RESEARCH ARTICLES

Developmental Stage and Level of Codon Usage Bias in Drosophila

Saverio Vicario,[*1] *Christopher E. Mason,*[†] *Kevin P. White,*[†] *and Jeffrey R. Powell*[*]

*Department of Ecology and Evolutionary Biology, Yale University; and †Department of Genetics, School of Medicine, Yale University

Codon usage bias (CUB) is a ubiquitous observation in molecular evolution. As a model, Drosophila has been particularly well-studied and indications show that selection at least partially controls codon usage, probably through selection for translational efficiency. Although many aspects of Drosophila CUB have been studied, this is the first study relating codon usage to development in this holometabolous insect with very different life stages. Here we ask the question: What developmental stage of *Drosophila melanogaster* has the greatest CUB? Genes with maximum expression in the larval stage have the greatest overall CUB when compared with embryos, pupae, and adults. (The same pattern was observed in *Drosophila pseudoobscura*, see Supplementary Material online.) We hypothesize this is related to the very rapid growth of larvae, placing increased selective pressure to produce large amounts of protein: a 300-fold increase requiring an approximate doubling of protein content every 10 h. Genes with highest expression in adult males and early embryos, stages with the least de novo protein synthesis, display the least CUB. These results are consistent with the hypothesis that CUB is caused (at least in part) by selection for efficient protein production. This seems to hold on the individual gene level (highly expressed genes are more biased than lowly expressed genes) as well as on a more global scale where genes with maximum expression during times of very rapid growth and protein synthesis are more biased than genes with maximum expression during times of low growth.

Introduction

Codon usage bias (CUB), the unequal use of synonymous codons, is a ubiquitous observation in virtually every organism in which it has been examined. In unicellular organisms, that is, bacteria and yeast, a major route of selection for uneven codon usage is via translational efficiency related to the relative abundance of isoaccepting transfer RNAs (tRNAs) (Gouy and Gautier 1982; Grosjean and Fiers 1982; Sharp et al. 1995). Codons most efficiently translated by the most abundant tRNAs are favored. By efficiency, we subsume speed and accuracy because it has been shown that messenger RNA (mRNA) with a preponderance of optimal codons are both more rapidly translated as well as more accurately, that is, fewer misincorporations of amino acids (Dix and Thompson 1989; Akashi 1994). Because all genes share the same protein synthesizing machinery, including tRNA pools, genes in unicellular organisms converge on a common pattern of CUB, especially strong for highly expressed genes.

When considering multicellular eukaryotes, two complexities arise with regard to the tRNA hypothesis to explain codon usage. First, different tissues may have different relative levels of isoaccepting tRNAs and thus select for different codons. In Drosophila, at least, there is little evidence for this. For example, genes with tissue-specific expression such as amylase in midguts, alcohol dehydrogenase in malphigian tubules, and chorion and yolk protein genes in ovaries all have the same pattern of codon usage, as do genes expressed in all tissue types such as myosin, actin, ribosomal proteins, or genes involved in general metabolism (Powell and Moriyama 1997). Exceptions to this seeming lack of tissue

specificity in Drosophila are the silk glands of both the silk moth and spiders where tRNA pools are adjusted specifically to maximize production of the silk protein (Garel 1974; Sprague 1995), although in this case the proportions of isoaccepting tRNAs do not change but rather the relative level of different families is adjusted to match the amino acid content of the silk protein.

A second confounding factor in complex multicellular eukaryotes is that they often undergo distinct developmental stages so that it is conceivable that tRNA pools could change during development and/or that the strength of selection on codon usage may vary at different life stages. Here we examine a well-studied model organism that has very distinct developmental stages both temporally and ecologically, the holometabolous insect Drosophila. Drosophila has been particularly well analysed for codon usage. A large number of studies have examined such issues as patterns of CUB (Shields et al. 1988; Akashi and Schaeffer 1997; Moriyama and Powell 1997), effects on translational accuracy (Akashi 1994), relationship to tRNA abundance (Shields et al. 1988; Moriyama and Powell 1997; Akashi 2001), effects of recombination (Kliman and Hey 1994, 2003; Powell and Moriyama 1997; Comeron et al. 1999; Marais et al. 2001; Comeron and Kreitman 2002), relation to levels of gene expression (Shields et al. 1988; Sharp and Lloyd 1993; Moriyama and Powell 1997; Powell and Moriyama 1997; Carlini and Stephan 2003), gene length (Moriyama and Powell 1998; Comeron et al. 1999), effect of CUB on rates of silent substitutions (Sharp and Li 1989; Moriyama and Gojobori 1992; Powell and Moriyama 1997), effects of population size (Akashi 1996; Kliman 1999; Maside et al. 2004), and variation in CUB within genes (Gleason and Powell 1997; Comeron and Kreitman 2002; Qin et al. 2004). Virtually all of these studies have concluded that natural selection affects codon usage and several implicate translational efficiency mediated by relative abundances of isoaccepting tRNAs as the selective agent.

For the studies presented here, we took advantage of the complete genome sequences and gene expression data

Fig. 1.—General scheme of microarray experiments to determine levels of gene expression. Letters in circles indicate stages of development from which cDNA was prepared. Numbered arrows indicate each hybridization experiment with head of arrow indicating red dye and tail of arrow green dye. Note the balance of scheme: each stage is assayed four times.

