DNA Variability and Divergence at the \textit{Notch} Locus in \textit{Drosophila melanogaster} and \textit{D. simulans}: A Case of Accelerated Synonymous Site Divergence

Vanessa Bauer DuMont,* Justin C. Fay,*1 Peter P. Calabrese†,2 and Charles F. Aquadro*,3

*Department of Molecular Biology and Genetics and †Department of Applied Mathematics, Cornell University, Ithaca, New York 14853

Manuscript received February 4, 2003
Accepted for publication January 14, 2004

ABSTRACT

DNA diversity in two segments of the \textit{Notch} locus was surveyed in four populations of \textit{Drosophila melanogaster} and two of \textit{D. simulans}. In both species we observed evidence of non-steady-state evolution. In \textit{D. simulans} we observed a significant excess of intermediate frequency variants in a non-African population. In \textit{D. melanogaster} we observed a disparity between levels of sequence polymorphism and divergence between one of the \textit{Notch} regions sequenced and other neutral X chromosome loci. The striking feature of the data is the high level of synonymous site divergence at \textit{Notch}, which is the highest reported to date. To more thoroughly investigate the pattern of synonymous site evolution between these species, we developed a method for calibrating preferred, unpreferred, and equal synonymous substitutions by the effective (potential) number of such changes. In \textit{D. simulans}, we find that preferred changes per “site” are evolving significantly faster than unpreferred changes at \textit{Notch}. In contrast we observe a significantly faster per site substitution rate of unpreferred changes in \textit{D. melanogaster} at this locus. These results suggest that positive selection, and not simply relaxation of constraint on codon bias, has contributed to the higher levels of unpreferred divergence along the \textit{D. melanogaster} lineage at \textit{Notch}.

To discern where and how natural selection has shaped levels of genomic diversity, we need estimates of the underlying “neutral” level of variation within and between species. We have been studying sequence polymorphism in regions of high crossing-over in an effort to define the levels of variability in regions presumably free from the effect of linked selection to estimate neutral levels and patterns of variation for \textit{Drosophila melanogaster} and \textit{D. simulans}. In this article we present data for nucleotide variability within and between species at two segments of the \textit{Notch} locus. \textit{Notch} was chosen for two reasons. First, it is located on the X chromosome in a region of relatively high recombination (2.1 \times 10^{-3} recombinants/generation/kb; reported as \textit{R} in Hey and Kliman 2002). Second, a previous course-scale (six-cutter) restriction site survey of a 60-kb region encompassing \textit{Notch} suggested this region of the genome to be evolving neutrally (Schaeffer et al. 1988; Begun and Aquadro 1992).

\textit{Notch} encodes a single-pass, transmembrane protein that serves as a receptor for cell-to-cell communication during metazoan development. \textit{Notch} homologs are present in diverse organisms including sea urchins, fruit flies, and humans. Due to its sequence and functional conservation, \textit{Notch}'s role during development has been extensively studied (e.g., Beatus and Lendahl 1998; Fleming 1998; Artavanis-Tsakonas et al. 1999). As a result, functional domains of the \textit{Notch} protein have been characterized, and many proteins involved in the \textit{Notch} pathway have been identified.

The \textit{Notch} transcript spans 30 kb of genomic DNA. We sequenced two segments (Figure 1), which we call the 5’ region (exons 3 and 4 and neighboring introns) and the 3’ region (the 3’ end of exon 6). These segments are 10 kb apart. Exons 3 and 4 encode several of the epidermal growth factor (EGF)-like repeats of the extracellular domain of \textit{Notch}. EGF repeats are involved in the interaction of \textit{Notch} with its ligands (i.e., Delta and Serrate) and the initiation of the \textit{Notch} signaling pathway (Rebay et al. 1991). The 3’ region of exon 6 represents the beginning of the intracellular domain of \textit{Notch} that mediates the ligand-dependent response of the \textit{Notch} pathway within the cell (through interactions with a number of other proteins).

Below we report that in \textit{D. melanogaster} there are significant differences in the ratio of polymorphism to divergence between \textit{Notch} 3’ and other neutral X-chromosome loci. The trend is similar, though not as dramatic, in \textit{D. simulans}. Of particular note is the high level of synonymous site divergence at \textit{Notch} 3’, a level three times the average observed between these species. Consideration of synonymous substitutions per effective
number of preferred, unpreferred, and equal synonymous preference sites suggests that at Notch 3’ positive selection has played a role in accelerating different types of synonymous site evolution in both of these species.

MATERIALS AND METHODS

Samples: Four population samples of D. melanogaster were surveyed for nucleotide sequence variability: Zimbabwe (Sen-gwa Wildlife Research Institute), United States (California), Ecuador (Atacame), and China (Beijing). Two population samples of D. simulans from the United States (North Carolina) and Zimbabwe (Harare) were also surveyed. Collection data for these populations have been reported previously (BEGUN and AQUADRO 1991, 1994, 1995). For all D. melanogaster populations, extracted X chromosome lines (BEGUN and AQUADRO 1994) were used. For D. simulans we used inbred lines. Sample sizes for D. melanogaster were 12, 15, 11, and 12 chromosomes for Zimbabwe, United States, Ecuador, and China, respectively. For D. simulans sample sizes were 12 and 10 for the United States and Zimbabwe, respectively.

Generation and analysis of sequence polymorphism data at Notch: Cesium chloride gradient-purified genomic DNA was used for the D. melanogaster Zimbabwe and United States samples and for the D. simulans United States sample. For Ecuador and China in D. melanogaster and Zimbabwe in D. simulans, genomic DNA was extracted using the Gentra systems DNA isolation kit by Puregene. For the 3′ region, 1656 bp was amplified using PCR (SAKI et al. 1988) with the following primers: N5452 5′-CGGTATAGTCGACGTCAGCAG and RN7087 5′-GCAATTCAGTCTGCGCAGCG. This region corresponds to bases 5452–7087 of GenBank accession M16152. For D. simulans lines 5452–7087 of GenBank accession M16152. For D. simulans lines 5452–7087 of GenBank accession M16152.

Sequencing was done manually using the Amersham (Arlington Heights, IL) cycle sequencing kit (as directed by the manufacturers). Sequencing was done using internal primers. PCR products were cleaned using QIAGEN’s PCR purification kit. Sequencing was done manually using the Amersham (Arlington Heights, IL) cycle sequencing kit (as directed by the manufacturer), which utilizes the dyeoxy chain terminator method (SANGER et al. 1977).

For what we term Notch 3′, 1525 bp was amplified using PCR with the following primers: NFI 5′-GCTAATCCGGATTATC TATCTG and NRIB 5′-TGTGCTAGTAATGTAATTCG. This corresponds to bases 536–1860 of GenBank accession K05908. Using internal primers, 1488 bp were sequenced as above.

Sequences are deposited in GenBank under accession nos. AF361372–AF361422 for Notch 3′ and AF360581–AF360631 for Notch 3′ in D. melanogaster. For both regions the D. simulans and D. yakuba accessions are AV191369–AV191414.

Sequences were aligned using MegAlign of the DNASTAR software package and analyzed using the DnaSP 3.0 program (ROZAS and ROZAS 1997). The effective number of synonymous and nonsynonymous sites was estimated using DnaSP 3.0. This program was also used to calculate θ and π, estimates of 3NXμ (since Notch is on the X chromosome). P values for Tajima’s D, Fu and Li’s D, and Fu’s Fs tests were obtained using the coalescent simulator of DnaSP 3.0, assuming either no recombination or R = 3NXμ = 94. This choice of R is based on the genetic map-based recombination rate at Notch of r = 2.1 × 10−5 recombinants/generation/bp (HEY and KLIMAN 2002) and an estimate of N of 1 × 105 (KREITMAN 1983) for D. melanogaster.

FAY and WU’S (2000) H test was performed on-line at http:// crimp.lbl.gov/hittest.html in the following manner. Ten thousand simulations were performed with R = 3NXμ = 94 and assuming the probability of back mutation is 0.01 and 0.02 for Notch 5′ and Notch 3′, respectively (chosen on the basis of observed levels of divergence). Segregating sites for which the derived state was ambiguous were not considered.

Nucleotide divergence between species: Sequences were aligned using MegAlign of the DNASTAR software package, with some alignments adjusted manually to keep gaps in-frame in coding regions. We estimated synonymous site divergence between D. melanogaster and D. simulans for both regions of Notch plus 79 other protein-coding regions. We considered only regions for which at least 50 synonymous sites could be compared between species. Notch 5′ plus 26 other gene regions also contained intron data where at least 100 bp of intron sequence could be aligned between these species. Pairwise divergence for synonymous and intron sites was estimated using Kimura’s two-parameter model (KIMURA 1980), using the program Sequencer 6.1.0 (written by B. Kessing and available at http://nmg.si.edu/).

