Hitchhiking effects of advantageous mutations have been invoked to explain reduced polymorphism in regions of low crossing-over in *Drosophila*. Besides reducing DNA heterozygosity, hitchhiking effects should produce strong linkage disequilibrium and a frequency spectrum skewed toward an excess of rare polymorphisms (compared to the neutral expectation). We measured DNA polymorphism in a Zimbabwe population of *D. melanogaster* at three loci, *yellow*, *achaete*, and *suppressor of forked*, located in regions of reduced crossing-over. Similar to previously published surveys of these genomic regions in other populations, we observed low levels of nucleotide variability. However, the frequency spectrum was compatible with a neutral model, and there was abundant evidence for recombination in the history of the *yellow* and *ac* genes. Thus, some aspects of the data cannot be accounted for by a simple hitchhiking model. An alternative hypothesis, background selection, might be compatible with the observed patterns of linkage disequilibrium and the frequency spectrum. However, this model cannot account for the observed reduction in nucleotide heterozygosity. Thus, there is currently no satisfactory theoretical model for the data from the tip and base of the X chromosome in *D. melanogaster*.

Introduction

Studies of *Drosophila melanogaster* have revealed reduced DNA polymorphism in genomic regions experiencing low recombination rates, yet no reduction in sequence divergence between *D. melanogaster* and the sibling species, *D. simulans* (Begun and Aquadro 1991, 1993; Berry et al. 1991; Martin-Campos et al. 1992; Langley et al. 1993; Aquadro et al. 1994). These results have been interpreted in terms of the hitchhiking effect, a phenomenon whereby the substitution of advantageous mutants reduces levels of linked, neutral variation (Maynard Smith and Haigh 1974; Kaplan et al. 1989; Stephan et al. 1992).

Besides predicting reduced heterozygosity in genomic regions experiencing low rates of crossing-over, a simple hitchhiking model predicts that the frequency distribution of variation in these genomic regions will be skewed toward an excess of rare polymorphisms (Hudson 1990; Langley 1990). This second prediction is based on the notion that most observed mutations would be recent, having accumulated subsequent to the "selective sweep" which had removed most (if not all) of the neutral variation. One can think of this nonneutral distribution of polymorphism as resulting from the unusual history of such genomic regions. However, testing the theoretical prediction of a skewed frequency spectrum is complicated by the fact that population history can also affect the frequency spectrum.

We previously reported summary statistics of DNA polymorphism in a *D. melanogaster* population from Zimbabwe (Begun and Aquadro 1993). Given that the species is thought to have originated in and spread from Africa (David and Capy 1988), this unusually variable population may be more representative of an "ancestral" population and may be closer to mutation-drift equilibrium (though there is no guarantee this is the case). Thus, studies from this population could provide additional insight into the forces determining variation within and between natural *Drosophila* populations and different chromosomal regions. Here, we present the data from the *yellow* (*y*), *achaete* (*ac*), and *suppressor of forked* (*su[f]*) gene regions (all residing in regions of low crossing-over) in the Zimbabwe *D. melanogaster* population. *Yellow* and *ac* are located at the tip of the X chromosome and are about 10 kb apart, while *su(f)* is at the base of the X chromosome. We use estimates of the frequency spectrum, linkage disequilibrium, and population differentiation in these gene regions to eval-
uate population genetic models purporting to explain
the observed reduction in DNA heterozygosity in regions
of low crossing-over in the Zimbabwe population and
other populations of *D. melanogaster*.

**Material and Methods**

**Samples and Restriction Mapping**

We used the previously described set of *X* chromo-
somes (*n* = 50; Begun and Aquadro 1993). High
resolution four-cutter analysis was carried out as pre-
viously described (Kreitman and Aguadé 1986; Begun
and Aquadro 1993) using 10 four-cutter restriction en-
*Sau* 3A I, *Scr* FI, and *Taq* I. The probe for *yellow* was a
4.5-kb *BamHI*/*Bgl* II fragment starting at nucleotide
position 192 of Geyer et al. (1986); this is a slightly
smaller region than was probed by Martin-Campos et
al. (1992). All comparisons between populations for the
*yellow* region include only data from the region spanned
by the 4.5-kb *BamHI*/*Bgl* II fragment. The probe for
*achaete* was a 2.2-kb *EcoRI* fragment from position 1
to 2232 of Villares and Cabrera (1987); the probe for
*su(f)* was a 6.4-kb *BamHI*/*Xba* I (Langley et al. 1993;
Mitchelson et al. 1993). Coordinates of restriction sites
follow Geyer et al. (1986), Villares and Cabrera (1987),
and the GenBank submission for *y*, *ac*, and *su(f)*,
respectively.

**Analysis**

Nucleotide heterozygosity was calculated as de-
scribed elsewhere (Nei and Li 1979; Hudson 1982).
Linkage disequilibrium was estimated by *D'* (Lewontin
1964) and tested for significance by Fisher's exact test.

