High Nucleotide Sequence Variation in a Region of Low Recombination in Drosophila simulans is Consistent with the Background Selection Model

Martha T. Hamblin and Charles F. Aquadro
Section of Genetics and Development, Cornell University

We surveyed nucleotide sequence variation at glucose dehydrogenase (Gld), in a region of low recombination on chromosome 3R, from a population sample of Drosophila simulans. The levels of nucleotide variation were surprisingly high. There was no departure from the expectation of a neutral model for the level of polymorphism, indicating no evidence of a selective sweep in this region. There was a significant deficiency of singleton polymorphisms according to the Fu and Li test, although Tajima and Hudson, Kreitman, and Aguade (HKA) tests do not provide evidence of a significant elevation of variation due to balancing selection. Genetic map data for the D. simulans third chromosome were used to calculate expected values of \( \pi \) for Gld under a current model of background selection, varying the values for the parameter sh (selection coefficient against deleterious mutations). We show that the recombinational landscape of D. simulans is sufficiently different from that of D. melanogaster that we expect higher variation under the background selection model, even when effective population sizes are assumed to be equal. The data for Gld were tested against the predictions using computer simulations of the distribution of the number of segregating sites conditioned on \( \pi \). Background selection alone can explain our observations as long as \( sh \) is larger than 0.005 and species-level effective population size is assumed to be several-fold larger than in D. melanogaster. Alternatively, the deleterious mutation rate may be smaller in D. simulans, or balancing selection may be acting nearby, thereby reducing the effect of background selection.

Introduction

Two models have been proposed to explain the well-documented correlation between rates of recombination and levels of neutral sequence variation in D. melanogaster (Begun and Aquadro 1992; Aquadro, Begun, and Kindahl 1994): the hitchhiking effects of adaptive substitutions (Maynard Smith and Haigh 1974; Kaplan, Hudson, and Langley 1989) and “background selection” against deleterious mutations (Charlesworth, Morgan, and Charlesworth 1993). Both of these phenomena are undoubtedly occurring simultaneously in nature, but which one plays a greater role in shaping patterns of molecular polymorphism? The answer to this question is currently of great interest, because of the very different picture of the evolutionary process that each one paints. If hitchhiking is largely responsible, new adaptive mutations must be arising and going to fixation quite frequently, since all regions of low recombination examined so far in D. melanogaster have reduced levels of variation (Hudson 1994). If, however, background selection is the major cause, a more strict neutralist view of evolution is supported, wherein most mutations are deleterious, many are neutral, and very few are advantageous.

Low variation in regions of reduced crossing over was initially attributed to hitchhiking effects (e.g., Begun and Aquadro 1991; Berry, Ajoika, and Kreitman 1991; Martin-Campos et al. 1992). Early theoretical formulations based on complete linkage (Charlesworth, Morgan, and Charlesworth 1993) suggested that background selection was not capable of causing the sorts of dramatic reductions in variation observed in D. melanogaster. However, recent theoretical work taking into account recombination (Hudson and Kaplan 1995; Nordborg, Charlesworth, and Charlesworth 1996) and the contribution of transposable elements to deleterious mutations (B. Charlesworth, personal communication) suggests that the effects of background selection may in fact be sufficient to explain the data. These theoretical developments have all been aimed at showing that background selection is strong enough to account for most of the relationship observed to date between neutral variation and recombination, even in regions of moderate to high recombination. Thus, background selection, rather than neutrality, might be considered the appropriate null hypothesis against which molecular polymorphism data should be tested. Since hitchhiking and background selection models now both predict a positive correlation between recombination and variation, the current challenge is to find ways to analyze our data that can distinguish between them. One approach to this problem is to try to better delimit the parameter space in which background selection operates.

We have surveyed DNA sequence variation at the glucose dehydrogenase (Gld) locus, at 84C on chromosome 3R, in D. simulans. A comparison of genetic map distance versus chromosomal band position along the third chromosome (fig. 1) shows that rates of recombination in the centromeric region including Gld are low in D. simulans, as in D. melanogaster. Our intention in carrying out this study was to compare the effects of low recombination on levels and patterns of silent and replacement polymorphism between D. simulans and D. melanogaster (data from D. melanogaster will be published separately). To our surprise, we observed quite high levels of nucleotide variation, with only limited evidence for a departure from the neutral expectation for the distribution of variation. We use genetic map data from the third chromosome of D. simulans to find the
expected effect of background selection at Gld, and to determine what values of \( sh \) produce predictions consistent with our observations.

