

# Maximum Likelihood Estimation of Ancestral Codon Usage Bias Parameters in *Drosophila*

Rasmus Nielsen,\*† Vanessa L. Bauer DuMont,‡ Melissa J. Hubisz,† and Charles F. Aquadro‡

\*Institute of Biology and Centre for Bioinformatics, University of Copenhagen, Copenhagen, Denmark; †Department of Biological Statistics and Computational Biology and ‡Department of Molecular Biology and Genetics, Cornell University

We present a likelihood method for estimating codon usage bias parameters along the lineages of a phylogeny. The method is an extension of the classical codon-based models used for estimating  $d_N/d_S$  ratios along the lineages of a phylogeny. However, we add one extra parameter for each lineage: the selection coefficient for optimal codon usage ( $S$ ), allowing joint maximum likelihood estimation of  $S$  and the  $d_N/d_S$  ratio. We apply the method to previously published data from *Drosophila melanogaster*, *Drosophila simulans*, and *Drosophila yakuba* and show, in accordance with previous results, that the *D. melanogaster* lineage has experienced a reduction in the selection for optimal codon usage. However, the *D. melanogaster* lineage has also experienced a change in the biological mutation rates relative to *D. simulans*, in particular, a relative reduction in the mutation rate from A to G and an increase in the mutation rate from C to T. However, neither a reduction in the strength of selection nor a change in the mutational pattern can alone explain all of the data observed in the *D. melanogaster* lineage. For example, we also confirm previous results showing that the *Notch* locus has experienced positive selection for previously classified unpreferred mutations.

## Introduction

The existence of codon bias, the nonrandom use of synonymous codons, is well documented in *Drosophila* (i.e., Shields et al. 1988). Although biased gene conversion and/or mutational pressures cannot be completely discounted (i.e., Takano-Shimizu 2001; Maside et al. 2004; Bartolomé et al. 2005; Singh et al. 2005), there is a wealth of evidence suggesting an important role of selection in establishing the codon bias in these species (i.e., Begun 2001; Kern et al. 2002; Bartolomé et al. 2004; Bauer DuMont et al. 2004). For example, in *Drosophila melanogaster*, codon bias appears to covary with the rate of recombination (Kliman and Hey 1993; Comeron et al. 1999; Duret 2002; but see Singh et al. 2005), suggesting that selection for codon bias is not as efficient in regions of low recombination. In addition, codon bias is more pronounced in highly expressed genes (reviewed in Powell and Moriyama 1997) and is positively correlated with the degree of conservation at an amino acid position (Akashi 1996) indicating that translational efficiency and/or accuracy plays some part in the selective pressure toward preferred codons. By comparing codon usage between high and low bias loci, preferred codons are predicted to have the nucleotides C or G in their third position (Akashi 1995).

Although codon bias appears to some degree in a number of *Drosophila* species, a difference in the magnitude of the bias has been observed at the level of individual loci and across genomes (i.e., Munté et al. 1997, 2001; Kliman 1999). For instance, a genome-wide reduction of codon bias has been observed between *D. melanogaster* and *Drosophila simulans* (Akashi 1996). Many more unpreferred mutations (i.e., those from a preferred codon to an unpreferred codon) have fixed along the *D. melanogaster* lineage. Sixty to seventy percent of synonymous changes between these species that differ between unpreferred and preferred codons have the unpreferred codon in *D. melanogaster*. In

addition, when rooted with an outgroup, many loci have a significant relative rate test at synonymous sites due in large part to the fixation of many unpreferred synonymous codons along the *D. melanogaster* lineage (Akashi 1996; Bauer DuMont et al. 2004). Relaxation of constraint, presumably due to a reduction in population size in *D. melanogaster*, has been the favored explanation for these observations.

However, although relaxation of constraint can explain most of the differences between *D. melanogaster* and *D. simulans* in synonymous codon usage, Bauer DuMont et al. (2004) found evidence for the role of positive selection in the fixation of unpreferred mutations at the *Notch* locus in *D. melanogaster*. This was done by devising a counting method similar to the common  $d_N/d_S$  estimation methods (e.g., Nei and Gojobori 1986). If the change in codon usage between these species at *Notch* was completely governed by relaxation of constraint, then the ratios in the number of preferred fixations per site and the number of unpreferred fixations per site should be equal. However, they observed significantly more unpreferred fixations per site than preferred fixations per site in *D. melanogaster*. One objective of this paper is to formalize the inference procedure of Bauer DuMont et al. (2004) in a codon-based likelihood framework.

The first direct estimators of selection coefficients affecting synonymous sites were obtained by McVean and Vieira (2001). They compared DNA data from *D. melanogaster* and *D. simulans* and from *D. melanogaster* and *Drosophila virilis* using a Markov Chain Monte Carlo method to estimate the strength of codon usage bias in these organisms. This study was a vast improvement over previous studies; in that, it allowed direct estimation of parameters relating to both mutation and selection. McVean and Vieira (2001) concluded that *D. melanogaster* shows no evidence of positive selection, whereas *D. simulans* experiences only half the selection pressure for codon usage of their common ancestor.

The analysis by McVean and Vieira (2001) was important in establishing appropriate models for statistical inferences of patterns and evolution of codon usage bias. However, by only using pairs of species, very little power is retained to make inferences regarding the pattern and evolution of codon usage bias in each of the 2 ancestral

Key words: codon usage bias, codon usage, maximum likelihood, codon-based models, selection.

E-mail: rasmus@binf.ku.dk.

*Mol. Biol. Evol.* 24(1):228–235. 2007

doi:10.1093/molbev/msl146

Advance Access publication October 13, 2006

phylogenetic lineages. Conventional wisdom in the field would argue that an outgroup is needed to make lineage-specific inferences, and from a statistical standpoint, it may be argued that lineage-specific inferences in the absence of an outgroup may not be desirable because they may suffer from low power or are very model dependent because most inferences will rely on the nonreversible aspects of the model. In addition, McVean and Vieira (2001) used a model that allowed selection, but not mutation, to vary among lineages.

