Mammalian phylogeny inferred from multiple protein data

Kei-ichi KUMA and Takashi MIYATA

*Department of Biophysics, Faculty of Science, Kyoto University,
Kyoto 606, Japan*

(Received 1 July 1994)

ABSTRACT

On the basis of multiple protein data, the phylogenetic relationships among the major clades of eutherian mammals, Primates, Cetacea, Artiodactyla, Carnivora, Lagomorpha, Myomorpha, and Caviomorpha, were analyzed by the maximum likelihood (ML) method, together with the maximum parsimony (MP) method and the neighbor joining (NJ) method. Using nineteen protein data, we first reexamined the hypothesis of rodent polyphyly proposed by Graur et al. (1991) based on the four taxon system, consisting of human, mouse (or rat), guinea-pig and an outgroup. The ML analysis does not support the rodent polyphyly, but strongly favors the traditional view of rodent monophyly representing the Myomorpha-Caviomorpha association. In the second analysis, the phylogenetic position of Lagomorpha were reexamined using three sets of multiple protein data, 13 mitochondrial DNA-coded proteins from rodents, rabbit, human, seals, bovine, whales and opossum, 25 protein data from human, rabbit and rodents and an outgroup, and 20 protein data from human, rabbit, bovine (or pig) and rodents. Our analysis favors the closer affinity of Lagomorpha to Primates than to Rodentia. The branching sequence of seven mammalian groups revealed by the present analyses is as follows: Myomorpha-Caviomorpha clade (Rodentia), rabbit, Primates, Carnivora, and Artiodactyla-Cetacea clade. Judging from the branch lengths measured by the numbers of synonymous substitutions, a series of divergence of these mammals is likely to be bush-like. The rapid rate of rodent evolution was reexamined in the light of the new phylogeny.

1. INTRODUCTION

The phylogenetic position of the guinea-pig is currently controversial issue: From a molecular phylogenetic analysis of fifteen different protein species by the maximum parsimony (MP) method, Graur et al. (1991) recently proposed the separate origins of the guinea-pig-like rodents (Caviomorpha) and the rat-like rodents (Myomorpha); the guinea-pig diverged before the separation of the primates from the myomorph rodents (rodent polyphyly). The rodent polyphyly was also supported by a mitochondrial 12S rRNA tree (Allard and Miyamoto, 1991). Phylogenetic analyses by the maximum likelihood (ML) method using almost the same protein data as used by Graur et al. (1991), however, did not give a significant support for the rodent polyphyly (Hasegawa et al., 1992). In contrast, evidence that a retroposed sequence (BC1 RNA) is present in both Myomorpha and Caviomorpha, but not in other mammalian orders supports the

rodent monophyly (Martignetti and Brosius, 1993).

The phylogenetic position of Lagomorpha is also controversial. Most paleontologists believe the Lagomorpha-Rodentia association, although they differ to one another in the divergence time (Li et al., 1990; Novacek, 1992 for reviews). In contrast, some molecular data favor the close affinity between Lagomorpha and Primates, whereas others support the branching order that Lagomorpha is an outgroup to Primates and Rodentia (Easteal, 1989; Bulmer et al., 1991; Li et al., 1990 for review). Recently a large amount of protein sequence data has been accumulated for the major groups of eutherian mammals. Using these data, the above molecular data should be reexamined on a statistically solid background.

Using multiple protein data sets, we here reexamine the phylogenetic relationships among Myomorpha, Caviomorpha, Primates, Lagomorpha, Carnivora, Cetacea and Artiodactyla by the ML method, together with the MP method and the neighbor joining (NJ) method (Saitou and Nei, 1987). On the basis of the newly reconstructed tree, the evolutionary rates of synonymous substitutions are compared among guinea-pig, myomorphs, bovine and human.

2. MATERIALS AND METHODS

Protein sequence data

To test the hypothesis of rodent polyphyly on a statistically solid background, we accumulated 19 protein data, each consisting of sequences from human, mouse (or rat), and guinea-pig as representatives of Primates (Pr), Myomorpha (My), and Caviomorpha (Ca), respectively, and an outgroup, largely from marsupials or chicken. To examine the phylogenetic position of Lagomorpha, three sets of multiple protein data were accumulated; thirteen mitochondrial (mt) DNA-coded proteins from human, mouse, rat, whales (*Balaenoptera musculus* and *B. physalus*), bovine, seals (*Phoca vitulina* and *Halichoerus grypus*) and opossum, twenty five different proteins from rabbit, human and mouse (or rat), and twenty protein data from rabbit, human, bovine and mouse or rat (outgroup). All the sequence data were taken from GenBank release 80.0.

