J Mol Evol (1995) 40:107–113

Journal of Molecular Evolution

Frequencies of Synonymous Substitutions in Mammals Are Gene-Specific and Correlated with Frequencies of Nonsynonymous Substitutions

Dominique Mouchiroud,[1] Christian Gautier,[1] Giorgio Bernardi[2]

[1] Laboratoire de Biométrie, Génétique et Biologie des Populations, U.R.A. 243, Université Claude Bernard, 69600 Villeurbanne, France
[2] Laboratoire de Génétique Moléculaire, Institut Jacques Monod, 2 Place Jussieu, 75005 Paris, France

Received: 5 October 1994

Abstract. The frequencies of synonymous substitutions of mammalian genes cover a much wider range than previously thought. We report here that the different frequencies found in homologous genes from a given mammalian pair are correlated with those in the same homologous genes from a different mammalian pair. This indicates that the frequencies of synonymous substitutions are gene-specific (as are the frequencies of nonsynonymous substitutions), or, in other words, that "fast" and "slow" genes in one mammal are fast and slow, respectively, in any other one. Moreover, the frequencies of synonymous substitutions are correlated with the frequencies of nonsynonymous substitution in the same genes.

Key words: DNA repair — DNA replication — DNA transcription — Isochores

Introduction

Very recent results (Bernardi et al. 1993) indicate that the frequencies of synonymous substitutions of mammalian genes are remarkably different for different genes. Although the range of variation (namely, the ratio of the highest and lowest values) for K_s, the substitution rate per synonymous site, was reported to be about fivefold by Li and Graur (1991), synonymous substitution frequencies cover in fact at least an up-to-20-fold range. The narrowest ranges (as low as threefold) were found (Bernardi et al. 1993) in comparisons comprising small data sets (like man/sheep, man/pig), the widest one in the mouse/rat comparison, which concerns a very large sample (in this case, an almost identical 19-fold range was reported for K_s values of 363 genes; Wolfe and Sharp 1993). The intermediate values (around tenfold) concern comparisons of both large (man/rat, man/calf) and small (man/primates) samples. These results provide a qualitative indication that the estimated range of synonymous substitution frequencies is dependent upon the gene sample studied and strongly suggest that the low and intermediate values mentioned above are underestimates, the actual ranges being probably wider than 20-fold.

The possibility that different frequencies of synonymous substitutions in different genes depend upon the isochore families in which the genes are located (Wolfe et al. 1989) was recently ruled out (Bernardi et al. 1993; see also Discussion). The results mentioned above raise, therefore, the question of whether the different frequencies are distributed at random over genes or are gene-specific. Here, we investigated this problem by comparing synonymous substitution frequencies of homologous genes from a given pair of mammals with those of the same homologous genes from another pair. We found that the synonymous substitution frequencies of homologous genes from different mammals are highly correlated with each other, indicating that the synonymous substitution frequency is a gene-specific property (as is the case for the nonsynonymous substitution frequency). In other words, high or low frequencies of synonymous

Correspondence to: D. Mouchiroud

This paper was presented at the Workshop "Open Problems in Molecular Evolution" organized by the International Society of Molecular Evolution (ISME) in Guanacaste (Costa Rica), 18–23 April 1994

substitutions are specific for any given gene, independent of the mammalian species under consideration. While a gene specificity of synonymous substitutions was suggested by Bulmer et al. (1991) on the basis of data that will be critically discussed in the following section, the present work unequivocally demonstrates this point. Moreover, we report that the frequency of synonymous substitutions is significantly correlated with the frequency of nonsynonymous substitutions in the same genes.

Materials and Methods

Sequences from GenBank (release 82) were managed using the HOVERGEN software (Duret et al. 1994), which is based upon the ACNUC system (Gouy et al. 1985) and provides a very efficient way to select homologous genes from vertebrates, to test if they are orthologous, and to predict the isochore location of genes on the basis of the data of Mouchiroud et al. (1991). Only homologous sequences having more than 150 synonymous codons were taken into account in this work, in order to avoid a too-large variance due to sampling. Mnemonics are available upon request.

The silent, or synonymous, difference frequency (SDF; Mouchiroud and Gautier 1990) was used to quantify dissimilarity between homologous coding sequences. SDF is the percentage of synonymous codons that are different in third positions of aligned sequences. In contrast to K_s, the synonymous substitution rate, SDF, simply provides a measure of the dissimilarity that exists between sequences. By definition, SDF does not rely on any hypothesis concerning the nature of the substitution process, whereas this obviously is not the case for K_s (which was calculated here according to Li et al. 1985). However, when relatively small ranges are considered (for example, for K_s values corresponding to less than 50% substitutions in the rat/mouse comparison), K_s and SDF are correlated nearly linearly (Bernardi et al. 1993). When the substitution process is stationary, the SDF results may explain process characteristics (for example, divergence rate). The relationships between SDFs for homologous genes or SDF and NSDF for the same genes (see below), especially the slopes, may, however, depend upon the statistical properties (number and/or size) of the sequence set under consideration.

