The Rate of Synonymous Substitution in Enterobacterial Genes Is Inversely Related to Codon Usage Bias[1]

Paul M. Sharp[2] *and Wen-Hsiung Li*
Center for Demographic and Population Genetics, University of Texas

Gene sequences from *Escherichia coli, Salmonella typhimurium,* and other members of the Enterobacteriaceae show a negative correlation between the degree of synonymous-codon usage bias and the rate of nucleotide substitution at synonymous sites. In particular, very highly expressed genes have very biased codon usage and accumulate synonymous substitutions very slowly. In contrast, there is little correlation between the degree of codon bias and the rate of protein evolution. It is concluded that both the rate of synonymous substitution and the degree of codon usage bias largely reflect the intensity of selection at the translational level. Because of the high variability among genes in rates of synonymous substitution, separate molecular clocks of synonymous substitution might be required for different genes.

Introduction

Comparative studies of protein sequence data have provided ample evidence that the rate of molecular evolution is inversely related to the degree of selective constraint on the sequences or sites involved (Dayhoff 1972; Kimura 1983). Thus, given that there is much variation among proteins in the functional stringency of the precise amino acid sequence, it was not surprising to discover that the rate of nonsynonymous (i.e., amino acid–replacing) nucleotide substitution varies greatly among genes (Miyata et al. 1980; Kimura 1983). On the other hand, it was commonly thought that most synonymous mutations are subject to no strong selective constraints and that therefore the rate of synonymous substitution should not vary much among genes. Analyzing the then available DNA sequence data, Miyata et al. (1980) found that this was indeed the case, and they suggested that synonymous substitutions from different genes could be pooled to serve as a molecular clock. However, subsequent analysis of a larger number of mammalian genes has revealed that the synonymous rate varies considerably among genes (Li et al. 1985*b*). Although there was a positive correlation between the synonymous and nonsynonymous rates in the same gene, this correlation could explain only part of the variation in the synonymous rate (Graur 1985; Li et al. 1985*b*).

On the other hand, a large body of evidence has been gathered that demonstrates that the use of alternative synonymous codons is far from random (Grantham et al. 1981; Ikemura 1985). The pattern of codon usage is species specific (Grantham et al. 1981), but within each species there is considerable variation in the degree of codon usage bias. For example, in *Escherichia coli* and the yeast *Saccharomyces cerevisiae* the bias is much stronger in highly expressed genes than in moderately and lowly

1. Key words: synonymous codon usage, synonymous substitution rate, molecular clock, translational efficiency, *Escherichia coli, Salmonella typhimurium.*

2. Permanent address: Department of Genetics, Trinity College, Dublin 2, Ireland.

Address for correspondence and reprints: Dr. Paul M. Sharp, Department of Genetics, Trinity College, Dublin 2, Ireland.

Mol. Biol. Evol. 4(3):222–230. 1987.
© 1987 by The University of Chicago. All rights reserved.
0737-4038/87/0403-0002$02.00

expressed genes (Bennetzen and Hall 1982; Gouy and Gautier 1982; Sharp and Li 1986; Sharp et al. 1986). Selective differences among synonymous codons have been hypothesized, largely related to the relative abundance of isoaccepting tRNAs and their interaction with different codons (Ikemura 1981, 1985; Grosjean and Fiers 1982). In *E. coli* and yeast, very highly expressed genes use almost exclusively the small group of codons that are presumed to be optimal for efficient translation (Ikemura 1981, 1985; Bennetzen and Hall 1982). Apparently, the large variation in codon usage bias in *E. coli* and yeast genes reflects variation in selective constraints mediated through the process of translation.

Possibly the variation among genes in the rate of synonymous substitution and in the degree of codon usage bias are two aspects of the same phenomenon, i.e., both may reflect the variation among genes in selective constraints on synonymous-codon usage. If this is true, then the rate of synonymous substitution should be inversely related to the degree of codon usage bias. A preliminary study based on a few genes (Ikemura 1985) has suggested that this is true. Here we use recently determined DNA sequence data from *E. coli, S. typhimurium,* and a few related species to demonstrate that a significant negative correlation indeed exists between the rate of synonymous nucleotide substitution and the degree of codon usage bias.