The relative rates test of TAJIMA (1993) was used to evaluate differences in the rate of substitution along the D. melanogaster and D. simulans lineages using D. yakuba as the outgroup. A list of the loci used in this study and their GenBank accession numbers are given in supplemental Appendix S1 at http://www.genetics.org/supplemental/

Estimating the effective number of preferred, unpreferred, and equally preferred synonymous sites: For D. melanogaster, D. simulans, and D. pseudoobscura, preferred codons had previously been determined by comparing codon usage between the 10% lowest and 10% highest biased genes (SHEILDs et al. 1988; AKASHI 1994). Depending on the amino acid, G-and/or C-ending codons were deemed preferred (indicated in Table S1 at http://www.genetics.org/supplemental/) as they were found to be used significantly more often at the high-biased loci compared to low-biased ones. Following AKASHI (1995), an “unpreferred change” is a change within a synonymous family from a preferred to an unpreferred codon. Changes from an unpreferred to a preferred codon are called “preferred,” and those among unpreferred or preferred codons (a few synonymous families have two preferred codons) are called “equal.”

We attempt to estimate the “effective number” of each type of synonymous change (in the same vein as the effective number of synonymous or nonsynonymous sites). We call these
the effective number of synonymous preference sites, and they represent the mutation potential toward each type of change. As such, any bias in the mutation process can have profound effects on the estimation of the number of sites. We have used the mutation rate estimates obtained by Petrov and Hartl (1999) from the substitution pattern at pseudogenes (importantly these authors report relative mutation rates for each type of change). The mutation bias was incorporated by adjusting the proportion of substitutions at each codon position that would result in a preferred synonymous, unpreferred synonymous, or equal synonymous change [using Akashi’s (1995) classification]. The effective numbers of synonymous preference sites are listed for each of the 61 sense codons in Table S1 at http://www.genetics.org/supplemental/ (stop codons are ignored in these calculations). The number of each type of synonymous change (preferred, unpreferred, or equal) is then compared relative to the effective number of synonymous preference sites for that change. Our goal is to evaluate whether differences in the number of preferred and unpreferred fixations between these species’ lineages represent positive selection or changes in selective constraint. We discuss this logic more in the results section.

We first describe how we estimated the number of synonymous preference sites for each codon. Petrov and Hartl (1999) used observed substitutions in “dead on arrival” transposable elements (considered to be pseudogenes) that were located throughout the genome to infer the underlying mutational pattern in Drosophila. After accounting for base frequency differences, they document that C to T and G to A mutations occur 2.2 times more frequently than the average of all other changes. They found this bias to be consistent across two subgenera of the Drosophila radiation (Sophophora and Drosophila). This mutation bias will affect estimates of both the effective number of synonymous and nonsynonymous sites and the effective number of synonymous preference sites in ways. First, when a nucleotide mutates, the likelihood that it mutates to each of the other three bases is not equal. Second, codon positions made up of G’s and C’s will have a higher mutation rate than those made up of A’s and T’s.

When one assumes that all mutations occur at an equal rate, all nucleotide positions in a sequence are considered equal (i.e., each represents 1 “site” of the sequence) and each of the three changes possible for a nucleotide are considered equal (i.e., each change is considered as one-third of a site). To incorporate the mutation bias, we scaled the value given to each change from one nucleotide to another (0.333) by the percentage of difference between the number of times that mutation was observed by Petrov and Hartl (1999) to that expected if all mutations occurred at an equal rate (i.e., 1/12 or 0.0833, see Figure 2). As an example consider all possible mutations from the C nucleotide. From Petrov and Hartl (1999), a C to T mutation occurs 15% of the time, while with no bias we would expect this value to be 8.33%. Thus, C to T mutations are now considered as (0.333(0.15 + 0.0833)) = 0.60 of a site instead of 0.333. Likewise, C to A mutations represent (0.333(0.09 + 0.0833)) = 0.36 and C and G mutations (1/3(0.06 + 1/12)) = 0.24 of a site. As a result codon positions occupied by C and G are now considered (0.60 + 0.24 + 0.36) = 1.2 sites and T and A positions as (0.32 + 0.28 + 0.20) = 0.8 sites.

To illustrate how we incorporate these mutation-biased pathways into our estimate of the number of synonymous preference sites we consider the codon CGG. One, zero, and three of the three possible changes at sites 1, 2, and 3, respectively, of the codon are synonymous. The C to A change at the first site is an equal synonymous change (see Table S1 at http://www.genetics.org/supplemental/), so 0.36 of first-position changes are equal preference. The only other synonymous changes are at site 3: G to A and G to T are both equal (so 0.96) and the G to C is preferred (thus 0.24). On a per codon basis (summing over the three codon positions), the number of “preferred synonymous sites” is thus 0 + 0 + 0.24 = 0.24, the number of “equal synonymous sites” is 0.36 + 0 + 0.96 = 1.32, and there are no “unpreferred synonymous sites.” So, summing across the codon, CGG is considered as 0.24 preferred and 1.32 equal synonymous preference sites.

To apply this method to a coding region, we inferred the ancestral sequence of D. melanogaster and D. simulans assuming parsimony, using D. yakuba as the outgroup. Given our use of parsimony, multiple mutational hits are not taken into account. To determine whether multiple hits affect our reconstruction of the ancestral sequence, we also used a maximum-likelihood method (using PAML; Yang and Nielsen 1998, 2000) to reconstruct the ancestor at Notch 3 (the coding region for which we observe the highest level of divergence). The highly supported ancestral sequence constructed using maximum likelihood was identical to the parsimony sequence. This finding supports the use of parsimony to reconstruct the ancestral sequence between these closely related species. However, the congruence of the two methods does not completely eliminate the uncertainty in estimating ancestral states, which may increase the variance of our estimate. The small number of codons for which the ancestral state of any of the three positions could not be inferred was excluded from further analysis. For each codon, the effective numbers of preferred, unpreferred, and equal synonymous preference sites (given in Table S1 at http://www.genetics.org/supplemental/) were multiplied by the number of times that codon occurred in the ancestral sequence for each coding region. These numbers were summed across codons, leading to an overall estimate of the effective number of preferred, unpreferred, and equal synonymous preference sites for the entire coding region. Two-by-two contingency tables were used to compare the rates of preferred and unpreferred fixations per site. Significance was determined using Fisher’s exact test. We point out that unless an equal number of purines and pyrimidines are found within a coding region, summing the

Figure 2.—Mutation pathways for Drosophila. Numbers not in parentheses are values taken from Petrov and Hartl (1999) divided by two, to represent the proportion of times such a mutation was observed out of the total of 12 possible changes in the pathway. Numbers in parentheses are the values used to incorporate the mutation bias when estimating the effective number of synonymous preference sites (see Materials and Methods).
number of synonymous and nonsynonymous sites will not equal the total length of the region when the mutations bias is taken into account. These values do, however, reflect the differential mutational potential between the different types of sites.

A program was written to construct the ancestral sequence and then partition the effective number of synonymous sites into the effective number of preferred, unpreferred, and equal synonymous preference sites as detailed above. This program also determines the number of preferred, unpreferred, and equal changes that have occurred along each species lineage for both fixed and pairwise comparisons and is available upon request from the authors.

**Correcting for multiple tests:** In this study a number of tests are applied first to the two *Notch* regions and subsequently to a number of other loci to determine the generality of the *Notch* results. For the test results at *Notch*, we apply the Bonferroni correction, considering that only those two regions have been tested. For example, with the relative rates test on the synonymous sites the necessary *P* values required to obtain significance are 0.05 ÷ 2 (since two tests were performed). For the remainder of the loci, 0.05 is divided by the total number of loci compared (including the *Notch* regions) to determine the significant *P* value. For tests applied along only one species lineage we consider only the number of loci compared for that species.

**RESULTS**

**Nucleotide variability — *D. melanogaster***: Polymorphic sites at the 3′ and 5′ regions of *Notch*, respectively, are presented in Figures S1 and S2 at http://www.genetics.org/supplemental/ for *D. melanogaster*. Estimates of nucleotide variability within each population are summarized in Table 1. As has been seen for other X-linked genes in this species (e.g., Begun and Aquadro 1993; Andolfatto 2001), there tends to be more nucleotide and haplotype diversity in Zimbabwe than in non-African populations. While the levels of variability observed at *Notch* are within the range seen at other loci in this species (e.g., Moriyama and Powell 1996), the 3′ region is less variable than the 5′ region for three of the four populations of *D. melanogaster*. This is in contrast to synonymous site divergence between *D. melanogaster* and *D. simulans*, which is higher at *Notch* 3′ than at *Notch* 5′ (Table 1, Figure 3).

