Rodent polyphyly?

SIR — From phylogenetic analyses of amino-acid sequence data of several proteins, Graur *et al.*[1] suggested that the order Rodentia may not be monophyletic, and that the guinea-pig-like rodents (Caviomorpha) may have a separate origin within mammalian evolution from that of the rat-like rodents (Myomorpha) and the squirrel-like rodents (Sciuromorpha). This suggestion radically contradicts the traditional view of rodent monophyly, based most on comparative morphology. Graur *et al.* used a maximum parsimony (MP) method, but it is known that this method is sometimes misleading, particularly when the evolutionary rate differs among lineages[2,3]. Therefore, we have re-examined the data by the maximum likelihood (ML) method[4,5], which is robust against the violation of rate constancy[3].

The sequence data used in our analysis are from SWISSPROT data library, and are mostly the same as those used by Graur *et al.* We additionally used factor IX sequences[6] with factor X as an outgroup. Three different models for amino-acid substitutions were used in our ML analysis, and the results were compared with those of the MP analysis by the PROTPARS program of Felsenstein's PHYLIP package also used by Graur *et al.* The table shows the results of our analysis on the relationship among Caviomorpha, Myomorpha and Primates. Although the MP method strongly favours the rodent polyphyly tree III: ((Myomorpha, Primates), Caviomorpha) with the bootstrap probability of 0.96 over the traditional tree I: ((Caviomorpha, Myomorpha), Primates) and another rodent polyphyly tree II: ((Caviomorpha, Primates), Myomorpha) consistently with Graur *et al.*, the ML method does not give any significant preference of tree III irrespective of the assumed model for amino-acid substitutions and does not discriminate between trees I and III. When Artiodactyla data are added (lipocortin I is excluded because only a short segment has been sequenced for cow), the MP method gives a bootstrap probability of as high as 0.92 for a rodent polyphyly tree: (Caviomorpha (Myomorpha (Primates, Artiodactyla))) which Graur *et al.* preferred among 15 possible trees. On the other hand, the ML methods based on the Dayhoff and Proportional models give low probabilities of 0.20 and 0.18 for the tree, and give subtotals of bootstrap probabilities of 0.69 and 0.76 for the rodent monophyly trees.

From these results, we conclude that Graur *et al.*'s claim of rodent polyphyly is premature from the data accumulated up to now, and that their suggestion may represent an example of the fact that unequal evolutionary rates can mislead MP analysis[2,3]. Nevertheless, we admit the possibility that the phylogenetic distance beween Caviomorpha and Myomorpha is large enough to warrant a separate ordinal status for the Caviomorpha because of the difficulty in discriminating between trees I and III.

MASAMI HASEGAWA
YING CAO
JUN ADACHI
Institute of Statistical Mathematics, 4-6-7 Minami-Azabu, Minato-ku, Tokyo 106, Japan

TAKA-AKI YANO
College of Arts and Sciences, Showa University, Fuji-Yoshida, Yamanashi 403, Japan

ML AND MP ANALYSES OF AMINO-ACID SEQUENCES OF PROTEINS

Tree	Crys	Lact	Hbα	Hbβ	NGF	Fac9	Ribo	Insu	Lipa	Cort	Total	*Pr*
ML method with Dayhoff model												
I	ML	ML	−5.6	−0.5	ML	−4.0	−3.5	−1.6	−3.1	−8.2	−4.7	.3608
			±7.9	±5.0		±3.5	±3.0	±2.3	±3.5	±6.9	±16.0	
II	−0.4	−4.2	−10.5	−3.3	−9.1	ML	ML	ML	ML	−10.5	−16.1	.0515
	±7.0	±5.3	±6.0	±3.6	±6.0					±6.1	±14.0	
III	−5.6	−5.5	ML	ML	−2.2	−3.8	−3.0	−1.6	−0.2	ML	ML	.5877
	±4.5	±4.5			±8.3	±3.5	±3.5	±2.3	±4.7			
ML method with Proportional model												
I	ML	ML	−6.6	−1.9	ML	−2.9	−3.8	ML	−3.8	−8.8	ML	.5441
			±10.2	±5.4		±4.0	±3.9		±6.2	±7.9		
II	−6.3	−5.8	−13.4	−4.3	−12.1	ML	ML	−1.7	−0.7	−10.5	−26.9	.0118
	±9.1	±4.7	±7.8	±4.1	±7.1			±2.5	±7.3	±7.3	±16.8	
III	−11.0	−5.7	ML	ML	−5.6	−4.0	−2.9	−1.2	ML	ML	−2.5	.4441
	±7.6	±4.6			±9.1	±3.5	±4.5	±2.8			±20.2	
ML method with Poisson model												
I	ML	ML	−3.3	−2.8	ML	−3.2	−3.2	ML	−4.9	−8.5	ML	.5598
			±9.4	±5.5		±4.1	±3.5		±6.6	±8.5		
II	−5.9	−6.3	−10.9	−4.8	−11.6	ML	ML	−1.9	−1.5	−10.6	−27.5	.0110
	±9.5	±5.1	±6.5	±4.4	±7.1			±2.7	±7.8	±7.7	±17.5	
III	−11.2	−5.9	ML	ML	−5.0	−4.3	−2.0	−1.5	ML	ML	−4.0	.4292
	±7.7	±5.2			±9.1	±3.6	±4.3	±2.9			±20.4	
MP method												
I	MP	5	4	1	1	1	3	1	9	4	17	.0303
II	2	MP	8	5	5	MP	MP	1	6	5	20	.0108
III	3	4	MP	MP	MP	3	2	MP	MP	MP	MP	.9589

