Proc. Natl. Acad. Sci. USA
Vol. 78, No. 9, pp. 5739–5743, September 1981
Genetics

Extraordinarily high evolutionary rate of pseudogenes: Evidence for the presence of selective pressure against changes between synonymous codons

(nucleotide sequence comparison/functional constraint/bias in code word usage/neutral theory)

TAKASHI MIYATA AND HIDENORI HAYASHIDA

Department of Biology, Faculty of Science, Kyushu University, Fukuoka 812, Japan

Communicated by Motoo Kimura, May 7, 1981

ABSTRACT **Comparisons of nucleotide sequences of several pseudogenes described to date, including α- and β-globin and immunoglobulin κ-type variable domain pseudogenes, with those of functional counterparts revealed that pseudogenes accumulate mutations at an extremely high rate uniformly over their entirety. It is remarkable that the evolutionary rate exceeds the rate of changes between synonymous codons, the highest known rate, in functional genes. Because no pseudogenes appear to function, this result strongly supports the neutral theory. In addition this result apparently indicates the presence of selective pressure against changes between synonymous codons in functional genes. Close examinations of codon utilization patterns in pseudogenes and functional genes revealed a significant correlation between the rate of changes at synonymous codon sites and the strength of bias in code word usage. This implies that even synonymous codon changes are not completely free from selective pressure but are constrained in part, although presumably weakly, depending on the degree of bias in code word usage. We also reexamined alignment between mouse *βh3* (pseudogene) and *βmaj* sequences and found a unique structure of the *βh3* that is homologous in sequence to the *βmaj* gene overall but contains a long deletion (about 150 base pairs) in the middle of the gene.**

Pseudogene is defined to be a region of DNA that shows significant homology in sequence to a functional gene but has several mutational changes that render it unable to produce a functional product (1). Since the first report of 5S RNA pseudogene in the 5S DNA repeat unit of *Xenopus laevis* oocytes (2), several pseudogenes have been identified in other eukaryotic gene clusters, including mammalian α- and β-globin gene families (1, 3–9) and the human immunoglobulin κ-type variable domain gene cluster (10). Recently, more pseudogenes have been identified in *Euglena gracilis* chloroplast ribosomal RNA transcription units (11) and the mouse 4.8S small nuclear RNA gene family (Y. Ohshima, N. Okada, T. Tani, Y. Itoh, and M. Itoh, personal communication), which apparently are unable to encode functional RNAs. Thus, the occurrence of pseudogenes appears to be a common feature of eukaryotic gene clusters.

A pseudogene is considered to be a naturally occurring mutant gene whose sequence resembles sequences of functional genes. Comparison of a pseudogene with its functional counterpart would therefore provide much insight into mechanism by which genes evolve. A comparison of the DNA sequence of mouse α-globin pseudogene to the sequences of normally functioning α-globin genes from mouse and rabbit has shown a remarkable feature of the pseudogene evolution (12): The loss of functions enables it to escape the selective pressures that operate to preserve the sequence almost completely throughout the entire region, thereby allowing accumulation of mutations at an extremely high rate, 1.9 times as large as the rate of changes between synonymous codons in functional genes which is the most rapidly evolving component so far examined (for review, see ref. 13). Kimura also obtained a similar result (14). This characteristic feature is well understandable from the viewpoint of neutral theory (15–17). In addition, these results apparently indicate the presence of some unknown functional constraint against changes between synonymous codons in functional genes (12).

In this report, we confirm the above results on the basis of recently established pseudogene sequence data from a much wider variety of sources. Furthermore, we show that there is a significant correlation between the rate of evolutionary changes between synonymous codons and the strength of bias in synonymous codon usage. This implies that, even at the synonymous codon site, nucleotide changes are not completely free from selective pressure but are constrained to some extent depending on the degree of bias in code word usage.

METHODS

Sequence Alignment. The procedure was described elsewhere (18).

Calculation of Sequence Difference (K). The procedures have been described (18, 19). For each pair of genes, the sequence difference or simply "difference," defined as the number of mismatches per nucleotide site, was calculated at every functional or structural block such as intron, exon, 5'- and 3'-noncoding regions, etc. (13). For coding regions, we carried out calculations for the two types of difference, synonymous difference (K_S) and amino acid difference (K_A). The K_S (K_A) is defined as the number of synonymous (amino acid) substitutions relative to that of the synonymous (amino acid) sites, the sum of fractions of nucleotide sites leading to synonymous (amino acid) change by single nucleotide replacement per nucleotide position of codon (19).