Here, we present codon-based likelihood models akin to the models of McVean and Vieira (2001) but applicable to more than 2 species. To reduce the number of parameters, our models assume that the same strength of codon bias is acting in all amino acids. However, in contrast to McVean and Vieira (2001), we allow for different mutational processes among evolutionary lineages, and we also use a more complex mutation model. Our implementation allows for standard numerical optimization methods (McVean and Vieira [2001] used a stochastic optimization algorithm), and it takes the full complexities of the genetic code into account avoiding the possible confounding effects of non-synonymous substitutions. We apply the method to previously published data from *Drosophila yakuba*, *D. simulans*, and *D. melanogaster*.

## Materials and Methods

### Likelihood Models

The model we will develop is an extension of the codon-based likelihood models by Goldman and Yang (1994) and Muse and Gaut (1994) that explicitly takes codon usage bias into account. As in previous analyses, we will assume that codons can be divided into 2 sets: unpreferred and preferred. Although this assumption may be an oversimplification, it significantly reduces the number of parameters of the model, and thereby, the computational and statistical complexity of the problem. The codon-based models of molecular evolution are Markov models with state space on the set of the 61 codons in the genetic code ( $\Omega$ ). The model is specified in terms of transition rates from state  $i$  to state  $j$ ,  $q_{ij}$  ( $i, j \in \Omega$ ). Any existing Markov model with rates  $q_{ij}$  can be modified to incorporate codon usage bias, forming a new Markov chain with rates  $q_{ij}^*(i, j \in \Omega)$ , by a consideration of the underlying population genetics.

If the selection coefficient acting on a mutation from an unpreferred to a preferred codon is  $s$ , the probability of fixation of such a mutant is (Kimura 1962)

$$P_{s+}/N = \frac{S/N}{1 - e^{-S}}, \quad (1)$$

assuming  $s$  is small (e.g.,  $s \ll 1$ ),  $N$  is the chromosomal population size, and  $S = 2Ns$ . Likewise, the probability of fixation of a mutation from a preferred to an unpreferred codon is

$$P_{s-}/N = \frac{-S/N}{1 - e^S} \quad (2)$$

and the probability of fixation of a new mutation from an unpreferred codon to an unpreferred codon, or from a preferred codon to a preferred codon, is  $1/N$ . Scaling time in

terms of the population size, the transition rates of the modified Markov chain are then given by

$$q_{ij}^* = \begin{cases} q_{ij}P_{s+}, & \text{if } i \text{ unpreferred, } j \text{ preferred} \\ q_{ij}P_{s-}, & \text{if } i \text{ preferred, } j \text{ unpreferred, } i \neq j, i, j \in \Omega. \\ q_{ij} & \text{else} \end{cases} \quad (3)$$

We see that, if the process defined by  $q_{ij}$  is time reversible with stationary distribution  $\pi_i$ ,  $i \in \Omega$ , then the process defined by  $q_{ij}^*$  is time reversible with stationary distribution

$$\pi_i^* \propto \begin{cases} \pi_i P_{s+}, & \text{if } i \text{ preferred} \\ \pi_i P_{s-}, & \text{if } i \text{ unpreferred, } i \in \Omega. \end{cases} \quad (4)$$

When time reversibility is not assumed, the stationary codon frequencies can be obtained using standard Markov chain methods, for example, by solving a system of 61 linear equations or as a byproduct of the Eigenvector decomposition often used in the calculation of transition probabilities. In the following, we will assume that  $q_{ij}$  is given by the nucleotide mutation process, modified to take selection at the amino acid level into account. Notice that selection for optimal codon usage will also act on amino acid changes. We represent codon  $i$  as a triplet  $i_1i_2i_3$  and codon  $j$  as  $j_1j_2j_3$  ( $i_1, i_2, i_3, j_1, j_2, j_3, \in \{T, C, A, G\}$ ). If codons  $i$  and  $j$  differ by exactly one nucleotide substitution in position  $k$ , then

$$q_{ij} = \begin{cases} \alpha_{i_kj_k}, & \text{if } i_k \neq j_k \text{ by a synonymous mutation} \\ \omega\alpha_{i_kj_k}, & \text{if } i_k \neq j_k \text{ by a nonsynonymous mutation} \end{cases} \quad (5)$$

If codons  $i$  and  $j$  differ in more than one nucleotide position,  $q_{ij} = 0$ . Here  $\alpha_{i_kj_k}$  is the nucleotide mutation rate from  $i_k$  to  $j_k$  ( $i_k, j_k \in \{T, C, A, G\}$ ), and  $\omega$  is the rate ratio of nonsynonymous to synonymous substitutions. The parameters of the model are then  $\omega$ ,  $S$ , and  $\alpha = \{\alpha_{AT}, \alpha_{AC}, \alpha_{AG}, \alpha_{TA}, \dots, \alpha_{GA}\}$ , in addition to parameters related to the phylogenetic tree such as the relative branch lengths. Additionally, it is possible to allow  $\omega$ ,  $S$ , or  $\alpha$  to vary among branches in the phylogenetic tree. The full model (FM) is here defined as a model in which  $\omega$ ,  $S$ , or  $\alpha$  vary independently among branches, with the exception of the 2 branches around the root in the (binary) tree. In these 2 branches, the values of  $\omega$ ,  $S$ , and  $\alpha$  are assumed to be identical. The total number of parameters of the FM is then 42 for 3 species.

Superimposing this stochastic process on a phylogenetic tree, sampling probabilities of the data can be calculated, and parameters can be estimated using maximum likelihood and numerical optimization (see e.g., Felsenstein 1981; Goldman and Yang 1994). In this way, it is possible to estimate lineage-specific parameters of the model, such as parameters pertaining to codon usage bias. Because the most general model we use is not time reversible, estimation is performed on a rooted tree in which *D. yakuba* is treated as an outgroup. Although the placement of the root along the *D. yakuba* lineage in principle could be estimated from the data, we fixed it using a molecular clock assumption between *D. yakuba* and *D. simulans*, and we assumed that the mutational process was identical on the 2 *D. yakuba* lineages to reduce the number of free parameters. All the parameter estimates are very robust to the placement of