We use the Hudson-Kreitman-Aguadé (HKA) test (Hudson et al. 1987) to evaluate deviations from the neutral expectation of a consistent ratio of variability and divergence among gene regions. Both regions of *Notch* are compared to one another as well as to other published X chromosome loci (Table 2). Only X chromosomal loci are used to avoid assumptions about differences in effective population size between the X chromosome and autosomes (a contrast complicated by assumptions about sex ratios in natural populations; e.g., Andolfatto 2001). We perform the test in two ways. First we consider total pairwise differences between *D. melanogaster* and *D. simulans* and then we consider only the differences that have occurred specifically along the *D. melanogaster* lineage (using *D. yakuba* as an outgroup). A significant difference between the two regions of *Notch* is observed in Ecuador with the U.S. results being marginally significant when we consider total divergence. In no population is there a significant difference between the two regions of *Notch* with the lineage-specific analyses. For the majority of the comparisons to other loci there is a significant departure with *Notch 3′* but not with *Notch 5′*. In general the results are more strongly significant with the lineage-specific analyses. The exceptions are nonsignificant results with analysis between both *Notch 3′* and other loci for which variability has previously been shown to be reduced. For example, *Pgd* has previously been shown to have significantly reduced variability compared to divergence (Begun and Aquadro 1994), and Labate et al. (1999) reported a reduced level of variability at *runt*. Therefore, in general, *Notch 3′* appears to be evolving in a manner different from other presumably neutral X chromosome loci. The HKA results suggest too little variability and/or too much divergence at *Notch 3′* compared to strict neutral expectations (Table 2). The pattern of sequence evolution along the *D. melanogaster* lineage appears to be the primary cause of this departure. In China there is a significant difference between *Notch 5′* and *vermilion* with the lineage-specific analysis. China differs from the other populations in that there is no marked increase in variation between *Notch 5′* and *Notch 3′* (Table 1).

Tajima’s (1989) *D* and *F* and Li’s (1993) *D* statistics tend to be negative for the 5′ region in all populations and for the 3′ region in Zimbabwe (Table 1), but the departures are not significant. In contrast, for *Notch 3′* most of the non-African populations have positive statistics for both Tajima’s and Fu and Li’s tests (Table 1). Only *Notch 3′* in Ecuador is significantly positive (marginally), with two intermediate-frequency segregating sites (Table 1 and Figure S1 at http://www.genetics.org/supplemental/). *F*’s (1997) *F* statistic is negative at both regions of *Notch* but is not significant in any population when recombination is taken into account. Fay and Wu’s (2000) *H* test reveals that no individual population sample or region has a significant excess of high-frequency-derived variants.

Application of the McDonald-Kreitman test (McDonald and Kreitman 1991) to each *Notch* region revealed no significant departure from the neutral prediction of equivalent ratios of synonymous and nonsynonymous polymorphism to divergence. This is also true when we combine data from the two regions (36 synonymous polymorphisms: 101 synonymous differences compared to 2 nonsynonymous polymorphisms: 3 nonsynonymous differences). The very small number of nonsynonymous variants, however, gives this test little power.

**Nucleotide variability — *D. simulans***: Figures S3 and S4 at http://www.genetics.org/supplemental/ present the polymorphic sites at *Notch 3′* and *5′*, respectively, in *D. simulans*. Estimates of nucleotide variability within each population are summarized in Table 1. We observe
**TABLE 1**
Levels of nucleotide sequence variability and results of neutrality tests at the 5' and 3' regions of Notch

|                  | Synonymous | Replacement | Intron | Tajima's D | Fu and Li's D | Fu's Fs |
|------------------|------------|-------------|--------|-------------|---------------|---------|
|                  | n | Hap | S | 0 | π | S | 0 | π | S | 0 | π | 5' region | 3' region |
| **D. melanogaster:** polymorphism | | | | | | | | | | | | | |
| Zimbabwe         | | 12 | 12 | 9 | 0.027 | 0.028 | 0 | 0.000 | 0.000 | 17 | 0.007 | 0.006 | -0.611 | 0.109 | -4.431 |
|                  | | | | | | | | | | | | | 0.308 | 0.446 | 0.012 |
|                  | | | | | | | | | | | | | (0.059) | (0.412) | (0.111) |
| United States    | | 15 | 12 | 4 | 0.011 | 0.008 | 1 | 0.001 | 0.0004 | 6 | 0.002 | 0.001 | -0.613 | -1.022 | -6.085 |
|                  | | | | | | | | | | | | | 0.236 | 0.181 | 0.002 |
|                  | | | | | | | | | | | | | (0.141) | (0.069) | (0.515) |
| Ecuador          | | 11 | 6  | 6 | 0.003 | 0.004 | 0 | 0.000 | 0.000 | 6 | 0.002 | 0.002 | -0.174 | -0.377 | -0.578 |
|                  | | | | | | | | | | | | | 0.451 | 0.351 | 0.207 |
|                  | | | | | | | | | | | | | (0.395) | (0.319) | (0.971) |
| China            | | 12 | 4  | 2 | 0.006 | 0.005 | 0 | 0.000 | 0.000 | 1 | 0.0004 | 0.0002 | -0.729 | -0.007 | -1.70 |
|                  | | | | | | | | | | | | | 0.259 | 0.469 | 0.166 |
|                  | | | | | | | | | | | | | (0.225) | (0.448) | (0.454) |
| **D. simulans:** polymorphism | | | | | | | | | | | | | |
| Zimbabwe         | | 10 | 10 | 5 | 0.017 | 0.018 | 0 | 0.000 | 0.000 | 19 | 0.008 | 0.008 | -0.239 | -0.691 | -3.023 |
|                  | | | | | | | | | | | | | 0.432 | 0.441 | 0.046 |
|                  | | | | | | | | | | | | | (0.285) | (0.108) | (0.260) |
| United States    | | 12 | 10 | 5 | 0.016 | 0.014 | 0 | 0.000 | 0.000 | 16 | 0.006 | 0.008 | 0.858 | 1.125 | -1.864 |
|                  | | | | | | | | | | | | | 0.185 | 0.107 | 0.101 |
|                  | | | | | | | | | | | | | (0.035) | (0.027) | (0.910) |
| Divergence       | | 21 | 0  | 21 | 0  | 21 | 21 | 0  | 21 | 21 | 0  | 21 | 21 | 0  | 21 |
| **D. melanogaster:** polymorphism | | | | | | | | | | | | | |
| Zimbabwe         | | 12 | 12 | 22 | 0.020 | 0.020 | 0 | 0.000 | 0.000 | -0.053 | -0.012 | -6.334 |
|                  | | | | | | | | | | | | | 0.514 | 0.529 | 0.002 |
|                  | | | | | | | | | | | | | (0.467) | (0.507) | (0.169) |
| United States    | | 15 | 6  | 6 | 0.005 | 0.005 | 0 | 0.000 | 0.000 | -0.547 | 0.620 | -0.649 |
|                  | | | | | | | | | | | | | 0.272 | 0.341 | 0.187 |
|                  | | | | | | | | | | | | | (0.234) | (0.297) | (0.905) |
| Ecuador          | | 11 | 4  | 2 | 0.002 | 0.003 | 0 | 0.000 | 0.000 | 1.665 | 0.966 | -0.627 |
|                  | | | | | | | | | | | | | 0.105 | 0.221 |
|                  | | | | | | | | | | | | | (0.019*) | (0.069) | (0.421) |
| China            | | 12 | 7  | 6 | 0.005 | 0.005 | 1 | 0.0003 | 0.0003 | -0.025 | 0.829 | -2.123 |
|                  | | | | | | | | | | | | | 0.530 | 0.206 | 0.077 |
|                  | | | | | | | | | | | | | (0.514) | (0.143) | (0.746) |
| **D. simulans:** polymorphism | | | | | | | | | | | | | |
| Zimbabwe         | | 10 | 8  | 29 | 0.028 | 0.026 | 0 | 0.000 | 0.000 | -0.304 | -0.141 | -0.398 |
|                  | | | | | | | | | | | | | 0.410 | 0.469 | 0.251 |
|                  | | | | | | | | | | | | | (0.256) | (0.390) | (0.956) |
| United States    | | 12 | 8  | 20 | 0.018 | 0.024 | 1 | 0.0003 | 0.0003 | 1.316 | 1.569 | 0.477 |
|                  | | | | | | | | | | | | | 0.073 | 0.025* | 0.255 |
|                  | | | | | | | | | | | | | (0.001**) | (0.005**) | (0.998) |

n is the number of chromosomes sequenced for each population. S is the number of segregating sites. Hap is the number of haplotypes found within each population. Divergence is uncorrected pairwise differences between D. melanogaster no. 1 and U.S. D. simulans no. 18. The 3' region consists of 1581 bp of coding sequence, with 386 and 1195 effective numbers of synonymous and nonsynonymous sites, respectively. The 5' region consists of 465 bases of coding (112 and 353 effective numbers of synonymous and nonsynonymous sites, respectively) and 1023 bp of intron sequence. Due to insertion/deletion differences only 844 intron sites could be directly compared between the species. Intron variability reported here is only for polymorphic sites that occur within these 844 bases. Tajima's D, Fu and Li's D, and Fu's Fs values are given with associated P values obtained from the coalescent simulator of DNAsp (Rozas and Rozas 1997) with no recombination, or in parentheses, with R = 94 as explained in text. Significance is considered at the 0.025 level as the simulation results are for a two-tailed test. *Significant at the 0.025 level; **significant at the 0.010 level.
slightly more variation in the Zimbabwe population than in the U.S. one. We note the pronounced haplotype structuring in the U.S. population, especially at Notch 3’ (Figures S3 and S4 at http://www.genetics.org/supplemental/). There is much less structure in the Zimbabwe population. In fact, only one of the haplotypes observed in the U.S. population at Notch 5’ is well represented in the African sample.