RESULTS AND DISCUSSION

Mouse β-Globin Pseudogene *βh3* Contains a Long Deletion (About 150 Nucleotides) in the Middle of the Sequence. Jahn *et al.* (5) described a mouse β-globin pseudogene *βh3* which is closely related to adult β-globin genes in sequence beyond codon position 75. The second intron is present at a precise position. However, short deletions and insertions within the coding sequence put the amino acid sequence out of phase. In addition, the alignment by Jahn *et al.* demonstrates the loss of continuity in the homology between the *βh3* and an adult β-globin gene sequence at codon position 75. We found an additional homology region between the two sequences before the

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.

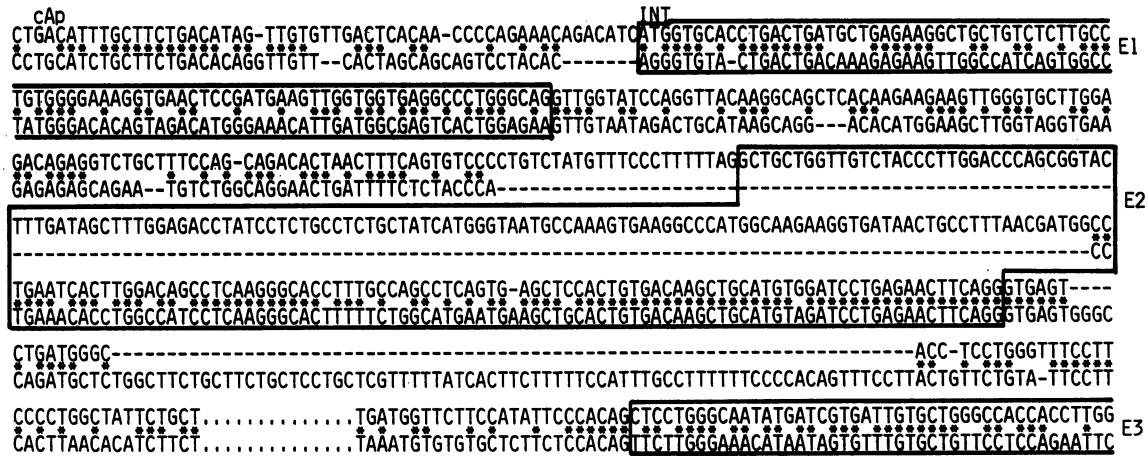

FIG. 1. Alignment between mouse globin βmaj (upper line) (20) and βh3 (lower line) (5) sequences. The coding sequence is enclosed. *, Nucleotides that are identical between the two; -, gap; INT and cAp, translation initiation and capping sites, respectively; E1, E2 and E3, first, second, and third exons, respectively.

position. The revised alignment obtained by our method is shown in Fig. 1. Clearly, βh3 contains a sequence corresponding to all regions of the adult β-globin gene. It is remarkable that this gene has deleted from it a long stretch corresponding to the region from around the last 20 nucleotides of the first intron to the middle of the second exon. It therefore appears to lack the putative splicing signals (21) for the first intron so that no transcripts produced by this gene could be spliced correctly. Furthermore, it lacks the initiator ATG, changed by mutation, to AGG. A similar change is found in the human α-globin pseudogene ψα1 (1). Due to lack of important signals essential to globin function together with the presence of several deletions and insertions within the coding sequence, βh3 apparently cannot produce a functional product.

