DOI: 10.1534/genetics.105.049676

Molecular Evolution in the *Drosophila melanogaster* Species Subgroup: Frequent Parameter Fluctuations on the Timescale of Molecular Divergence

Hiroshi Akashi,[1] Wen-Ya Ko, Shengfu Piao, Anoop John, Piyush Goel, Chiao-Feng Lin and Alexa P. Vitins[2]

Institute of Molecular Evolutionary Genetics and Department of Biology, Pennsylvania State University, University Park, Pennsylvania 16802

Manuscript received August 16, 2005
Accepted for publication December 19, 2005

ABSTRACT

Although mutation, genetic drift, and natural selection are well established as determinants of genome evolution, the importance (frequency and magnitude) of parameter fluctuations in molecular evolution is less understood. DNA sequence comparisons among closely related species allow specific substitutions to be assigned to lineages on a phylogenetic tree. In this study, we compare patterns of codon usage and protein evolution in 22 genes (>11,000 codons) among *Drosophila melanogaster* and five relatives within the *D. melanogaster* subgroup. We assign changes to eight lineages using a maximum-likelihood approach to infer ancestral states. Uncertainty in ancestral reconstructions is taken into account, at least to some extent, by weighting reconstructions by their posterior probabilities. Four of the eight lineages show potentially genomewide departures from equilibrium synonymous codon usage; three are decreasing and one is increasing in major codon usage. Several of these departures are consistent with lineage-specific changes in selection intensity (selection coefficients scaled to effective population size) at silent sites. Intron base composition and rates and patterns of protein evolution are also heterogeneous among these lineages. The magnitude of forces governing silent, intron, and protein evolution appears to have varied frequently, and in a lineage-specific manner, within the *D. melanogaster* subgroup.

UNDERSTANDING the forces governing the origins and evolutionary fates of DNA mutations is central to the study of molecular evolution. A great deal of attention has been focused on determining the relative contributions of genetic drift and natural selection to patterns of divergence among genomes (reviewed in Ohta 2002). However, the magnitude, timescale, and genomic breadth of fluctuations in molecular evolutionary forces remain to be studied systematically. Such knowledge is critical for modeling the causes of molecular evolution and is necessary for designing tests of adaptive and deleterious evolution and methods for phylogenetic inference and ancestral state reconstruction.

Determinants of molecular evolution include mutation rates and patterns, effective population sizes, rates of recombination and biased gene conversion, and the fitness effects of mutations. Strict constancy of all these factors is implausible. However, the timescale of parameter fluctuations determines their relevance to molecular evolution; variability in evolutionary forces cause heterogeneous substitution patterns if parameters changes occur on a similar timescale as molecular evolution (Gillespie 1993, 1994; Cutler 2000a,b). Although theoretical concerns suggest that appropriately scaled parameter fluctuations should not be common, numerous studies have invoked nonstationarity to explain variable rates of protein evolution at particular loci or in specific lineages. These include fluctuations in neutral mutation rates (Fitch and Markowitz 1970; Fitch 1971; Takahata 1987), effective population sizes (*e.g.*, Ohta 1987, 1993; Moran 1996; Johnson and Seger 2001; Woolfit and Bromham 2003), and distributions of fitness effects (*e.g.*, Gillespie 1991; Eanes *et al.* 1993; Messier and Stewart 1997; Zhang *et al.* 2002a).

Synonymous codon usage appears to be particularly amenable to microevolutionary analysis of parameter fluctuations. A relatively simple model of "major codon preference" (MCP) that incorporates mutation pressure, genetic drift, and weak selection favoring translationally preferred codons (Li 1987; Bulmer 1991) is supported by both biochemical and population genetic evidence (Ikemura 1985; reviewed in Andersson and Kurland 1990; Sharp *et al.* 1995; Akashi 2001; Duret 2002). A similar scenario (with variation in parameter values) appears to apply across synonymous families and protein-coding genes within a given genome. MCP

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under accession nos. DQ167723–DQ167805.

[1]*Corresponding author:* Department of Biology, Penn State University, 208 Mueller Laboratory, University Park, PA 16802.
E-mail: akashi@psu.edu

[2]*Present address:* Department of Genetics, Washington University School of Medicine, St. Louis, MO 63110.

TABLE 1

Genes examined in this study

Gene	Full name	No. codons	Chromosome	Map	MCU	No. introns	Intron bp
Adh	*Alcohol dehydrogenase*	256	2L	35B3	0.80	3	581
Adhr	*Alcohol dehydrogenase related*	272	2L	35B3	0.48	2	290
Amy-p	*Amylase proximal*	494	2R	53F12	0.89	0	
Amyrel	*Amyrel*	493	2R	53C14	0.65	1	56
AP-50	*AP-50*	425	3R	94A15–16	0.72	2	117
boss	*bride of sevenless*	485	3R	96F8	0.62	2	80
Cyp28c1	*Cyp28c1*	429	X	10F1	0.65	4	109
dpp	*decapentaplegic*	505	2L	22F1–3	0.66	0	
Fur2	*Furin 2*	540	X	14C1	0.61	6	278
g	*garnet*	528	X	12B4–6	0.63	4	152
Gpdh	*Glycerol 3 phosphate dehydrogenase*	254	2L	26A3	0.73	3	380
LanB2	*Laminin B2*	653	3L	67C2	0.65	2	112
Mcm3	*Minichromosome maintenance 3*	496	X	4F5	0.65	2	72
ND75	*NADH:ubiquinone reductase 75-kDa subunit precursor*	547	X	7E1	0.71	3	149
Osbp	*Oxysterol binding protein*	654	3R	96B10	0.66	1	71
pav	*pavarotti*	535	3L	64A10–11	0.54	3	149
per	*period*	618	X	3B1–2	0.74	3	157
RpII215	*RNA polymerase II 215-kDa subunit*	743	X	10C6–7	0.58	1	104
run	*runt*	488	X	19E2	0.72	1	187
ry	*rosy*	1318	3R	87D9	0.59	2	201
twi	*twist*	432	2R	59C2	0.77	1	52
Zw	*Zwischenferment*	517	X	18D13	0.82	2	140

Gene symbols, full names, chromosomal locations, and map positions (in *D. melanogaster)* are from FlyBase (Drysdale *et al.* 2005). No. codons is the number of aligned codons. MCU, major codon usage, is the average value across the six species in the *D. melanogaster* subgroup for the aligned codons. The number of introns studied from these genes and the numbers of aligned nucleotides summed across the introns for a given gene are shown.

predicts fitness classes of silent DNA mutations (translationally preferred and unpreferred changes) (Akashi 1995) and comparisons of evolutionary patterns between the classes can identify weak selection as well as departures from equilibrium and their causes (Akashi 1996). Finally, relatively high levels of variation at silent sites both within and between populations allow inference of patterns and causes of codon bias evolution among closely related species. Evolution under a balance among weak forces is expected to show great sensitivity to parameter fluctuations. Variation in mutation patterns (*e.g.*, Osawa *et al.* 1988; Sueoka 1988; Shields 1990; Rodríguez-Trelles *et al.* 2000; Takano-Shimizu 2001), biased gene conversion (Duret *et al.* 2002; Webster *et al.* 2003), selection intensity ($N_e s$) (*e.g.*, Shields *et al.* 1988; Akashi 1996; Llopart and Aguadé 1999; Zhang *et al.* 2002b; Bachtrog 2003; Peréz *et al.* 2003), recombination (*e.g.*, Takano-Shimizu 1999; Munté *et al.* 2001), and codon preferences (Powell *et al.* 2003) have been argued as causes of among-lineage variation in codon usage and synonymous substitution patterns.

Here, we examine molecular evolution among lineages within the *Drosophila melanogaster* subgroup to identify departures from steady-state silent and protein evolution and to determine their causes. Sequences for 22 genes from six species (*D. melanogaster*, *simulans*, *teissieri*, *yakuba*, *erecta*, and *orena*) were used to reconstruct nucleotides at ancestral nodes within the phylogeny and infer changes on six terminal and two internal lineages. Departures from equilibrium codon bias appear to have been common in the *D. melanogaster* subgroup; three lineages show declines of codon bias and one lineage shows an increase in major codon usage. These patterns appear to be consistent among genes. Intron base composition has also changed within the subgroup and both rates of protein evolution and the nature of nucleotide changes underlying amino acid changes vary among these lineages.

MATERIALS AND METHODS

Drosophila strains: We examined a combination of available DNA sequences and sequences obtained for this study. The strains sequenced for this study were *D. simulans* (251.6), *D. teissieri* (257.0), *D. yakuba* (261.0), and *D. orena* (245.0) from the National Drosophila Species Resource Center (Department of Biological Sciences, Bowling Green State University, Bowling Green, OH) and the Tucson Stock Center (University of Arizona, Tucson, AZ). *D. erecta* (S-18) was kindly provided by M. Ashburner (Department of Genetics, University of Cambridge, U K).

DNA sequences: Table 1 gives gene names, chromosomal regions, and map positions for the 22 genes examined. Sources for available sequences are given in supplemental Table 1 at http://www.genetics.org/supplemental/. Sequences obtained for this study can be found under accession nos. DQ167723–DQ167805.

A number of criteria were employed to select genes for this study. For several of these genes, previous studies had sequenced orthologs from multiple species within the *D. melanogaster* subgroup. Because region-specific molecular evolution may be associated with reduced crossing over in Drosophila (Takano-Shimizu 1999; Begun 2002), only genes located in regions of relatively high recombination in *D. melanogaster* (according to Hey and Kliman 2002) and that are not located close to centromeres or telomeres in other *melanogaster* subgroup species (based on Ashburner *et al.* 2005) were included in this study. In our sequencing efforts, we attempted to sequence ≥400 codons from each gene (to allow detection of gene-specific evolutionary patterns); small genes were not included in data collected for this study and genes with available data were included if ≥200 codons could be aligned among the six species under study. Data were collected for genes with at least one intron and ≥100 bp of intron sequence (in *D. melanogaster*) to allow comparisons of base composition changes in coding and noncoding regions. Uncertainty in alignments is often associated with "simple" and repetitive regions (they are prone to insertion/deletion evolution) and genes were not selected if SEG and NSEG (Wootton and Federhen 1996) detected large proportions of simple amino acid or nucleotide sequences. Genes were chosen to cover a range of codon bias; the sample includes a large fraction of genes with intermediate levels of codon bias, but very highly biased genes were not chosen because ancestral state reconstruction is less robust. Finally, to avoid complications of distinguishing paralogs and orthologs, candidate genes were rejected if BLAST (Altschul *et al.* 1990) searches of the *D. melanogaster* genome revealed more than a single close match. Paralogs of the *Amylase* locus have undergone extensive gene conversion within the *D. melanogaster* subgroup (Shibata and Yamazaki 1995) and available data for one of the two paralogs, *Amy-p*, were included in the analysis.

DNA sequences were obtained through a combination of genomic and "vectorette" PCR and automated sequencing using Beckman (Fullerton, CA) CEQ 8800 and Applied Biosystems (Foster City, CA) 3730 sequencers. The protocols are described in Ko *et al.* (2003). DNA sequences were aligned using ClustalW (Thompson *et al.* 1994) and modified by eye. Codons containing sites that aligned with a gap in any of the sequences were excluded from the analysis. A "conservative" approach was taken in the alignments; regions with questionable alignments were excluded from the analysis since we sought to study nucleotide changes rather than insertion/deletion events. Alignment files employed in this study are available from H. Akashi.

