

Hill-Robertson Interference is a Minor Determinant of Variations in Codon Bias Across *Drosophila melanogaster* and *Caenorhabditis elegans* Genomes

Gabriel Marais* and Gwenaél Piganeau†¹

*Laboratoire “Biométrie et biologie évolutive,” UMR CNRS 5558, Université Claude Bernard Lyon 1, Villeurbanne, France; and †Center for the Study of Evolution, University of Sussex, Brighton

According to population genetics models, genomic regions with lower crossing-over rates are expected to experience less effective selection because of Hill-Robertson interference (HRI). The effect of genetic linkage is thought to be particularly important for a selection of weak intensity such as selection affecting codon usage. Consistent with this model, codon bias correlates positively with recombination rate in *Drosophila melanogaster* and *Caenorhabditis elegans*. However, in these species, the G+C content of both noncoding DNA and synonymous sites correlates positively with recombination, which suggests that mutation patterns and recombination are associated. To remove this effect of mutation patterns on codon bias, we used the synonymous sites of lowly expressed genes that are expected to be effectively neutral sites. We measured the differences between codon biases of highly expressed genes and their lowly expressed neighbors. In *D. melanogaster* we find that HRI weakly reduces selection on codon usage of genes located in regions of very low recombination; but these genes only comprise 4% of the total. In *C. elegans* we do not find any evidence for the effect of recombination on selection for codon bias. Computer simulations indicate that HRI poorly enhances codon bias if the local recombination rate is greater than the mutation rate. This prediction of the model is consistent with our data and with the current estimate of the mutation rate in *D. melanogaster*. The case of *C. elegans*, which is highly self-fertilizing, is discussed. Our results suggest that HRI is a minor determinant of variations in codon bias across the genome.

Introduction

Synonymous codon usage bias commonly observed in living forms is usually assumed to be under a selection-mutation-drift balance (Bulmer 1991). In many unicellular organisms, invertebrates and plants, codon bias is thought to be mainly the result of small selective effects (Sharp et al. 1993; Hartl, Moriyama, and Sawyer 1994; Akashi 1995; Chiapello et al. 1998). In these species, including *Drosophila melanogaster* and *Caenorhabditis elegans*, highly expressed genes preferentially use optimal codons, corresponding to the most abundant tRNAs in cells, because of weak selection for translational efficiency (Shields et al. 1988; Stenico, Lloyd, and Sharp 1994; Moriyama and Powell 1997; Duret and Mouchiroud 1999; Duret 2000). But in such species, it is also recognized that mutation pressure can partly be responsible for variations in codon bias across the genome (Kliman and Hey 1994; Akashi, Kliman, and Eyre-Walker 1998). Codon bias is positively correlated with recombination rate in *D. melanogaster* (Kliman and Hey 1993; Comeron, Kreitman, and Aguadé 1999; Marais, Mouchiroud, and Duret 2001) and *C. elegans* (Marais, Mouchiroud, and Duret 2001).

Two models have been proposed to explain this observation.

The first model proposes that the positive correlation between codon bias and recombination rate is caused by Hill-Robertson interference (HRI) (Kliman and Hey 1993; Comeron, Kreitman, and Aguadé 1999; McVean and Charlesworth 2000). HRI leads to a decrease of selection efficacy. This is because the 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 other loci (Hill and Robertson 1966; Felsenstein 1974). Simulation studies suggest that the effect of genetic linkage should be particularly damaging in the case of weak selection, such as selection acting on codon usage (Li 1987; Comeron, Kreitman, and Aguadé 1999; McVean and Charlesworth 2000).

The second model proposes that the positive correlation between codon bias and recombination rate is a byproduct of mutational bias variations (MBV) associated with recombination (Marais, Mouchiroud, and Duret 2001). Consistent with this model, in *D. melanogaster* and *C. elegans* the G+C content of both noncoding DNA and synonymous sites correlates positively with recombination rate (Marais, Mouchiroud, and Duret 2001). In the *D. melanogaster* subgroup, local changes in crossing-over frequencies between species are correlated with changes in MBV (Takano-Shimizu 2001). Because most of the optimal codons end in G or C in both *D. melanogaster* and *C. elegans* (Shields et al. 1988; Stenico, Lloyd, and Sharp 1994; Duret and Mouchiroud 1999), the high frequency of optimal codons observed in regions of high recombination may be the result of MBV associated with recombination (Marais, Mouchiroud, and Duret 2001). A positive correlation between G+C content and recombination has also been observed in other organisms, such as yeast (Baudat and Nicolas

¹ Both authors contributed equally to this work.

Abbreviations: Fop, frequency of optimal codons; R_s , Spearman's rank correlation coefficient; MBV, mutational bias variations; HRI, Hill-Robertson interference; Fop-MBV, Fop corrected for MBV; r , recombination rate; m , mutation rate; N , effective population size; L , number of selected sites; s , selection coefficient, CDS, coding sequence; Fop-max, Fop value expected with independent selected sites; EST, expressed sequence tag.

Key words: codon usage, recombination, mutation patterns, Hill-Robertson interference, *Drosophila*, *Caenorhabditis*.

Address for correspondence and reprints: Gabriel Marais, Laboratoire “Biométrie et biologie évolutive,” UMR CNRS 5558, Université Claude Bernard Lyon 1, 43 Bvd du 11 novembre 1918, 69622 Villeurbanne, France. E-mail: marais@biomserv.univ-lyon1.fr.

Mol. Biol. Evol. 19(9):1399–1406. 2002

© 2002 by the Society for Molecular Biology and Evolution. ISSN: 0737-4038

1997; Gerton et al. 2000), mouse (Perry and Ashworth 1999), and human (Eyre-Walker 1993; Eisenbarth et al. 2000; Fullerton, Bernardo Carvalho, and Clark 2001; Yu et al. 2001). In such eukaryotic organisms, the recombination machinery induces genetic conversion between parental chromosomes during meiosis (Smith and Nicolas 1998). Experimental evidence in mammals suggests that genetic conversion associated with recombination favors the copy of the most GC-rich sequence over the other (Brown and Jiricny 1988; Bill et al. 1998). Biased gene conversion might explain why MBV are associated with recombination in many organisms (Galtier et al. 2001).