**Table 1**  
**Parameter Estimates from the Concatenated Data**

Model	Log Like	Number of Parameters	AIC	$s_m$	$s_s$	$s_y$	$\omega_m$	$\omega_s$	$\omega_y$
FM	-35810.4	42	71704.8	0.13	1.17	1.11	0.15	0.19	0.14
Symmetric FM	-36009.7	24	72067.5	0.11	1.19	0.94	0.15	0.19	0.13
$S_y = 0$	-36517.1	41	73116.1	0.32	1.48	0.00	0.15	0.18	0.13
$S_m = 0$	-35810.6	41	71703.2	0.00	1.15	1.11	0.16	0.19	0.14
$S_s = 0$	-35821.5	41	71725.0	0.01	0.00	1.12	0.15	0.18	0.14
$\alpha_m = \alpha_y$	-35840.3	31	71742.7	-0.28	1.14	1.14	0.14	0.18	0.13
$\alpha_m = \alpha_y, S_s = S_m$	-35853.3	30	71766.6	0.10	0.10	1.15	0.13	0.18	0.13
$\alpha_m = \alpha_y, S_s = 0$	-35851	30	71762.0	-0.37	0.00	1.15	0.14	0.18	0.13
$\alpha_m = \alpha_y, S_m = 0$	-35841.8	30	71743.6	0.00	1.19	1.15	0.14	0.18	0.13
$\alpha_s = \alpha_y$	-35826.5	31	71715.1	0.15	1.37	1.13	0.15	0.18	0.13
$\alpha_s = \alpha_y, S_s = S_m$	-35836.7	30	71733.3	0.77	0.77	1.12	0.14	0.20	0.13
$\alpha_s = \alpha_y, S_s = 0$	-35849	30	71758.0	-0.01	0.00	1.13	0.15	0.24	0.13
$\alpha_s = \alpha_y, S_m = 0$	-35826.8	30	71713.6	0.00	1.35	1.13	0.16	0.18	0.13
$\alpha_s = \alpha_m = \alpha_y$	-35850.1	20	71740.1	-0.31	1.42	1.14	0.14	0.18	0.13
PH	-36738.9	9	73495.8	0.01	1.74	1.48	0.15	0.18	0.13
w/polymorph.	-40080.1	42	—	0.15	0.82	1.12	0.15	0.15	0.14

NOTE.—The parameters indicated by  $s_i$  and  $\omega_i$  are the scaled selection coefficient acting on synonymous sites and the  $d_N/d_S$  ratio, respectively, on the lineage leading to species  $i$ , and the species are  $m$ : *D. melanogaster*,  $s$ : *D. simulans*, and  $y$ : *D. yakuba*. The row labeled “PH” gives the parameter estimates when using the mutation rates published by Petrov and Hartl (1999). The row labeled “w/polymorph.” gives the maximum likelihood value and parameters estimates for the FM when polymorphisms in *D. simulans* and *D. melanogaster* are included (this is equivalent to pairwise divergence between single sequences sampled from each species). Because the data is different in this row, the likelihood value is not comparable with the other values shown in the table.

the root, except for the mutational matrix in the *D. yakuba* lineage itself (under assumptions of non-time reversibility). For example, when *D. yakuba* is (erroneously) assumed to be the direct ancestor of *D. melanogaster* and *D. simulans*, parameter estimates of the mutation rates,  $\omega$  and  $S$  on the lineages leading to *D. melanogaster* and *D. simulans* differ from the estimates based on the molecular clock assumption by less than 1%. The estimates of selection coefficients and  $\omega$  on the *D. yakuba* lineage itself changes by only 1–2%.

In addition to analyses where  $\alpha$  is considered a parameter, we also perform analyses using the previously published estimates of  $\alpha$  by Petrov and Hartl (1999). Petrov and Hartl (1999) used observed substitutions in “dead on arrival” transposable elements (considered to be pseudogenes) that were located throughout the genome to infer the underlying mutational pattern in *Drosophila*. We also analyze a symmetric model without strand biases, that is, where the mutation rates between complementary nucleotides are identical.

We concatenated the sequences from all the loci and estimated parameters under the full model. In a subsequent gene-by-gene analysis, we used mutation parameters from the concatenated data to reduce the number of parameters. Branch lengths and values of  $\omega$  and  $S$  were then estimated for each of the 3 lineages.

Notice that we have so far assumed that the strength of the codon usage bias is the same in all amino acids. Obviously, this may not be a realistic assumption and can be relaxed if sufficiently large data sets are available to allow the estimation of additional parameters.

## Data

The method was applied to loci for which polymorphism data was available in both *D. melanogaster* and *D. simulans*, and polymorphic codons are removed from the analysis, thereby reducing the chance of confounding polymorphisms with fixed substitutions. The maximum likelihood estimates

are then only based on (apparently) fixed differences, except where otherwise noted. Eighteen loci were identified from the literature to fit our criteria. A list of the loci and the GenBank accession numbers used can be found in Appendix 4. For 9 of these loci, polymorphism data from intronic regions was also available. We followed Akashi (1996) in our classification of which codons are optimal. However, the analysis was also repeated assuming only one codon can be optimal for each amino acid. We also present results from the analyses of intronic sequences from 9 of the loci.

## Population Genetic Analyses

To further elucidate the role of selection, we construct a test based on comparing variability within and between species at different types of mutations akin to the McDonald and Kreitman test (1991). We construct a 2-by-2 contingency table of unpreferred/preferred polymorphism and fixed mutations. A test of homogeneity is then performed using a  $G$ -test.

## Results

### Estimation of $\omega$ and ancestral codon usage bias

We first estimated parameters for the 3 species by concatenating the sequences, assuming that *D. yakuba* is the outgroup and that the ancestor of these species was at mutation–selection–drift equilibrium at the time of speciation. The results of this analysis are given in table 1.

Estimates of  $\omega$  are highest in *D. simulans* (0.19) and smaller in *D. melanogaster* (0.15) and *D. yakuba* (0.14). In contrast, the estimates of  $S$  on codon usage bias is similar in *D. yakuba* ( $S_y = 1.11$ ) and *D. simulans* ( $S_s = 1.17$ ) but close to zero ( $S_m = 0.13$ ) for *D. melanogaster*. Assuming a constant mutation rate, the relative small value of  $\omega$  on the *D. melanogaster* lineage appears primarily to be caused by an increased synonymous rate ratio on this lineage. The expected number of synonymous substitutions per codon

**Table 2**  
**Likelihood Ratio Tests**

Test	Parameters	2LR	Degees of Freedom	Approximate P Value
Symmetric mutation rates	$H_0: \alpha_{ij} = \alpha_{vk}$ if $(i, v)$ and $(j, k)$ are Watson–Crick pairs.	398.7	18	<0.0001
Same amount of selection in <i>melanogaster</i> and <i>simulans</i>	$H_0: S_s = S_m$	20.4	1	<0.0001
Selection on <i>melanogaster</i> lineage	$H_0: S_m = 0$	0.4	1	0.527
Selection on <i>simulans</i> lineage	$H_0: S_s = 0$	22.2	1	<0.0001
Selection on <i>yakuba</i> lineage	$H_0: S_y = 0$	1413.3	1	<0.0001
Mutation <i>mel</i> = mutation <i>yak</i>	$H_0: \alpha_m = \alpha_y$	59.9	11	<0.0001
Mutation <i>sim</i> = mutation <i>yak</i>	$H_0: \alpha_s = \alpha_y$	32.3	11	0.0007
PH mutational model	$H_0: \alpha = \alpha_{PH}$	1857.0	33	<0.0001

on the *D. melanogaster* lineage is 35% larger than on the *D. simulans* lineage.