Molecular phylogeny inference

For each protein data, the ML analysis was made by applying the extended version of ML method recently developed by Hasegawa's group (Kishino et al., 1990; Adachi and Hasegawa, 1992). To evaluate the statistical significance of tree topologies inferred by the ML method, we calculated the difference Δli ($= li - lmax$) of log-likelihood li of tree i from that $lmax$ of the maximum likelihood tree and the standard error $\pm SE$ by the method of Kishino and Hasegawa (1989). A bootstrap probability pi for a particular tree of being the highest likelihood tree among the alternatives during bootstrap resamplings (Felsenstein, 1985) was estimated approximately by REL (resampling estimated log-likelihood) method

(Kishino et al., 1990). We also calculated the total value of log-likelihoods of tree i ($Li = \sum Li$, where the summation runs over all the data sets used). In addition, the difference ΔLi ($= Li - Lmax$) of Li from $Lmax$ (the highest total log-likelihood), and the standard error $\pm SE$, as well as the total value of bootstrap probability P_i were also estimated. Details for the method of analysis have been described elsewhere (Nikoh et al., 1994).

Using the same data, analyses by the MP method (PROTPARS in Felsenstein's program package PHYLIP, version 3.5c) and the NJ method (Saitou and Nei, 1987) were also carried out. In the MP analysis, the difference Δn_i ($= n_i - n_{min}$) of the number n_i of substitutions of a tree i from that n_{min} of the most parsimonious tree as well as the bootstrap probability p_i were calculated for each of the data. Finally the total number of substitutions and bootstrap probabilities of tree i (ΔN_i and P_i , respectively) was estimated.

3. RESULTS AND DISCUSSION

Test for the hypothesis of rodent polyphyly

To test the hypothesis of rodent polyphyly, we analyzed four taxonomy system consisting of human (as a representative of Primates), mouse or rat (Myomorpha), guinea-pig (Caviomorpha), and an outgroup, by applying the ML method. For each data, the difference Δli of log-likelihood li of a tree i ($i=I, III$) and the bootstrap probability p_i were calculated. The results are summarized in Table 1. The ML method has advantages over other known tree-making methods in that it allows one to synthesize results on tree topologies inferred from different protein data: It is possible to estimate the total value of log-likelihoods (Li) and that (P_i) of bootstrap probabilities of different data, and thus the reliability of a particular tree topology can be evaluated overall (Kishino and Hasegawa, 1989; Kishino et al., 1990). The total values of ΔLi and P_i are also shown in Table 1.

The tree I of Table 1 represents the traditional view of rodent monophyly that Myomorpha and Caviomorpha form a natural clade. The tree III represents the rodent polyphyly, originally proposed by Graur et al. (1991). Judging from the total value of log-likelihoods of 19 protein data sets, consisting of 6504 sites in total, the ML method gives the significant preference of tree I. The total value of bootstrap probabilities is 0.88 which is remarkably higher than those of the other two. The protein data analyzed here include sequences from birds, reptile and amphibian as outgroups. Even if six data sets with marsupial sequences as outgroups were used, the total values of log-likelihoods and bootstrap probabilities are significantly higher in tree I ($Lmax = -7160.5$ and $P_I = 0.78$) than in tree III ($\Delta L_{III} = -12.6 \pm 12.0$ and $P_{III} = 0.11$) and tree II ($\Delta L_{II} = -12.8 \pm 12.1$ and $P_{II} = 0.11$).

The MP analysis based on the same data sets showed that the tree III is the MP tree, but is not significant; the tree I is also possible within the confidence interval

Table 1. Maximum likelihood analysis for the phylogenetic relationships among guinea-pig, myomorphs and primates

