

Synonymous Codon Usage, Accuracy of Translation, and Gene Length in *Caenorhabditis elegans*

Gabriel Marais, Laurent Duret

Laboratoire "Biométrie et Biologie Evolutive," UMR 5558, Bâtiment 711, Université Claude Bernard, Lyon 1,
69622 Villeurbanne Cedex, France

Received: 18 June 2000 / Accepted: 10 November 2000

Abstract. In many unicellular organisms, invertebrates, and plants, synonymous codon usage biases result from a coadaptation between codon usage and tRNAs abundance to optimize the efficiency of protein synthesis. However, it remains unclear whether natural selection acts at the level of the speed or the accuracy of mRNAs translation. Here we show that codon usage can improve the fidelity of protein synthesis in multicellular species. As predicted by the model of selection for translational accuracy, we find that the frequency of codons optimal for translation is significantly higher at codons encoding for conserved amino acids than at codons encoding for nonconserved amino acids in 548 genes compared between *Caenorhabditis elegans* and *Homo sapiens*. Although this model predicts that codon bias correlates positively with gene length, a negative correlation between codon bias and gene length has been observed in eukaryotes. This suggests that selection for fidelity of protein synthesis is not the main factor responsible for codon biases. The relationship between codon bias and gene length remains unexplained. Exploring the differences in gene expression process in eukaryotes and prokaryotes should provide new insights to understand this key question of codon usage.

Key words: Codon usage — Codon bias index — Accuracy of translation — Gene length — *Caenorhabditis elegans*

Introduction

Synonymous codon usage is found to vary both among species and among genes from the same genome (for review see Sharp et al. 1995). These differences among genes and among species are explained by variation in the balance between mutational patterns and natural selection efficacy (for review see Sharp et al. 1993). It is now clearly demonstrated that natural selection shapes codon usage in both unicellular (*Escherichia coli*, *Saccharomyces cerevisiae*) and multicellular organisms (*Drosophila melanogaster*, *Caenorhabditis elegans*, and *Arabidopsis thaliana*) (Grantham et al. 1981; Ikemura 1985; Shields et al. 1988; Stenico et al. 1994; Moriyama and Powell 1997; Duret and Mouchiroud 1999). In these species, codon usage in highly expressed genes is biased toward "optimal" codons corresponding to the most abundant tRNAs (Ikemura 1985; Moriyama and Powell 1997; Duret 2000). However, the exact mechanism of selection acting on codon usage remains unknown.

At least two facets of translation could be affected by codon usage: the rate of elongation and the accuracy of mRNA species translation (Bulmer 1991; Akashi 1994; for review see Akashi and Eyre-Walker 1998). Enhancing the fidelity of protein synthesis at optimal codons could be due to more specific amino acid acylation of the most abundant tRNAs, greater fidelity in the initial discrimination step of protein synthesis at optimal codons, or more efficient proofreading in the subsequent step at these codons (Akashi 1994). In the rest of the text, the terms "accuracy of translation" and "fidelity of protein synthesis" include the three aspects of translation described above.

The effect of translational errors on the function of the protein is not the same for all amino acids. Indeed, some

errors may be tolerated at those positions that are not crucial for protein function, whereas other positions are highly constrained. The nature and proportion of constrained amino acids vary widely according to the proteins. In absence of biochemical or structural data, these amino acids can be predicted by comparative sequence analysis: conserved residues in orthologous proteins from distantly related species are likely to be those that are essential for protein function. The model of accuracy of translation thus predicts that selection on codon usage should be stronger at “constrained” codons encoding for amino acids that are evolutionary conserved than at “nonconstrained” codons encoding for nonconserved amino acids. Until now, this prediction was tested and validated only in *Drosophila* by Akashi (1994) with a method based on comparative statistical tests. Here we show that a new index defined as the difference in frequencies of optimal codons between constrained and nonconstrained codons (ΔF_{op}) is significantly positive for the 548 genes in *C. elegans* for which a human orthologue is available. Hence, we confirm in the nematode that codon usage can improve the accuracy of translation in multicellular species.

