Proc. Natl. Acad. Sci. USA
Vol. 88, pp. 5974–5978, July 1991
Evolution

Synonymous nucleotide substitution rates in mammalian genes: Implications for the molecular clock and the relationship of mammalian orders

(mammalian phylogeny/primates/rodents/artiodactyls)

MICHAEL BULMER*†, KENNETH H. WOLFE‡§, AND PAUL M. SHARP‡

*Department of Statistics, Oxford University, Oxford OX1 3TG, United Kingdom; and ‡Department of Genetics, Trinity College, Dublin 2, Ireland

Communicated by John Maynard Smith, April 3, 1991

ABSTRACT **Synonymous substitution rates have been estimated for 58 genes compared among primates, artiodactyls, and rodents. Although silent sites might be expected to be neutral, there is substantial rate variation among genes within each lineage. Some of the rate variation is associated with G+C content: genes with intermediate G+C values have the highest rates. Nevertheless, considerable heterogeneity remains after correcting for G+C content. Synonymous substitution rates also vary among lineages, but the relative rates of genes are well conserved in different lineages. Certain genes have also been sequenced in a fourth order (lagomorph or carnivore), and these data have been used to investigate mammalian phylogeny. Data on lagomorphs are consistent with a star phylogeny, but there is evidence that carnivores and artiodactyls are sister groups. Genes sequenced in both rat and mouse suggest that the increased substitution rate in rodents has occurred since the rat/mouse divergence.**

Mammalian gene sequences have been the focus of many investigations of molecular evolutionary rates. There has been much discussion of two particularly controversial and interrelated topics—namely, (*i*) the constancy (or otherwise) of amino acid, and latterly nucleotide, substitution rates (i.e., over the possible extent and reliability of a "molecular clock") and (*ii*) the neutral theory of molecular evolution (1). In recent years, attention has focused on synonymous, or silent, substitutions, which are thought unlikely to be under natural selection. Kimura's (2) and subsequent (3) analyses suggesting that silent sites evolve not only at very high rates, but also at similar rates in genes encoding proteins that evolve at very different rates, were influential in promoting more widespread acceptance of the neutral theory. More recently, it has been suggested that synonymous rates are similar not only in different genes, but also in different lineages (4–6).

Both aspects of this universal silent molecular clock have been contested. Selective constraint on synonymous codon usage clearly leads to synonymous rate variation among genes in bacteria (7, 8) and perhaps also in *Drosophila* (9). There have been many reports that synonymous rates vary among lineages, particularly in mammals, where rodents are thought to have evolved faster (10–13), while the rate in higher primates is thought to have slowed down (14, 15). Several of these analyses, as well as other tests of the molecular clock using the ratio of variance to mean (16, 17), rely on assumptions about the relationships among orders of mammals. However, there are rather few points of agreement concerning the divergence of eutherian orders (18), so that a star-like radiation is normally assumed; if this assumption is incorrect, estimates of rates and their variability will be unreliable.

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

Here we exploit the burgeoning mammalian DNA sequence data base to examine the question of whether synonymous rates vary among genes and/or among lineages. Finding that both types of variation exist, we then ask whether there is any interaction; i.e., do quickly or slowly evolving genes have the same relative rates in different lineages? The data are gene sequences available for members of the three orders Primates, Rodentia, and Artiodactyla, some of which are also available for Lagomorpha and Carnivora. The methodology also allows for testing the star phylogeny assumption for these orders and the determination of the order of branching when it fails.

DATA AND METHODS

DNA sequences were taken from GenBank (release 60.0) and the literature, for 58 genes (see Table 1) sequenced from representatives of the three orders Primates, Rodentia, and Artiodactyla. Only one sequence was used from each order, with the cow given preference over other artiodactyls and the rat given preference over mouse (all primate sequences used were human). For 10 of these genes, data were also available from the Lagomorpha (rabbit), and for 5 of the genes data were available from the Carnivora (dog or seal).