Proteins	No. of Sites Compared	outgroup	l_{max}	Δl_i		
				tree I ((Gp,My),Pr)	tree II ((Gp,Pr),My)	tree III ((Pr,My),Gp)
1 α -1 antitrypsin	409	Opossum	-3045.5	ML	-5.7 ± 6.0	-6.6 ± 5.6
2 α -crystallin A chain*	172	Kangaroo	-622.3	ML	-0.5 ± 7.0	-5.6 ± 4.5
3 β -globin	146	Kangaroo	-857.0	ML	-2.9 ± 3.1	-0.0 ± 4.8
4 α -globin	141	Kangaroo	-758.9	-2.5 ± 6.8	-6.7 ± 5.2	ML
5 α -lactalbumin	139	Kangaroo	-1046.7	ML	-0.2 ± 6.1	-3.0 ± 4.8
6 Pancreatic RNase	122	Kangaroo	-826.8	-0.8 ± 1.3	ML	-0.7 ± 1.4
7 Complement C3	1633	Cobra	-10511.8	ML	-0.2 ± 11.5	-9.7 ± 10.0
8 Transglutaminase	663	Chicken	-3923.3	-2.4 ± 7.6	-3.0 ± 7.3	ML
9 Lipoprotein lipase	462	Chicken	-2385.2	-1.1 ± 4.0	-1.3 ± 3.8	ML
10 Lipocortin	342	Pigeon	-1714.6	-4.5 ± 7.0	-8.4 ± 5.5	ML
11 NGF β chain	236	Chicken	-1343.9	<u>ML</u>	-9.2 ± 6.2	-9.3 ± 6.2
12 VP-NP precursor	142	Chicken	-841.9	-8.8 ± 5.2	-8.8 ± 5.2	<u>ML</u>
13 Phospholipase A ₂ *	140	Cobra	-977.4	-4.7 ± 3.8	-4.0 ± 4.2	ML
14 Glucagon*	131	Chicken	-632.1	-0.1 ± 1.2	ML	-0.3 ± 0.9
15 IGF1*	127	Chicken	-522.8	-1.5 ± 2.3	-0.8 ± 2.7	ML
16 Insulin	107	Chicken	-705.6	-0.0 ± 0.3	-0.0 ± 0.2	ML
17 H ⁺ -, K ⁺ -ATPase*	985	Bufo	-4731.0	<u>ML</u>	-10.4 ± 9.1	-16.3 ± 7.6
18 C-reactive protein	219	Xenopus	-1602.7	-0.6 ± 4.4	ML	-3.2 ± 3.1
19 Corticotrophin	188	Xenopus	-1184.4	ML	-1.8 ± 2.3	-0.8 ± 3.0
Total	6504					
ML method	ΔLi		-38260.7	ML	-37.0 ± 22.4	-28.7 ± 23.0
	P_i			0.88	0.03	0.09
MP method	ΔN_i			13 ± 18.9	14 ± 18.9	MP
	P_i			0.17	0.17	0.66
NJ method	M_i			9	5	5

$\Delta l_i = l_i - l_{max}$, where l_i and l_{max} are the log-likelihood of tree i and that of the maximum likelihood tree, respectively. For each protein data, the values of Δl_i and l_{max} are shown for possible three trees. ML, the maximum likelihood tree with the highest log-likelihood value (i.e., $\Delta l_i = 0.0$). Gp, Guinea-pig; My, myomorphs; Pr, primates. ((Gp,My),Pr), for example, represents the association of Gp and My to form a cluster and Pr is an outgroup. For myomorphs mouse or rat (indicated by*) sequences were used; for primates human sequences were used. The ML tree whose l_{max} is significantly higher than those of the other trees is underlined. In "Total" the total values of 19 data are shown for each of the ML, maximum parsimony (MP) and neighbor joining (NJ) methods; $\Delta Li = Li - Lmax$, where $Li = \sum l_i$, the total value of log-likelihoods of tree i over 19 data, and $Lmax$ ($= -38260.7$) is the total log-likelihood of ML tree; P_i , total bootstrap probability; $\Delta N_i = N_i - N_{min}$, where N_i ($= \sum n_i$) and N_{min} are the total numbers of substitutions of tree i over 19 data, and the total numbers of substitutions of the MP tree, respectively; M_i , the numbers of proteins by which the highest bootstrap probabilities are given for tree i.

(Table 1). An additional analysis was also made by the NJ method. The NJ method supports the tree I: Although in NJ method, it is not possible to evaluate results on tree topologies overall, the tree I has the highest bootstrap probability in 9 cases out of 19 protein data examined (Table 1). It has already been known that the MP method is often misleading, when the rate of evolution differs for different lineages (Felsenstein, 1978; Iwabe et al., 1991; Hasegawa et al., 1991).

From these results we conclude that the molecular phylogenetic analysis based on new protein data strongly supports the Caviomorpha-Myomorpha association, which is consistent with the traditional view of rodent monophyly.

Phylogenetic relationships among Primates, Myomorpha, Cetacea, Artiodactyla, and Carnivora

At present, the complete nucleotide sequences of mitochondrial genomes from human, mouse, rat, whales (*Balaenoptera musculus* and *B. physalus*), bovine, seals (*Phoca vitulina* and *Halichoerus grypus*) and opossum are available, and thus it is possible to infer the phylogenetic relationships among five eutherian orders, Primates (Pr), Myomorpha (My), Cetacea (Ce), Artiodactyla (Ar), and Carnivora (Ca), based on thirteen different protein data, considering Marsupialia (Ma) as an outgroup; in Ce, My and Ca, the two species or genera were assumed to form a clade in each order in the ML analysis. An ML analysis using the chicken sequence as an outgroup supports that the divergence of Marsupialia antedates the divergence of the five eutherian mammals at the significance level of 2SE (Data not shown).

