree-1 is the case as our analysis prefers marginally, the average evolutionary rate of proteins coded for in mtDNA along the bovine line is lower than that along the murid line by a factor of 0.54 (0.55) for the 2955 (2737) sites data. This ratio could be an under-estimate, since the effect of multiple substitutions may be more effectively evaluated in murids than in bovine because of the additional branching (mouse/rat branching) in murids. However, since the estimated branch length of 2-mouse (2-rat) in tree-1 (the 2955 sites data) by the ML analysis of Dayhoff model excluding rat (mouse) is 0.1128 ± 0.0069 (0.1200 ± 0.0071), which approximately coincides with 0.1203 ± 0.0072 (0.1271 ± 0.0074) estimated by adding 2-3 and 3-mouse (3-rat), the effect of multiple substitutions seems to be minor in comparing the rate in bovine

Fig. 2. Candidate trees of proteins coded for in mtDNA. The horizontal length of each branch is proportional to the estimated number of substitutions by Dayhoff model (the 2955 sites data). The place of the root is arbitrary between the node 1 and chicken.

with that in murids.

From a phylogenetic analysis of protein-coding genes in nuclear DNA, Li et al. (1990) suggested that the order Rodentia is an outgroup to Primates and Artiodactyla (tree-2 in this paper). If this is the case, the rate of proteins coded for in mtDNA along bovine line is lower than that along human line by a factor of 0.49 (0.55) for the 2955 (2737) sites data. Anyway, slower rate in bovine than in human or in murids is obvious. However, relative rate between human and murids is uncertain because of the uncertainty of the branching order.

In spite of the large data set of complete mtDNA genomes, we could not resolve the branching order among Primates, Artiodactyla and Rodentia. This may be due to the time length separating two branching events among these three lines was short, and because chicken is too distant from the placental mammals to be used as an outgroup. The numbers of differences in COII deserve attention. The numbers of differences of human from bovine and murids (55 and 57) are much larger than that between bovine and murids (19), consistently with the rapid rate of human COII suggested by Brown and Simpson (1982) and Cann et al. (1984). However, chicken differs from human, bovine and murids at 69, 68 and 69 among 218 sites, and the plausible rapid rate in the human line cannot be

detected when chicken is used as an outgroup in the relative rate test. This may suggest that the accumulation of amino acid substitutions at some of the sites under study have nearly reached at the saturation level between mammals and chicken. Therefore, a more closely related outgroup than chicken is needed. Data from a marsupial mtDNA are highly desirable to resolve the branching order among the three mammalian lines studied in this work and to evaluate rate heterogeneity of mtDNA evolution among them.

Although we could not resolve the branching order among Primates, Artiodactyla and Rodentia with any statistical confidence, the rate heterogeneity of amino acid substitutions of proteins coded for in mtDNA are obvious among these lines, and the pattern of heterogeneity differs from that of nuclear DNA, in which Rodentia evolve more rapidly than Primates and Artiodactyla. The evolutionary rate can vary due to two causes; one is the change of the mutation rate, and the other is the alteration of the selective constraint (Kimura, 1987). Synonymous substitution rate of nucleotides is likely to be directly related to mutation rate. But in order to evaluate the synonymous rate, data from closely related species are needed. Further works are necessary in order to establish the phylogenetic relationship among the mammalian orders and to decompose the rate heterogeneity into these two factors.

This work was carried out under the Institute of Statistical Mathematics Cooperative Research Program (90-ISM CRP-69 and 91-ISM CRP-57), and was supported by grants from the Ministry of Education, Science, and Culture of Japan to M.H.