For all 3 species, we tested the hypothesis of no selection for optimal codon usage using a likelihood ratio test (table 2). In the case of *D. simulans* and *D. yakuba*, we rejected the null hypothesis of  $S = 0$  with strong statistical significance. There is much more power to reject this hypothesis in *D. yakuba* than in the other lineages, predominantly, because the root is located on this lineage. In the case of *D. melanogaster*, we could not reject the hypothesis of  $S = 0$ . This conclusion also holds up under other assumptions regarding the mutational model, for example, if we assume that the mutation process is identical in *D. melanogaster* and *D. yakuba* or if we assume that the mutations rates are identical in *D. simulans* and *D. yakuba*. Likewise, this result also holds up if we assume that only one codon is optimal for each amino acid (not shown).

McVean and Vierra (2001) suggested that *D. simulans* experiences half the level of selection on codon usage than its ancestor. In addition, results of Akashi et al. (2006) suggest a decrease in major codon usage in *D. yakuba*. Our estimates of  $S_s$  and  $S_y$  are similar and are significantly greater than one. Thus, using our method, selection in favor of preferred mutations appears to be active along both species' lineages. The difference between studies could be explained by the fact that McVean and Vierra (2001) did not use an outgroup species and, therefore, could not easily distinguish between changed processes on the *D. simulans* and the *D. melanogaster* lineage, and Akashi et al. (2006) only compared raw numbers of differences not calibrated by the number of sites.

When the analysis is performed including polymorphisms, the estimate of  $S_s$  is reduced from 1.17 to 0.82. This is expected if mutations from preferred to unpreferred codons are slightly deleterious because the polymorphism data will contain an excess of such mutations, and they will erroneously be considered fixed differences. This further confirms the existence of codon usage bias in the *D. simulans* lineage and illustrates the confounding effects of including segregating polymorphism when estimating selection parameters. However, the estimate of  $S_m$  changes only from 0.13 to 0.15, again suggesting that the *D. melanogaster* lineage is either largely unaffected by selection for optimal codon usage or affected by opposite directed selection in different genes (as will be detailed below).

#### Estimation of mutation parameters

The estimates of mutation rate parameters are quite different between the *D. melanogaster* and the *D. simulans* lineages (table 3). In particular, there appears to be a reduction in the A to G and T to C mutations rate and an increase in the C to T mutation rate in *D. melanogaster* compared with *D. simulans*. We do not list results for the *D. yakuba* lineage because these results are sensitive to the placement of the root in a model that is not time reversible. A likelihood ratio tests rejects the hypothesis of equal mutation rates of *D. yakuba* and *D. melanogaster* with strong statistical significance (table 2). It also rejects the hypothesis of equal mutation rates between the *D. yakuba* and *D. simulans*. The Akaike information criterion (AIC) score for the models assuming  $\alpha_m = \alpha_y$  and  $\alpha_s = \alpha_y$  are  $-35840.3$  and  $-35826.5$ , respectively, suggesting that the primary cause of the difference between the mutational processes on the 2 lineages is a change in the mutational process on the *D. melanogaster* lineage.

Assuming strand symmetry, we would expect that the rate of mutation from nucleotide  $i$  to  $j$ ,  $i \neq j$ , equals the rate of mutation from  $v$  to  $k$ ,  $v \neq k$ , if  $i$  and  $v$ , and  $j$  and  $k$  form Watson–Crick pairs. We tested this hypothesis using a likelihood ratio test (table 2) and could reject the hypothesis of strand symmetry with strong statistical confidence. This is in agreement with recent studies by Singh et al. (2005), although the pattern of the asymmetry is not identical between the studies. The cause of this asymmetry is still unknown, but transcription associated mutation processes are not likely to be the cause as transcription-coupled repair does not appear to operate in *D. melanogaster* and direct estimates of mutation patterns have not detected it (De Cock et al. 1992; Van Der Helm et al. 1997; Singh et al. 2005). Although the strand asymmetry may conceivably be caused by modeling inadequacies relating to the effect of selection, the fact that asymmetries also are observed for intronic sequences (see table 3), and that such asymmetries also are observed in other studies, suggests that this is not the case.

We also tested if the mutation model of Petrov and Hartl (1999; PH) fits the data, and we could reject the hypothesis of this mutational model ( $\alpha = \alpha_{PH}$ ) with strong statistical confidence. However, it is interesting to note that our estimates of the mutation matrix in *D. melanogaster* is very similar to the PH model, whereas the estimates in *D. simulans* are much more different from the PH model.

**Table 3**  
**Estimates of Mutation Rate Parameters**

PH				
	A	C	G	T
A	—	0.05	0.08	0.07
C	0.09	—	0.06	0.15
G	0.15	0.06	—	0.09
T	0.07	0.08	0.05	—
<i>melanogaster</i> Coding				
	A	C	G	T
A	—	0.063	0.086	0.082
C	0.107	—	0.036	0.189
G	0.146	0.057	—	0.054
T	0.071	0.066	0.045	—
<i>simulans</i> Coding				
	A	C	G	T
A	—	0.050	0.167	0.059
C	0.075	—	0.053	0.088
G	0.131	0.057	—	0.049
T	0.068	0.129	0.074	—
<i>melanogaster</i> Intronic				
	A	C	G	T
A	—	0.026	0.072	0.021
C	0.089	—	0.055	0.201
G	0.205	0.033	—	0.106
T	0.073	0.102	0.018	—
<i>simulans</i> Intronic				
	A	C	G	T
A	—	0.032	0.223	0.077
C	0.056	—	0.065	0.092
G	0.140	0.066	—	0.025
T	0.082	0.118	0.022	—

NOTE.—To facilitate easy comparison, the values of  $\alpha_{ij}$  are scaled so that they add to one. Results for *D. yakuba* are not shown because these estimates are sensitive to the placement of the root. PH is the mutation matrix estimated by Petrov and Hartl (1999).