Tree 1 ((Caviomorpha, Myomorpha), Primates); tree II ((Caviomorpha, Primates), Myomorpha); tree III, ((Myomorpha, Primates), Caviomorpha). ML analyses were performed by three models for amino-acid substitutions; the Dayhoff model[5] assumes an empirical transition matrix, compiled by Dayhoff *et al.*[7], and the Proportional and Poisson models assume that the probability of an amino-acid i being replaced by another amino acid j during an infinitesimally short time interval of dt is given by $udt\pi_j$ (π_j is the frequency of j, and u is a parameter that determines the substitution rate)[4] and by udt, respectively. For the ML analyses, the highest likelihood tree for each protein is indicated as 'ML', and the differences of log-likehihoods of alternative trees from that of the ML tree are shown with their s.e. following ± estimated by Kishino and M. H.'s formulae[8]. 'Total' refers to summation of the log-likelihood differences from the diverse proteins. For the MP analyses, the most parsimonious tree is indicated as 'MP', and the differences of numbers of substitutions of alternative trees from that of the MP tree are shown. *Pr*, bootstrap probability for being the ML or MP tree among alternatives during bootstrap resampling estimated by the RELL (resampling of estimated log-likelihoods of respective sites or, for the MP method, resampling of estimated numbers of substitutions) method given in ref. 5. (sample size of 10^4). Investigated proteins: Crys, α-crystallin A; Lact, α-lactalbumin; Hbα, α-globin; Hbβ, β-globin, NGF, β-nerve growth factor; Fac9, factor IX; Ribo, pancreatic ribonuclease; Insu, proinsulin; Lipa, lipoprotein lipase; Cort, lipocortin I.

1. Graur, D., Hide, W. A. & Li, W.-H. *Nature* **351**, 649–652 (1991).
2. Felsenstein, J. *Syst. Zool.* **27**, 401–410 (1978).
3. Hasegawa, M., Kishino, H. & Saitou, N. *J. molec. Evol.* **32**, 443–445 (1991).
4. Felsenstein, J. *J. molec. Evol.* **17**, 368–376 (1981).
5. Kishino, H., Miyata, T. & Hasegawa, M. *J. molec. Evol.* **31**, 151–160 (1990).
6. Bottema, C. D. K. et al. *Am. J. hum. Genet.* **49**, 820–838 (1991).
7. Dayhoff, M. O., Schwartz., R. M. & Orcutt, B. C. in *Atlas of Protein Sequence and Structure* (ed. Dayhoff, M. O.) 345–352 (National Biomedical Research Foundation, Washington, DC, 1978).
8. Kishino, H. & Hasegawa, M. *J. molec. Evol.* **29**, 170–179 (1989).

Biased DNA repair

SIR — Hennecke *et al.*[1] report the first *in vitro* characterization of a DNA mismatch endonuclease. This enzyme, the *vsr* gene product, initiates very short patch repair of *E. coli* by nicking one strand of the duplex next to the T at a mismatched T:G in the sequence C$\underline{T}$WG or $\underline{T}$WGG. This specificity confirms extensive *in vivo* data on the *vsr*-initiated system[2]. Sequence-specific nicking is the first step in the repair of certain T:G mismatches to C:G and is apparently designed to target the mutation hot spot caused by 5-methylcytosine deamination to thymine at the *dcm* site: 5′ C $\underline{T}$ W G G 3′, where the underlined T is mismatched with G; and 3′ G G W ^{m}C C 5′, where W is A or T and ^{m}C is 5-methylcytosine.

However, methylation on the unmutated strand, C^{m}CWGG, in the *dcm* site is not required for the function of the enzyme[1]. There is an interesting con-