Phylogenetic analysis: Phylogenetic relationships among the six species under investigation were studied using outgroup sequences from *D. ananassae* and *D. pseudoobscura*. Sources of each sequence are listed in supplemental Table 1 at http://www.genetics.org/supplemental/. Details of the phylogenetic analysis are given in supplemental Table 2 at http://www.genetics.org/supplemental/.

Reconstructing ancestral states: Ancestral nucleotides were inferred at internal nodes in the assumed species phylogeny (Figure 1) using maximum-parsimony (MP) and maximum-likelihood (ML) methods. BASEML version 2.0k (Yang 1997) was employed for ancestral state reconstructions at four internal nodes in a specified unrooted tree given sequence data from six species. Given the tree and the sequences, BASEML calculates maximum-likelihood estimates (MLEs) for branch lengths and substitution model parameters and employs the estimates to determine posterior probabilities of states at interior nodes (Yang *et al.* 1995). The output includes joint reconstructions (sets of ancestral nucleotides at internal nodes) and their probabilities. A Jukes–Cantor (Jukes and Cantor 1969) one-parameter model with small and equal branch lengths was used to generate MP reconstructions. For nucleotide sites where BASEML gave multiple joint reconstructions, only the most probable set of ancestral nucleotides (the "most parsimonious") was employed in the analysis. In cases of "ties" between most probable reconstructions for a given nucleotide position ("equally parsimonious" reconstructions), the site was excluded from the analysis. This model-based approach gave very similar results to an iterative parsimony inference approach (data not shown).

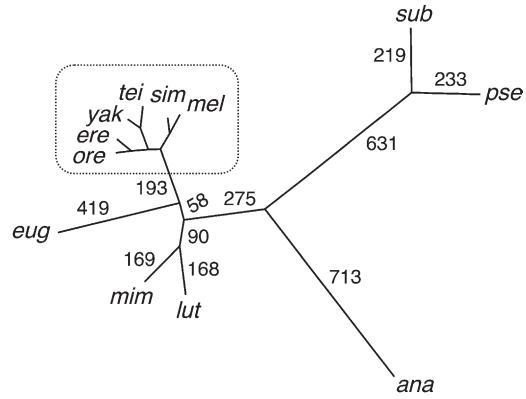

Figure 1.—Phylogenetic relationships within the *melanogaster* subgroup. Unrooted synonymous distance tree of species in the *melanogaster* group. Silent distances, d_s, were calculated individually for *Adh*, *Adhr*, *Gld*, and *ry* genes using CODEML (F3×4 model; Yang 1997). Branch lengths are the averages of silent distances across these four genes (×1000). The *D. melanogaster* subgroup is boxed. All nodes within the subgroup were supported with 100% bootstrap scores for neighbor-joining, parsimony, and ML methods (Ko *et al.* 2003). Abbreviations for species (that are not given in Table 2) are: *eug*, *D. eugracilis*; *mim*, *D. mimetica*; *lut*, *D. lutescens*; *ana*, *D. ananassae*; *pse*, *D. pseudoobscura*; and *sub*, *D. subobscura* (*pse* and *sub* are employed as outgroups). Branch lengths, d_s, are shown except for lineages within the *melanogaster* subgroup species: *mel* (0.073), *sim* (0.052), *tei* (0.076), *yak* (0.050), *ere* (0.063), *ore* (0.053), *melsim* (0.061), *teiyak* (0.077), *ereore* (0.056), and *teiyakereore* (0.042). Lineages are named according to the most recent node (*i.e.*, *melsim* refers to the lineage from the common ancestor of the subgroup to the common ancestor of *D. melanogaster* and *D. simulans*).

The HKY85 (Hasegawa *et al.* 1985) model was employed for maximum-likelihood analysis. The model includes base composition parameters as well as a transition:transversion rate ratio parameter. BASEML allows separate estimation of parameters at the three codon positions based on the nucleotide composition of the extant sequences and patterns of substitution. The instantaneous rate matrix is assumed to be constant throughout the tree. Joint probabilities of ancestral nucleotides at each codon position were multiplied to yield joint probabilities of ancestral codons (assuming independence among positions). For a given lineage, pairs of ancestral and derived codons that differ at a single position are referred to as "reconstructed changes." The probability of each reconstructed change was treated as its count; this approach attempts to account for uncertainty in ancestral state inference by weighting ancestral reconstructions by their probabilities. Reconstructed changes that include a stop codon were removed from the analysis. This method is similar to one employed in Galtier and Boursot (2000) except that model parameters are estimated separately for the three codon positions and rates are assumed to be homogeneous at each position.

TABLE 2

Abbreviations

Abbreviation	Description
MCP	Major codon preference
MCU	Major codon usage = (no. major codons)/(no. major + no. minor codons)
up	Unpreferred → preferred silent change
pu	Preferred → unpreferred silent change
$d_{up,pu}$	(no. up – no. pu)/(no. up + no. pu)
MCU_H	High MCU regions within genes
MCU_L	Low MCU regions within genes
SW	G or C → A or T change
WS	A or T → G or C change
$d_{WS,SW}$	(no. WS – no. SW)/(no. WS + no. SW)
mel	*D. melanogaster*
sim	*D. simulans*
tei	*D. teissieri*
yak	*D. yakuba*
teiyak	*D. teissieri–D. yakuba* (ancestral node)
ere	*D. erecta*
ore	*D. orena*
ereore	*D. erecta–D. orena* (ancestral node)
WSR	Wilcoxon signed-rank test

In cases where ancestral and derived codons differed at two positions, relative probabilities were calculated for the two minimum step paths between the codons. For a given gene, total counts of silent and replacement changes across lineages were obtained for codons that differed at a single position. The ratio of these counts was used to weight the pathways between codons that differed at two positions (as described in Nei and Gojobori 1986). Analyses were also conducted assuming equal probabilities of silent and replacement changes. Codons where ancestral and derived states differed at all three nucleotide positions were excluded from the data for a given lineage (such cases were very rare, see below). Because the lineages examined are short (4–8% silent divergence), probabilities of multiple hits at a given nucleotide site within a lineage were assumed to be negligible.

Classification of DNA changes: Reconstructed codon changes were classified into functional and/or putative fitness classes depending on their ancestral and derived states. "Synonymous families" were defined as sets of codons that encode the same amino acid and that are connected by single-nucleotide changes. Under this definition, sixfold redundant serine codons are treated as separate twofold and fourfold families. Among silent changes, mutations from minor to major codons (according to *D. melanogaster* codon preferences in Akashi 1995) were classified as preferred (up) and changes in the opposite direction were classified as unpreferred (pu). Abbreviations employed in this study are given in Table 2. We employ the difference in the proportions of up and pu changes as a measure of departures from equilibrium, $d_{up,pu} = (up - pu)/(up + pu)$. This measure has a range of −1.0 and 1.0 and its magnitude can be compared directly for increases and declines in codon bias (*e.g.*, $d_{up,pu} = 0.33$ and −0.33 indicate twofold excess of up and up changes, respectively). Mutations within classes were classified as either preferred to preferred (pp) or unpreferred to unpreferred (uu). Within-class changes were included in the total counts for silent changes but were not included in tests of equilibrium because their fitness effects are not predicted.

Replacement changes were also classified by their effect on GC content. Using the ambiguity code for nucleotides, we denote A or T as "weak" (W) and G or C as "strong" (S). A or T → G or C changes are abbreviated WS and changes in the reverse direction are denoted SW. We use $d_{WS,SW} = (WS - SW)/(WS + SW)$ as a measure of departures from equilibrium GC content.

Polymorphism analysis: DNA changes were partitioned into polymorphic and fixed differences using data for multiple alleles in the *D. melanogaster*, *simulans*, and *yakuba* lineages for subsets of the 22 genes. Details of the method are given in the supplemental materials at http://www.genetics.org/supplemental/.

Partitioning codons into high and low major codon usage regions: To divide the data into regions differing in levels of codon bias, within-gene major codon usage was analyzed in sliding windows of 100 codons. The windows were shifted sequentially through the coding sequence and if a given window showed major codon usage (MCU) ≥ 0.7, all codons included in the window were assigned to the high MCU class. Codons that were never included within MCU ≥ 0.7 windows were pooled into a low MCU class (codons that belonged to both high and low MCU windows were assigned to the high MCU category). Each gene was partitioned into high and low MCU regions, MCU_H and MCU_L, respectively, and ancestral state reconstructions were conducted separately for the partitions. High or low MCU regions for a given gene were excluded if they consisted of <30 codons (this led to exclusion of 9 and 8 low MCU codons in *Gpdh* and *per*, respectively). Synonymous family-specific comparisons between MCU_H and MCU_L would control for heterogeneity in fitness effects among silent changes, but the numbers of inferred changes were not sufficient to allow such analysis.

Intron analysis: Sequences from 42 introns from the 19 genes included in the restricted data set for coding region analysis were examined. Additional sequences from 18 introns were available from genes that were not included in the coding region analysis either because they appeared to be outliers with respect to codon bias evolution (*AP-50*, *Osbp*, and *pav*) or because the available coding regions were small [*fruitless* (*fru*), *Cecropin C* (*CecC*), *janusA* (*janA*), *janusB* (*janB*), *ocnus* (*ocn*), *roughex* (*rux*), and *Superoxide dismutase* (*Sod*)]. The introns examined and their sources are given in supplemental Table 3 at http://www.genetics.org/supplemental/. Introns were aligned using a combination of available software [ClustalW (Thompson *et al.* 1994) and MCALIGN (Keightley and Johnson 2004)] and extensive manual adjustment. Many of the alignments included a high proportion of gaps; data are presented for regions of the 60 introns where alignments appeared satisfactory.

Statistical analysis: *G*-tests (adjusted by Williams' correction for continuity) were conducted as described in Sokal and Rohlf (1995). Wilcoxon signed-rank (WSR) tests (Snedecor and Cochran 1989) were employed to test null hypotheses of distributions of $d_{up,pu}$ centered on zero across genes. Tests were performed on the differences between the counts of up and pu changes. Probabilities for WSR tests were calculated by simulating the distribution of the test statistic. For each of 10^6 iterations, the test statistic was recalculated after random assignment of positive and negative signs to the data. The two-tailed probability was the number of simulated test statistics with equal or greater values than the test statistic for the data. All computer programs (written in C) employed in these analyses are available from H. Akashi.