This suggests that the PH estimates of the mutation rate are relatively accurate but do not fit the data well simply because the mutation matrices are different in the 3 species. We tested this by imposing the PH model on the *D. melanogaster* lineage only. Using a likelihood ratio test, we could not reject this model against the general model (LR = 0.7;  $P = 0.17$ ), demonstrating that the PH model provides a good approximation to the mutational process on this lineage. The results presented here would, therefore, be qualitatively similar if we imposed the PH mutation matrix on the *D. melanogaster* lineage.

We tested if a model involving only a change in the mutation process but no change in  $S_m$  could explain the shift codon usage in *D. melanogaster* (table 1). Again, the hypothesis of no difference between  $S_m$  and  $S_s$  can be rejected with strong statistical significance (LR = 20.4,  $P < 0.0001$ ), suggesting that mutation alone is not driving the difference in codon usage between these species.

Among the models examined here, the best performing model, according to the AIC, is a model with no selection for optimal codon usage in the *D. melanogaster* lineage but with different mutation matrices among all 3 lineages and selection on the *D. yakuba* and *D. simulans* lineages ( $S_m = 0$ ). The second best model is the FM, and the third best model is a model with equal mutation matrices on the *D. yakuba* and *D. simulans* lineages and no selection on the *D. melanogaster* lineage ( $\alpha_s = \alpha_y$ ,  $S_m = 0$ ). Changing

the assumptions so that only one codon is optimal for each amino acid gives an Aikake score of 71619.4 for the full model, 71618.1 for the model with  $S_m = 0$ , and 71617.8 for the model with  $\alpha_s = \alpha_y$  and  $S_m = 0$ , suggesting that the latter model is preferable. In general, the conclusions of no (or very weak) selection in the *D. melanogaster* lineage and a shift in the mutation process along the *D. melanogaster* lineage seems to be relative robust to assumptions regarding the set of optimal codons. Also, the much higher AIC value for a model assuming  $\alpha_m = \alpha_y$  than for a model assuming  $\alpha_s = \alpha_y$  (table 1), suggests that the major cause for the difference in mutation process between *D. melanogaster* and *D. simulans* is a shift in *D. melanogaster*.

We estimated mutation matrices for intronic regions from 9 of the loci analyzed to evaluate if the hypothesis of a change in the mutational process is also supported by this data (table 3). The results are roughly compatible with the results obtained from the exonic regions but show an even stronger mutational shift between *D. melanogaster* and *D. simulans* in the C to T and A to G mutation rates.

### Gene-by-Gene Analysis

Parameters were estimated for each gene based on the species-specific mutation matrices obtained from the concatenated data. This analysis reveals evidence for codon usage bias in all genes in the ancestral *D. yakuba* lineage (table 4) where a likelihood ratio test is significant at the 5% level (before correction for multiple tests) in all cases but *anon1g5*. In *D. simulans*, many genes show evidence for selection for optimal codon usage, and for 3 genes, we can reject the null hypothesis of  $S_s = 0$  at the 5% level using a likelihood ratio test. In all cases where the null hypothesis of  $S_s = 0$  or  $S_y = 0$  can be rejected, the estimates of  $S_s$  and  $S_y$  are positive (selection for the optimal codon).

In the *D. melanogaster* lineage, we can only reject the hypothesis of  $S_m = 0$ , for one gene: N3, the 3' end of the *Notch* locus with strong statistical confidence ( $P = 0.0003$ ). Curiously, the estimate of  $S_m$  for this gene is negative (−1.9), meaning that unpreferred mutations are being favored on the *D. melanogaster* lineage. Clearly, the evolution of this gene is rather different from the other genes. *Notch* 3' also has the largest likelihood ratio against the hypothesis of  $S_s = 0$  (30.3) but with a positive estimate of  $S_s$  (3.6). Thus, there seems to be selection favoring different sets of codons in this gene region in the *D. simulans* and the *D. melanogaster* lineages, as was also inferred by Bauer DuMont et al. (2004). In the *D. melanogaster*, lineage selection is now acting against apparent “preferred” mutations. This result also suggests that the inference of no selection in the *D. melanogaster* lineage may be caused by a cancellation of effects of very weak codon usage bias in the *D. melanogaster* lineage in many genes and strong opposite directed codon usage bias in the *Notch* 3' region.

### Population Genetic Analysis

The contingency tables for unpreferred/preferred polymorphisms and fixations are shown in table 5. Homogeneity cannot be rejected in *D. melanogaster* ( $P = 0.66$ ) but can be rejected in *D. simulans* ( $P < 0.0001$ ). This illustrates that the evolutionary processes are different in these 2

**Table 4**  
The Parameter Estimates for Each Gene

Gene	2LR( $S_y = 0$ )	2LR( $S_m = 0$ )	2LR( $S_s = 0$ )	$S_y$	$S_m$	$S_s$
<i>adh</i>	198.5	0.4	1.1	1.9	1.0	2.8
<i>anon1a3</i>	18.9	1.3	0.4	0.6	0.8	0.8
<i>anon1e9</i>	28.7	1.2	0.3	0.6	0.8	-0.5
<i>anon1g5</i>	4.8	0.0	0.8	0.3	-0.1	-0.9
<i>hex-a</i>	203.6	2.1	2.3	1.3	1.3	1.3
<i>hex-c</i>	119.9	0.1	0.3	1.0	0.2	-0.6
<i>hex-T1</i>	117.3	0.1	0.4	1.0	-0.3	0.7
<i>hex-T2</i>	187.4	0.0	0.7	1.3	0.0	0.5
<i>nth</i>	26.1	0.0	0.2	0.5	0.1	0.4
<i>N3'</i>	99.1	13.0	30.3	0.9	-1.9	3.6
<i>N5'</i>	67.0	0.9	1.4	1.3	-0.8	3.7
<i>per</i>	290.4	0.2	3.0	1.5	0.4	2.0
<i>pgi</i>	296.6	2.8	11.4	1.5	1.3	10.0
<i>pgm</i>	96.3	0.8	0.0	0.9	0.5	-0.1
<i>rel</i>	131.0	0.0	4.0	0.8	0.1	1.1
<i>tpi</i>	230.6	0.3	4.2	2.2	-10.0	3.2
<i>z</i>	62.0	2.3	0.6	1.1	-10.0	0.8
<i>zw</i>	412.4	0.4	9.5	1.9	0.6	1.9

NOTE.—The likelihood ratio test is significant at the 5% level if  $2LR > 3.84$  and at the 1% level if  $2LR > 6.63$  before correction for multiple testing.

species, and the result is consistent with the hypothesis of a relaxation of selection in the *D. melanogaster* lineage.