Fig. 2.—Base composition effects of probes used in microarray experiments. Mean observed intensity considering probe base composition of all four nucleotides individually.

as assessed by microarray studies for both *Drosophila melanogaster* (and *Drosophila pseudoobscura* presented in Supplementary Material online) available to us at the initiation of these studies (Stolc et al. 2004; Mason 2006). Because most genes are expressed at more than one developmental stage, we had to define "stage of expression." We first identified those genes that had statistically significant changes in level of expression through development and defined stage of expression as the stage of maximum expression. The reasoning is that selection for codon usage, if linked to translational efficiency, should be strongest when a gene has highest expression. This approach, although very straightforward, does not take into account that genes differentially expressed could have more than one developmental stage in which they are expressed at high levels. For this reason, we also implemented a weighted approach, in which the impact of individual genes on the average CUB of each developmental stage is proportional to the percent of its total expression at that stage.

Materials and Methods
Data

Coding sequences from *D. melanogaster* annotation 4.2 were downloaded from FlyBase and checked for presence of start and stop codons in the correct location. For each transcript, a codon usage table was built.

Expression data for six developmental stages for all known coding sequences were obtained for *D. melanogaster* (Stolc et al. 2004). Probes were designed based on annotation 4.0 in *D. melanogaster*. The six stages compared in microarray experiments are early embryo (E0, first zygotic division to 3 h), late embryo (E3, 3 h to hatching), larvae (L, all three instars), pupae (P), and adult stages, both female (F) and male (M), were up to 10 days posteclosion. (The term "stage" here refers to both the immature developmental stages as well as the different sexes of adults.) For each stage, several individuals were sampled, spaced in time of development within the boundary of each stage. Female adults were collected for total RNA extraction with eggs present in their abdomen. The design of the experiments (fig. 1) pro-

vides that each stage was tested four times, using two differently colored dyes twice, that is, the study is balanced; this should minimize bias due to some stages being more thoroughly examined than others. Details on the collection of the individuals, production of the cDNA libraries, and the hybridization can be retrieved in Stolc et al. (2004).

To minimize codon sampling error for the calculation of CUB indices and the estimation of CUB at gene and stage level, we only used transcripts $\geq$200 codons ($\geq$600 bp).

Analysis of Expression Data

First we corrected for base composition of the probes used. Generally, this is done by simply considering the GC:AT ratio. Here we show that each base has a unique effect on strength of signal (fig. 2), so all data were corrected considering individual base effects.

The data available for *D. melanogaster* that we analysed was the expression level of each transcript. This approach is quite novel (Wolfinger et al. 2001), given that generally analysis of microarray data producing results is at the level of a single probe per gene. With the emerging evidence that a very high fraction of protein-coding genes have multiple transcripts due to alternative splicing and other complications, it is difficult to precisely define "level of gene expression" in any detailed and accurate manner. This new challenge is being addressed by the design of the microarray experiments that had several probes for the same gene and often several probes for the same exon. Before the analyses, all probes that in light of the last annotations (4.2) were no longer located within an exon were deleted. Probes that were shared by more than one transcript were deleted. This cleaning procedure shrunk our data set in *D. melanogaster* from 18,488 transcripts of annotation 4.2 to 13,081 transcripts for which we have an unambiguous probe set.

Analysis of Variance to Identify Significant Changes in Expression

The data were analysed after being dye linearized and log2 transformed. This transformation was performed in

Table 1
Number of Genes with Maximum Expression at Each Developmental Stage

	E0	E3	L	P	M	F	Total
Number of genes	895	493	657	379	793	144	3291
Mean ENC	50.66	48.79	46.23	49.01	50.86	49.28	49.29
Mean F_{op}	0.4720	0.4970	0.5241	0.4880	0.4548	0.4919	0.4851

order to have the data normally distributed and to interpret analysis of variance (ANOVA) coefficients as log ratio (Kerr et al. 2000). This also allows for the multiplicative correction of the signal for probe composition. This correction is equivalent to a Languimir process (Hekstra et al. 2003) but without correcting for chemical saturation and noise caused by sequence-specific mismatch, although probe composition is taken into account. The ANOVA model was the following:

$$
\begin{aligned}
\mathrm{Log}(I_{ijkp}) &= \bar{x} + A_i + D_j + AD_{ij} + V_k + C_A n_A \\
&+ C_C n_C + C_G n_G + G_g + VG_{kt} + AP_{ip} + \epsilon_{ijkp}, \\
i &= 1, \ldots, 12, \quad j = 1, 2, \quad k = 1, \ldots, 6, \\
t &= 1, \ldots, 13081, \quad p = 1, \ldots, 44030.
\end{aligned} \quad (1)
$$

The $\bar{x}$ is the overall mean. The three upper case Cs are the mean effect of substituting a nucleotide (subscript) multiplied by the number of nucleotides of that sort (n) to a theoretical probe of only thymines. A, D, and AD are the mean effect of array, dye, and array by dye interaction. G and VG give the mean effect of gene and gene by variety (=developmental stage). AP indicates the mean effect of the interaction probe array. The ϵ represents the residuals. The subscripts $i, j,$ and k denote array number, type of dye, kind of variety (=developmental stage), respectively. The subscript p denotes which probe or spot on the array was used. The subscript t denotes the transcript. The goal is to estimate overall behavior for a transcript across stages and probes (Wolfinger et al. 2001), with the addition of explicit corrections for sequence effect. The parameters of major interest are the coefficients VG that give the relative expression of each transcript in each stage. An estimation of absolute expression (comparable across transcripts) in each stage could be obtained summing the G terms with the corresponding VG terms and subtracting the average of the AP term for that transcript. For downstream use, these measures are then transformed for each stage into percentage over the total expression of the gene/transcript over the six developmental stages.