Extraordinarily High Evolutionary Rate of Pseudogene. Pseudogenes described to date, including α- and β-globin pseudogenes and immunoglobulin κ-type variable domain pseudogene, are closely related in sequence to their respective functional counterparts, suggesting that they were derived from common ancestors by gene duplication (1, 3–6, 10). To reveal the different modes of evolution between a pseudogene ($A_{pseudo.}$) and its functional counterpart ($A_{funct.}$) from a comparison of their temporal sequences, it may be effective to introduce a third homologous gene ($B_{funct.}$), the common ancestor of the three being far remote from that of the $A_{funct.}$ and the $A_{pseudo.}$ (Fig. 2). Since the divergence time of a gene pair $A_{pseudo.}/B_{funct.}$ is the same as that of another pair $A_{funct.}/B_{funct.}$, the sequence difference of the former [designated as $K(A_{pseudo.}/B_{funct.})$] can be compared directly with that of the latter [i.e., $K(A_{funct.}/B_{funct.})$] without using paleontological data which is often subject to considerable uncertainty. Because both the pairs contain the $B_{funct.}$ in common, the difference of values between $K(A_{pseudo.}/B_{funct.})$ and $K(A_{funct.}/B_{funct.})$ is clearly responsible for the difference in the numbers of mutations accumulated in the respective lines to the temporal genes $A_{pseudo.}$ and $A_{funct.}$—that is, the difference of evolutionary rates between the pseudogene and the functional gene.

According to the above idea, we carried out analysis for the evolutionary divergence of three distinct types of pseudogenes, including α- and β-globin pseudogenes from mouse and immunoglobulin κ-type variable domain pseudogene from human (Table 1). An index $R = K(A_{pseudo.}/B_{funct.}) \div K(A_{funct.}/B_{funct.})$, reflecting the difference between evolutionary rates of the pseudogene and its functional counterpart, was also introduced. A characteristic pattern of nucleotide substitutions in several mammalian genes, including α- and β-globin genes, insulin genes, and growth hormone genes, has been found (13): with respect to the extent of sequence divergence, each gene is divided into two distinct blocks, the rapidly evolving blocks including the synonymous codon sites, intron, and 5′ portion of 3′ noncoding region, and the slowly evolving blocks including the 5′ noncoding region, 3′ portion of the 3′ noncoding region, and amino acid alternating sites, where the sequences are strongly conserved due to several functional requirements.

Immunoglobulin variable domain genes, however, show a slightly different pattern in sequence variation (unpublished data). The 5′ noncoding region shows considerable sequence divergence, the value of K being almost comparable to that of K_S. This may be partially due to unusually large nucleotide-length containing a less-constrained stretch in most of this region. The 3′ flanking sequence shows strong sequence preservation due to a requirement for V-J recombination characteristic for immunoglobulin genes (22–25). Table 1 also contains the same classification for functional blocks. It is remarkable that the *R* values in Table 1 are always larger than unity, even in the rapidly evolving regions. Furthermore, the *R* values are always larger in strongly constrained blocks than in less-constrained blocks. These results imply that the loss of function enables the pseudogenes to escape several selective pressures almost completely throughout the entire region, thereby allowing accumulation of mutations uniformly over all the regions at a rate

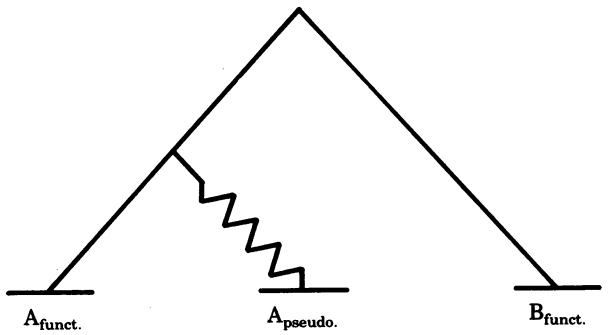

FIG. 2. Schematic representation of phylogenetic relationship among three genes. $A_{pseudo.}$ and $A_{funct.}$, a pseudogene and its functional counterpart from the same species; $B_{funct.}$, a functional gene from different species, B. According to the phylogenetic relationship, the time since divergence between $A_{pseudo.}$ and $B_{funct.}$ is the same as that between $A_{funct.}$ and $B_{funct.}$

Table 1. Comparison of nucleotide sequence of pseudogene with that of functional gene