To calculate synonymous substitution rates, the rate was first estimated separately for 4-fold and 2-fold degenerate sites in codons for each gene compared between each pair of species. Only third position sites in those codons in which the first two positions were identical in the two species were used. Let p be the proportion of these codons differing at the third position. Then, for 4-fold degenerate sites, to correct for multiple hits, calculate

$$b = 1 - (A_1A_2 + C_1C_2 + G_1G_2 + T_1T_2), \qquad [1]$$

where A_1 is the relative frequency of A (adenine) at the third position in these codons in the first species, and so on. The corrected substitution rate at 4-fold degenerate sites is calculated as

$$d_4 = -b \ln(1 - p/b). \qquad [2]$$

This is an extension of the formula of Tajima and Nei (19), allowing for possible differences in nucleotide frequencies between the two species. Rodent genes have a shift in G+C content (20), and so it is necessary to use this method to avoid any possible problem arising due to differences in base composition (21). Lewontin (22) has justified the use of Tajima and Nei's empirical correction formula when the

†To whom reprint requests should be addressed at: Department of Biological Sciences, Rutgers University, Piscataway, NJ 08855-1059.
§Present address: Department of Biology, Indiana University, Bloomington, IN 47405.

nucleotide frequencies are at the same equilibrium level in both species. Our extension of this method can be justified in a similar way when the nucleotide frequencies differ between species, if it can be assumed that they have reached an equilibrium in each species, but must be viewed with caution in case this condition is not true.

For 2-fold degenerate sites, only transitions were considered. Let p_R and p_Y be the proportions of codons differing at the third position among the purine-ending and pyrimidine-ending groups, respectively, and define

$$b_R = 1 - (A_1A_2 + G_1G_2)$$

$$b_Y = 1 - (C_1C_2 + T_1T_2), \quad [3]$$

where A_1 is the relative frequency of A at the third position in the purine group in the first species, and so on. The corrected substitution rate at the 2-fold degenerate sites is then calculated as

$$d_2 = -b \ln(1 - p/b), \quad [4]$$

where b and p are the averages of b_R and b_Y, and p_R and p_Y, respectively. Write d_{4ij} and d_{2ij} for the estimated substitution rates at 4-fold and 2-fold degenerate sites for a particular gene between species i and j. An overall estimate of the substitution rate can be obtained as

$$d_{ij} = 0.75(d_{4ij} + d_{2ij}). \quad [5]$$

The choice of equal weights in combining these estimates is somewhat arbitrary. The factor 0.75 is used to make the rate value comparable to other commonly used estimators, such as those of Li *et al.* (23) and Nei and Gojobori (24), which consider a 2-fold degenerate site as equivalent to one-third of a synonymous site.

Having estimated the substitution rates between different pairs of orders, we wish to estimate the rates within each lineage. We use the method of unweighted least squares (as in ref. 25). For three orders, the estimate of the rate within the first lineage is

$$a_1 = (d_{12} + d_{13} - d_{23})/2, \quad [6]$$

with similar expressions for the other two lineages. For four orders, if we assume a star phylogeny the corresponding estimate for the first lineage is

$$a_1 = [(d_{12} + d_{13} + d_{14})/3] - [(d_{23} + d_{24} + d_{34})/6]. \quad [7]$$

If we do not assume a star phylogeny, there are three possible unrooted trees [12,34], [13,24], and [14,23], with an internal branch between the two pairs of orders separated by the comma; number these trees 1, 2, and 3. For tree t, let b_{it} be the length of the branch from species i to the nearest node, and let c_t be the internal branch length. Then

$$c_1 = [(d_{13} + d_{14} + d_{23} + d_{24})/4] - [(d_{12} + d_{34})/2], \quad [8]$$

with similar expressions for c_2 and c_3, and

$$b_{it} = a_i - c_t, \quad [9]$$

where a_i is the estimate under a star phylogeny.