Materials and Methods

Cloning and Sequencing of Gld from D. simulans

An EMBL3 (Stratagene) library of genomic DNA from an inbred isofemale line (CAS 54), partially digested with \( Sa_{u}3A \), was constructed and screened by standard methods (Sambrook, Fritsch, and Maniatis 1989). The library was screened with a 1.9-kb \( EcoRI/XbaI \) fragment containing exon 4 from D. melanogaster plasmid pSG9 (Cavener et al. 1986). One positive plaque was obtained. Two \( HindIII/EcoRI \) fragments containing all of the D. simulans exons 2, 3, and 4 were subcloned into pBluescript. Double-stranded plasmid DNA was sequenced using a Sequenase kit (USB) and custom-synthesized primers. Both strands were sequenced. The sequence has been entered into GenBank as accession numbers U63324 and U63325.

Population Sample

D. simulans females collected in 1984 in Raleigh, N.C. were used to establish inbred lines by 10 generations of sib mating (Aquadro, Lado, and Noon 1988). Genomic DNAs were prepared by CsCl gradient centrifugation. A 1.6-kb fragment of exon 4 was PCR-amplified using as primers:

4521: \( 5'\)-CTTAAACCCTTTACACAGGTTGG-3'
6157: \( 5'\)-TACTCGCTGTAATTTCGCTTG-3'.

The numbers refer to the 5' nucleotide position of the D. melanogaster sequence as reported in GenBank accession M29298. Nine hundred seventy base pairs (4600-5570) of DNA were sequenced from each of 11 inbred lines. PCR products were sequenced directly with Sequenase (USB) after digestion of one strand with lambda exonuclease (BRL)(Higuchi and Ochman 1989). Polymorphisms were scored by the method of Nachman et al.(1994). This approach is appropriate for regions where insertions and deletions are rarely found. Except in cases where sequences were ambiguous, DNAs were sequenced on only one strand.

In Situ Hybridization

In situ hybridization was performed according to Lim (1993), using as a probe biotinylated plasmid DNA corresponding to the 4.5-kb \( EcoRI/HindIII \) fragment of D. simulans Gld containing exon 4. The probe was detected with a Vectastain ABC kit from Vector Laboratories and 3,3'diaminobenzidine tetrahydrochloride from Sigma.

Estimation of Recombination Rate

Two molecularly characterized genes (labial and doublesex) flanking Gld were chosen to confirm low rates of recombination in the immediate vicinity of Gld. DNA was extracted from individual D. simulans by the procedure of Gloor et al. (1993) A 400-bp fragment spanning the intron of the labial (lab) gene was amplified using as primers:

1854: \( 5'\)-CACGCTTAATGAAACGCAGGTGAG-3'
1962: \( 5'\)-TGGAGTGCTGCGTCAGGATGTC-3'.

1268:
2288:

The numbers refer to the 5' nucleotide position of the D. melanogaster sequence as reported in Mlodzik, Fjose, and Gehring (1988). Size variation was scored by electrophoresis through 5% polyacrylamide TBE gels. A 1-kb fragment was amplified from the coding region of doublesex (dsx) using as primers:

1268: \( 5'\)-GAGACTGGGAATAGCGCACACGATG-3'
2288: \( 5'\)-TTGCGCCGTCTTTCGATGAC-3'.

