The Pattern of Mammalian Evolution and the Relative Rate of Molecular Evolution

Simon Easteal

Human Genetics Group, John Curtin School of Medical Research, The Australian National University, Canberra ACT 2601, Australia

Manuscript received May 5, 1989
Accepted for publication September 15, 1989

ABSTRACT

The rates of nucleotide substitution at four genes in four orders of eutherian mammals are compared in relative rate tests using marsupial orthologs for reference. There is no evidence of systematic variation in evolutionary rate among the orders. The sequences are used to reconstruct the phylogeny of the orders using maximum likelihood, parsimony and compatibility methods. A branching order of rodent then ungulate then primate and lagomorph is overwhelmingly indicated. The nodes of the nucleotide based cladograms are widely separated in relation to the total lengths of the branches. The assumption of a star phylogeny that underlies Kimura's test for molecular evolutionary rate variation is shown to be invalid for eutherian mammals. Excess variance in nucleotide or amino acid differences between mammalian orders, above that predicted by neutral theory is explained better by variation in divergence time than by variation in evolutionary rate.

THE neutral theory prediction that, assuming a constant mutation rate, the mean rate of molecular evolution is stochastically uniform among lineages can be tested by comparing the interlineage variance in rate with that expected assuming a Poisson distribution (KIMURA 1968). Several authors (OHTA and KIMURA 1971; LANGLEY and FITCH 1974; KIMURA 1983; GILLESPIE 1984, 1986a, b) have reported that the prediction is inconsistent with data for mammalian proteins.

This observation has presented a challenge to neutral theory. It has led, on the one hand, to the development of elaborate theories of molecular evolution episodically driven by natural selection (GILLESPIE 1984, 1986a, b); and on the other hand, to attempts to modify neutral theory to make it consistent with the observation by allowing for variable rates of mutation and variation in degrees of functional constraint among lineages (KIMURA 1987; TAKAHATA 1987).

Early approaches to testing for rate variance involved assumptions about either the divergence times of the compared species (OHTA and KIMURA 1971) or the branching order of the species (LANGLEY and FITCH 1974). More recently KIMURA (1983) devised a test for evolutionary rate variance based on the assumption that the species, proteins or genes being compared diverged from each other within a short period of time relative to the total length of time over which they have been evolving independently of each other, *i.e.*, that assumes a star phylogeny. The orders of eutherian mammals are often assumed to have diverged in this way, and it is for this reason (as well as for reasons of sequence availability) that the test has been applied to mammalian proteins.

The star-phylogeny assumption is critical to the appropriateness of the test. If the assumption is invalid then variance in nucleotide change might be explained by divergence-time variation rather than by evolutionary rate variation. There is in fact no sound empirical basis for assuming a star phylogeny for eutherian mammals. The pattern of early eutherian evolution is poorly documented in the fossil record (CLEMENS *et al.* 1979; NOVACEK 1982; KIELAN-JAWOROWSKA, BROWN and LILLEGRAVEN 1987) and the assumption is more a reflection of this lack of knowledge than of any pattern in the fossil data (WYSS, NOVACEK and MCKENNA 1987). The conclusion, derived from this or the earlier approaches, that the relative rates of mammalian protein evolution are inconsistent with neutral theory must remain tentative until the relative divergence times of mammalian orders can be more accurately determined.

Cladistic analysis of nucleotide sequences can be used to determine the branching order of taxa, but not their relative divergence times unless rate constancy is assumed. However, once a branching order is established, the hypothesis of rate constancy can be evaluated in another way–by the relative rate test (SARICH and WILSON 1967; WU and LI 1985; LI and TANIMURA 1987).

Using this approach to study published globin gene sequences in four mammalian orders (EASTEAL 1988) I demonstrated a branching sequence for these orders of rodents then ungulates then lagomorphs and primates. This branching order is consistent with some

The publication costs of this article were partly defrayed by the payment of page charges. This article must therefore be marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

derived from protein sequence data (CZELUSNIAK *et al.* 1982; GOODMAN *et al.* 1982) but not with others (PENNY and HENDY 1985). It is different from that proposed by KOOP and GOODMAN (1988) based on globin gene sequences using marsupial sequences as outgroups.

In my analysis I used maximum parsimony algorithms (HENDY and PENNY 1982; METROPOLIS *et al.* 1953) with paralogous genes as outgroups. The same branching order was indicated by three sets of genes (α-globins, β-globins and ϵ-globins), by a high proportion of bootstrap samples (FELSENSTEIN 1985) and by maximum likelihood (FELSENSTEIN 1981) analysis (my unpublished data). However as FELSENSTEIN (1978) and PENNY, HENDY and HENDERSON (1987) have pointed out, under some circumstances, particularly unequal rates of change, parsimony methods may result in convergence to an incorrect tree. LAKE's (1987) evolutionary parsimony method appears to overcome this problem; however, LI *et al.*'s (1987) computer simulations have shown that its performance is substantially worse than maximum parsimony unless rate differences are very great (approximately fourfold or more). Use of evolutionary parsimony would thus only seem appropriate when evolutionary rates are known to be highly variable.

As a check for rate constancy in my previous analysis (EASTEAL 1988) I compared the rate of divergence of orthologous (homologous via speciation) mouse and human genes with paralogs (homologs via gene duplication) from four different mammalian orders. No evidence of rate variation was observed; however, since the paralogs are substantially diverged, the stochastic error in the comparisons was high and moderate rate differences may not have been detected, although, if such differences exist, nonsignificant faster rates for mouse genes would be expected to be consistently observed, which was not the case.

Genomic sequences of marsupial β- and ϵ-globin genes (KOOP and GOODMAN 1988) as well as cDNA sequences of marsupial α-globin (WAINWRIGHT and HOPE 1985) and α-lactalbumin (C. COLLET, unpublished data) are now available. These provide the most closely related sequences that can be used without question as references for comparison of the evolutionary rates of eutherian genes.