Functional block	Sequence difference (K) per site K(I)	K(II)	R	d.c.
α-Globin gene	*Mψα/Rα*	*Mα/Rα*		
Amino acid site	0.225	0.105	2.14	S
5′ Noncoding	0.273	0.212	1.29	S
	(0.351)	(0.297)	(1.18)	
3′ Noncoding:				
3′ Portion	0.344	0.219	1.57	S
	(0.344)	(0.219)	(1.57)	
5′ Portion	0.500	0.463	1.08	W
	(0.510)	(0.473)	(1.08)	
Synonymous site	0.530	0.486	1.09	W
β-Globin gene*	*Mψβ/Rβ*	*Mβ/Rβ*		
Amino acid site	0.244	0.126	1.94	S
5′ Noncoding	0.362	0.236	1.53	S
	(0.474)	(0.250)	(1.90)	
Intron†	0.448	0.319	1.40	W
	(0.547)	(0.415)	(1.32)	
Synonymous site	0.427	0.342	1.25	W
Ig κ-type V gene	*Hψκ100/Mκ2*	*Hκ101/Mκ2*		
Amino acid site	0.281	0.202	1.39	S
5′ Flanking	0.355	0.339	1.05	W
	(0.365)	(0.349)	(1.05)	
Intron	0.396	0.322	1.23	W
	(0.464)	(0.360)	(1.29)	
3′ Flanking	0.444	0.259	1.71	S
	(0.444)	(0.259)	(1.71)	
Synonymous site	0.449	0.374	1.20	W

The pseudogenes include several deletions, insertions, or both relative to the functional counterparts. These alternations result in frameshifts. When more than one gap site was found within nucleotide positions of each codon, the corresponding codon was excluded from calculations of K_S and K_A. Two methods that differ in treating gaps were applied for calculating K in noncoding regions. When gaps were found in one of the sequences being compared, the corresponding sites were excluded from comparison (method 1). In method 2, a gap was counted as a "mismatch" (results are given in parenthesis). Nucleotide sites with more than 10 consecutive gaps were ignored in the calculation for both methods. K(I) and K(II), sequence differences for pair I ($A_{pseudo.}/B_{funct.}$) and pair II ($A_{funct.}/B_{funct.}$), respectively (see Fig. 2); $R = K(I)/K(II)$; d.c., degree of constraint; S, strong constraint; W, weak constraint (for the classification, see text); *Mψα* and *Mψβ*, mouse α-globin ψα30.5 and β-globin βh3 pseudogenes, respectively; *Mα* and *Rα*, mouse and rabbit functional adult α-globin genes, respectively; *Mβ* and *Rβ*, adult β-globin genes from mouse (*βmaj*) and rabbit, respectively; *Hψκ100* and *Hκ101*, human κ-type pseudogene and functional gene, respectively; *Mκ2*, *Mκ41* and *Mκ21*, κ-type functional genes *κ2*, *κ41*, and *κ21*, respectively, from mouse. For references for these sequences, see footnote of Table 2.

* Sequences of the region shown in Fig. 1 were compared among *Mψβ*, *Mβ*, and *Rβ*.

† Value for short intron.

that is extraordinary, being higher than the rate of changes between synonymous codons in functioning genes which is the most rapidly evolving component described to date (for references, see ref. 13). These results also confirm a previous argument (12) on pseudogene evolution.

Although the *R* values in Table 1 are consistently larger than unity for all the cases examined, there is still a possibility that this is only a result of statistical fluctuation. To avoid such effect, we extended our analysis by using α- and β-globin genes from the chicken as the $B_{funct.}$ and averaging the K_S and K_A for all pairs between mammalian genes and a chicken gene. For the case of the immunoglobulin gene, two more genes from the mouse were introduced for the $B_{funct.}$, one of which, *κ21*, is distantly related in sequence to *Hκ101* from man.

Table 2 shows the *R* values of the synonymous sites and amino acid sites for five pseudogene cases. The values of R are always larger than unity except for a case of human *ψα1*, in which the value of the synonymous sites is slightly lower than unity. Furthermore, the *R* values are larger in amino acid sites than in synonymous sites, without exception. These results are consistent with the above argument. From the results of Tables 1 and 2 in addition to those of other workers (12, 14, 34), it is clear that pseudogenes evolve at a higher rate than do synonymous substitutions of functional genes but not at an equal rate (1, 6). In previous work (12), the evolutionary rate of mouse α-globin pseudogene *ψα30.5* was estimated to be 12.6×10^{-9} per site per yr, which corresponds to 1.9 times as large as the rate of synonymous substitutions of functional genes, although one should wait for a reliable estimate for elucidation of more pseudogene sequences among which orthogonal comparisons (35) are possible.