Material and Methods

DNA sequence data are available for 23 pairs of homologous genes from *Escherichia coli* and *Salmonella typhimurium* (table 1) and for several pairs from some related species (table 2). Estimated numbers of synonymous and nonsynonymous nucleotide substitutions between genes were calculated by a method that takes account of both the degree of degeneracy of nucleotide sites and the different rates of transitions and transversions (Li et al. 1985*b*).

The degree of synonymous-codon usage bias was measured by the "codon adaptation index" (CAI; Sharp and Li 1987), which estimates the extent of bias toward codons that are known to be favored in highly expressed genes. Each codon is assigned a relative "adaptiveness" value according to its frequency of use in a species-specific reference set of very highly expressed genes, and the CAI for a gene is then calculated as the geometric-mean relative adaptiveness of its codons. A CAI value of 1.0 indicates that the gene in question contains only optimal codons and thus has the greatest possible degree of bias. A value close to zero indicates extensive use of codons that normally are rare. Examination of very weakly expressed *E. coli* genes suggests (1) that the frequencies of these rare codons are usually even lower than would be expected on the basis of random codon usage (Sharp and Li 1986) and (2) that the observed CAI values for *E. coli* genes are all higher than the CAI value (~0.17) for a sequence in which all 61 sense codons are equiprobable. *Escherichia coli* and *S. typhimurium* exhibit similar patterns of codon bias (Ikemura 1985), presumably because they are so closely related that similar constraints operate on codon usage in the two species. A reference set of 27 very highly expressed *E. coli* genes (mainly ribosomal protein, elongation factor, and outer-membrane-protein genes; see Sharp and Li 1986) was used to calculate CAI values for both the *E. coli* and *S. typhimurium* genes (table 1). For each gene, the CAI values for the two species are quite similar.

Results

Since the divergence time is the same for all orthologous pairs of *Escherichia coli* and *Salmonella typhimurium* genes, we can compare the relative rates of evolution

Table 1
Comparison of Genes: *Escherichia coli* and *Salmonella typhimurium*

Gene	No. of Codons	Mean K_A	Mean ± SE K_S	CAI	
				E. coli	*S. typhimurium*
trpA	267	0.08	1.77 ± 0.33	0.34	0.32
cheY	128	0.01	1.49 ± 0.31	0.30	0.34
trpC	451	0.07	1.39 ± 0.18	0.31	0.31
tar	552	0.12	1.37 ± 0.14	0.32	0.32
araC	280	0.04	1.27 ± 0.17	0.25	0.23
aroA	425	0.07	1.26 ± 0.14	0.36	0.30
pabA	186	0.08	1.24 ± 0.28	0.28	0.29
dnaG	580	0.08	1.18 ± 0.11	0.27	0.28
hisIE	202	0.06	1.11 ± 0.17	0.39	0.38
trpE	519	0.07	1.06 ± 0.10	0.36	0.33
trpD	530	0.02	1.06 ± 0.10	0.34	0.32
trpB	396	0.02	1.03 ± 0.11	0.41	0.36
crp	209	0.00	0.89 ± 0.16	0.49	0.46
orf1[a,b] ...	108	0.04	0.84 ± 0.17	0.35	0.39
ilvY[a]	257	0.01	0.72 ± 0.10	0.34	0.30
metB	385	0.02	0.57 ± 0.06	0.34	0.33
rpoD	612	0.01	0.49 ± 0.05	0.58	0.52
ilvM	86	0.03	0.47 ± 0.12	0.23	0.23
glnA[a]	71	0.07	0.39 ± 0.12	0.64	0.51
ompA ...	345	0.04	0.35 ± 0.05	0.77	0.71
metJ	104	0.01	0.29 ± 0.08	0.41	0.36
rpoB[a]	958	0.01	0.31 ± 0.03	0.63	0.62
rpsU	70	0.00	0.04 ± 0.03	0.73	0.71
5′ *tar*	280	0.20	1.36 ± 0.21	0.31	0.30
3′ *tar*	235	0.01	1.11 ± 0.17	0.34	0.37

NOTE.—Gene sequences were taken from GenBank (1986), except for *cheY* (Stock et al. 1985), *trpC* and *trpD* (Horowitz et al. 1983), *pabA* (Kaplan et al. 1985), *dnaG, rpoD,* and *rpsU* (Erickson et al. 1985), *hisIE* (Chiariotti et al. 1986), *crp* (Cossart et al. 1986), *orf1* (Neuhard et al. 1985), *ilvY* (Wek and Hatfield 1986; R. C. Wek, personal communication), *metB* (P. McWilliam and S. Thompson, personal communication), *ilvM* (Lopes and Lawther 1986), *glnA* (Rocha et al. 1985), *ompA* (Freudl and Cole 1983), *metJ* (Saint-Girons et al. 1984; Urbanowski and Stauffer 1985), and *rpoB* (Sverdlov et al. 1986).