In this paper I compare these genes from four eutherian orders (rodents, ungulates, lagomorphs and primates) with their marsupial orthologs. I demonstrate that the rate of sequence evolution does not vary among the eutherian orders. I then use the same sequences to determine the phylogeny of the orders and to investigate whether the star-phylogeny assumption is valid.

MATERIALS AND METHODS

Nucleotide sequences were obtained from the EMBL (1988 release 14) data base or directly from the published literature. The sources of the sequences are in EASTEAL (1988) except for: opossum ϵ^m- and β^m- (KOOP and GOODMAN 1988) and *Dasyurus* α-globin (WAINWRIGHT and HOPE 1985), human and guinea pig α-lactalbumin (HALL *et al.* 1982), goat α-lactalbumin (KUMAGAI *et al.* 1987) and Tammar wallaby α-lactalbumin (C. COLLET, unpublished data).

Multiway sequence alignment, using an iterative procedure (FENG and DOOLITTLE 1987) from the computer package ALIGN (D. SMITH, The Australian National University) was performed with a gap penalty of 2.5 and gap-length penalty of 0. Estimates of synonymous and nonsynonymous substitution rates in coding regions were made using the method of LI, WU and LUO (1985). In noncoding regions corrections were made for multiple substitutions as described by KIMURA (1983). Formal analysis of relative evolutionary rates was by the method of WU and LI (1985).

Phylogenetic analysis was performed using maximum parsimony, maximum likelihood and maximum compatibility methods. The algorithms used were all from the phylogenetic reconstruction computer package, PHYLIP (J. FELSENSTEIN, University of Washington). The maximum parsimony algorithms, DNAMETRO, DNAPENNY and DNAPARS use Metropolis-annealing (METROPOLIS *et al.* 1953), branch and bound (HENDY and PENNY 1982) and the method of FITCH (1971), respectively, to find the most parsimonious trees. Maximum compatibility trees obtained using the DNACOMP algorithm did not differ in any substantial way from the maximum parsimony trees and they are not presented. Maximum likelihood trees (FELSENSTEIN 1981) were obtained using the DNAML algorithm.

Of the four genes for which marsupial sequences were available only the three globin genes had been sequenced in all four eutherian orders. No rabbit (lagomorph) α-lactalbumin gene was available so that α-lactalbumin was not included in the phylogenetic analysis.

RESULTS

Relative rates of nucleotide substitution: The alignment of noncoding sequences from the β-like globin genes is shown in Figure 1. Alignment of the long second exons of these genes was problematic and they were not included in the analysis. For the α-globin and α-lactalbumin genes only the coding regions were compared. The coding region of the Tammar wallaby α-lactalbumin gene is three codons shorter than that of the goat and human genes and 39 codons shorter than that of the guinea pig gene at the 3′ end. Only the regions of the eutherian genes that aligned within the boundaries of the marsupial gene were compared.

The substitution rates between the eutherian and marsupial genes for synonymous and nonsynonymous sites in coding regions and noncoding regions are shown in Tables 1, 2 and 3, respectively. In all comparisons the mean synonymous rates are similar to the noncoding rates suggesting that the synonymous sites are evolving largely without functional constraint.

For all three kinds of site, evolutionary rates were compared using the relative rate test. The results (Tables 4–6) show that there are only three significant differences between orders. These are the nonsynonymous sites for the combined globin genes and the globin and α-lactalbumin genes combined between