The numbers refer to the 5' nucleotide position of the D. melanogaster sequence as reported in Burtis and Baker (1989). The PCR products were analyzed by digestion with Hae III (NEBiolabs) and electrophoresis through 6% polyacrylamide, or 2% agarose, TBE gels. Isofemale lines from a sample of D. simulans from Homosassa, Fla., collected in January 1994, were surveyed for variation at lab and dsx. Isofemale lines 44
Table 1

Divergence at GM between two randomly chosen alleles of
D. melanogaster and D. simulans

<table>
<thead>
<tr>
<th>Substitutions</th>
<th>No. of Sites</th>
<th>% Divergence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coding</td>
<td>5 replacement</td>
<td>1,564</td>
</tr>
<tr>
<td></td>
<td>49 silent</td>
<td>475</td>
</tr>
<tr>
<td>Intron</td>
<td>46</td>
<td>353</td>
</tr>
<tr>
<td>5’ and 3’ UTR</td>
<td>0</td>
<td>120</td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
<td>2,312</td>
</tr>
</tbody>
</table>

Amino Acid Substitutions

<table>
<thead>
<tr>
<th>Substitutions</th>
<th>Dayhoff Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asparagine to serine at aa 123</td>
<td>0</td>
</tr>
<tr>
<td>Aspartic acid to glutamic acid at aa 167</td>
<td>3</td>
</tr>
<tr>
<td>Isoleucine to valine at aa 286</td>
<td>4</td>
</tr>
<tr>
<td>Valine to alanine at aa 312</td>
<td>0</td>
</tr>
<tr>
<td>Glutamic acid to aspartic acid at aa 471</td>
<td>3</td>
</tr>
</tbody>
</table>

Abbreviations:
- GM: Gene Map
- CAS: Cambridge Alumni Society

a Corrected for multiple hits by the method of Jokes and Cantor (Li and Graur 1991) given in parentheses.

b From Dayhoff (1978). Numbers indicate whether a substitution is as common (0) or more common (>0) than would be expected by chance.

and 50 had near-fixed differences. Mating pairs were set up using females from isofemale line 44 and males from isofemale line 50. After 5 days, parents were removed and analyzed for their genotypes at lab and dsx. Two matings in which the parents were homozygous for different alleles at both loci were chosen. The females all had a smaller fragment at lab, and contained a Hae III site at dsx that was not present in the males. F1 progeny were collected from those matings. Females were held as virgins for 1-5 days prior to mating with their male siblings. Single pairs were placed in 10 ml vials and left for 5 days, at which time the F1 parents were removed and analyzed to confirm that they were both heterozygous at both lab and dsx. Throughout the experiment, vials were maintained at 25°C on a 12 hour light/12 hour dark cycle. All F2 progeny from each successful mating were collected. Genotypes of 404 F2 progeny were determined.

Results

Normal Levels of Divergence Between
D. melanogaster and D. simulans at Gld

One allele of D. simulans (CAS 54) Gld was cloned, and the entire protein coding region (the second, third and fourth exons), all of intron II, and small amounts of introns I and III and 3’ UTR were sequenced and compared to the sequence of the same region from the Oregon R strain of D. melanogaster (Krasney, Carr, and Cavener 1990). Results are summarized in Table 1. The coding regions show typical levels of divergence between these closely related species (Hudson et al. 1994). None of the five amino acid replacements is radical, suggesting no change in function or selective constraint in the two lineages.

Polymorphism at Gld in D. simulans

970 nucleotides of exon 4 were sequenced from 11 D. simulans chromosomes sampled from a population in Raleigh, N.C. Twenty-seven polymorphisms were observed (table 2), 26 of which are synonymous. Three nucleotides are segregating at position 5197. Variation at site 5510 causes a leucine (CUC) to isoleucine (AUC) replacement polymorphism. Estimates of nucleotide diversity (π and θ) for total sites at Gld in D. simulans are 0.011 and 0.009, respectively, about 15- to 20-fold higher than estimates of these parameters for Gld in a sample of D. melanogaster collected at the same location (unpublished data).