[a] Partial sequence.

[b] Open reading frame upstream of *pyrE.*

in different genes (table 1). Interestingly, there is a large range in synonymous-substitution rates among genes. The two very highly expressed genes—*rpsU* and *ompA*—have both the greatest bias in codon usage and a very low degree of synonymous divergence (table 1). Three other genes with a high codon bias—*rpoB, rpoD,* and *glnA*—also show comparatively low rates of synonymous substitution. Among genes with a low codon bias (i.e., those in which CAI < 0.4), there is considerable variability in the synonymous-substitution rate, K_S; but high K_S values predominate among the longer sequences (i.e., those in which $I > 200$), which have a smaller SE for the estimates of K_S.

K_S has been plotted against the degree of codon bias (fig. 1). There is a significant negative linear correlation (coefficient 0.68, $P < 0.01$) between these two statistics, confirming that genes with a more extreme synonymous-codon bias undergo synonymous substitution at a slower rate. Among the genes with a low codon bias (CAI

Table 2
Comparison of Genes: Species of Enterobacteriaceae

Comparison	CAI[a]	K_S [b]	K_A [b]
A. Species Pairs with a Divergence Time Equal to That between *Escherichia coli* and *Enterobacter aerogenes*			
E. aerogenes ompA gene vs.:			
E. coli	0.77	0.45	0.09
Shigella dysenteriae	0.75	0.51	0.10
Salmonella typhimurium	0.71	0.53	0.09
Klebsiella pneumoniae pabA gene vs.:			
E. coli	0.28	1.77	0.13
S. typhimurium	0.29	1.26	0.12
K. pneumoniae trpA gene vs.:			
E. coli	0.34	1.61	0.08
S. typhimurium	0.32	1.80	0.11
B. Species Pairs with a Divergence Time Equal to That between *E. coli* and *Serratia marcescens*			
S. marcescens lpp gene vs.:			
E. coli	0.85	0.33	0.07
S. marcescens pabA gene vs.:			
E. coli	0.28	1.47	0.16
S. typhimurium	0.29	1.57	0.19
K. pneumoniae	...	1.88	0.17
S. marcescens trpG gene vs.:			
E. coli	0.33	2.56	0.14
S. dysenteriae	0.37	1.74	0.15
S. typhimurium	0.31	1.60	0.14

Note.—References for sequences are as in table 1.
[a] Computed using a reference set of very highly expressed genes from *E. coli*.
[b] K_S and K_A = number of synonymous and nonsynonymous substitutions, respectively, between species.

< 0.4), the two with the lowest K_S values are rather short, and so the low K_S values could have arisen from sampling errors. The *ilvM* gene, which has a K_S value much lower than would be expected on the basis of its CAI value, is a short, previously unrecognized gene sandwiched between *ilvG* and *ilvE* and contains within its coding sequence a promoter for the *ilvE* gene (Lopes and Lawther 1986). The need to conserve this regulatory sequence may serve as an extra constraint reducing K_S in *ilvM*.

Sequence data are available for some genes from species of Enterobacteriaceae other than *E. coli* and *S. typhimurium,* notably *Klebsiella pneumoniae, Enterobacter aerogenes,* and *Serratia marcescens. Klebsiella pneumoniae* and *E. aerogenes* are considered to be more closely related to each other than either is to *E. coli* (Brenner 1984), and so the divergence times between *E. coli* (or *S. typhimurium*) genes and *K. pneumoniae* or *E. aerogenes* genes should be equivalent. *Serratia marcescens* is considered to be more distantly related to all of the above species (Brenner 1984), and so any comparisons between *S. marcescens* and the other species should involve equivalent time scales. Estimates of nucleotide substitution among these species are detailed in table 2. For the genes with low CAI values (i.e., *pabA, trpA*, and *trpG*), the K_S values are generally high—and near the level of saturation of synonymous substitutions. Again the genes with a high synonymous-codon bias (*ompA* and *lpp*) show a much