Recently, both models have been tested in *C. elegans* and *D. melanogaster* by considering separately codons ending in G or C and codons ending in A or U (Marais, Mouchiroud, and Duret 2001). In both invertebrates, the frequency of GC-ending codons correlates positively with recombination rate, and the frequency of AU-ending codons correlates negatively with recombination rate, in agreement with the MBV model but not with the HRi model. Thus, the positive correlation between codon bias and recombination rate is mainly caused by MBV in *C. elegans* and *D. melanogaster* (Marais, Mouchiroud, and Duret 2001). An important question remains: is it possible to detect HRi on codon usage in *C. elegans* and *D. melanogaster* once the effect of MBV has been accounted for?

Introns are often considered good indicators of mutation patterns (Kliman and Hey 1993, 1994; Akashi, Kliman, and Eyre-Walker 1998). Thus, in our previous work, we used introns as indicators of MBV, but we failed to detect any HRi on codon usage (Marais, Mouchiroud, and Duret 2001). However, introns may be poor indicators of MBV affecting synonymous sites in such compact genomes as *D. melanogaster* and *C. elegans*. Because selection on codon usage is not expected to act on lowly expressed genes, we used the synonymous sites of lowly expressed genes to account for the effect of MBV on codon bias. We measured the differences between codon biases of highly expressed genes and their lowly expressed neighbors. This measure of codon bias should therefore be independent of the MBV occurring at synonymous sites. In *D. melanogaster* we find that HRi probably affects selection on codon usage of genes located in regions of very low recombination (<1 cM/Mb). Under the assumption that highly expressed genes are representative of the genes experiencing selection on codon usage, only 4% of genes are affected by less effective selection on codon usage because of HRi in this species. In *C. elegans* we do not find any evidence for the effect of recombination on selection for codon bias. We suggest that the correlation between codon bias and recombination rate is a consequence of MBV in this species. Computer simulations indicate that HRi only affects selection on codon usage when the local recombination rate is below the mutation rate. This prediction of the model is consistent with our data and the current estimate of the mutation rate in *D. melanogaster*. The case of *C. elegans*, which is highly self-fertilizing, is discussed. Finally, our results suggest that

HRi is a minor determinant of variations in codon bias across the genome.

Materials and Methods

Genomic Data Analysis

The sequence data are from the complete genomes of *C. elegans* (Release May 2, 1999 downloaded by FTP at ftp://ncbi.nlm.nih.gov/genbank/genomes/C_elegans/; The *C. elegans* Sequencing Consortium 1998) and *D. melanogaster* (Release October 2, 2000 downloaded by FTP at <http://www.fruitfly.org/sequence/download.html>; Adams et al. 2000). The expression level was determined with a method based on expressed sequence tags (ESTs) counting (Duret and Mouchiroud 1999). Genes were classified according to the number of matching ESTs. The top 30% of genes with ESTs were classified as highly expressed genes (for *C. elegans*, $n = 1,768$ with more than 17 detected ESTs; for *D. melanogaster*, $n = 2,399$ with more than 12 detected ESTs). Genes without ESTs were classified as lowly expressed genes (for *C. elegans*, $n = 9,392$; for *D. melanogaster*, $n = 5,219$). Other genes are considered moderately expressed (for *C. elegans*, $n = 4,034$; for *D. melanogaster*, $n = 6,132$). Recombination rate (cM/Mb) was estimated with a previously described procedure (Kliman and Hey 1993; Barnes et al. 1995; Comeron, Kreitman, and Aguadé 1999; Marais, Mouchiroud, and Duret 2001). For each chromosome (or chromosome arm in *D. melanogaster*), we chose the least-order polynomial that fits the data with $R^2 \geq 0.99$. Recombination rate as a function of chromosomal location is estimated by taking the derivative of the polynomial curve. For *C. elegans*, 780 markers localized both in genetic maps and in whole genome sequences have been used (available at <http://wormbase.sanger.ac.uk>), and third- to fifth-order polynomial curves are used to model the relationship between genetic positions and physical positions. The wild populations of *C. elegans* are mainly self-fertile hermaphrodites and should have less recombination than the laboratory strains used to construct the genetic maps. In this species, we point out that recombination rates should be accurate relative to each other, but not as an absolute value. For *D. melanogaster*, 898 markers have been used to estimate recombination rate (available at <http://flybase.bio.indiana.edu>), and fourth-order polynomial curves are used to model the relationship between genetic positions and physical positions (except for the left arm of the third chromosome, where we retained the second-order polynomial because the use of a higher-order polynomial did not increase R^2 , which is 0.98). The recombination rate on the fourth chromosome was considered to be zero. Genes are classified into 10 classes of recombination rate with nearly 10% of the total number of genes per class for both *C. elegans* and *D. melanogaster* (except for 0–0.25 and 0.25–0.5 classes in *D. melanogaster*, 5% each). We measured codon bias by the frequency of optimal codons (Fop): Fop ranges from 0.33, when codon usage is uniform, to 1, when genes use only optimal codons (Stenico, Lloyd, and Sharp 1994; Duret and Mouchiroud 1999).