REFERENCES

- Adachi, J. and Hasegawa, M. (1992). Computer Science Monographs, Vol. 27. MOLPHY: Programs for Molecular Phylogenetics I—PROTML: Maximum Likelihood Inference of Protein Phylogeny. Institute of Statistical Mathematics, Tokyo.
- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Trans. Autom. Contr.* **19**, 716–723.
- Anderson, S., Bankier, A. T., Barrell, B. G., de Bruijn, M. H. L., Coulson, A. R., Drouin, J., Eperon, I. C., Nierlich, D. P., Roe, B. A., Sanger, F., Schreier, P. H., Smith, A. L. H., Staden, R. and Young, I. G. (1981). Sequence and organization of the human mitochondrial genome. *Nature* **290**, 457–464.
- Anderson, S., Bruijn, M. H. L., Coulson, A. R., Eperon, I. C., Sanger, F. and Young, I. G. (1982). The complete sequence of bovine mitochondrial DNA: conserved features of the mammalian mitochondrial genome. *J. Mol. Biol.* **156**, 683–717.
- Bibb, M. J., Van Etten, R. A., Wright, C. T., Walberg, M. W. and Clayton, D. A. (1981). Sequence and gene organization of mouse mitochondrial DNA. *Cell* **26**, 167–180.
- Brown, G. G. and Simpson, M. V. (1982). Novel features of animal mtDNA evolution as shown by sequences of two rat cytochrome oxidase subunit II genes. *Proc. Natl. Acad. Sci. USA* **79**, 3246–3250.
- Cann, R. L., Brown, W. M. and Wilson, A. C. (1984). Polymorphic sites and the mechanism of evolution in human mitochondrial DNA. *Genetics* **106**, 479–499.

- Dayhoff, M. O., Schwartz, R. M. and Orcutt, B. C. (1978). A model of evolutionary change in proteins. In: Atlas of Protein Sequence and Structure (ed: M. O. Dayhoff), Vol. 5, Suppl. 3, pp. 345–352, National Biomedical Research Foundation, Washington, D.C.
- Felsenstein, J. (1981). Evolutionary trees from DNA sequences: a maximum likelihood approach. *J. Mol. Evol.* **17**, 368–376.
- Gadaleta, G., Pepe, G., De Candia, G., Quagliariello, C., Sbisà, E. and Saccone, C. (1989). The complete nucleotide sequence of the *Rattus norvegicus* mitochondrial genome: cryptic signals revealed by comparative analysis between vertebrates. *J. Mol. Evol.* **28**, 497–516.
- Hasegawa, M. and Kishino, H. (1989). Heterogeneity of tempo and mode of mitochondrial DNA evolution among mammalian orders. *Jpn. J. Genet.* **64**, 243–258.
- Hasegawa, M., Kishino, H. and 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. and Horai, S. (1990). Mitochondrial DNA evolution in primates: transition rate has been extremely low in lemur. *J. Mol. Evol.* **31**, 113–121.
- Hasegawa, M., Cao, Y., Adachi, J. and Yano, T. (1992a). Rodent polyphyly? *Nature* **355**, 595–595.
- Hasegawa, M., Hashimoto, T., Adachi, J., Iwabe, N. and Miyata, T. (1992b). Early divergences in the evolution of eukaryotes: ancient divergence of *Entamoeba* that lacks mitochondria revealed by protein sequence data. *J. Mol. Evol.* (submitted).
- Kikuno, R., Hayashida, H. and Miyata, T. (1985). Rapid rate of rodent evolution. *Proc. Japan Acad.* **B61**, 153–156.
- 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 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. and Hasegawa, M. (1990). Converting distance to time: an application to human evolution. *Methods in Enzymology* **183**, 550–570.
- Kishino H., Miyata T. and Hasegawa M. (1990). Maximum likelihood inference of protein phylogeny and the origin of chloroplasts. *J. Mol. Evol.* **30**, 151–160.
- Li, W.-H., Gouy, M., Sharp, P.M., O'hUigin, C. and Yang, Y.-W. (1990). Molecular phylogeny of Rodentia, Lagomorpha, Primates, Artiodactyla, and Carnivora and molecular clocks. *Proc. Natl. Acad. Sci. USA* **87**, 6703–6707.
- Loomis, W.F. and Smith, D.W. (1990). Molecular phylogeny of *Dictyostelium discoideum* by protein sequence comparison. *Proc. Natl. Acad. Sci. USA* **87**, 9093–9097.
- Sakamoto, Y., Ishiguro, M. and Kitagawa, G. (1986). *Akaike Information Criterion Statistics*. D. Reidel Publ. Co., Dordrecht.
- Wu, C.-I. and Li, W.-H. (1985). Evidence for higher rates of nucleotide substitution in rodents than in man. *Proc. Natl. Acad. Sci. USA* **82**, 1741–1745.