Selection for fidelity of protein synthesis is expected to be greater in longer genes because the cost of producing a protein is proportional to its length (Eyre-Walker 1996). Thus, the model of selection for translational accuracy predicts a positive correlation between codon bias and gene length. However, in *S. cerevisiae*, *D. melanogaster*, *C. elegans*, and *A. thaliana*, a negative correlation between codon bias and gene length has been observed (Moriyama and Powell 1998; Comeron et al. 1999; Duret and Mouchiroud 1999). One hypothesis resolving this apparent contradiction is proposed: if long genes are less constrained than short ones, selection for translational accuracy is expected to be weaker in long genes than in short ones. Here we show that there is a positive correlation between the rate of nonsynonymous substitutions (K_a) and the length of 548 genes studied in *C. elegans*. If the relationship between codon bias and gene length is due to a small proportion of constrained codons in long genes, then it is expected that codon bias at constrained codons only should be the same for all the genes whatever their length. However, we find that the frequency of optimal codons at the constrained codons is negatively correlated with the gene length. Consequently, we propose that selection for translational accuracy acts on synonymous codon usage in multicellular species with a low intensity.

Materials and Methods

Human/Nematode Orthologous Pairs and Alignments

We used 819 human/nematode orthologous pairs and protein alignments identified and computed by Whealan et al. (1999). In this data

set, 117 nematode genes had several human homologues with close protein alignment scores. To conduct analysis on independent observations, we selected only the human homologue with the best protein alignment score. In this way, we removed 256 pairs of the initial data set and we conserved only 563 pairs. Identical amino acids in nematode and human orthologues were considered evolutionary conserved. Other amino acids were considered nonconserved. The rate of nonsynonymous substitutions (K_a) was calculated according to the Li (1993) procedure with JaDis (software developed by Gonçalves et al. 1999).

Codon Bias Measurements by F_{op}

Codon bias was measured by the frequency of optimal codons used in a gene (F_{op}) (Stenico et al. 1994). Twenty-one optimal codons for *C. elegans* were identified for the 18 amino acids where alternative synonyms exist by correspondence analysis of relative synonymous codon usage (RSCU) by Stenico et al. (1994). They are strongly used by highly expressed genes and correspond to the most abundant tRNAs in *C. elegans* (Duret and Mouchiroud 1999; Duret 2000). F_{op} is calculated as the number of occurrences of the 21 optimal codons divided by the number of occurrences of the corresponding 18 amino acids. Values range from 0.36 when codon usage is uniform to 1 when codon usage is highly biased (Stenico et al. 1994).

A New Index to Study the Difference in Codon Bias Between Constrained and Nonconstrained Codons in a Gene: ΔF_{op}

ΔF_{op} is the difference in frequencies of optimal codons used in a gene at constrained codons and at nonconstrained codons. ΔF_{op} is calculated for one gene as follows:

$$\Delta F_{op} = \sum_{i=1}^{18} \left[\left(\frac{n_{oc,i}}{n_{c,i}} - \frac{n_{onc,i}}{n_{nc,i}} \right) \left(\frac{n_{c,i} + n_{nc,i}}{n_{tot}} \right) \right]$$

where i is 1 of the 18 amino acids with synonymous codons (all the amino acids except methionine and tryptophane); $n_{oc,i}$, the number of occurrences of an optimal codon(s) of the i th amino acid analyzed at conserved sites in the nematode gene; $n_{onc,i}$, the number of occurrences of an optimal codon(s) of the i th amino acid analyzed at nonconserved sites in the nematode gene; $n_{c,i}$, the number of occurrences of the i th amino acid analyzed at conserved sites in the nematode gene; $n_{nc,i}$, the number of occurrences of the i th amino acid analyzed at nonconserved sites in the nematode gene; and n_{tot} , the total number of codons in the nematode gene excluding amino acids with a unique codon (methionine and tryptophane) and stop codons (UAG, UAA, UGA).

