Current Biology *16*, 2053–2057, October 24, 2006 DOI 10.1016/j.cub.2006.08.067

Report

Selection Intensity on Preferred Codons Correlates with Overall Codon Usage Bias in *Caenorhabditis remanei*

Asher D. Cutter[1,2,*] and Brian Charlesworth[1]
[1] Institute of Evolutionary Biology
School of Biological Sciences
University of Edinburgh
Edinburgh EH9 3JT
United Kingdom

Summary

Adaptive codon usage provides evidence of natural selection in one of its most subtle forms: a fitness benefit of one synonymous codon relative to another. Codon usage bias is evident in the coding sequences of a broad array of taxa, reflecting selection for translational efficiency and/or accuracy as well as mutational biases. Here, we quantify the magnitude of selection acting on alternative codons in genes of the nematode *Caenorhabditis remanei*, an outcrossing relative of the model organism *C. elegans*, by fitting the expected mutation-selection-drift equilibrium frequency distribution of preferred and unpreferred codon variants to the empirical distribution. This method estimates the intensity of selection on synonymous codons in genes with high codon bias as $N_e s = 0.17$, a value significantly greater than zero. In addition, we demonstrate for the first time that estimates of ongoing selection on codon usage among genes, inferred from nucleotide polymorphism data, correlate strongly with long-term patterns of codon usage bias, as measured by the frequency of optimal codons in a gene. From the pattern of polymorphisms in introns, we also infer that these findings do not result from the operation of biased gene conversion toward G or C nucleotides. We therefore conclude that coincident patterns of current and ancient selection are responsible for shaping biased codon usage in the *C. remanei* genome.

Results and Discussion

Codon usage bias describes the nonrandom incorporation of alternative synonymous codons into coding sequences, which can be driven both by neutral factors (mutation, genetic drift, biased gene conversion) and by natural selection [1, 2]. Strong patterns of codon usage bias pervade bacterial and yeast genomes [3, 4], and evidence of selection on codon usage bias is found in a variety of plants, arthropods, and nematodes [5–10]. Selection for translational efficiency and/or accuracy in highly expressed genes is the most likely explanation for adaptive codon usage bias, such that a greater abundance of codons in a sequence matching the most abundant tRNAs in the cell allows translation to occur more quickly and with less error [9]. The codons that are overrepresented in highly expressed genes are termed preferred or optimal codons [3]. Because only a very small fitness difference is likely to result from incorporating one synonymous codon rather than another, theory predicts that selection for codon usage bias requires a large effective population size for a population to respond to such weak selection (so that $N_e s \approx 1$, where N_e is the effective population size and s is the selection coefficient against an unpreferred codon at a particular site [1, 2]). However, relatively few studies have quantified the strength of selection favoring such preferred codons, and in eukaryotes such estimates are limited to species of *Drosophila* [11–17] and humans [18, 19].

Here, we quantify the selection intensity on codon bias from the frequency distribution of preferred and unpreferred codon variants in the nematode *Caenorhabditis remanei*, employing a little-used method that does not require information about the ancestral state of variant sites. Recent studies surveyed nucleotide polymorphism in this species [20–23], from which we extracted the status of variant sites with respect to whether they define preferred and unpreferred codons in nine genes, by using the *C. elegans* optimal codon table as a reference [7, 24]. We classify diallelic polymorphisms at synonymous sites into three categories, depending on whether the base difference results in (1) two different unpreferred codons encoding the amino acid (UU), (2) two different preferred codons encoding the amino acid (PP), or (3) one preferred and one unpreferred codon encoding the same amino acid (PU). Both the UU and PP cases probably involve neutral or nearly neutral changes, whereas selection may be able to distinguish between the alternative bases associated with a PU polymorphism. The *C. elegans* optimal codon table is probably appropriate for *C. remanei*, given that *C. briggsae* shares the same set of optimal codons with *C. elegans* [25], that *C. briggsae* and *C. remanei* are equally divergent from *C. elegans* [26], and that optimal codon identity evolves slowly in nematodes and other taxa [27]. However, these various congeners are too divergent at silent sites to permit outgroup polarization of intraspecific polymorphisms relative to their ancestral state, which forms the basis of the "ratio of polymorphism to divergence" (r_{pd}) approaches that have been used to estimate $N_e s$ in several species of *Drosophila* [12, 14]. Consequently, we employ an alternative method to identify the value of $N_e s$ that provides the best fit of the mutation-selection-drift equilibrium distribution of preferred codon variants [28] to that observed empirically.