RESULTS

Phylogenetic relationships within the *D. melanogaster* subgroup: Inferred phylogenetic relationships among

TABLE 3

Inference methods and silent and protein changes in the *D. melanogaster* subgroup

Method	Silent	up	pu	$d_{up,pu}$	Replacement
ML *s/r*	3145.7	937.5	1475.5	−0.22	624.7
ML equal	3140.5 (1.00)	936.6 (1.00)	1473.4 (1.00)	−0.22	629.8 (1.01)
ML emp	3096.2 (0.98)	929.1 (0.99)	1453.9 (0.99)	−0.22	560.3 (0.90)
Parsimony	2603.9 (0.83)	642.6 (0.69)	1383.1 (0.94)	−0.37	538.1 (0.86)

The total numbers of inferred silent (as well as preferred and unpreferred) and replacement changes are shown. Counts are summed across eight lineages in the *D. melanogaster* subgroup. ML *s/r*, maximum-likelihood inference using silent/replacement ratios to weight paths between codons differing in multiple positions; equal, equal weights for paths between such codons; emp, data excluding codons differing in multiple positions. The total numbers of inferred changes in each class are also given for parsimony inference. The numbers in parentheses are the ratios of the counts under each inference method and the counts under ML *s/r*. $d_{up,pu}$ is also given for each method.

the six *D. melanogaster* subgroup species studied are shown in Figure 1. Results for a concatenated sequence that includes 22 genes support the division of the subgroup into two major clades, *D. melanogaster–D. simulans* and *D. teissieri–D. yakuba–D. erecta–D. orena*. These results are in agreement with previous analysis of a subset of these genes using different sets of outgroup species (Ko *et al.* 2003) although a number of studies have supported a *D. melanogaster–D. simulans–D. teissieri–D. yakuba* clade (Lachaise *et al.* 1988; Powell 1997). Silent distances among the *D. melanogaster* subgroup species and outgroups are shown in Figure 1. Both outgroups employed in the phylogenetic analysis, *D. ananassae* and *D. pseudoobscura*, are quite distantly related to the species in the *D. melanogaster* subgroup. However, support for this topology is consistent across methods and is consistent with analyses using closer outgroup sequences for smaller sets of genes (Kopp and True 2002; Ko *et al.* 2003) as well as protein trees (Parsch 2003). Gene-specific results are given in supplemental Table 2 at http://www.genetics.org/supplemental/. Larger genes tend to support the topology shown in Figure 1, and alternative topologies tended to be weakly supported by smaller genes. However, data from *boss* favor a *mel–sim–ere–ore* clade with 100% bootstrap support under three different phylogenetic reconstruction methods. The cause of this discrepancy is unclear.

Summary of inferred changes within the *D. melanogaster* subgroup: A total of 11,682 codons were aligned among the six species across the 22 genes and summary data for the inferred numbers of silent and replacement changes under different methods are shown in Table 3. Under our ML inference approach, each pair of ancestral and derived codons that differ at one or more positions is assigned a probability. The sum of such probabilities (across genes and lineages) was 4615.7. Of this total, 98.5% of the probability was assigned to codons that differed at a single position between the ancestral and derived states, 1.4% of the probability was assigned to codons that differed at two positions, and 0.06% of the total probability was assigned to codon reconstructions where the ancestral and derived states differed at all three codon positions (the latter class were not included in the analysis). For codons differing at two positions between ancestral and derived states, equal weighting and silent/replacement weighting gave similar counts for silent and protein changes (most pathways between such codons consisted of equal numbers of silent and replacement changes). However, a significant proportion of replacement changes occurred in codon reconstructions that differed at two sites. Elimination of such reconstructions resulted in a 1.6% reduction in the counts of silent changes and a 10.3% reduction in the counts of inferred replacement changes (Table 3). Reductions in replacement counts were >15% for *Adhr*, *Osbp*, and *Zw*.

Results from MP and ML inference differ considerably for these data (Table 3). The inferred numbers of silent and protein changes were 17 and 14% lower for parsimony than for ML, respectively. This difference partly reflects the exclusion, under MP, of codons with equally parsimonious reconstructions. MP underestimation of changes to more common states (*i.e.*, major codons) (Collins *et al.* 1994) appeared to also have a large effect; compared to ML inference, MP counts were reduced by 6% for pu changes and by >30% for up changes. These results suggest that inference methods must be considered carefully even in analysis of close lineages and that reconstructing pathways for codons that differ at multiple sites may be important for studies of protein evolution among these species. Because simulation results indicated greater inference biases for MP than for ML under the HKY85 model for the levels of codon bias and divergence in these data (our unpublished data), the following analysis employs codon changes inferred under the ML method. Silent: replacement ratios were used to reconstruct changes for codons that differed at multiple positions (all results given below were similar in analyses excluding such codons).

Gene-specific patterns of silent evolution in the *D. melanogaster* subgroup: Table 4 shows gene-specific numbers of inferred silent and replacement changes

TABLE 4

Gene-specific silent and protein changes in the *D. melanogaster* subgroup

Gene	Sil	up	pu	$d_{up,pu}$	Lin dir	Rep	Rep/sil
Adh	55.0	24.4	20.2	0.09	5/3	16.9	0.31
Adhr	103.6	45.1	32.4	0.16	5/3	10.1	0.10
Amy-p	67.3	23.9	29.9	−0.11	4/4	47.8	0.71
Amyrel	157.9	48.7	71.9	−0.19	2/6	55.0	0.35
AP-50	90.4	12.0	56.9	−0.65	0/8	1.0	0.01
boss	117.9	46.5	46.0	0.01	4/4	12.0	0.10
Cyp28c1	201.9	82.7	69.9	0.08	3/5	65.1	0.32
dpp	53.9	10.7	24.8	−0.40	2/5	20.0	0.37
Fur2	113.8	40.5	46.7	−0.07	5/3	32.5	0.29
g	148.5	41.6	66.4	−0.23	2/6	9.3	0.06
Gpdh	39.1	7.3	25.0	−0.55	2/6	2.0	0.05
LanB2	172.2	60.1	76.9	−0.12	3/5	51.9	0.30
Mcm3	188.5	67.5	85.8	−0.12	4/4	2.0	0.01
ND75	124.6	36.0	66.6	−0.30	2/6	9.3	0.07
Osbp	161.2	22.9	102.3	−0.63	0/8	33.1	0.21
pav	194.8	40.9	106.4	−0.44	0/8	9.9	0.05
per	220.9	64.6	101.4	−0.22	2/6	99.0	0.45
RpII215	217.3	49.8	112.6	−0.39	0/8	5.0	0.02
run	82.8	17.8	41.3	−0.40	3/5	6.0	0.07
ry	408.3	131.4	163.5	−0.11	4/4	68.5	0.17
twi	88.4	31.9	46.5	−0.19	2/6	18.9	0.21
Zw	137.6	31.2	82.4	−0.45	1/7	49.6	0.36
Total	3145.7	937.5	1475.5	−0.22		624.7	0.20

Numbers of inferred silent (preferred and unpreferred) and replacement changes are summed across lineages. The data are from maximum-likelihood inference. Sil, silent changes; Rep, replacement changes; Lin dir, the numbers of lineages that show positive/negative $d_{up,pu}$; Rep/sil, the ratio of the counts for replacement and silent changes for a given gene.

summed across eight lineages. Lineage-specific data for each gene are given in supplemental Table 4 at http://www.genetics.org/supplemental/. Across the 22 genes, 3145.7 silent and 624.7 replacement changes were assigned to the eight lineages. Although 18 of the 22 genes show negative overall $d_{up,pu}$ across lineages, almost all genes show positive $d_{up,pu}$ in at least one or two of the lineages. However, several of the genes show large excesses of pu changes in all lineages. *AP-50*, *Osbp*, and *pav* show $d_{up,pu} < -0.2$ in each of the eight lineages examined (*RpII215* also showed consistent declines in MCU, but $d_{up,pu}$ values were closer to zero in some lineages). These three genes were treated as outliers and are candidates for parameter changes prior to the radiation of the *D. melanogaster* subgroup. They were excluded from the analysis because their patterns of evolution would complicate the analysis of parameter changes since the radiation of the subgroup. Potential causes of such cladewide patterns are reductions in expression level (leading to a gene-specific reduction in $N_e s$), regional changes in mutation patterns, and changes in rates of recombination or biased gene conversion. Examination of expression patterns, recombination rates, and patterns of base composition evolution in noncoding regions of these genes within the *D. melanogaster* subgroup and in close outgroups may reveal the cause(s) of these cladewide patterns. None of the 22 genes showed positive $d_{up,pu}$ across the eight lineages.

Tests of lineage-specific departures from equilibrium MCU: Table 5 gives summary data for silent changes among lineages in the *D. melanogaster* subgroup. Lineages are referred to by the upper node. Extant nodes and terminal lineages are abbreviated as follows: *D. melanogaster*, *mel*; *D. simulans*, *sim*; *D. teissieri*, *tei*; *D. yakuba*, *yak*; *D. erecta*, *ere*; and *D. orena*, *ore*, respectively. Ancestral nodes are abbreviated *teiyak* for the common ancestor of the *D. teissieri* and *D. yakuba* lineages and *ereore* for the common ancestor of the *D. erecta* and *D. orena* lineages.

Figure 2 shows gene-specific patterns of silent changes for the eight lineages. The decline of codon bias in *D. melanogaster* is striking; all 19 genes show at least twice as many pu changes as up changes ($d_{up,pu} \leq -0.33$). These strong declines of MCU confirm previous findings (Akashi 1995, 1996; McVean and Vieira 1999; Begun 2001; DuMont *et al.* 2004). Declines of MCU are also consistent among genes in *yak*. Sixteen genes show negative $d_{up,pu}$ and a large fraction of these show large excesses. Three genes show small positive $d_{up,pu}$ but the trend of negative $d_{up,pu}$ is consistent across genes (WSR $P = 0.00028$). *D. orena* shows a similar pattern of

TABLE 5

Lineage-specific silent changes in the *D. melanogaster* subgroup

Lineage	up	pu	$d_{up,pu}$	*G*	Gene dir	WSR *P*
mel	63.6	296.9	−0.65	163.6***	0/19	<0.00001***
sim	85.8	126.3	−0.19	7.7*	6/13	0.039*
tei	160.1	75.1	0.36	31.3***	15/4	0.00032***
yak	94.3	204.4	−0.37	41.5***	3/16	0.00028***
teiyak	140.6	83.5	0.25	14.9***	14/5	0.02*
ere	129.0	112.2	0.07	1.2	10/9	0.24
ore	89.4	161.5	−0.29	21.0***	3/16	0.0045**
ereore	98.6	150.3	−0.21	10.7**	4/14	0.045*

The numbers of inferred unpreferred and preferred silent changes (summed across 19 genes) are shown for each lineage. The data are from maximum-likelihood inference. $d_{up,pu}$ for the total pu and up counts are shown. *G*-test statistics for goodness-of-fit tests with expectations of equal numbers of up and pu changes are shown. Gene dir, the numbers of genes that show positive/negative $d_{up,pu}$; WSR *P*, Wilcoxon signed-rank test probabilities. *$P < 0.05$, **$P < 0.005$, ***$P < 0.0005$.

consistent decline of MCU; although one gene, *ry*, shows a strong trend in the opposite direction, the overall excess of pu changes is consistent across genes ($P = 0.0045$). Evidence for overall declines of codon bias is weaker in *sim* ($P = 0.039$) and *ereore* ($P = 0.045$), where $d_{up,pu}$ values are generally negative but several genes show strong trends in the opposite direction. Only one lineage, *tei*, shows a consistent *increase* in preferred codon usage ($P = 0.00032$). Most genes also show large positive $d_{up,pu}$ in *teiyak*, but several have relatively strong trends in the opposite direction and statistical support for an overall excess of up is weak ($P = 0.023$). The small numbers of silent changes on a given lineage for most loci preclude detection of gene-specific departures from equilibrium. However, *ry*, which contributes the largest number of codons among the 19 genes, shows trends in a direction opposing the rest of the genes in the three lineages in the *D. erecta–D. orena* clade (*ere*, *ore*, and *ereore*). Lineage-specific codon bias changes are consistent among synonymous families (supplemental Table 5 at http://www.genetics.org/supplemental/). However, patterns in the three lineages in the *ereore* clade are strongly dependent on data from *ry*.

