Genet. Res., Camb. (2001), **78**, *pp.* 171–175. *With 2 figures.* © 2001 Cambridge University Press
DOI: 10.1017/S0016672301005249 Printed in the United Kingdom

Multiplicative versus additive selection in relation to genome evolution: a simulation study

GWENAËL PIGANEAU[1]*, ROLAND WESTRELIN[2], BERNARD TOURANCHEAU[2] AND CHRISTIAN GAUTIER[1]

[1] *Centre National de la Recherche Scientifique, Unité Mixte de Recherche 5558, Biométrie et Biologie Évolutive, Université Claude Bernard Lyon 1, 43 boulevard du 11 novembre 1918, 69622 Villeurbanne cedex, France*
[2] *Laboratoire Réseaux hauts Débits et Supports d'Applications Multimedias, Institut National de Recherche en Informatique et en Automatique, Université Claude Bernard Lyon 1, 69622 Villeurbanne cedex, France*

(*Received 8 January 2001 and in revised form 12 April and 3 May 2001*)

Summary

The evolution of molecular quantitative traits, such as codon usage bias or base frequencies, can be explained as the result of mutational biases alone, or as the result of mutation and selection. Whereas mutation models can be investigated easily, realistic modelling of selection-directed genome evolution is analytically intractable, and numerical calculations require substantial computer resources. We investigated the evolution of optimal codon frequency under additive and multiplicative effects of selected linked codons. We show that additive selective effects of many linked sites cannot be effective in genomes when the number of selected sites is greater than the effective population size, a realistic assumption according to current molecular data. We then discuss the implications of these results for isochore evolution in vertebrates.

1. Introduction

The efficacy of selection acting simultaneously at linked sites (a site can be a codon, a nucleotide or a gene, depending on what contributes to fitness) is reduced compared with the same selection pressure acting at independent sites (Hill & Robertson, 1966; Li, 1987). This is because linkage disequilibrium between alleles at selected loci, generated by the stochastic nature of mutation and sampling in a finite population, 'interferes' with the action of selection at any one locus (Felsenstein, 1974). Recent work on the effect of linkage on the efficiency of selection has stressed the fact that Hill–Robertson (HR) interference plays a major role in the effect of selection on codon usage bias and the pattern of silent polymorphism within the genome (Comeron *et al.*, 1999; McVean & Charlesworth, 2000). The efficiency of selection can be predicted assuming independence between sites (Li, 1987; McVean & Charlesworth, 1999). Let v be the mutation rate from optimal codons to non-optimal codons and u the reverse mutation rate. Let x be the optimal codon frequency, s be the selection coefficient for optimal codons, and N the effective size of the diploid population. Then, at mutation–selection–drift equilibrium:

$$E(x) = \sum_{n=0}^{\infty} \frac{(4Ns)^n \Gamma(4Nu+n+1)}{n!\Gamma(4N(u+v)+n+1)} \Big/ \sum_{n=0}^{\infty} \frac{(4Ns)^n \Gamma(4Nu+n)}{n!\Gamma(4N(u+v)+n)}, \quad (1)$$

where $\Gamma(.)$ denotes the Gamma function (Li, 1987). At mutation–selection–drift equilibrium, the optimal codon frequency is determined by Ns, $N(u+v)$ (the mean mutation rate for one codon per generation), and $u/(u+v)$ the average frequency of optimal codons without selection. There are two difficulties when several sites participate in a trait (e.g. the optimal codon frequencies of genes or of a genome). The first one is the reduced selection efficiency, e.g. codon bias, due to HR interference between the selected sites (Comeron *et al.*, 1999; McVean & Charlesworth, 2000). The second difficulty is the way to define the contribution of each codon to the fitness of a genome.

* Corresponding author. Telephone: + 33 4 72 44 81 96. Fax: + 33 4 78 89 27 19. e-mail: piganeau@biomserv.univ-lyon1.fr.