The variant frequency distribution of preferred codons for PU polymorphisms is expected to be skewed toward high frequencies if they are subject to selection for translational efficiency and/or accuracy; otherwise, a neutral distribution is expected, i.e., the frequency distribution will be symmetrically U-shaped [28]. For the

*Correspondence: asher.cutter@utoronto.ca
[2] Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, Ontario M5S 3G5, Canada.

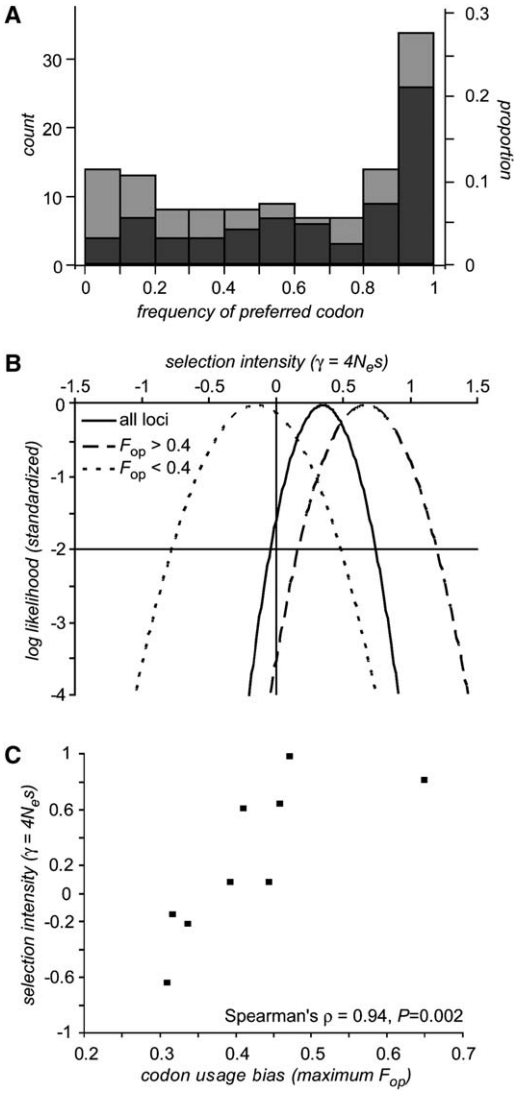

Figure 1. Selection on Codon Usage

(A) Frequency distribution of preferred codons at synonymous sites polymorphic for preferred and unpreferred codons, pooled across all loci (light gray, n = 122 variant sites) and for loci with F_{op} > 0.4 (dark gray, n = 75 variant sites).

(B) Log-likelihood curves for $\gamma = 4N_es$ from frequencies of preferred and unpreferred codons, standardized to make each maximum log-likelihood zero. The likelihood curve for the pooled sample is given in black, with dashed and dotted lines corresponding to sets of loci with high or low codon bias, respectively.

(C) Plot of maximum likelihood selection intensity estimates for each locus against the codon bias of the locus.

synonymous polymorphisms in *C. remanei*, preferred codon variants demonstrate a striking skew toward high frequencies (Figure 1A). This observation supports the hypothesis that ongoing selection shapes codon usage bias in *C. remanei*. Consequently, we fitted the frequency distribution of preferred codon variants to that expected under mutation-selection-drift equilibrium, in order to calculate maximum likelihood estimates for the scaled selection coefficient, $\gamma = 4N_es$ [28]. For all PU polymorphisms considered together, γ = 0.35, ranging between ~0 and 0.98 for individual loci (Table 1; Figure 1B). We also considered separately loci with high overall codon bias (F_{op} > 0.4) and with low codon bias (F_{op} < 0.4) (Figure 1B), for which a likelihood-ratio test detected significant heterogeneity in γ ($2\Delta \ln L$ = 4.08, df = 1, p = 0.043). Based on a 2-log-likelihood confidence limit, we demonstrate significant selection on preferred codons ($4N_es > 0$) for the subset of loci with high codon bias (Figure 1B; Table 2). Table 2 also shows that there is a much larger increase in ln*L* when fitting separate γ values for GC ↔ AT mutations in low- and high-biased coding sequences (see below), further supporting the existence of an effect of codon usage on N_es. Furthermore, the per-locus selection intensity estimates correlate strongly with the overall codon usage bias of genes (Figure 1C; Spearman's ρ = 0.94, p = 0.002). In other words, loci with stronger overall codon usage bias that has built up historically also exhibit stronger ongoing selection for the incorporation of preferred codons.

