J Mol Evol (1982) 19:28–35

Journal of Molecular Evolution

Molecular Clock of Silent Substitution: At Least Six-Fold Preponderance of Silent Changes in Mitochondrial Genes Over Those in Nuclear Genes

Takashi Miyata[1], Hidenori Hayashida[1], Reiko Kikuno[1], Masami Hasegawa[2], Midori Kobayashi[3], and Katsuro Koike[3]

[1]Department of Biology, Faculty of Science, Kyushu University, Fukuoka 812, Japan
[2]The Institute of Statistical Mathematics, 4–6–7 Minami-Azabu, Minato-ku, Tokyo 106, Japan
[3]Cancer Institute, Japanese Foundation for Cancer Research, Toshima-ku, Tokyo 170, Japan

Summary. For each of eleven different types of nuclear genes, comparisons of the protein coding sequences were made between human, mouse and rat pairwisely, and the evolutionary rate of silent substitution, $v_S^{nucl.}$, was estimated. It is shown that the $v_S^{nucl.}$ is not only very high (=5.37 x 10^{-9}/site/yr), but also approximately uniform for different genes regardless of the types, which confirms our previous results (Miyata et al. 1980b). This is in sharp contrast to the rate of protein evolution which differs greatly from protein to protein. Furthermore the $v_S^{nucl.}$ is shown to be approximately constant with respect to different divergence times, at least within a short time period ($\leqslant$ 75 Myr). Based on these observations, we propose a new molecular clock which has several advantages over a protein clock. Using this clock, we show that the rate of amino acid replacement in the immunoglobulin Cκ gene of b4 rabbit is unexpectedly high, almost comparable to the rate of silent changes. This rate may be the highest one for protein evolution that we know so far. We further examine the rate of silent substitutions in mitochondrial genes comparing mouse and rat. Surprisingly the rate is extremely high ($\geqslant$ 35 x 10^{-9}/site/yr), at least 6-times as high as the corresponding rate of nuclear genes. Based on the estimate, we discuss a possible origin of the rapid rate found in mitochondrial DNA.

Key words: Synonymous substitution – Uniform Rate of Evolution – Rapid Evolution of mtDNA – Cκ Gene of b4 Rabbit

Introduction

The preponderance of silent or synonymous changes in protein coding regions has been well documented for many genes (e.g., Perler et al. 1980; Jukes 1980; Miyata et al. 1980b) since the first observations in globin (Salser et al. 1976) and histone (Grunstein et al. 1976) genes. Kimura (1977) presented a first reliable estimate for the rate of synonymous substitutions: He showed that synonymous changes have occurred with the highest rate, almost comparable to the rate of fibrinopeptides, and emphasized the importance of these changes as an evidence for the neutral theory (Kimura 1968; Ohta 1974). The discovery of a mouse α-globin pseudogene (Nishioka and Leder 1980; Vanin et al. 1980) revealed a further important evolutionary characteristic of synonymous changes. By comparing the pseudogene with its functional counterpart, Miyata and Yasunaga (1981) and independently Kimura (1980) showed that the pseudogene evolves at a rapid rate, about 1.9 times as high as the rate of synonymous substitution in functional genes. Subsequent works on pseudogene evolution supported this result (Miyata and Hayashida 1981; Li et al. 1981; Takahata and Kimura 1981). These results suggest the operation of a functional constraint upon these changes. Indeed, Miyata and Hayashida (1981) showed that even synonymous changes are not completely free from selective pressure but are constrained in part, although weakly, depending on the degree of bias in code word usage.

Comparison of the rate of synonymous substitution between different genes revealed another evolutionary aspect. Previously we showed that the rate is not only very high but also roughly uniform for different genes regardless of the types (Miyata et al. 1980b; Miyata and Yasunaga 1980; Miyata 1981). This is in sharp contrast to the rate of protein evolution which differs greatly

Abbreviations: URF, Unidentified reading frame; yr, Year.

Offprint requests to: T. Miyata

0022–2844/82/0019/0028/$01.60

from protein to protein (e.g., Dickerson 1971; Wilson et al. 1977). These properties are useful as a molecular clock to show evolutionary branching orders of closely related spcies or genes in a multigene family, even when the encoded proteins evolve at different rates (Miyata et al. 1980b; Miyata and Hayashida 1982). Furthermore this clock would also provide additional information for determining whether two genes of identical function from different species are orthologous, i.e., their DNA sequence reflects the time passed since the speciation of the two species, or paralogous, i.e., they diverged long before the two species splited (Fitch and Margoliash 1970). It is however required for a molecular clock that the rate be constant with respect to different divergence times.

In this report, we confirm the rate constancy of synonymous substitution for different genes from a much wider variety of sources, and also show that the rate is actually constant with respect to divergence time. Furthermore we compare three mitochondrial (mt) DNA genes from rat (Saccone et al. 1981; Koike et al. manuscript in preparation) with mouse equivalents (Bibb et al. 1981) and show that, in mtDNA genes, the rate of synonymous substitution is uniform for different genes as in the case of nuclear genes, but is surprisingly high, at least six-fold greater than that of nuclear genes. This high rate of synonymous substitution in mtDNA genes would be a good clock particularly for knowing the time passed since the speciation of such closely related species as man and African apes and also would provide deeper insight into mechanisms by which mammalian mtDNA evolves at an extreme rate (Brown et al. 1979).