At equilibrium, there should be equally many fixed preferred and unpreferred mutations. This is observed in *D. simulans* but not in *D. melanogaster* (table 5), again indicating that the process is at equilibrium in *D. simulans* but not in *D. melanogaster*. The fact that *D. melanogaster* have equal ratios of fixed to polymorphic unpreferred and preferred mutations, but an almost 20-fold increase in unpreferred fixations over preferred fixations, demonstrates that the *D. melanogaster* is not at equilibrium but has experienced a change in either mutation bias or intensity of selection. In contrast, *D. simulans* has an equal ratio of unpreferred and preferred fixed mutations, indicating that the difference between the 2 species cannot be explained by nonequilibrium conditions in *D. simulans*. These results are in agreement with Kern and Begun (2005) who reported nonequilibrium evolution at both synonymous and intron sites in *D. melanogaster* but equilibrium in *D. simulans* when comparing GC with AT versus AT to GC fixations. Combining this with the evidence from maximum likelihood analysis and the test of homogeneity, we conclude that *D. melanogaster* has experienced reduced selection for optimal codon usage.

## Discussion

A change in codon usage can be due to shifts in selection pressures or a change in the mutation process. Both types of changes have been documented in *Drosophila* either at the level of individual loci or genome wide (e.g., Munté et al. 1997, 2001; Rodríguez-Trelles et al. 1999, 2000; Takano-Shimizu 1999, 2001; Akashi et al. 2006). One method for teasing these effects apart is to make explicit models that incorporate both of them. Because mutational biases also affect amino acid changes and changes between preferred and between unpreferred codons, these models make it possible to estimate both the effect of mutation biases and parameters relating to the selection on optimal codon usage. The method we have developed in this

**Table 5**  
Preferred and Unpreferred Polymorphisms and Fixations in the 2 Species

	<i>Drosophila simulans</i> fixed	<i>Drosophila simulans</i> polymorphic	<i>Drosophila melanogaster</i> fixed	<i>Drosophila melanogaster</i> polymorphic
Preferred	98	83	22	16
Unpreferred	92	221	196	166

paper relies on an a priori assignment of preferred codons. If this assignment is wrong, the method will lead to false inferences. It should also be noted that in our implementation, all preferred changes (or unpreferred changes) have the same selection coefficient acting on them and that all equal mutations (changes between unpreferred states) are neutral. These are potentially unrealistic assumptions. However, changing our assumption that for some amino acids multiple codons are optimal, to one codon being optimal leads to qualitatively similar conclusions, suggesting that our results are largely robust to small perturbations of the model. Nonetheless, we cannot exclude that the specific parameter estimates presented here are somewhat affected by various inadequacies of the model assumptions. The models used here assume a constant effect among sites and among amino acids. Many previous studies have found that modeling of variation in the substitution pattern among sites, or among amino acids, can significantly improve the model fit (e.g., Yang et al. 1994; Goldman et al. 1998; Yang et al. 2000). The methods presented here have made no attempts at modeling this type of variation and can therefore in this regard still be substantially improved. The analysis based on the concatenated data has the additional drawback that variation in the parameters among genes is not taken into account. Although it is unlikely that any of our major conclusions have been affected by this, we note that more sophisticated methods for meta-analyses of multiple genes, that do take variation among genes into account, could improve the current analysis.

Bauer DuMont et al. (2004) concluded that there has been an acceleration of unpreferred changes in the *Notch* locus along the *D. melanogaster* lineage. Our analysis confirms this observation, with a strongly significant likelihood ratio test indicating positive selection for apparent unpreferred synonymous mutations in this locus. Simple relaxation of constraint and the apparent change in mutation bias along this lineage cannot explain the data observed for the 3' end of the *Notch* locus. Also, the difference in the likelihood ratio from the 3' end of the *Notch* locus compared with all other loci is so extreme that the *Notch* locus must be considered a clear outlier. It is also interesting to notice that the selection on the *Notch* 3' end is also extremely strong in *D. simulans* but is favoring apparent preferred codons. This raises the possibility that lineage-specific changes in the expression and/or function of the *Notch* locus is being modulated by codon usage in this locus.

In the other loci, a clear picture is emerging from this likelihood analysis. Selection for optimal codon usage is affecting *D. yakuba* and *D. simulans* but is only weakly affecting *D. melanogaster*, if at all. However, there appears to have been a change in the mutational process in the *D. melanogaster* toward lower mutation rate of T to C mutations and

A to G mutations. This change is observed in both exonic and intronic regions and can, therefore, not be explained by possible model inadequacies. On the other hand, a model involving only a change in the mutation process can also not explain the data. The reduction in selection intensity in the *D. melanogaster* lineage inferred from the phylogenetic analysis is corroborated by the population genetic data both using the number of preferred to unpreferred changes within and between species and by considering the frequency spectrum on unpreferred and preferred polymorphisms.

The conventional explanation for the apparent reduction in codon bias along the *D. melanogaster* lineage has been a decrease in population size in this species. However, inferring effective population size for a species is notoriously difficult, and while differences in genomic patterns of variations between *D. melanogaster* and *D. simulans* suggests a smaller effective population size in the former species, there are caveats to this interpretation (as discussed in Capy and Gibert 2004; Morton et al. 2004). It should also be emphasized that the gene-by-gene analyses reveals that a model assuming no selection on the *D. melanogaster* lineage cannot explain all our observations as  $S_m = 0$  can be rejected for the *Notch* locus. The fact that *D. melanogaster*, presumably, has experienced both a change in the mutation process and a reduction of the intensity of selection may explain why the pattern of codon usage in *D. melanogaster* has remained such an enigma in studies of molecular evolution. The possibility that selection is working to fix apparently unpreferred mutations in some loci may be a further contributing factor.