As shown in Table 2, the total value of log-likelihoods of 13 protein data gives the significant preference of the branching sequence (My, (Pr, (Ca, (Ce, Ar)))). This tree is significantly higher in the total values of log-likelihoods and bootstrap probabilities than the second ML tree (My, (Pr, (Ce, (Ca, Ar)))) (i.e., $\Delta L_I = 0.0$ and $P_I = 0.86$, and $\Delta L_{II} = -33.7 \pm 25.3$ and $P_{II} = 0.08$). The MP method also support the former tree, although it is not significant. In addition, the NJ method also favors the former tree; the former has the largest bootstrap probability in 7 protein data out of 13. Thus, because Rodentia is monophyletic as shown in Table 1, the branching sequence of the five mammals is highly likely to be Rodentia, Primates, Carnivora, and Artiodactyla-Cetacea clade. The branching sequence is consistent with that by Janke et al. (1994).

Phylogenetic position of Lagomorpha

Because molecular data contradict fossil data on the evolutionary position of Lagomorpha (Li et al., 1990; Novacek, 1992, for reviews), a phylogenetic tree among Rodentia, Primates, Artiodactyla and Lagomorpha were reexamined using newly accumulated protein sequences. We first examined the relationship among mouse (or rat), human, and rabbit using 25 protein data. The results are shown in Table 3. The ML method strongly favors the tree I, representing the Lago-

Table 2. Maximum likelihood analysis for the phylogenetic relationships among myomorphs, human, seals, whales, and bovine, based on thirteen mtDNA-coded protein data

Proteins	No. of Sites Compared	l_{max}	Δl_i		
			tree I (((B,W),S),H),M)	tree II (((B,S),W),H),M)	tree III (((B,W),S),M),H)
1 ND5	597	-5953.3	ML	-5.1±8.6	-16.2±10.9
2 CO I	513	-2378.1	-11.7±7.6	-24.9±12.9	-5.8±9.2
3 ND4	459	-3856.8	ML	-11.2±12.8	-8.6±9.0
4 Cyt b	379	-2628.8	ML	-7.2±7.9	-5.1±7.2
5 ND2	345	-3778.7	-4.3±6.3	-7.3±5.2	-13.7±8.9
6 ND1	315	-2290.7	-11.5±9.0	-4.8±10.5	-12.5±8.8
7 CO III	261	-1442.9	-6.3±15.4	-14.2±18.8	-19.3±12.5
8 CO II	227	-1347.9	-2.4±7.1	-2.5±6.2	-6.0±9.5
9 ATP6	226	-1766.3	-20.3±14.4	-13.9±15.5	-8.4±9.7
10 ND6	162	-1688.4	-3.1±5.9	-3.0±6.0	-2.6±3.0
11 ND3	113	-936.2	-3.0±9.3	-1.5±9.8	-8.4±7.4
12 ND4L	97	-850.5	-6.5±5.1	-0.8±2.9	-6.5±5.2
13 ATP8	63	-792.7	-1.0±2.7	-7.3±5.6	ML
Total	3757				
ML method	ΔLi	-29781.2	ML	-33.7±25.3	-43.0±24.0
	P_i		0.86	0.08	0.03
MP method	ΔN_i		MP	6±11.7	23±15.2
	P_i		0.63	0.31	0.06
NJ method	M_i		7	2	1

$\Delta l_i = l_i - l_{max}$, where l_i and l_{max} are the log-likelihood of tree i and that of the maximum likelihood tree, respectively. ML, the maximum likelihood tree with the highest log-likelihood value (i.e., $\Delta l_i = 0.0$). For each protein data, the values of Δl_i and l_{max} are shown only for the ML tree (tree I) and the second (tree II) and third (tree III) ML trees among 105 possible trees. Opossum was used as an outgroup. (((B,W),S),H),M), for example, represents the branching sequence of M (mouse and rat), H (human), S (seals), W (whales) and B (bovine). The ML tree whose l_{max} is significantly higher than those of the other trees is underlined. In "Total" the total values of 13 data are shown for each of the ML, maximum parsimony (MP) and neighbor joining (NJ) methods; $\Delta Li = Li - L_{max}$, where $Li = \sum l_i$, the total value of log-likelihoods of tree i over 13 data, and $L_{max} (= -29781.2)$ is the total log-likelihood of ML tree; P_i , total bootstrap probability; $\Delta N_i = N_i - N_{min}$, where $N_i (= \sum n_i)$ and N_{min} are the total numbers of substitutions of tree i over 13 data, and the total numbers of substitutions of the MP tree, respectively; M_i , the numbers of proteins by which the highest bootstrap probabilities are given for tree i.

morpho-Primates association; both the total values of log-likelihoods and bootstrap probabilities are significantly higher than tree II (the traditional view of Rodentia-Lagomorpha association) and tree III (the Rodentia-Primates association). The MP analysis also supports the tree I (Table 3). In addition, tree I is also favored by the NJ method; in the 13 protein data out of 25, tree I has the highest