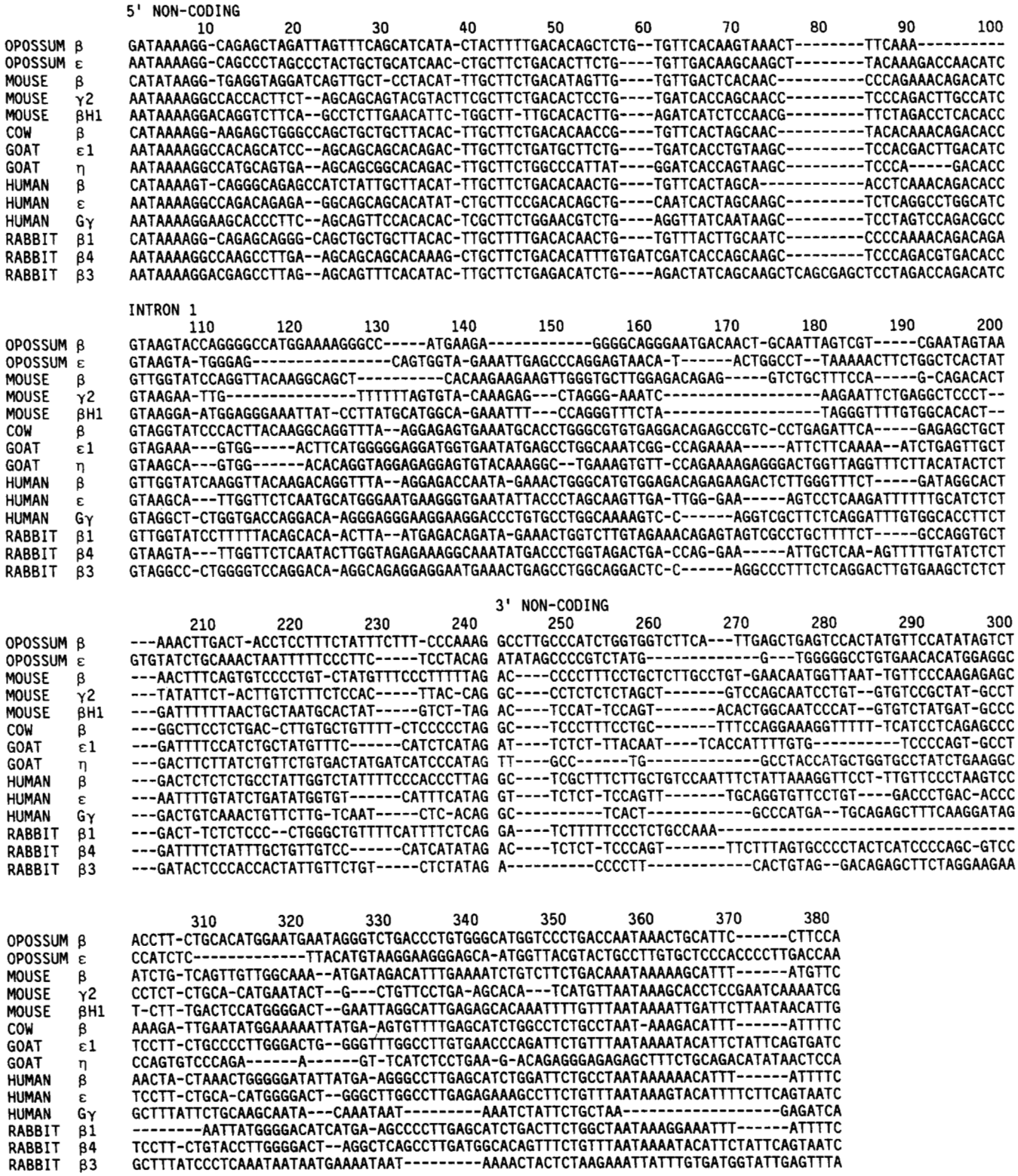

FIGURE 1.—Multiway alignment of 5′ (nucleotides 1–100) 3′ (nucleotides 242–381) noncoding sequences and intron 1 (nucleotides 101–241) for the mammalian functional β-like globin genes analyzed. A gap penalty of 2.5 was used for alignment. Gaps are indicated by dashes.

rodents and ungulates, and for the combined genes between rodents and primates. In all three cases the rodent genes have evolved faster. The differences are all significant at the 0.05 level. Since these differences are nonsynonymous, they may result from the differential effects of natural selection; however, three differences at this level of significance may be expected by chance from the total number of comparisons made particularly since they are not independent of each other. There are no significant differences in silent-site or noncoding region substitution rates. Overall there is no consistent pattern of variation among

TABLE 1

Rates of nucleotide substitution at *synonymous* sites between eutherian and marsupial genes

Gene	N	Eutherian order			
		Rodent	Ungulate	Lagomorph	Primate
α-Globin	104	102.7 ± 20.4	97.1 ± 17.6	106.5 ± 19.6	114.9 ± 24.1
β-Globin	100	90.0 ± 16.8	106.9 ± 19.4	101.2 ± 17.4	103.1 ± 17.9
ε-Globin	101	92.6 ± 16.3	61.5 ± 11.2	69.3 ± 12.4	67.2 ± 12.2
Globin mean	305	96.2 ± 10.1	85.2 ± 8.5	89.2 ± 9.0	91.2 ± 9.4
α-Lactalbumin	80	86.8 ± 17.5	97.2 ± 20.0		97.0 ± 20.1
Mean	385	94.3 ± 8.7	87.5 ± 7.8		92.1 ± 8.0

Rodent genes are all from mouse except α-lactalbumin from guinea pig. Ungulate genes are all from goat except β-globin from cow. Lagomorph genes are from rabbit and primate genes are from human.

TABLE 2

Rates of nucleotide substitution at *nonsynonymous* sites between eutherian and marsupial genes

Gene	N	Eutherian order			
		Rodent	Ungulate	Lagomorph	Primate
α-Globin	324	17.5 ± 2.6	13.9 ± 2.2	17.9 ± 2.6	13.8 ± 2.2
β-Globin	343	23.3 ± 3.0	20.7 ± 2.9	20.3 ± 2.7	20.8 ± 2.8
ε-Globin	344	13.2 ± 2.1	10.5 ± 1.9	11.5 ± 1.9	12.2 ± 2.0
Globin mean	1012	18.0 ± 1.5	14.9 ± 1.3	16.3 ± 1.4	15.5 ± 1.4
α-Lactalbumin	334	42.6 ± 4.7	38.5 ± 4.3		36.4 ± 4.1
Mean	1347	23.3 ± 1.5	20.1 ± 1.4		20.1 ± 1.4

TABLE 3

Rates of nucleotide substitution in noncoding regions between eutherian and marsupial genes

Gene	N	Eutherian order			
		Rodent	Ungulate	Lagomorph	Primate
β-Globin	314	98.3 ± 11.2	93.4 ± 10.3	101.9 ± 11.8	91.4 ± 9.7
ε-Globin	311	82.4 ± 8.8	77.3 ± 8.1	74.0 ± 7.5	80.3 ± 8.3
Mean	625	89.9 ± 6.9	84.9 ± 6.4	85.8 ± 6.5	85.6 ± 6.3

orders; all four orders have the highest and the lowest rate in different comparisons. There does not appear therefore to be any taxon-specific effect on rate such as cell-generation time or global mutation rate.

The magnitude of rate differences between taxa can be estimated approximately. Following LI and TANIMURA (1987), for a relative rate test the number of substitutions per site (K) between the reference sequence (3) and the point of divergence (0) of the two compared sequences (1 and 2) is obtained as: $K_{03} = (K_{13} + K_{23} - K_{12})/2$. The numbers of substitutions per site in the lineages leading to the compared sequences are estimated as: $K_{01} = K_{13} - K_{03}$ and $K_{02} = K_{23} - K_{03}$. The ratios, K_{01}:K_{02} for the comparisons showing significant differences, range from 1.65 to 1.81. For other "total" comparisons (Tables 4–6) and for the different genic regions combined (data not shown) the ratios are closer to 1.0. Overall the degree of differences in rate between taxa appear to be substantially less than twofold. For this degree of difference the simulations of LI *et al.* (1987) indicate that maximum parsimony is a more effective method of estimating phylogenies than is evolutionary parsimony.