Comparison of Codon Usage Patterns Between Pseudogenes and Functional Genes. The fact that the evolutionary rate of pseudogenes is much higher than that of changes between synonymous codons in functional genes indicates that, in functional genes, even the synonymous codon sites, the most rapidly evolving component, are not completely free from selective pressure but are subject to constraints against nucleotide changes due to requirements for some functions. Nonrandom use of degenerate codons found in many genes (e.g., see ref. 36) may be a candidate for such selective pressure by which the rate of changes between synonymous codons is reduced. Several arguments have already been noted on the primary factors determining the bias of codon utilization (for review, see refs. 37 and 38). Although the functional significance of specific codon utilization is still not understood fully, it might be related to selective pressure against changes between synonymous codons, thereby reducing the evolutionary rate to some extent (12, 13, 19). Conversely, if this is really so, then it is expected that synonymous codons are found to be used more uniformly in pseudogenes than in functional counterparts.

To elucidate the relationship between the degree of bias in

Table 2. Ratio of sequence difference (K) between pseudogene and functional gene to that between two functional genes

	Synonymous site	Amino acid site
β-Globin gene		
K(*Rψβ/Cβ*):K (mammal *β/Cβ*)	1.12	1.27
K(*Mψβ/Cβ*):K (mammal *β/Cβ*)*	1.19	1.33
α-Globin gene		
K(*Hψα/Cα*):K (mammal *α/Cα*)	0.99	1.61
K(*Mψα/Cα*):K (mammal *α/Cα*)	1.06	1.33
Ig κ-type V gene		
K(*Hψκ100/Mκ2*):K(*Hκ101/Mκ2*)	1.20	1.39
K(*Hψκ100/Mκ41*):K(*Hκ101/Mκ41*)	1.08	1.51
K(*Hψκ100/Mκ21*):K(*Hκ101/Mκ21*)	1.03	1.21

K(mammal *β/Cβ*), average of K(*Hβ/Cβ*), K(*Rβ/Cβ*), and K(*Mβ/Cβ*); K(mammal *α/Cα*), average of K(*Hα/Cα*), K(*Rα/Cα*), and K(*Mα/Cα*). *Hβ*, *Rβ*, *Mβ*, and *Cβ* are functional β-globin genes from man (26), rabbit (27), mouse (*βmaj*) (20), and chicken (28), respectively. *Hα*, *Rα*, *Mα*, and *Cα* are functional α-globin genes from man (α2) (1), rabbit (29), mouse (30), and chicken (31), respectively. *Rψβ*, rabbit pseudogene ψβ2(6); *Mψβ*, mouse pseudogene *βh3*(5); *Hψα*, human pseudogene *ψα1*(1); *Mψα*, mouse pseudogene *ψα30.5*(4); *Hψκ100*, human pseudogene *HK100* (10); *Hκ101*, human functional gene *HK101* (10); *Mκ2*, *Mκ41*, and *Mκ21*, mouse functional genes *κ2* (32), *κ41* (24), and *κ21* (33), respectively.

* Boxed regions in Fig. 1 were compared among the four sequences.

the use of synonymous codon and the rate of changes between them, we introduce an index, f, which is defined as the frequency of a degenerate codon group ending with a particular base divided by the expected frequency if all the degenerate codons were used uniformly. By definition, it is apparent that the mean of f_U (f of U-ending codons), f_C, f_A, and f_G equals unity. The standard deviation σ_f of the fs may be a good measure representing the degree of bias in codon usage. σ_f is expected to be larger in heavily biased genes than in less-biased genes.

Table 3 shows the values of f and σ_f for α- and β-globin genes and immunoglobulin variable domain genes from mammalian species together with their pseudogene counterparts. It is remarkable that the values of σ_f are consistently smaller in pseudogenes than in their respective functional counterparts. That is, in pseudogenes, there is an appreciable tendency to use synonymous codons more uniformly than in functional genes. Functional α- and β-globin genes avoid using A in the third position of the codon, whereas a significant increase is found in the values of f_A for all the globin pseudogenes examined. Functional α-globin genes are distinguished from β-globin genes in that they show a strong preference for the use of C-ending codons, whereas the values of f_C are decreased considerably in the respective pseudogenes. Immunoglobulin variable domain genes prefer to use all the codons more uniformly than do globin genes. Even in such less-biased genes, the pseudogene counterpart still appears to show a decreased value for σ_f. These results together with the fact that pseudogenes evolve at a higher rate than do changes between synonymous codons in functional genes support an argument that nonrandom use of synonymous codons affects the rate of changes between them.