Methods

Calculation of Sequence Difference at the DNA Level. The sequence difference K or simply "difference" between a pair of homologous sequences is defined as a fraction of the number of mismatchs relative to the total number of sites compared. In coding regions, a single-base change in a codon either results in a replacement of the encoded amino acid (replacement substitution) or leads to the appearence of synonymous codon (synonymous substitution). We shall refer to a site where a single-base change leads to a synonymous (replacement) substitution as synonymous (replacement) site. We have calculated the difference at the synonymous site, K_S, and that at the replacement site, K_A, for coding regions as described previously (Miyata and Yasunaga 1980). For noncoding regions, the calculation procedure is straightforward except for treating gaps: A gap containing more than one nucleotide position was considered as a single base substitution in a single site. The estimated difference was corrected, if necessary, for multiple substitutions from the formula (Jukes and Cantor 1969; Kimura and Ohta 1972): $K^c = -(3/4)\ln[1 - (4/3)K]$, (by superscript, we mean corrected difference). For the genetic code table of mammalian mitochondria, we followed Anderson et al. (1981).

Calculation of Sequence Difference at the Protein Level. The procedure is the same as that for noncoding region at the DNA level. We will represent the amino acid difference per residue as k_{aa} to distinguish it from the K_A at the DNA level. Correction for multiple hits was made, if necessary, by the formula (Dickerson 1971; Kimura and Ohta 1972) : $k^c_{aa} = -\ln(1 - k_{aa})$.

Results and Discussion

Evolutionary Rate of Synonymous Substitution is Approximately Uniform for Different Genes

Previously we showed that the evolutionary rate of synonymous substitutions (v_S) is not only very high but also approximately constant between some different genes, including α- and β-globin, growth hormone and insulin genes (Miyata et al. 1980b; Miyata and Yasunaga 1980; Miyata 1982). To confirm this with a larger variety of genes, we examined the rates of synonymous substitution for 11 different genes comparing the coding sequences from human and rodents for each gene. Assuming the human/rodent split took place 75 Myr ago (Dayhoff 1978), the v_S was evaluated from the formula: $v_S = K^c_S / 2T$, here T is the time of divergence of sequences compared. Table 1 shows the estimated values of K_S and v_S for 11 mammalian nuclear genes. Apparently the K_S and v_S are similar for the different genes examined so far. The average values of K_S corresponding to the comparison between human and rodents, and of v_S are calculated to be 0.492 ± 0.032 and $5.37 \pm 0.63 \times 10^{-9}$/site/yr, respectively. This is in sharp contrast to the rate of evolution at the replacement site (v_A), the values of which vary from gene to gene (data not shown) and are distributed over a much wider range, varying from 0.23×10^{-9} for the insulin gene to 5.9×10^{-9} for the immunoglobulin Cκ gene of b4 rabbit (see Table 3).

Although the v_S values are roughly constant for different genes, they seem to differ slightly depending on the types of genes. There may be several reasons for this: First the difference may be a result of fluctuation due to the small number of sites compared. Secondly, since eukaryotic genes are known to be often duplicated, we cannot exclude the possibility that the compared sequences are paralogous for some genes, for example, prolactin and chorionic gonadotropin genes where the estimated values of v_S are slightly larger than those of the other genes. Thirdly, and presumably most importantly, nonrandom use of degenerate codons would influence variability at the synonymous sites. From the comparison of the pseudogene with its functional counterpart, we previously showed that base alternations at the synonymous sites are not completely free but are constrained to some extent, although weakly, depending on the degree of bias in code word usage (Miyata and Hayashida 1981). Although codon usage pattern seems to differ even between genes from mammals, such a constraint would be very weak, considering the rapid rate of synonymous substitution.

Table 1. The rate of evolution at the synonymous site of coding region (v_S) for various mammalian genes

Gene	Species pair	Refs.	K_S	$\bar{K}_S$	$v_S \times 10^{-9}$
a) Nuclear genes					
(i) T = 75 Myr					
α-globin	h/m	(1/ 2)	0.518	0.502	5.53
	m/ra	(2/ 3)	0.486		
β-globin[1]	h/m	(4/ 5)	0.459	0.483	5.17
	m/ra	(5/ 6)	0.507		
GH	h/r	(7/ 8)	0.471		4.95
CS	hCS/rGH	(9/ 8)	0.475		5.01
Preproinsulin (insulin)	h/r	(10/11)	0.474		5.00
			(0.474)		(5.00)
Prolactin	h/r	(12/13)	0.547		6.52
CG α-subunit	h/m	(14/15)	0.534		6.23
Ig V_κ[2]	h/m		0.432		4.29
Ig C_κ	h/m	(16/16)	0.513		5.77
Ig C_μ	h/m	(17/18)	0.484		5.19
Average			0.492		5.37
(S.D.)			(0.032)		(0.63)
(ii) T = 46 Myr					
α-globin	h/ra	(1/ 3)	0.291		
β-globin[1]	h/ra	(4/ 6)	0.360	0.331	4.75
Ig C_κ	h/ra	(16/19)	0.342		
(iii) T = 37 Myr					
β-globin[3]	h/c	(4/20)	0.261		4.34
(iv) T = 17 Myr					
Amylase	r/m	(21/22)	0.166	0.176	5.90
Ig C_κ	r/m	(23/16)	0.186		
b) mtDNA genes					
(i) T = 75 Myr					
Overall[4]	h/m	(24/25)	0.712		---
(ii) T = 17 Myr					
cyt b	r/m	(26/25)	0.595		
COI[3]	r/m	(26/25)	0.560	0.593	⩾ 34.5
URF1[3]	r/m	(27/25)	0.625		

GH, growth hormone; CS, chorionic somatomammotropin; CG, chorionic gonadotropin; IgV and IgC, immunoglobulin variable and constant domain genes, respectively; cyt. b, cytochrome b; COI, cytochrome oxidase subunit I; URF1, unidentified reading frame 1; h, human; r, rat; ra, rabbit; c, cebus; K_S and v_S, difference (per site) and evolutionary rate (per site per yr) at the synonymous site, respectively; $\bar{K}_S$, average value of K_S; T, divergence time of species compared. The values of T were taken from Dayhoff (1978) for h/m and h/r, from Martin et al. (1981) for h/c and from Table 2 for r/m and h/ra.