The optimization of the parameters of the ANOVA was performed in two steps given the large number of parameters relative to each gene. In the first step, the model to optimize was the following:

$$
\begin{aligned}
\mathrm{Log}(I_{ijkp}) &= \bar{x} + A_i + D_j + AD_{ij} + V_k + C_A n_A \\
&+ C_C n_C + C_G n_G + e_{ijkp}.
\end{aligned}
$$

Then the residuals were used as observed value in the second level of optimization that was performed for each gene independently:

$$
\begin{aligned}
\epsilon_{ijkp} &= G_0 + VG_{kt} + AP_{ip} + \iota_{ijkp}, \\
\epsilon_{ijkp} &= G_1 + VG_{kt} + AP_{ip} + \iota_{ijkp}, \\
&\vdots \\
\epsilon_{ijkp} &= G_n + VG_{kt} + AP_{ip} + \iota_{ijkp}.
\end{aligned}
$$

This was possible only because the design of the experiment is equalized with respect to each probe in each of the 12 hybridizations (fig. 1); thus, parameters A, D, AD, V, C_A, C_C, C_G are orthogonal to the parameters relative to the probe and probe sets relative to one gene.

A summary of the statistically significantly differentially expressed genes divided into six groups depending on when expression peaked is in table 1.

Analysis of Overall CUB through Development

To address the issue of overall (average) CUB at different developmental stages, it is possible to use a nonparametric approach to test deviations from the null hypothesis of equal selection pressure across stages. Based on the criteria above, we subdivided our genes into six nonoverlapping groups based on the six stages (table 1). We tested for differences in mean and overall distribution shape among the six groups. For the study of the means, we estimated both the mean effective number of codon (ENC, Wright 1990) and mean frequency of optimal codon, F_{OP} (Sharp and Lloyd 1993), of each gene for each stage. Note that ENC is a nondirectional measure of unevenness of use of codons, whereas F_{OP} measures deviation from an optimal state and is thus directional. The definition of the optimal set of codon, necessary to calculate F_{OP}, was defined using the results of Vicario et al. (2007). For each amino acid, we choose only the most preferred codon disregarding secondary preferred codons. The significance of the difference across stages was assessed by randomizing the assignment of the genes to the stages. This null distribution that represent the distribution values if there were no correlation between genes CUB and the stages when the genes are most expressed was used to test the existence of the correlation and to test the significance of the difference of the stage characterized by gene with higher CUB with the other stages. We compared the six distributions over the mean, median, and first and third quartiles, this to appreciate the overall difference among distributions.

To identify stages that had similar mean values, a parametric linear model was used to assess what stages had statistically indistinguishable levels of CUB and grouped them. A model for comparison was constructed in a least

squares framework. All stages that had coefficients with overlapping confidence intervals were tentatively grouped; the groups were then confirmed by comparing the residual sum of squares, given the number of predictors used.

To avoid shortcomings caused by classifying genes into nonoverlapping groups, we took another approach. The percentage of total expression per stage was used to provide an estimated weighted mean of CUB per stage. The significance of the difference in the weighted mean was estimated using a bootstrap protocol. This approach allowed maintaining the totality of the structure of the data (i.e., possible coexpression of genes with similar CUB, possible correlation between CUB and profile of expression, etc.) although still sensitive to whether individual genes were unduly influencing the mean.

To visually inspect the distribution of ENC, a weighted version of the empirical cumulative distribution (ECDF) was derived for each stage. In the ECDF, the abscissa value for each gene is simply the ENC or F_{OP}, whereas the ordinate is a weighted version of the percentile. The weight of a gene is its percentage of total expression at the stage being considered divided by the sum of all the percentages of total expression at that stage. The weighted percentile of a given gene is the sum of the weight of all the genes more biased plus its weight.

Ribosomal Proteins

Changes in overall intensity of CUB across development, if they are caused by pressure for optimal translation due to requirements for high levels of protein synthesis, should be coupled with an increase in ribosomes. In order to verify this prediction, we monitored the expression profile in *melanogaster* of 122 proteins classified in gene ontology as being ribosomal proteins and for which expression data are available (Stolc et al. 2004).

Results

Correction for Base Composition

The effects of base composition of probes on strength of signal are shown in figure 2. Because the effect of each of the four bases is unique, three base composition parameters capture better the influence of probe sequences than simply GC%. In all subsequent analyses, corrections for base composition of probes were incorporated.