NSDF, the nonsynonymous difference frequency, is defined, similarly to SDF, as the percentage difference in amino acids.

Figure 1 displays a plot of SDF values for 162 pairs of homologous genes from calf and man against SDF values for the same homologous genes from calf and murids (rat or mouse; SDF values for calf/mouse and calf/rat are practically identical). These data expand those of Bulmer et al. (1991), which concerned 58 mammalian genes from the same orders, and provide a first indication that frequencies of synonymous substitutions in mammals are gene-specific. A problem of Fig. 1, as well as of any three-order comparison, is, however, the existence of a common branch in the phylogenetic tree leading from the common ancestor to the species under consideration. Indeed, this may artefactually contribute to the dissimilarity measurements concerning two pairs of species. There are two possible ways to eliminate the influence of this common branch.

The first one requires a modeling of the substitution process, which involves hypotheses on the topology of the tree and on the characteristics of the substitution process. Bulmer et al. (1991) did a very careful and mathematically relevant analysis of these hypotheses. However, the existence of the minor shift, exhibited by the genes of murids, which were included in their analysis, were the source of a problem concerning the hypothesis of the homogeneity of the process. Indeed, Mouchiroud and Gautier (1988) showed that 44% of murid genes exhibit a significant codon usage change when compared with other mammalian lineages displaying the "general compositional pattern" (Bernardi et al. 1988; Mouchiroud et al. 1988; Mouchiroud and Gautier 1990). This difficulty led Bulmer et al. (1991) to propose a correction, which is, however, not valid when the equilibrium of the Markov process is not reached in all species.

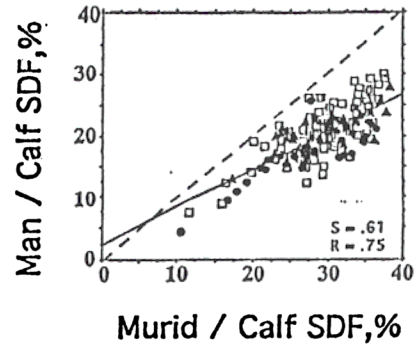


Fig. 1. SDF values, the synonymous difference frequencies, for 162 pairs of homologous genes from man/calf are plotted against SDF values for the same genes from rat (or mouse)/calf. The orthogonal regression line (D'Onofrio et al. 1991), through the points is shown. *Solid circles* correspond to human GC_3 values higher than 80%, *solid triangles* to GC_3 values comprised between 80% and 50%, *squares* to GC_3 values lower than 50% (GC_3 is the GC level at third codon positions.) The slope (S) and correlation coefficient (R) are indicated here, as in all other figures.

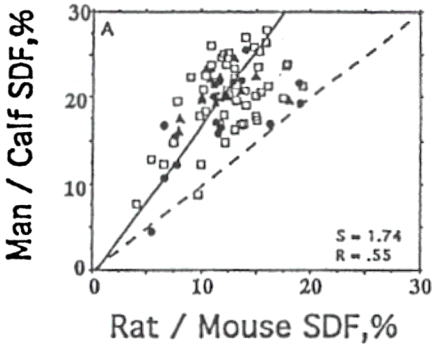
An alternative solution of the common branch problem was made possible through the recent increase of available homologous sequences. Indeed, this allowed us to use a four-species strategy that permits a direct comparison of gene pair dissimilarity (Fig. 2) and that is independent of the nonhomogeneity of the substitution process, even if equilibrium is not reached in all the species under consideration.

Results

The plot of synonymous difference frequencies (SDF) values for 85 homologous genes from the man/calf pair *vs* the mouse/rat pair (Fig. 2A) compares two species with two different ones. The correlation coefficient (R = 0.55) is highly significant ($P < 10^{-4}$). As expected, very similar results were obtained when plotting K_s values (Fig. 2B).

Several features of Fig. 2A are of interest. First of all, the SDF ranges are about sixfold in the human/bovine comparison and about fourfold in the mouse/rat comparison. In all likelihood, these SDF ranges are underestimates, due to the particular, relatively small, gene samples used in the comparison under consideration (See Introduction.)

Second, as expected from previous work (Bernardi et al. 1993), there is no correlation between synonymous substitution frequencies and GC levels in third codon positions, like the compositional differences that exist among the isochores and the genes exhibiting high or low GC levels, respectively, of man and murids (Salinas et al. 1986; Zerial et al. 1986; Bernardi et al. 1988; Mouchiroud et al. 1988; Mouchiroud and Gautier 1988,



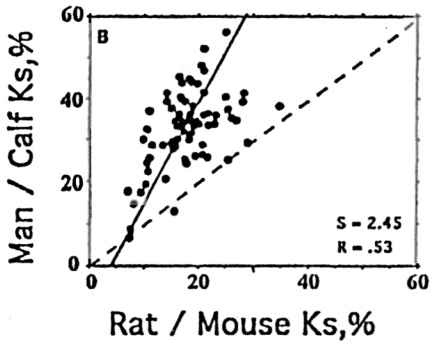


Fig. 2. SDF (A) and K_s (B) values for 85 homologous genes from calf/man are plotted against SDF or K_s values, respectively, for the same genes from rat/mouse. In the top figure, *solid circles* correspond to human GC_3 values higher than 80%, *solid triangles* to values between 80% and 50%, *squares* to values lower than 50%. The orthogonal regression lines (D'Onofrio et al. 1991) are shown in both figures.