[1]The segments (codons 22–35 and 92–107) showing strong sequence conservation at the DNA level were excluded from the analysis (Miyata et al. 1980b)

[2]Average value for four Vκ genes. For sequence data, see Hayashida and Miyata (submitted)

[3]Only the fragment length sequence is available

[4]All the coding regions excluding URF6AL were compared

References: 1, Wilson et al. (1980); 2, Nishioka and Leder (1979); 3, Heindell et al. (1978); 4, Lawn et al. (1980); 5, Konkel et al. (1979); 6, Van Ooen et al. (1979); Martial et al. (1979); 8, Seeberg et al. (1977); 9, Shine et al. (1977); 10, Bell et al. (1980); 11, Lomedico et al. (1979); 12, Cooke et al. (1981); 13, Cooke et al. (1980); 14, Fiddes and Goodman (1979); 15, Chin et al. (1981); 16, Hieter et al. (1980); 17, Rabbitts et al. (1981); 18, Kawakami et al. (1980); 19, Heidmann et al. (1981); 20, Martin et al. (1981); 21, MacDonald et al. (1980); 22, Hagenbüchle et al. (1980); 23, Sheppard and Gutman (1981); 24, Anderson et al. (1981); 25, Bibb et al. (1981); 26, Koike et al. (unpublished data); 27, Saccone et al. (1981)

Table 2. The estimate for the times (T) of separation of (I) rat and mouse and of (II) human and rabbit from comparisons of mammalian gene sequences at the DNA and protein levels

	K_1	K_{2a}	K_{2b}	T (Myr)
(I)	Rat/ Mouse	Rat/ Human	Human/ Mouse	
DNA sequence (replacement)	(Difference per base)			
1) cytochrome b	0.04	0.14	0.14	20
2) COI	0.01	0.05	0.05	15
3) URF1	0.05	0.23	0.20	15
Average				16.7 ± 2.4
(II)	Human/ Rabbit	Human/ Others	Rabbit/ Others	
a) DNA sequence	(Difference per base)			
1) α-globin				
Noncoding	0.25	0.32	0.38	48
Synonymous	0.27	0.52	0.49	31
Replacement	0.11	0.08	0.11	87
2) β-globin				
Noncoding	0.22	0.31	0.32	49
Intron	0.31	0.44	0.45	43
Synonymous	0.36	0.46	0.51	35
Replacement	0.07	0.15	0.16	32
b) Protein sequence	(Difference per residue)			
3) Cytochrome b5	0.11	0.20	0.19	46
4) Cytochrome c	0.09	0.11	--	56
5) Myoglobin	0.10	0.15	0.14	53
6) Carbonic anhydrase	0.15	0.21	0.25	47
7) Insulin	0.02	0.08	0.07	19
8) Immunoglobulin Cγ	0.33	0.38	0.39	61
Average*				45.5 ± 9.2

For each of the different proteins and genes, "T" was estimated as $T = (K_1^c/K_2^c) \times 75$ Myr, where $K^c = -(3/4)\ln[1 - (4/3)K]$ for the comparison at the DNA level and $K^c = -\ln[1 - K]$ for the comparison at the protein level, and $K_2 = (K_{2a} + K_{2b})/2$. "Others" means mammalian species other than human and rabbit: mouse for the cases of 1), 2) and 8) of (II), bovine and pig for 3), horse, donkey, camel and seal for 4), horse, bovine and sheep for 5), bovine and sheep for 6), mouse and sheep for 7). Since in cytochrome c, the rate of protein evolution is clearly higher in the primate line than in the other mammalian lines (Dickerson 1971), T was estimated as $T = (K_1^c/K_{2a}^c) \times 75$ Myr in this case. Since the values of T estimated from the replacement sites of α-globin gene and from insulin are unusual compared with others, these two cases were excluded from the calculation of the average value of T. For references of DNA sequence data, see footnote of Table 1. The values of differences for 3)–8) of (II) were taken from Dayhoff (1978)

Table 3. Comparison of IgCκ gene sequence of b4 rabbit with those of human and mouse and the evolutionary rate of the replacement sites

a)	IgCκ		Globin		
			α	β	
	K_A	K_S	K_S	K_S	$\bar{K}_S$
Human/Rabbit	0.327	0.342	0.291	0.360	0.331
Rabbit/Mouse	0.312	0.541	0.518	0.459	0.506
Human/Mouse	0.241	0.513	0.486	0.507	0.502
b)	K_A^c	σ_A	T (Myr)	$v_A \times 10^{-9}$	
O/Rabbit	0.272	0.034	46	5.9 ± 0.7	
O/Human	0.158	0.026	46	3.4 ± 0.6	
O/Mouse	0.133	0.024	110	1.2 ± 0.2	