The methods in this paper are readily applicable to the large genomic data sets currently being generated in various *Drosophila* species. Applications to these large data sets will help further elucidate the evolution of codon usage bias in *Drosophila*. One note of caution, however, arises from the current study. It is clear that in lineages affected by selection for optimal codon bias, the strength of selection will be underestimated because polymorphisms and interspecific substitutions will be confounded. The collection of appropriate polymorphism data will allow this key discrimination to be made.

A program performing the analyses discussed in this paper will be available from <http://www.binf.ku.dk/~rasmus/webpage/programs.html>.

## Acknowledgments

This work was supported by National Science Foundation/National Institutes of Health (NIH) grant DMS/NIGMS—0201037 to R. Durrett, R.N., and C.F.A, a grant from the Danish National Science Foundation to R.N., and by NIH grant GM36431 to C.F.A.

## Appendix 1 GenBank Accession Numbers of Sequences Analyzed

Listed below are the genes used in this study and their GenBank accession numbers (*D. melanogaster* (mel), *D. simulans* (sim), *D. yakuba* (yak)): *Adh* (mel: M17827, M17837, M19547, M17828, M17830, M17831, M17832, M17833, M17834, M17835, M17836; sim: M19263, X57364, X57363, X57362, X57361; yak: X57366), *anon1A3* (mel: AF161745, AF161746, AF161727, AF161735, AF161736,

AF161737, AF161743, AF161744, AF161729, AF161731, AF161738, AF161739; sim: AF161749, AF161759, AF161750, AF161752, AF161753, AF161754, AF161756, AF161758, AF161760, AF161755, AF161757, AF161751; yak: AF005844), *anon1E9* (mel: AF161764, AF161765, AF161767, AF161770, AF161771, AF161772, AF161773, AF161766; sim: AF161776, AF161777, AF161778, AF161779, AF161780, AF161781, AF161782, AF161783; yak: AF005848), *anon1G5* (mel: AF005865, AF005879, AF005880, AF161786, AF005866, AF005867, AF005868, AF005869, AF005870, AF005871, AF005873, AF005878; sim: AF005874, AF005875, AF005876, AF161787, AF161788, AF161789, AF161790; yak: AF005852), *Hex-A* (mel: AF257523, AF257532, AF257533, AF257534, AF257524, AF257525, AF257526, AF257527, AF257528, AF257529, AF257530, AF257531; sim: AF257609, AF257618, AF257619, AF257620, AF257610, AF257611, AF257612, AF257613, AF257614, AF257615, AF257616, AF257617; yak: AF257650), *Hex-C* (mel: AF257540, AF257549, AF257550, AF257551, AF257541, AF257542, AF257543, AF257544, AF257545, AF257546, AF257547, AF257548; sim: AF257623, AF257632, AF257633, AF257634, AF257635, AF257636, AF257624, AF257625, AF257626, AF257627, AF257628, AF257629, AF257630, AF257631; yak: AF257651), *Hex-T1* and *Hex-T2* (mel: AF257590, AF257599, AF257600, AF257601, AF257591, AF257592, AF257593, AF257594, AF257595, AF257596, AF257597, AF257598, AF257602; sim: AF257637, AF257642, AF257646, AF257647, AF257648, AF257649, AF257638, AF257640, AF257641, AF257643, AF257645, AF257639, AF257644; yak: AF257652), *mtH*: (mel: AF280552, AF280561, AF280563, AF280553, AF280554, AF280555, AF280556, AF280557, AF280558, AF280559, AF280560; sim: AF280602, AF280593, AF280592, AF280591, AF280601, AF280600, AF280599, AF280598, AF280597, AF280596, AF280595, AF280594; AF280583), *N3'* (mel: AF360583, AF360581, AF360582, AF360584, AF360585, AF360586, AF360587, AF360588, AF360589, AF360590, AF360591, AF360592, AF360594, AF360595; sim: AY191373, AY191369, AY191370, AY191371, AY191372, AY191374, AY191375, AY191376, AY191377, AY191378, AY191379, AY191380; yak: AY191414), *N5'* (AF361407, AF361408, AF361409, AF361410, AF361411, AF361412, AF361413, AF361414, AF361415, AF361416, AF361417, AF361418, AF361419, AF361420, AF361421; sim: AY191395, AY191391, AY191392, AY191393, AY191394, AY191395, AY191396, AY191397, AY191398, AY191399, AY191400, AY191401, AY191402; yak: AY191413), *per* (mel: L07817, L07818, L07819, L07821, L07823, L07825; sim: L07826, L07832, L07828, L07829, L07830, L07831; yak: X61127), *Pgi* (mel: L27539, L27554, L27555, U20566, U20567, U20568, U20569, U20570, U20571, U20572, U20573, L27540, U20574, U20575, L27541, L27542, L27543, L27544, L27545, L27546, L27553; sim: L27547, U20559, U20560, U20561, U20564, U20565, L27548, L27549, L27550, L27551, L27552, U20556, U20557, U20558; yak: L27673), *Pgm* (mel: AF290313, AF290328, AF290330, AF290331, AF290315, AF290316, AF290317, AF290323, AF290324, AF290325, AF290326, AF290327; sim: AF290366,

AF290367, AF290368, AF290369, AF290358, AF290359, AF290360, AF290361, AF290362, AF290363, AF290364, AF290365, AF290357; yak: AF290370), *Rel* (mel: AF204284, AF204286, AF204287, AF204285, AF204288, AF204289; sim: AF204277, AF204278, AF204279, AF204280, AF204281, AF204282, AF204283; yak: AF204290), *Tpi* (mel: U60836, U60845, U60846, U60847, U60837, U60838, U60839, U60840, U60841, U60842, U60843, U60844, U60851, U60853, U60854; sim: U60861, U60862, U60863, U60864, U60865, U60866, U60867, U60868), U60869; yak: U60870), *z* (mel: L13045, L13043, L13044, L13046, L13047, L13048; sim: L13050, L13049, L13051, L13052, L13053, L13055; yak: AF255327), *Zw* (mel: U42738, U42747, U42748, U42749, U43165, U43166, U43167, U44721, U45985, U42739, U42740, U42741, U42742, U42743, U42744, U42745, U42746; sim: L13891, L13892, L13893, L13894, L13876, L13877, L13878, L13879, L13881, L13882, L13883, L13884; yak: U42750).