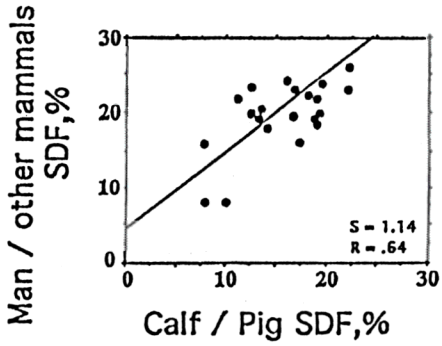


Fig. 3. SDF values for homologous genes from the calf/pig pair are plotted against SDF values of man/rabbit or man/carnivore pairs. The latter include dog, cat, and mink.

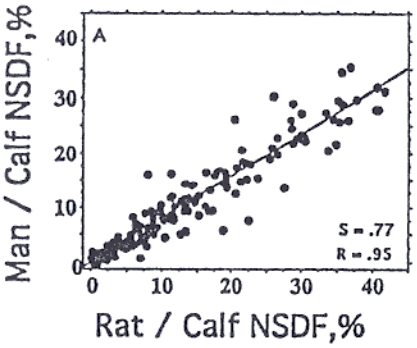
1990) Since a correlation exists between GC levels in third codon positions and GC levels of the isochores (see Bernardi 1989; 1993a,b, for reviews on isochores) containing the corresponding genes (Bernardi et al. 1985; Aota and Ikemura 1986; Ikemura and Aota 1988; Aïssani et al. 1991), there is also no correlation between synonymous substitution frequencies and the GC levels of isochores containing the genes under consideration (see Bernardi et al. 1993).

Third, the fact that the orthogonal regression line has a slope higher than unity is due to the larger divergence time between calf and man compared to that between rat and mouse. The higher synonymous substitution rate of the latter pair (Wu and Li 1984; Li and Wu 1987) obviously is not large enough to compensate for the larger divergence time of the former pair (80 Myrs *vs* 12 Myrs; see Bulmer et al. 1991).

As no common branch exists in the comparison of the two pairs of species under consideration, the correlations observed in Fig. 2 directly demonstrate the existence of a gene-specific component of SDF variability. The square of the correlation coefficient, 0.30, indicates that 30% of the variability of SDF between man and calf can be predicted when SDF values between mouse and rat are known. This gene-specific component is, however, underestimated, probably because the substitution process in the murid lineage is peculiar in showing both acceleration and nonhomogeneity. Indeed, if only the 63 largest homologous genes, with more than 300 silent sites, are considered, the percentage of predicted variance increases to 37% (not shown), an increase which is difficult to explain by sampling variability only. Moreover, if SDF values between calf and pig are compared with SDF values between man/rabbit and carnivores (dog, cat, or mink), the correlation coefficient is even higher ($R = 0.64$; $R^2 = 0.41$), in spite of the small sample size (Fig. 3). This result should, however, be confirmed using a larger gene sample.

Figure 4 displays diagrams in which nonsynonymous difference frequencies (NSDF) for homologous gene pairs from man/calf are plotted against NSDF values of rat/calf (Fig. 4A) or rat/mouse (Fig. 4B). The correlation coefficients were very high (0.95 and 0.89, respectively) and the slopes were equal to 0.77 and 1.81, respectively, in the two cases. Interestingly, the slopes are close to those exhibited by SDF plots (Figs. 1 and 2A), indicating a parallelism between the synonymous and nonsynonymous substitution processes.

Figure 5 shows that significant ($P = 10^{-3}$) correlation coefficients (0.57 and 0.46) are found when plotting SDF values for homologous genes from man/calf (Fig. 5A) and rat/mouse (Fig. 5B), respectively, against log NSDF values for the same genes of Fig. 3. The lower slope of the latter plot (3.15 *vs* 4.82) is due to the lower variability of both SDF and NSDF values in the rat/mouse case. Log NSDF was used instead of NSDF because the relationship between synonymous and nonsynonymous substitutions is not linear. If NSDF values are used instead of log NSDF, the correlation coefficients slightly decrease (to 0.46 and 0.36, respectively), but remain significant ($P = 10^{-3}$ and $P = 7.10^{-3}$, respectively). The plots of Fig. 5A and 5B concern the same 85 homologous genes and are, therefore, strictly comparable. Incidentally, a plot of K_s *vs* log K_a for the same genes of Fig. 2 displayed correlation coefficients of 0.61 and 0.51, respectively (not