K_S and K_A, sequence differences at the synonymous and replacement sites, respectively. $\bar{K}_S$, the average value of K_Ss of α- and β-globin genes. K_A^c, corrected difference at the replacement sites (i.e., $K_A^c = -(3/4)\ln[1-(4/3)K_A]$). Assuming that total number of base substitutions be distributed in poisson fashion, the standard error σ_A was estimated as $\sigma_A = \sqrt{K_A^c/N_A}$ (Miyata and Hayashida 1982), where N_A is the average number of replacement sites in the sequence compared (N_A = 240.6). "O", the common ancestor of rabbit and human, the date of which goes back to 46 Myr ago (Table 2). T, the time span between "O" and the temporal species. The rate of evolution at the replacement site v_A (per site per yr) is given by $v_A = (K_A^c \pm \sigma_A)/T$

Thus an approximate constancy might occur. This relatively high and constant character of synonymous substitution is reasonably understandable from the viewpoint of the neutral theory (Kimura 1968; Ohta 1974; Kimura and Ohta 1974).

The Rate of Synonymous Substitution is Approximately Constant for Different Times of Divergence at Least within a Short Time Period

To determine whether or not the extent of sequence divergence at the synonymous site or K_S^c depends linearly on time T since the separation of the compared sequences, we further compared the coding sequences between human and rabbit, between human and cebus and between rat and mouse for several genes for which DNA sequences are available, and calculated differences at the synonymous sites for each gene. Since it is difficult to obtain reliable estimates for the time since separation of these species pairs from paleontological data, an alternative approach has been used: It has been established that each protein evolves at a roughly constant rate in most cases (e.g., Dickerson 1971; Kimura and Ohta 1974; Wilson et al. 1977). This constancy is particularly useful for fixing a branching point in the time at which species separate, when many different proteins are used. From the pairwise comparison of mtDNA gene sequences of human, mouse and rat, we have calculated K_A of each gene for each species pair. Assuming that human and rodents diverged 75 Myr ago, the time T since the separation of rat and mouse is estimated as $T = 75 \times k_1^c / [(k_{2a}^c + k_{2b}^c) / 2]$ (Myr), here k_1^c, k_{2a}^c and k_{2b}^c are corrected differences at the replacement sites for species pairs rat/mouse, rat/human and mouse/human, respectively. Averaging the estimated values of T over three different genes, cytochrome b, cytochrome oxidase subunit I (COI) and URFl genes, we finally have T = 17 ± 2 Myr for the rat/mouse divergence time.

Comparison of DNA sequences of several regions for α- and β-globin genes together with amino acid sequences of several proteins show that rabbit is more closely related to human than to other mammals. By the same procedure as used for rat and mouse, we estimated the separation time of human and rabbit as being on average 46 Myr ago (Table 2).

Using the estimated times of species separation (for human/cebus, we used data from Martin et al. (1981)), we calculated the average rate of synonymous substitution for each of three different cases where divergence times are different from each other (Table 1). Fig. 1 shows a plot of corrected differences at the synonymous sites, K_S^c, of nuclear genes against the divergence times, together with the cases of pseudogene (Miyata and Yasunaga 1981) and mtDNA genes (see below), for comparison. All the points fall nearly on a straight line for nuclear genes, implying that the rate of evolu-

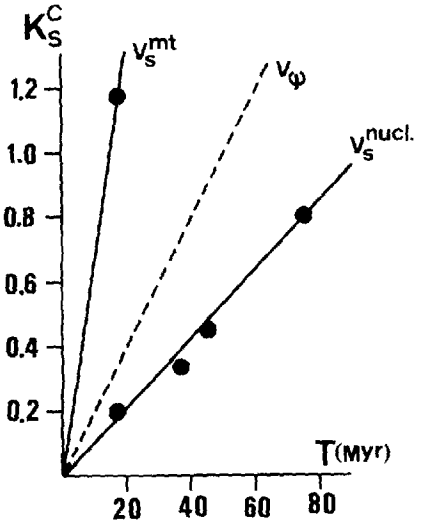

Fig. 1. The rates of evolution at the synonymous sites of coding region in nuclear genes ($v_S^{nucl.}$) and mtDNA genes (v_S^{mt}), showing approximate constancy with respect to time T. K_S^c, corrected difference at the synonymous site. The K_S^c and T values corresponding to each point were taken from Table 1. According to Miyata and Yasunaga (1981), the rate of pseudogene evolution (v_ψ) is estimated to be approximately 1.9 times as large as the $v_S^{nucl.}$, which also included in the figure for comparison

tion at the synonymous site is approximately constant with respect to divergence time, at least within a short time period (i.e., $T \leqslant 75$ Myr).

The Rate of Synonymous Substitution of mtDNA Genes is Extremely High, At Least 6 Times as High as the Corresponding Rate of Nuclear Genes.