The HKA test results for D. simulans are given in Table 3. We detect no departure from neutrality in the relationship between levels of variability and divergence between Notch 5’ and Notch 3’ in either population. Neither region of Notch differs significantly from other X chromosomal loci although there is a trend for Notch 3’ to have a lower ratio of variability to divergence. The results are the same whether the HKA test is applied using total divergence or when considering only the divergence along the D. simulans lineage (as described above).

Tajima’s D and Fu and Li’s D tests (Table 1) are both significant at Notch 3’ in the U.S. population of D. simulans. Fu and Li’s D is also marginally significant at Notch 5’ in this population. All departures are positive, suggesting too many intermediate-frequency variants. This is in agreement with the visual pattern observed in the polymorphic site tables (Figures S3 and S4 at http://www.genetics.org/supplemental/). No test is significant in the Zimbabwe population. Fu’s Fs statistic tends to be negative in these populations but is not significant when recombination is considered. Also, no departures from neutrality were detected in either population with Fay and Wu’s (2000) H test.

We also applied the McDonald-Kreitman test (McDonald and Kreitman 1991) to the D. simulans data. Again we found no significant departures from neutral predictions. As with D. melanogaster, this is true even when the data across the two Notch regions are combined (41 synonymous polymorphisms:101 synonymous differences, compared to 1 nonsynonymous polymorphism:3 nonsynonymous differences).

**Synonymous site divergence at Notch:** We observe low to moderate levels of synonymous site polymorphism in both species yet extremely high levels of divergence. In contrast, intron divergence at Notch is less than the mean of 27 intron regions compared (Figure 3). This suggests that a regionally high mutation rate does not explain the high level of synonymous divergence at Notch. Could these results indicate a role of positive selection in the accelerated fixation of synonymous mutations at this locus? We present a number of analyses to attempt to answer this question.

Relative rate tests on pairwise sequence divergence (rooted with D. yakuba) demonstrate that significantly more synonymous substitutions have occurred along the D. melanogaster lineage than along D. simulans at both regions of Notch (Table 4). This result is in agreement with those of Akashi (1996) who noted a trend of more
TABLE 2

HKA test data and \( P \) values in the four \( D. \ melanogaster \) populations

| Region          | Notch 5'   | \( v^* \)  | \( P_{gt}^b \) | \( G_6pd^c \) | \( runt^d \) |
|-----------------|------------|------------|----------------|--------------|--------------|
| Zimbabwe        |            |           |                |              |              |
| Notch 5'        | [43, 67 (32)] | \( P = 0.678 \ (0.432) \) | \( P = 0.562 \ (0.224) \) | \( P = 0.983 \) |
| [26, 52 (36)]   |            |           |                |              |              |
| Notch 3'        | [43, 67 (32)] | \( P = 0.678 \ (0.432) \) | \( P = 0.562 \ (0.224) \) | \( P = 0.983 \) |
| [22, 112 (61)]  | \( P = 0.119 \ (0.300) \) | \( P = 0.052 \ (0.069) \) | \( P = 0.023^* \ (0.017^*) \) | \( P = 0.115 \) |
| USA             |            |           |                |              |              |
| Notch 5'        | [17, 67 (32)] | \( P = 0.620 \ (0.300) \) | \( P = 0.184 \ (0.120) \) |              |
| [10, 50 (36)]   |            |           |                |              |              |
| Notch 3'        | [17, 67 (32)] | \( P = 0.620 \ (0.300) \) | \( P = 0.184 \ (0.120) \) |              |
| [6, 108 (61)]   | \( P = 0.070 \ (0.164) \) | \( P = 0.024^* \ (0.015^*) \) | \( P = 0.506 \ 0.013^* \ (0.002^*\) |              |
| Ecuador         |            |           |                |              |              |
| Notch 5'        | [12, 67 (32)] | \( P = 0.969 \ (0.527) \) | \( P = 0.969 \ (0.527) \) |              |
| [7, 48 (36)]    |            |           |                |              |              |
| Notch 3'        | [12, 67 (32)] | \( P = 0.969 \ (0.527) \) | \( P = 0.969 \ (0.527) \) |              |
| [2, 108 (61)]   | \( P = 0.023^* \ (0.066) \) | \( P = 0.024^* \ (0.015^*) \) | \( P = 0.506 \ 0.013^* \ (0.002^*\) |              |
| China           |            |           |                |              |              |
| Notch 5'        | [16, 67 (32)] | \( P = 0.139 \ (0.045^*) \) | \( P = 0.139 \ (0.045^*) \) |              |
| [3, 47 (36)]    |            |           |                |              |              |
| Notch 3'        | [16, 67 (32)] | \( P = 0.139 \ (0.045^*) \) | \( P = 0.139 \ (0.045^*) \) |              |
| [6, 107 (61)]   | \( P = 0.887 \ (0.859) \) | \( P = 0.045^* \ (0.028^*) \) | \( P = 0.045^* \ (0.028^*) \) |              |

Comparisons were made only between loci containing the same population sample and with similar rates of recombination. For each gene region we give the number of polymorphic and divergent sites in brackets. The number of polymorphic sites is given before the comma. After the comma we give the total number of pairwise differences between \( D. \ melanogaster \) and \( D. \ simulans \) (not in parentheses) and the number of differences that have occurred specifically along the \( D. \ melanogaster \) lineage (in parentheses). Lineage-specific analyses were not possible for all gene regions. \( P \) values in parentheses are for lineage-specific comparisons. \( ^*P < 0.05; ^{**}P < 0.01. \)

\( ^a \)Data are from Begun and Aquadro (1995).
\( ^b \)Data are from Begun and Aquadro (1994).
\( ^c \)Data are from Eanes et al. (1996).
\( ^d \)Data are from Larget et al. (1999).

Synonymous substitutions along the \( D. \ melanogaster \) lineage compared to \( D. \ simulans \). We applied the relative rate test to a number of other loci sequenced in these species (Table 4). Without a Bonferroni correction, four other loci are significant at the 5% level: three with more synonymous substitutions occurring in \( D. \ melanogaster \) (per, Amy-P, and Amyrel) and one with more substitutions in \( D. \ simulans \) (mei-218). Only Notch remains significant after a Bonferroni correction is applied. Summing across loci we observe significantly more synonymous substitutions along the \( D. \ melanogaster \) lineage even after correcting for multiple tests.

Interestingly, there is no significant difference between lineages in intron divergence at Notch or at any of the other loci after a Bonferroni correction. Also, when summing across introns there is no trend for more substitutions along the \( D. \ melanogaster \) lineage. In addition, intron divergence is significantly lower than fourfold synonymous divergence (Mann-Whitney \( P = 0.023^* \)). (The comparison of intron to fourfold synonymous divergence is chosen to avoid the need to assume a particular transition/transversion bias when estimating synonymous site divergence.) Comparatively lower intron divergence could reflect stronger functional constraint or lower mutation rate in introns and/or positive selection accelerating divergence at synonymous sites. We thus compare levels of polymorphism and divergence between intron data at Notch 5' and synonymous data at Notch 5' and Notch 3' in Zimbabwe. While there is no significant difference between the ratio of polymorphism and divergence at intron and synonymous sites within Notch 5' (17 intron polymorphisms and 34 intron differences vs. 9 synonymous polymorphisms and 18 synonymous differences, Fisher’s exact test \( P \) value \( \geq 0.999 \)), the ratios are significantly different between Notch 5' intron and Notch 3' synonymous sites (17 intron polymorphisms and 34 intron differences vs. 22 synonymous polymorphisms and 113 synonymous differences, Fisher’s exact test \( P \) value = 0.015). Note that the large number of synonymous differences at Notch 3' appears to be the outlier. The 5' intron and 3' synonymous data are not, however, significant with the more conservative HKA test, which takes into account evolutionary variance (\( P \) value = 0.113). Thus, constraint or mutation rate differences cannot be completely discounted for the lower intron divergence, yet the tendency toward...
greater synonymous site divergence at Notch 3’ suggests a potential role of positive selection acting on synonymous sites in this region. Below we perform additional analyses to further investigate this possibility.

Akashi (1996) showed that on average 60–70% of synonymous differences involving unpreferred and preferred codons between these species have the unpreferred codon in D. melanogaster. At Notch 5’ and Notch 3’, respectively, 89 and 98% of such synonymous divergent sites have the unpreferred codon in D. melanogaster. Relaxation of selective constraint on codon bias (due to a smaller effective population size) has been proposed as the cause of the greater number of unpreferred fixations in D. melanogaster compared to D. simulans (Akashi 1995, 1996; McVean and Vieira 2001). Given that Notch appears to be an extreme example of this general trend we wanted to discriminate between relaxation of constraint and positive selection as the cause for the codon usage patterns at this locus. We did this by developing a method to estimate the effective number of synonymous preference sites.

If relaxation of constraint is the sole explanation for the difference in synonymous evolution between D. melanogaster and D. simulans, the ratio of the number of preferred differences to the effective number of preferred synonymous preference sites should equal the ratio of unpreferred differences per unpreferred synonymous preference site along the D. melanogaster lineage. The class of changes with a significantly lower ratio could be deleterious, and ones with a higher ratio may be advantageous.