Table 3. Maximum likelihood analysis for the phylogenetic relationships among Lagomorpha, Primates and Myomorpha

Proteins	out group	No. of Sites Compared	l_{max}	Δl_i		
				tree I ((La,Pr),My)	tree II ((La,My),Ar)	tree III ((Pr,My),La)
1 $\alpha 1$ -antitrypsin	Opossum	407	-2924.4	-7.2 $\pm$ 4.5	ML	-4.8 $\pm$ 5.5
2 CO II	Opossum	227	-1202.4	-0.0 $\pm$ 1.7	-0.1 $\pm$ 1.6	ML
3 α -crystallin A	Kangaroo	172	-624.4	ML	-2.5 $\pm$ 7.9	-8.5 $\pm$ 6.1
4 β Hemoglobin	Kangaroo	144	-767.0	-0.2 $\pm$ 2.1	-0.9 $\pm$ 1.5	ML
5 α Hemoglobin	Kangaroo	141	-722.5	-8.1 $\pm$ 6.1	ML	-7.7 $\pm$ 6.1
6 RAG-1	Chicken	1034	-4988.4	ML	-3.1 $\pm$ 5.4	-4.0 $\pm$ 5.0
7 Anion exchange protein	Chicken	842	-4234.9	ML	-8.1 $\pm$ 5.5	-8.9 $\pm$ 5.2
8 Progesterone receptor	Chicken	771	-4244.6	ML	-2.4 $\pm$ 6.0	-6.6 $\pm$ 4.7
9 Prolactin receptor	Chicken	538	-3642.6	ML	-9.0 $\pm$ 6.6	-5.0 $\pm$ 7.7
10 Thyroid hormone BP	Chicken	487	-2094.4	-6.8 $\pm$ 7.6	-6.7 $\pm$ 7.5	ML
11 Creatine kinase	Chicken	381	-1439.8	-1.5 $\pm$ 5.1	ML	-2.7 $\pm$ 4.5
12 α -actin	Chicken	377	-1111.3	-0.0 $\pm$ 0.0	-0.0 $\pm$ 0.0	ML
13 γ -actin	Chicken	375	-1140.4	-8.6 $\pm$ 6.7	-8.6 $\pm$ 6.7	ML
14 LDH	Chicken	332	-1422.4	-1.4 $\pm$ 5.6	ML	-4.4 $\pm$ 4.5
15 GLUT-1	Chicken	331	-1237.3	-1.0 $\pm$ 1.8	ML	-1.0 $\pm$ 1.8
16 ADH	Chicken	328	-1712.3	ML	-3.3 $\pm$ 3.3	-1.3 $\pm$ 4.2
17 T cell receptor β	Chicken	275	-2251.9	ML	-1.7 $\pm$ 5.4	-5.4 $\pm$ 3.7
18 Prothrombin	Chicken	234	-1231.2	-10.8 $\pm$ 6.5	-10.9 $\pm$ 6.4	ML
19 Carbonic anhydrase II	Chicken	232	-1362.6	-0.8 $\pm$ 2.0	-0.0 $\pm$ 2.6	ML
20 Glutathion S-transferase	Chicken	221	-1338.3	ML	-15.6 $\pm$ 7.9	-12.4 $\pm$ 8.9
21 MARCKS	Chicken	192	-956.7	-3.6 $\pm$ 2.8	ML	-3.4 $\pm$ 2.9
22 C3	Cobra	722	-4651.8	ML	-9.2 $\pm$ 6.5	-10.0 $\pm$ 6.3
23 EF 1 α *	<i>Xenopus</i>	462	-1448.4	ML	-4.3 $\pm$ 4.2	-4.3 $\pm$ 4.3
24 Tryptophan hydroxylase	<i>Xenopus</i>	444	-1993.4	-1.8 $\pm$ 6.7	-7.2 $\pm$ 5.0	ML
25 Calreticulin	<i>Xenopus</i>	315	-1390.9	-11.7 $\pm$ 6.1	-6.8 $\pm$ 7.9	ML
Total		9984				
ML method			-50197.7	ML	-37.0 $\pm$ 23.8	-26.9 $\pm$ 24.7
				0.85	0.04	0.11
MP method				MP	59 $\pm$ 18.8	52 $\pm$ 18.8
NJ method				13	6	6