Random Sampling of the Data Set

To resolve the problem of the covariations of gene length and recombination rate, we forced the distribution of gene length to be the same for the different classes of recombination rate for both lowly and highly expressed genes. We chose the distribution of gene length of the recombination rate class with the smallest sample size among lowly and highly expressed genes to be the reference distribution of gene length for all other recombination rate classes for both lowly and highly expressed genes. For *C. elegans*, this distribution corresponds to 23% of genes with coding sequence (CDS) length <1,000 nucleotides, 18% of genes with CDS length of 1,000–1,750 nucleotides, and 59% of genes with CDS length >1,750 nucleotides. For *D. melanogaster* this distribution corresponds to 18% of genes with CDS length <800 nucleotides, 21% of genes with CDS length of 800–1,550 nucleotides, and 61% of genes with CDS length >1,550 nucleotides. We generated 10 new data sets by random sampling of genes in each class of sequence length for each class of recombination for both lowly and highly expressed genes. In *D. melanogaster*, $n = 4,159$ for each data set corrected for gene length variations; in *C. elegans*, $n = 3,100$ for each data set corrected for gene length variations.

Computer Simulations

The simulation process is close to that of previous simulations studies of HRi (Li 1987; Comeron, Kreitman, and Aguadé 1999; McVean and Charlesworth 2000): we assumed that each individual is represented by L biallelic sites (e.g., optimal and nonoptimal codons). The haploid population size is N . If not specified, the mutation rate from nonoptimal toward optimal codons is u , the reverse mutation rate is $v = 2u$ leading to an equilibrium value of 0.33 without selection (Fop = 0.33 when codon usage is uniform), and the global mutation rate (number of mutation per site per generation) is $m = u(1 - \text{Fop}) + v\text{Fop}$. The number of mutations follows a Poisson distribution of mean NLu and NLv . The number of crossing-overs per generation also follows a Poisson distribution of mean NLr where r is the recombination rate (number of crossing-over per site per generation). The N individuals of the next generation are randomly chosen by multinomial sampling among the N individuals of the present generation, given their relative fitness in the population. The absolute fitness of a sequence with i optimal sites is given by $(1 + s)^i$, which is equivalent to negative selection on nonoptimal codons, given a simple transformation of selection coefficient s (Piganeau et al. 2001). The process is run for $4/(u + v)$ generations to reach equilibrium. The mean and variance of the equilibrium optimal codons frequency are calculated from 100 values checked every $2N$ generations, and each simulation is run at least four times. Without linkage between selected sites, selection efficiency is known to depend on the scaled mutation rates Nu and Nv and selection coefficients Ns (Li 1987). In the rest of the text, the Fop value expected without linkage between selected sites is referred as Fop-max.

Table 1
The Correlation Between the G + C Content of Noncoding DNA and the Recombination Rate in *D. melanogaster* and *C. elegans*

	<i>D. melanogaster</i>	<i>C. elegans</i>
Intron sequences >200 bp.....	0.048*** (7,340) [0.05*** (6,948)]	0.263*** (13,448)
5' flanking regions	0.029*** (14,092) [0.032*** (13,235)]	0.099*** (15,500)
3' flanking regions	0.026** (14,081) [0.024* (13,129)]	0.115*** (15,724)

NOTE.—Only large intron sequences were selected because short intron sequences are not expected to evolve neutrally (see text). Main values correspond to the Spearman's rank correlation coefficient (R_s). * $P < 0.05$, ** $P < 0.005$, *** $P < 0.0005$. () = sample size. [] = without subtelomeric regions.

Under complete linkage, the selection efficiency depends on the scaled mutation rates NLu and NLv and Ns (McVean and Charlesworth 2000).

Results and Discussion

In *D. melanogaster* and *C. elegans* we have previously shown that the G+C content of noncoding DNA positively correlates with the recombination rate, suggesting that MBV varies with recombination in those organisms (Marais, Mouchiroud, and Duret 2001). Here we confirm these results with another estimate of recombination rate (see table 1). Note that the positive correlation between the G+C content of noncoding DNA and the recombination rate is weak but statistically significant and comparable to the positive correlation between the frequency of optimal codons (Fop) and the recombination rate (in *D. melanogaster*: $R_s = 0.058$ with $P < 10^{-4}$, $n = 13,750$; in *C. elegans*: $R_s = 0.105$ with $P < 10^{-4}$, $n = 15,194$). In *D. melanogaster* the association between the G+C content of noncoding DNA and the recombination rate remains unchanged when subtelomeric regions with controversial recombination rate estimates (Hey and Kliman 2002) are excluded (see table 1).

Therefore, HRi on selection on codon usage can only be detected once MBV have been accounted for. Introns are often considered good indicators of mutation patterns because their evolution is assumed to be neutral (Kliman and Hey 1993, 1994; Akashi, Kliman, and Eyre-Walker 1998). On the basis of this assumption, we used introns as indicators of MBV; thus, we computed the residuals of the regression between introns G+C content and codon bias, but we failed to detect any HRi on these residuals (Marais, Mouchiroud, and Duret 2001). However, most introns are short both in *C. elegans* (The *C. elegans* Sequencing Consortium 1998) and *D. melanogaster* (Adams et al. 2000). The base composition of these short introns is constrained by the presence of elements for splicing reaction (Fields 1990; Mount et al. 1992). Intergenic regions and introns at the first position in genes often contain regulatory elements of gene expression (Maroni 1994; Duret and Bucher 1997). Thus, most of the noncoding DNA may not really