In Table 5 we report the number of changes observed and the effective number of preferred, unpreferred, and equal preference sites in D. melanogaster and D. simulans, for both regions of Notch. For these comparisons we consider fixed differences as we are evaluating per site rates of evolution for different types of synonymous mutations. In this manner we minimize the effects of segregating deleterious mutations that will never go to fixation but could be counted as such in pairwise comparisons. With Fisher’s exact test, Notch 3’ in D. melanogaster shows significantly more unpreferred fixations than preferred fixations per site (Table 5). We also report equal substitutions per site for comparison. No significant difference is observed at Notch 5’ in D. melanogaster although the trend is in the same direction as Notch 3’ (for Notch 5’ the power of the comparison is limited by the small number of preferred synonymous sites). In D. simulans, we observe significant differences at both Notch 3’ and 5’ (Table 5). However, in this species we observe a significant excess of preferred compared to unpreferred fixations per site. To determine whether these results at Notch are part of a genome-wide phenomenon, we repeated our analysis for loci for which D. melanogaster, D. simulans, and D. yakuba have been sequenced and for which there are at least 450 bases of coding sequence (Tables 6 and 7). Again, we consider only fixed differences that have occurred along each species lineage. Differences in the number of loci between the species reflect availability of polymorphism data.

First we consider the results in D. melanogaster. Only Notch 3’ has a significant excess of unpreferred compared to preferred substitutions per site in this species. While more loci have a higher unpreferred than preferred rate of substitution, the Wilcoxon signed rank

**TABLE 3**

HKA test data and P values in the two D. simulans populations

| Notch 5’ | v \(^a\) | dec-1 \(^b\) | G6pd \(^c\) | mei-218 \(^d\) |
|---------|------|--------|----------|------------|
| Zimbabwe |  |  |  |  |
| Notch 5’ | [32, 33] |  |  |  |
| [24, 61 (14)] | P = 0.212 | — | P = 0.561 | — |
| Notch 3’ | [32, 33] |  |  |  |
| [29, 108 (38)] | P = 0.542 (0.359) | P = 0.041* | — | P = 0.224 |
| United States |  |  |  |  |
| Notch 5’ | [41, 67 (23)] | [11, 37 (15)] | [14, 48 (22)] | [6, 27 (22)] |
| [21, 50 (14)] | P = 0.572 (0.867) | P = 0.856 (0.615) | P = 0.712 (0.411) | P = 0.549 (0.101) |
| Notch 3’ | [41, 67 (23)] | [11, 37 (15)] | [14, 48 (22)] | [6, 27 (22)] |
| [20, 108 (38)] | P = 0.179 (0.212) | P = 0.055 (0.148) | P = 0.307 (0.496) | P = 0.386 (0.696) | P = 0.645 (0.543) |

Comparisons were made only between loci containing the same population sample and with presumably similar rates of recombination. For each gene region we give the number of polymorphic and divergent sites in brackets. The number of polymorphic sites is given before the comma. After the comma we give the total number of pairwise differences between D. melanogaster and D. simulans (not in parentheses) and the number of differences that have occurred specifically along the D. simulans lineage (in parentheses). Lineage-specific analyses were not possible for all gene regions. P values in parentheses are for lineage-specific comparisons. *P < 0.05.

\(^a\)Zimbabwe data are from Hamblin and Vieille (1999); U.S. data are from Begun and Aquadro (1995).

\(^b\)Data are from Begun and Whitley (2000).

\(^c\)Zimbabwe data are from Hamblin and Vieille (1999); U.S. data are from Eanes et al. (1996).
test is not significant (P value = 0.322). Therefore, in *D. melanogaster* it appears that most but not all of the apparent higher rate of unpreferred substitutions can be explained by *Drosophila*’s mutation bias together with relaxation of constraint on codon bias along the *D. melanogaster* lineage. The exception is *Notch*. Not only are unpreferred substitutions occurring faster than preferred at *Notch 3’*, but unpreferred divergence at both regions of *Notch* is significantly higher than that of all the other loci individually or combined in this species (*Notch 3’*, 47 unpreferred fixations out of 264 sites vs. total across other loci, 146 unpreferred fixations out of 4965 sites, P value < 0.0001; *Notch 5’*, 12 unpreferred fixations out of 97 sites vs. total across other loci, 146 unpreferred fixations out of 4965 sites, P value < 0.0001). No difference in the rate of preferred and equal divergence is observed between *Notch* and the other loci.

Could this extreme rate of unpreferred divergence at *Notch* (specifically *Notch 3’*) in *D. melanogaster* simply reflect an even more extreme mutation bias toward A’s and T’s than that inferred by Petrow and Hartl (1999)? One approach to this question is to examine substitutions at nearby intron sites (hoping they more closely reflect the underlying mutation process). We do not have intron data immediately adjacent to *Notch 3’*. However, at *Notch 5’* we can compare the number of G/C to A/T substitutions relative to the total number of G’s and C’s in the ancestor between fourfold degenerate and adjacent intron sites. Note that while we could not detect a significantly elevated number of unpreferred vs. preferred substitutions per site along the *D. melanogaster* lineage for *Notch 5’* alone, unpreferred divergence per site at this region is significantly greater than that of the other genes analyzed. This is consistent with the *Notch 5’* pattern of evolution being a less extreme case of the pattern seen at *Notch 3’*. We found that G/C to A/T fixations have occurred significantly more often at fourfold degenerate sites than at introns in the *Notch 5’* region (Fisher’s exact P value = 0.004; six fourfold G/C to A/T substitutions, with 60 G’s and C’s in ancestor vs. nine intron G/C to A/T substitutions, with 462 G’s and C’s in ancestor). There is no difference between fourfold and intron sites in the rate of A/T to G/C mutations on a per base level (Fisher’s exact P value > 0.999; no fourfold A/T to G/C substitutions, with 13 A’s and T’s in ancestor vs. three intron A/T to G/C substitutions, with 642 A’s and T’s in ancestor). These data suggest that substitution of A and T-ending codons (un preferred codons) at *Notch 5’* in *D. melanogaster* has occurred more frequently than the substitution of A’s and T’s in introns, suggesting that mutation bias alone cannot explain our results.

The pattern of synonymous site substitution is very different along the *D. simulans* lineage. Here, *Notch*, like many other loci, shows significantly more preferred than unpreferred fixations per site (Tables 5 and 7). Many loci remain significant after applying the Bonferroni

| Locus | Along mel | Along sim | χ² | Along mel | Along sim | χ² |
|-------|-----------|-----------|----|-----------|-----------|----|
| *Notch 3’* | 61 | 38 | 5.34** | 18 | 11 | 1.69 |
| *Notch 5’* | 18 | 3 | 10.71** | 4 | 9 | 1.92 |
| *ac* | 11 | 9 | 0.20 | 8 | 3 | 2.27 |
| *Adh* | 17 | 13 | 0.53 | 21 | 10 | 3.90* |
| *Amy-d* | 26 | 12 | 5.16* | 9 | 5 | 0.14 |
| *Amy-f* | 9 | 12 | 0.43 | 4 | 9 | 1.92 |
| *AP50* | 13 | 11 | 0.17 | 13 | 11 | 0.17 |
| *Cen190* | 8 | 12 | 0.80 | 14 | 10 | 0.67 |
| *dec-I* | 13 | 15 | 0.14 | 6 | 4 | 0.40 |
| *Est-6* | 23 | 21 | 0.09 | 9 | 7 | 0.25 |
| *g* | 14 | 10 | 0.67 | 4 | 3 | 0.14 |
| *Gdd1* | 6 | 4 | 0.40 | 7 | 3 | 1.60 |
| *Hex-A* | 9 | 10 | 0.05 | 7 | 3 | 1.60 |
| *Hex-C* | 20 | 10 | 3.35 | 7 | 3 | 1.60 |
| *HexIV* | 21 | 13 | 1.88 | 7 | 3 | 1.60 |
| *Hex2* | 22 | 13 | 2.31 | 7 | 3 | 1.60 |
| *His3* | 5 | 7 | 0.35 | 7 | 3 | 1.60 |
| *Hsc70-4* | 8 | 8 | 0.00 | 7 | 3 | 1.60 |
| *Hyd* | 9 | 7 | 0.25 | 7 | 3 | 1.60 |
| *Isc* | 4 | 3 | 0.14 | 7 | 3 | 1.60 |
| *mir218* | 10 | 16 | 1.38 | 7 | 3 | 1.60 |
| *mth* | 15 | 20 | 1.00 | 7 | 3 | 1.60 |
| *nu* | 7 | 3 | 1.60 | 7 | 3 | 1.60 |
| *osa* | 4 | 6 | 0.40 | 7 | 3 | 1.60 |
| *Osbp* | 8 | 13 | 1.19 | 4 | 2 | 0.67 |
| *per* | 29 | 11 | 8.10* | 1 | 8 | 5.44* |
| *Pgd* | 14 | 10 | 0.67 | 4 | 2 | 0.67 |
| *Pgi* | 17 | 15 | 0.15 | 4 | 2 | 0.67 |
| *Pgsc* | 22 | 28 | 0.72 | 4 | 2 | 0.67 |
| *Rel* | 24 | 19 | 0.58 | 4 | 2 | 0.67 |
| *sc* | 14 | 7 | 2.35 | 4 | 2 | 0.67 |
| *sn* | 16 | 11 | 0.95 | 4 | 2 | 0.67 |
| *sod* | 6 | 4 | 0.40 | 4 | 2 | 0.67 |
| *sog* | 12 | 8 | 0.80 | 4 | 2 | 0.67 |
| *Tpi* | 4 | 8 | 1.35 | 4 | 2 | 0.67 |
| *v* | 22 | 13 | 2.31 | 4 | 2 | 0.67 |
| *z* | 7 | 9 | 0.25 | 4 | 2 | 0.67 |
| *Zav* | 14 | 17 | 0.29 | 4 | 2 | 0.67 |
| Total mel | 615 | 502 | 11.43** | 536 | 461 | 5.64* |
| Total sim | 596 | 461 | 5.64* | 219 | 172 | 5.65* |