### Literature Cited

- Akashi H. 1995. Inferring weak selection from patterns of polymorphism and divergence at "silent" sites in *Drosophila* DNA. *Genetics*. 139:1067–1076.
- Akashi, H. 1996. Molecular evolution between *Drosophila melanogaster* and *D. simulans*: reduced codon bias, faster rates of amino acid substitution, and larger proteins in *D. melanogaster*. *Genetics*. 144:1297–1307.
- Akashi H, Ko W-Y, Piao S, John A, Goel P, Lin C-F, Vitins AP. 2006. Molecular evolution in the *Drosophila melanogaster* species subgroup: frequent parameter fluctuations on the time-scale of molecular divergence. *Genetics*. 172:1711–1726.
- Bartolomé C, Maside X, Yi S, Grant AL, Charlesworth B. 2005. Patterns of selection on synonymous and non-synonymous variants in *Drosophila miranda*. *Genetics*. 169:1495–1507.
- Bauer DuMont V, Fay JC, Calabrese PP, Aquadro CF. 2004. DNA variability and divergence at the *Notch* locus region of *Drosophila melanogaster* and *D. simulans*: a case of accelerated synonymous site divergence. *Genetics*. 167:171–185.
- Capy P, Gilbert P. 2004. *Drosophila melanogaster*, *Drosophila simulans*: so similar yet so different. *Genetica*. 120:5–16.
- 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.
- De Cock JGR, Klink EC, Ferro W, Lohman PHM, Eeken JCJ. 1992. Neither enhanced removal of cyclobutane pyrimidine dimers nor strand-specific repair is found after transcriptoin induction of the beta-3-tubulin gene in a *Drosophila* embryonic cell line Kc. *Mutat. Res.* 293:11–20.
- Duret L. 2002. Evolution of synonymous codon usage in metazoans. *Curr Opin Genet Dev.* 12:640–649.
- Felsenstein J. 1981. Evolutionary trees from DNA sequences—a maximum-likelihood approach. *J Mol Evol.* 17:368–376.
- Goldman N, Thorne JL, Jones DT. 1998. Assessing the impact of secondary structure and solvent accessibility on protein evolution. *Genetics*. 149:445–458.
- Goldman N, Yang Z. 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol Biol Evol.* 11:725–736.
- Kern AD, Jones CD, Begun DJ. 2002. Genomic effects of nucleotide substitutions in *Drosophila simulans*. *Genetics*. 162:1753–1761.
- Kimura M. 1962. On the probability of fixation of mutant genes in a population. *Genetics*. 47:713–719.
- Kliman RM. 1999. Recent selection on synonymous codon usage in *Drosophila*. *J Mol Evol.* 49:343–351.
- Kliman RM, Hey J. 1993. Reduced natural selection associated with low recombination in *Drosophila melanogaster*. *Mol Biol Evol.* 10:1239–1258.
- Maside X, Lee AW, Charlesworth B. 2004. Selection on codon usage in *Drosophila americana*. *Curr Biol.* 14:150–154.
- McDonald JH, Kreitman M. 1991. Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature.* 20:652–654.
- McVean GAT, Vieira J. 2001. Inferring parameters of mutation, selection, and demography from patterns of synonymous site evolution in *Drosophila*. *Genetics*. 157:245–257.
- Morton RA, Choudhary M, Cariou M-L, Singh RS. 2004. A reanalysis of protein polymorphism in *Drosophila melanogaster*, *D. simulans*, *D. sechellia* and *D. mauritiana*: effects of population size and selection. *Genetica*. 120:101–114.
- Munté A, Aguadé M, Segarra C. 1997. Divergence of the *yellow* gene between *Drosophila melanogaster* and *D. subobscura*: recombination rate, codon bias and synonymous substitution. *Genetics*. 147:165–175.
- Munté A, Aguadé M, Segarra C. 2001. Changes in the recombinational environment affect divergence in the *yellow* gene of *Drosophila*. *Mol Biol Evol.* 18:1045–1056.
- Muse SV, Gaut BS. 1994. A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome. *Mol Biol Evol.* 11:715–724.
- Nei M, Gojobori T. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol.* 3:418–426.
- Petrov DA, Hartl DL. 1999. Patterns of nucleotide substitution in *Drosophila* and mammalian genomes. *Proc Natl Acad Sci USA.* 96:1475–1479.
- Powell JR, Moriyama EN. 1997. Evolution of codon usage bias in *Drosophila*. *Proc Natl Acad Sci.* 94:7784–7790.
- Rodríguez-Trelles F, Tarrío R, Ayala FJ. 1999. Switch in codon bias and increased rates of amino acid substitution in the *Drosophila saltans* species group. *Genetics*. 153:339–350.
- Rodríguez-Trelles F, Tarrío R, Ayala FJ. 2000. Fluctuating mutation bias and the evolution of base composition in *Drosophila*. *J Mol Evol.* 50:1–10.
- 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.
- Singh ND, Arndt PF, Petrov DA. 2005. Genomic heterogeneity of background substitutional patterns in *Drosophila melanogaster*. *Genetics*. 169:709–722.
- Takano-Shimizu T. 1999. Local recombination and mutation effects on molecular evolution in *Drosophila*. *Genetics*. 153:1285–1296.
- Takano-Shimizu T. 2001. Local changes in GC/AT substitution biases and in crossover frequencies on *Drosophila* chromosomes. *Mol Biol Evol.* 18:606–619.
- Van Der Helm PJJ, Klink EC, Lohman PHM, Eeken JCJ. 1997. The repair of UV-induced cyclobutane pyrimidine dimers in the individual genes *Gart*, *Notch* and *white* from isolated brain tissue of *Drosophila melanogaster*. *Mutat Res.* 383:113–124.
- Yang Z, Goldman N, Friday AE. 1994. Comparison of models for nucleotide substitution used in maximum likelihood phylogenetic estimation. *Mol Biol Evol.* 11:316–324.
- Yang Z, Nielsen R, Goldman N, Pedersen A-MK. 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics*. 155:431–449.

Arndt von Haeseler, Associate Editor

Accepted October 10, 2006