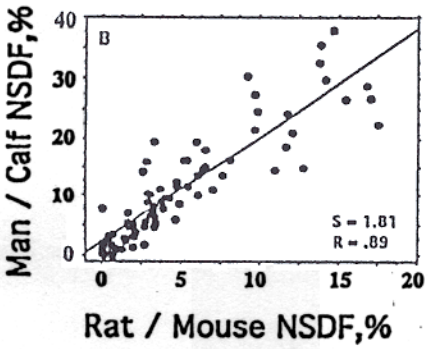
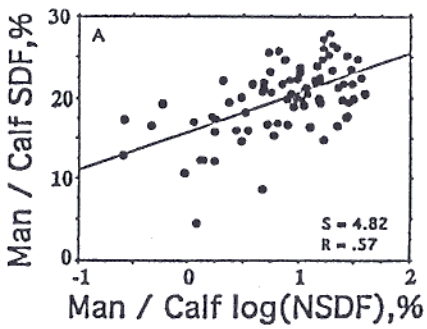


Fig. 4. NSDF values, the nonsynonymous difference frequencies, for homologous genes from man/calf are plotted against NSDF values for the same genes from the rat/calf (A) and rat/mouse (B) comparisons of Fig. 2.



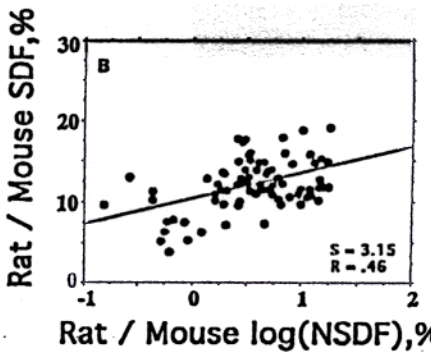


Fig. 5. SDF values for homologous genes from calf/man (A) and rat/mouse (B) are plotted against log NSDF values for the same genes.

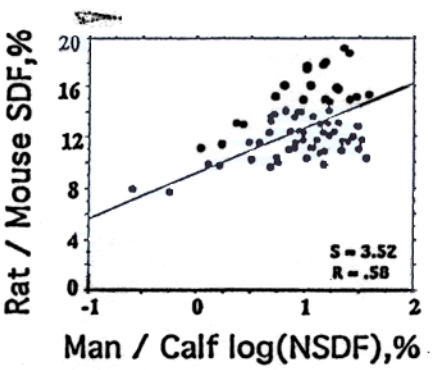


Fig. 6. SDF values for homologous genes from the rat/mouse pair are plotted against log NSDF values for the same genes from the calf/man pair.

shown). If different, larger sets of genes are used, correlation coefficients of 0.55 and 0.44 are found for 159 genes from man/calf and for 303 genes from rat/mouse, respectively.

Figure 6 is similar to Fig. 5, but SDF and NSDF data from different pairs (rat/mouse and calf/man) were used in the comparisons. The significant ($P < 10^{-4}$) correlation found ($R = 0.58$) is identical to that of Fig. 5A, thus demonstrating that it is the gene-specific component of SDF that correlates with NSDF and not a property associated with the lineages under consideration.

Discussion

Synonymous Substitution Frequencies of Homologous Mammalian Genes are Gene-specific

The main point made in the present work is the unequivocal demonstration that the frequencies of synonymous substitution are very strikingly correlated for homologous genes from all mammals that could be tested. In other words, the frequencies of synonymous substitutions found when comparing homologous mammalian coding sequences are gene-specific. Indeed, genes showing high or low frequencies of synonymous substitutions are "fast" or "slow," respectively, in all mammalian species tested. The correlations between synonymous substitution frequencies and the characteristics of the gene products or the X/autosomal localization of genes will be discussed elsewhere (paper in preparation).

The variation in synonymous rate may be due, as in the case of nonsynonymous substitutions (see following section), to differences in mutation rate and/or to negative selection. This point will be further discussed after taking into consideration two other results reported here, namely (1) the gene-specificity of the frequencies of nonsynonymous substitutions in homologous mammalian genes and (2) the high correlation of the frequencies

of synonymous and nonsynonymous substitutions in the same homologous genes.

Nonsynonymous Substitution Frequencies of Homologous Mammalian Genes Are Gene-Specific

It is well known (Dickerson 1971; Dayhoff 1972; Nei 1987) that the rate of nonsynonymous substitutions is extremely variable among genes, since it ranges, in a human/murid comparison, from zero (in the genes for histones 3 and 4) to $279 \cdot 10^{-9}$ substitutions per site per year (in the gene for interferon γ; Li and Graur 1991). The average for the small set of genes taken into consideration by Li and Graur (1991) is $0.85 \pm 0.73 \cdot 10^{-9}$ substitutions per site per year. The close-to-1,000-fold range in nonsynonymous substitution rates found in different genes (neglecting those showing no substitution) may be ascribed to differences in the rate of mutation and/or to the intensity of selection.