Base composition evolution in high and low MCU regions: Base composition changes were compared among MCU_H, MCU_L, and intron regions to investigate the cause(s) of nonequilibrium codon bias evolution. Because all major codons in *D. melanogaster* are G- or C-ending, $d_{WS,SW}$ and $d_{up,pu}$ measures are similar for these data. Synonymous DNA changes were partitioned into those that occurred in MCU_H and MCU_L regions. Sliding windows of 100 codons and an MCU cutoff of 0.7 divided the data into 6024 codons from high-bias regions of 17 genes and 4027 codons from low-bias regions of 13 genes (11 genes contributed codons to both partitions). This division gives similar overall numbers of silent fixations (1472.8 and 1224.3, respectively) in MCU_H and MCU_L. The average MCU values (across species) are 0.745 and 0.564 for the partitions.

Summary data for silent changes in the MCU partitions are given in Table 6. Among the four lineages that show consistent departures from equilibrium codon bias, *tei*, *yak*, and *ore* show stronger departures from equilibrium in high-bias than in low-bias regions (Figure 3). Although the trend is similar in *mel*, the magnitude of the difference in $d_{WS,SW}$ between MCU_H and MCU_L is small. *ry*, an apparent outlier in the *ere*, *ore*, and *ereore* lineages (Figure 2), contributes 1211 of 4027 codons to MCU_L. Exclusion of *ry* data in these lineages has little effect on the $d_{WS,SW}$ of MCU_H regions (*ry* contributes 107 of the 6024 codons), but $d_{WS,SW}$ values for MCU_L regions change from −0.01 to 0.20 in *ere*, from −0.30 to −0.20 in *ore*, and from 0.1 to −0.12 in *ereore*. In *ere*, exclusion of *ry* leads to similar $d_{WS,SW}$ for high- and low-bias regions (0.24 and 0.20, respectively). The nonpartitioned data gave weak support for a consistent decline of codon bias in the *ereore* lineage; the partitioned analysis suggests a larger GC decline in high-bias than in low-bias regions, but the consistency of the decline in high-bias genes remains only weakly supported (WSR $P = 0.011$).

Intron base composition evolution: Summary data for base composition evolution among 60 introns in the *D. melanogaster* subgroup are given in Table 6 (data for individual introns are given in supplemental Table 7 at http://www.genetics.org/supplemental/). The *sim* lineage shows a close fit to equilibrium base composition evolution in these introns, consistent with the findings of Kern and Begun (2005). The most notable departure from equilibrium intron GC content is a nearly twofold excess of WS changes ($d_{WS,SW} = 0.28$, WSR $P = 0.001$) in *ere*. The *ore* lineage shows a consistent increase in intron GC ($d_{WS,SW} = 0.24$, $P = 0.0047$) in contrast to negative $d_{WS,SW}$ at silent sites. A weaker signal for increased intron GC was observed in the *tei* lineage for the 42 introns of the 19 genes employed in the coding-region analysis ($P = 0.02$). However, data from the larger set of 60 introns did not support a consistent departure from

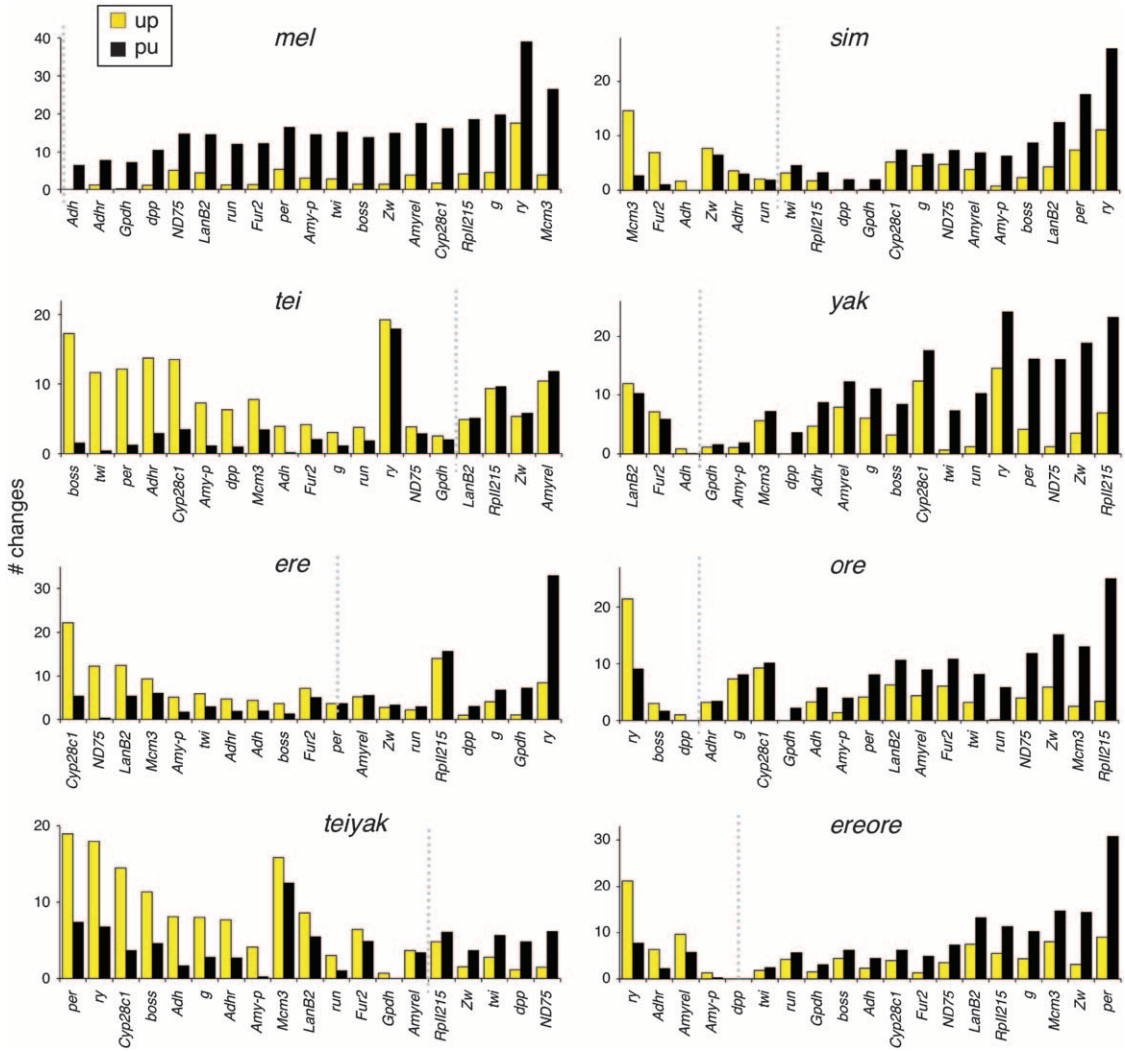

FIGURE 2.—Lineage-specific silent evolution in the *D. melanogaster* subgroup. For each lineage, counts of unpreferred and preferred silent changes are shown for each gene. Data are plotted in the histograms in decreasing order of (up − pu) for each lineage. The transition from positive to negative values is marked with a dotted line on each graph. Abbreviations for lineages are given in Table 2.

equilibrium ($P = 0.3$). More data will be required to confirm possible increases and decreases in intron GC content in the *teiyak* and *yak* lineages, respectively.

AP-50, *Osbp*, and *pav* have undergone large declines of GC content at silent sites in all eight *melanogaster* subgroup lineages, but the introns of these genes show only weak indications of base composition changes (35.2 WS and 43.4 SW changes for data pooled across six introns). Biases toward AT are significantly larger in coding regions (98.35 WS and 297.4 SW changes) than in the introns of these genes, suggesting a reduction in selection intensity at silent sites at these loci prior to the radiation of the *melanogaster* subgroup.

Intron *vs.* coding-region base composition changes: We compared $d_{WS,SW}$ in MCU_H, MCU_L, and intron regions to test the cause(s) of changes in codon usage within the *D. melanogaster* subgroup. In the *tei* lineage, MCU_H regions show a strong increase in GC, MCU_L regions show a smaller increase, and introns show little evidence for a change in base composition (in the set of 60 introns). These trends are consistent with an increase in selection intensity at silent sites in this lineage; SW:WS ratios are heterogeneous among the three classes (2×3 test of independence, $G = 18.8$, $P < 0.001$). In contrast, *yak*, the sibling lineage to *tei*, shows overall $d_{WS,SW}$ patterns consistent with a decline in selection intensity at silent sites; high bias regions have declined in GC to a greater extent than low bias regions and introns ($G = 11.0$, $P = 0.004$). The ancestral *ereore* lineage shows similar patterns of $d_{WS,SW}$ ($G = 17.2$, $P = 0.0002$) and the heterogeneity remains after removal of *ry* data. The *ore* lineage also shows a strong decline of MCU in high-bias regions and a smaller decline in low-bias regions, but introns show a significant *increase* in GC; the statistical evidence for heterogeneity among these classes is strong ($G = 32.5$, $P < 10^{-5}$) and removal of *ry* data has little effect on the pattern. The *mel* lineage shows the strongest decline of GC content at silent sites of the

TABLE 6

Lineage-specific base composition changes in the *D. melanogaster* subgroup

Lineage	Class	Total	WS	SW	$d_{WS,SW}$	*G*	Gene dir	WSR *P*
mel	MCU_H	251.0	34.2	189.5	−0.69	118.4***	0/17	0.00001***
	MCU_L	211.9	40.0	130.7	−0.53	50.7***	0/13	0.0003***
	Rep	67.6	14.0	29.1	−0.35	5.3*	5/9	0.27
	Int	145.4	50.6	64.5	−0.12	1.7	16/23	0.77
sim	MCU_H	151.4	48.5	75.0	−0.21	5.7*	3/12	0.044*
	MCU_L	141.7	51.7	63.2	−0.10	1.2	4/8	0.28
	Rep	45.1	12.8	18.7	−0.19	1.1	4/8	0.66
	Int	152.9	56.3	59.3	−0.03	0.1	19/20	0.71
tei	MCU_H	149.5	98.4	28.7	0.55	40.3***	14/2	0.0002***
	MCU_L	157.1	79.8	47.6	0.25	8.2**	9/4	0.094
	Rep	57.7	30.1	12.0	0.43	8.0**	9/1	0.004**
	Int	201.6	77.9	70.4	0.05	0.4	24/19	0.31
yak	MCU_H	213.7	44.3	133.1	−0.50	46.4***	3/14	0.0006**
	MCU_L	161.7	51.1	85.7	−0.25	8.8**	2/10	0.065
	Rep	42.1	18.9	9.7	0.32	2.9	8/3	0.034*
	Int	209.3	64.8	92.1	−0.17	4.7*	16/32	0.078
teiyak	MCU_H	171.0	79.0	49.5	0.23	6.8*	11/6	0.25
	MCU_L	141.7	83.1	39.1	0.36	16.2***	8/3	0.027*
	Rep	61.5	21.4	17.9	0.09	0.3	8/6	0.37
	Int	147.4	66.0	49.9	0.14	2.2	27/12	0.0079*
ere	MCU_H	163.1	85.4	52.9	0.24	7.7*	10/7	0.17
	MCU_L	162.0	62.5	64.2	−0.01	0.0	8/3	0.24
	Rep	87.3	36.1	18.7	0.32	5.6*	7/7	0.56
	Int	182.3	82.0	46.0	0.28	10.2**	31/13	0.001**
ore	MCU_H	164.5	39.5	104.0	−0.45	29.9***	2/14	0.0004***
	MCU_L	132.7	54.9	58.0	−0.03	0.1	5/6	1.0
	Rep	107.6	31.5	42.8	−0.15	1.7	6/10	0.20
	Int	158.7	71.4	43.5	0.24	6.8*	26/10	0.0047**
ereore	MCU_H	209.5	55.1	118.0	−0.36	23.3***	4/12	0.011*
	MCU_L	115.5	54.5	44.6	0.10	1.0	3/6	0.75
	Rep	114.9	35.8	39.9	−0.05	0.2	5/9	0.59
	Int	203.9	71.1	72.2	−0.01	0.0	26/17	0.21