Overall Codon Usage Bias through Development

The differentially expressed genes with more than 200 codons were subdivided into six stage categories following the methodology described above and shown in table 1. The empirical raw and weighted cumulative distributions of ENC and F_{OP} for the six stages of development are shown in figure 3. It is conspicuous that larvae have a distribution that is skewed toward low values of ENC (high bias) and higher values of F_{OP}. A formal comparison of the mean ENC and F_{OP} by randomization is shown in table 2 which shows that the larval stage has a lower mean (and median) ENC and higher F_{OP} than expected if no correlation between CUB and stages is assumed (P values $< 10^{-4}$). A more detailed permutation analysis (table 3) that focus on the comparison between larva and the other stages distributions gives slightly different results between the two indices: F_{OP} indicates that the full larva distribution (tested on first and third quartiles, mean, and median) is highly significantly (P values between $<10^{-4}$ and 0.0067) displaced compared with all other stages, whereas for ENC the third quartile (the least biased genes) is not significantly different from the other stages (P value = 0.2899) whereas the rest of the distribution it is (P values between $<10^{-4}$ and 0.0037). This shows that the larva's gene CUB distribution values is overall different than the ones of the others stages and is not the effect of few influential genes.

To check that the effect is indeed relative to the stage—and not to the genes that happen to be expressed in that stage—we performed the same permutation test as in table 3 but using a different rule for the grouping by assigning genes to a stage based on where they had the minimal expression and not the maximal as in previous test. The results of this test confirm our view (cf. table 4) that the new larva gene group consistently has lower CUB than the other stages and this for all four descriptor statistics of the distribution (mean, median, first and third quartiles).

The result from the weighted means approach confirmed the general pattern that larvae have higher CUB than any other stage and early embryo the least bias, along with males in *melanogaster* (lower panes of fig. 3). In all 10,000 pseudoreplicates of the bootstrap, the larvae is the most biased stage (table 5), The Jack–knife analysis (table 5) using samples of 50 and 10% of the total data produced 10,000 pseudoreplicates that had larvae as the most biased nearly 100% of the time for both indices (ENC and F_{OP}). A Jack–knife of 2.5%, equivalent to 82, still had 95.58 and 89.71% of the pseudoreplicates with larvae as the most biased stage for *melanogaster* for ENC and F_{OP}, respectively. This indicates that the patterns observed are not due to a few outstanding genes in the larval stage (e.g., perhaps a few with exceptionally high expression) but is a general pattern across genes. Note that this analysis was done on the weighted data set so all genes contributed to the Jack–knife results depending on the relative level of expression in the larval stage.

It is clear that genes with maximum expression in the larval stage have the greatest level of CUB. (Adult males [for both ENC and F_{OP}] and early embryos [only for ENC] tend to have the less CUB in *D. melanogaster*, although this pattern is weaker than the higher larval pattern and is not consistent in *D. pseudoobscura*, unlike the larval pattern [see Supplementary Material online].)

Both measures of CUB have one or two stages with significantly lower bias than expected (fig. 3 and table 2). The significance of the relative order among stages is confirmed by the more formal test performed using linear model (table 6). Using both indices, models with only three and four groups explained an amount of variance not significantly different from the full model with six groups (P values = 0.66, 0.3183, ENC and F_{op} models, respectively). The grouping found with the two indices are very similar:
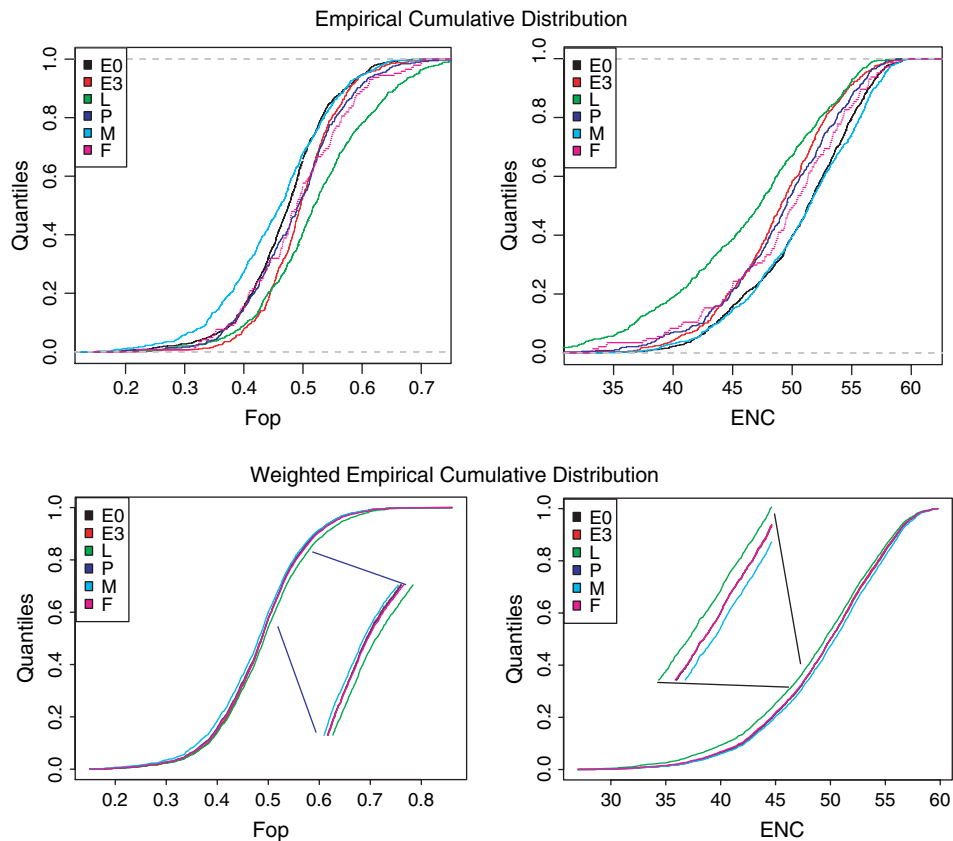
FIG. 3.—Cumulative distributions of genes expressed at different stages (coded by colors indicated in upper right) with different level of CUB measured by ENCs. Upper two graphs are for "raw" or unweighted ENC where a gene is classified by the stage where it has the maximum expression. The lower two graphs are ENC weighted by percentage expression at each stage over the total gene expression at all stages.