$\Delta l_i = l_i - l_{max}$, where l_i and l_{max} are the log-likelihood of tree i and that of the maximum likelihood tree, respectively. For each protein data, the values of Δl_i and l_{max} are shown for possible three trees. ML, the maximum likelihood tree with the highest log-likelihood value (i.e., $\Delta l_i = 0.0$). La, Lagomorpha; My, Myomorpha; Pr, Primates. ((La,Pr),My), for example, represents the association of La and Pr to form a cluster and My is an outgroup. For Lagomorpha rabbit sequences were used; for Myomorpha mouse or rat (indicated by*) sequences were used; for Primates human sequences were used. The ML tree whose l_{max} is significantly higher than those of the other trees is underlined. In "Total" the total values of 25 data sets are shown for each of the ML, maximum parsimony (MP) and neighbor joining (NJ) methods; $\Delta Li = Li - Lmax$, where $Li = \sum l_i$, the total value of log-likelihoods of tree i over 25 data, and $Lmax$ ($= -50197.7$) is the total log-likelihood of ML tree; P_i , total bootstrap probability; $\Delta N_i = N_i - N_{min}$, where N_i ($= \sum n_i$) and N_{min} are the total numbers of substitutions of tree i over 25 data, and the total numbers of substitutions of the MP tree, respectively; M_i , the numbers of proteins by which the highest bootstrap probabilities are given for tree i . Abbreviations of proteins: RAG-1, Recombination activating gene-1; C3, Complement component C3; Thyroid hormone BP, Thyroid hormone binding protein; EF 1 α , Elongation factor 1 α ; LDH-A, Lactate dehydrogenase A; GLUT-I, Glucose transporter type-I; ADH, Alcohol dehydrogenase; CO II, Cytochrome oxidase II; MARCKS, Myristoylated aranine-rich C kinase substrate.

Table 4. Maximum likelihood analysis for the phylogenetic relationships among Lagomorpha, Artiodactyla and Primates

Proteins	No. of Sites Compared	l_{max}	Δl_i			
			tree I ((La,Ar),Pr)	tree II ((La,Pr),Ar)	tree III ((Ar,Pr),La)	
1 PKC β	673	-2105.6	ML	-0.1 $\pm$ 0.4	-0.1 $\pm$ 0.4	
2 PKC α	672	-2201.9	ML	-3.0 $\pm$ 3.8	-3.0 $\pm$ 3.8	
3 Prolactin receptor	555	-3410.3	-5.4 $\pm$ 6.2	ML	-6.2 $\pm$ 6.1	
4 Thyroid hormone BP	506	-2027.1	-4.1 $\pm$ 4.4	-4.7 $\pm$ 4.1	ML	
5 GLUT-I	492	-1672.0	-10.9 $\pm$ 7.7	ML	-6.1 $\pm$ 9.5	
6 α 1-antitrypsin	411	-2754.7	-9.6 $\pm$ 7.2	-12.6 $\pm$ 6.1	ML	
7 Dipeptidase*	409	-2277.5	-11.3 $\pm$ 6.2	ML	-3.1 $\pm$ 8.7	
8 D-amino acid oxidase*	346	-1785.4	-0.4 $\pm$ 6.6	-3.0 $\pm$ 5.6	ML	
9 LDH-A	332	-1234.0	ML	-2.4 $\pm$ 6.7	-1.1 $\pm$ 7.3	
10 Calreticulin	316	-1216.9	-3.7 $\pm$ 8.5	-7.3 $\pm$ 7.4	ML	
11 Apo E	307	-1840.4	ML	-0.9 $\pm$ 5.8	-4.3 $\pm$ 4.5	
12 IL-1 α	264	-1743.6	ML	-8.4 $\pm$ 5.1	-2.4 $\pm$ 7.1	
13 Prothrombin	235	-1086.3	ML	-4.6 $\pm$ 5.6	-4.7 $\pm$ 5.5	
14 TNF α	232	-1262.1	-6.9 $\pm$ 6.0	-7.6 $\pm$ 5.7	ML	
15 CO II	227	-1122.1	-5.4 $\pm$ 4.9	-6.4 $\pm$ 4.5	ML	
16 TIMP	203	-1113.9	-4.0 $\pm$ 3.8	-4.4 $\pm$ 3.6	ML	
17 TNF β	194	-1041.8	ML	-8.2 $\pm$ 9.3	-11.5 $\pm$ 8.2	
18 α -crystallin A	172	-578.2	-1.7 $\pm$ 2.5	-1.7 $\pm$ 2.5	ML	
19 β Hemoglobin	144	-707.9	-0.5 $\pm$ 3.4	-0.6 $\pm$ 3.0	ML	
20 α Hemoglobin	141	-665.2	-4.4 $\pm$ 6.4	-3.0 $\pm$ 7.0	ML	
Total						
ML method	ΔL_i	6831	-31889.1	-25.8 $\pm$ 25.0	-36.4 $\pm$ 24.2	ML
	P_i			0.14	0.04	0.82
MP method	ΔN_i			17 $\pm$ 15.2	31 $\pm$ 14.5	MP
	P_i			0.13	0.01	0.86
NJ method	M_i			6	2	12