be neutral in both *C. elegans* and *D. melanogaster*. Large introns that are not at the first position in genes are probably neutral; but these introns often contain transposable elements, which are often AT-rich (Shields and Sharp 1989; Lerat, Biéumont, and Capy 2000; Lerat, Capy, and Biéumont 2002). Thus, the G+C content of introns and intergenic regions may be poor indicators of MBV affecting synonymous sites (Duret and Hurst 2001). Genes with a low expression level are not expected to undergo selection on codon usage. Consistent with this, these genes have a weak codon bias (Shields et al. 1988; Stenico, Lloyd, and Sharp 1994; Duret and Mouchiroud 1999) and a high number of synonymous substitutions (Shields et al. 1988; Sharp and Li 1989; Powell and Moriyama 1997; but see Dunn, Bielawski, and Yang 2001). Thus, the codon bias of lowly expressed genes should solely reflect mutation patterns. We can therefore estimate the effect of MBV with the codon bias of lowly expressed genes. In contrast, genes with a very high expression level are expected to undergo selection on codon usage. Consistent with this, these genes have a highly biased codon usage (Shields et al. 1988; Stenico, Lloyd, and Sharp 1994; Duret and Mouchiroud 1999) and a low number of synonymous substitutions (Shields et al. 1988; Sharp and Li 1989; Powell and Moriyama 1997; but see Dunn, Bielawski, and Yang 2001). Thus, the codon bias of highly expressed genes should be affected by both selection and mutation patterns. For these genes, the correlation between codon bias and recombination rate should be a consequence of both MBV and HRi if any. HRi can be brought to the fore by comparing codon biases of highly expressed genes (HRi + MBV) with lowly expressed genes (MBV) for different recombination rates.

We studied the complete genomes of *C. elegans* (The *C. elegans* Sequencing Consortium 1998) and *D. melanogaster* (Adams et al. 2000). We measured codon bias by the frequency of optimal codons (Fop) (Stenico, Lloyd, and Sharp 1994; Duret and Mouchiroud 1999). For each highly expressed gene, we measured the average difference between its Fop and the Fop of its lowly expressed neighbors over an interval of 100 kb centered on the midpoint of the highly expressed gene. In this way, we removed the local effect of MBV on Fop of highly expressed genes. In figure 1, we show the residuals of Fop after the removal of the MBV effect on codon usage (noted Fop-MBV for Fop corrected for MBV) according to recombination rate. The overall relationship between Fop-MBV and recombination rate is clearly not linear (see fig. 1). In *D. melanogaster* we observed a weak but significant increase of Fop-MBV with recombination rate for highly expressed genes located in regions of recombination rate of 0–1 cM/Mb (Spearman's rank correlation coefficient $R_s = 0.129$ with $P = 0.0033$) and no relationship between Fop-MBV and recombination rate for the other highly expressed genes (1 to > 3.9 cM/Mb, $R_s = -0.019$ with $P = 0.32$). This observation suggests that codon usage of highly expressed genes located in regions with recombination rate under ~1 cM/Mb in *D. melanogaster* probably experiences HRi. The same is found for moderately expressed

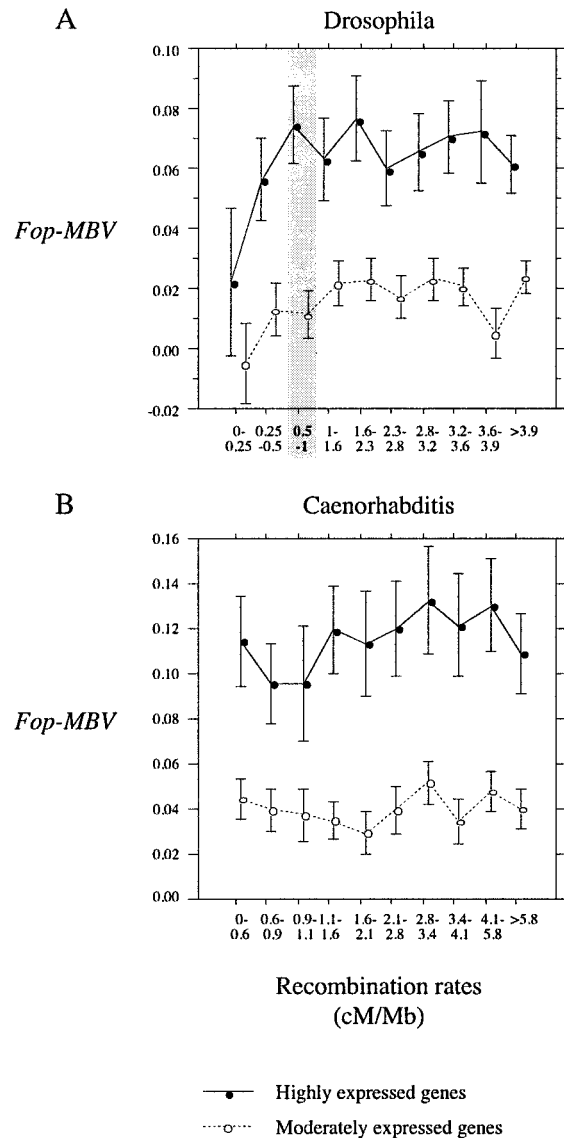


FIG. 1.—Relationship between codon bias corrected for MBV and recombination in (A) *D. melanogaster* and (B) *C. elegans*. To remove the local effect of MBV on Fop of each highly expressed gene, we measured the average difference between its Fop and the Fop of its lowly expressed neighbors over an interval of 100 kb centered on the midpoint of the highly expressed gene. The results are unchanged by using intervals ranging from 50 to 500 kb, by using other codon bias index, and by excluding subtelomeric regions with controversial recombination rates (Hey and Kliman 2002). The residuals of Fop after removal of the MBV effect on codon usage are noted Fop-MBV (for Fop corrected for MBV). The same approach has been applied to moderately expressed genes. Error bars correspond to the 95% interval.

genes, although variations in Fop-MBV induced by HRi tend to be weaker (see fig. 1). For these genes, variations in Fop-MBV in regions of recombination rate of 0–1 cM/Mb are not significant ($R_s = 0.021$ with $P = 0.46$). Thus, we do not consider them in the rest of the analysis. In *C. elegans* the relationship between Fop-MBV and recombination rate for highly expressed genes is not convincing, although there is a global correlation between the two parameters ($R_s = 0.064$ with $P < 0.0075$). For moderately expressed genes, the relationship is not