In comparing mtDNA sequences of human and two other mammals, bovine and mouse, we previously showed that the sequence differences at the synonymous sites are already at saturation level for the two genes (Miyata et al. submitted). Comparison of the complete sequences of mtDNA between human and mouse revealed that, for all the genes except for the URF6AL, the synonymous sites are extremely divergent, on average 0.71 in K_S value, which supports the previous result. For estimating the v_S of mtDNA genes, it is therefore necessary to compare sequences of more closely related species. At present, nucleotide sequences of cytochrome b, COI and URFl genes are available for rat mtDNA (Saccone et al. 1981; Koike et al. manuscript in preparation). We have compared the corresponding gene sequences of mouse and rat and calculated the K_S for each gene (Table 1). The estimated K_S values are similar for the three different genes as is the case in nuclear genes. Surprisingly, the synonymous sites of mtDNA genes are extremely divergent, when compared with the corresponding sites of nuclear genes for the same comparison between rat and mouse: The K_S^c value of the former (= 1.17 on average) is about 6 times larger than that of the latter (=0.20 on average for amylase and IgCκ genes). Using the divergence time between rat and mouse, the rate of synonymous substitution of mtDNA genes was estimated as on average 35×10^{-9}/site/yr (Table 1). This value is about 6 times larger than 5.37×10^{-9}, the corresponding rate of nuclear genes. Note that the Jukes-Cantor formula could not fully correct for multiple hits for such a large K value ($K \sim 0.6$), and thus the actual value may be slightly larger than 35×10^{-9}. (Recently more elaborate formulae for correcting for the multiple hits were developed by Kimura (1980,1981) and Takahata and Kimura (1981). Unfortunately these formulae could not be applied directly to our K_S and K_A for the correction of multiple hits. Computer simulation showed that the values estimated by Jukes-Cantor formula are in good agreement with the actual values for $K_S \leqslant 0.55$).

The Rapid Evolution of Protein Coding Genes in MtDNA Possibly is Due to an Elevated Mutation Rate

From the comparison of restriction endonuclease cleavage maps and melting temperatures of heteroduplex DNAs for different species, Brown et al. (1979) showed that the rate of mtDNA evolution is approximately 10 times as high as that of nuclear DNA. The evolutionary rate of a gene in general is proportional to the rate of mutations occurring on the DNA and inversely proportional to the degree of functional constraint against the mutations (Kimura and Ohta 1974; Dickerson 1971; Wilson et al. 1977). Thus the rapid rate of mtDNA evolution is due to an elevated mutation rate, to a relaxed functional constraint or to both (Brown et al. 1979). Brown et al. (1979) and Brown (1981) suggested the possibility of an elevated mutation rate for mtDNA. Comparison of the rates of synonymous substitutions between mtDNA and nuclear genes would provide us with an important insight here: Since synonymous sites are known to be constrained very weakly, the major difference between them may be responsible for the elevated mutation rate in mtDNA. From the observation of an unusually strong divergence at the synonymous sites of mtDNA genes, we previously suggested the possibility of an elevated mutation rate for mtDNA (Miyata et al. submitted). The present observation that the v_S of mtDNA genes is extremely high, at least 6 times higher than that of nuclear genes and at least 3 times higher than that of pseudogenes (see Fig.1) strongly supports the above argument at least for protein coding genes. Furthermore the latter indicates an at least 3-fold elevation of mutation rate in mtDNA relative to nuclear DNA.

In contrast, the rate of substitution at replacement sites (v_A) for mtDNA genes is as low as for nuclear genes: The v_A value estimated from the comparison between human and mouse is on the average 1.25×10^{-9}/site/yr, excluding the overlapping gene segments and URF6AL genes (the URF6AL is unusual in that, unlike the other genes, it has an apparently small K_S value (= 0.55) relative to those of other genes, which is contrasted with its large K_A value (= 0.36) between human and mouse). This value ($v_A = 1.25 \times 10^{-9}$) is comparable to the corresponding rate of β-globin gene ($= 1.13 \times 10^{-9}$). Considering the 3-fold elevation of mutation rate of mtDNA, this figure divided by a factor 3 ($= 0.42 \times 10^{-9}$) is only 2 times the rate of insulin ($v_A = 0.23 \times 10^{-9}$) at the most, one of the slowly evolving molecules. Thus the replacement sites of mtDNA genes are constrained strongly as compared with the corresponding sites of nuclear genes. It is very difficult to explain these results by relaxed functional constraint alone. We rather prefer an alternative interpretation that, at least for protein coding genes, rapid rate of mtDNA evolution is due largely to its elevated mutation rate.

The Evolutionary Rate of Synonymous Substitution As a New Molecular Clock

The evolutionary features characteristic of synonymous substitutions could be summerized for both nuclear and mitochondrial DNA genes as follows: (i) The evolutio-

nary rate of synonymous substitution, v_S, is greater than that of replacement substitution, v_A, for most genes. (ii) The v_S is approximately constant for many different types of genes. (iii) The v_S remains essentially unchanged for different divergence times at least within a short evolutionary period. (iv) The estimated value of v_S is on the average 5.37×10^{-9}/site/yr for nuclear genes and 35×10^{-9}/site/yr for mtDNA genes. These properties are useful as a "clock" to show the evolutionary branching orders of closely related species or genes in multigene families. For example, knowing the sequence difference at the synonymous sites, K_S, between different species or between different genes in a multigene family, the divergence time (T) between them could be estimated as

$$T = -(3/4)\ln[\,1 - (4/3)K_S\,] / 2v_S \quad , \qquad (1)$$

here v_S is 5.37×10^{-9} for nuclear genes and 35×10^{-9} for mtDNA genes.