The total counts of SW and WS changes are shown for high (MCU_H) and low (MCU_L) codon bias regions (the definitions of the regions are given in the text). Total counts are also given for SW and WS replacement (Rep) and intron (Int) changes. The counts are summed across 19 genes and 60 introns for each lineage. $d_{WS,SW}$ values and *G*-test statistics for 2×2 tests of goodness of fit to expectations of equal numbers of SW and WS changes are shown. The numbers of genes (or introns) that show positive/negative $d_{WS,SW}$ are given and probabilities of Wilcoxon signed-rank tests (WSR *P*) are shown. $^*P < 0.05$, $^{**}P < 0.005$, $^{***}P < 0.0005$.

eight lineages examined. Introns in this lineage do not show a strong departure from equilibrium GC content and $d_{WS,SW}$ is heterogeneous among the three classes ($G = 32.5$, $P < 10^{-5}$), suggesting a decrease in selection intensity. However, similar $d_{WS,SW}$ in high- and low-bias regions are not consistent with a uniform decrease in $N_e s$ across genes. Finally, there is little evidence in these data for heterogeneity in $d_{WS,SW}$ among high-bias, low-bias, and intron regions in the *sim*, *teiyak*, and *ere* lineages. Interestingly, *ere* shows similar positive $d_{WS,SW}$ values in MCU_H, MCU_L (excluding *ry*), intron, and replacement sites, suggesting that all regions may be affected by a change in mutational bias or biased gene conversion (however, only the change in introns is statistically significant). The results given above were not strongly dependent on MCU cutoffs or window sizes; finding were similar in analyses of 80 codon windows and a cutoff of MCU = 0.75.

Intron size evolution: Several cases of size–class variation within the *D. melanogaster* subgroup were observed among the 60 introns examined (see supplemental Table 9 at http://www.genetics.org/supplemental/).

Silent, intron, and protein divergence in the *D. melanogaster* subgroup: The total counts of silent, intron, and replacement nucleotide changes (summed across classes) in the eight lineages examined within the *D. melanogaster* subgroup are given in Table 6. The proportions of changes assigned to each lineage are shown in Figure 4A. The histograms allow comparisons of relative distances among these lineages but are not scaled to the numbers of "sites" and are not corrected for multiple hits.

Intron distances were compared between first introns [which contain an excess of regulatory elements in both Drosophila and mammals (Marais *et al.* 2005 and references therein)] and non-first introns. Distances (summed over lineages) were 40% greater (498.5

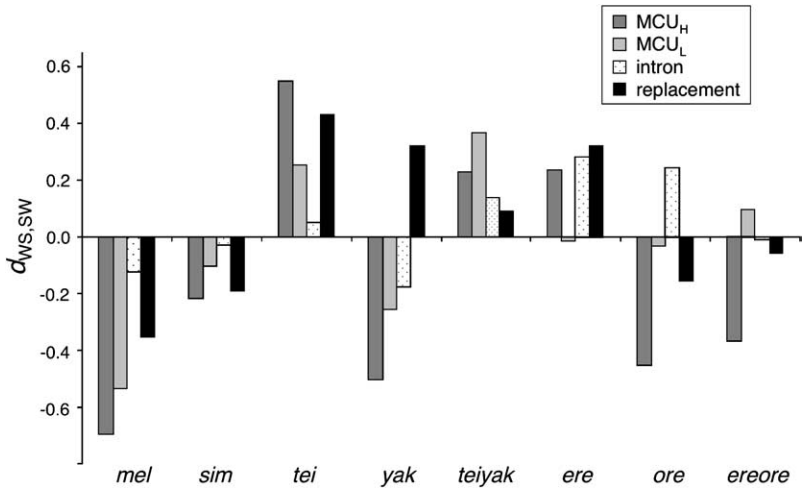

FIGURE 3.—Nonstationary base composition in the *D. melanogaster* subgroup. Differences in the percentage of SW and WS changes, $d_{WS,SW}$, are shown for MCU_H, MCU_L, introns, and replacement changes for lineages in the *D. melanogaster* subgroup. Data and statistical analyses are given in Table 4.

changes at 1961 sites) in the 45 non-first introns than in the 15 first introns (902.8 changes at 2413 sites) ($G = 37.7$, $P < 10^{-5}$), consistent with greater constraint in first introns. However, relative distances among lineages in the *D. melanogaster* subgroup were not affected by inclusion of first introns and data for all 60 introns were employed in the analyses below.

Comparisons among distances for MCU_H, MCU_L, and intron changes demonstrate the difficulty of estimating "neutral divergence" among these species; the distances are clearly heterogeneous among classes (3×8 *G*-test, $G = 65.2$, $P < 10^{-5}$). The most notable differences are fast evolution of MCU_H regions in *ereore* (largely caused by high divergence at *per*) and a large difference in the relative distances for silent sites and introns in *mel* (intron distances are similar in *mel* and *sim*, but silent distances are >50% higher in *mel*). MCP theory predicts accelerated silent evolution following parameter changes (TAKANO-SHIMIZU 1999; supplemental Figure 1 at http://www.genetics.org/supplemental/).

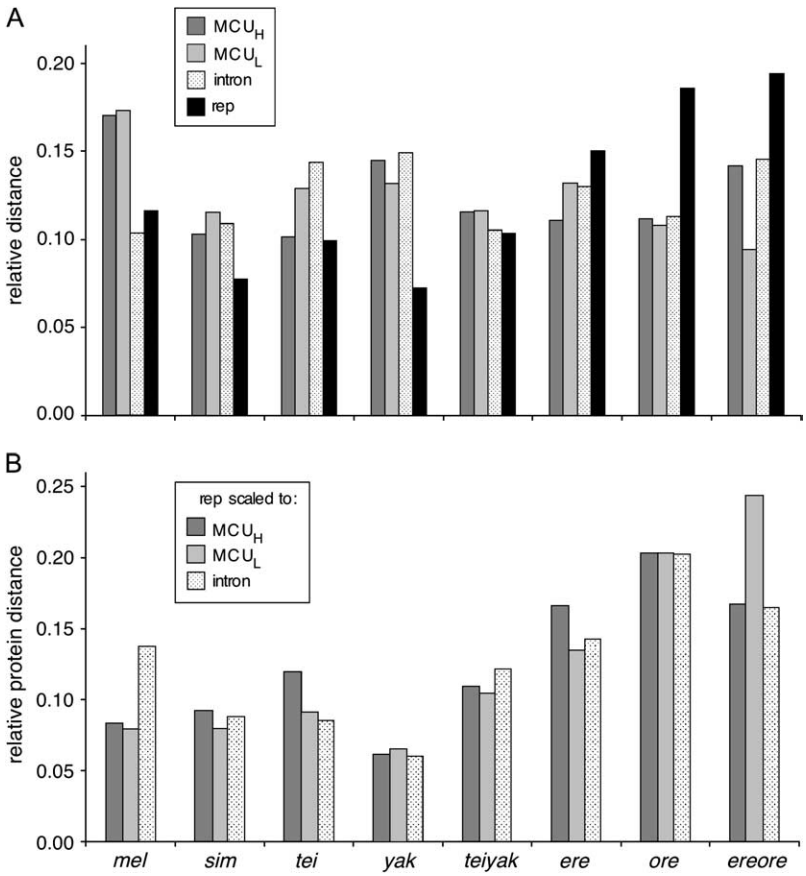

FIGURE 4.—Molecular distances among lineages in the *D. melanogaster* subgroup. (A) Relative distances for silent changes within MCU_H and MCU_L regions, changes in introns, and replacement changes are shown for lineages within the *D. melanogaster* subgroup. For each class, the number of changes within each lineage is divided by the total for eight lineages. (B) Scaled protein distance among lineages within the *D. melanogaster* subgroup. Relative protein distances are scaled by silent distances in MCU_H and MCU_L regions and by intron distances.

Faster silent evolution in *mel* (declining codon bias) than in *sim* (no rejection of equilibrium) is consistent with this prediction. However, overall silent distances appear to be similar between *ere* (no rejection of equilibrium) and *ore* (MCU reduction). This similarity appears to reflect large contributions from fast-evolving outlier genes (especially *ry*) in the *ere* lineage (Figure 2). It is unclear which, if any, of these measures can be employed to correct for neutral distance (the product of *per* generation mutation rate and the number of generations on a lineage) given ambiguities in intron alignments and evidence for both selection and departures from equilibrium at silent sites.

Figure 4A shows the relative numbers of inferred replacement changes for different lineages and Figure 4B plots relative distances for protein evolution scaled to silent distances in high- and low-bias regions as well as intron distance. Several features of protein evolution appear to be consistent among these scalings. The most notable patterns are relatively short protein distances in *yak* and greater protein divergence in the *ere–ore* clade (especially in *ore*). All scalings support a roughly threefold difference in protein divergence between *yak* and *ore*. This difference is consistent among genes; of 16 genes that show different protein distances in *ore* and *yak*, 14 show higher replacement counts in *ore* (WSR $P = 0.0004$; unscaled comparisons are conservative because silent and intron distances are greater in *yak*). Differences among scalings affect comparisons of protein evolution between *mel* and *sim*; intron calibrations suggest faster protein evolution in *mel* than in *sim*, but scaling to silent sites shows similar rates (in either case, differences in rates of protein evolution are not consistent among genes). There is no evidence for associations between rates of protein evolution and departures from equilibrium codon bias. The fastest (*ore*) and slowest (*yak*) protein evolutions occur in lineages that show declines in MCU consistent with reductions in $N_e s$ at silent sites. Intermediate levels of protein divergence occurred in lineages showing close to equilibrium (*sim*) or increasing MCU (*tei*).

Considerable gene-specific rate heterogeneity in both silent and protein evolution in the *D. melanogaster* subgroup is shown in supplemental Table 10 at http://www.genetics.org/supplemental/.