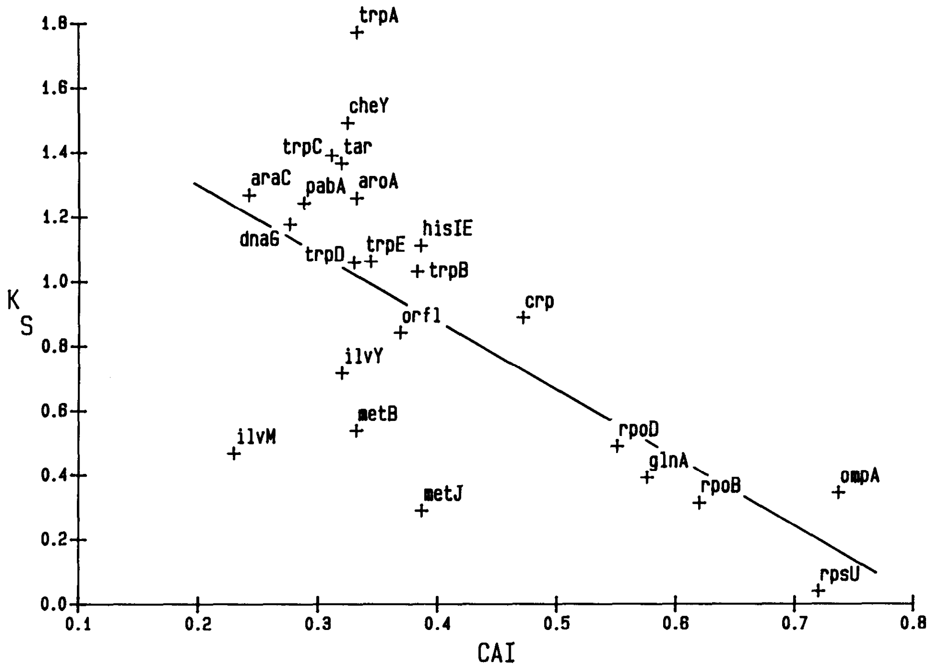

FIG. 1.—Relationship between synonymous-codon usage bias and K_S in *Escherichia coli* and *Salmonella typhimurium*. K_S = Estimated number of nucleotide substitutions per synonymous site; CAI = mean of the CAI values for the two species. For a sequence in which all 61 codons are equiprobable, CAI = ~0.17. The least-squares linear regression of K_S on CAI is drawn; the linear correlation coefficient is 0.68, $P < 0.01$.

lower synonymous divergence. These data are entirely consistent with those in table 1 and figure 1, bearing in mind the longer divergence times.

Discussion

The correlation coefficient computed in figure 1 should be taken with caution because the estimated K_S values may not be reliable when the true values are considerably higher than one (Li et al. 1985*b*). However, a significant correlation between CAI and K_S is observed even if one uses the observed proportion of differences at synonymous sites instead of the K_S value. For simplicity, we have assumed a linear correlation (fig. 1). We tried various transformations (e.g., logs, squares) of the axes, but the correlation coefficients obtained were very similar to one another.

Comparison of the rates of nucleotide substitution at noncoding sites (e.g., pseudogenes), coding but degenerate sites (i.e., sites at which synonymous mutations can occur), and amino acid–determining sites has revealed an inverse relationship between the rate and the degree of functional constraint (Kimura 1983; Li et al. 1985*a*). Here we extend that observation to synonymous sites in different genes that appear to be under different degrees of constraint, as interpreted on the basis of the observed degree of synonymous-codon bias. This confirms the preliminary observations of Ikemura (1985), which were made on the basis of data for only a few genes.

Horizontal transfer of genes between *Escherichia coli* and *Salmonella typhimurium* would produce gene pairs with surprisingly low synonymous divergence for a particular degree of codon bias. In figure 1 a few genes seem to be in this category,

but the two outstanding examples can be explained by the small number of codons examined (*metJ*) and/or extra sequence constraints (*ilvM*), so that interspecific exchange does not appear to be an important confounding factor here.