These levels of polymorphism were shown by HKA tests (Hudson, Kreitman, and Aguade 1987) of Gld versus per (Kliman and Hey 1993) and rosy (unpublished data) to be entirely consistent with the neutral expectation (P values greater than 0.90 and 0.75, respectively, table 3). Tajima’s test (Tajima 1989) indicates that the frequency spectrum of variants is also consistent with the neutral expectation (Tajima’s D = 0.972, not significantly different from zero). A departure from a neutral allele frequency spectrum is not predicted under

Table 2

Nucleotide Polymorphism in Exon 4 of Gld in D. simulans

<table>
<thead>
<tr>
<th>NUCLEOTIDE POSITION (KRASNEY, CARR, AND CAVENER 1990)</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
</tr>
<tr>
<td>15</td>
</tr>
<tr>
<td>54</td>
</tr>
<tr>
<td>60</td>
</tr>
<tr>
<td>34</td>
</tr>
<tr>
<td>18</td>
</tr>
<tr>
<td>39</td>
</tr>
<tr>
<td>63</td>
</tr>
<tr>
<td>65</td>
</tr>
</tbody>
</table>

Nom.-The polymorphism at site 5510 leads to a replacement (Leu to Ile).
Table 3
H.K.A. Tests of Neutrality of Gld in D. simulans, Based on Silent Sites Only

<table>
<thead>
<tr>
<th>Gene</th>
<th>Bases Surveyed and Compared</th>
<th>Polymorphic Within D. melanogaster and D. simulans</th>
<th>Differences Between D. melanogaster and D. simulans</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gld .......</td>
<td>11</td>
<td>230.7</td>
<td>26</td>
</tr>
<tr>
<td>per</td>
<td>6</td>
<td>385.6</td>
<td>40</td>
</tr>
<tr>
<td>x² for Gld vs. per = 0.05, P &gt; 0.9</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>boss</td>
<td>5</td>
<td>358.6</td>
<td>38</td>
</tr>
<tr>
<td>x² for Gld vs. boss = 0.48, P &gt; 0.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>rosy</td>
<td>8</td>
<td>321.0</td>
<td>51</td>
</tr>
<tr>
<td>x² for Gld vs. rosy = 0.01, P &gt; 0.9</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

There is evidence of only moderate levels of linkage disequilibrium at Gld in D. simulans. There are nine polymorphic sites at which the less common nucleotide is present in at least 4 of 11 alleles. Pairwise examination of these nine sites reveals that, in 24 of 36 pairs, all four genic types are present. Similarly, the various haplotypes are not related to each other in a simple stepwise series of mutational differences, suggesting extensive recombination in the history of the sampled alleles.

Cytological Localization of Gld and Rate of Recombination

Failure to observe the expected low variation in a region of low recombination led us to confirm the inferred cytological location of Gld in D. simulans by in situ hybridization. A single hybridizing band was observed at 84C, as in D. melanogaster, confirming the centromeric location of Gld in D. simulans. We then estimated the actual rate of recombination at Gld in flies from a natural population, using as markers molecular variants at loci flanking Gld: labial, at 84A, and doublesex, at 84E, 35 cytological bands apart. Isofemale lines from a Florida population of D. simulans were screened for variation at these loci (the Raleigh lines were no longer available). Size variation in the intron of lab, and a Hae III site polymorphism at dsx, were identified, and the appropriate crosses were set up as described in Materials and Methods. Of 404 F2 progeny scored, four were recombinants (table 4). These data give an estimate of 1.0 centimorgans (cM) across these 35 cytological bands, with a 95% confidence interval of 0.2 cM.

Discussion

The level of DNA sequence polymorphism at Gld in D. simulans is comparable to that found at genes in D. simulans that experience normal levels of recombination (Moriyama and Powell 1996). Estimates of \( \pi \) and \( \theta \) are 15 to 20-fold higher than those for Gld sampled from four D. melanogaster populations (unpublished data). These results are surprising, given our knowledge of recombination in this genomic region (fig. 1 and Results), and contrast with previous observations that other genes from regions of very reduced recombination in D. melanogaster also show a significant reduction in variability in D. simulans (Begun and Aquadro 1991; Berry, Ajoika, and Kreitman 1991; Moriyama and Powell 1996).

A possible explanation for our observations was that recombination in the Gld region in D. simulans might not actually be particularly low. Gld is close to the proximal breakpoint of the large segment of 3R that is inverted in D. melanogaster relative to D. simulans. Different reports of the breakpoints have placed Gld both within the inversion (84B3 to 92C, Ohnishi and Voelker 1979) and outside (84F1 to 93F6-7, Lemeunier, David, and Tsacas 1986). Inclusion of Gld in the inversion, and its consequent displacement to a region of high recombination away from the centromere, could have explained our unexpected results. The in situ hybridization, however, confirmed the location of Gld in D. simulans at 84C, identical to that in D. melanogaster.