Table 1. Codon Bias Statistics

Locus[a]	Max F_{op}[b]	Min ENC[b]	Frequency of Preferred Variants: Mean	Frequency of Preferred Variants: Median	γ[c]	Max(ln*L*)[d]
p22	0.650	39.3	0.672	0.818	0.81	5.3
p17	0.472	40.0	0.705	0.875	0.98	19.8
p15	0.460	47.3	0.639	0.667	0.64	22.9
Cr-glp-1	0.445	61.0	0.517	0.542	0.08	94.5
p21	0.411	50.2	0.633	0.712	0.61	19.0
P24	0.394	50.3	0.518	0.419	0.08	36.0
Cr-odr-3	0.337	49.8	0.452	0.500	−0.22	65.6
Cr-fem-3	0.317	52.5	0.467	0.364	−0.15	171.8
p18	0.311	52.1	0.361	0.206	−0.64	−0.6
pooled			0.576	0.606	0.35	431.5

[a] Data from [20]–[23].

[b] Computed for sequenced regions only.

[c] Maximum likelihood estimate of selection intensity $\gamma = 4N_es$ on codon usage.

[d] Maximum log-likelihood value for estimation of γ.

The widespread operation of gene conversion biased toward AT → GC base replacement also could lead to a prevalence of high frequencies of preferred codon variants [29, 30], because preferred codons tend to have guanine or cytosine at third positions in *Caenorhabditis* (16 of 21 preferred codons end in G or C; [7, 24]). Consequently, we also assessed the frequency distribution of synonymous PU sites in coding regions with a GC ↔ AT polymorphism relative to GC ↔ AT polymorphic sites in noncoding regions. Consistent with the expectations of selection rather than biased gene conversion, GC ↔ AT polymorphisms at PU sites yielded an estimate of the selection intensity that is nominally greater than that observed for noncoding GC ↔ AT variants (γ = 0.30 versus −0.08), but not significantly so, according to the 2-log-likelihood confidence intervals (Table 2). Similarly, GC ↔ AT variants at other classes of sites that should not be subject to differential selection (PP and UU sites, all synonymous or PU sites in low F_{op} loci) show no evidence of the high γ values seen for PU sites in high F_{op} loci (Table 2). The negative sign of the estimated γ for GC variants in noncoding regions implies that, if anything, AT variants occur at higher frequency than GC variants, in contrast to reports for *D. melanogaster* and

Table 2. Selection Intensities Inferred for Different Classes of Sites

Site Class	Variant Type	Total Variant Sites	γ (2 ln*L* Interval)	Max ln*L*
PU sites	P/U	122	0.35 (−0.05–0.75)	431.5
	P/U (high F_{op})	75	0.67 (0.16–1.21)[a]	161.0
	P/U (low F_{op})	47	−0.14 (−0.78–0.49)	272.5
Synonymous sites	GC/AT	174	0.09 (−0.23–0.42)	651.7
	GC/AT (high F_{op})	98	0.47 (−0.17–0.71)	105.1
	GC/AT (low F_{op})	76	−0.12 (−0.63–0.37)	477.4
PU sites	GC/AT	108	0.30 (−0.12–0.72)	398.2
	GC/AT (high F_{op})	68	0.55 (0.02–1.10)[a]	153.5
	GC/AT (low F_{op})	40	−0.12 (−0.81–0.56)	245.9
Noncoding sites	GC/AT	157	−0.08 (−0.43–0.26)	1272.9
PP + UU sites	GC/AT	66	−0.23 (−0.78–0.30)	254.8

[a] Significant difference from zero.