$\Delta l_i = l_i - l_{max}$, where l_i and l_{max} are the log-likelihood of tree i and that of the maximum likelihood tree, respectively. For each protein data, the values of Δl_i and l_{max} are shown for possible three trees. ML, the maximum likelihood tree with the highest log-likelihood value (i.e., $\Delta l_i = 0.0$). La, Lagomorpha; Ar, Artiodactyla; Pr, Primates. ((La,Ar),Pr), for example, represents the association of La and Ar to form a cluster and Pr is an outgroup. For Lagomorpha rabbit sequences were used; for Artiodactyla bovine or pig (indicated by*) sequences were used; for Primates human sequences were used. Rodentia was used as an outgroup. The ML tree whose l_{max} is significantly higher than those of the other trees is underlined. In "Total" the total values of 20 data are shown for each of the ML, maximum parsimony (MP) and neighbor joining (NJ) methods; $\Delta Li = Li - L_{max}$, where $Li = \sum l_i$, the total value of log-likelihoods of tree i over 20 data, and $L_{max} (= -31889.1)$ is the total log-likelihood of ML tree; P_i , total bootstrap probability; $\Delta N_i = N_i - N_{min}$, where $N_i (= \sum n_i)$ and N_{min} are the total numbers of substitutions of tree i over 20 data, and the total numbers of substitutions of the MP tree, respectively; M_i , the numbers of proteins by which the highest bootstrap probabilities are given for tree i. Abbreviations of proteins: PKC, protein kinase C; Thyroid hormone BP, Thyroid hormone binding protein; GLUT-I, Glucose transporter type-I; LDH-A, Lactate dehydrogenase A; Apo E, Apolipoprotein E; IL-1 α , Interleukin-1 α ; TNF, Tumor necrosis factor; CO II, Cytochrome oxidase II; TIMP, Tissue inhibitor of metalloproteinases.

bootstrap probabilities (Table 3).

Next, we considered the relationship among human, rabbit and bovine, based on 20 protein data. From the results of Table 2 and Table 3, it is reasonable to consider the mouse or rat as an outgroup of human, rabbit and bovine. The results by the ML analysis are shown in Table 4. Table 4 strongly supports that Lagomorpha is an outgroup of Primates and Artiodactyla. This tree topology was also supported by the MP method and the NJ method (Table 4). Combining the results of Table 3 and Table 4, we have the branching sequence of Rodentia, Lagomorpha, Primates-Artiodactyla clade. This branching order is consistent with that suggested by Li et al. (1990).

From the results of Tables 1–4, the branching order of the seven eutherian mammals was inferred as follows: After the separation of eutherian mammals from Marsupialia, the Myomorpha-Caviomorpha clade (Rodentia) branched off first, and in turn Lagomorpha, Primates and Carnivora branched off in order, and finally Artiodactyla-Cetacea clade branched off (Fig. 1). Both the rodent monophyly and the close affinity between Artiodactyla and Cetacea is compatible with the morphological perspective. The phylogenetic position of Lagomorpha, however, contradicts the traditional view that Rodentia and Lagomorpha are closely related to each other to form a superordinal clade (Novacek, 1992 for review). Although this result is significant at the level of 1SE and is consistent

Fig. 1. Phylogeny of eight mammals inferred from four sets of multiple protein data. This tree was inferred from the results of four analyses shown in Tables 1–4. The branch lengths are arbitrary. The divergence leading to rodents, rabbit, human, seal, and bovine-whale clade is bush-like (dotted area), judging from the short evolutionary distance between the nodes (see Fig. 2).

a)

	Human	Bovine	Mouse	Guinea pig
Human	————	0.448 ± 0.015	0.612 ± 0.019	0.562 ± 0.018
Bovine	————	————	0.692 ± 0.022	0.594 ± 0.019
Mouse	————	————	————	0.699 ± 0.022
Guinea pig	————	————	————	————

b)

Fig. 2. Unrooted tree of four mammals and the evolutionary distances measured by the numbers k_s 's of synonymous substitutions per site. a), The distance matrix between pairs of mammals. The average values of k_s 's over 11 genes and their standard errors $\pm$ SE between species are shown. The k_s 's were calculated by Miyata and Yasunaga's method (1980). The sequence data were taken from GenBank release 80.0. b), The unrooted tree of human, bovine, mouse/rat and guinea-pig. The branch lengths were calculated by the k_s 's shown in Fig. 2a.

with that of Li et al. (1990), the place of Lagomorpha should be confirmed by further analyses based on molecular data.