ΔF_{op} is a useful index to test directly the hypothesis of selection for translational accuracy. It corresponds to the sum of the difference in F_{op} between constrained and nonconstrained codons for each of the 18 amino acids in which alternative synonyms exist weighted by their frequencies in the gene. For statistical reasons, 15 pairs were excluded because they contained fewer than 20 constrained codons or 20 nonconstrained codons. Finally, the data set included 548 human/nematode pairs.

Expression Profiles Measuring with ESTs

Expression profiles were determined by counting the number of occurrences of each gene among EST sequences from *C. elegans* libraries (embryo, 29,872 ESTs; adult, 38,115 ESTs) as described by Duret and Mouchiroud (1999). Then we classified genes according to their expression profiles: very low expression (0 ESTs; $N = 123$), low ex-

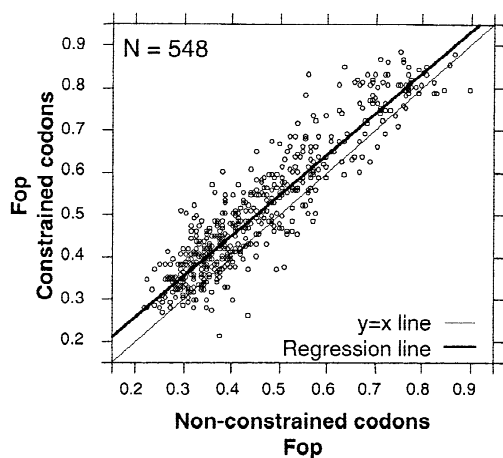


Fig. 1. Comparison of F_{op} at constrained and nonconstrained codons. The *thick line* represents the regression lines of F_{op} at constrained codons versus F_{op} at nonconstrained codons ($y = 0.06 - 0.97x$, $R^2 = 0.86$, $p < 0.0001$) and the *thin line* represents the first diagonal ($y = x$). The means of the difference between F_{op} at constrained codons and at nonconstrained codons is 0.048.

pression (1 to 5 ESTs; $N = 83$), moderate expression (6 to 16 ESTs; $N = 109$), and high expression (more than 16 ESTs; $N = 233$).

Results

Variation of Codon Bias Between Constrained and Nonconstrained Sites in Protein Coding Sequences of 548 Genes in C. elegans

The model of selection for translational accuracy predicts that the frequency of optimal codons should be higher at “constrained” codons encoding for amino acids that are evolutionary conserved than at “nonconstrained” codons encoding for nonconserved amino acids. It is the only model that predicts such a relationship between codon usage in DNA and constraints at the protein level. Selection to increase elongation rates does not predict such an association. For each gene, we determined constrained codons and nonconstrained codons by comparing human/nematode orthologous proteins (see Materials and Methods). To test the hypothesis of selection for translational accuracy, we compared F_{op} at constrained codons and F_{op} at nonconstrained codons. Figure 1 shows that the regression line is above the first diagonal line, which means that F_{op} is significantly higher for constrained codons than for nonconstrained ones. The frequency of optimal codons is about 0.048 (12%) higher at constrained codons than at nonconstrained ones.