D. simulans [29], a result that conflicts with a biased gene conversion interpretation of the patterns in *C. remanei*.

The expected proportion of sites containing preferred codons at mutation-selection-drift equilibrium can be approximated by the expression $E(x) = \frac{1}{1+\kappa e^{-\gamma}}$, where κ is the ratio of mutation rates from and to preferred codons [1, 2]. Although it is difficult to use this expression to infer γ, because of problems in estimating κ, the same relation should describe expected synonymous-site GC-content, $E(x')$, by letting κ' be the GC ↔ AT mutation bias and γ' the selection intensity on synonymous GC ↔ AT changes. In this case, κ' can be estimated from the GC content of noncoding sequence (GC_{nc}), assuming that it is neutrally evolving and at equilibrium; for the sample of genes considered here, average GC_{nc} = 0.33, so we infer $\kappa' = 2.0$. The value of γ' for GC variants is estimated from the maximum likelihood method to be 0.30 from the above-mentioned GC ↔ AT variants at PU sites (Table 2). With these estimates of κ' and γ', the expected average GC-content at synonymous sites, $E(x')$, is 40.3%, a value slightly less than the observed 41.4% mean GC-content at 4-fold degenerate sites. Because GC-biased gene conversion would cause underestimation of κ', biased gene conversion should cause the above calculations to predict greater than observed levels of GC at silent sites [14, 17]; the slightly greater observed than predicted equilibrium GC content thus lends additional support to the notion that biased gene conversion does not drive the observed patterns in *C. remanei*. Note that a similar method to that used here for estimating γ' has been used to quantify the intensity of biased gene conversion in *D. melanogaster* and *D. simulans* [29].

Among the sample of nine genes considered here, we found the selection intensity on preferred codons, $N_e s$, to average 0.09, and for the loci with strong overall codon usage bias, we compute a slightly higher value of $N_e s$ = 0.17. This value is somewhat smaller than the $N_e s$ values of 0.19, 0.38, 0.63, and 0.65 reported, respectively, for *D. melanogaster*, *D. virilis*, *D. miranda*, and *D. americana* [14, 15, 17] and for a combined analysis of several *Drosophila* species ($0.34 \leq N_e s \leq 0.78$) [16]. Even greater estimates of $N_e s$ have been reported for *D. simulans* ($N_e s \approx 2$) and *D. pseudoobscura* ($N_e s \approx 4$) [12, 13]. However, values of $N_e s > 1$ are expected to yield genes that only use preferred codons (which is not observed empirically), so it has been argued that the effects of departure from equilibrium may confound inferences of selection intensity on codon bias in these species [14]. Also, a reanalysis of *D. simulans* data yielded a substantially lower estimate of $N_e s$ = 0.69 [15], as did an analysis of multiple *Drosophila* species that includes *D. simulans* [16].

Several possible factors may explain the difference between the selection intensity estimates for *C. remanei* and *Drosophila*. First, it may simply reflect lower overall levels of codon bias among the sample of genes included in this study. Indeed, McVean and Vieira [15] report a 4-fold range of variation in selection intensity among loci for two species of *Drosophila*, and short versus long genes in *Drosophila* exhibit nearly a 2-fold difference in selection intensity on codon usage [16]. Second, variance in selection intensity across sites may cause underestimation of mean γ [17]. We extended our likelihood method to include a distribution of γ values across sites, with a second-order Taylor series approximation to obtain log-likelihood surfaces for both the mean and variance of γ. While including a variance term always increased the likelihood, the effects were nonsignificant and the location of the maximum with respect to mean γ was not affected. Third, the effective population sizes of *Drosophila* species may, in general, be greater than that of *C. remanei*, leading to a reduced value of the combined parameter $N_e s$ in the worm. However, estimates of N_e derived from measures of diversity (θ) and mutation rate imply comparable effective population sizes for *C. remanei* and *Drosophila* species [23]. Fourth, the magnitude of the selective advantage of incorporating preferred codons might be greater in flies than in worms. For example, if *Drosophila* have substantially longer larval and adult life spans in nature than do *C. remanei*, then the opportunity for selection on codon usage may be greater in flies. Alternatively, a shorter half-life of transcripts or proteins in *Drosophila* cells would require higher rates of transcription, which could result in added selection on codon usage—although we know of no data that speak to this hypothesis. The *D. melanogaster* genome also has fewer genes than *C. elegans*, but they are more frequently alternatively spliced [31, 32], which could result in greater per-locus transcription rates in *Drosophila* relative to *Caenorhabditis* if the transcription rate per nuclear genome is similar in the two groups.