Table 3. Relative frequency (f) of degenerate codons used in α-globin, β-globin, and immunoglobulin κ-type variable domain genes

	f_U	f_C	f_A	f_G	σ_f
α-Globin:					
Functional gene					
Human ($\alpha 2$)	0.41	2.12	0.09	1.39	0.80
Rabbit	0.40	2.16	0.14	1.30	0.79
Mouse ($\alpha 1$)	1.01	1.76	0.29	0.95	0.52
Chicken	0.60	2.20	0.26	0.94	0.73
Pseudogene					
Human ($\psi\alpha 1$)	0.81	1.56	0.33	1.30	0.47
Mouse ($\psi\alpha 30.5$)	1.06	1.63	0.44	0.88	0.43
β-Globin:					
Functional gene					
Human	1.15	1.26	0.22	1.37	0.46
Rabbit	1.18	1.12	0.25	1.46	0.45
Mouse (βmaj)	1.13	1.49	0.23	1.16	0.47
Chicken	0.60	2.07	0.17	1.16	0.71
Pseudogene					
Rabbit ($\psi\beta 2$)	1.16	1.19	0.44	1.22	0.33
Mouse ($\beta h3$)	0.81	1.28	0.64	1.28	0.28
Ig κ					
Functional gene					
Human (*HK101*)	1.07	1.29	0.75	0.89	0.20
Mouse ($\kappa 2$)	1.42	0.76	0.98	0.84	0.25
Mouse ($\kappa 41$)	1.49	1.10	0.78	0.64	0.33
Mouse ($\kappa 21$)	1.36	1.03	0.77	0.84	0.23
Pseudogene					
Human (*HK100*)	1.10	1.21	0.77	0.92	0.17

f, Frequency of degenerate codons ending with a specific base, divided by the expected frequency if all the degenerate codons were used uniformly. For human α-globin gene, for example, U-, C-, A-, and G-ending codons appear 14, 73, 3, and 48 times, excluding nondegenerate codons. Thus, f_U is $14/[(14 + 73 + 3 + 48)/4] = 0.41$. σ_f, standard deviation of f_U, f_C, f_A, and f_G. For sources of sequence data used, see legend of Table 2.

Table 3 also reveals an interesting feature of codon utilization pattern in functional α- and β-globin genes. With respect to the value of σ_f, α-globin genes from mammals and a bird are clearly classified into two groups, a heavily biased group (αII) including α-globin genes from man, rabbit, and chicken and a less-biased group (αI) with mouse α-globin gene. A similar classification is possible for β-globin genes. Human, rabbit, and mouse β-globin genes (βI) have different values of σ_f from chicken β-globin gene (β-II). The αII and βII groups are similar in codon utilization pattern and σ_f value (0.7–0.8) but they are quite distinct from αI and βI which again are almost identical in σ_f value (~0.5). This indicates that, although globin genes could be classified into two groups, the classification does not necessarily correspond to gene types (i.e., α vs. β) but depends rather on species.

Further Evidence for the Correlation Between Constraint Against Synonymous Changes and Degree of Bias in Synonymous Codon Usage. More supporting evidence for the possibility that bias in synonymous codon usage is responsible for a selective pressure against changes between synonymous codons comes from comparison of α- and β-globin gene sequences. The above classification for globin genes with respect to σ_f allows us to compare the K_S values among three categories, heavily biased pairs (i.e., αII/βII), moderately biased pairs (αI/βII and αII/βI), and less-biased pairs (αI/βI). It should be noted that, because α- and β-globin gene divergence is far remote from the divergence of mammals and birds (39), all the pairs have the same divergence time.

The results are shown in Table 4. A significant correlation is found between K_S and the σ_f. The heavily biased pairs show the lowest K_S values (the mean ±SD within the category is 0.51 ± 0.02), and the less-biased pairs show the highest (0.66 ± 0.00). For moderately biased pairs, the K_S values (0.59 ± 0.02) lie between the other two. This result together with the results that pseudogenes have a higher evolutionary rate and lower σ_f values than do the respective values of functional genes clearly indicates that, in functional genes, even the synonymous sites could not accumulate mutations freely but are subject to selective pressure which relates in part to bias in synonymous codon utilization. This type of selective pressure, however, must be weak compared with another type of selective pressure operating on the amino acid sites to preserve correct tertiary structures of encoded protein molecules because synonymous sites are known to evolve at much higher rates than amino acid sites (for references, see ref. 13).