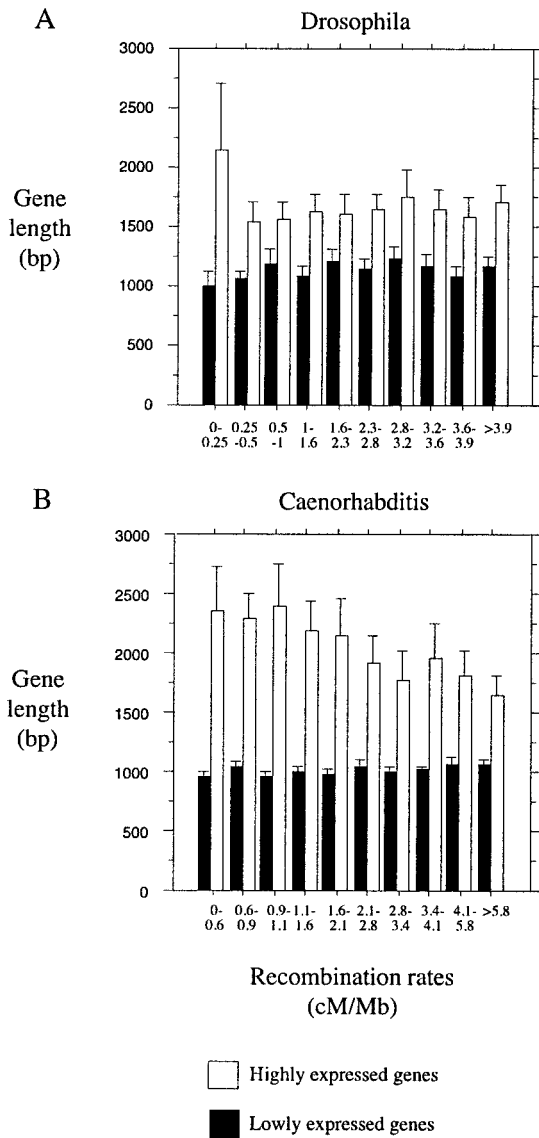


FIG. 2.—Relationship between gene length (bp) and recombination in (A) *D. melanogaster* and (B) *C. elegans*. Error bars correspond to the 95% interval.

convincing, and there is no global correlation (see fig. 1).

In *D. melanogaster* and *C. elegans*, codon bias is strongly correlated with gene length (Moriyama and Powell 1998; Comeron, Kreitman, and Aguadé 1999; Duret and Mouchiroud 1999; Marais and Duret 2001). Because the distribution of gene length is not random with respect to recombination rate in either genome (see fig. 2), we forced the distribution of gene length to be the same in each class of recombination rate. We generated 10 new data sets by random sampling of genes. In figure 3, we show the reevaluation of the relationship between Fop-MBV and recombination rate for *D. melanogaster* and *C. elegans* for 10 data sets corrected for gene length variations. We still observed a significant linear relationship between Fop-MBV and recombination rate for highly expressed genes located in regions of recombination rate of 0–1 cM/Mb, and no relation-

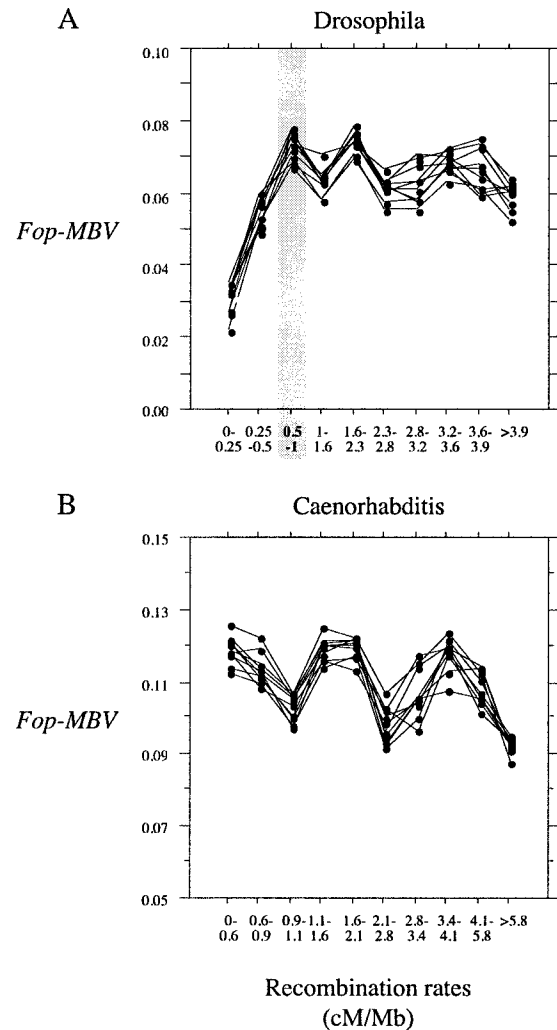


FIG. 3.—Relationship between codon bias corrected for MBV and recombination in (A) *D. melanogaster* and (B) *C. elegans*. The distributions of gene lengths are the same for the different classes of recombination rate for all the 10 data sets generated by random sampling. For *D. melanogaster* we found a significant linear relationship between Fop-MBV and recombination rate for highly expressed genes located in regions of recombination rate of 0–1 cM/Mb and no relationship for the other highly expressed genes for 7 of the 10 sampled data sets. For *C. elegans* we found no significant relationship between Fop-MBV and recombination rate for highly expressed genes for all the sampled data sets.

ship for the other highly expressed genes in *D. melanogaster*. We did not observe any significant relationship between Fop-MBV and recombination rate for all highly expressed genes in *C. elegans*. Thus, the relationship between Fop-MBV and recombination rate for highly expressed genes primarily detected in *C. elegans* (see fig. 1) is a byproduct of the variations of gene lengths along the genome. We have no evidence for the effect of recombination on selection for codon bias in this species. However, we detected HRi on codon usage in *D. melanogaster*. Note that variations in codon bias induced by HRi are only ~5% for highly expressed genes (and ~1.5% for moderately expressed genes, see fig. 1). Moreover, HRi influences codon usage only in regions of low recombination rate (<1 cM/Mb). These regions