Phylogenetic relationships: The single most parsimonious branching order of the functional genes of the β-globin complex in the four eutherian orders and in marsupials (obtained using DNAMETRO) is shown in Figure 2. The numbers at the branches in the tree indicate the percentage of bootstrap samples (of 100 samples) that gave the branch or cluster indicated. All branches are found in a high percentage of bootstrap samples except those leading to: (1) human β and rabbit β1 (58%); (2) human Gγ, rabbit β3, goat η and opossum ε (33%); (3) goat η and opossum ε (52%).

The topology of the tree is broadly consistent with that obtained previously by the same method (EASTEAL 1988) for the eutherian genes only. The only difference is that goat η (ε2) now appears to cluster with human Gγ and rabbit β3, whereas previously it clustered with the other nonadult expressed genes. The two branches leading to goat (and opossum η) are

TABLE 4

Comparisons of the numbers of *synonymous* substitutions per 100 nucleotides between orthologous eutherian genes by the relative rate test using marsupial genes for reference

	Gene					
	α-Globin	β-Globin	ε-Globin	Globin total	α-Lactalbumin	Total
N	104	105	101	310	80	390
K_{ru}	66.9	56.6	76.7	67.1	69.3	66.8
$K_{rm} - K_{um}$	5.6 ± 21.1	−16.9 ± 18.5	31.1 ± 16.5	11.0 ± 10.1	−10.4 ± 20.7	6.8 ± 8.9
$K_{r0}:K_{l0}$	1.18	0.54	2.36	1.39	0.85	1.23
K_{rl}	74.0	59.5	66.9	66.4		
$K_{rm} - K_{lm}$	−3.8 ± 22.6	−11.2 ± 17.7	23.3 ± 15.9	7.0 ± 10.3		
$K_{r0}:K_{l0}$	0.90	0.68	2.07	1.23		
K_{rp}	79.8	48.7	57.0	62.5	54.2	60.6
$K_{rm} - K_{pm}$	−12.2 ± 26.5	−13.1 ± 16.7	25.4 ± 15.1	5.0 ± 10.4	−10.2 ± 19.1	2.2 ± 8.5
$K_{r0}:K_{p0}$	0.73	0.58	2.61	1.17	0.68	1.07
K_{ul}	35.1	37.8	56.3	41.8		
$K_{um} - K_{lm}$	−9.4 ± 15.4	5.7 ± 14.6	−7.8 ± 12.8	−4.0 ± 7.7		
$K_{u0}:K_{l0}$	0.58	1.36	0.76	0.82		
K_{up}	38.5	36.6	46.0	40.0	34.1	39.4
$K_{um} - K_{pm}$	−17.8 ± 20.1	3.8 ± 14.7	−5.7 ± 12.0	−6.0 ± 8.0	0.2 ± 16.7	−4.6 ± 6.6
$K_{u0}:K_{p0}$	0.37	1.23	0.78	0.74	1.01	0.94
K_{lp}	31.5	35.6	55.4	40.7		
$K_{lm} - K_{pm}$	0.0 ± 23.6	−1.9 ± 13.4	2.1 ± 13.2	−2.0 ± 8.2		
$K_{l0}:K_{p0}$	0	0.90	1.08	0.91		

K_{ij} is the substitution rate between orders i and j; $i, j = m, r, u, p$ and l where m = marsupial, r = rodent, u = ungulate, p = primate, l = lagomorph. $K_{i0}:K_{j0}$ is the ratio of substitution rates between orders i and j since their divergence (at point 0) estimated as described in the text.

TABLE 5

Comparisons of the numbers of *nonsynonymous* substitutions per 100 nucleotides between orthologous eutherian genes by the relative rate test using marsupial genes for reference

	Gene					
	α-Globin	β-Globin	ε-Globin	Globin total	α-Lactalbumin	Total
N	320	337	345	1002	337	1339
K_{ru}	9.4	15.7	7.4	10.7	19.6	13.1
$K_{rm} - K_{um}$	3.6 ± 1.9	2.6 ± 2.7	2.7 ± 1.7	3.1 ± 1.2*	4.1 ± 3.9	3.2 ± 1.3*
$K_{r0}:K_{l0}$	2.24	1.40	2.15	1.81	1.52	1.65
K_{rl}	11.1	12.8	10.4	11.4		
$K_{rm} - K_{lm}$	−0.4 ± 2.2	2.5 ± 2.5	1.7 ± 1.9	1.7 ± 1.3		
$K_{r0}:K_{l0}$	0.93	1.48	1.11	1.35		
K_{rp}	8.7	12.5	8.8	9.9	17.0	11.6
$K_{rm} - K_{pm}$	3.7 ± 1.9	3.0 ± 2.3	1.0 ± 1.8	2.5 ± 1.3	6.2 ± 3.6	3.2 ± 1.2*
$K_{r0}:K_{p0}$	2.49	1.63	1.26	1.49	2.15	1.76
K_{ul}	10.3	9.2	7.6	9.1		
$K_{um} - K_{lm}$	−4.0 ± 2.0	0.1 ± 2.1	−1.0 ± 1.7	−1.4 ± 1.1		
$K_{u0}:K_{l0}$	0.44	1.02	0.77	0.73		
K_{up}	8.6	8.3	4.9	7.1	15.2	9.0
$K_{um} - K_{pm}$	0.1 ± 1.8	0.4 ± 1.9	−1.7 ± 1.4	−0.6 ± 1.0	2.1 ± 3.3	0 ± 1.1
$K_{u0}:K_{p0}$	1.02	1.10	0.49	0.84	1.32	1.0
K_{lp}	11.3	5.2	7.7	8.0		
$K_{lm} - K_{pm}$	4.1 ± 2.1	0.5 ± 1.5	−0.7 ± 1.6	0.8 ± 0.1		
$K_{l0}:K_{p0}$	2.14	1.21	8.33	1.22		

Abbreviations are as in Table 4. * $P < 0.05$.

the least reliable of all those in the tree. The clustering of goat η with opossum ϵ and of these genes with human Gγ and rabbit $\beta 3$ is therefore uncertain.