The estimators in Eqs. **6–9** are linear functions of the d_{ij} values defined in Eq. **5**, so that their covariance matrix can be calculated from that of the d_{ij} values; the latter is 0.75^2 times the sum of the covariance matrices of the d_{4ij} and the d_{2ij} values, since these two terms are independent. An approximate general formula for the covariances of the d_{xij} values (x = 2 or 4) is

$$\mathrm{Cov}(d_{xij}, d_{xkl}) = b\{[(1 - b)\exp 2\delta/b] + [(2b - 1)\exp \delta/b] - b\}/n, \quad [10]$$

(26–28). In this formula b is the quantity defined in Eq. **1** for d_4 or in Eqs. **3** and **4** for d_2: if different b values are used in the calculation of d_{xij} and d_{xkl}, it is here taken as their geometric mean; n is the sample size; and δ is the branch length in common between the pair of species i and j and the pair k and l. [If different sets of sites are used for the two pairs of species, then n is replaced by $n_{ij}n_{kl}/n_{ijkl}$, where n_{ij} (or n_{kl}) is the number of sites used for species i and j (or k and l), and n_{ijkl} is the number of sites available for comparison in all species.]

If the true internal branch length is γ, then the expected value of c_t is γ if t is the true tree and -0.5γ otherwise. The variability of the c_t values can be used to test for a star phylogeny ($\gamma = 0$) as follows. If **V** is the covariance matrix of c_1 and c_2, calculated as in the previous paragraph, then

$$(c_1\ c_2)\mathbf{V}^{-1}(c_1\ c_2)^T \quad [11]$$

is a χ^2 statistic with 2 degrees of freedom under a star phylogeny. The three c_t values represent only 2 degrees of freedom, since they sum to 0 identically; Eq. **11** has the same value whichever two of them are chosen to calculate it.

RESULTS

Synonymous substitution rates were calculated for the primate, artiodactyl, and rodent lineages for 58 genes, ignoring information from other orders (see Table 1). Table 1 summarizes the average rates at 4-fold and 2-fold degenerate

Table 1. Synonymous substitution rates in 58 mammalian genes

Degeneracy	Primate	Artiodactyl	Rodent
	Average rates		
4-fold	0.145 (0.010)	0.189 (0.011)	0.355 (0.014)
2-fold	0.100 (0.013)	0.126 (0.013)	0.220 (0.016)
4-fold − 2-fold	0.045 (0.016)	0.063 (0.017)	0.136 (0.021)
4-fold + 2-fold	0.244 (0.016)	0.315 (0.017)	0.575 (0.021)
K_S*	0.183 (0.012)	0.236 (0.013)	0.431 (0.016)
	χ^2 values for heterogeneity among genes†		
	116.5	95.3	141.6
	198.0‡		
Corrected for G+C	162.6‡		116.4

The following genes were used (gene symbols are from ref. 29): LALBA, ALB, FGFB, ALPL [R], CCK, CYP17, corticotropin-releasing factor, ELA2 [R], DBI, PENKA, PENKB, UDP-NAG galactosyltransferase, GH [R], GCG, CGA [R], G proteins ($G_{o\alpha}$, $G_{i1\alpha}$, $G_{s\alpha}$), IL2, IL2R, nicotinic acetylcholine receptors (α and γ), M1 [R] and M2 muscarinic acetylcholine receptors, matrix Gla protein, ATPSB, ATP1A1, ATP1B [C], neuroleukin/phosphohexose isomerase, RHO, OT, PTH, protein disulfide isomerase [R], PRKAR1 [R], PRL [R], RLN1, SPARC/ostonectin, DNTT, THBD, TF, TSHB [R], atrial natriuretic factor [CLR], HBB [LR], GPX1 [LR], IL1A [L], IL1B [L], LDLR [L], PRKCA [LR], PRKCB-II [L], PRKCG [L], ACP2 [L], LHB [C], myoglobin [C], PLA2 [C], POMC [R], PLP [R], urokinase-type plasminogen activator, vasopressin/neurophysin II. Symbols in brackets indicate that the sequence is available for carnivores [C], lagomorphs [L], or both mouse and rat [R]. Full details and sequences are available on request. Average rates are the corrected number of substitutions per site in each lineage. Values in parentheses are standard errors; these do not take into account differences between genes and are conditional on the sample of genes used.

*Synonymous substitution rate = 0.75 × (4-fold + 2-fold) (see Eq. 5).

†With 56 or 54 (values corrected for G+C) degrees of freedom; all six values are significant at the 0.1% level.