The evolutionary rates of synonymous substitutions of mouse, guinea-pig, human and bovine

Graur et al. (1991) examined the evolutionary rate of the guinea-pig relative to that of myomorphs based on the assumption of the rodent polyphyly (tree III of Table 1). Now that the traditional view of rodent monophyly has been confirmed, their discussion on the rapid evolutionary rates of rodents (Kikuno et al., 1985; Wu and Li, 1985) should be reexamined. According to Fig. 1, an unrooted tree among mouse (or rat), guinea-pig, human, and bovine is represented as shown in Fig. 2b. The root of this tree is expected to be placed somewhere between the nodes A and B of Fig. 2b.

Based on the unrooted tree, the branch lengths measured by the number k_S of synonymous substitutions per site were calculated using the distance matrix of Fig. 2a, which were obtained by comparing the nucleotide sequences of eleven nuclear genes between species by Miyata and Yasunaga's method (1980). From the branch lengths of Fig. 2b, the average evolutionary rate of synonymous substitutions of mouse genes is estimated to be 1.4 ($=0.44/0.31$) times faster than that of guinea-pig genes. Also the corresponding rate of bovine genes is 1.4 times faster than that of human genes. The corresponding ratio of mouse to bovine is in a range of 1.9–1.6, depending on the position of the root. The estimated distance between nodes A and B is very small, suggesting bush-like divergence among Myomorpha, Caviomorpha, Lagomorpha, Primates, Carnivora, and Artiodactyla-Cetacea clade.

After this work has been completed, we have learned that Cao et al. (1994a, b) independently made similar analyses as those shown in Table 1 and Table 2, and obtained essentially identical results.

We thank Prof. M. Hasegawa for showing us his manuscripts before publication. This work was supported in part by grants from the Ministry of Education, Science and Culture of Japan.

REFERENCES

- Adachi, J. and Hasegawa, M. (1992). MOLPHY: programs for molecular phylogenetics I. PROTML: maximum likelihood inference of protein phylogeny. *Computer Science Monographs*, **27**, Institute of Statistical Mathematics, Tokyo.
- Allard, M. W. and Miyamoto, M. M. (1991). Tests for rodent polyphyly. *Nature* **353**, 610–611.
- Bulmer, M., Wolfe, K. H. and Sharp, P. M. (1991). Synonymous nucleotide substitution rates in mammalian genes: Implications for the molecular clock and the relationship of mammalian orders. *Proc. Natl. Acad. Sci. USA* **88**, 5974–5978.
- Cao, Y., Adachi, J., Yano, T. and 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. and 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.* (in press).
- Easteal, S. (1990). The pattern of mammalian evolution and the relative rate of molecular evolution. *Genetics* **124**, 165–173.
- Felsenstein, J. (1978). Cases in which parsimony and compatibility methods will be positively misleading. *Syst. Zool.* **27**, 401–410.
- Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39**, 783–791.
- Graur, D., Hide, W. A. and Li, W.-H. (1991). Is the guinea-pig a rodent? *Nature* **351**, 649–652.
- Hasegawa, M., Cao, Y. and Adachi, J. (1992). Rodent polyphyly? *Nature* **355**, 595.
- Hasegawa, M., Kishino, H. and Saitou, N. (1991). On the maximum likelihood method in molecular phylogenetics. *J. Mol. Evol.* **32**, 443–445.
- Iwabe, N., Kuma, K., Kishino, H., Hasegawa, M. and Miyata, T. (1991). Evolution of RNA polymerases and branching patterns of the three major groups of Archaeobacteria. *J. Mol. Evol.*

- 32, 70–78.
- Janke, A., Feldmaier-Fuchs, G., Thomas, W. K., von Haeseler, A. and Pääbo, S. (1994). The marsupial mitochondrial genome and the evolution of placental mammals. *Genetics* **137**, 243–256.
- Kikuno, R., Hayashida, H. and Miyata, T. (1985). Rapid rate of rodent evolution. *Proc. Japan Acad.* **61(B)**, 153–156.
- 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., 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.
- Martignetti, J. A. and Brosius, J. (1993). Neural BC1 RNA as an evolutionary marker: Guinea pig remains a rodent. *Proc. Natl. Acad. Sci. USA* **90**, 9698–9702.
- Miyata, T. and Yasunaga, T. (1980). Molecular evolution of mRNA: A method for estimating evolutionary rates of synonymous and amino acid substitutions from homologous nucleotide sequences and its application. *J. Mol. Evol.* **16**, 23–36.
- Nikoh, N., Hayase, N., Iwabe, N., Kuma, K. and Miyata, T. (1994). Phylogenetic relationship of the kingdoms Animalia, Plantae, and Fungi inferred from twenty three different protein species. *Mol. Biol. Evol.* (in press).
- Novacek, M. J. (1992). Mammalian phylogeny: shaking the tree. *Nature* **356**, 121–125.
- Saitou, N. and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* **4**, 406–425.
- 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.