An important feature of the tree is that the eutherian branching order indicated by the adult-expressed genes, shown branching to the left, and the

TABLE 6

Comparisons of the numbers of nucleotide changes per 100 sites in noncoding regions between orthologous eutherian genes by the relative rate test using marsupial genes for reference

	Gene		
	β-Globin	ε-Globin	Total
N	314	311	625
K_{ru}	52.4	69.1	60.0
K_{rm}-K_{um}	4.9 ± 11.7	5.1 ± 9.7	5.0 ± 7.4
K_{r0}:K_{u0}	1.21	1.16	1.18
K_{rl}	44.8	73.0	62.2
K_{rm}-K_{lm}	−3.6 ± 11.5	8.5 ± 9.6	4.1 ± 7.4
K_{r0}:K_{u0}	0.85	1.26	1.14
K_{rp}	44.8	80.7	59.8
$K_{rm} - K_{pm}$	6.9 ± 10.8	2.1 ± 10.1	4.3 ± 7.2
K_{r0}:K_{p0}	1.36	1.05	1.15
K_{ul}	35.1	36.1	35.9
$K_{um} - K_{lm}$	−8.5 ± 10.4	3.3 ± 7.2	−0.9 ± 6.0
$K_{u0} - K_{l0}$	0.61	1.20	0.95
K_{up}	30.3	44.7	36.9
$K_{um} - K_{pm}$	2.0 ± 9.2	−3.0 ± 8.2	−0.7 ± 6.0
K_{u0}:K_{p0}	1.14	0.87	0.96
K_{lp}	31.4	34.0	33.0
$K_{lm} - K_{pm}$	10.5 ± 9.6	−6.3 ± 7.3	0.2 ± 5.8
K_{l0}:K_{p0}	2.0	0.69	1.01

Abbreviations are as in Table 4.

embryonic-expressed genes, shown branching to the top right, is the same. The order is rodent then ungulate then lagomorph and primate.

The tree was constructed by comparison of coding and noncoding sequences. The alignment of the noncoding sequences (Figure 1) was made using an iterative multiway method. While this is a more rational approach than the intuitive combining of two-way alignments, the alignments produced are not necessarily correct as they depend on an arbitrary choice of gap penalties. The noncoding sequences were included in the analysis to maximize the length of the compared sequences. However, because of alignment problems, their inclusion does introduce some uncertainty about the results. For this reason the coding and noncoding sequences were analysed separately.

The maximum likelihood trees, which have the same topologies as the corresponding maximum parsimony trees (using DNAPENNY) for the β- and α-globin genes and for the coding regions of three classes of genes combined (Figure 3) are the same as those in Figure 2. The proportions of bootstrap samples indicating this branching order for β-globin and total coding regions are approximately the same as the coding and noncoding regions combined (Figure 2). For the α-globin coding regions the branching order is indicated by a smaller proportion of bootstrap samples (59% indicate a primary branching of rodents and 46% indicate a lagomorph-primate monophyletic group). A different branching order is indicated for the ε-globin gene coding regions with the ungulates and primates forming a monophyletic group separate from the rodents and lagomorphs which form a separate group. Both of these groups however are indicated by a minority of bootstrap samples. Their validity is thus highly questionable and, rather than indicating an alternative branching order, analysis of the globin coding region alone would appear to be insufficient to resolve the branching order.

The maximum likelihood trees for the β- and ε-globin noncoding regions (Figure 4) both have topologies consistent with Figure 2. In both cases, a primary branching of rodents is indicated in a high proportion of bootstraps (100% and 95% for ε-globin and β-globin, respectively). Primates and lagomorphs cluster in 84% of bootstrap samples of β-globin noncoding regions but in only a low proportion of bootstrap samples of β-globin noncoding regions; the proportion is not shown in Figure 4 because an equally parsimonious topology exists that groups ungulates with primates, and which was indicated by a greater proportion of bootstrap samples (36%).

The rodent/ungulate/primate and lagomorph branching order is thus clearly indicated by one (β-globin) of three coding regions as well as the combined coding regions and by one (ε-globin) of two noncoding regions. The α-globin coding region also indicates this topology but with lower probability, and the β-globin noncoding region indicates this topology when analyzed by maximum likelihood and indicates that it is one of two most parsimonious topologies. Only one region (ε-globin coding region) positively identifies an alternative topology with both maximum likelihood and maximum parsimony methods, however the validity of this alternative topology has low probability as indicated by bootstrap sampling. The branching order is thus indicated by different genes and by both coding and noncoding regions. It does not appear to be an artifact arising from analysis of particular genes or from the incorrect alignment of noncoding sequences.

The topologies of the maximum parsimony and maximum likelihood trees based on the combined coding and noncoding regions are the same (Figure 5). A primary branching of rodents and a clustering of lagomorphs and primates are indicated in 100% and 86% of bootstrap samples respectively. The nodes of both trees are widely separated. The maximum parsimony and likelihood methods, respectively, indicate that in the ungulate, lagomorph and primate lineages, 43% and 37% of the average genetic change since their divergence from rodents occurred before the divergence of ungulates. Similarly the two methods respectively indicate that 34% and 22% of the average genetic change in the lagomorph and primate lineages, arising since their divergence from ungulates, occurred before they split (Figure 5).