larva alone, late embryo, pupae, and female grouped in the middle and male and early embryo with less CUB grouped or not depending from the index used.

In both species, the order of bias of stages does not change with the exception of male. In the Supplementary Material online, we present results from *D. pseudoobscura* confirming that genes with maximum expression at the larval stage have the highest mean CUB. In both species, larvae are the most biased, then female, pupae, late embryo, and finally early embryo. This regularity is perturbed only by the jump of male from the least biased in *melanogaster* to similar to pupae in *pseudoobscura* (see Supplementary Material online).

Ribosomal Proteins

Of the 122 *D. melanogaster* ribosomal proteins examined, 32 vary significantly in level of expression across stages at significance level 0.05 with sequential Bonferroni correction; of these, 26 of them peak in larvae.

Discussion

The results show very clearly that genes with maximum expression at different times in development differ in overall intensity of CUB, using two different measure of CUB and both raw and weighted measures proportional to the level of expression in each stage [fig. 3, tables 2, 3, and 5]). The most striking pattern is the high CUB at the larval stage, with less than average codon bias in early embryos and adult males. Three scenarios could account for a different mean CUB in the different stages and in particular the high CUB in the larval stage. In the first scenario, a small group of genes coexpressed at high level in larvae could cause a shift in the mean value of ENC. Under this

Table 2
Percentage of Randomized Distributions that Have the Value CUB Lower than the Observed

Index	Statistics	E0	E3	L	P	M	F
	Mean	0	0.62	100	0	85.01	2.24
F_{op}	Median	0	14.93	100	0.42	52.58	1.54
	Mean	0	98.26	100	84.30	0	50.36
ENC	Median	0.00	10.95	100	0.14	52.68	6.28

NOTE.—The randomization procedure reassigned genes to stages keeping the same number of gene per stages. A total of 10,000 re-sampling were performed.

Table 3
Permutation Tests on the Displacement of the Distribution of Larva Genes' CUB towards Higher Values

Index	Descriptor	L-Max (no L Stages)	P Value
F_{op}	First quartile	0.0074209	0.0067
	Median	0.02257262	0
	Mean	0.02711801	0
	Third quartile	0.03543438	0
ENC	First quartile	−4.137309	0
	Mean	−2.557924	0
	Median	−1.755515	0.0037
	Third quartile	−0.61258	0.2899

NOTE.—The randomization procedure reassigned genes to stages, keeping the same number of genes per stages. A total of 10,000 resamplings were performed. The statistic used is the difference for a given descriptor of the distribution between larva and the stage, of the remaining ones, that has higher CUB level for that given descriptor. This statistic was calculated for two CUB index (ENC and F_{op}) using mean, median, and first and third quartile as descriptor of the distribution. To obtain a P value, we compared the null distribution obtained by randomization with the observed values of the eight statistics.

Table 4
Permutation Tests on the Displacement of the Distribution of Minimally Expressed in Larva Genes's CUB towards Lower Values

Index	Descriptor	L-Min (not L)	P Value
F_{op}	First quartile	−0.0062	0.018
	Median	−0.018	0
	Mean	−0.021	0
	Third quartile	−0.034	0
ENC	First quartile	5	0
	Median	3.2	0
	Mean	3.6	0
	Third quartile	2.46	0

NOTE.—The tests were performed as in table 3 but the genes were classified depending on where they had minimal relative expression.

scenario, the stage effect is in reality only due to this small subgroup of coexpressed genes. In a second scenario, a majority of the genes that are preferentially expressed in larvae have high level of absolute expression that causes the relatively high level of CUB. This scenario would also predict that other genes that have maximum expression in larvae but with relatively low absolute expression would maintain a level of CUB comparable to genes equally expressed at other stages. Both these scenarios are unlikely in view of the results of the permutation (table 3) and resampling (table 5) analyses in the discrete and weighted classification of the genes, respectively. In fact in the first analysis, using F_{OP}, even the first quartile (low CUB genes) has significantly higher CUB than the others stages, whereas in the second analysis, the same patterns appear when only 2.5% of the genes are randomly sampled. Both this results are compatible with a stage effect that impacts the quasi-totality of the genes that have some expression in larval stage.