The "silent substitution clock" has several advantages for determining branching orders of closely related species and genes as compared with a "protein clock". First the constancy of v_S for different genes allows one to use the same v_S value for any genes regardless of the types. This is in sharp contrast to the "protein clock", the rate of which varies for different proteins (Dickerson 1971; Kimura and Ohta 1974; Wilson et al. 1977). Second this property is also useful for determining the divergence time of duplicated genes, even when the encoded proteins evolve at different rates (Miyata et al. 1980b; Miyata and Hayashida 1982). Third one occasionally finds cases at which a certain protein lineage shows anomalously large amounts of sequence change relative to the homologous proteins in other lineages (Wilson et al. 1977). Theoretically, two alternative interpretations are possible for this: One is the possibility of different rates for different lineages, and the other is that, although the rate may have remained essentially unchanged throughout evolution, the comparison is paralogous but not orthologous. The silent substitution clock might provide additional information in this regard. Forth the high evolutionary rate results in sufficient amounts of base substitutions at the synonymous sites, even when comparing such closely related species as man and African apes (i.e., chimpanzee and gorilla), to allow one to obtain statistically reliable estimates. MtDNA genes are particularly suited for such a purpose. For example, according to biochemical estimates based on the protein clock and cleavage maps of DNA, the time of speciation of man and African apes was shown to be about 5 Myr ago (see Wilson et al. (1977) for review). If this estimate is really correct, it is expected to have a value 0.28 for K_S of mtDNA genes between these species. On the other hand, many anthropologists believe that the speciation goes back to a more remote time, about 30 Myr ago (see Wilson et al. (1977) for review). This divergence time leads to the K_S of 0.70 in value, close to a saturation level. In any case, the observed number of synonymous substitutions is expected to be large enough for a reliable estimate of the separation time. If a complete or partial mtDNA sequence is available either for chimpanzee or gorilla, we can test with a certain amount of confidence which of the estimates is really the right.

The Replacement Sites in the IgCκ Gene of b4 Rabbit are Unexpectedly Divergent

Serological analysis of rabbit immunoglobulin κ chain revealed the presence of four allotypic forms designated by the symbols b4, b5, b6 and b9 (Farnsworth et al. 1976). The amino acid sequences of the Cκ region of b4 (Heidmann et al. 1981), b5 (Chersi et al. 1980) and b9 (Farnsworth et al. 1976) show a strong divergence; they differ by 26–37% of their amino acid sequences. Furthermore, comparison of the nucleotide sequences of Cκ genes between the b4 rabbit and mouse showed that they differ by 39 % in their coding regions (Heidmann et al. 1981). This extensive divergence of allelic forms of rabbit Cκ sequences together with the observation that a single rabbit can express three b allotypes have led to a hypothesis that b allotypes in the rabbit are encoded by closely linked duplicated genes regulated by polymorphic control mechanism (Farnsworth et al. 1976; Chersi et al. 1980). This hypothesis affords a possibility that the extensive divergence found between b4 rabbit and mouse is a result of paralogous comparison. To test this, we have calculated the K_S and K_A of the Cκ gene comparing b4 rabbit with two other mammals, human and mouse, (Hieter et al. 1980), for which DNA sequences are available.

Table 3 shows the K_S and K_A of Cκ gene for three species pairs rabbit/human, rabbit/mouse and human/mouse, together with the K_S values of α- and β-globin genes for the same species pairs for comparison. The K_S values are similar between the Cκ gene and the globin genes for each pair, indicating that the comparisons between the b4 rabbit and the two other mammals are orthologous for the Cκ gene. The comparison of Cκ genes of the b4 rabbit and human shows a surprising divergence at the replacement sites; the K_A is almost the same in value to the K_S. This is quite unusual in that K_S is much larger than K_A in many genes examined so far (Miyata et al. 1980b). Similar cases showing $K_A \simeq K_S$ were reported for the comparisons between duplicated IgVλ genes (Miyata et al. 1980b) and between duplicated H-2 antigen genes (Bregegere et al. 1981). In these two cases, however, the base substitutions are very rare, so that we can not exclude the possibility that this is simply the result of fluctuation due to the small amount of base changes.

It is of particular interest to know the rate of replacement substitutions for the Cκ gene of b4 rabbit. To calculate the rate, we adopted the same procedure as described previously (Miyata and Hayashida 1982; Miyata et al. 1982): Writing "O" for an ancestral node of rabbit and human, we have

$$K_A^c(0/r) = (1/2)\,[K_A^c(r/h) + K_A^c(r/m) - K_A^c(h/m)], \quad (2a)$$

$$K_A^c(0/h) = (1/2)[K_A^c(r/h) + K_A^c(h/m) - K_A^c(r/m)], \quad (2b)$$

$$K_A^c(0/m) = (1/2)[K_A^c(r/m) + K_A^c(h/m) - K_A^c(r/h)], \quad (2c)$$

here r, h and m stand for rabbit, human and mouse, respectively, and $K_A^c(0/r)$, for example, for a corrected difference between 0 and r. Using the differences calculated by (2a) - (2c), we could estimate the rate of replacement substitution, v_A, for each of the three different lineages (Table 3). Surprisingly the v_A of b4 rabbit is unusually high (5.9×10^{-9}/site/yr), almost comparable to the rate of synonymous substitution. This rate is the highest one for replacement substitution that we so far know. We further compared the Cκ protein sequences of the allotypes b4, b5 and b9 with those of other species, human and mouse. It was shown that the protein sequences of b5 and b9 evolve at a similar rate to that of b4 (data not shown). Although we have no satisfactory explanation for this unexpected divergence, it may be possible that, during immunoglobulin κ-type gene evolution, the Cκ exon had been exchanged for the corresponding exon of a pseudogene in the rabbit lineage by a mechanism such as has been found in mouse immunoglobulin γ-type constant domain genes (Miyata et al. 1980a).

Acknowledgements. We thank Prof. H. Matsuda and Dr. T. Yasunaga for discussions and comments. This work was supported in part by grants from the Ministry of Education, Science and Culture of Japan.