However, amino acid compositional biases between functional and nonfunctional sites exist in many genes. For instance, uncommon amino acids such as cysteine occur frequently at constrained sites. Cysteine has two synonymous codons (UGC = optimal codon and UGU = nonoptimal codon), and the frequency of the optimal

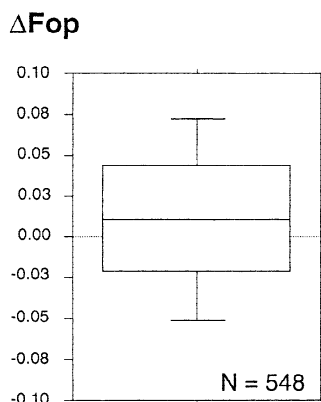


Fig. 2. Difference in frequencies of optimal codons between constrained and nonconstrained codons in a gene (ΔF_{op} ; see Materials and Methods). ΔF_{op} is significantly positive ($\Delta F_{op} = +0.013$, $df = 547$, with theoretical $\Delta F_{op} = 0.0$, Student's $t = 5.2$, $p < 0.0001$). The *central box* depicts the middle 50% of the data between the 25th and the 75th percentiles and the enclosed *horizontal line* represents the median value of the distribution.

codon of this amino acid due to chance is 0.5, compared to 0.36 for all amino acids (see Materials and Methods). Consequently, an excess of cysteine at constrained sites could increase the difference in F_{op} between constrained codons and nonconstrained ones. To take this problem into account, we created a new index defined as the difference in frequencies of optimal codons between constrained and nonconstrained codons (ΔF_{op}). It corresponds to the sum of the difference in F_{op} between constrained and nonconstrained codons for each of the 18 amino acids in which alternative synonyms exist weighted by their frequencies in a gene (see Materials and Methods). Thus, it is not sensitive to amino acid compositional biases between constrained and nonconstrained sites existing in many genes. Then we measured ΔF_{op} for 548 genes in *C. elegans* to study variation of codon bias with functional constraints within genes. Figure 2 shows that ΔF_{op} is significantly greater than zero ($\Delta F_{op} = +0.013$, with theoretical $\Delta F_{op} = 0.0$, Student's $t = 5.2$, $df = 547$, and $p < 0.0001$). Thus, independently of variations in amino acid composition, the frequency of optimal codons is about 0.013 (3%) higher at constrained codons than at nonconstrained ones.

Relationship Between Codon Bias and Length of 548 Genes in C. elegans

With regard to the latter result, we proposed one hypothesis to explain the negative correlation between codon bias and gene length observed in *S. cerevisiae*, *D. melanogaster*, *C. elegans*, and *A. thaliana* (Moriyama and Powell 1998; Comeron et al. 1999; Duret and Mouchiroud 1999). If long genes contain a smaller proportion of constrained codons than short ones, then F_{op} should decrease with gene length. This hypothesis predicts that the rate of nonsynonymous substitutions (K_a) should be posi-

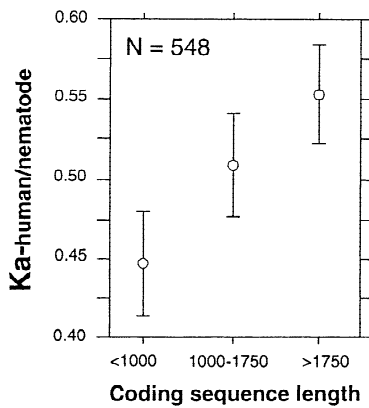


Fig. 3. K_a according to gene length. An analysis of variance (ANOVA) showed that gene length has a significant effect on K_a ($F = 9.8$, $p < 0.0001$). Scheffe's test showed that K_a increases with gene length (short vs. medium genes, $p < 0.05$; short vs. long genes, $p < 0.0001$; medium vs. long genes, nonsignificant). Each item corresponds to the mean K_a for short genes (coding sequence, < 1000 nucleotides; $N = 173$), medium genes (coding sequence, $1000\text{--}1750$ nucleotides; $N = 203$), and long genes (coding sequence, >1750 nucleotides; $N = 172$). Error bars indicate the 95% confidence interval.

tively correlated with gene length. Figure 3 shows that there is a positive correlation between K_a and gene length in *C. elegans* (ANOVA— $F = 9.8$, $p < 0.0001$; Scheffe's test—short vs medium genes, $p < 0.05$; short vs long genes, $p < 0.0001$; medium vs long genes, nonsignificant).