A third scenario would be that the translation machinery is used at or near saturation level in larvae, meaning that virtually all ribosomes are attached to mRNA. Consequentially, the selection for optimal translation increases for all the mRNA present in the cytoplasm in order to free up ribosomes as quickly as possible. This scenario does not specify if the genes that cause the translation machinery to be very busy are a few very highly expressed genes or several moderately highly expressed. This is consistent with Kurland's (1991) view that selection for efficient translation does not act only on relatively few genes with high expression, but rather selection is for overall efficiency of cellular protein production. Thus, selection for rapid and accurate translation acts on all genes expressed at times of high demand for protein synthesis, albeit genes with higher concentrations of mRNA would be under greater selection to rapidly free up ribosomes. Our results seem to be most consistent with this view.

That the larval stage is characterized by high level of translational selection is corroborated by the fact that the majority of the ribosomal protein genes that vary in level of expression during development peak in expression at the larval stage (see last paragraph in Results), suggesting that the number of ribosomes is a limiting factor at the larval stage more so than at any other time in development. The reason why translation machinery is under such heavy usage is probably caused by the high rate of growth of larvae. A measure of the translation machinery saturation would be the ratio of free ribosomes to ribosomes attached to mRNA (Bulmer 1991). Unfortunately, this information is available only for oogenesis and embryogenesis (Ruddell and Jacobs-Lorena 1983) and indicates a decrease of attached ribosomes from 79% in late oogenesis to 49% in 1-h embryos consistent with the observed difference between early embryo and female (≈oogenesis because gravid females were used with much of there mass consisting of eggs). For the other stages, it is possible to estimate an indirect measure of pressure for rapid protein synthesis by examining the change in biomass and protein content.

The mass increase between a newly hatched larva to full-grown third instar larva is approximately 200-fold in wet mass and 266-fold in dry mass (Siard et al. 1991). This last value is estimated based on an average egg weight of 0.01125 mg in wet mass, with dry mass 23% the weight of the wet mass (Schreuders et al. 1996) and an average weight for late third instar larvae of 2.25 mg in wet mass and 0.6 mg in dry mass (Santos et al. 1997). The increase in wet biomass is an underestimate of the increase in protein because protein content of wet mass increases from 4% in early larvae to 6% in third instar (Church and Robertson 1966), indicating an approximate 300-fold increase in protein quantity during the larval stage. Given that the larval stage lasts about 4 days (25 °C on laboratory medium), this requires a doubling of protein content approximately every 10 hours.

None of the other stages has an augmentation of protein content comparable to the larvae. Early embryonic development (here defined as up to 3 hours after fertilization) is characterized by very little protein production, although there are large quantities of ribosomes of maternal origin in the embryo (Ruddell and Jacobs-Lorena 1983). Genes with maximum expression in early embryos have the least CUB. Late embryo rebuilds the quasi-totality of the body protein in the ~20 h of development before hatching but with overall protein content constant; this is accomplished by demolition of yolk protein for production of new protein. During the pupal stage, the fly rebuilds the totality of the body, but this replacement of biomass occurs in approximately 4–5

Table 5
Frequency of Time that the Different Stages Had Larger Weighted Average CUB in Jack–Knife Analyses with 10,000 Re-sampling

CUB Index	% Genes Sampled	E0	E3	L	P	M	F
ENC	100%[a]	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000
	50%	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000
	10%	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000
	5%	0.0003	0.0013	0.9967	0.0004	0.0002	0.0011
	2.5%	0.0044	0.0141	0.9558	0.0083	0.0041	0.0133
	1%	0.0336	0.0611	0.7871	0.0446	0.0320	0.0416
F_{op}	100%[a]	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000
	50%	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000
	10%	0.0001	0.0007	0.9992	0.0000	0.0000	0.0000
	5%	0.0024	0.0127	0.9782	0.0012	0.0001	0.0054
	2.5%	0.0178	0.0459	0.8971	0.0124	0.0039	0.0229
	1%	0.0640	0.0943	0.7078	0.0436	0.0366	0.0537

Note.—The first column indicates the CUB index used in the analysis. The second column indicates the percentage of genes sampled in each replicate.

[a] Resampling procedures that allowed sampling with replacement. The weighted average used the relative expression to quantify the contribution of each gene to the estimate of stage CUB.

days at 25 °C, although in those days the protein content roughly doubles (from 6% to 12%, Church and Robertson 1966). Disregarding metabolic turnover, as we did for larvae, in adults the only significant production of biomass is the production of gametes and proteins from gonad-associated accessory cells. Adult males most likely have a small, although unknown, production of net protein per day. This is not true for the female, in which egg production is quite substantial. Stearns et al. (1993) found in *D. melanogaster* colonies reared at 25 °C that between day 4 and day 10 from eclosion females have a peak production of 70 eggs per day, equivalent to 0.7875 mg per day or about 50–70% the female's body mass. But, this biomass is not produced by all cells; two-thirds come from the fat bodies and one-third from the egg chamber (Ruddell and Jacobs-Lorena 1983). Also it is important to note that this increase protein production in gravid females is largely due to a few highly expressed genes such as the three yolk protein genes and seven chorion protein genes (as identified in FlyBase). So this high rate of protein synthesis in gravid females is unlike the overall high protein production in larvae in that it is confined to relatively few genes in a few tissues.