2. Definition of the additive and multiplicative selection schemes

We limit the present analysis to two previously published selection schemes. From Li's (1987) definition, in the case of L diallelic loci, the absolute fitness of a genome consisting of L codons with i optimal codons is equal to $F_i = 1-(L-i)s'$ for negative selection, and $F_i = 1+is$ for positive selection when additivity over codons is assumed. The positive and negative additive selection models are mathematically equivalent with $s' = s/(1+Ls)$, $Ls' < 1$ under this definition. This can be deduced from the expression of the ratio of the fitness of a sequence with $i+j$ optimal codons (F_{i+j}) over the ratio of a sequence with i optimal codons (F_i) under both schemes (Table 1). Under the multiplicative selection scheme, the absolute fitness is $F_i = (1-s')^{L-i}$ or $F_i = (1+s)^i$. The positive and negative multiplicative selection models are mathematically equivalent if $s' = s/(1+s)$.The additive and the multiplicative selection schemes are nearly equivalent when the product $Ls \ll 1$, since $(1+s)^i \approx 1+is$ (Li, 1987).

For a given s', negative additive selection against non-optimal codons is more effective on codon usage bias than multiplicative selection (McVean & Charlesworth, 2000). This arises from the definition of the additive and the negative selection scheme (Table 1). In the case of positive selection, the ratio F_{i+j}/F_i is greater under the multiplicative selection scheme than under the additive selection scheme, so that one expects multiplicative selection to be more effective than additive selection. In the case of negative selection, the contrary is true: multiplicative effects are less effective than additive effects (Table 1). The negative additive selection scheme is restricted to the parameter space $Ls' < 1$, so that we have used the positive selection scheme. The total number of synonymous sites is estimated to be greater than the effective population size, N, in many species (Kondrashov, 1995). For example, the total number of codons in *Drosophila* is about $L \approx 7\times10^6$ (total number of genes: 13600 (Adams *et al.*, 2000), average protein length 500 amino acids) and the effective population size is about $N \approx 10^6$ (Sawyer & Hartl, 1992); in the worm *Caenorhabditis elegans* $L \approx 7\times10^6$, and in the plant *Arabidopsis thaliana* $L \approx 10^7$. Estimates of effective population sizes are lacking in the latter species. Selection on codon usage has been demonstrated in *Drosophila* (Shields *et al.*, 1998), *Caenorhabditis* and *Arabidopsis* (Stenico *et al.*, 1994; Duret & Mouchiroud, 1999). As a consequence, since selection is effective when $Ns \approx 1$ (Kimura, 1983) it is reasonable to study the case $Ls > 1$ (where L is the total number of codons in a genome).

3. Results

We explored the efficiency of selection under positive directional selection on optimal codons under both selection schemes, using Li's (1987) multisite model. The equilibrium frequency of the optimal codon x_{eq} was computed 100 times every $2N$ generations to ensure independence between values (Li, 1987), allowing $10(u+v)$ generations to reach equilibrium (Tachida, 2000). The corresponding value predicted from the single-site theory was calculated from each run (for this haploid model, $4N$ is replaced by $2N$ in equation 1). We used $N(u+v) = 0{\cdot}05$ and $0{\cdot}005$, $N = 10^2$, 10^3 and 10^4, L from 32 to 38400 and $Ns = 0{\cdot}1, 0{\cdot}4$, 1, 4 and 10. The number of recombination events per generation was assumed to follow a Poisson process with parameter cNL. Each simulation was run at least four times. Fig. 1 shows that, for positive selection over sites, additive effects of optimal codons are less effective than multiplicative effects. An increase in Ls, L constant, equivalent to a decrease in N keeping Ns and $N(u+v)$ constant, dramatically reduces the efficacy of additive selection over sites (Fig. 1*A*, *B*). Thus, the selection–mutation–drift equilibrium optimal codon frequency depends on N and L with additive effects of selected sites.

A methodological consequence is that the calculation times (which increase with increasing population size) cannot be reduced by keeping $N(u+v)$ and Ns constant under additive effects when $Ls > 1$. This restriction has not been noted in the previous simulation studies comparing additive and multi-

Table 1. *Comparison of multiplicative versus additive selection under positive and negative selection. F_i is the absolute fitness of a sequence of L sites with i optimal sites*

	Positive selection			Negative selection		
Fitness	Multiplicative		Additive	Multiplicative		Additive
F_{i+j}	$(1+s)^{i+j}$		$1+is+js$	$(1-s')^{L-i-j}$		$1-(L-i-j)s'$
F_i	$(1+s)^i$		$1+is$	$(1-s')^{L-i}$		$1-(L-i)s'$
F_{i+j}/F_i	$(1+s)^j$	$>$	$1+sj(1+is)$	$(1-s')^{-j}$	$<$	$1+s'j/(1-(L-i)s')$

https://doi.org/10.1017/S0016672301005249 Published online by Cambridge University Press