If this effect is responsible for the whole negative correlation between codon bias and gene length, then it is expected that codon bias at constrained codons only should be the same for all genes whatever their length. To test this prediction, we studied F_{op} at constrained codons and nonconstrained codons for 233 highly expressed genes (because they exhibit a strong association between codon bias and gene length in *C. elegans*). Figure 4 shows that the frequency of optimal codons at the constrained codons is negatively correlated with the gene length as observed for all codons and for nonconstrained codons. Moreover, slopes of F_{op} and gene length regression lines for all codons, constrained codons, and nonconstrained codons are quite similar (slopes are 0.32 ± 0.02 , 0.32 ± 0.02 , and 0.30 ± 0.02 , respectively). This means that the variability of K_a with gene length cannot explain the relationship between codon bias and gene length for highly expressed genes and thus for all genes. The hypothesis proposed is thus rejected.

Discussion

Methodological Interests and Limit of ΔF_{op}

ΔF_{op} is a useful index to test directly the hypothesis of selection for fidelity of protein synthesis. It has several

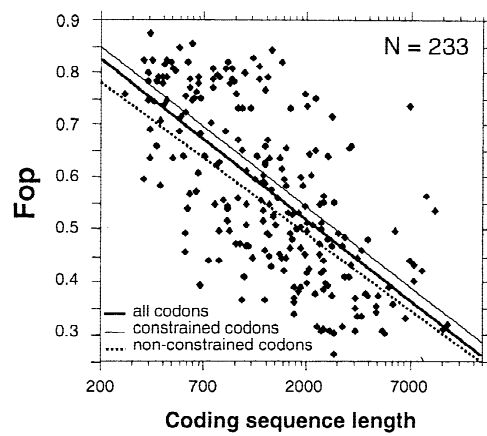


Fig. 4. F_{op} according to gene length for 233 highly expressed genes for constrained, nonconstrained, and all codons. The *thick line* represents the regression lines of F_{op} according to gene length for all codons ($y = 1.6 - 0.32x$, $R^2 = 0.46$, $p < 0.0001$); the *dashed line*, for nonconstrained codons ($y = 1.5 - 0.3x$, $R^2 = 0.41$, $p < 0.0001$); and the *thin line*, for constrained codons ($y = 1.6 - 0.32x$, $R^2 = 0.46$, $p < 0.0001$).

advantages: it is not sensitive to amino acid compositional biases between constrained and nonconstrained sites existing in some genes, it eliminates the problem of the variation in fitness effects due to differences of expression level, and it allows us to quantify the variation of frequencies of optimal codons between constrained and nonconstrained codons. However, as shown in Fig. 2, the mean ΔF_{op} is significantly positive but its variance is high because many genes have a negative ΔF_{op} .

This is likely to be due to the species chosen as a reference to detect functional constraints in *C. elegans* genes. The human/nematode comparisons reveal constrained codons in *C. elegans* genes with a high stringency but they cannot reveal all of them because the evolutionary distance between the two species is too large. Consequently, the set of codons we considered to be nonconstrained because the amino acids they encode were not conserved between nematode and human may include true nonconstrained codons and false nonconstrained codons. The false nonconstrained codons can be classified into two groups: (i) codons encoding for amino acids that lost their function in human proteins or won a new function in nematode proteins since the divergence between *H. sapiens* and *C. elegans* and (ii) codons encoding for amino acids that have experienced conservative changes. To detect with a high stringency the true nonconstrained codons, a species closely related to *C. elegans* such as *Caenorhabditis briggsae* should be used as a reference. Unfortunately, there are not enough *C. briggsae* sequences in the public databases to conduct an analysis with statistical significance. However, it is important to note that the methodological limit described above mixes constrained and nonconstrained codons and produces a decrease in ΔF_{op} . Thus, the ΔF_{op} measured previously is likely to be underestimated.