A number of branching orders different from the

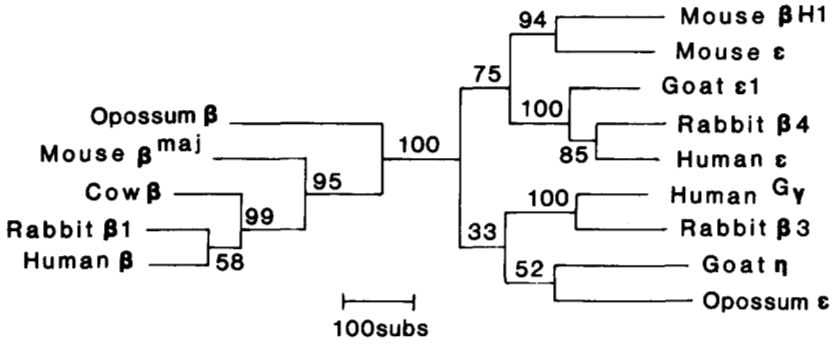

FIGURE 2.—Single most parsimonious unrooted cladogram of mammalian β-like globin genes (obtained using DNAMETRO). Coding sequences, and noncoding sequences shown in Figure 1 were analyzed. the percentage of (100) bootstrap samples showing the branches in the tree are indicated.

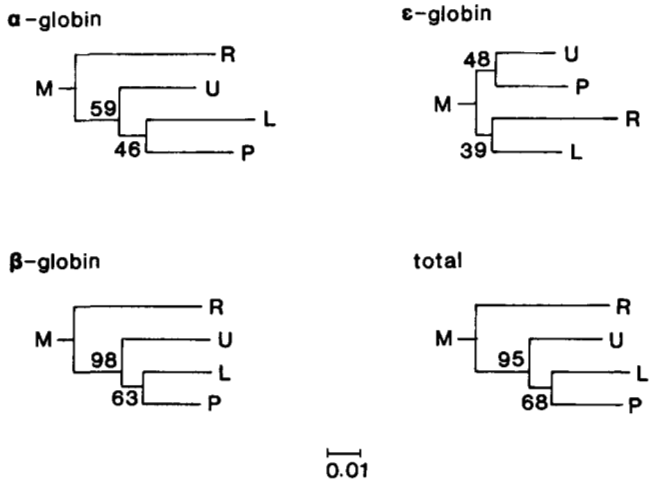

FIGURE 3.—Maximum likelihood cladograms for four eutherian orders (R = rodent, U = ungulate, L = lagomorph, P = primate) rooted by marsupial (M) outgroup based on the *coding regions* of individual and combined globin genes. The percentage of bootstrap samples showing the branches in the trees are indicated. The scale indicates the expected number of substitutions per site.

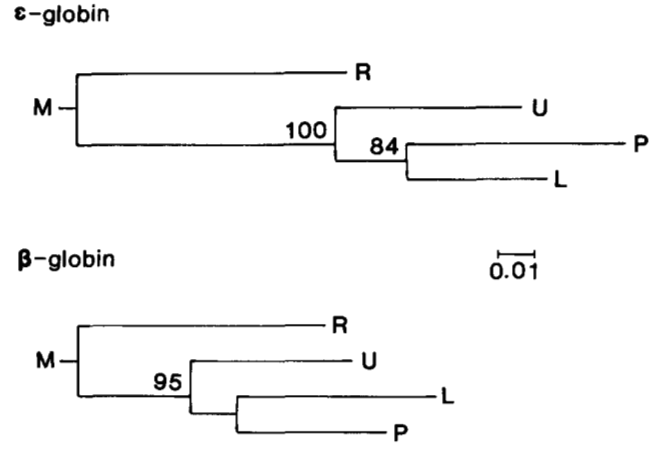

FIGURE 4.—Maximum likelihood cladograms for four eutherian orders rooted by marsupial outgroup based on the *noncoding regions* of ε and β-globin genes. The percentage of bootstrap samples showing the branches in the trees are indicated. Abbreviations are as in Figure 3. The scale indicates the expected number of substitutions per site.

one obtained here have been proposed. Three of these are compared in Table 7. Tree 1 is that obtained in the present study. Tree 2 is that proposed by KOOP and GOODMAN (1988) in which ungulates diverge first, followed by primates then rodents and lagomorphs. Tree 2 also indicates that primates and rodents are more closely related to each other than either is to ungulates, and thus that ungulates are an appropriate reference species for a relative rate test of rodents and primates as suggested by WU and LI (1985). Trees 3 and 4 are two other trees indicating a relatively close relationship between rodents and primates. Tree 3 has a branching order of ungulate then lagomorph then rodent and primate. In tree 4 there are two separate monophyletic groups, one consisting of ungulates and lagomorphs and the other of primates and rodents.

The number of nucleotide changes occurring in each of the trees was determined using DNAPARS with the topologies defined. For each of the three globin genes, for the combined coding regions and the combined noncoding regions, and for all regions combined, the numbers of nucleotide changes occurring in trees 2, 3 and 4 are similar and substantially greater than in tree 1. Trees 2, 3 and 4 do not therefore appear to be probable alternatives to tree 1.

DISCUSSION

An attempt has been made here to investigate the phylogenetic relations of four orders of eutherian mammals, using maximum parsimony, likelihood and compatibility methods. The appropriateness of these methods depends on there being approximate uniformity of evolutionary rate among the orders. A lack of rate variation has been demonstrated by relative rate comparisons of the eutherian genes with their marsupial orthologues. The phylogenetic analysis indicates that the orders are related in a dichotomously rather than a polychotomously branching pattern and that their branching points are widely separated; *i.e.*, that they do not form a star phylogeny. The observation (KIMURA 1983; GILLESPIE 1984, 1986a, b) that there is greater variance in evolutionary rate among eutherian orders than is predicted by neutral theory depends on the assumption of a star phylogeny. Since this assumption is not valid the increased variance cannot be taken as evidence of evolutionary rate variation (resulting either from natural selection or other

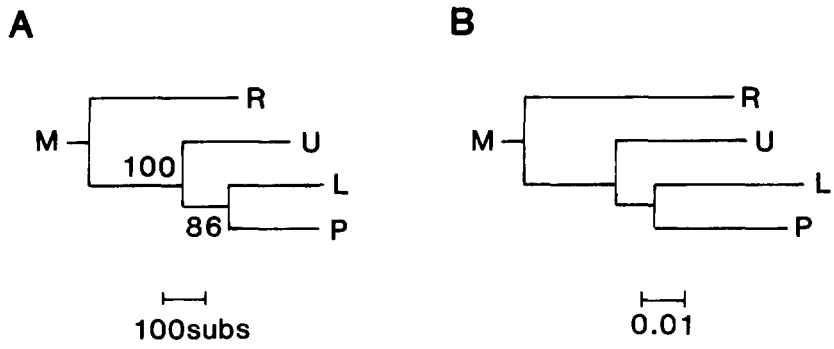

FIGURE 5.—Maximum parsimony (A) and maximum likelihood (B) cladograms for four eutherian orders rooted by marsupial outgroup based on the coding regions of α-, β- and ϵ-globin genes and the noncoding regions of β- and ϵ-globin genes combined. The percentage of bootstrap samples showing the branches of the maximum parsimony tree are indicated. Abbreviations are as in Figure 3. The scale for the maximum likelihood tree indicates the expected number of substitutions per site.