‡Pooled over primates and artiodactyls.

sites. The 2-fold rate estimates the frequency of transitions, while the difference between the 4-fold and 2-fold rates estimates the frequency of transversions. Transitions are found to be about twice as frequent as transversions, which is in line with other findings (23). The substitution rate is slightly higher in the artiodactyl lineage than in the primate lineage but is increased by a factor of ≈2 in the rodent lineage, again in line with earlier findings (10). Further analysis will be based on the sum of the 4-fold and 2-fold substitution rates, multiplied by a factor of 0.75; this quantity is termed the synonymous substitution rate, K_S.

Heterogeneity Between Genes. We first investigate whether there are differences in substitution rates among genes within a lineage, and if so whether these differences are consistent in different lineages. To remove the overall effect of the lineage, the substitution rates within a lineage were divided by the average value for that lineage shown in Table 1; these relative substitution rates have an average value of unity within each lineage.

One gene (PLP) in the data set is situated on the X chromosome, and it has been suggested that X chromosome-linked genes have a mutation rate about two-thirds that in autosomal genes because most germ-line mutations occur in males (30). The relative substitution rate of the PLP gene is substantially less than 0.67 in all three lineages, with the average value over the three lineages being 0.22 ± 0.06. This gene has been excluded from the remainder of the analysis in order not to confound differences among autosomal genes with a possible difference between autosomal and sex-linked genes.

Table 1 also shows a statistical analysis of the heterogeneity among the remaining genes. The first line shows χ^2 tests for heterogeneity of the substitution rates among the 58 genes within each of the three lineages; they are all highly significant, showing that some genes evolve faster than others, even at synonymous sites. This conclusion is strengthened by the PLP gene, which clearly evolves slowly even after making allowance for its sex linkage.

Table 2 tests whether these differences are consistent between lineages by analyzing the differences of the relative substitution rates of each gene between a pair of lineages. There is no evidence of a difference between homologous genes in primates and artiodactyls; genes behave in the same way in both lineages. However, there is evidence of a difference in behavior between both these orders and rodents; some genes behave in a different way in rodents than they do in primates and artiodactyls.

To analyze these differences further, the relative substitution rates for primates and artiodactyls were averaged for each gene to reduce sampling noise, since there is no evidence of a difference in behavior between these orders. Results for the combined primate/artiodactyl rates are shown in Tables 1 and 2.

Effect of G+C Content. It has been suggested that mutation rates (and hence substitution rates) vary with G+C content, such that genes with intermediate G+C values have the highest rates (31). To determine whether this accounts for some or all of the variability in substitution rates, a quadratic regression on G+C content was fitted to the data. The fitted equations are as follows:

$$\text{for primates/artiodactyls, } y = 1.11 - 3.3\ (x - 0.72)^2$$

$$\text{for rodents, } y = 1.11 - 5.6\ (x - 0.55)^2,$$

where y is the relative substitution rate and x is the G+C content in the third position (as a proportion). In both cases, there is a regression of the predicted form, with the quadratic term being significant at the 5% level, although the highest substitution rate occurs at a G+C content somewhat higher than 0.50.

Heterogeneity After Correction for G+C Content. The regression of G+C content accounts for some of the variability in substitution rates between genes within a lineage and may also account for some of the differences between the primate/artiodactyl and the rodent lineages, since changes in G+C content have occurred in some rodent genes (20). In Tables 1 and 2, the analysis for heterogeneity is repeated after correcting the data for G+C content from the estimated regression. Substantial variability remains both within and between lineages. Fig. 1 shows the standardized residuals for the rodent lineage plotted against the standardized residuals for the primate/artiodactyl lineages (see legend for details of the calculation of standardized residuals). There is a consistent relationship between them with two exceptions, which account for the significant interaction term in Table 2. Thus, for 56 of the 58 genes examined, the relative synonymous substitution rates seen in genes (after compensation for lineage-specific and G+C content effects) are consistent in the three mammalian orders. The two exceptional genes are ATP1A1 (encoding Na^+/K^+-ATPase α) and PRKCA (protein kinase C α), which have relatively fast rates in primates/artiodactyls and relatively slow rates in rodents. Three genes (two encoding G proteins $G_{o\alpha}$ and $G_{s\alpha}$, and ACP2 encoding protein phosphatase 2Aα) have particularly slow rates in all orders and one (ALB, encoding serum albumin) has a fast rate. None of the slow genes are X chromosome-linked (29). PRKCA has been sequenced in both rat and mouse, and it is therefore of interest to consider its behavior in the rodent lineage before and after the rat/mouse divergence. The