Evolutionary Rate of Pseudogene May Correspond To a Limiting Rate Predicted by Neutral Theory. The neutral theory stresses that most evolutionary changes that result in nu-

Table 4. Correlation between sequence difference (K_S) and degree of bias in code word usage in α- and β-globin genes

			Sequence difference (K_S)			
			βI			βII
			$H\beta$ (0.46)	$R\beta$ (0.45)	$M\beta$ (0.47)	$C\beta$ (0.71)
αI	$M\alpha$	(0.52)	0.663	0.661	0.661	0.598
	$C\alpha$	(0.73)	0.617	0.633	0.601	0.537
αII	$R\alpha$	(0.79)	0.581	0.575	0.598	0.484
	$H\alpha$	(0.80)	0.569	0.553	0.563	0.505

αI and αII, Classification of α-globin genes from man ($H\alpha$), rabbit ($R\alpha$), mouse ($M\alpha$), and chicken ($C\alpha$) with respect to σ_fs (the values enclosed in parentheses). A similar classification was also made for β-globin genes. For sources of sequence data, see legend of Table 2.

cleotide substitutions in functional genes are caused not by selection but by random drift of mutant genes that are selectively equivalent or neutral (15–17). The probability that a mutational change is selectively neutral depends strongly on functional constraints. That is, the weaker the constraint, the larger the probability of random change being selectively neutral (16, 17). According to this view, functionally less- important genes or parts of genes are expected to evolve faster than are more important ones. Because no pseudogenes appear to have a function, the result that they evolve at a highest known rate is completely consistent with this argument. Furthermore, the neutral theory reveals an remarkable feature on gene evolution. Kimura (17) has already pointed out that a gene evolves at an upper rate which is almost equivalent to the rate of mutations occurring on the DNA sequence, when its functional constraint disappears completely. It seems likely that the evolutionary rate of pseudogenes is corresponding to this limiting rate (12). No gene could evolve faster than this rate. Conversely, if genes or parts of genes evolve at lower rate than the limiting rate, the neutral theory suggests a possibility that their nucleotide sequences are subject to constraints related to known or unknown biological functions. This allows more detailed pictures about mechanisms by which genes evolve. Under this viewpoint, we indeed have shown that synonymous substitutions in functional genes are not completely free from constraints but are subject to constraint which is related to bias in synonymous codon usage to some extent, although presumably weak compared with selection pressure operating on protein tertiary structure.

Note Added in Proof. We have just learned that Edgell *et al.* (40) obtained an alignment similar to that in Fig. 1.

We thank Dr. M. Kimura for stimulating discussions and continuous encouragement throughout this work. Thanks are also due to Prof. H. Matsuda and Dr. T. Yasunaga for helpful discussions and comments. This work was supported in part by grants from the Ministry of Education, Science and Culture of Japan.