References

Anderson S, Bankier AT, Barrell GB, de Bruijn MHL, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJH, Staden R, Young IG (1981) Sequence and organization of the human mitochondrial genome. Nature 290:457–464

Bell GI, Pictet RL, Rutter WJ, Cordell B, Tischer E, Goodman HM (1980) Sequence of the human insulin gene. Nature 284:26–32

Bibb MJ, Van Etten RA, Wright CT, Walberg MW, Clayton DA (1981) Sequence and gene organization of mouse mitochondrial DNA. Cell 26:167–180

Bregegere F, Abastodo JP, Kvist S, Rask L, Lalanne JL, Garoff H, Cami B, Wiman K, Larhammar D, Peterson PA, Gachelin G, Kourilsky P, Dobberstein B (1981) Structure of C-terminal half of two H-2 antigens from cloned mRNA. Nature 292:78–81

Brown WM, George M Jr, Wilson AC (1979) Rapid evolution of animal mitochondrial DNA. Proc Natl Acad Sci USA 76: 1967–1971

Brown WM (1981) Mechanisms of evolution of animal mitochondrial DNA. Annals NY Acad Sci 361:119–134

Brown WM, Prager EM, Wang A, Wilson AC (1982) Mitochondrial DNA sequences of primates: Tempo and mode of evolution. J Mol Evol 18:225–239

Chersi A, Alexander CB, Mage R (1980) Partial primary structure of the immunoglobulin light chain constant region of a single rabbit of b5 allotype. Mol Immunol 17:1515–1523

Chin WW, Kronenberg HM, Dee PC, Maloof F, Habener JF (1981) Nucleotide sequence of the mRNA encoding the pre-α-subunit of mouse thyrotropin. Proc Natl Acad Sci USA 78:5329–5333

Cooke NE, Coit D, Weiner RI, Baxter JD, Martial JA (1980) Structure of cloned DNA complementary to rat prolactin messenger RNA. J Biol Chem 255:6502–6510

Cooke NE, Coit D, Shine J, Baxter JD, Martial JA (1981) Human prolactin. J Biol Chem 256:4007–4016

Dayhoff MO (1978) In Atlas of protein sequence and structure. Dayhoff MD (ed) vol 5, suppl 3, National Biomedical Research Foundation, Silver Spring, MD

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

Fransworth V, Goodfliesh R, Rodkey S, Hood L (1976) Immunoglobulin allotypes of rabbit kappa chains: Polymorphism of a control mechanism regulating closely linked duplicated genes? Proc Natl. Acad Sci USA 73:1293–1296

Fiddes JC, Goodman HM (1979) Isolation, cloning and sequence analysis of the cDNA for the α-subunit of human chorionic gonadotropin. Nature 281:351–356

Fitch WM, Margoliash E (1970) The usefulness of amino acid and nucleotide sequences in evolutionary studies. Evol Biol 4:76–109

Grunstein M, Schede P, Kedes L (1976) Sequence analysis and evolution of sea urchin (*Lytechinus pictus* and *Strongylocentrotus purpuratus*) histone H4 messenger RNAs. J Mol Biol 104:351–369

Hagenbuchle O, Bovey R, Young RA (1980) Tissue-specific expression of mouse α-amylase genes: Nucleotide sequence of isoenzyme mRNAs from pancreas and salivary gland. Cell 21: 179–187

Heidemann O, Auffray C, Cazenave P, Rougeon F (1981) Nucleotide sequence of constant and 3' untranslated regions of a κ immunoglobulin light chain mRNA of a homozygous b4 rabbit. Proc Natl Acad Sci USA 78:5802–5806

Heindell HC, Paddock GV, Studnicka GM, Salser WA (1978) The primary sequence of rabbit α-globin mRNA. Cell 15: 43–54

Hieter PA, Max EE, Seidmann JG, Maizel JV Jr, Leder P (1980) Cloned human and mouse kappa immunoglobin constant and J region genes conserved homology in functional segments. Cell 22:197–207

Jukes TH, Cantor CR (1969) Evolution of protein molecules. In: Munro HN (ed) Mammalian protein metabolisms II, vol III, Academic Press, New York, pp 21–132

Jukes TH (1980) Silent nucleotide substitutions and the molecular evolutionary clock. Science 210:973–978

Kawakami T, Takahashi N, Honjo T (1980) Complete nucleotide sequence of mouse immunoglobulin μ gene and comparison with other immunoglobulin heavy chain genes. Nucleic Acids Res 8:3933–3945

Kimura M (1968) Evolutionary rate at the molecular level. Nature 217:624–626

Kimura M, Ohta T (1972) On the stochastic model for estimation of mutational distance between homologous proteins. J Mol Evol 2:87–90

Kimura M, Ohta T (1974) On some principles governing molecular evolution. Proc Natl Acad Sci USA 71:2848–2852

Kimura M (1977) Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. Nature 267:275–276

Kimura M (1980) A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16:111–120

Kimura M (1981) Estimation of evolutionary distances between homologous nucleotide sequences. Proc Natl Acad Sci USA 78:454–458

Konkel DA, Maizel JV Jr, Leder P (1979) The evolution and sequence comparison of two recently diverged mouse chromosomal β-globin genes. Cell 18:865–873

Lawn RM, Efstratiadis A, O'Connell C, Maniatis T (1980) The nucleotide sequence of the human β-blobin gene. Cell 21: 645–651

Li W, Gojobori T, Nei M (1981) Pseudogenes as a paradigm of neutral evolution. Nature 292:237–239