Does Codon Usage Improve the Accuracy, the Speed of Translation, or Both?

F_{op} is higher at constrained codons than at nonconstrained ones as expected according to the model of selection for translational accuracy. However, mutational hypotheses could be proposed to explain this observation. Akashi (1994) tested mutational alternatives that could potentially explain the higher frequency of optimal codons at constrained codons than at nonconstrained ones, and all of them were clearly rejected. We could not test these mutational hypotheses with our data set but we assumed that they were probably invalid in the nematode as in *Drosophila*. Furthermore, selection to increase elongation rates does not predict a relationship between codon usage in DNA and functional constraints at the protein level. Thus, selection for fidelity of protein synthesis is the most likely model to explain the result obtained.

However, a question remains: Is it possible to exclude selection for speed of translation to explain codon usage pattern in *C. elegans*? The first aspect of codon usage pattern which speed and accuracy models have to explain concerns the positive correlation between codon bias and level of gene expression (Duret and Mouchiroud 1999). On one hand, if it is assumed that the metabolic cost of a misincorporation is a function of the number of nonfunctional proteins synthesized, selection for translational accuracy should act more strongly in highly expressed genes (Akashi 1994). On the other hand, if it is considered that the cellular content of free ribosomes is a limiting factor of translation as suggested by experiments conducted on yeast (for a review see Warner 1999), selection for translational rate of elongation should act more strongly in highly expressed genes because they can potentially occupy many ribosomes (Bulmer 1991). Hence, both speed and accuracy models can explain the association between expression profiles and codon bias in many species including *C. elegans*.

The second aspect of codon usage pattern which speed and fidelity models have to explain is how variations in codon usage could affect differential survival of individuals in natural populations. On one hand, enhancing the accuracy of translation provides economy of energy consumption, which may increase the cell growth rate and, in turn, fitness (Bulmer 1991). On the other hand, enhancing the speed of translation could maintain the pool of free ribosomes and thus optimize the growth rate (Kurland et al. 1996) and, finally, fitness (Bulmer 1991; Berg and Kurland 1997). Hence, selection for speed and accuracy of protein synthesis explains how codon usage may act on the growth rate and then on fitness in many species including *C. elegans*. Consequently, we cannot exclude that codon usage could also improve the rate of elongation of protein synthesis. Moreover, several experiments conducted on *E. coli*

show that optimal codons could increase both the accuracy and the speed of translation (Robinson et al. 1984; Precup and Parker 1987; Sorensen et al. 1989). Thus codon usage is possibly related to both the elongation rate and the accuracy of translation.

The Relationship Between Gene Length and Codon Bias Stays Unresolved

Selection to maximize the rate of elongation is independent of gene length, and selection for translational accuracy predicts a positive correlation between codon bias and gene length (Eyre-Walker 1996; Duret and Mouchiroud 1999). Thus, speed and accuracy models do not predict the negative correlation between codon bias and gene length in *S. cerevisiae*, *D. melanogaster*, *C. elegans*, and *A. thaliana*. In the present work, we show that long genes are less constrained than short ones but this cannot explain the negative correlation between codon bias and gene length in *C. elegans*.

Furthermore, all the hypotheses classically proposed in the literature to explain the negative correlation between codon bias and gene length have been actually rejected. Short genes are not more expressed than long ones in *D. melanogaster*, *C. elegans*, and *A. thaliana* (Duret and Mouchiroud 1999). The observed decrease in the global F_{op} with protein length could be explained if the functional constraints responsible for selection against nonoptimal codons were restricted to a limited portion of the gene. However, Duret and Mouchiroud (1999) rejected this hypothesis in both nematode and fly.