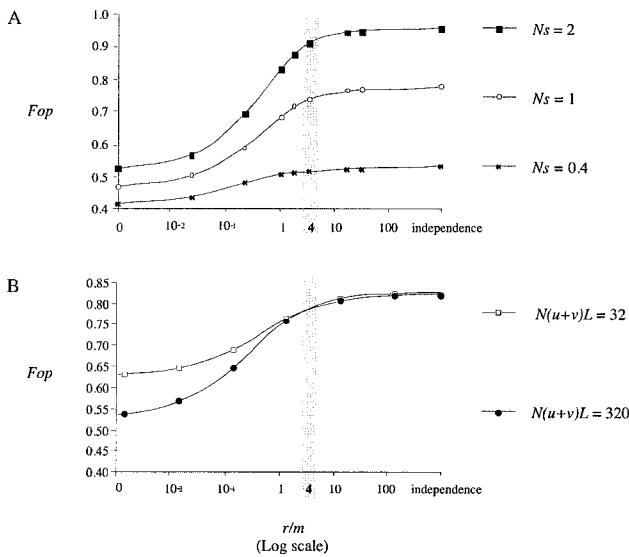


FIG. 4.—Simulation study of HRI on selection on codon usage. We observed that the selection efficiency depends on the ratio of the recombination rate (r) over the mutation rate (m). The r/m ratio above which 95% of Fop-max (see text) is reached, is obtained for $r \geq 4m$. This did not change with (A) different N_s (with $Nm = 0.01$, $N(u + v)L = 320$, $u/(u + v) = 0.33$) and (B) different $N(u + v)L$ (with $Nm = 0.01$, $N_s = 1$, $u/(u + v) = 0.4$). See *Materials and Methods* for the correspondence between u , v , and m . Fop values for $r/m > 4$ with $N(u + v)L = 3,200$ are not available because of nonpractical simulation times. Data on synonymous polymorphism in *Drosophila* species gave $N_s = 2$ (Akashi 1995), although these values may be underestimated (Andolfatto and Przeworski 2000; Andolfatto 2001). A rough estimate of NmL in *C. elegans* is 20 (number of highly expressed genes per chromosome = 1,768/6, mean length of highly expressed genes = 688 codons, and $Nm = 0.0001$ derived from Koch et al. 2000) and 3,375 in *D. melanogaster* (number of highly expressed genes per chromosome = 2,443/4, mean length of highly expressed genes = 551 codons, and $Nm = 0.01$ [McVean and Charlesworth 2000]).

contain 20.1% of the total number of genes in *D. melanogaster*: 7.2% are lowly expressed genes, 9.1% are moderately expressed genes, and 3.8% are highly expressed genes. However, lowly expressed genes can be excluded because they probably do not undergo selection on codon usage. Moderately expressed genes can also be excluded because of the limited impact of HRI on codon bias of these genes. Thus, few genes ($\sim 4\%$) are affected by less effective selection on codon usage in this species. These genes are located in genomic regions corresponding mainly to the fourth chromosome and to the subtelomeric and pericentromeric regions of the other chromosomes (as defined in Kliman and Hey 1993). Thus, HRI is a minor determinant of variations in codon bias across the genome.

The relationship between selection efficiency on codon usage and recombination was also investigated by simulation. In figure 4, we show the nonlinear relationship between HRI and the ratio of the recombination rate (r) over the mutation rate (m). For a lower recombination rate ($r < 4m$, from our simulations), recombination increases selection efficiency. For a higher recombination rate, associations between the alleles are broken down sufficiently fast so that they behave as if there were independent: the mutation-selection equilibrium optimal codon frequency reaches 95% of the value with

independent codons (Fop-max). Thus, we show through computer simulations that selection efficiency depends on recombination for a reduced range of recombination rates, depending on the r/m ratio. Furthermore, our simulations enable us to estimate that 95% of Fop-max is reached if $r \geq 4m$. This does not change for different N_s and $N(u + v)L$ (see fig. 4). In *D. melanogaster* the recombination rate above which Fop no longer increases with recombination is 1 cM/Mb = 10^{-8} recombinations per site per generation (see fig. 1). Using the formula $r = 4m$, this gives a mean mutation rate equal to 2.5×10^{-9} mutations per site per generation, which is consistent with current estimates of the mean mutation rate in this species (Drake et al. 1998; Keightley and Eyre-Walker 1999). In *C. elegans* the current estimates of the mean mutation rate are known to be lower (Drake et al. 1998). Accordingly, HRI should have an effect in regions where the recombination rate is lower than 1 cM/Mb in *C. elegans*, which is consistent with our failure to detect genes experiencing the effect of recombination on codon bias in this species. Although the sex ratio in the wild populations of *C. elegans* is largely unknown, they are thought to be mainly self-fertile hermaphrodites and should have less recombination than the laboratory strains used to construct genetic maps. Recombination rates should be accurate relative to each other but are probably overestimated (Nordborg 2000). Thus, in *C. elegans* the expected limit $r = 4m$ cannot be easily compared with our recombination data; hence, the absence of a relationship between Fop-MBV and recombination rates reflects either that the whole genome experiences HRI or that the whole genome does not experience HRI. As the selection strengths necessary to have a high codon bias in a genome composed of totally linked sites should be very important, it is more likely that the genome of *C. elegans* does not experience HRI; but only a comparison between a closely related and cross-fertilizing nematode species such as *Caenorhabditis remanei* (Haag and Kimble 2000) may allow discrimination between those two hypotheses. Unfortunately, the sequence data available for this species are poor. Thus, our analysis suggests that in *C. elegans* genes could be concentrated in regions of relatively low recombination rate without suffering of HRI as observed in the pericentromeric regions (Barnes et al. 1995; The *C. elegans* Sequencing Consortium 1998). As a conclusion, our results suggest that HRI is a minor determinant of variations in codon bias across the genome because it has small effects and influences few genes in *D. melanogaster*, and it entails no detectable variation in *C. elegans*.