Among mammalian genes there is a tendency for genes with a high nonsynonymous rate, K_A, to have a high K_S (Graur 1985; Li et al. 1985*b*). Similarly here the correlation of K_A and K_S for the *E. coli–S. typhimurium* comparison is 0.57. Whereas K_A is influenced by selection at the protein level, K_S is presumed to be influenced by selection on synonymous-codon usage, i.e., at the level of translation. The correlation of K_A and K_S indicates that, among the genes studied, those that are highly expressed and tend to have a high CAI and a low K_S also tend to encode conserved proteins. However, a direct relationship between protein sequence constraint and codon bias seems unlikely, since the (negative) correlation of K_A and CAI is only 0.36. The importance of the precise amino acid sequence to protein function—and hence the degree of sequence conservation (and K_A)—can vary along a peptide. In contrast, the synonymous-codon composition is a property of an mRNA as a whole, so that the degree of codon bias (and hence K_S) should be comparatively uniform along a gene. This is dramatically illustrated by the *tar* gene. The amino-terminal half of the *tar*-gene product has diverged more rapidly than any other protein examined here, whereas the majority of the carboxy-terminal half is quite conserved (table 1, bottom). In contrast to the rate of amino acid replacement (reflected in K_A), K_S is very similar in the two halves of the gene (table 1).

It has been suggested that in some cases the level of gene expression is modulated evolutionarily by the selection of rare codons to reduce the rate of translation (see, e.g., Grosjean and Fiers 1982; Konigsberg and Godson 1983). On the basis of this assumption it would be expected that this constraint on codon usage would reduce K_S, just as selection for optimal codons in highly expressed genes reduces the rate. In particular, it has been reported that the *dnaG, lacI, trpR,* and *araC* genes of *E. coli* have an excess of rare codons, a situation that has been explained as a mechanism to maintain low expression (Konigsberg and Godson 1983). We have shown elsewhere (Sharp and Li 1986) that these genes do not have significantly more rare codons than do a large number of other *E. coli* genes expressed at moderate to low levels. From the data presented here (table 1, fig. 1) it can be seen that *dnaG* and *araC* are accumulating synonymous substitutions at a rate typical of genes with a low codon bias. This suggests that the incidence of rare codons, yielding a low CAI, in these genes results from a comparative lack of negative (purifying) selection rather than from the presence of positive selection.

The approximate time (t_{ES}) of divergence of the *E. coli* and *S. typhimurium* lineages has been estimated previously by using rRNA molecular clocks (Hori and Osawa 1978; Ochman and Wilson 1986). Hori and Osawa cited Hori's (1976) data for 5S rRNA, which was used to produce a phylogeny of eukaryotes and prokaryotes. Assuming that the ancestors of man and *Xenopus* diverged 300 Myr ago, they estimated t_{ES} to be 37 ± 26 Myr ago. This assumes that 5S rRNA evolves at the same rate in vertebrate and enterobacterial lineages. Ochman and Wilson (1986), using similarity values (S_{ab}) for 16S rRNA comparisons, estimated t_{ES} to be 110–150 Myr ago. The average number of substitutions per synonymous site between *E. coli* and *S. typhimurium* from the 23 genes examined here is 0.90, and the rate of synonymous substitutions per year (V_S) can be derived for the above estimates of t_{ES}. If t_{ES} = 37 Myr, then $V_S = 12.4 \times 10^{-9}$ substitutions/site/year. This is ~2–3 times higher than the

average rate (4.7×10^{-9}) for 35 mammalian genes (Li et al. 1985*b*). If t_{ES} = 130 Myr ago, then $V_S = 3.5 \times 10^{-9}$, which is somewhat lower than the mean value for mammalian genes. These rate estimates are perhaps surprising. Since bacteria have a short generation time, their number of DNA replications per unit time could be much greater than that of mammals. Then, unless mammals have a comparatively inefficient DNA repair system, the rate of mutation should be much higher in bacteria. In turn, V_S should also be considerably higher in bacteria. Therefore, we conclude *either* (1) that the estimates of t_{ES} cited above are too high *or* (2) that selection against synonymous mutations is more effective in these bacteria than in mammals, perhaps owing to a much larger population size or stronger selective constraints on synonymous-codon usage (Li, accepted).