Our initial qualitative assessment of recombination rate was based on a plot of genetic distance versus physical distance across the third chromosome, where we assumed that cytological locations of D. simulans genes are the same as in D. melanogaster, with the exception of the large inversion (fig. 1). In such a plot, the slope of the line reflects the rate of recombination. The slope of the line across the centromere is determined by only two loci: radius incompletus at 77E–F on 3L mapped at 58 cM, and ebony, at the proximal end of the inversion on 3R, mapped at 60 cM (Ohnishi and Voelker 1979). Two centimorgans over seven major cytological divisions containing 275 cytological bands gives an adjusted coefficient of exchange (ACE) of 0.0036, less than for the comparable region in D. melanogaster (ACE is cM per cytological band, divided by two to account for the lack of recombination in males). Because this estimate is based entirely on two mapped loci, and because mapping data based on marked laboratory stocks might not be representative of natural populations (Brooks 1988),
we made an independent estimate of recombination around \( Gld \) in \( D. \) \textit{simulans} from a natural population. Our estimate of \( 1.0 \) (\( \pm 0.5 \)) \(\text{cM} \) over 35 cytological bands is consistent with previous mapping data (Sturtevant 1929; Ohnishi and Voelker 1979) and the estimate of \( 0.6 \) \(\text{cM} \) over the same region in \( D. \) \textit{melanogaster} (Lindsley and Zimm 1992). Thus, recombination has been confirmed to be low.

\( GM \) is not the only example of high variation in this region. Ayala, Chang, and Harti (1993) found even higher variation in a sample of five \( Rh3 \) alleles in \( D. \) \textit{simulans} @ (silent) = 0.052). While \( Rh3 \) in \( D. \) \textit{melanogaster} is at band position 1540 in \( D. \) \textit{melanogaster}, a region of high recombination, it is close to the distal breakpoint of the fixed inversion, and hence in \( D. \) \textit{simulans} is at band 1071, just 75 bands distal to \( Gld \) (table 5). This result suggests that high variation at \( Gld \) may be a regional, rather than a locus-specific, phenomenon.

We failed to detect a departure from the expectations of a neutral model according to the HKA test, based on a contrast between variation within and between species, and Tajima’s test, based on a comparison of the number of segregating sites and pairwise differences. A recent episode of directional selection would have caused significantly reduced variation and possibly a skewed frequency distribution, and thus a negative Tajima’s \( D \) (Braverman et al. 1995; Simonsen, Churchill, and Aquadro 1995). Thus, in contrast to other studies of genes in regions of low recombination, we do not see evidence of a recent selective sweep at \( Gld \).

The \( Fu \) and \( Li \) test detected a deficiency of mutations in the external branches of the phylogeny of these \( Gld \) alleles (\( P < 0.025 \)), a pattern that may be produced by an old balanced polymorphism. The HKA and Tajima test results, however, do not suggest that a strong elevation of variation has occurred due to balancing selection. The issue of balancing selection is discussed further below.

Below we consider whether the high level of variation at \( Gld \) is consistent with a current model of background selection alone, which is expected to reduce neutral variation but not to cause a departure from a neutral allele frequency spectrum (Charlesworth, Charlesworth, and Morgan 1995). Current models of background selection (Hudson and Kaplan 1995; Nordborg, Charlesworth, and Charlesworth 1996) predict dramatic reductions of nucleotide variation in the centromeric regions of the autosomes of \( D. \) \textit{melanogaster}. We tested whether these models are consistent with our observations at \( Gld \) in \( D. \) \textit{simulans} by making a prediction for nucleotide diversity (\( \tau \)) at \( Gld \) using equation 15 of Hudson and Kaplan (1995):

\[
\tau = \pi_o \exp[\gamma G]
\]

where

\[
G = \sum_i \frac{u \cdot sh}{2} \frac{[x_i + 1 - x_i]}{[M(x_i + 1) - M(x_i)]^2},
\]

\[
\pi_o = \text{the level of neutral variation expected in the absence of background selection},
\]

\[
u = \text{deleterious mutation rate per band},
\]

\[
sh = \text{product of selection coefficient and dominance factor},
\]

\[
x_i = \text{cytological band position of locus \( i \)},
\]

\[
M(x_i) = \text{map position in morgans of locus \( i \) at position \( x \)}.
\]