Table 2. χ^2 values for heterogeneity of relative substitution rates among lineages

Lineage	Artiodactyl	Rodent
Primate	49.4	75.9*
Artiodactyl		98.2***
Primate/artiodactyl		106.7***
Primate/artiodactyl (corrected for G+C)		77.9*

With 56 or 54 (value corrected for G+C) degrees of freedom. Significance levels: *, $P < 0.05$; ***, $P < 0.001$.

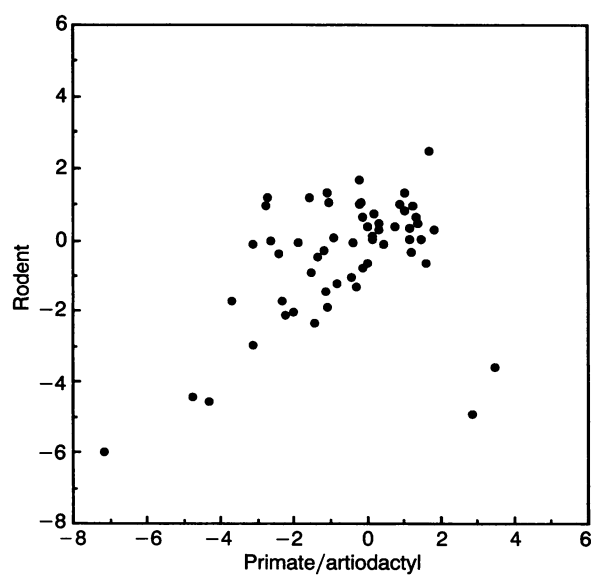

FIG. 1. Correlation between the standardized residuals in rodents and in primates/artiodactyls for 57 genes. The standardized residual is $(d - D)/\text{SE}$, where d is the synonymous substitution rate, D is its expected value given the G+C content of the gene, and SE is the standard error of d.

Table 3. Phylogeny and branch lengths from four order comparisons

	Primate/artiodactyl/ rodent/lagomorph	Primate/artiodactyl/ rodent/carnivore	Primate/artiodactyl/ rat/mouse
Number of genes	10	5	15
χ^2 for star phylogeny†	0.38	7.25*	312.9***
c_1	0.000 ± 0.016	−0.039 ± 0.032	+0.311 ± 0.028
c_2	+0.008 ± 0.016	+0.076 ± 0.032	−0.145 ± 0.024
c_3	−0.008 ± 0.016	−0.037 ± 0.030	−0.166 ± 0.024
Inferred tree	Star	[13,24]	[12,34]
Internal branch length	0.000 ± 0.016	0.076 ± 0.032	0.311 ± 0.028
b_1 (primate)	0.192 ± 0.016	0.173 ± 0.033	0.217 ± 0.024
b_2 (artiodactyl)	0.246 ± 0.019	0.166 ± 0.031	0.256 ± 0.025
b_3 (rodent/rat)	0.332 ± 0.023	0.412 ± 0.046	0.126 ± 0.019
b_4 (lagomorph/carnivore/mouse)	0.244 ± 0.020	0.224 ± 0.035	0.101 ± 0.020

The c values are defined in Eq. **8**; the b values are defined in Eq. **9**. Significance levels: *, $P < 0.05$; ***, $P < 0.001$.
†Defined in Eq. **11**.

relative substitution rate was 0.42 ± 0.14 before the rat/mouse split, and 1.34 ± 0.28 and 0.89 ± 0.32, respectively, in rat and mouse after their divergence. Thus, it seems that the anomalous behavior of this gene occurred mostly in the common ancestor of rat and mouse.