Classifications of amino acid changes: Replacement changes were classified according to their effect on nucleotide (GC) content (Table 6; supplemental Table 8 at http://www.genetics.org/supplemental/). For data summed across lineages, there are roughly equal numbers of WS and SW replacement changes. However, the ratios of the two classes are variable among lineages (2×8 *G*-test, $G = 24.5$, $P = 0.0005$); *tei* shows an almost threefold excess of WS replacement changes and *mel* shows a twofold excess in the opposite direction. The numbers of replacement changes in these two categories on individual lineages are quite small (from 28.6 in *yak* to 74.3 in *ore*), and only *tei* shows a consistent pattern among genes; nine genes deviate toward excess WS replacement changes and 1 deviates in the opposite direction (WSR $P = 0.004$). Trends toward SW replacement changes in *mel* and WS changes in *yak* are stronger among "fixed" differences on these lineages (see below).

The ratios of conservative and nonconservative amino acid changes (using Grantham 1974 and Miyata *et al.* 1979 chemical distance matrices) were not heterogeneous among the eight lineages (data not shown).

Polymorphic and fixed changes: The analyses above employed assignments of changes to eight lineages within the *D. melanogaster* subgroup using a single allele from each species. Changes assigned to the "lineages" leading to the six extant nodes combine those that have been fixed within a species and those that segregate within populations. Under neutral evolution and a constant mutation rate, expected patterns and rates of evolution are identical throughout the gene tree, but natural selection can dramatically alter pre- and post-most recent common ancestor (MRCA) patterns of molecular evolution (Kimura 1983). Because slightly deleterious unpreferred changes segregate at appreciable frequencies within populations but have reduced fixation probabilities (relative to neutral and advantageous changes) (Akashi 1995), analysis of single alleles among closely related species can lead to false inference of reductions of codon bias (Bartolomé *et al.* 2005).

Inferences of polymorphic (post-MRCA) and fixed (pre-MRCA) mutations were conducted for a subset of genes (8, 8, and 18, respectively) in the *mel*, *sim*, and *yak* lineages. Under major codon preference, the expected pu:up ratio is greater for polymorphic mutations than for fixed differences (Akashi 1995) and, at equilibrium, $d_{\mathrm{up,pu}}$ should be zero among fixations. The *mel* lineage showed the largest negative $d_{\mathrm{up,pu}}$ in the single-allele analysis and pu/up ratios are similar (80.7/13.7 and 61.4/9.4, respectively) on pre- and post-MRCA parts of the lineage (supplemental Table 11 at http://www.genetics.org/supplemental/). Such a pattern is consistent with a complete relaxation of selection at silent sites.

In the *sim* lineage, pu:up ratios differ considerably for polymorphic and fixed changes (89.0/34.9 and 21.5/19.4, respectively). There is a greater than twofold excess of pu over up silent polymorphisms, but approximately equal pu/up ratios among fixed differences. This is consistent with previous findings (Akashi 1995, 1999) and supports the maintenance of near-equilibrium codon bias under weak selection in *sim*. Examination of a larger number of genes will be required to determine if this is a general pattern in *sim* (small declines in MCU have been noted by McVean and Vieira 1999 and Begun 2001). The strong dependence of *sim* lineage ancestral inference on evolutionary patterns on the ancestral *melsim* lineage (our unpublished data) will need to be taken into account, however.

In the *yak* lineage, the analysis incorporates a greater number of genes, but is limited by a sample size of only two alleles per gene. A larger sample of alleles may change the assignments of a considerable fraction of fixed mutations to the polymorphic class. At silent sites, pu:up ratios are higher for polymorphic than for fixed mutations (102.6/35.1 and 140.2/73.6, respectively), but the difference is not statistically significant and pu:up ratios remain high for fixed differences. Additional polymorphism data will be necessary to test for selection at silent sites in this lineage and to determine whether negative $d_{up,pu}$ will diminish on the pre-MRCA branch.

DISCUSSION

Evolutionary forces appear to have fluctuated repeatedly on the timescale of molecular evolution among closely related species in the *D. melanogaster* subgroup. Both the composition and evolutionary rates of DNA and proteins show departures from steady state. Some of these patterns are consistent with lineage-specific changes in selection intensity, but other patterns are not readily explained by simple models of genomewide, lineage-specific parameter changes.

Departures from equilibrium silent evolution: Population genetic analyses in *D. simulans*, *pseudoobscura*, *subobscura*, *miranda*, and *americana* support the maintenance of codon bias through interactions among weak selection, mutation pressure, and genetic drift. Differences in the ratios of preferred and unpreferred changes within and between species and in their site frequency spectra within populations (Akashi 1995, 1999; Akashi and Schaeffer 1997; Kliman 1999; Llopart and Aguadé 2000; Begun 2001; Maside *et al.* 2004; Bartolomé *et al.* 2005; Comeron and Guthrie 2005) are consistent with a balance among weak evolutionary forces. Such a balance is expected to be highly sensitive to parameter fluctuations (Akashi 1996; Takano-Shimizu 1999).

Of the eight lineages examined from the *D. melanogaster* subgroup, three appear to have undergone declines of codon bias (*D. melanogaster*, *yakuba*, and *orena*) and one (*D. teissieri*) shows evidence for an increase in MCU. The directions of these departures from equilibrium are consistent across genes and synonymous families and the magnitudes are quite large. *mel* shows by far the strongest departure, an almost fivefold greater number of pu changes than up changes. *yak* and *ore* show roughly twofold excesses of pu changes and *tei* shows a greater than twofold excess of up changes. A decline in codon bias in the ancestral *ereore* lineage and an increase in the *teiyak* were weakly supported. The *sim* and *ere* lineages do not show consistent departures from equilibrium codon bias among the 19 genes examined.

Because the approach to equilibrium base composition can be quite slow (on the order of $1/u$ generations where u is the mutation rate), detection of departures from equilibrium in multiple short lineages does not necessarily indicate frequent parameter fluctuations. For example, for the three outlier genes, a single parameter change could be responsible for codon bias declines in all eight lineages. In our analysis, each of the three pairs of extant sibling lineages shows different patterns of codon usage evolution, suggesting at least three parameter changes among the six extant lineages. In addition, if we assume shared increases of codon bias in *teiyak* and *tei* and shared MCU reductions in *ereore* and *ore*, then an additional parameter change is required in at least one of the two ancestral lineages. Thus, a minimum of four parameter changes in the eight lineages appear to be required to explain the observations.

Gillespie (1993) and Cutler (2000a,b) have noted the parameter sensitivity of models that invoke variability of forces in molecular evolution. For such models to account for heterogeneity among lineages, parameter fluctuations must occur on a similar timescale as molecular evolution (roughly $1/u$ generations). A curious tuning of parameter fluctuations appears to be required to explain molecular evolution among *D. melanogaster* and its close relatives. However, the patterns may be less surprising if parameter changes occur on multiple timescales; rapid fluctuations have little effect on patterns of divergence, but slower changes result in lineage-specific molecular evolution.

Nonequilibrium silent evolution has several implications for studies of molecular evolution. Rates of silent evolution in the approach to equilibrium are expected to differ both from neutral rates and from equilibrium rates under selection. Relationships between codon bias and silent divergence can be negative, positive, or flat depending on the status of silent sites with respect to steady state (Takano-Shimizu 1999; Piganeau *et al.* 2002; supplemental Figure 1 at http://www.genetics.org/supplemental/). Departures from equilibrium may have contributed to the lack of negative relationships between silent divergence and codon bias in Drosophila (Dunn *et al.* 2001; see also Bierne and Eyre-Walker 2003). Attempts to detect selection at silent sites should first test for equilibrium codon bias evolution among the lineages under study. However, similar levels of codon bias between extant nodes may not indicate steady-state evolution in the lineages connecting them. In pairwise comparisons, *D. simulans* and *D. yakuba* genes show similar MCU, but our analyses indicate overall increases of codon bias in *teiyak* and decreases in *yak* (changes were not assigned in the two interior ancestral lineages).

Synonymous changes are often employed in molecular evolutionary studies as a form of internal calibration to infer mechanisms of protein or noncoding DNA evolution (reviewed in Li 1997). Patterns of synonymous variation are assumed to reflect a combination of local mutation processes (Hughes and Nei 1988) and the recent evolutionary history of a given genetic region (Sawyer *et al.* 1987; McDonald and Kreitman 1991).

However, evidence for weak selection and observations of strong and frequent departures from equilibrium suggest that greater consideration of the processes underlying silent DNA evolution will be critical for advancing our understanding of both protein and regulatory region evolution. Frequent departures from equilibrium on the short lineages examined here demonstrate that assumptions of parameter constancy *within* as well as among lineages should be treated with caution. Such changes can have a strong impact on patterns of polymorphism and divergence (Eyre-Walker 1997, 2002; Kern and Begun 2005).

This study was designed to detect lineage-specific departures from stationary codon bias evolution and the limited sample of genes and the generally small gene-specific numbers of changes on a given lineage place a strong limit on our ability to identify region- or gene-specific evolution. Our sample is biased toward genes on the X and is deficient in genes from the third chromosome (because *AP-50*, *pav*, and *Osbp* are outliers, only 3 genes from the third chromosome are included among the 19 genes investigated). At least one gene in our study, *ry*, shows strong gene-specific patterns of codon bias evolution in several lineages. Singh *et al.* (2005) have shown elevated codon bias on Drosophila X chromosomes relative to autosomes and increases in GC content were noted by Takano-Shimizu (1999, 2001) for genes on the tip (telomeric region) of the X chromosome in the *D. teissieri–D. yakuba* and *D. erecta–D. orena* clades. These departures appear to be region specific; genes in the present analysis show a decline of preferred codon usage in lineages showing strong GC increases at the tip of the X chromosome. These patterns will be described in detail in a separate study. Limited data suggest that gene- and region-specific departures from stationarity may also be frequent.

Causes of nonequilibrium codon bias evolution: Under major codon preference, departures from equilibrium could reflect altered fixation probabilities caused by changes in selection intensity ($N_e s$) or changes in the mutational input of preferred and unpreferred changes (u/v). If changes in MCU reflect uniform changes in selection intensity (*i.e.*, all $N_e s$ values scaled by the same factor), then codon bias changes should be greater for codons under stronger selection than for codons under weaker selection or for neutrally evolving regions (Akashi 1996). Uniform changes in u/v predict similar codon bias changes ($d_{up,pu}$) for strongly and weakly selected codons as well as neutral regions because fixation rates are linearly dependent on mutation rates. Predictions for changes in biased gene conversion (BGC) depend on the relative contributions of selection and BGC to codon bias and are less clear. However, changes in BGC would presumably affect intron base composition (Galtier *et al.* 2006).