Another hypothesis based on population genetics models proposes that there might be a small-scale Hill–Robertson effect acting on codon usage in a gene (Li 1987; Comeron et al. 1999). The Hill–Robertson effect corresponds to a decrease in natural selection efficacy with increasing genetic linkage (Hill and Robertson 1966; Maynard-Smith and Haigh 1974; Charlesworth et al. 1993): natural selection cannot choose the best variants for many genes if these genes are tightly linked. Many simulations have shown that this kind of effect could occur in virtual sequences (Li 1987; Comeron et al. 1999). However, these simulations are not realistic because sequences were considered to evolve independently among generations. Thus, the results obtained by simulations have to be validated by the prediction that several sequences of various lengths should have the same codon bias when they are genetically linked (Comeron et al. 1999). However, this prediction has been refuted in *C. elegans* (Duret and Mouchiroud 1999).

An interesting aspect of the surprising negative correlation between codon bias and gene length is that it occurs only in eukaryotes. In *E. coli*, the beginning and the end of the protein coding sequence of the genes are constrained because of a selective pressure to maintain promoting translation signals (Eyre-Walker 1996). If the

beginning and the end of the sequence are removed, no correlation between codon bias and gene length is observed (Comeron et al. 1999). The relationship between codon bias and gene length could result in differences in the gene expression process in eukaryotes and prokaryotes. Presently, all the hypotheses tested have been rejected and this observation stays unresolved. It seems to be a key question to understand the exact mechanism of selection on codon usage in various species such as *C. elegans*, *D. melanogaster*, *A. thaliana*, and *S. cerevisiae*.

Conclusion

Akashi (1994) showed that the model of selection for fidelity of protein synthesis is valid for codon usage in *Drosophila* and we show that it is also valid for codon usage in the nematode. However, the model of selection to increase elongation rates cannot be excluded for these invertebrates. Furthermore, the model of the selection for accuracy of translation predicts a strong positive correlation between codon bias and gene length (Eyre-Walker 1996), whereas the model of selection for speed of translation predicts no correlation (Duret and Mouchiroud 1999). However, a strong negative correlation between codon bias and gene length has been observed in several eukaryotes including *D. melanogaster* and *C. elegans* (Moriyama and Powell 1998; Comeron et al. 1999; Duret and Mouchiroud 1999). This observation suggests that selection for translational accuracy is not the main factor responsible for codon biases. It should be noted that the model of selection for translational speed cannot explain the effect of gene length on codon biases either. Finally, the relationship between codon bias and gene length remains unexplained. Since the relationships between codon usage bias and gene length are different in eukaryotes and prokaryotes (Eyre-Walker 1996; Moriyama and Powell 1998; Comeron et al. 1999; Duret and Mouchiroud 1999), exploring the differences in the gene expression process in eukaryotic and prokaryotic organisms should provide new insights to understand this major point of codon usage.

Acknowledgments. We thank Wojciech Makalowski for the data set of human/nematode orthologous pairs and Dominique Mouchiroud and Christian Gautier for helpful comments and discussions. This work was supported by the Ministère de la Recherche and the Centre National de la Recherche Scientifique.