As far as the first possibility is concerned, the difference in mutation rate is simply not large enough, by far, to account by itself for the phenomenon under consideration. This can be judged from the fact that synonymous mutation rates for mammalian genes vary by a value which is very far from the 1000-fold range of nonsynonymous substitutions. The most important factor in determining the rate of nonsynonymous substitution appears, therefore, to be the selection intensity, which is determined, in turn, by selective constraints on amino acids (Li and Graur 1991). The higher the selective constraints, the higher the chances of amino acid substitutions causing deleterious effects on protein function(s) and being eliminated by negative selection (advantageous mutations being very rare compared to deleterious mutations, they can be neglected for the sake of the present discussion). It should be noted that negative selection is also indicated by the actual amino acid changes observed, functionally crucial amino acids never being replaced and conservative substitutions being predominant over nonconservative ones. These conclusions are reinforced by the observation, reported in the present work, that the nonsynonymous substitution frequencies of all homologous genes available are comparable in all the mammalian species which could be tested (Fig. 4).

Synonymous and Nonsynonymous Substitution Frequencies of Homologous Genes Are Correlated

A correlation between synonymous and nonsynonymous substitution frequencies was first reported for small mammalian gene sets by several authors (Fitch 1980; Lipman and Wilbur 1985; Graur 1985; Li et al. 1985; Miyata et al. 1987; Tieher and Graur 1989) and, more recently, for a large set of genes from the mouse/rat pair (Wolfe and Sharp 1993). In the present work, this correlation was found in all pairs of homologous mammalian genes that could be tested.

It has been argued that the correlation between synonymous and nonsynonymous substitution frequencies can exist either (1) because the two rates are similar over the whole gene, such that conserved proteins have low divergence at silent sites, for some reason (Li and Graur 1991; Wolfe and Sharp 1993), or (2) because substitutions at adjacent nucleotide positions have occurred at a frequency greater than would be predicted from either substitution rate alone (Wolfe and Sharp 1993).

Wolfe and Sharp (1993) suggested that doublet substitutions are responsible, because, if the synonymous substitution rate is recalculated, ignoring those codons where the species differ by more than a single nucleotide substitution, the correlation coefficient between K_s and K_a for the mouse/rat pair drops from a significant value, 0.45, to a nonsignificant one, 0.10. Their approach did not distinguish, however, between adjacent (1–2, 2–3) and nonadjacent (1–3) changes, provided that the changes were located on the same codon. It is difficult to see how this approach can disprove the existence of a correlation between K_a and K_s, because the loss of a significant correlation was also caused by the elimination of the relatively frequent codons carrying 1–3 changes, which are irrelevant as far as doublet substitutions are concerned. Moreover, codons carrying doublet substitutions in positions 2–3 represent only 2% of all fourfold degenerate codons (which, in turn, represent only about half of the codons) carrying a third position change (see Table 2 of Wolfe and Sharp 1993).

Likewise, the argument that third position changes concern 25% of fourfold degenerate codons carrying a 2–3 change, but only 15% of codons where position 2 is unchanged, should be weighed against the fact that the first case concerns 2% of all such codons carrying a third codon position change and the second 98%. Moreover, whether this minute amount of doublet changes arises from mutational events involving simultaneous substitutions at two adjacent nucleotides (Wolfe and Sharp 1993) or from two separate events (Lipman and Wilbur 1985) is not clear.

In any case, the possibility that doublet substitutions are responsible for the correlation between synonymous and nonsynonymous substitutions appears to be essentially ruled out by the fact that if all codons comprising consecutive substitutions (in positions 1–2, 2–3, and 3–1) are ignored, the correlation coefficients of the human/bovine plot (Fig. 5) and of the rat/mouse plot only decrease from 0.57 to 0.30 and from 0.46 to 0.32, respectively, both values still remaining highly significant ($P = 10^{-3}$). As in Fig. 5, these correlations were calculated between SDF and log NSDF, because the relation between the two substitution rates is not linear. Using the same data, the correlation coefficient of a plot of K_s *vs* log K_a remains significant ($R = 0.31$ in the human/bovine

and $R = 0.38$ in the rat/mouse case). Therefore, even if a modest contribution to the correlation by doublet substitutions still remains a real possibility, there is no doubt about the existence of a significant correlation between synonymous and nonsynonymous substitutions, independently of doublet substitutions.

Synonymous Substitution Frequencies of Homologous Mammalian Genes Are not Mainly due to Differences in Mutation Rates

Differences in Mutation Rates Associated with Compositional Differences

Differences in average mutation rates (and in mutational biases) in third codon positions of coding sequences located in different isochores of mammalian genomes were claimed by several authors (Filipski 1988; Ticher and Graur 1989; Wolfe et al. 1989), but they could not be confirmed (Bernardi et al. 1993). The differences under consideration appear to be mainly due to using small gene samples when individual fluctuations in silent substitution frequency from gene to gene are relatively large (Bernardi et al. 1993). A lack of correlation between synonymous site divergence and GC levels in the mouse/rat comparison (as well as between K_4 and GC_4) was also independently reported by Wolfe and Sharp (1993). These authors found, however, a variation in K_4 (more specifically, a peak of K_4 values at 60% GC), but only when K_4 (the mutation rate at fourfold degenerate sites) was averaged over genes comprised within each 1% interval of GC_4 (the GC levels at fourfold degenerate sites). In any case, this effect is admittedly small—in fact, small enough to no longer support the claim "that the substitution rate and the base composition of silent sites vary together in a systematic way" and the ensuing speculations on the maintenance and origin of isochores (Wolfe et al. 1989).