Chi-square test with 1 d.f. is shown (TAJMA 1993). mel, *D. melanogaster*; sim, *D. simulans*. *Significant at the 0.05 level without Bonferroni correction; **significant at the 0.05 level with Bonferroni correction.

* Totals without *Notch* regions.

* Totals when considering only loci with intron data.
TABLE 5
Ratios of the number of preferred, unpreferred, and equal substitutions per number of synonymous preference sites at both regions of Notch along each species' lineage

|                | Preferred | Unpreferred | P value | Equal |
|----------------|-----------|-------------|---------|-------|
| **D. melanogaster** |           |             |         |       |
| Notch 5'       | 1/17 = 0.059 | 12/97 = 0.124 | 0.688   | 1/23 = 0.043 |
| Notch 3'       | 0/63 = 0.000 | 47/264 = 0.178 | <0.0001** | 10/122 = 0.082 |
| Total across all loci | 26/948 = 0.027 | 205/5326 = 0.038 | 0.111   | 65/1544 = 0.042 |
| **D. simulans**  |           |             |         |       |
| Notch 5'       | 2/17 = 0.118 | 0/97 = 0.000 | 0.021*  | 0/23 = 0.000 |
| Notch 3'       | 23/63 = 0.365 | 1/264 = 0.004 | <0.0001** | 5/122 = 0.041 |
| Total across all loci | 92/1332 = 0.069 | 92/7006 = 0.013 | <0.0001** | 49/2183 = 0.022 |

P values are from Fisher’s exact tests comparing preferred and unpreferred mutations. **Significant with the Bonferroni correction; *significant only without the Bonferroni correction.

As with D. melanogaster, we compared the number of

TABLE 6
Numbers of preferred, unpreferred, and equal substitutions and number of synonymous preference sites for many loci along the D. melanogaster lineage

| Locus  | Preferred fixations | Preferred sites | Unpreferred fixations | Unpreferred sites | P value | Equal fixations | Equal sites |
|--------|---------------------|-----------------|-----------------------|-------------------|---------|-----------------|-------------|
| Acp26 Aa | 1.5                 | 42              | 2                     | 64                | 0.648   | 4               | 45          |
| Adh     | 0                   | 16              | 2                     | 191               | >0.9999 | 0               | 32          |
| Amy-d   | 1                   | 13              | 6                     | 387               | 0.208   | 1               | 48          |
| Amy-p   | 1                   | 13              | 9                     | 388               | 0.284   | 2               | 52          |
| anon1A3 | 2.5                 | 35              | 2                     | 94                | 0.287   | 2.5             | 38          |
| anon1E9 | 2                   | 80              | 1                     | 221               | 0.174   | 3               | 98          |
| anon1G5 | 0                   | 38              | 1                     | 72                | >0.9999 | 0               | 44          |
| Est-6   | 4                   | 87              | 11                    | 204               | >0.9999 | 2               | 106         |
| Hex-A   | 1                   | 42              | 5                     | 285               | 0.565   | 3               | 71          |
| Hex-C   | 1                   | 53              | 8                     | 243               | >0.9999 | 5               | 84          |
| Hex-t1  | 1                   | 51              | 6                     | 231               | >0.9999 | 2               | 80          |
| Hex-t2  | 0                   | 43              | 16                    | 262               | 0.141   | 6               | 67          |
| mth     | 1                   | 74              | 11                    | 196               | 0.189   | 2               | 101         |
| Notch 3' | 0                   | 63              | 47                    | 264               | <0.0001** | 10             | 122         |
| Notch 5' | 1                   | 17              | 12                    | 97                | 0.688   | 1               | 23          |
| per     | 2                   | 29              | 10                    | 364               | 0.606   | 7               | 100         |
| Pgi     | 5                   | 65              | 12                    | 306               | 0.206   | 3               | 93          |
| Pgm     | 2                   | 101             | 15.5                  | 383               | 0.389   | 7.5             | 146         |
| Rel     | 0                   | 6               | 2                     | 203               | >0.9999 | 0               | 22          |
| Tpi     | 0                   | 20              | 6                     | 124               | 0.596   | 0               | 32          |
| z       | 0                   | 22              | 9                     | 369               | >0.9999 | 1.5             | 65          |
| Zw      | 0                   | 22              | 9                     | 369               | >0.9999 | 1.5             | 65          |
| Total   | 26                  | 948             | 204.5                 | 5326              | 0.111   | 64.5            | 1544        |

P values are from Fisher’s exact tests comparing preferred and unpreferred mutations. **Significant with the Bonferroni correction.
TABLE 7
Numbers of preferred, unpreferred, and equal substitutions and number of synonymous preference sites for many loci along the D. simulans lineage

| Locus | Preferred fixations | Preferred sites | Unpreferred fixations | Unpreferred sites | P value | Equal fixations | Equal sites |
|-------|--------------------|----------------|----------------------|------------------|---------|----------------|-------------|
| Adh   | 0                  | 16             | 0                    | 191              | >0.9999 | 1              | 32          |
| anon1E9 | 2                  | 80             | 4                    | 221              | 0.66    | 1              | 98          |
| anon1A3 | 1                  | 35             | 1                    | 94               | 0.476   | 1              | 38          |
| anon1G5 | 1                  | 38             | 1                    | 72               | >0.9999 | 2              | 44          |
| AP50  | 1                  | 31             | 2                    | 270              | 0.279   | 1              | 68          |
| Cen190 | 4                  | 58             | 5                    | 178              | 0.229   | 1              | 57          |
| dec-1 | 4                  | 56             | 6                    | 143              | 0.472   | 4              | 85          |
| g     | 4                  | 34             | 2                    | 197              | 0.005*  | 2              | 60          |
| Hex-A | 2                  | 42             | 3                    | 285              | 0.125   | 5              | 71          |
| Hex-C | 1                  | 53             | 5                    | 243              | >0.9999 | 2              | 84          |
| Hex-41 | 3                  | 51             | 2                    | 231              | 0.043*  | 2              | 80          |
| Hex-42 | 4                  | 43             | 8                    | 262              | 0.073   | 1              | 67          |
| Hsc-70-4 | 2                  | 26             | 3                    | 312              | 0.049*  | 1              | 50          |
| hyd   | 1                  | 92             | 2                    | 165              | >0.9999 | 1              | 109         |
| mir218 | 2                  | 44             | 9                    | 150              | >0.9999 | 5              | 67          |
| mir   | 3                  | 20             | 3                    | 256              | 0.006*  | 1              | 48          |
| mth   | 4                  | 74             | 2                    | 196              | 0.05    | 1              | 101         |
| Notch 3’ | 23                 | 63             | 1                    | 264              | <0.0001*** | 5         | 122         |
| Notch 5’ | 2                  | 17             | 0                    | 97               | 0.021*  | 0              | 23          |
| osa   | 0                  | 44             | 0                    | 139              | >0.9999 | 0              | 94          |
| Osdp  | 2                  | 34             | 2                    | 197              | 0.104   | 0              | 60          |
| per   | 2                  | 29             | 3                    | 364              | 0.046*  | 2              | 100         |
| Pgdl  | 1                  | 19             | 5                    | 205              | 0.416   | 0              | 45          |
| Pg1   | 5                  | 38             | 1                    | 378              | <0.0001*** | 1         | 75          |
| Pgm   | 2                  | 65             | 6                    | 306              | 0.654   | 2              | 93          |
| Rel   | 3                  | 101            | 3                    | 383              | 0.109   | 2              | 146         |
| sn    | 3                  | 39             | 2                    | 259              | 0.017*  | 1              | 72          |
| sog   | 3                  | 42             | 3                    | 252              | 0.040*  | 1              | 75          |
| Tpi   | 2                  | 6              | 1                    | 203              | 0.002** | 0              | 22          |
| z     | 1                  | 20             | 3                    | 124              | 0.454   | 0              | 32          |
| Zw    | 4                  | 22             | 4                    | 369              | 0.0005*** | 3         | 65          |
| Total | 92                 | 1332           | 92                   | 7006             | <0.0001*** | 49        | 2183        |

P values are from Fisher’s exact tests comparing preferred and unpreferred mutations. **Significant with a Bonferroni correction; *significant only without the Bonferroni correction.