The present study indicates that V_S in Enterobacteria is influenced by natural selection differentiating between alternative synonymous codons. This is also likely to be true in multicellular organisms, although in the latter the determinants of synonymous-codon usage remain to be elucidated. Selection on synonymous codons is perhaps the most subtle form of natural selection detected; yet, because of the small selective differences involved, it is probably the most pervasive. Since the V_S apparently can vary widely among genes, separate molecular clocks of synonymous substitution should be used for genes with different levels of synonymous-codon usage bias.

Acknowledgments

We are grateful to Ronald C. Wek, Shahla Thompson, and Peter McWilliam for access to their unpublished DNA sequence data and to Robert J. Schwartz for plotting the figure. This work was supported by National Institutes of Health grant GM 30998.

LITERATURE CITED

BENNETZEN, J. L., and B. D. HALL. 1982. Codon selection in yeast. J. Biol. Chem. **257**:3026–3031.

BRENNER, D. J. 1984. Enterobacteriaceae. Pp. 408–420 *in* N. R. KRIEG, ed. Bergey's manual of systematic bacteriology. Vol. **1**. Williams & Wilkins, Baltimore.

CHIARIOTTI, L., P. ALIFANO, M. S. CARLOMAGNO, and C. B. BRUNI. 1986. Nucleotide sequence of the *Escherichia coli hisD* gene and of the *Escherichia coli* and *Salmonella typhimurium hisIE* region. Mol. Gen. Genet. **203**:382–388.

COSSART, P., E. A. GROISMAN, M.-C. SERRE, M. J. CASADABAN, and B. GICQUEL-SANZEY. 1986. *crp* Genes of *Shigella flexneri, Salmonella typhimurium,* and *Escherichia coli.* J. Bacteriol. **167**:639–646.

DAYHOFF, M. O. 1972. Atlas of protein sequence and structure. Vol. **5**. National Biomedical Research Foundation, Silver Spring, Md.

ERICKSON, B. D., Z. F. BURTON, K. W. WATANABE, and R. R. BURGESS. 1985. Nucleotide sequence of the *rpsU-dnaG-rpoD* operon from *Salmonella typhimurium* and a comparison of this sequence with the homologous operon of *Escherichia coli.* Gene **40**:67–78.

FREUDL, R., and S. T. COLE. 1983. Cloning and characterization of the *ompA* gene from *Salmonella typhimurium.* Eur. J. Biochem. **134**:497–502.

GenBank. 1986. Genetic sequence data bank. Release 42.0. Bolt, Beranek & Newman, Cambridge, Mass.

GOUY, M., and C. GAUTIER. 1982. Codon usage in bacteria: correlation with gene expressivity. Nucleic Acids Res. **10**:7055–7074.

GRANTHAM, R., C. GAUTIER, M. GOUY, M. JACOBZONE, and R. MERCIER. 1981. Codon catalog usage is a genome strategy modulated for gene expressivity. Nucleic Acids Res. **9**:r43–r79.

GRAUR, D. 1985. Amino acid composition and the evolutionary rates of protein-coding genes. J. Mol. Evol. **22**:53–63.

GROSJEAN, H., and W. FIERS. 1982. Preferential codon usage in prokaryotic genes: the optimal codon-anticodon interaction energy and the selective codon usage in efficiently expressed genes. Gene **18**:199–209.

HORI, H. 1976. Molecular evolution of 5S RNA. Mol. Gen. Genet. **145**:119–123.

HORI, H., and S. OSAWA. 1978. Evolution of ribosomal proteins in *Enterobacteriaceae.* J. Bacteriol. **133**:1089–1095.

HOROWITZ, H., J. VAN ARSDELL, and T. PLATT. 1983. Nucleotide sequence of the *trpD* and *trpC* genes of *Salmonella typhimurium.* J. Mol. Biol. **169**:775–797.

IKEMURA, T. 1981. Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes. J. Mol. Biol. **146**:1–21.

———. 1985. Codon usage and tRNA content in unicellular and multicellular organisms. Mol. Biol. Evol. **2**:13–34.

KAPLAN, J. B., W. K. MERKEL, and B. P. NICHOLS. 1985. Evolution of glutamine amidotransferase genes: nucleotide sequences of the *pabA* genes from *Salmonella typhimurium, Klebsiella aerogenes* and *Serratia marcescens.* J. Mol. Biol. **183**:327–340.