Thirteen loci across the third chromosome, for which genetic and cytological positions are known (table 5), were used as the \( x_i \) in the above equation.

The background selection model of Hudson and Kaplan is intended to explain patterns of variation in \( D. \) \textit{melanogaster}, and their analysis employs estimates of the parameters \( sh \) and \( u \) that are based on empirical studies of this species. Because we had no comparable data from \( D. \) \textit{simulans}, we used the same values as Hudson and Kaplan: \( u = 0.0002 \) mutations per chromosomes band; \( sh = 0.005, 0.02, \) or 0.03. Hudson and Kaplan used \( \pi_o = 0.014 \) for \( D. \) \textit{melanogaster} because it produced a good fit to estimates of \( \tau \) generated by restriction site surveys of a mix of coding and noncoding regions. We have used a similar approach to choosing \( \pi_o \), based on DNA sequence polymorphism data available for a total of five loci on the \( D. \) \textit{simulans} third chromosome, including \( Gld \). Considering only silent sites, and using the intermediate value of \( sh = 0.02 \), we determined that \( \pi_o = 0.11 \) gives the best fit of all five observations to the model (fig. 2).

Predicted levels of \( \tau \), and the corresponding number of segregating sites (\( S \)) that would be expected, given \( \pi \), for 230.7 silent sites in 11 individuals, are shown in table 6. (Note that because background selection is expected to reduce \( N_e \) but not change the frequency distribution, the estimate of \( \pi \) should also be an estimate...
Our empirical observation of $S = 26$ is higher than any of the predictions. To test whether the number of segregating sites observed in our sample is significantly different from that predicted by the model, we used computer simulations as described in Hudson (1993) to generate a distribution of values of $S$ for each prediction of $\pi$. This method is conservative because it overestimates the variance of $S$ when there is some recombination within the locus. Because our hypothesis was that $S$ was too large to be consistent with the model selection, we found one-tailed probabilities by counting the number of times out of 1,000 trials that $S$ was 26 or greater. We rejected the null hypothesis of selection only when $sh = 0.005$ ($P < 0.001$). These tests suggest that this particular model of selection can explain variation at $Gld$ in $D. simulans$ as long as $\pi_o$ is very large and $sh$ is not very small. A lower value of $u$ could make the data consistent even if $sh$ were as small as 0.005.

If variation at $GM$ has been elevated by balancing selection, as suggested by the Fu and Li test, the data might then be fully consistent with the background selection model. To explore this possibility further, we conducted additional HKA tests where background selection is assumed in the model. How- ever, because the level of variation at $Gld$ is largely consistent with the predictions of the background selection model, we need not invoke a major effect of balancing selection to explain the data.

It is worth noting that the five gene regions on the third chromosome used to estimate $\pi_o$ all show higher-than-average variation: the average $\pi$ for these regions is 0.057, as opposed to an average of 0.031 for 12 gene regions across the genome (Moriyama and Powell 1996). Thus, the range of parameter values found to be consistent with the data and background selection may change as additional data become available.