Four Taxon Comparisons: Phylogenetic Inference. Of the genes analyzed above, 10 have been sequenced also in lagomorphs, and 5 have been sequenced in carnivores (see Table 1). Furthermore, 15 genes have been sequenced in both rat and mouse, in addition to primates and artiodactyls. We shall here consider these data from four orders (or in the last case four species) with primary interest in making phylogenetic inferences, although some additional information about the variability of substitution rates will also be obtained.

Table 3 shows the results of a statistical analysis performed after averaging the d_{ij} values with weights proportional to the sizes of the genes. There is no evidence of a departure from a star phylogeny for the primate/rodent/artiodactyl/lagomorph grouping. In contrast, there is evidence of departure from a star phylogeny in the primate/artiodactyl/rodent/carnivore analysis: primates are associated with rodents, and artiodactyls are associated with carnivores, although the χ^2 test is only significant at the 5% level. As expected, there is strong evidence for an association between rat and mouse.

Heterogeneity in Lagomorphs. There is clear evidence of heterogeneity among genes for the lagomorph rates (χ^2 = 35.7 with 9 degrees of freedom; $P < 0.001$). This heterogeneity follows more closely the primate/artiodactyl pattern than the rodent pattern. For example, two of the genes identified above as being at the extremes of the distribution in Fig. 1 can also be examined in lagomorphs: for PRKCA the relative substitution rates are 1.42 ± 0.15 for primate/artiodactyl, 1.20 ± 0.19 for lagomorph, and 0.75 ± 0.13 for rodent; for ACP2 the corresponding figures are 0.37 ± 0.07, 0.37 ± 0.10, and 0.76 ± 0.13. The χ^2 value for the difference between primate/artiodactyl and lagomorph rates is 6.7 with 9 degrees of freedom (not significant). It is tentatively concluded that primates, artiodactyls, and lagomorphs behave in the same way, with the same sets of "fast" and "slow" genes in all three orders, with rodents being the odd order in having a small number of exceptions to this rule. There are insufficient data on carnivores to allow any conclusions to be drawn.

DISCUSSION

The enhanced codon usage bias observed in highly expressed genes in unicellular organisms suggests that synonymous codons are under translational selective constraint (7, 8, 32). However, one would not expect these constraints to be effective in mammals (33) because of their small effective population sizes (34). Thus, differences in mammalian synonymous substitution rates may primarily reflect differences in mutation rates (31). Therefore, it is rather surprising that there is substantial heterogeneity among genes within the same lineage (Table 1). We have previously (31, 35) proposed a model that predicts higher mutation rates in genes with intermediate G+C content than in G+C- or A+T-rich genes, and we have thus suggested an explanation of the observed relationship between synonymous substitution rate and G+C content in rodents. This effect also occurs in the present data set but explains only part of the heterogeneity among genes (Table 1).

It is difficult to determine whether the residual heterogeneity (after correction for G+C content) reflects differences in mutation rate unrelated to G+C content or some unidentified selection pressure. The only clue to the origin of this heterogeneity is that it appears to be nearly independent of lineage, so that fast genes go fast and slow genes go slow in all lineages (Table 2 and Fig. 1). This is true without exception for primates and artiodactyls and for the smaller number of lagomorph genes available, and it is true with only 2 exceptions (of 58 genes) in rodents (Fig. 1). Our prejudice is to believe that this reflects differences among genes in relative mutation rates that are conserved between lineages, but it is also consistent with some form of selective constraint. However, it does not appear to be consistent with Gillespie's episodic clock model (36).

The best phylogeny consistent with the data presented in Tables 1 and 3 is shown in Fig. 2. This tree has been rooted

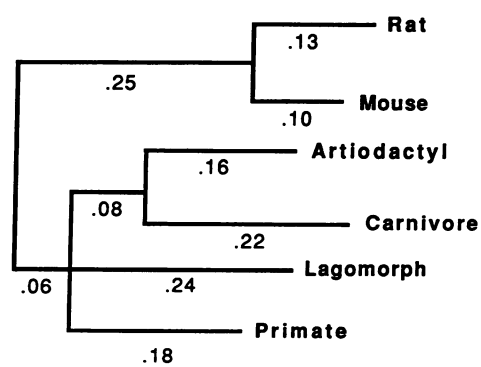

FIG. 2. Relationships among mammalian taxa inferred from the data in Tables 1 and 3. Values are the numbers of synonymous substitutions per site along each branch, estimated by combination of the results in Tables 1 and 3. Horizontal branch lengths are drawn to scale, but vertical separation is for clarity only. The tree is rooted according to ref. 37.