The pattern of polymorphisms at synonymous sites demonstrates that selection differentiates among alternative codons in *C. remanei*. Specifically, among sites that are polymorphic for preferred and unpreferred codons, we observe that (1) the preferred codon is typically present at higher frequency and (2) the frequency distribution is more skewed for loci that are composed of a higher fraction of optimal codons, indicating that historical and ongoing patterns of selection are coincident. This does not appear to be a result of biased gene conversion toward G and C nucleotides, as shown by variants in noncoding regions not conforming to the pattern expected if biased gene conversion shapes base composition and codon usage. An estimate of the *C. remanei* effective population size of $N_e = 1.6 \times 10^6$, based on an application of the estimated *C. elegans* neutral mutation rate to *C. remanei* silent-site diversity [23], implies that we are able to detect a selection coefficient (s) on preferred codons of approximately 1.1×10^{-7}. These patterns illustrate how even weak natural selection can be recorded in genomes over both short and long evolutionary timescales by skewing allele frequencies in a population and by altering codon usage and base composition in genes.

Experimental Procedures

In this study, we used polymorphism information for nine nuclear loci from recent surveys of genetic diversity in *C. remanei* [20–23]. Six loci derive from uncharacterized genes that share putative orthologs with *C. elegans* genes Y25C1A.5, ZK430.1, E01G4.6, R160.7 (*lst-2*), T24D11.1, and D1005.1, given the arbitrary identifiers *p15*, *p24*, *p17*, *p18*, *p21*, and *p22*, respectively [23]. The regions are composed primarily of coding sequence, accounting for approximately a quarter of the coding portion of the corresponding genes, and range from 551 bp to 693 bp in length, sampled from 31 to 34 individuals (GenBank accessions DQ897370–DQ897565). The remaining three loci have been surveyed for diversity in samples of 11 to 12 individuals [20–22] and range in length from 426 bp partial sequence of *Cr-glp-1* to the >2 kbp full-length of *Cr-odr-3* and *Cr-fem-3* (GenBank accessions AF491520–AF491531, AY146567–AY146577, AY196906, AY142113, AY769072–AY769081). One nuclear locus (*Cr-tra-2*) also reported in [20] was excluded due to lack of sufficient variant sites. Overall, these loci have an average diversity at silent sites (θ_{si}) of 0.047 [23]. In addition, the pattern of polymorphisms in the sample reveals little evidence of population structure or demographic size changes [23]. We identified UU, PP, PU, and GC – AT variant sites and their frequencies in the samples of sequences by using DnaSP v. 4.10.04, where P denotes a preferred codon and U an unpreferred codon, and by applying the *C. elegans* codon usage table [33]. The numbers of PU sites per locus ranged from 7 to 35, with 122 PU variant sites in total. By using the program CodonW (J. Peden, http://codonw.sourceforge.net), we calculated the frequency of optimal codons (F_{op}) and the effective number of codons (ENC) from the sequenced region of each locus (i.e., not the entire predicted coding region of the gene), retaining the maximum value observed among the sequences of different individuals. Individuals differed very little in the observed values of F_{op} and ENC, and use of means or medians of these measures of codon bias did not qualitatively affect the analyses.

The theoretical distribution of preferred codon variants, $\phi(x)$, is described by the relation $\phi(x) = \frac{Ce^{\gamma x}}{x(1-x)}$, where x is the frequency of the preferred variant, $\gamma = 4N_e s$ (with additive selection assumed), and C is a constant such that $\int_{\frac{1}{2N}}^{1-\frac{1}{2N}} \phi(x)dx = 1$ is satisfied [28]. A Perl or Fortran program that calculates the likelihood surface for γ according to this method is available upon request. Summation of log-likelihood values for different genes permits the combination of data from multiple loci. Importantly, this approach is independent of mutational biases, provided that the population is at equilibrium [28]. Both this frequency distribution method and outgroup polarization methods may be influenced by departures from demographic equilibrium, although it is unknown by how much this might affect inferences of $N_e s$; however, population genetic statistics suggest that *C. remanei* approximates demographic equilibrium [23]. Violations of the model assumptions of independence among variant sites and the infinite sites model of mutation [34] could also compromise the inferences of $N_e s$. Pervasive recombination in *C. remanei* appears to be frequent enough to cause rapid decay of linkage disequilibrium within a locus [23], so sequence variants in *C. remanei* may roughly conform to the independence assumption. The high diversity (silent-site $\theta \sim 5\%$), and therefore large effective population size [23], for *C. remanei* facilitates the use of this intraspecific frequency distribution approach for even relatively few loci. However, the high synonymous and noncoding sequence diversity in this species also results in some sites with more than two segregating variants [23]. Sites that depart from the assumption of an infinite sites mutation process (i.e., sites with more than two variants) were excluded from the analysis.