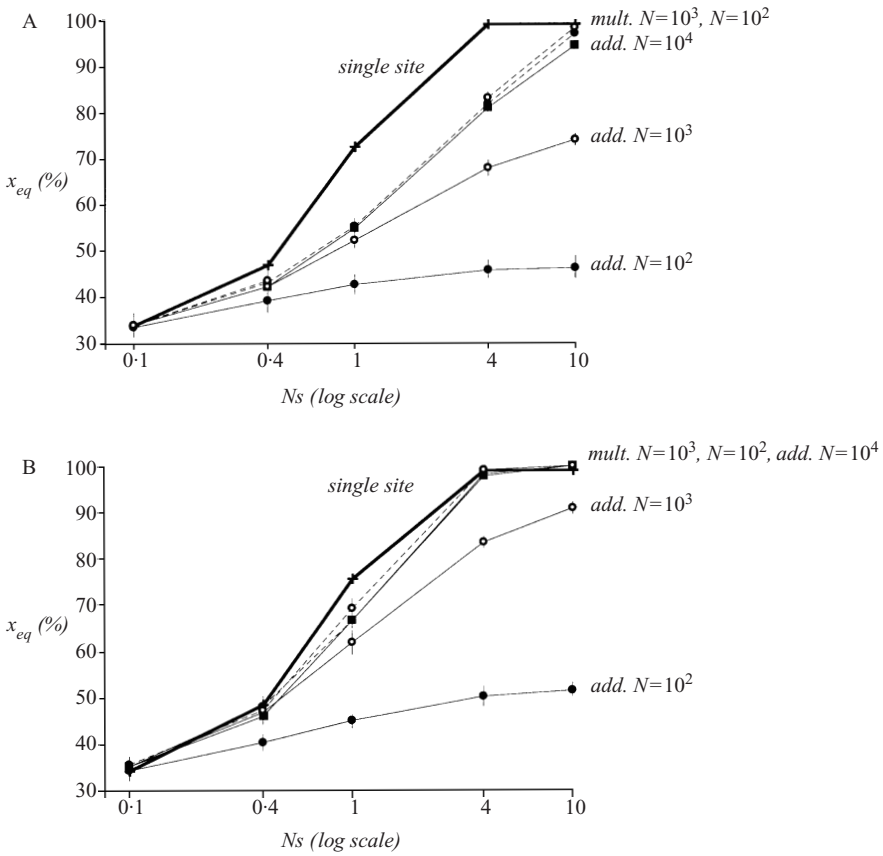

Fig. 1. Evolution of the optimal codon frequency *xeq* in a sequence of 384 completely linked codons for $N(u+v) = 0{\cdot}05$ (*A*) and $N(u+v) = 0{\cdot}005$ (*B*), under mutational bias, $u/(u+v) = 30\,\%$, increasing selection and different population sizes. Bold lines, (+) prediction from the single-site theory (equation (1), see text); dashed lines, multiplicative effects (open circles, $N = 10^3$; filled circles, $N = 10^2$); continuous lines, additive effects (filled squares, $N = 10^4$; open circles, $N = 10^3$; filled circles, $N = 10^2$).

plicative selection schemes (Li, 1987; McVean & Charlesworth, 2000). When effects of optimal codons are multiplicative, the relevant parameters are $N(u+v)$, Ns and $L(u+v)$ (Tachida, 2000; McVean & Charlesworth, 2000). The effect of HR interference increases with $L(u+v)$, the mean number of mutations per genome per generation (Fig. 2). HR interference becomes negligible, that is x_{eq} is well predicted from the single-site theory, for small $L(u+v)$ under multiplicative or additive codon effects when $Ls < 1$ (Fig. 1*B*; Fig. 2 with $L = 32$).