Using this very approximate method to evaluate the effort of protein synthesis, the least protein synthesis is in early embryo and adult male, followed by late embryo, followed by pupae, then gravid females with high protein production limited to a few tissues and genes. Finally, larvae have an average increase in protein content of 300-fold in 4 days, representing all genes required for cell division and growth. This order of protein synthesis output matches the order found for stage effect on CUB: male and early embryo have the lowest CUB, followed by late embryo and pupae, then female, and larvae with the highest. The changes in intensity of CUB are less extreme than the changes in rate of protein synthesis probably because concentrations of ribosomes, especially at the peak of protein synthesis in larvae, attenuate the difference. (*D. pseudoobscura* displays a nearly identical pattern with the exception that males do not show a lower level of CUB relative to females, late embryos, and pupae [see Supplementary Material online].)

This work although novel in metazoa, finds corroboration in previous studies. Sharp et al. (2005) found that intensity of CUB between different bacterial species correlated positively with number of rRNA operons in the genome of each species, even after correction for phylogenetic correlations. The number of rRNA operons was found to be a predictor of the average bacterial growth rate (Klappenbach et al. 2002); higher numbers of rRNA operons are associated with higher growth rates and thus with greater CUB.

The above is concerned with quantitative changes in the strength of selection for codon usage across developmental stages. It does not address qualitative changes in favored codons possibly due to changes in relative abundances of isoaccepting tRNAs. Studies have indicated the relative abundance of isoaccepting tRNAs is remarkably stable across developmental stages of *D. melanogaster* for 19/20 amino acids (White et al. 1973); the exception is aspartic acid. In fact, there is evidence that this single amino acid

Table 6
Grouping of the Stage as Indicated by the Linear Model

	ENC				F_{op}			
	E0–M	E3–P–F	L	P Value	E0	E3–P–F	M	L
Number of genes	1618	1016	657	0.6618	895	1016	657	723
Mean	50.75	48.94	46.23		0.4720	0.4929	0.5241	0.4549

Note.—*P* value express the probability that the excess of the stage grouped model to explain as well that the full stages model.

may change qualitatively in codon preference during development (Vicario 2006; Vicario S, Powell JR, in preparation).

Literature Cited

Akashi H. 1994. Synonymous codon usage in *Drosophila melanogaster*. Natural selection and translational accuracy. Genetics. 136:927–935.

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. 2001. Gene expression and molecular evolution. Curr Opin Genet Dev. 11:660–666.

Akashi H, Schaeffer SW. 1997. Natural selection and the frequency distributions of "silent" DNA polymorphisms in Drosophila. Genetics. 146:295–307.

Bulmer M. 1991. The selection-mutation-drift theory of synonymous codon usage. Genetics. 129:897–907.

Carlini DB, Stephan W. 2003. *In vivo* introduction of unpreferred synonymous codons into the Drosophila *Adh* gene results in reduced levels of ADH protein. Genetics. 163:239–243.

Church RB, Robertson FW. 1966. A biochemical study of the growth of *Drosophila melanogaster*. J Exp Zool. 162:337–351.

Comeron JM, Kreitman M. 2002. Population, evolutionary and genomic consequences of interference selection. Genetics. 161:389–410.

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.

Dix DB, Thompson RC. 1989. Codon choice and gene expression: synonymous codons differ in translational accuracy. Proc Natl Acad Sci USA. 86:6888–6892.

Garel J-P. 1974. Functional adaptation of tRNA population. J Theor Biol. 43:211–225.

Gleason JM, Powell JR. 1997. Interspecific and intraspecific comparisons of the *period* locus in the *Drosophila willistoni* sibling species. Mol Biol Evol. 14:741–753.

Gouy M, Gautier C. 1982. Codon usage in bacteria: correlation with gene expressivity. Nucleic Acids Res. 10:7055–7074.

Grosjean H, Fiers W. 1982. Preferential codon usage in prokaryotic genes: the optimal codon-anticodon interaction energy and the selective codon usage in efficiently expressed genes. Gene. 18:199–209.

Hekstra D, Taussig AR, Magnasco M, Naef F. 2003. Absolute mRNA concentrations from sequence-specific calibration of oligonucleotide arrays. Nucleic Acids Res. 31:1962–1968.

Kerr MK, Martin M, Churchill GA. 2000. Analysis of variance for gene expression microarray data. J Comput Biol. 7:819–837.

Klappenbach JA, Dunbar JM, Schmidt TM. 2002. rRNA operon copy number reflects ecological strategies of bacteria. Appl Environ Microbiol. 66:1328–1333.

Kliman RM. 1999. Recent selection on synonymous codon usage in Drosophila. J Mol Evol. 49:343–351.

Kliman RM, Hey J. 1994. The effects of mutation and natural selection on codon bias in genes of Drosophila. Genetics. 137:1049–1056.