The likelihood model was extended to included the possibility that γ values for different sites within a gene are drawn from distribution $\phi(\gamma)$, with mean $\bar{\gamma}$ and standard deviation σ. The likelihood expression for a given γ, $\ln L(\gamma)$, is integrated over the distribution of γ to get $\int \ln L\left(\gamma\right)\phi\left(\gamma\right)d\gamma \approx \ln L\left(\bar{\gamma}\right) + \frac{1}{2}\sigma^2(d^2 \ln L(\gamma)/d\gamma^2)_{\bar{\gamma}}$. After evaluating the derivatives, this approximation was used to obtain a log-likelihood surface for $\bar{\gamma}$ and σ. Because of the minor effect of this variance term on the likelihood estimates, the values for γ reported here assume $\sigma = 0$ to simplify the calculations.

Acknowledgments

We thank D. Charlesworth for valuable discussions relating to this work and two anonymous reviewers for comments on the manuscript. A.D.C. received support from the National Science Foundation with International Research Fellowship Program grant #0401897.

Received: June 20, 2006
Revised: August 23, 2006
Accepted: August 24, 2006
Published: October 23, 2006

References

1. Li, W.H. (1987). Models of nearly neutral mutations with particular implications for nonrandom usage of synonymous codons. J. Mol. Evol. *24*, 337–345.
2. Bulmer, M. (1991). The selection-mutation-drift theory of synonymous codon usage. Genetics *129*, 897–907.
3. Ikemura, T. (1982). Correlation between the abundance of yeast transfer RNAs and the occurrence of the respective codons in protein genes: differences in synonymous codon choice patterns of yeast and *Escherichia coli* with reference to the abundance of isoaccepting transfer RNAs. J. Mol. Biol. *158*, 573–597.
4. Merkl, R. (2003). A survey of codon and amino acid frequency bias in microbial genomes focusing on translational efficiency. J. Mol. Evol. *57*, 453–466.
5. Moriyama, E.N., and Powell, J.R. (1997). Codon usage bias and tRNA abundance in *Drosophila*. J. Mol. Evol. *45*, 514–523.
6. Chiapello, H., Lisacek, F., Caboche, M., and Alain, H. (1998). Codon usage and gene function are related in sequences of *Arabidopsis thaliana*. Gene *209*, GC1–GC38.
7. Duret, L., and 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.
8. Kreitman, M., and Antezana, M. (1999). The population and evolutionary genetics of codon bias. In Evolutionary Genetics: From Molecules to Morphology, *Volume 1*, R.S. Singh and C.B. Krimbas, eds. (New York: Cambridge University Press), pp. 82–101.
9. Duret, L. (2002). Evolution of synonymous codon usage in metazoans. Curr. Opin. Genet. Dev. *12*, 640–649.
10. Wright, S.I., Yau, C.B.K., Looseley, M., and Meyers, B.C. (2004). Effects of gene expression on molecular evolution in

Arabidopsis thaliana and *Arabidopsis lyrata*. Mol. Biol. Evol. *21*, 1719–1726.