Table 4. Synonymous substitution rates in mammalian lineages

Lineage	K_S*	Time†	Rate‡
Primate	0.18	80	2.2
Lagomorph	0.24	80	3.0
Carnivore	0.30	80	3.8
Artiodactyl	0.24	80	3.0
Rodent			
Before rat/mouse split	0.25	88	2.8
After rat/mouse split	0.12	12	10.0

*Number of synonymous substitutions per site.
†Million years (based an fossil records).
‡Number of synonymous substitutions per site per 10^9 years.

with rodents as an outgroup to the four other mammalian orders considered, and the distance of 0.31 between the rat/mouse divergence and the primate/lagomorph/artiodactyl divergence has been apportioned to the two sides of the root following the recent results of Li *et al.* (37), who used birds as an outgroup to root the tree. Other analyses, albeit based on much smaller data sets (5, 6, 11), have also suggested that rodents diverged first. Our conclusions about mammalian phylogeny are similar to those of Li *et al.*; the data sets used are similar, although the methods of analysis are quite different.

There are two main features in the tree. The first concerns the phylogenetic affinities of lagomorphs. Many taxonomists, going back to Linnaeus, have grouped rodents and lagomorphs as sister groups in the taxon Glires (38, 39). This grouping was not supported by immunological studies (40), or by a recent analysis based on globin gene sequences (6). The present study clearly fails to support the Glires concept; lagomorphs do not group with rodents (see Fig. 2 and the first column of Table 3). The study of globin genes suggested that artiodactyls are an outgroup to a primate–lagomorph clade (6); this too is contradicted in Fig. 2, which is based on a larger data set.

The second feature is that there is evidence that artiodactyls and carnivores are sister groups (see Fig. 2 and the second column of Table 3). This grouping (the Ferungulata) was suggested in earlier classifications (41, 42) but has received little recent support (38).

To convert the rates in Fig. 2 into rates per unit time, we must place a time scale on the tree. We follow others (16, 37) in assuming that the main mammalian radiation (the primate/lagomorph/artiodactyl divergence in Fig. 2) occurred about 80 million years ago (Mya). Li *et al.* (37) have suggested that the rodent divergence occurred around 100 Mya. Recent paleontological evidence places the rat/mouse divergence at 12 Mya (43). Using these approximate dates, we have calculated the synonymous substitution rate (per 10^9 years) for each lineage (Table 4). These estimates strongly suggest that the increased rate in the rodent lineage has occurred since the rat/mouse divergence. Apart from this, the rates are reasonably uniform, with a small decrease in the primate lineage; the carnivore rate estimate is somewhat higher but is based on only five genes and so may be unreliable in view of the heterogeneity among genes.

This is a publication from the Irish National Centre for Bioinformatics.