These predictions can be tested by comparing $d_{up,pu}$ among synonymous families or genes or gene regions that experience different intensities of selection. McVean and Vieira (2001) found evidence for heterogeneity in selection strength among synonymous families, but the numbers of within-family changes were too small in our data for comparisons among codon families undergoing similar mutation patterns. We compared silent changes in MCU_H, MCU_L, and putatively neutrally evolving introns to determine the cause(s) of nonequilibrium codon bias evolution. This approach assumes that variation in selection intensity causes both within- and among-gene variation in MCU. Correlations between MCU and expression levels (Duret and Mouchiroud 1999) are consistent with increases in the fitness effects of silent mutations as a function of translation rates and weak correlations between synonymous and intron GC contents support a relatively small contribution of variation in mutation biases to codon bias variation among Drosophila genes (Hey and Kliman 2002). The latter finding also suggests that the role of biased gene conversion may be small. Begun (2001) found lower average frequencies of pu changes in highly biased genes, consistent with higher $N_e s$. In addition, several lines of evidence support within-gene variation of selection intensity at silent sites (Akashi 1994; Kliman and Eyre-Walker 1998; Llopart and Aguadé 1999, 2000; Iida and Akashi 2000; Comeron and Kreitman 2002; Qin *et al.* 2004; Comeron and Guthrie 2005).

Differences in base composition evolution among MCU_H, MCU_L, and introns are consistent with departures from equilibrium codon bias caused by changes in selection intensity in *tei*, *yak*, *ore*, and *ereore*. The cause(s) of the dramatic decline of codon bias in *mel* is less clear; high- and low-bias regions show similar departures from equilibrium. Although intron base composition appears to be consistent with equilibrium GC in the single-allele analysis, Kern and Begun (2005) found striking differences in pre- and post-MRCA intron base composition evolution in *mel*. More data (especially for low-bias regions) and contrasts between polymorphic and fixed changes will be necessary to confirm patterns of codon bias change among MCU partitions and introns. In addition, the notion that MCU variation within genes reflects variation in fitness benefits to major codons requires further investigation. Other selection pressures (such as selection on mRNA secondary structure) in low-bias regions could complicate the interpretation of tests of the translational selection model.

Departures from equilibrium intron base composition were detected in the *ere* and *ore* lineages. We interpret such changes as indications of changes in mutation processes or biased gene conversion, but closer examination of the locations of the changes and their effects on mRNA structure, splicing enhancers, and regulatory elements is necessary (Bergman and Kreitman 2001; Halligan *et al.* 2004; Haddrill *et al.* 2005). In addition, intron regions with ambiguous alignments were

excluded from the analysis. This filtering may increase the proportion of constrained sites in the data.

Protein evolution in the *D. melanogaster* subgroup: Both rates of protein evolution and types of amino acid changes are variable among lineages in the *D. melanogaster* subgroup. The *mel* lineage shows an increase in SW changes and *tei* and *yak* show excesses of WS replacement changes. These changes in GC content could reflect selection for nucleotide or amino acid composition. For example, in the *tei* lineage, an excess of HNN → GNN over GNN → HNN (WSR test, $P = 0.008$) accounts for more than half of the excess of WS replacement changes. This pattern is notable because GNN codons are used preferentially in highly expressed genes in a number of prokaryotic (Gutiérrez *et al.* 1996; Akashi and Gojobori 2002) and eukaryotic (Akashi 2003) genomes and may confer fitness advantages related to efficient protein synthesis.

Rates of protein evolution are heterogeneous among lineages in the *D. melanogaster* subgroup but no clear associations were observed with changes in codon bias. Interestingly, among the species with current distributions within Africa, rates of protein evolution appear to be inversely related with the extent of their current biogeographic ranges (according to Lachaise *et al.* 1988; Cobb *et al.* 2000). *D. orena* shows the most restricted distribution [it is known only from a single collection (of 11 individuals) from the West Cameroon mountains] and the fastest rate of protein evolution. *D. erecta*, which is confined to equatorial western Africa and is thought to be a rare species, shows the next fastest rate of protein evolution. *D. teissieri* and *D. yakuba* are widely distributed across the Afrotropical region and show slower rates of protein evolution than *D. erecta* and *D. orena*. Historical distributions of *D. melanogaster* and *D. simulans* are less clear. Partitioning of changes on pre- and post-MRCA lineages will test whether rate differences are consistent with the accumulation of slightly deleterious amino acid substitutions. These patterns raise the possibility that protein divergence, but not codon bias evolution, is associated with long-term population sizes. Among prokaryotes, major codon preference appears to be an adaptation for fast growth (Dong *et al.* 1996; Rocha 2004). Differences in the ecology of Drosophila species may contribute to species-specific fitness benefits to major codon usage.

Genome sequences for *D. melanogaster*, *D. simulans*, *D. yakuba*, and *D. erecta* will allow large-scale analysis of a limited number of lineages within this subgroup. With such data, changes can be assigned to the *mel* and *sim* lineages and to two combined lineages, *ereore* + *ere* and *teiyak* + *yak*. Pooling data for parent and child lineages that differ in codon bias evolution may mask lineage-specific patterns. For the data examined here, overall patterns of silent changes in the combined *ereore* + *ere* and *teiyak* + *yak* lineages are consistent with equilibrium codon bias evolution. Including sequences from *D. teissieri* and *D. orena* doubles the number of lineages on which changes can be assigned (from four to eight) and reveals differences between *yak* and *ere* and their parental lineages. The two additional terminal lineages show compelling departures from steady state (an increase in codon bias in *tei* and accelerated protein evolution in *ore*).

Our conclusions here are limited by the analysis of single alleles for most genes and lineages. Polymorphism and divergence data from multiple lineages will allow discrimination between departures from equilibrium (among substitutions) and effects of pooling pre- and post-MRCA nonneutral changes. Such data will also help to distinguish between selection- and mutation-driven changes in base composition. The combination of comparisons of evolutionary patterns for classes of changes among closely related lineages and between fixed and polymorphic mutations may shed light on both the relative contributions of evolutionary forces and the nature and frequency of parameter fluctuations in molecular divergence.

We are grateful to Ziheng Yang for advice on ancestral state reconstruction using BASEML as well as comments that helped to improve this article. We thank Claude dePamphilis and members of his lab for use of their automated sequencer. This work also benefited from valuable suggestions from two anonymous reviewers. This research was supported by National Science Foundation grant DEB-0521964 to H.A.

LITERATURE CITED

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

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., 1999 Inferring the fitness effects of DNA mutations from polymorphism and divergence data: statistical power to detect directional selection under stationarity and free recombination. Genetics **151:** 221–238.

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

Akashi, H., 2003 Translational selection and yeast proteome evolution. Genetics **164:** 1291–1303.

Akashi, H., and T. Gojobori, 2002 Metabolic efficiency and amino acid composition in the proteomes of *Escherichia coli* and *Bacillus subtilis*. Proc. Natl. Acad. Sci. USA **99:** 3695–3700.

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

Altschul, S. F., W. Gish, W. Miller, E. W. Myers and D. J. Lipman, 1990 Basic local alignment search tool. J. Mol. Biol. **215:** 403–410.

Andersson, S. G., and C. G. Kurland, 1990 Codon preferences in free-living microorganisms. Microbiol. Rev. **54:** 198–210.

Ashburner, M., K. G. Golic and R. S. Hawley, 2005 *Drosophila: A Laboratory Handbook*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Bachtrog, D., 2003 Protein evolution and codon usage bias on the neo-sex chromosomes of *Drosophila miranda*. Genetics **165:** 1221–1232.

Bartolomé, C., X. Maside, S. Yi, A. L. Grant and B. Charlesworth, 2005 Patterns of selection on synonymous and nonsynonymous variants in *Drosophila miranda*. Genetics **169:** 1495–1507.

Begun, D. J., 2001 The frequency distribution of nucleotide variation in *Drosophila simulans*. Mol. Biol. Evol. **18:** 1343–1352.

Begun, D. J., 2002 Protein variation in *Drosophila simulans*, and comparison of genes from centromeric versus noncentromeric regions of chromosome 3. Mol. Biol. Evol. **19:** 201–203.

Bergman, C. M., and M. Kreitman, 2001 Analysis of conserved noncoding DNA in Drosophila reveals similar constraints in intergenic and intronic sequences. Genome Res. **11:** 1335–1345.

Bierne, N., and A. Eyre-Walker, 2003 The problem of counting sites in the estimation of the synonymous and nonsynonymous substitution rates: implications for the correlation between the synonymous substitution rate and codon usage bias. Genetics **165:** 1587–1597.

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

Cobb, M., M. Huet, D. Lachaise and M. Veuille, 2000 Fragmented forests, evolving flies: molecular variation in African populations of *Drosophila teissieri*. Mol. Ecol. **9:** 1591–1597.

Collins, T. M., P. H. Wimberger and G. J. P. Naylor, 1994 Compositional bias, character-state bias, and character-state reconstruction using parsimony. Syst. Biol. **43:** 482–496.

Comeron, J. M., and T. B. Guthrie, 2005 Intragenic Hill-Robertson interference influences selection intensity on synonymous mutations in Drosophila. Mol. Biol. Evol. **22:** 2519–2530.

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

Cutler, D. J., 2000a Understanding the overdispersed molecular clock. Genetics **154:** 1403–1417.

Cutler, D. J., 2000b The index of dispersion of molecular evolution: slow fluctuations. Theor. Popul. Biol. **57:** 177–186.

Dong, H., L. Nilsson and C. G. Kurland, 1996 Co-variation of tRNA abundance and codon usage in *Escherichia coli* at different growth rates. J. Mol. Biol. **260:** 649–663.

Drysdale, R. A., M. A. Crosby and the FlyBase Consortium, 2005 FlyBase: genes and gene models. Nucleic Acids Res. **33:** D390–395.

Dumont, V. B., J. C. Fay, P. P. Calabrese and C. F. Aquadro, 2004 DNA variability and divergence at the notch locus in *Drosophila melanogaster* and *D. simulans*: a case of accelerated synonymous site divergence. Genetics **167:** 171–185.

Dunn, K. A., J. P. Bielawski and Z. Yang, 2001 Substitution rates in Drosophila nuclear genes: implications for translational selection. Genetics **157:** 295–305.

Duret, L., 2002 Evolution of synonymous codon usage in metazoans. Curr. Opin. Genet. Dev. **12:** 640–649.

Duret, L., and D. Mouchiroud, 1999 Expression pattern and, surprisingly, gene length shape codon usage in Caenorhabditis, Drosophila, and Arabidopsis. Proc. Natl. Acad. Sci. USA **96:** 4482–4487.

Duret, L., M. Semon, G. Piganeau, D. Mouchiroud and N. Galtier, 2002 Vanishing GC-rich isochores in mammalian genomes. Genetics **162:** 1837–1847.

Eanes, W. F., M. Kirchner and J. Yoon, 1993 Evidence for adaptive evolution of the *G6pd* gene in the *Drosophila melanogaster* and *Drosophila simulans* lineages. Proc. Natl. Acad. Sci. USA **90:** 7475–7479.

Eyre-Walker, A., 1997 Differentiating between selection and mutation bias. Genetics **147:** 1983–1987.

Eyre-Walker, A., 2002 Changing effective population size and the McDonald–Kreitman test. Genetics **162:** 2017–2024.

Fitch, W. M., 1971 Rate of change of concomitantly variable codons. J. Mol. Evol. **1:** 84–96.

Fitch, W. M., and E. Markowitz, 1970 An improved method for determining codon variability in a gene and its application to the rate of fixation of mutations in evolution. Biochem. Genet. **4:** 579–593.

Galtier, N., and P. Boursot, 2000 A new method for locating changes in a tree reveals distinct nucleotide polymorphism vs. divergence patterns in mouse mitochondrial control region. J. Mol. Evol. **50:** 224–231.