11. Hartl, D.L., Moriyama, E.N., and Sawyer, S.A. (1994). Selection intensity for codon bias. Genetics *138*, 227–234.
12. Akashi, H. (1995). Inferring weak selection from patterns of polymorphism and divergence at silent sites in *Drosophila* DNA. Genetics *139*, 1067–1076.
13. Akashi, H., and Schaeffer, S.W. (1997). Natural selection and the frequency distributions of "silent" DNA polymorphism in *Drosophila*. Genetics *146*, 295–307.
14. Maside, X.L., Lee, A.W.S., and Charlesworth, B. (2004). Selection on codon usage in *Drosophila americana*. Curr. Biol. *14*, 150–154.
15. McVean, G.A.T., and Vieira, J. (1999). The evolution of codon preferences in *Drosophila*: a maximum-likelihood approach to parameter estimation and hypothesis testing. J. Mol. Evol. *49*, 63–75.
16. Comeron, J.M., and Guthrie, T.B. (2005). Intragenic Hill-Robertson interference influences selection intensity on synonymous mutations in *Drosophila*. Mol. Biol. Evol. *22*, 2519–2530.
17. Bartolome, C., Maside, X., Yi, S., Grant, A.L., and Charlesworth, B. (2005). Patterns of selection on synonymous and nonsynonymous variants in *Drosophila miranda*. Genetics *169*, 1495–1507.
18. Williamson, S.H., Hernandez, R., Fledel-Alon, A., Zhu, L., Nielsen, R., and Bustamante, C.D. (2005). Simultaneous inference of selection and population growth from patterns of variation in the human genome. Proc. Natl. Acad. Sci. USA *102*, 7882–7887.
19. Comeron, J.M. (2006). Weak selection and recent mutational changes influence polymorphic synonymous mutations in humans. Proc. Natl. Acad. Sci. USA *103*, 6940–6945.
20. Graustein, A., Gaspar, J.M., Walters, J.R., and Palopoli, M.F. (2002). Levels of DNA polymorphism vary with mating system in the nematode genus *Caenorhabditis*. Genetics *161*, 99–107.
21. Jovelin, R., Ajie, B.C., and Phillips, P.C. (2003). Molecular evolution and quantitative variation for chemosensory behaviour in the nematode genus *Caenorhabditis*. Mol. Ecol. *12*, 1325–1337.
22. Haag, E.S., and Ackerman, A.D. (2005). Intraspecific variation in *fem-3* and *tra-2*, two rapidly coevolving nematode sex-determining genes. Gene *349*, 35–42.
23. Cutter, A.D., Baird, S.E., and Charlesworth, D. (2006). High nucleotide polymorphism and rapid decay of linkage disequilibrium in wild populations of *Caenorhabditis remanei*. Genetics, in press.
24. Stenico, M., Lloyd, A.T., and Sharp, P.M. (1994). Codon usage in *Caenorhabditis elegans*: delineation of translational selection and mutational biases. Nucleic Acids Res. *22*, 2437–2446.
25. Stein, L.D., Bao, Z., Blasiar, D., Blumenthal, T., Brent, M.R., Chen, N., Chinwalla, A., Clarke, L., Clee, C., Coghlan, A., et al. (2003). The genome sequence of *Caenorhabditis briggsae*: a platform for comparative genomics. PLoS Biol. *1*, 166–192.
26. Kiontke, K., Gavin, N.P., Raynes, Y., Roehrig, C., Piano, F., and Fitch, D.H.A. (2004). *Caenorhabditis* phylogeny predicts convergence of hermaphroditism and extensive intron loss. Proc. Natl. Acad. Sci. USA *101*, 9003–9008.
27. Cutter, A.D., Wasmuth, J., and Blaxter, M.L. (2006). The evolution of biased codon and amino acid usage in nematode genomes. Mol. Biol. Evol., in press.
28. McVean, G.A.T., and Charlesworth, B. (1999). A population genetic model for the evolution of synonymous codon usage: patterns and predictions. Genet. Res. *74*, 145–158.
29. Galtier, N., Bazin, E., and Bierne, N. (2006). GC-biased segregation of noncoding polymorphisms in *Drosophila*. Genetics *172*, 221–228.
30. Marais, G. (2003). Biased gene conversion: implications for genome and sex evolution. Trends Genet. *19*, 330–338.
31. Adams, M.D., Celniker, S.E., Holt, R.A., Evans, C.A., Gocayne, J.D., Amanatides, P.G., Scherer, S.E., Li, P.W., Hoskins, R.A., Galle, R.F., et al. (2000). The genome sequence of *Drosophila melanogaster*. Science *287*, 2185–2195.
32. C. elegans Sequencing Consortium. (1998). Genome sequence of the nematode *C. elegans*: a platform for investigating biology. Science *282*, 2012–2018.
33. Rozas, J., Sanchez-DelBarrio, J.C., Messeguer, X., and Rozas, R. (2003). DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics *19*, 2496–2497.
34. Kimura, M. (1971). Theoretical foundation of population genetics at the molecular level. Theor. Popul. Biol. *2*, 174–208.