Acknowledgments

Special thanks to Laurent Duret and Adam Eyre-Walker for helpful comments on the manuscript and friendly support. We also thank Vincent Daubin, Christian Gautier, Dominique Mouchiroud, and Stephen Wright for stimulating discussions. We are grateful to Roland Westrelin and Bernard Tourancheau for their help in the optimization of the simulation program. This

work is supported by the Ministère de la Recherche et de l'Enseignement Supérieur and the French bioinformatics programme.

LITERATURE CITED

- ADAMS, M. D., S. E. CELNIKER, R. A. HOLT et al. (95 co-authors). 2000. The genome sequence of *Drosophila melanogaster*. *Science* **287**:2185–2195.
- AKASHI, H. 1995. Inferring weak selection from patterns of polymorphism and divergence at “silent” sites in *Drosophila* DNA. *Genetics* **139**:1067–1076.
- AKASHI, H., R. M. KLIMAN, and A. EYRE-WALKER. 1998. Mutation pressure, natural selection, and the evolution of base composition in *Drosophila*. *Genetica* **102/103**:49–60.
- ANDOLFATTO, P. 2001. Contrasting patterns of X-linked and autosomal nucleotide variation in *Drosophila melanogaster* and *Drosophila simulans*. *Mol. Biol. Evol.* **18**:279–290.
- ANDOLFATTO, P., and M. PRZEWORSKI. 2000. A Genome-wide departure from the standard neutral model in natural populations of *Drosophila*. *Genetics* **156**:257–268.
- BARNES, T. M., Y. KOHARA, A. COULSON, and S. HEKIMI. 1995. Meiotic recombination, noncoding DNA and genomic organization in *Caenorhabditis elegans*. *Genetics* **141**:159–179.
- BAUDAT, F., and A. NICOLAS. 1997. Clustering of meiotic double-strand breaks on yeast chromosome III. *Proc. Natl. Acad. Sci. USA* **94**:5213–5218.
- BILL, C. A., W. A. DURAN, N. R. MISELIS, and J. A. NICKOLOFF. 1998. Efficient repair of all types of single-base mismatches in recombination intermediates in Chinese Hamster ovary cells. Competition between long-patch and G-T glycosylase-mediated repair of G-T mismatches. *Genetics* **149**:1935–1943.
- BROWN, T. C., and J. JIRICNY. 1988. Different base/base mismatches are corrected with different efficiencies and specificities in monkey kidney cells. *Cell* **54**:705–711.
- BULMER, M. 1991. The selection-mutation-drift theory of synonymous codon usage. *Genetics* **129**:897–907.
- THE C. ELEGANS SEQUENCING CONSORTIUM. 1998. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* **282**:2012–2018.
- CHIAPELLO, H., F. LISACEK, M. CABOCHE, and A. HENAUT. 1998. Codon usage and gene function are related in sequences of *Arabidopsis thaliana*. *Gene* **209**:GC1–GC38.
- COMERON, J. M., M. KREITMAN, and M. AGUADÉ. 1999. Natural selection on synonymous sites is correlated with gene length and recombination in *Drosophila*. *Genetics* **151**:239–249.
- DRAKE, J. W., B. CHARLESWORTH, D. CHARLESWORTH, and J. CROW. 1998. Rates of spontaneous mutation. *Genetics* **148**:1667–1686.
- DUNN, K. A., J. P. BIELAWSKI, and Z. YANG. 2001. Substitution rates in *Drosophila* nuclear genes: implications for translational selection. *Genetics* **157**:295–305.
- DURET, L. 2000. tRNA gene number and codon usage in *C. elegans* genome are co-adapted for optimal translation of highly expressed genes. *Trends Genet.* **16**:287–289.
- DURET, L., and P. BUCHER. 1997. Searching for regulatory elements in human noncoding sequences. *Curr. Opin. Struct. Biol.* **7**:399–406.
- DURET, L., and L. D. HURST. 2001. The elevated GC content at exonic third sites is not evidence against neutralist models of isochore evolution. *Mol. Biol. Evol.* **18**:757–762.
- DURET, L., and D. MOUCHIROUD. 1999. Expression pattern and, surprisingly, gene length shape codon usage in *Caenorhabditis*, *Drosophila*, and *Arabidopsis*. *Proc. Natl. Acad. Sci. USA* **96**:4482–4487.
- EISENBARTH, I., G. VOGEL, W. KRONE, W. VOGEL, and G. ASSUM. 2000. An isochore transition in the NF1 gene region coincides with a switch in the extent of linkage disequilibrium. *Am. J. Hum. Genet.* **67**:873–880.
- EYRE-WALKER, A. 1993. Recombination and mammalian genome evolution. *Proc. R. Soc. Lond. B* **252**:237–243.
- FELSENSTEIN, J. 1974. The evolutionary advantage of recombination. *Genetics* **78**:737–756.
- FIELDS, C. 1990. Information content of *Caenorhabditis elegans* splice site sequences varies with intron length. *Nucleic Acids Res.* **18**:1509–1512.
- FULLERTON, S. M., A. BERNARDO CARVALHO, and A. G. CLARK. 2001. Local rates of recombination are positively correlated with GC content in the human genome. *Mol. Biol. Evol.* **18**:1139–1142.
- GALTIER, N., G. PIGANEAU, D. MOUCHIROUD, and L. DURET. 2001. GC-content evolution in mammalian genomes: the biased gene conversion hypothesis. *Genetics* **159**:907–911.
- GERTON, J. L., J. DERISI, R. SHROFF, M. LICHTEN, P. O. BROWN, and T. D. PETES. 2000. Global mapping of meiotic recombination hotspots and coldspots in the yeast *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA* **97**:11383–11390.
- HAAG, E. S., and J. KIMBLE. 2000. Regulatory elements required for development of *Caenorhabditis elegans* hermaphrodites are conserved in the *tra-2* homologue of *C. remanei* a male/female sister species. *Genetics* **155**:105–116.
- HARTL, D. L., E. N. MORIYAMA, and S. A. SAWYER. 1994. Selection intensity for codon bias. *Genetics* **138**:227–234.
- HEY, J., and R. M. KLIMAN. 2002. Interactions between natural selection, recombination and gene density in the genes of *Drosophila*. *Genetics* **160**:595–608.
- HILL, W. G., and A. ROBERTSON. 1966. The effect of linkage on limits to artificial selection. *Genet. Res.* **8**:269–294.
- KEIGHTLEY, P. D., and A. EYRE-WALKER. 1999. Deleterious mutations and the evolution of sex. *Science* **290**:331–333.
- KLIMAN, R. M., and J. HEY. 1993. Reduced natural selection associated with low recombination in *Drosophila melanogaster*. *Mol. Biol. Evol.* **10**:1239–1258.
- . 1994. The effects of mutation and natural selection on codon bias in the genes of *Drosophila*. *Genetics* **137**:1049–1056.
- KOCH, R., H. G. VAN LUENEN, M. VAN DER HORST, K. L. THIJSSSEN, and R. H. PLASTERK. 2000. Single nucleotide polymorphisms in wild isolates of *Caenorhabditis elegans*. *Genome Res.* **10**:1690–1696.
- LERAT, E., C. BIÉMONT, and P. CAPY. 2000. Codon usage and the origin of P elements. *Mol. Biol. Evol.* **17**:467–468.
- LERAT, E., P. CAPY, and C. BIÉMONT. 2002. Codon usage by transposable elements and their host genes in five species. *J. Mol. Evol.* **54**:625–637.
- LI, W.-H. 1987. Models of nearly neutral mutations with particular implications for nonrandom usage of synonymous codons. *J. Mol. Evol.* **24**:337–345.
- MARAIS, G., and L. DURET. 2001. Synonymous codon usage, accuracy of translation, and gene length in *Caenorhabditis elegans*. *J. Mol. Evol.* **52**:275–280.
- MARAIS, G., D. MOUCHIROUD, and L. DURET. 2001. Does recombination improve selection on codon usage? Lessons from nematode and fly complete genomes. *Proc. Natl. Acad. Sci. USA* **98**:5688–5692.
- MARONI, G. 1994. The organization of *Drosophila* genes. *DNA Seq.* **4**:347–354.