TABLE 7

Numbers of nucleotide changes for phylogenetic trees with different topologies

		Tree[a]			
Sequence	Length	1	2	3	4
α-Globin	429	239	244	247	247
β-Globin	841	706	728	724	730
ϵ-Globin	841	689	735	729	730
Coding	1317	697	715	717	717
Noncoding	794	937	992	983	990
Total	2111	1634	1707	1700	1707

[a] Tree topologies are as follows:

1. (Marsupial, (((Lagomorph, Primate), Ungulate), Rodent)).
2. (Marsupial, (((Lagomorph, Rodent), Primate), Ungulate)).
3. (Marsupial, (((Primate, Rodent), Lagomorph), Ungulate)).
4. (Marsupial, ((Primate, Rodent), (Lagomorph, Ungulate)).

factors), and may be better explained as reflecting variation in divergence time.

Earlier studies (OHTA and KIMURA 1971; LANGLEY and FITCH 1974) also reported greater variance in the rate of mammalian protein evolution than expected from neutral theory, although different approaches were used in arriving at this conclusion. OHTA and KIMURA's (1971) approach relied on assumptions about the geological divergence times of species derived from fossil record interpretation which may not be reliable. For instance they assumed a polychotomous divergence of eutherian orders (occurring 80 mya) which, as has been shown here, is an incorrect assumption. Their other assumed divergence times may be similarly incorrect.

LANGLEY and FITCH's (1974) analysis is based on an assumed eutherian phylogeny that has been shown here to be incorrect. Their approach to testing for evolutionary rate variation should be more sensitive than the relative rate test. However, although both methods involve assumptions about phylogeny, the only assumption made here in applying the relative rate test has been that eutherians comprise a monophyletic group when compared with marsupials. On the other hand, LANGLEY and FITCH's (1974) method requires a complete knowledge of the phylogeny of all compared species. It would be inappropriate to use the phylogeny derived from the DNA sequence data as a basis for testing for rate variation for the same data using the LANGLEY and FITCH approach.

The branching order indicated by the cladistic analysis (rodent then ungulate then primate and lagomorph) is quite different from that indicated by KOOP and GOODMAN's (1988) analysis of many of the same genes studied here (ungulate then primate then rodent and lagomorph), and from that assumed by WU and LI (1985) in their investigation of evolutionary rate variation (ungulate then primate and rodent).

The difference with KOOP and GOODMAN's (1988) phylogeny may result from their use of "local-branch-swapping" algorithm (GOODMAN *et al.* 1979) which is not guaranteed to find the most parsimonious tree. The branch-and-bound algorithm used here (DNAPENNY) is guaranteed to do so. In neither of the other studies was any attempt made to assess the reliability of the trees obtained. In the present study the appropriateness of the use of maximum parsimony, compatibility and likelihood methods was established by a lack of evidence of evolutionary rate variation among taxa. This was assessed independently of the cladistic analysis. The tree obtained here is shown to be reliable by its being indicated from analysis of different genes and of both coding and noncoding regions, and by a high proportion of bootstrap samples. It is also a substantially more parsimonious tree than the alternatives discussed above.

The results reported here emphasize the importance of accurate phylogenetic information in investigations of evolutionary processes and demonstrate how this information can be reliably obtained from DNA sequence data. Previous comparative studies of molecular data from different eutherian orders (LANGLEY and FITCH 1974; KIMURA 1983; GILLESPIE 1984, 1986a, b; WU and LI 1985) have concluded that the rate of molecular evolution is variable among lineages. These conclusions depend on phylogenetic assumptions that appear from the present analysis to be incorrect. The phylogeny presented here is consistent with the results of the relative rate tests which showed that overall there is stochastic uniformity of molecular evolutionary rate among the eutherian orders.

I am grateful to C. COLLET for making available the sequence of

the Tammar wallaby α-lactalbumin gene before publication and to J. H. GILLESPIE and D. PENNY for their valuable comments. I also thank D. SMITH, W.-H. LI and J. FELSENSTEIN for providing their computer programs for sequence alignment, estimation of rates of synonymous and nonsynonymous substitutions and phylogenetic reconstruction respectively.

LITERATURE CITED

CLEMENS, W. A., J. A. LILLEGRAVEN, E. H. LINDSAY and G. G. SIMPSON, 1979 Where, when and what–a survey of known Mesozoic mammal distribution. pp. 7–58 in *Mesozoic Mammals*, edited by J. A. LILLEGRAVEN, Z. KIELAN-JAWOROWSKA and W. A. CLEMENS. University of California Press, Berkeley.

CZELUSNIAK, J., M. GOODMAN, D. HEWETT-EMMETT, M. L. WEISS, P. J. VENTA and R. E. TASHIAN, 1982 Phylogenetic origins and adaptive evolution of avian and mammalian haemoglobin genes. Nature **298:** 297–300.

EASTEAL, S., 1988 Rate constancy of globin gene evolution in placental mammals. Proc. Natl. Acad. Sci. USA **85:** 7622–7626.

FELSENSTEIN, J., 1978 Cases in which parsimony and compatibility methods will be positively misleading. Syst. Zool. **27:** 401–410.