**Conclusions**

Our analysis of neutral variation in the $Gld$ region of $D. simulans$ demonstrates (1) the complexity of the background selection process, such that single-gene comparisons are likely to be uninformative, and (2) the importance of knowing the extended recombinational environment of a gene for interpreting levels and patterns of polymorphism. The high levels of nucleotide variation at $Gld$ in $D. simulans$ appeared, on the surface, to be irreconcilable with the background selection model. However, the effects of background selection accumulate over large chromosomal regions, particularly if $sh$ is not small. Figure 1 shows that, while recombination is very low in $D. melanogaster$ from about band 600 to band 1300, the comparable region of centromeric repression of recombination in $D. simulans$ extends only from about band 750 to band 1100. Because this region

![Graph showing observed and predicted levels of DNA variation](https://example.com/graph.png)

**Table 6**

<table>
<thead>
<tr>
<th>Predicted Levels of $\pi$ and $S$ at the Silent Sites in Exon 4 of $Gld$ Under the Background Selection Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>$SH$</td>
</tr>
<tr>
<td>$\pi$</td>
</tr>
<tr>
<td>$S$</td>
</tr>
<tr>
<td>$P$</td>
</tr>
</tbody>
</table>

Note: 230.7 silent sites; observed $\theta = 0.0381$, $S = 26$, $n = 11$, $\pi_{out} = 0.11$, $u = 0.0002$. Tests are one-tailed.

Using the approach described above for $Gld$, we calculated the theoretical expectation for the ratio $\pi/\pi_o$ at $boss$ and $rosy$, the reference loci, by using genetic data available for the nearby loci $Ald$ and $Men$, respectively (see table 5). With a factor of $\pi/\pi_o (Gld) = (0.338)\pi/\pi_o (rosy)$, the $x^2$ statistic was 0.86 ($P > 0.25$). With a factor of $\pi/\pi_o (Gld) = (0.367)\pi/\pi_o (boss)$, the $x^2$ statistic was 0.053 ($P > 0.75$). Thus, while the Fu and Li test detects an unusual distribution of variation, we fail to find the excess of polymorphism expected under balancing selection, even when the effects of background selection are assumed in the model. However, because the level of variation at $Gld$ is largely consistent with the predictions of the background selection model, we need not invoke a major effect of balancing selection to explain the data.
is only half as big in *D. simulans*, we need not invoke any major differences in μ or sh between the species to explain a smaller effect of background selection in *D. simulans*. In fact, when μ, sh, and π₀ are held constant, expectations of π in *D. simulans* are almost twice those for *D. melanogaster*, due to differences in recombination alone (data not shown). Note that the assumption of equal values of π₀ is equivalent to an assumption of equal species-level effective population size (Nₑ). Knowledge of recombinational landscapes is therefore essential if we are to make inferences about species-level Nₑ based on average levels of nucleotide variation. For example, higher levels of polymorphism in *D. simulans* may be due, in part, to the fact that the total genetic map in *D. simulans* is longer, reflecting higher rates of recombination on average (Sturtevant 1929; Ohnishi and Voelker 1979; True, Mercer, and Laurie 1996). Similarly, *D. pseudoobscura* has one metacentric and three acrocentric chromosomes, rather than two metacentrics and one acrocentric as in *D. melanogaster*, and an overall larger total genetic map (at least 446 cm; Anderson 1990). Levels of DNA sequence variation are generally highest in *D. pseudoobscura* (Riley, Kaplan, and Veuille 1992; Schaeffer and Miller 1993), consistent with the prediction that smaller linkage groups lead to a reduced effect of background selection (Nordborg, Charlesworth, and Charlesworth 1996).

We have shown that, under a reasonable set of parameter values, high variation in a region of low recombination will be consistent with the effects of background selection alone. What remains to be seen is whether these same parameter values can simultaneously explain observations of very low variation in other regions of the *D. simulans* genome, where the lack of genetic map data currently precludes the sort of analysis presented here. The cause of the nonequilibrium frequency distribution detected by the Fu and Li test also deserves additional study.

**Acknowledgments**

We thank Doug Cavena for clones and sequencing primers; Brian Charlesworth and Dick Hudson for advice, discussion, and permission to use their unpublished work; Mike Ford for providing the program to simulate distributions of S; and Michael Nachman for discussion and comments on the manuscript. This work was supported by a grant from the National Institutes of Health (NIH) to C.F.A., and an NIH Predoctoral Traineeship to M.T.H.

**LITERATURE CITED**


———. 1994. How can the low levels of DNA sequence variation in regions of the *Drosophila* genome with low recombination be explained? PNAS USA 91: 6815-6818.


Jeffrey R. Powell, reviewing editor

Accepted June 28, 1996