Introduction of recombination between sites reduces the effects of HR interference under the multiplicative fitness scheme (Comeron *et al.*, 1999; McVean & Charlesworth, 2000). The effect of recombination at the rate of one crossover per sequence per generation is sufficient to remove all the effects of interference (Fig. 2*B*, *D*), except for high values of $L(u+v)$, when $(u+v) \approx c$ (Fig. 2*B*; $L = 38\,400$, $(u+v) = 5\times10^{-5}$, $c = 3\times10^{-5}$). HR interference on genomes depends on the $(u+v)/c$ ratio: when the recombination rate per genome Lc is not sufficient to make the $L(u+v)$ segregating sites independent ($u+v \approx c$ was not sufficient, from our simulations), there is interference and reduction of selection efficiency. When c is greater than $(u+v)$, the segregating sites are independent and there is no interference: the frequency of optimal codon is well predicted from the single-site theory (Fig. 2*B*, *D*). For the additive selection scheme, the effect of recombination is not sufficient to predict x_{eq} from the single-site model (Fig. 2*A*, *C*), and recombination does not increase the efficiency of selection at all for $L = 38\,400$. More exactly, recombination cannot increase the efficacy of selection for $Ls > 1$, as a simple consequence of the definition of the additive fitness: the relative fitness of a sequence with $i+1$ optimal sites appearing from mutation in a haploid population of size N compared with the relative fitness of a sequence containing i optimal sites is equal to the ratio of the absolute fitness F_{i+1}/F_i:

$$\frac{F_{i+1}}{F_i} = 1 + \frac{s}{1+is}.$$

It follows that for a given selection coefficient over site and L very large, so that the number of optimal sites i (at least equal to $uL/(u+v)$) is also very large, the difference between the fitness of a sequence with $i+1$ optimal sites and the fitness of a sequence with i

https://doi.org/10.1017/S0016672301005249 Published online by Cambridge University Press

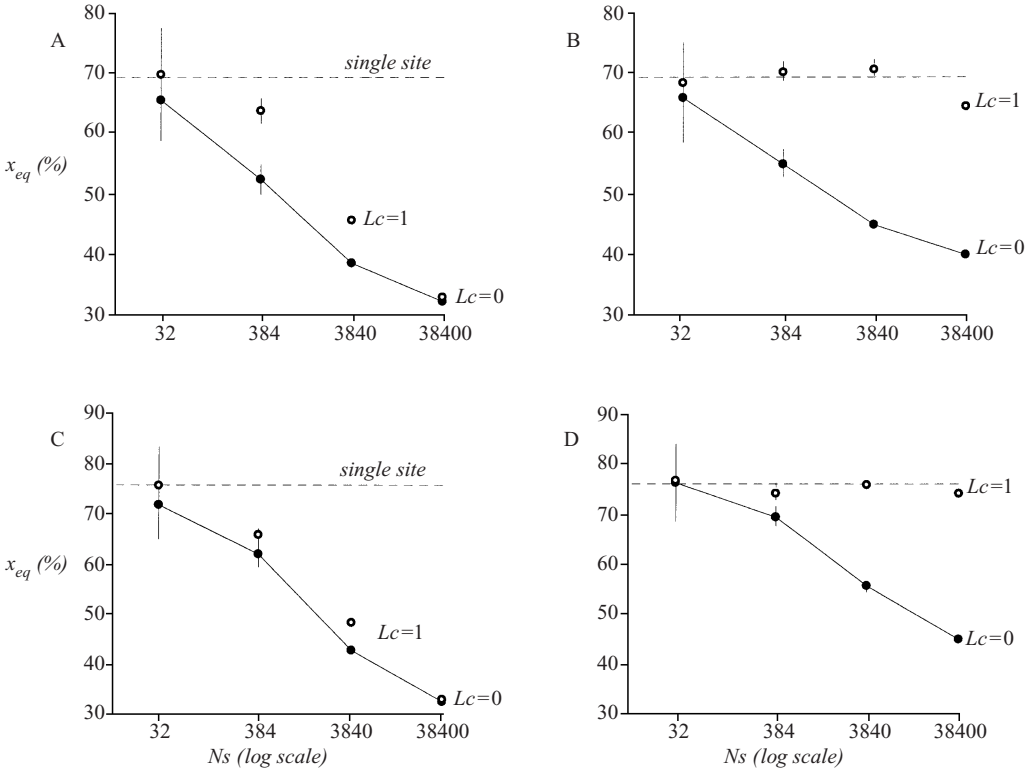

Fig. 2. Evolution of the optimal codon frequency *xeq* with increasing L under total linkage (continuous lines) and one crossover per sequence per generation (open circles), compared with the predictions from single-site theory (dashed line), $Ns = 1$. Additive effects: $N(u+v) = 0{\cdot}05$ (*A*) and $N(u+v) = 0{\cdot}005$ (*C*). Multiplicative effects: $N(u+v) = 0{\cdot}05$ (*B*) and $N(u+v) = 0{\cdot}005$ (*D*). $N = 10^3$, $u(u+v) = 30\,\%$.

optimal sites is negligible, as $F_{i+1}/F_i \approx 1$. More generally, selection can not distinguish between genomes containing i optimal sites and genomes containing $i+j$ optimal sites when the number of selected sites is very large.