References

- Akashi H (1994) Synonymous codon usage in *Drosophila melanogaster*: Natural selection and translational accuracy. *Genetics* 136:927–935
- Akashi H, Eyre-Walker A (1998) Translational selection and molecular evolution. *Curr Opin Genet Dev* 8:688–693
- Berg OG, Kurland CG (1997) Growth rate-optimised tRNA abundance and codon usage. *J Mol Biol* 270:544–550
- Bulmer M (1991) The selection-mutation-drift theory of synonymous codon usage. *Genetics* 129:897–907
- Charlesworth B, Morgan MT, Charlesworth D (1993) The effect of deleterious mutations on neutral evolution. *Genetics* 134:1289–1303
- Comeron JM, Kreitman M, Aguadé M (1999) Natural selection on synonymous sites is correlated with gene length and recombination in *Drosophila*. *Genetics* 151:239–249
- Duret L (2000) tRNA gene number and codon usage in *C. elegans* genome are co-adapted for the optimal translation of highly expressed genes. *Trends Genet* 16:287–289
- Duret L, Mouchiroud D (1999) Expression pattern and, surprisingly, gene length shape codon usage in *Caenorhabditis*, *Drosophila*, and *Arabidopsis*. *Proc Natl Acad Sci USA* 96:4482–4487
- Eyre-Walker A (1996) Synonymous codon bias is related to gene length in *Escherichia coli*: Selection for translational accuracy? *Mol Biol Evol* 13:864–872
- Goncalves I, Robinson M, Perrière G, Mouchiroud D (1999) JaDIS: Computing distances between nucleic acid sequences. *Bioinformatics* 15:424–425
- Grantham R, Gautier C, Gouy M, Jacobzone M, Mercier R (1981). Codon catalog usage is a genome strategy modulated for gene expressivity. *Nucleic Acids Res* 9:43–74
- Hill WG, Robertson A (1966) The effect of linkage on limits to artificial selection. *Genet Res* 8:269–294
- Ikemura T (1985) Codon usage and tRNA content in unicellular and multicellular organisms. *Mol Biol Evol* 2:13–34
- Kurland CG, Hughes D, Ehrenberg M (1996) Limitations of translational accuracy. In *Escherichia coli* and *Salmonella*. In: Neidhart FC (ed) Cellular and molecular biology. ASM Press, Washington, DC, pp 979–1004
- Li WH (1987) Models of nearly neutral mutations with particular implications for nonrandom usage of synonymous codons. *J Mol Evol* 24:337–345
- Li WH (1993) Unbiased estimation of the rates of synonymous and nonsynonymous substitutions. *J Mol Evol* 36:96–99
- Maynard-Smith J, Haigh J (1974) The hitch-hiking effect of a favourable gene. *Genet Res* 23:23–35
- Moriyama EN, Powell JR (1997) Codon usage bias and tRNA abundance in *Drosophila*. *J Mol Evol* 45:514–523
- Moriyama EN, Powell JR (1998) Gene length and codon usage bias in *Drosophila melanogaster*, *Saccharomyces cerevisiae* and *Escherichia coli*. *Nucleic Acids Res* 26:3188–3193
- Precup J, Parker J (1987) Missense misreading of asparagine codons as a function of codon identity and context. *J Biol Chem* 262:11351–11356
- Robinson M, Lilley R, Little S, Emtage JS, Yamamoto G, Stephans P, Milligan A, Eaton M, Humphreys G (1984) Codon usage can affect efficiency of translation of genes in *Escherichia coli*. *Nucleic Acids Res* 12:6663–6671
- Sharp PM, Stenico M, Peden JF, Lloyd AT (1993) Codon usage: Mutational bias, translational selection, or both? *Biochem Soc Trans* 21:835–841
- Sharp PM, Averof M, Lloyd AT, Matassi G, Peden JF (1995) DNA sequence evolution: The sounds of silence. *Philos Trans R Soc Lond B Biol Sci* 349:241–247
- Shields DC, Sharp PM, Higgins DG, Wright F (1988) “Silent” sites in *Drosophila* genes are not neutral: Evidence of selection among synonymous codons. *Mol Biol Evol* 5:704–716
- Sorensen MA, Kurland CG, Pedersen S (1989) Codon usage determines translation rate in *Escherichia coli*. *J Mol Biol* 207:365–377
- Stenico M, Lloyd AT, Sharp PM (1994) Codon usage in *Caenorhabditis elegans*: Delineation of translational selection and mutational biases. *Nucleic Acids Res* 22:2437–2446
- Warner JR (1999) The economics of ribosome biosynthesis in yeast. *TIBS* 24:437–440
- Wheelan SJ, Boguski MS, Duret L, Makalowski W (1999) Human and nematode orthologs—Lessons from the analysis of 1800 human genes and the proteome of *Caenorhabditis elegans*. *Gene* 238:163–170