We would like to suggest here that the small effect under discussion is, however, not simply due to sampling reasons, but to an underestimate of rates at the two ends of the compositional spectrum. In those regions, changes mainly being A $\leftrightarrows$ T and G $\leftrightarrows$ C, respectively, chances of back mutations are high, leading to an underestimate of rates. Indeed, in the human/bovine comparison, A $\leftrightarrows$ T and G $\leftrightarrows$ C changes largely increase, as expected, at either end of the compositional distribution, unlike other changes (paper in preparation).

An alternative explanation has been suggested by Gu and Li (1994), who developed a model based on the variation of nucleotide pools during the cell cycle (Mathews and Ji 1992). Some serious problems exist, however, with the idea that this variation is responsible for the formation of GC-rich isochores, like the late replication of both GC-poor and GC-rich satellite DNAs as well as of the inactive X chromosome and the lack of formation of GC-rich isochores in cold-blooded vertebrates, which share early and late replication with warm-blooded vertebrates (Bernardi et al. 1988; Bernardi 1993a).

While the results of Bernardi et al. (1993) and Wolfe and Sharp (1993) rule out differences in mutation rates that are associated with base composition of synonymous sites, they do not say anything about differences that are not associated with base composition. These differences should, however, be gene-specific in order to account for the findings reported in this paper.

An alternative view about such differences might be that they are not "gene-specific," but "region-specific," a view apparently receiving some support by the finding (Wolfe et al. 1989) that six pairs of genes which are physically linked in mouse and rat show close K_4 values. The pairs of genes under consideration (metallothionein I and II, albumin and α-fetoprotein, cytochromes P_1 450 and P_3 450, γ_A and γ_D crystallin, Ly-2 and Ly-3 antigens, immunoglobulin heavy chains Cδ, $C\gamma_3$, and Cε) are, however, functionally very close. The closeness of K_4 values may, therefore, be due to their common function rather than to their physical linkage. For this reason, the data of Wolfe et al. (1989) do not contradict our conclusion about the "gene specificity" of synonymous mutation rates.

Differences in Mutation Rates Not Associated with Compositional Differences

The mechanisms behind nucleotide substitutions are repair and replication errors. The most widespread and general repair mechanism, the nucleotide excision repair, is known to be much more efficient for certain expressed genes than for nonexpressed genes and other silent DNA sequences (Bohr et al. 1985; Mellon et al. 1986). Moreover, the preferential repair of an active gene concerns the transcribed DNA strand (Mellon et al. 1987), suggesting a mechanism that directly couples nucleotide excision repair to transcription.

Although there is no indication that repair is different for genes characterized by different transcription levels, one may wonder whether or not differential repair is the explanation for the gene specificity of synonymous and nonsynonymous substitution frequencies and for the correlation between the former and the latter. As far as nonsynonymous substitutions are concerned, it is clear that their frequencies are not simply determined by higher or lower repair efficiencies (nor by different rates of replication errors), but by negative selection (as generally admitted), because the changes in amino acids in any given protein are not random changes, but specific ones. As already mentioned, this is demonstrated by the fact that functionally crucial amino acids cannot be replaced and conservative amino acid replacements predominate over nonconservative changes. These observations indicate beyond any doubt that nonsynonymous

substitution frequency is not simply determined by the mutation process itself, and that negative selection plays a role.

In contrast, one might speculate that synonymous substitution frequencies are the result of different repair efficiencies and/or of gene-specific replication errors. These explanations should be judged, however, taking into account the observation that synonymous substitutions in fourfold degenerate positions exhibit specific patterns (paper in preparation), a finding paralleling the specific pattern of nonsynonymous substitutions (see above), and the demonstration (Eyre-Walker 1994) that differences in synonymous codon usage between mammalian genes are not due to differences in the efficiencies of DNA mismatch repair.

Acknowledgments. We wish to thank our colleagues Adam Eyre-Walker, Wen-Hsiung Li, Tomoko Ohta, and Ken Wolfe for their comments. Two of us (D.M., C.G.) acknowledge the financial support of GIP GREG.