Galtier, N., E. Bazin and N. Bierne, 2006 GC-biased segregation of noncoding polymorphisms in Drosophila. Genetics **172:** 221–228.

Gillespie, J. H., 1991 *The Causes of Molecular Evolution*. Oxford University Press, New York.

Gillespie, J. H., 1993 Substitution processes in molecular evolution. I. Uniform and clustered substitutions in a haploid model. Genetics **134:** 971–981.

Gillespie, J. H., 1994 Substitution processes in molecular evolution. II. Exchangeable models from population genetics. Evolution **48:** 1101–1113.

Grantham, R., 1974 Amino acid difference formula to help explain protein evolution. Science **185:** 862–864.

Gutiérrez, G., L. Marquez and A. Marin, 1996 Preference for guanosine at first codon position in highly expressed *Escherichia coli* genes. A relationship with translational efficiency. Nucleic Acids Res. **24:** 2525–2527.

Haddrill, P. R., B. Charlesworth, D. L. Halligan and P. Andolfatto, 2005 Patterns of intron sequence evolution in Drosophila are dependent upon length and GC content. Genome Biol. **6:** R67.

Halligan, D. L., A. Eyre-Walker, P. Andolfatto and P. D. Keightley, 2004 Patterns of evolutionary constraints in intronic and intergenic DNA of Drosophila. Genome Res. **14:** 273–279.

Hasegawa, M., H. Kishino and T. Yano, 1985 Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J. Mol. Evol. **22:** 160–174.

Hey, J., and R. M. Kliman, 2002 Interactions between natural selection, recombination and gene density in the genes of Drosophila. Genetics **160:** 595–608.

Hughes, A. L., and M. Nei, 1988 Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. Nature **335:** 167–170.

Iida, K., and H. Akashi, 2000 A test of translational selection at 'silent' sites in the human genome: base composition comparisons in alternatively spliced genes. Gene **261:** 93–105.

Ikemura, T., 1985 Codon usage and tRNA content in unicellular and multicellular organisms. Mol. Biol. Evol. **2:** 13–34.

Johnson, K. P., and J. Seger, 2001 Elevated rates of nonsynonymous substitution in island birds. Mol. Biol. Evol. **18:** 874–881.

Jukes, T. H., and C. R. Cantor, 1969 Evolution of protein molecules, pp. 21–132 in *Mammalian Protein Metabolism III*, edited by H. N. Munro. Academic Press, New York.

Keightley, P. D., and T. Johnson, 2004 MCALIGN: stochastic alignment of noncoding DNA sequences based on an evolutionary model of sequence evolution. Genome Res. **14:** 442–450.

Kern, A. D., and D. J. Begun, 2005 Patterns of polymorphism and divergence from non-coding sequences of *D. melanogaster* and *D. simulans*: evidence for non-equilibrium processes. Mol. Biol. Evol. **22:** 51–62.

Kimura, M., 1983 *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge, UK/New York.

Kliman, R. M., 1999 Recent selection on synonymous codon usage in Drosophila. J. Mol. Evol. **49:** 343–351.

Kliman, R. M., and A. Eyre-Walker, 1998 Patterns of base composition within the genes of *Drosophila melanogaster*. J. Mol. Evol. **46:** 534–541.

Ko, W. Y., R. M. David and H. Akashi, 2003 Molecular phylogeny of the *Drosophila melanogaster* species subgroup. J. Mol. Evol. **57:** 562–573.

Kopp, A., and J. R. True, 2002 Phylogeny of the Oriental Drosophila melanogaster species group: a multilocus reconstruction. Syst. Biol. **51:** 786–805.

Lachaise, D., M. L. Cariou, J. R. David, F. Lemeunier, L. Tsacas *et al.*, 1988 Historical biogeography of the *Drosophila melanogaster* species subgroup. Evol. Biol. **22:** 159–225.

Li, W. H., 1987 Models of nearly neutral mutations with particular implications for nonrandom usage of synonymous codons. J. Mol. Evol. **24:** 337–345.

Li, W.-H., 1997 *Molecular Evolution*. Sinauer Associates, Sunderland, MA.

Llopart, A., and M. Aguadé, 1999 Synonymous rates at the *RpII215* gene of Drosophila: variation among species and across the coding region. Genetics **152:** 269–280.

Llopart, A., and M. Aguadé, 2000 Nucleotide polymorphism at the *RpII215* gene in *Drosophila subobscura*. Weak selection on synonymous mutations. Genetics **155:** 1245–1252.

Marais, G., P. Nouvellet, P. D. Keightley and B. Charlesworth, 2005 Intron size and exon evolution in Drosophila. Genetics **170:** 481–485.

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

McDonald, J. H., and M. Kreitman, 1991 Adaptive protein evolution at the *Adh* locus in Drosophila. Nature **351:** 652–654.

McVean, G. A., and J. Vieira, 1999 The evolution of codon preferences in Drosophila: a maximum-likelihood approach to parameter estimation and hypothesis testing. J. Mol. Evol. **49:** 63–75.

McVean, G. A., and J. Vieira, 2001 Inferring parameters of mutation, selection and demography from patterns of synonymous site evolution in Drosophila. Genetics **157:** 245–257.

Messier, W., and C. B. Stewart, 1997 Episodic adaptive evolution of primate lysozymes. Nature **385:** 151–154.

Miyata, T., S. Miyazawa and T. Yasunaga, 1979 Two types of amino acid substitutions in protein evolution. J. Mol. Evol. **12:** 219–236.

Moran, N. A., 1996 Accelerated evolution and Muller's ratchet in endosymbiotic bacteria. Proc. Natl. Acad. Sci. USA **93:** 2873–2878.

Munté, A., M. Aguadé and C. Segarra, 2001 Changes in the recombinational environment affect divergence in the *yellow* gene of Drosophila. Mol. Biol. Evol. **18:** 1045–1056.

Nei, M., and T. Gojobori, 1986 Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. **3:** 418–426.

Ohta, T., 1987 Very slightly deleterious mutations and the molecular clock. J. Mol. Evol. **26:** 1–6.

Ohta, T., 1993 An examination of the generation-time effect on molecular evolution. Proc. Natl. Acad. Sci. USA **90:** 10676–10680.

Ohta, T., 2002 Near-neutrality in evolution of genes and gene regulation. Proc. Natl. Acad. Sci. USA **99:** 16134–16137.

Osawa, S., T. Ohama, F. Yamao, A. Muto, T. H. Jukes *et al.*, 1988 Directional mutation pressure and transfer RNA in choice of the third nucleotide of synonymous two-codon sets. Proc. Natl. Acad. Sci. USA **85:** 1124–1128.

Parsch, J., 2003 Selective constraints on intron evolution in Drosophila. Genetics **165:** 1843–1851.

Peréz, J. A., A. Munté, J. Rozas, C. Segarra and M. Aguadé, 2003 Nucleotide polymorphism in the *RpII215* gene region of the insular species *Drosophila guanche*: reduced efficacy of weak selection on synonymous variation. Mol. Biol. Evol. **20:** 1867–1875.

Piganeau, G., D. Mouchiroud, L. Duret and C. Gautier, 2002 Expected relationship between the silent substitution rate and the GC content: implications for the evolution of isochores. J. Mol. Evol. **54:** 129–133.

Powell, J. R., 1997 *Progress and Prospects in Evolutionary Biology: The Drosophila Model*. Oxford University Press, New York.

Powell, J. R., E. Sezzi, E. N. Moriyama, J. M. Gleason and A. Caccone, 2003 Analysis of a shift in codon usage in Drosophila. J. Mol. Evol. **57**(Suppl. 1): S214–S225.

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

Rocha, E. P., 2004 Codon usage bias from tRNA's point of view: redundancy, specialization, and efficient decoding for translation optimization. Genome Res. **14:** 2279–2286.

Rodríguez-Trelles, F., R. Tarrío and F. J. Ayala, 2000 Fluctuating mutation bias and the evolution of base composition in Drosophila. J. Mol. Evol. **50:** 1–10.

Sawyer, S. A., D. E. Dykhuizen and D. L. Hartl, 1987 Confidence interval for the number of selectively neutral amino acid polymorphisms. Proc. Natl. Acad. Sci. USA **84:** 6225–6228.

Sharp, P. M., M. Averof, A. T. Lloyd, G. Matassi and J. F. Peden, 1995 DNA sequence evolution: the sounds of silence. Philos. Trans. R. Soc. Lond. B Biol. Sci. **349:** 241–247.

Shibata, H., and T. Yamazaki, 1995 Molecular evolution of the duplicated *Amy* locus in the *Drosophila melanogaster* species subgroup: concerted evolution only in the coding region and an excess of nonsynonymous substitutions in speciation. Genetics **141:** 223–236.

Shields, D. C., 1990 Switches in species-specific codon preferences: the influence of mutation biases. J. Mol. Evol. **31:** 71–80.

Shields, D. C., P. M. Sharp, D. G. Higgins and F. Wright, 1988 "Silent" sites in Drosophila genes are not neutral: evidence of selection among synonymous codons. Mol. Biol. Evol. **5:** 704–716.

Singh, N. D., J. C. Davis and D. A. Petrov, 2005 X-linked genes evolve higher codon bias in Drosophila and Caenorhabditis. Genetics **171:** 145–155.

Snedecor, G. W., and W. G. Cochran, 1989 *Statistical Methods*. Iowa State University Press, Ames, IA.

Sokal, R. R., and F. J. Rohlf, 1995 *Biometry: The Principles and Practice of Statistics in Biological Research*. W. H. Freeman, New York.

Sueoka, N., 1988 Directional mutation pressure and neutral molecular evolution. Proc. Natl. Acad. Sci. USA **85:** 2653–2657.

Takahata, N., 1987 On the overdispersed molecular clock. Genetics **116:** 169–179.

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.

Thompson, J. D., D. G. Higgins and T. J. Gibson, 1994 CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22:** 4673–4680.

Webster, M. T., N. G. Smith and H. Ellegren, 2003 Compositional evolution of noncoding DNA in the human and chimpanzee genomes. Mol. Biol. Evol. **20:** 278–286.

Woolfit, M., and L. Bromham, 2003 Increased rates of sequence evolution in endosymbiotic bacteria and fungi with small effective population sizes. Mol. Biol. Evol. **20:** 1545–1555.

Wootton, J. C., and S. Federhen, 1996 Analysis of compositionally biased regions in sequence databases. Methods Enzymol. **266:** 554–571.

Yang, Z., 1997 PAML: a program package for phylogenetic analysis by maximum likelihood. Comput. Appl. Biosci. **13:** 555–556.

Yang, Z., S. Kumar and M. Nei, 1995 A new method of inference of ancestral nucleotide and amino acid sequences. Genetics **141:** 1641–1650.

Zhang, J., Y. P. Zhang and H. F. Rosenberg, 2002a Adaptive evolution of a duplicated pancreatic ribonuclease gene in a leaf-eating monkey. Nat. Genet. **30:** 411–415.

Zhang, Z., N. Inomata, T. Ohba, M. L. Cariou and T. Yamazaki, 2002b Codon bias differentiates between the duplicated *amylase* loci following gene duplication in Drosophila. Genetics **161:** 1187–1196.

Communicating editor: M. Aguadé