A/T to G/C mutations on a per base pair level between fourfold synonymous and intron positions in D. simulans. We found no difference (Fisher’s exact P value = 0.119; 1 fourfold A/T to G/C substitution, with 13 A’s and T’s in ancestor vs. 10 intron A/T to G/C substitutions, with 642 A’s and T’s in ancestor). However, the power of this comparison is compromised by the small number of A’s and T’s at fourfold degenerate sites.

Thus, specific types of synonymous changes have been accelerated in both species. In D. simulans there appears to be a genome-wide trend of an accelerated fixation of preferred changes per site, with Notch appearing to be an extreme example. In contrast, in D. melanogaster unpreferred substitutions appear accelerated at Notch. While there is a genome-wide trend in this same direction in D. melanogaster, it is not significant. However, in D. melanogaster the power of such comparisons is compromised by the small number of preferred sites inferred in the ancestor of these species.

DISCUSSION

Our interest in the Notch locus stemmed from previous data suggesting that this region was unaffected by positive selection in D. melanogaster. This is true at the amino acid level for the regions of Notch we surveyed. However, our data suggest that synonymous fixations at Notch have been accelerated by positive selection along both the D. melanogaster and D. simulans lineages.

At the level of polymorphism alone we do not detect any departures from neutral expectations in D. melanogaster. In D. simulans, we observe a significant excess of intermediate frequency variants in our U.S. sample. This result could be due to balancing selection and/or de-
mography. Haplotype structure as seen at Notch is also reported at other unlinked loci in non-African samples of D. simulans (e.g., Begun and Aquadro 1994; Eanes et al. 1996; Hamblin and Aquadro 1996; Hamblin and Veullie 1999; Labate et al. 1999). Thus, it is likely our result is due to demography, perhaps the result of historical admixture between divergent populations or population bottlenecks (Wall et al. 2002).

When comparing levels of polymorphism to divergence we detect significant non-neutral patterns in D. melanogaster. For example, the HKA test detects a significantly lower ratio of polymorphism to divergence at Notch 3′ compared to other “neutral” X chromosome loci in this species. The neutral theory predicts that regions with high divergence will also have high levels of polymorphism. However, if the variants examined tend to be advantageous on average themselves, then levels of divergence will be elevated relative to polymorphism. At the regions of Notch we studied, the most striking feature of the data is the extremely high level of synonymous site divergence, not the level or patterns of polymorphism. A closer examination of the synonymous site divergence data reveals a significantly greater number of unpreferred substitutions per site than preferred substitutions in D. melanogaster. In contrast, in D. simulans we observe significantly more preferred fixations per site than unpreferred ones. Thus, the pattern of synonymous site evolution at Notch suggests the influence of positive selection, but in the opposite directions in these closely related species.

What could be the cause of the acceleration of specific synonymous changes at Notch in these species? Synonymous codon usage bias is thought to be the result of a mutation or gene conversion bias, selection for translational accuracy/efficiency, and/or other forms of selection (i.e., selection for mRNA stability or regulation of transcription; reviewed in Akashi 2001 and Duret 2002). Although we have tried to address the issue of mutation bias by comparing the substitution pattern at fourfold synonymous and intron positions at Notch 5′, there remains the possibility that our results in D. melanogaster are simply due to an underestimation of the mutation bias at the 3′ end of Notch. Mutational processes in Drosophila have been shown to be context specific (Kliman and Eyre-Walker 1998). However, the base composition at Notch 3′ in the extant and ancestral sequences does not deviate from that observed at the other loci in our study. We note that the mutational biases that we use from Petrov and Hartl (1999) to infer the number of synonymous preference sites differ slightly from the biases estimated by McVean and Vieira (2001). To see if this difference could affect our results, we incorporated the McVean and Vieira (2001) mutational biases into our estimation of the number of sites and reapplied the test to Notch 3′. The conclusion of a higher level of unpreferred than preferred divergence per site at Notch 3′ in D. melanogaster remains unchanged (data not shown).

Our results in D. simulans support previous claims that positive selection is involved in the establishment of a bias toward preferred G- and C-ending codons in Drosophila (i.e., Akashi 1994; Akashi and Schaeffer 1997; Kliman 1999; Kern et al. 2002). With our method, positive selection appears to play a role not only at the Notch locus but also genome-wide in this species. We note that other studies (Begun 2001; McVean and Vieira 2001) have reported evidence of relaxation of constraint on codon bias along the D. simulans lineage. Such conclusions are drawn from observing an excess of unpreferred substitutions (i.e., more unpreferred substitutions than preferred) compared to that expected under mutation-selection-drift equilibrium. There are a number of potential differences between our method and those of Begun (2001) and McVean and Vieira (2001) to explain the contrasting results. For example, we compare the number of unpreferred and preferred substitutions on a per site basis while Begun (2001) does not. In addition, we draw our conclusions from fixed differences while McVean and Vieira (2001) use pairwise comparisons. McVean and Vieira (2001) note that pairwise comparisons may lead to an underestimation of selection coefficients.

As an aside, our data suggest not only that positive selection has shaped synonymous site evolution in D. simulans but also that this species is at mutation-selection-drift equilibrium. An equal number of preferred and unpreferred substitutions would be expected for a lineage at equilibrium (Bulmer 1991). This is what we observe when we sum fixed differences across loci. Also, due to evolutionary variance, at equilibrium some loci will have more preferred substitutions and others more unpreferred, but the number of loci that go one way or the other should be equal. This is indeed the case in D. simulans with 15 loci having more unpreferred fixations and 14 more preferred (Wilcoxon signed rank test P value = 0.370).

When considering the mutation-selection-drift equilibrium predictions detailed above, D. melanogaster does not appear to be at equilibrium. When summing across loci we observe approximately eight times more unpreferred fixations than preferred. Also, we observe more unpreferred fixations than preferred in 19 of the 21 loci studied (Wilcoxon signed rank test P value = 0.0001). This general observation of a greater number of unpreferred than preferred fixations in D. melanogaster has previously been interpreted as relaxation of constraint on codon bias along this species lineage (e.g., Akashi 1994, 1996). With our method we cannot reject this hypothesis genome-wide in D. melanogaster. However, relaxation of constraint does not appear sufficient to explain the large excess of unpreferred fixations per site in D. melanogaster at the Notch 3′ region.

Our results suggest that positive selection is involved
in the fixation of unpreferred mutations at Notch 3’, but the nature of that selection is presently unknown. There are a number of possible explanations. A recent switch in codon preference due to a change in tRNA abundance seems unlikely, given the resulting load of deleterious fixations across the genome as detailed by Akashi et al. (1998). Selection pertaining to Notch mRNA stability and/or gene expression and regulation could be involved. For example, a growing number of studies show the importance of synonymous sites in the functionality of proteins and in translation kinetics (Cortazzo et al. 2002; Duan et al. 2003; Oresic et al. 2003). Smith and Eyre-Walker (2001) have shown that in some cases the use of “suboptimal” codons in Escherichia coli appears to be due to some form of “conflicting selection” (i.e., regulation of gene expression or mRNA/DNA secondary structure). To the best of our knowledge the structure of the Notch mRNA is unknown, and no programs to predict the structure can input the entire 8-kb transcript.

The question remains as to why Notch is an extreme example of genome-wide trends in codon usage within each of these species and in synonymous codon usage differences between them. Notch resides in a region of high recombination, which may aid in the efficiency of weak selection due to a larger region-specific effective population size (i.e., Birky and Walsh 1988; Barton 1995). Pronounced differences in synonymous codon usage between the melanogaster and the ananassae and obscura species groups (i.e., more unpreferred fixations in the former) at the yellow locus are thought to be due to a marked reduction of recombination at the tip of the X chromosome in the melanogaster species group (Munte et al. 1997, 2001). If such an explanation were to account for the extreme differences in codon usage between the species used in this study, we would expect the recombination rate at Notch in D. melanogaster to appear relatively low. However, using the method of Wall (2000), we found that the population rate of recombination per base pair for both Notch regions is relatively high compared to other loci and is much higher than estimates for the Notch region based on integrated map methods (data not shown). The estimates of $R$ $(3N_e r$ for Notch) per base pair are 0.263 and 0.213 for Notch 5’ and Notch 3’, respectively.

It has also been noted that changes in mutation bias can cause different substitution processes even among closely related species and can lead to a brief burst of substitutions (Takano-Shimizu 1999, 2001). However, these studies conclude a strong role of mutation without calibrating the comparisons by the base composition of the ancestral sequence. As previously mentioned for D. melanogaster, when base composition is taken into account, GC to AT substitutions (the type of mutations that tend to result in unpreferred changes) have occurred significantly faster at fourfold degenerate sites than at introns at Notch 5’. This suggests to us that a recent switch in mutation bias does not explain our results.