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.

FENG, D.-F., and R. F. DOOLITTLE, 1987 Progressive sequence alignment as a prerequisite to correct phylogenetic trees. J. Mol. Evol. **25:** 351–360.

FITCH, W. M., 1971 Towards defining the course of evolution: minimum change for a specified tree topology. Syst. Zool. **20:** 406–416.

GILLESPIE, J. H., 1984 The molecular clock may be an episodic clock. Proc. Natl. Acad. Sci. USA **81:** 8009–8013.

GILLESPIE, J. H., 1986a Natural selection and the molecular clock. Mol. Biol. Evol. **3:** 138–155.

GILLESPIE, J. H., 1986b Variability of evolutionary rates of DNA. Genetics **113:** 1077–1091.

GOODMAN, M., J. CZELUSNIAK, G. W. MOORE, A. E. ROMERO-HERRERA and G. MATSUDA, 1979 Fitting the gene lineage into its species lineage. A parsimony strategy illustrated by cladograms constructed from globin sequences. Syst. Zool. **28:** 132–163.

GOODMAN, M., A. E. ROMERO-HERRERA, H. DENE, J. CZELUSNIAK and R. E. TASHIAN, 1982 Amino acid sequence evidence on the phylogeny of primates and other eutherians, pp. 115–191 in *Macromolecular Sequences in Systematic and Evolutionary Biology*, edited by M. GOODMAN. PLENUM, NEW YORK.

HALL, I. L., R. K. CRAIG, M. R. EDBROOKE and P. N. CAMPBELL, 1982 Comparison of the nucleotide sequence of cloned human and guinea-pig pre-alpha-lactalbumin cDNA with that of chicken pre-lysozyne cDNA suggests evolution from a common ancestral gene. Nucleic Acids Res. **10:** 3503–3515.

HENDY, M. D., and D. PENNY, 1982 Branch and bound algorithms to determine minimal evolutionary trees. Math. Biosci. **59:** 277–290.

KIELAN-JAWOROWSKA, A., T. M. BROWN and J. A. LILLEGRAVEN, 1979 Eutheria, pp. 221–258 in *Mesozoic Mammals*, edited by J. A. LILLEGRAVEN, Z. KIELAN-JAWOROWSKA and W. A. CLEMENS. University of California Press, Berkeley.

KIMURA, M., 1968 Evolutionary rate at the molecular level. Nature **217:** 624–626.

KIMURA, M., 1983 *The Neutral Theory of Molecular Evolution.* Cambridge University Press, Cambridge.

KIMURA, M., 1987 Molecular evolutionary clock and the neutral theory. J. Mol. Evol. **26:** 24–33.

KOOP, B. F., and M. GOODMAN, 1988 Evolutionary and developmental aspects of two hemoglobin β-chain genes (ϵ^m and β^m) of opossum. Proc. Natl. Acad. Sci. USA **85:** 3893–3897.

KUMAGAI, I., E. TAMAKI, S. KAKINUMA and K. MIURA, 1987 Molecular cloning and sequencing of cDNA encoding goat pre α-lactalbumin. J. Biochem. **101:** 511–517.

LAKE, J. A., 1987 A rate-independent technique for analysis of nucleic acid sequences: evolutionary parsimony. Mol. Biol. Evol. **26:** 59–73.

LANGLEY, C. H., and W. M. FITCH, 1974 An examination of the constancy of the rate of molecular evolution. J. Mol. Evol. **3:** 161–177.

LI, W.-H., and M. TANIMURA, 1987 The molecular clock runs more slowly in man than in apes and monkeys. Nature **326:** 93–96.

LI, W.-H., C.-I. WU and C. C. Luo, 1985 A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. Mol. Biol. Evol. **2:** 150–174.

LI, W.-H., K. H. WOLFE, J. SOURDIS and P. M. SHARP, 1987 Reconstruction of phylogenetic trees and estimation of divergence times under nonconstant rates of evolution. Cold Spring Harb. Symp. Quant. Biol. **52:** 847–856.

METROPOLIS, N., A. W. ROSENBLUTH, M. N. ROSENBLUTH, A. H. TELLER and E. TELLER, 1953 Equation of state calculations by fast computing machines. J. Chem. Phys. **21:** 1087–1092.

NOVACEK, M. J., 1982 Information for molecular studies from anatomical and fossil evidence on higher eutherian phylogeny, pp. 3–41 in *Macromolecular Sequences in Systematics and Evolutionary Biology*, edited by M. GOODMAN. Plenum, New York.

OHTA, T., and M. KIMURA, 1971 On the constancy of the evolutionary rate of cistrons. J. Mol. Evol. **1:** 18–25.

PENNY, D., and M. D. HENDY, 1985 Estimating the reliability of evolutionary trees. Mol. Biol. Evol. **3:** 403–417.

PENNY, D., M. D. HENDY and I. M. HENDERSON, 1987 Reliability of evolutionary trees. Cold Spring Harbor Symp. Quant. Biol. **52:** 847–856.

SARICH, V. M., and A. C. WILSON, 1967 Immunological time scale for hominid evolution. Science **158:** 1200–1203.

TAKAHATA, N., 1987 On the over-dispersed molecular clock. Genetics **116:** 169–179.

WAINWRIGHT, B., and R. HOPE, 1985 Cloning and chromosomal location of the α- and β-globin genes from a marsupial. Proc. Natl. Acad. Sci. USA **82:** 8105–8108.

WU, C.-I., and W.-H. LI, 1985 Evidence for higher rates of nucleotide substitution in rodents than in man. Proc. Natl. Acad. Sci. USA **82:** 1741–1745.

WYSS, A. R., M. J. NOVACEK and M. C. MCKENNA, 1987 Amino acid sequence versus morphological data and the interordinal relationships of mammals. Mol. Biol. Evol. **4:** 99–116.

Communicating editor: A. H. D. BROWN