4. Discussion

In vertebrate genomes, codon usage bias is related to the local G+C content of the genome, that is the isochore localization of the gene (Bernardi *et al.*, 1985). The debate simmers on whether the isochore structure (for review see Bernardi, 2000) of the vertebrate genome is due to selection (Bernardi *et al.*, 1985; Eyre Walker, 1999) or to neutral mechanisms (Wolfe *et al.*, 1989) Gu & Li, 1994). If selection for GC nucleotides is involved, *Ls* is far beyond 1 (the number of nucleotides in the GC-rich isochores is roughly 10^9 and the long-term effective population size 10^4 to 10^5; Nachman & Crowell, 2000), so that additive effects of selection on GC nucleotides cannot be considered to be responsible for the emergence of the isochore structure in vertebrates. The molecular significance of additive or multiplicative effects of selected GC nucleotides remains an open question. Increased thermostability of GC-rich isochores has been proposed to be advantageous for warm-blooded vertebrates (Bernardi *et al.*, 1985; but see Hughes *et al.*, 1999; Rodrigues-Trelles *et al.*, 2000). Biochemical studies suggest that DNA thermostability increases linearly with GC content (Marmur & Doty, 1959), whereas conformational properties of DNA would rather imply non-linear effects (the effect of replacing one AT by one GC depends on the amount of GC present in the sequence) (Foloppe & MacKerell, 1999). The estimated mutation rate in the human genome, of about 10^{-8} mutations per nucleotide per generation (Nachman & Crowell, 2000), is of the same order of magnitude as the estimated overall recombination rate in the human genome, 10^{-8} recombinations per nucleotide per generation (Weissenbach *et al.*, 1992), although recombination rates vary greatly along the genome (Lynn *et al.*, 2000). This suggests that HR interference may be an important process if selection plays a role in the evolution of the isochore structure.

Our analysis suggests that additive selection on codon usage over the whole genome of most eukaryotic genomes (for which the number of codons is greater than the effective population size) cannot account for codon usage bias. But, as codon usage is highly biased only in the more strongly expressed proteins (e.g. Duret & Mouchiroud, 2000), the number of selected sites in a genome must be smaller than the total number of codons, so that additive directional selection may be effective in species with large effective

https://doi.org/10.1017/S0016672301005249 Published online by Cambridge University Press

population sizes, such as *Drosophila*. However, we are aware that the definition of the fitness of the genome as an additive or multiplicative increasing and symmetrical function of the codon content is not satisfactory, although it is the simplest way to study HR interference. Many biochemical models of the effect of codon usage on protein translation rate or translation efficiency have been developed but do not take into account the effect of linkage between genes and the resulting expected interference (Bulmer, 1991; Zhang *et al*., 1994; Xia, 1996). We can therefore expect more realistic models of codon usage to be developed, taking into account both optimal codon usage and linkage between the genome sequences.

We thank L. Duret, D. Mouchiroud and E. Baake for many useful discussions and helpful comments on the manuscript. This work is supported by the French bioinformatics programme.

References

Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. D. *et al.* (2000). The genome sequence of *Drosophila melanogaster*. *Science* **287**, 2185–2195.

Bernardi, G. (2000). Isochores and the evolutionary genomics of vertebrates. *Gene* **241**, 3–17.

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

Bulmer, M. G. (1991). The selection–mutation–drift theory of synonymous codon usage. *Genetics* **129**, 897–907.

Comeron, J. M., Kreitman, M. & Aguadé, M. (1999). Natural selection of synonymous sites is correlated with gene length and recombination in *Drosophila*. *Genetics* **151**, 239–249.