1. Proundfoot, N. J. & Maniatis, T. (1980) *Cell* **21**, 537–544.
2. Jacq, C., Miller, J. R. & Brownlee, G. G. (1977) *Cell* **12**, 109–120.
3. Nishioka, Y. & Leder, P. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 2806–2809.
4. Vanin, E. F., Goldberg, G. I., Tucker, P. W. & Smithies, O. (1980) *Nature (London)* **286**, 222–226.
5. Jahn, C. L., Huchison, C. A. III, Philips, S. J., Weaver, S., Haigwood, N. L., Voliva, C. F. & Edgell, M. H. (1980) *Cell* **21**, 159–168.
6. Lacy, E. & Maniatis, T. (1980) *Cell* **21**, 545–553.
7. Haynes, J. R., Rosteck, P., Jr., Schon, E. A., Gallagher, P. M., Burks, D. J., Smith, K. & Lingrel, J. B. (1980) *J. Biol. Chem.* **255**, 6355–6367.
8. Leder, P., Hansen, J. N., Konkel, D., Leder, A., Nishioka, Y. & Talkington, C. (1980) *Science* **209**, 1336–1342.
9. Fritsch, E. F., Lawn, R. M. & Maniatis, T. (1980) *Cell* **19**, 959–972.
10. Bentley, D. L. & Rabbitts, T. H. (1980) *Nature (London)* **288**, 730–733.
11. Orozco, E. M., Jr., Rushlow, K. E., Dodd, J. R. & Hallick, R. B. (1980) *J. Biol. Chem.* **255**, 10997–11003.
12. Miyata, T. & Yasunaga, T. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 450–453.
13. Miyata, T., Yasunaga, T. & Nishida, T. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 7328–7332.
14. Kimura, M. (1980) *J. Mol. Evol.* **16**, 111–120.
15. Kimura, M. (1968) *Nature (London)* **217**, 624–626.
16. Kimura, M. & Ohta, T. (1974) *Proc. Natl. Acad. Sci. USA* **71**, 2848–2852.
17. Kimura, M. (1977) *Nature (London)* **267**, 275–276.
18. Miyata, T., Yasunaga, T., Yamawaki-Kataoka, Y., Obata, M. & Honjo, T. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 2143–2147.
19. Miyata, T. & Yasunaga, T. (1980) *J. Mol. Evol.* **16**, 23–36.
20. Konkel, D. A., Maizel, V., Jr. & Leder, P. (1979) *Cell* **18**, 865–873.
21. Lerner, M., Boyle, J., Mount, S., Wolin, S. & Steiz, J. (1980) *Nature (London)* **283**, 220–224.
22. Sakano, H., Huppi, K., Heinrich, G. & Tonegawa, S. (1979) *Nature (London)* **280**, 288–294.
23. Max, E. E., Seidman, J. G. & Leder, P. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 3450–3454.
24. Seidman, J. G., Max, E. E. & Leder, P. (1979) *Nature (London)* **280**, 370–375.
25. Sakano, H., Maki, R., Kurosawa, Y., Roeder, W. & Tonegawa, S. (1980) *Nature (London)* **286**, 676–683.
26. Lawn, R. M., Efstratiadis, A., O'Connell, C. & Maniatis, T. (1980) *Cell* **21**, 647–651.
27. Van Ooen, A., Van Den Berg, J., Mantei, N. & Weissmann, C. (1979) *Science* **206**, 337–344.
28. Richards, R. I., Shine, J., Ullrich, A., Wells, J. R. E. & Goodmann, H. M. (1979) *Nucleic Acids Res.* **7**, 1137–1146.
29. Heindell, H. C., Liu, A., Paddock, G. V., Studnicka, G. M. & Salser, W. A. (1978) *Cell* **15**, 43–54.
30. Nishioka, Y. & Leder, P. (1979) *Cell* **18**, 875–882.
31. Deacon, N. J., Shine, J. & Naora, H. (1980) *Nucleic Acids Res.* **8**, 1187–1199.
32. Nishioka, Y. & Leder, P. (1979) *J. Biol. Chem.* **255**, 3691–3694.
33. Seidman, J. G. & Leder, P. (1980) *Nature (London)* **286**, 779–783.
34. Takahata, N. & Kimura, M. (1981) *Genetics*, in press.
35. Fitch, W. M. & Margoliash, E. (1970) *Evol. Biol.* **4**, 76.
36. Grantham, R., Gautier, C. & Gouy, M. (1980) *Nucleic Acids Res.* **8**, 1893–1912.
37. Hasegawa, M., Yasunaga, T. & Miyata, T. (1979) *Nucleic Acids Res.* **7**, 2073–2079.
38. Miyata, T., Hayashida, H., Yasunaga, T. & Hasegawa, M. (1979) *Nucleic Acids Res.* **7**, 2431–2438.
39. Dayhoff, M. O. (1978) *Atlas of Protein Sequence and Structure*, ed. Dayhoff, M. O. (National Biomedical Foundation, Bethesda, MD), Vol. 5, Suppl. 3, pp. 229–249.
40. Edgell, M. H., Weaver, S., Jahn, C. L., Padgett, R. W., Phillips, S. J., Voliva, C. F., Comer, M. B., Hardies, S. C., Haigwood, N. L., Langley, C. H., Racine, R. R. & Hatchson, C. A., III (1981) *Organization and Expression of Globin Genes*, eds. Stamatoyannopoulos, G. & Nienhuis, A. (Liss, New York), pp. 69–88.