- MCVEAN, G. A. T., and B. CHARLESWORTH. 2000. The effects of Hill-Robertson interference between weakly selected mutations on patterns of molecular evolution and variation. *Genetics* **155**:929–944.
- MORIYAMA, E. N., and J. R. POWELL. 1997. Codon usage bias and tRNA abundance in *Drosophila*. *J. Mol. Evol.* **45**:514–523.
- . 1998. Gene length and codon usage bias in *Drosophila melanogaster*, *Saccharomyces cerevisiae* and *Escherichia coli*. *Nucleic Acids Res.* **26**:3188–3193.
- MOUNT, S. M., C. BURKS, G. HERTZ, G. D. STORMO, O. WHITE, and C. FIELDS. 1992. Splicing signals in *Drosophila*: intron size, information content, and consensus sequences. *Nucleic Acids Res.* **20**:4255–4262.
- NORDBORG, M. 2000. Linkage disequilibrium, gene trees and selfing: an ancestral recombination graph with partial self-fertilization. *Genetics* **154**:923–929.
- PERRY, J., and A. ASHWORTH. 1999. Evolutionary rate of a gene affected by chromosomal position. *Curr. Biol.* **9**:987–989.
- PIGANEAU, G., R. WESTRELIN, B. TOURANCHEAU, and C. GAUTIER. 2001. Multiplicative versus additive selection in relation to genome evolution: a simulation study. *Genet. Res.* **78**:171–175.
- POWELL, J. R., and E. N. MORIYAMA. 1997. Evolution of codon usage bias in *Drosophila*. *Proc. Natl. Acad. Sci. USA* **94**:7784–7790.
- SHARP, P. M., and W. H. LI. 1989. On the rate of DNA sequence evolution in *Drosophila*. *J. Mol. Evol.* **28**:398–402.
- SHARP, P. M., M. STENICO, J. F. PEDEN, and A. T. LLOYD. 1993. Codon usage: mutational bias, translational selection, or both? *Biochem. Soc. Trans.* **21**:835–841.
- SHIELDS, D. C., and P. M. SHARP. 1989. Evidence that mutation patterns vary among drosophila transposable elements. *J. Mol. Biol.* **207**:843–846.
- SHIELDS, D. C., P. M. SHARP, D. G. HIGGINS, and F. WRIGHT. 1988. “Silent” sites in *Drosophila* genes are not neutral: evidence of selection among synonymous codons. *Mol. Biol. Evol.* **5**:704–716.
- SMITH, K. N., and A. NICOLAS. 1998. Recombination at work for meiosis. *Curr. Opin. Genet. Dev.* **8**:200–211.
- STENICO, M., A. T. LLOYD, and P. M. SHARP. 1994. Codon usage in *Caenorhabditis elegans*: delineation of translational selection and mutational biases. *Nucleic Acids Res.* **22**:2437–2446.
- TAKANO-SHIMIZU, T. 2001. Local changes in GC/AT substitution biases and in crossover frequencies on *Drosophila* chromosomes. *Mol. Biol. Evol.* **18**:606–619.
- YU, A., C. ZHAO, Y. FAN et al. (11 co-authors). 2001. Comparison of human genetic and sequence-based physical maps. *Nature* **409**:951–953.

WOLFGANG STEPHAN, reviewing editor

Accepted March 14, 2002