1. Kimura, M. (1983) *The Neutral Theory of Molecular Evolution* (Cambridge Univ. Press, Cambridge, U.K.).
2. Kimura, M. (1977) *Nature (London)* **267,** 275–276.
3. Miyata, T., Yasunaga, T. & Nishida, T. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 7328–7332.
4. Wilson, A. C., Ochman, H. & Prager, E. M. (1987) *Trends Genet.* **3,** 241–247.
5. Easteal, S. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 7622–7626.
6. Easteal, S. (1990) *Genetics* **124,** 165–173.
7. Ikemura, T. (1985) *Mol. Biol. Evol.* **2,** 13–34.
8. Sharp, P. M. & Li, W.-H. (1986) *J. Mol. Evol.* **24,** 28–38.
9. Sharp, P. M. & Li, W.-H. (1989) *J. Mol. Evol.* **28,** 398–402.
10. Wu, C.-I. & Li, W.-H. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 1741–1745.
11. Kikuno, R., Hayashida, H. & Miyata, T. (1985) *Proc. Jpn. Acad.* **61B,** 153–156.
12. Britten, R. (1986) *Science* **231,** 1393–1398.
13. Li, W.-H., Tanimura, M. & Sharp, P. M. (1987) *J. Mol. Evol.* **25,** 330–342.
14. Li, W.-H. & Tanimura, M. (1987) *Nature (London)* **326,** 93–96.
15. Koop, B. F., Goodman, M., Xu, P., Chan, K. & Slightom, J. L. (1986) *Nature (London)* **319,** 234–237.
16. Kimura, M. (1987) *J. Mol. Evol.* **26,** 24–33.
17. Gillespie, J. H. (1986) *Genetics* **113,** 1077–1091.
18. Benton, M. J. (1988) *Trends Ecol. Evol.* **3,** 40–45.
19. Tajima, F. & Nei, M. (1984) *Mol. Biol. Evol.* **1,** 269–285.
20. Mouchiroud, D., Gautier, C. & Bernardi, G. (1988) *J. Mol. Evol.* **27,** 311–320.
21. Saccone, C., Pesole, G. & Preparata, G. (1989) *J. Mol. Evol.* **29,** 407–411.
22. Lewontin, R. C. (1989) *Mol. Biol. Evol.* **6,** 15–32.
23. Li, W.-H., Wu, C.-I. & Luo, C.-C. (1985) *Mol. Biol. Evol.* **2,** 150–174.
24. Nei, M. & Gojobori, T. (1986) *Mol. Biol. Evol.* **3,** 418–426.
25. Li, W.-H. (1989) *Mol. Biol. Evol.* **6,** 424–435.
26. Nei, M., Stephens, J. C. & Saitou, N. (1985) *Mol. Biol. Evol.* **2,** 66–85.
27. Bulmer, M. (1989) *Genetics* **123,** 615–619.
28. Nei, M. & Jin, L. (1989) *Mol. Biol. Evol.* **6,** 290–300.
29. Tenth International Workshop on Human Gene Mapping (1989) *Cytogenet. Cell Genet.* **51** (1–4).
30. Miyata, T., Hayashida, H., Kuma, K., Mitsuyasu, K. & Yasunaga, T. (1987) *Cold Spring Harbor Symp. Quant. Biol.* **52,** 863–867.
31. Wolfe, K. H., Sharp, P. M. & Li, W.-H. (1989) *Nature (London)* **337,** 283–285.
32. Bulmer, M. (1988) *J. Evol. Biol.* **1,** 15–26.
33. Shields, D. C., Sharp, P. M., Higgins, D. G. & Wright, F. G. (1988) *Mol. Biol. Evol.* **5,** 704–716.
34. Nei, M. & Graur, D. (1985) *Evol. Biol.* **17,** 73–118.
35. Wolfe, K. H. (1991) *J. Theor. Biol.*, in press.
36. Gillespie, J. H. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 8009–8113.
37. Li, W.-H., Gouy, M., Sharp, P. M., O'hUigin, C. & Yang, Y.-W. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 6703–6707.
38. Novacek, M. J., Wyss, A. R. & McKenna, M. C. (1988) in *The Phylogeny and Classification of the Tetrapods*, ed. Benton, M. J. (Oxford Univ. Press, Oxford, U.K.), pp. 31–71.
39. Benton, M. J. (1990) *J. Mol. Evol.* **30,** 409–424.
40. Sarich, V. M. (1985) in *Evolutionary Relationships Among Rodents*, eds. Luckett, W. P. & Hartenberger, J. L. (Plenum, New York), pp. 423–452.
41. Gregory, W. K. (1910) *Bull. Am. Mus. Nat. Hist.* **27,** 1–524.
42. Simpson, G. G. (1945) *Bull. Am. Mus. Nat. Hist.* **85,** 1–350.
43. Jaeger, J.-J., Tong, H. & Denys, C. (1986) *C.R. Acad. Sci. Paris* **302,** 917–922.