Lomedico P, Rosenthal N, Efstratiadis A, Gilbert W, Kolodner R, Tizard R (1979) The structure and evolution of the two nonallelic rat preproinsulin genes. Cell 18:545–558

MacDonald RJ, Crerar MM, Swain WF, Pictet RL, Thomas G, Rutter WJ (1980) Structure of a family of rat amylase genes. Nature 287:117–122

Martial JA, Hallewell RA, Baxter JD, Goodman HM (1979) Human growth hormone: Complementary DNA cloning and expression in bacteria. Science 205:602–607

Martin SL, Zimmer EA, Davidson WS, Wilson AC, Kan YW (1981) The untranslated regions of β-globin mRNA evolve at a functional rate in higher primates. Cell 25:737–741

Miyata T, Yasunaga T (1980) Molecular evolution of mRNA: A method for estimating evolutionary rates of synonymous and amino acid substitutions from homologous nucleotide sequences and its application. J Mol Evol 16:23–36

Miyata T, Yasunaga T, Yamawaki-Kataoka Y, Obata M, Honjo T (1980a) Nucleotide sequence divergence of mouse immunoglobulin γ1 and γ2b genes and the hypothesis of intervening sequence-mediated domain transfer. Proc Natl Acad Sci USA 77:2143–2147

Miyata T, Yasunaga T, Nishida T (1980b) Nucleotide sequence divergence and functional constraint in mRNA evolution. Proc Natl Acad Sci USA 77:7328–7332

Miyata T, Yasunaga T (1981) Rapidly evolving mouse α-globin related pseudo gene and its evolutionary history. Proc Natl Acad Sci USA 78:450–453

Miyata T, Hayashida H (1981) Extraordinarily high evolutionary rate of pseudogenes: Evidence for the presence of selective pressure against changes between synonymous codons. Proc Natl Acad Sci USA 78:5739–5741

Miyata T (1982) Evolutionary changes and functional constraint in DNA sequences. In: Kimura M (ed) Molecular Evolution, protein polymorphism and the neutral theory. Japan Scientific Press, Tokyo, pp 233–266

Miyata T, Hayashida H (1982) Recent divergence from a common ancestor of human IFN-α genes. Nature 295:165–168

Miyata T, Kikuno R, Ohshima Y (1982) A pseudogene cluster in the leader region of the *Euglena* chloroplast 16S-23S rRNA genes. Nucleic Acids Res 10:1771–1780

Nishioka Y, Leder P (1979) The complete sequence of a chromosomal mouse α-globin gene reveals elements conserved throughout vertebrate evolution. Cell 18:875–882

Nishioka Y, Leder A, Leder P (1980) Unusual α-globin-like gene that has cleanly lost both globin intervening sequences. Proc Natl Acad Sci USA 77:2806–2809

Ohta T (1974) Mutational pressure as the main cause of molecular evolution and polymorphism. Nature 252:351–354

Perler F, Efstratiadis A, Lomedico P, Gilbert W, Kolonder R, Dodgson J (1980) The evolution of genes: The chicken preproinsulin gene. Cell 20:555–566

Rabbitts TH, Forster A, Milstein CP (1981) Human immunoglobulin heavy chain genes: Evolutionary comparisons of Cμ, Cδ and Cγ genes and associated switch sequences. Nucleic Acids Res 9:4509–4524

Saccone C, Cantatore P, Gadaleta G, Gallerani R, Lanave C, Pepe G, Kroon AM (1981) The nucleotide sequence of the large ribosomal RNA gene and the adjacent tRNA genes from rat *mitochondria*. Nucleic Acids Res 9:4139–4148

Seeberg PH, Shine J, Martial JA, Baxter JD, Goodman HM (1977) Nucleotide sequence and amplification in bacteria of structural gene for rat growth hormone. Nature 270:486–494

Sheppard HW, Gutman GA (1981) *Allelic forms of rat κ chain* genes: Evidence for strong selection at the level of nucleotide sequence. Proc Natl Acad Sci USA 78:7064–7068

Shine J, Seeberg PH, Martial JA, Baxter JD, Goodman HM (1977) Construction and analysis of recombinant DNA for human chorionic somatomammotropin. Nature 270:494–499

Takahata N, Kimura M (1981) A model of evolutionary base substitutions and its applications with special reference to rapid change of pseudo genes. Genetics 98:641–657

Van Ooen A, Van Den Berg J, Mantei N, Weissmann C (1979) Comparison of total sequence of a cloned rabbit β-globin gene and its flanking regions with a homologous mouse sequence. Science 206:337–344

Vanin EF, Goldberg GI, Tucker PW, Smithies O (1980) A mouse α-globin-related pseudogene lacking intervening sequences. Nature 286:222–226

Wilson AC, Carlson SS, White TJ (1977) Biochemical evolution. Ann Rev Biochem 46:573–639

Wilson JT, Wilson JB, Reddy VB, Covallesco C, Ghosh PK, Riek JK, Forget BG, Weissman SM (1980) Nucleotide sequence of the coding portion of human α globin messenger RNA. J Biol Chem 255:2807–2815

Yarmush ML, Sogn JA, Kindt TJ (1979) Latent allotypes: A window to a genetic enigma. Ann Immunol Inst Pasteur 130C:143–156

Received March 29, 1982 / Revised June 15, 1982

Note Added in Proof

The rapid rate of synonymous substitution in mtDNA genes has recently been reported by Brown et al. (1982).