The substitution process is governed by the input of mutations and fixation due to drift and selection. With the available data, it does not appear that mutational processes alone can explain our results, unless the mutation bias is more extreme in exons compared to introns. Thus, positive selection appears to have accelerated the fixation of a subset of synonymous codons at Notch in D. melanogaster and D. simulans. These results add to the growing caution in the use of synonymous site evolution as a neutral proxy (e.g., Akashi and Kreitman 1995; Bustamante et al. 2002; Fay et al. 2002; Smith and Eyre-Walker 2002; Swanson et al. 2003). For example, $d_s/d_\circ$ comparisons for which synonymous changes are assumed neutral may underestimate the presence of positive selection at the amino acid level. Our results may also shed light as to the cause of the lower average level and variance in intron divergence compared to synonymous site divergence observed between these species (Figure 3; Bauer and Aquadro 1997; Takano-Shimizu 2001). A striking difference is the skew toward high levels of divergence for synonymous sites but not introns. This pattern could indicate that intron mutations are in general more selectively constrained than synonymous. Given that Drosophila genes tend to have short introns (Mount et al. 1992), this explanation is plausible. On the other hand, our Notch results suggest that positive selection has the potential to accelerate the fixation of some synonymous mutations for at least some of the loci with high levels of synonymous site divergence.

We thank Willie Swanson, Floyd Reed, Jennifer Calkins, Leila Hatch, Rasmus Nielsen, Guy Reeves, Rick Durrett, Gil McVean, Hiroshi Akashi, Montse Aguadé, and Andy Clark for valuable comments on the manuscript and/or participating in productive discussions. We particularly thank Martha Hamblin for discussions that played a key role in instigating our comparison of the number of synonymous preference changes on a per site basis. This research was supported by National Institutes of Health grant GM36431 to C.F.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., 2001 Gene expression and molecular evolution. Curr. Opin. Genet. Dev. 11: 660–666.

Akashi, H., and M. Kreitman, 1995 Molecular evidence of natural selection. Annu. Rev. Ecol. Syst. 26: 403–422.

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

Akashi, H., R. M. Kliman and A. Eyre-Walker, 1998 Mutation

Synonymous Site Divergence in the Notch Region
pressure, natural selection, and the evolution of base composition in Drosophila. Genetics 102/103: 49–60.

Andolfatto, P., 2001 Contrasting pattern of X-linked and autosomal nucleotide variation in Drosophila melanogaster and Drosophila simulans. Mol. Biol. Evol. 18: 279–290.

Artavanis-Tsakonas, S., M. D. Rand and R. J. Lake, 1999 Notch signaling: cell fate control and signal integration in development. Science 284: 770–776.

Barton, N. H., 1995 Linkage and the limits to natural selection. Genetics 140: 821–841.

Bauer, V. L, and C. F. Aquadro, 1997 Rates of DNA sequence evolution are not sex-biased in Drosophila melanogaster and D. simulans. Mol. Biol. Evol. 14: 1252–1257.

Beattie, P., and U. Lande, 1998 Notch and neurogenesis. J. Neurosci. Res. 54: 125–136.

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

Begun, D. J., and C. F. Aquadro, 1991 Molecular population genetics of the distal portion of the X chromosome in Drosophila: evidence for genetic hitchhiking of the yellow-african region. Genetics 129: 1147–1158.

Begun, D. J., and C. F. Aquadro, 1992 Levels of naturally occurring DNA polymorphism correlate with recombination rates in D. melanogaster. Nature 356: 519–520.

Begun, D. J., and C. F. Aquadro, 1993 African and North American populations of Drosophila melanogaster are very different at the DNA level. Nature 365: 548–556.

Begun, D. J., and C. F. Aquadro, 1994 Evolutionary inferences from DNA variation at the β-phosphoglucomutase dehydrogenase locus in natural populations of Drosophila: selection and geographic differentiation. Genetics 136: 155–171.

Begun, D. J., and C. F. Aquadro, 1995 Molecular variation at the vermilion locus in geographically diverse populations of Drosophila melanogaster and D. simulans. Genetics 140: 1019–1032.

Begun, D. J., and P. Whitley, 2000 Reduced X-linked nucleotide polymorphism in Drosophila simulans. Proc. Natl. Acad. Sci. USA 97: 5960–5965.

Birry, C. W., Jr., and J. B. Walsh, 1988 Effects of linkage on rates of molecular evolution. Proc. Natl. Acad. Sci. USA 85: 6414–6418.

Bulmer, M., 1991 The selection-mutation-drift theory of synonymous nucleotide substitutions in Drosophila melanogaster. Mol. Biol. Evol. 153: 1753–1761.

Kimura, M., 1980 A simple method for estimating evolutionary rate of base substitution through comparative studies of nucleotide sequences. J. Mol. Evol. 16: 111–120.

Kliman, R. M., and D. J. Begun, 2002 Genomic effects of nucleotide substitutions in Drosophila simulans. Genetics 162: 153–159.

Kern, A. D., C. D. Jones and D. J. Begun, 2002 Recent selection on synonymous codon usage in Drosophila. J. Mol. Evol. 49: 343–351.

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

Kreitman, M., 1983 Nucleotide polymorphism at the alcohol dehydrogenase locus of Drosophila melanogaster. Nature 304: 412–417.

Labate, J. A., C. H. Bierman and W. F. Eanes, 1999 Nucleotide variation at the rDNA locus in Drosophila melanogaster and Drosophila simulans. Mol. Biol. Evol. 16: 724–731.

McWhinney, J. H., and M. Kreitman, 1991 Adaptive protein evolution at the Adh locus in Drosophila. Nature 20: 652–654.

McVeagh, G. A. T., and J. Vrijns, 2001 Inferring parameters of mutation, selection, and demography from patterns of synonymous-site evolution in Drosophila. Genetics 157: 245–257.

Moriyama, E. N., and J. R. Powell, 1996 Intraspecific nuclear DNA variation in Drosophila. Mol. Biol. Evol. 13: 261–277.

Mount, S. M., C. Bures, G. Hertz, G. D. Stormo, O. Wutte et al., 1992 Splicing signals in Drosophila: intron size, information content, and consensus sequences. Nucleic Acids Res. 20: 4255–4262.

Munteé, A., M. Aguadé and C. Segarra, 1997 Divergence of the yellow gene between Drosophila melanogaster and D. subobscura: recombination rate, codon bias and synonymous substitution. Genetics 147: 165–175.

Munteé, 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.

Oresic, M., M. H. H. Dein, D. Korzenblum and D. Shalloway, 2003 Tracing specific synonymous codon-secondary structure correlations through evolution. J. Mol. Evol. 56: 473–484.

Petrov, D. A., and D. L. Hartl, 1999 Patterns of nucleotide substitution in Drosophila and mammalian genomes. Proc. Natl. Acad. Sci. USA 96: 1475–1479.

Rebay, I., R. J. Fleming, R. G. Feigon, L. Cherras, P. Cherras et al., 1991 Specific EGF repeats of Notch mediate interactions with Delta and Serrate, implications for Notch as a multifunctional receptor. Cell 67: 687–699.

Rozas, J., and R. Rozas, 1997 DnaSP version 2.0: a novel software package for extensive molecular population genetics analysis. Comput. Appl. Biosci. 13: 307–311.

Saiki, R. K., D. H. Geleland, D. Stoffel, S. J. Scharf, R. Higuchi et al., 1988 Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239: 487–491.

Sanger, F., S. Nickson and A. R. Coulson, 1977 DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA 74: 5463–5467.

Scarpulla, S. W., C. F. Aquadro and C. H. Langley, 1988 Restriction-map variation in the Notch region of Drosophila melanogaster. Mol. Biol. Evol. 5: 30–40.

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

Smith, N. G. C. and A. Eyre-Walker, 2001 Why are translationally suboptimal synonymous codons used in Escherichia coli? J. Mol. Evol. 53: 225–236.

Smith, N. G. C., and A. Eyre-Walker, 2002 Adaptive protein evolution in Drosophila. Nature 415: 1022–1024.

Swanson, W. J., R. Nielsen and Q. Yang, 2003 Pervasive adaptive evolution in mammalian fertilization proteins. Mol. Biol. Evol. 20: 18–20.

Takano-Shimizu, T., 1999 Local recombination and mutation ef-
fects 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.
Tajima, F., 1989 Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123: 585–595.
Tajima, F., 1993 Simple methods for testing the molecular evolutionary clock hypothesis. Genetics 135: 599–607.
Wall, J. D., 2000 A comparison of estimators of the population recombination rate. Mol. Biol. Evol. 17: 156–163.

Wall, J. D., P. Andolfatto and M. Przeworski, 2002 Testing models of selection and demography in Drosophila simulans. Genetics 162: 203–216.
Yang, Z., and R. Nielsen, 1998 Synonymous and nonsynonymous rate variation in nuclear genes of mammals. J. Mol. Evol. 46: 409–418.
Yang, Z., and R. Nielsen, 2000 Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. Mol. Biol. Evol. 17: 32–43.

Communicating editor: M. Aguadé