KIMURA, M. 1983. The neutral theory of molecular evolution. Cambridge University Press, Cambridge.

KONIGSBERG, W., and G. N. GODSON. 1983. Evidence for use of rare codons in the *dnaG* gene and other regulatory genes of *Escherichia coli.* Proc. Natl. Acad. Sci. USA **80**:687–691.

LI, W.-H. Models of nearly neutral mutations with particular implications for nonrandom usage of synonymous codons. J. Mol. Evol., accepted.

LI, W.-H., C.-C. LUO, and C.-I. WU. 1985*a*. Evolution of DNA sequences. Pp. 1–94 *in* R. J. MACINTYRE, ed. Molecular evolutionary genetics. Plenum, New York.

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

LOPES, J. M., and R. P. LAWTHER. 1986. Analysis and comparison of the internal promoter, pE, of the *ilvGMEDA* operons from *Escherichia coli* and *Salmonella typhimurium.* Nucleic Acids Res. **14**:2779–2798.

MIYATA, T., T. YASUNAGA, and T. NISHIDA. 1980. Nucleotide sequence divergence and functional constraint in mRNA evolution. Proc. Natl. Acad. Sci. USA **77**:7328–7332.

NEUHARD, J., E. STAUNING, and R. A. KELLN. 1985. Cloning and characterization of the *pyrE* gene of the *PyrE*::Mu*d*1(Ap *lac*) fusions from *Salmonella typhimurium.* Eur. J. Biochem. **146**:597–603.

OCHMAN, H., and A. C. WILSON. 1986. Evolutionary history of enteric bacteria. *In* J. L. INGRAHAM, K. B. LOW, B. MAGANSANIK, F. C. NEIDHART, M. SCHAECHTER, and H. E. UMBARGER, eds. *Escherichia coli* and *Salmonella typhimurium:* molecular and cellular aspects. ASM, Washington, D.C. (accepted).

ROCHA. M., M. VAZQUEZ, A. GARCIARRUBIO, and A. A. COVARRUBIAS. 1985. Nucleotide sequence of the *glnA-glnL* intercistronic region of *Escherichia coli.* Gene **37**:91–99.

SAINT-GIRONS, I., N. DUCHANGE, G. N. COHEN, and M. M. ZAKIN. 1984. Structure and autoregulation of the *metJ* regulatory gene in *Escherichia coli.* J. Biol. Chem. **259**:14282–14285.

SHARP, P. M., and W.-H. LI. 1986. Codon usage in regulatory genes in *Escherichia coli* does not reflect selection for "rare" codons. Nucleic Acids Res. **14**:7737–7749.

———. 1987. The codon adaptation index—a measure of directional synonymous codon usage bias, and its potential application. Nucleic Acids Res. **15**:1281–1295.

SHARP, P. M., T. M. F. TUOHY, and K. R. MOSURSKI. 1986. Codon usage in yeast: cluster analysis clearly differentiates between highly and lowly expressed genes. Nucleic Acids Res. **14**:5125–5143.

STOCK, A., D. E. KOSHLAND, and J. STOCK. 1985. Homologies between the *Salmonella typhimurium* Che Y protein and proteins involved in the regulation of chemotaxis, membrane protein synthesis, and sporulation. Proc. Natl. Acad. Sci. USA **82**:7989–7993.

SVERDLOV, E. D., N. A. LISITSYN, S. O. GURYEV, YU. V. SMIRNOV, V. M. ROSTAPSHOV, and G. S. MONASTYRSKAYA. 1986. Genes coding for B-subunit of bacterial RNA polymerases. I. Primary structure of the *rpoB* gene *Eco*RI-C fragment from *Salmonella typhimurium* (in Russian). Biooganicheskaya Khimiya **12**:699–707.

URBANOWSKI, M. L., and G. V. STAUFFER. 1985. Nucleotide sequence and biochemical characterization of the *metJ* gene from *Salmonella typhimurium* LT2. Nucleic Acids Res. **13**: 673–685.

WEK, R. C., and G. W. HATFIELD. 1986. Nucleotide sequence and *in vivo* expression of the *ilvY* and *ilvC* genes in *Escherichia coli* K12. J. Biol. Chem. **261**:2441–2450.

WALTER H. FITCH, reviewing editor

Received August 5, 1986; revision received November 20, 1986.