References

Aissani B, D'Onofrio G, Mouchiroud D, Gardiner K, Gautier C (1991) The compositional properties of human genes. J Mol Evol 32:497–503

Aota S, Ikemura T (1986) Diversity in G + C content at the third position of codons in vertebrate genes and its cause. Nucleic Acids Res 14:6345–6355

Bernardi G, Olofsson B, Filipski J, Zerial M, Salinas J, Cuny G, Meunier-Rotival M, Rodier F (1985) The mosaic genome of warm-blooded vertebrates. Science 228:953–958

Bernardi G, Mouchiroud D, Gautier C, Bernardi G (1988) Compositional patterns in vertebrate genomes: conservation and change in evolution. J Mol Evol 28:7–18

Bernardi G (1989) The isochore organization of the human genome. Ann Rev Genet 23:637–661

Bernardi G (1993a) The vertebrate genome: isochores and evolution. Mol Biol Evol 10:186–204

Bernardi G (1993b) The human genome organization and its evolutionary history: a review. Gene 135:57–66

Bernardi G, Mouchiroud D, Gautier C (1993) Silent substitutions in mammalian genomes and their evolutionary implications. J Mol Evol 37:583–589

Bohr VA, Smith CA, Okumoto DS, Hanawalt PC (1985) DNA repair in an active gene: removal of pyrimidine dimers from the DHFR gene of CHO cells is much more efficient than in the genome overall. Cell 40:359–369

Bulmer M, Wolfe KH, Sharp PM (1991) Synonymous nucleotide substitution rates in mammalian genes: implications for the molecular clock and the relationship of mammalian orders. Proc Natl Acad Sci U S A 88:5974–5978

Dayhoff MO (1972) Atlas of protein sequence and structure, vol 5. National Biomedical Research Foundation, Washington, DC

Dickerson RE (1971) The structure of cytochrome c and the rates of molecular evolution. J Mol Evol 1:26–45

D'Onofrio G, Mouchiroud D, Aïssani B, Gautier C, Bernardi G (1991) Correlations between the compositional properties of human genes, codon usage and amino acid composition of proteins. J Mol Evol 32:504–510

Duret L, Mouchiroud D, Gouy M (1994) HOVERGEN: Homologous Vertebrate Genes data base. Nucleic Acids Res 22:2360–2363.

Eyre-Walker A (1994) DNA mismatch repair and synonymous codon evolution in mammals. Mol Biol Evol 11:88–98

Filipski J (1988) why the rate of silent codon substitutions is variable within a vertebrate's genome. J Theor Biol 134:159–164

Fitch WM (1980) Estimating the total number of nucleotide substitutions since the common ancestor of a pair of genes: comparison of several methods and three beta hemoglobin messenger RNA's. J Mol Evol 16:153–209

Gouy M, Gautier C, Attimonelli M, Lanave C, di Paola G (1985) ACNUC—a portable retrieval system for nucleic acid sequence databases: logical and physical designs and usage. Comput Appl Biosci 1:167–172

Graur D (1985) Aminoacid composition and the evolutionary rates of protein-coding genes. J Mol Evol 22:5362

Gu X, Li W-H (1994) A model for the correlation of mutation rate with GC content and the origin of GC-rich isochores. J Mol Evol 38: 468–475

Ikemura T, Aota S (1988) Global variation in G + C content along vertebrate genome DNA. J Mol Biol 203:1–13

Li W-H, Wu C, Luo CC (1985) A new method for estimating synonymous and non-synonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. Mol Biol Evol 2:150–174

Li W-H, Wu C-I (1987) Rates of nucleotide substitution are evidently higher in rodents than in man. Mol Biol Evol 4:74–77

Li W-H, Graur D (1991) Fundamentals of molecular evolution. Sinauer, Sunderland, MA

Lipman DJ, Wilbur WJ (1985) Interaction of silent and replacement changes in eukaryotic coding sequences. J Mol Evol 21:161–167

Mathews CK, Ji J (1992) DNA precursor asymmetries, replication fidelity, and variable genome evolution. Bioessays 14:295–301

Mellon I, Bohr VA, Smith CA, Hanawalt PC (1986) Preferential DNA repair of an active gene in human cells. Proc Natl Acad Sci U S A 83:8878–8882

Mellon I, Spivak G, Hanawalt PC (1987) Selective removal of transcription-blocking DNA damage from the transcribed strand of the mammalian DHFR gene. Cell 51:241–249

Miyata T, Hayashida H, Kuma K, Mitsuyasu K, Yasunaga T (1987) Male-driven molecular evolution: a model and nucleotide sequence analysis. Cold Spring Harbor Symp Quant Biol 52:863–867

Mouchiroud D, Gautier C (1988) High codon usage changes in mammalian genes. Mol Biol Evol 5:192–194

Mouchiroud D, Gautier C, Bernardi G (1988) The compositional distribution of coding sequences and DNA molecules in humans and murids. J Mol Evol 27:311–320

Mouchiroud D, Gautier C (1990) Codon usage changes and sequence dissimilarity between human and rat. J Mol Evol 31:81–91

Mouchiroud D, D'Onofrio G, Aissani B, Macaya G, Gautier C, Bernardi G (1991) The distribution of genes in the human genome. Gene 100:181–187

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York

Salinas J, Zerial M, Filipski J, Bernardi G (1986) Gene distribution and nucleotide sequence organization in the mouse genome. Eur J Biochem 16:469–478