Kliman RM, Hey J. 2003. Hill-Robertson interference in *Drosophila melanogaster*: reply to Marais, Mouchiroud, and Duret. Genet Res. 81:89–90.

Kurland CG. 1991. Codon bias and gene expression. FEBS Lett. 285:165–169.

Marais G, Mouchiroud D, Duret L. 2001. Does recombination improve selection on codon usage? Lessons from nematode and fly complete genomes. Proc Nat Acad Sci USA. 98:5688–5692.

Maside X, Lee AW, Charlesworth B. 2004. Selection on codon usage in *Drosophila americana*. Curr Biol. 14:150–154.

Mason CE. 2006. Genome evolution between Drosophila melanogaster and Drosophila pseudoobscura [PhD dissertation]. [New Haven (CT)]: Yale University.

Moriyama EN, Gojobori T. 1992. Rates of synonymous substitution and base composition of nuclear genes in *Drosophila*. Genetics. 130:855–864.

Moriyama EN, Powell JR. 1997. Codon usage bias and tRNA abundance in Drosophila. J Mol Evol. 45:514–523.

Moriyama EN, Powell JR. 1998. Gene length and codon usage bias in *Drosophila melanogaster*, *Saccharomyces cerevisiae* and *Escherichia coli*. Nucleic Acids Res. 26:3188–3193.

Powell JR, Moriyama EN. 1997. Evolution of codon usage bias in *Drosophila*. Proc Natl Acad Sci USA. 94:7784–7790.

Qin H, Wu WB, Comeron JM, Kreitman M, Li W-H. 2004. Intragenic spatial patterns of codon usage bias in prokaryotic and eukaryotic genomes. Genetics. 168:2245–2260.

Ruddell A, Jacobs-Lorena M. 1983. Abrupt decline in the rate of accumulation of total protein and yolk in postvitellogenic egg chambers of Drosophila. Rouxs Arch Dev Biol. 192:189–195.

Santos M, Borash DJ, Amitabh J, Bounlutay N, Mueller LD. 1997. Density-dependent selection in *Drosophila*: evolution of growth rate and body size. Evolution. 5:420–432.

Schreuders PD, Kassis JN, Cole KW, Schneider U, Mahowald AP, Mazur P. 1996. The kinetics of embryo drying in *Drosophila melanogaster* as a function of the steps in permeabilization: experimental. J Insect Physiol. 42:501–516.

Sharp PM, Averof M, Lloyd AT, Matassi G, Peden JF. 1995. DNA-sequence evolution—the sounds of silence. Phil Trans R Soc Lond B Biol Sci. 349:241–247.

Sharp PM, Bailes E, Grocock RJ, Peden JF, Sockett RE. 2005. Variation in the strength of selected codon usage bias among bacteria. Nucleic Acids Res. 33:1141–1153.

Sharp PM, Li W-H. 1989. On the rate of DNA sequence evolution in *Drosophila*. J Mol Evol. 29:398–402.

Sharp PM, Lloyd AT. 1993. Codon usage. In: Maronni G, editor. An atlas of Drosophila genes. New York: Oxford University Press. p. 378–397.

Shields DC, Sharp PM, Higgins DG, Wright F. 1988. "Silent" sites in *Drosophila* are not neutral: evidence of selection among synonymous codons. Mol Biol Evol. 5:704–716.

Siard T, Jacobson KB, Farkas WR. 1991. Queuine metabolism and cadmium toxicity in *Drosophila melanogaster*. Biofactor. 3:41–47.

Sprague KU. 1995. Transcription of eukaryotic tRNA genes. In: Söll D, RajBhandary UL, editors. tRNA structure, biosynthesis, and function. Washington (DC): ASM Press. p. 31–50.

Stearns SC, Kaiser M, Hillesheim E. 1993. Effect on fitness components of enhanced expression of elongation factor EF-1a. I. The contrasting approaches of molecular and population biologists. Am Nat. 142:961–993.

Stolc V, Gauhar Z, Mason CE, Halsz G, Van Batenburg MF, Rifkin SA, Barbano PE, Bussemaker HJ, White KP. 2004. A gene expression map for the euchromatic genome of *Drosophila melanogaster*. Science. 306:655–660.

Townsend JP, Hartl DL. 2002. Bayesian analysis of gene expression levels: statistical quantification of relative mRNA

level across multiple strains or treatments. Genome Biol. 3:2–16.

Vicario S. 2006. Expressing genes in a complex world: an analysis of codons, growth, and chromosomes in Drosophila [PhD dissertation]. [New Haven (CT)]: Yale University.

Vicario S, Moriyama EN, Powell JR. 2007. Codon usage in twelve species of Drosophila. BMC Evol Biol. 7:226.

White B, Tener G, Holden J, Suzuki D. 1973. Analysis of tRNAs during the development of Drosophila. Dev Biol. 33: 185–195.

Wolfinger RD, Gibson G, Wolfinger ED, Bennett L, Hamadeh H, Busherl P, Afshari C, Paules RS. 2001. Assessing gene significance from cDNA microarray expression data via mixed models. J Comput Biol. 8:625–637.

Wright F. 1990. The effective number of codons used in a gene. Gene. 87:23–29.

Dr. Manolo Gouy, Associate Editor