Duret, L. & Mouchiroud, D. (1999). Expression pattern and, surprisingly, gene length shape codon usage in *Caenorhabditis*, *Drosophila*, and *Arabidopsis*. *Proceedings of the National Academy of Sciences of the USA* **96**, 4482–4487.

Eyre Walker, A. (1999). Evidence of selection on silent site base composition in mammals: potential implications for the evolution of isochores and junk DNA. *Genetics* **152**, 675–683.

Felsenstein, J. (1974). The evolutionary advantage of recombination. *Genetics* **78**, 737–756.

Foloppe, N. & MacKerell, A. D. (1999). Intrinsic conformational properties of deoxyribonucleosides: implicated role for cytosine in the equilibrium among A, B and Z forms of DNA. *Biophysical Journal* **76**, 3206–3218.

Gu, X. & Li, W. H. (1994). A model for the correlation of mutation rate with GC content and the origin of GC rich isochores. *Journal of Molecular Evolution* **38**, 468–475.

Hill, W. G. & Robertson, A. (1966). The effect of linkage on limits to artificial selection. *Genetical Research* **8**, 269–294.

Hughes, S., Zelus, D. & Mouchiroud, D. (1999). Warm-blooded isochore structure in Nile Crocodile and Turtle. *Molecular Biology and Evolution* **16**, 1521–1527.

Kimura, M. (1983). *The Neutral Theory of Molecular Evolution*. Cambridge: Cambridge University Press.

Kondrashov, A. S. (1995). Contamination of the genome by very slightly deleterious mutations: Why have we not died 100 times over? *Journal of Theoretical Biology* **175**, 583–594.

Li, W. H. (1987). Models of nearly neutral mutations with particular implications for nonrandom usage of synonymous codons. *Journal of Molecular Evolution* **24**, 337–345.

Lynn, A., Kashuk, C., Petersen, M. B., Bailey, J. A., Cox, D. R., *et al.* (2000). Patterns of meiotic recombination on the long arm of chromosome 21. *Genome Research* **10**, 1319–1332.

Marmur, J. & Doty, P. (1959). Heterogeneity in deoxyribonucleic acids. I. Dependence on composition of the configurational stability of deoxyribonucleic acids. *Nature* **183**, 1427–1429.

McVean, G. A. T. & Charlesworth, B. (1999). A population genetic model for the evolution of synonymous codon usage: patterns and prediction. *Genetical Research* **74**, 145–158.

McVean, G. A. T. & Charlesworth, B. (2000). The effects of Hill–Robertson interference between weakly selected mutations on patterns of molecular evolution and variation. *Genetics* **155**, 929–944.

Nachman, M. W. & Crowell, S. L. (2000). Estimate of mutation rate per nucleotide in humans. *Genetics* **156**, 297–304.

Rodrigues-Trelles, F., Tarrio, R. & Ayala, F. J. (2000). Fluctuating mutation bias and the evolution of base composition in *Drosophila*. *Journal of Molecular Evolution* **50**, 1–10.

Sawyer, S. A. & Hartl, D. L. (1992). Population genetics of polymorphism and divergence. *Genetics* **132**, 1161–1176.

Shields, D. C., Sharp, P. M., Higgins, D. G. & Wright, F. (1998). 'Silent' sites in *Drosophila* genes are not neutral: evidence of selection among synonymous codons. *Molecular Biology and Evolution* **5**, 704–716.

Stenico, M., Lloyd, A. T. & Sharp, P. M. (1994). Codon usage in *Caenorhabditis elegans*: delineation of translation selection and mutational biases. *Nucleic Acids Research* **22**, 2437–2446.

Tachida, H. (2000). Molecular evolution in a multisite nearly mutation model. *Journal of Molecular Evolution* **50**, 69–81.

Weissenbach, J., Gyapay, G., Dib, C., Vignal, A., Morissette, J., *et al.* (1992). A second-generation linkage map of the human genome. *Nature* **359**, 794–801.

Wolfe, K. H., Sharp, P. M. & Li, W. H. (1989). Mutation rates differ among regions of the mammalian genome. *Nature* **337**, 283–285.

Xia, X. (1996). Maximizing transcription efficiency causes codon usage bias. *Genetics* **144**, 1309–1320.

Zhang, S., Goldman, E. & Zubay, G. (1994). Clustering of low codons and ribosome movement. *Journal of Theoretical Biology* **170**, 339–354.