Ticher A, Graur D (1989) Nucleic acid composition, codon usage, and the rate of synonymous substitution in protein-coding genes. J Mol Evol 28:286–298

Wolfe KH, Sharp PM, Li W-H (1989) Mutation rates differ among regions of the mammalian genome. Nature 337:283–285

Wolfe KH, Sharp PM (1993) Mammalian gene evolution: nucleotide sequence divergence between mouse and rat. J Mol Evol 37:441–456

Wu CI, Li W-H (1984) Evidence for higher rates of nucleotide substitution in rodents than in man. Proc Natl Acad Sci U S A 82:1741–1745

Zerial M, Salinas J, Filipski J, Bernardi G (1986) Gene distribution and nucleotide sequence organization in the human genome. Eur J Biochem 160:479–485

STATEMENT OF OWNERSHIP, MANAGEMENT, AND CIRCULATION (Required by 39 U.S.C. 3685). (1) Publication title: Journal of Molecular Evolution. (2) Publication No. 002-047. (3) Filing Date: 10/1/94. (4) Issue Frequency: Monthly. (5) No. of Issues Published Annually: 12. (6) Annual Subscription Price: $745.00. (7) Complete Mailing Address of Known Office of Publication: 175 Fifth Avenue, New York, NY 10010. (8) Complete Mailing Address of Headquarters or General Business Office of Publisher: 175 Fifth Avenue, New York, NY 10010. (9) Full Names and Complete Mailing Addresses of Publisher, Editor, and Managing Editor: Publisher: Springer-Verlag New York, Inc., 175 Fifth Avenue, New York, NY 10010. Editor: Emile Zuckerkandl, Ph.D., Institute of Molecular and Medical Sciences, 460 Page Mill Road, Palo Alto, CA 94306. Managing Editor: Springer-Verlag New York, Inc., 175 Fifth Avenue, New York, NY 10010. (10) Owner: Springer-Verlag Export GmbH, Tiergartenstrasse 17, D-69121 Heidelberg, Germany, and Springer-Verlag GmbH & Co. KG, Heidelberger Platz 3, D-14197 Berlin, Germany. (11) Known Bondholders, Mortgagees, and Other Security Holders Owning or Holding 1 Percent or More of Total Amount of Bonds, Mortgages, or Other Securities: Dr. Konrad Springer, Heidelberger Platz 3, D-14197 Berlin, Germany. (12) The purpose, function, and nonprofit status of this organization and the exempt status for federal income tax purposes: Has Not Changed During Preceding 12 Months. (13) Publication Name: Journal of Molecular Evolution. (14) Issue Date for Circulation Data Below: September 1994. (15) Extent and Nature of Circulation: (a.) Total No. Copies (Net Press Run): Average No. Copies Each Issue During Preceding 12 Months, 1283; Actual No. Copies of Single Issue Published Nearest to Filing Date, 1300. (b.) Paid and/or Requested Circulation: (1) Sales Through Dealers and Carriers, Street Vendors, and Counter Sales: Average No. Copies Each Issue During Preceding 12 Months, 0; Actual No. Copies of Single Issue Published Nearest to Filing Date, 0. (2) Paid or Requested Mail Subscriptions: Average No. Copies Each Issue During Preceding 12 Months, 761; Actual No. Copies of Single Issue Published Nearest to Filing Date, 768. (c.) Total Paid and/or Requested Circulation: Average No. Copies Each Issue During Preceding 12 Months, 761; Actual No. Copies of Single Issue Published Nearest to Filing Date, 768. (d.) Free Distribution by Mail: Average No. Copies Each Issue During Preceding 12 Months, 123; Actual No. Copies of Single Issue Published Nearest to Filing Date, 124. (e.) Free Distribution Outside the Mail: Average No. Copies Each Issue During Preceding 12 Months, 95; Actual No. Copies of Single Issue Published Nearest to Filing Date, 95. (f.) Total Free Distribution: Average No. Copies Each Issue During Preceding 12 Months, 218; Actual No. Copies of Single Issue Published Nearest to Filing Date, 219. (g.) Total Distribution: Average No. Copies Each Issue During Preceding 12 Months, 979; Actual No. Copies of Single Issue Published Nearest to Filing Date, 987. (h.) Copies Not Distributed: (1) Office Use, Leftovers, Spoiled: Average No. Copies Each Issue During Preceding 12 Months, 304; Actual No. Copies of Single Issue Published Nearest to Filing Date, 313. (2) Return from News Agents: Average No. Copies Each Issue During Preceding 12 Months, 0; Actual No. Copies of Single Issue Published Nearest to Filing Date, 0. Percent Paid and/or Requested Circulation: Average No. Copies Each Issue During Preceding 12 Months, 77.83; Actual No. Copies of Single Issue Published Nearest to Filing Date, 77.81. (16) This Statement of Ownership will be printed in the Vol. 40 #1 issue of this publication. I certify that all information furnished on this form is true and complete.

Craig Van Dyck
Vice-President, Production