The HKA test (Hudson et al. 1987) tests the null
hypothesis that the ratio of polymorphism to divergence
at two (or more) independent loci is compatible with
single underlying values of *N* and *μ*. Our HKA tests
used *ac* and *su(f)* restriction site data from Zimbabwe
and previously published sequence divergence estimates
to *Drosophila simulans* (Martin-Campos et al. 1992;
Langley et al. 1993). The effective number of nucleotides
surveyed (Hudson 1982) for HKA tests in Zimbabwe
*D. melanogaster* were 536 and 1,092 for *ac* and *su(f)*,
respectively (some restriction sites were omitted for *su(f)*
because of insertion/deletion variation between *D. mel-
anogaster* and *D. simulans*; we used a *su(f)* alignment
provided by C. Langley). Estimates of divergence (dif-
fferences/bases surveyed) were 117/2,174 and 412/3,741
for *ac* and *su(f)*, respectively. Data from each locus were
compared to DNA sequence data from a random sample
(*n* = 11) of the 5′-flanking region of the *X*-linked *ver-
milion* gene in Zimbabwe. The number of sites surveyed
and segregating sites at *vermilion* were 535 and 30, re-
spectively. The number of differences between a randomly
selected allele from *D. melanogaster* and *D. simulans*
were 42 (there is no evidence that this region has been
influenced by selection in Africa; D. Begun and C.
Aquadro, unpublished data).

Tajima proposed a test of the neutral, equilibrium
model based on the idea that the parameter *4N* *μ* (3*N* *μ*
in the case of *X*-linked genes) can be estimated from the
number of segregating sites or from the number of pair-
wise differences. The difference between these two esti-
mators is expected to be zero under a neutral, equilib-
rium model (Tajima 1989). Tajima's test statistic, *D*,
will be negative or positive if there is an excess or deficit
of rare polymorphisms, respectively. Tajima's (1989)
test was carried out using restriction site polymorphisms
(indel variation not included).

The null hypothesis that samples from different
demographic locations were from a single, panmictic pop-
ulation was tested using permutation-based methods
(1,000 trials; Hudson et al. 1992a; Roff and Bentzen
1992). Estimates of *F* *ST* were carried out as described
by Hudson et al. (1992b) with intrapopulation hetero-
zygosity weighted by sample size. A cladogram of *su(f)*
haplotypes was constructed using PAUP (Swofford

**Results**

**Polymorphism**

We scored 102, 68, and 171 restriction sites in *y*,
*ac*, and *su(f)*, respectively. A summary of polymorphic
restriction sites and insertion/deletion variants is shown
in table 1. Estimates of nucleotide heterozygosity are
presented in table 2.

**Statistical Tests of Neutrality**

The *χ* *2* values in HKA tests for *ac* and *su(f)* com-
pared to *vermilion* were 8.39 and 27.69, respectively.
Both comparisons reject the null hypothesis of neutral,
equilibrium evolution (1 degree of freedom; *P* < 0.005).
However, the Tajima (1989) *D* values for *y*, *ac*, and
*su(f)* (table 3) were not significantly different from zero,
the expectation under a neutral, equilibrium model.

**Linkage Disequilibrium**

In table 4 we show estimates of the linkage dis-
equilibrium parameter, *D*', for the *y*-*ac* region, including
only restriction sites with frequency greater than 0.1.
Overall, only 3 of 10 pairwise comparisons were signif-
ificant. Four of six pairwise comparisons between *y* and
*ac* show four gametic types indicating that crossing-over,
gene conversion, or parallel mutation has occurred dur-
Table 1

Restriction Map Variation in the yellow, achaete, and su(f) Gene Regions in a Zimbabwe Sample of Drosophila melanogaster

<table>
<thead>
<tr>
<th>Site</th>
<th>Line</th>
</tr>
</thead>
<tbody>
<tr>
<td>y:</td>
<td>000000011111111112222222223333333344444444444555555556</td>
</tr>
<tr>
<td></td>
<td>13567901235678012345678901234569012345678903567890</td>
</tr>
<tr>
<td>Hha 345</td>
<td></td>
</tr>
<tr>
<td>Rsa 665-668</td>
<td></td>
</tr>
<tr>
<td>Taq 1247-1250</td>
<td></td>
</tr>
<tr>
<td>Hin 1314-1318</td>
<td></td>
</tr>
<tr>
<td>Hin 1901-1905</td>
<td></td>
</tr>
<tr>
<td>Alu 2124-2127</td>
<td></td>
</tr>
<tr>
<td>Sau 2256</td>
<td></td>
</tr>
<tr>
<td>Hae 3097-3100</td>
<td></td>
</tr>
<tr>
<td>del 571-806 (84 bp)</td>
<td></td>
</tr>
<tr>
<td>ins 1547-1639 (400 bp)</td>
<td></td>
</tr>
<tr>
<td>del 2929-3004 (38 bp)</td>
<td></td>
</tr>
<tr>
<td>ins 3004-3130 (125 bp)</td>
<td></td>
</tr>
<tr>
<td>ac:</td>
<td></td>
</tr>
<tr>
<td>Sau 1553</td>
<td></td>
</tr>
<tr>
<td>Alu 2147</td>
<td></td>
</tr>
<tr>
<td>del 150-250 (2 bp)</td>
<td></td>
</tr>
<tr>
<td>del 1855-1936 (13 bp)</td>
<td></td>
</tr>
<tr>
<td>su(f):</td>
<td></td>
</tr>
<tr>
<td>Hae 3586-3589</td>
<td></td>
</tr>
<tr>
<td>Hha 3791</td>
<td></td>
</tr>
<tr>
<td>Dde 5793</td>
<td></td>
</tr>
<tr>
<td>Hin 7217</td>
<td></td>
</tr>
<tr>
<td>Hae 7538-7541</td>
<td></td>
</tr>
<tr>
<td>Hae 7678-7681</td>
<td></td>
</tr>
<tr>
<td>ins 1319-2287</td>
<td></td>
</tr>
<tr>
<td>ins 2398-3076</td>
<td></td>
</tr>
<tr>
<td>del 3417-3525 (5 bp)</td>
<td></td>
</tr>
<tr>
<td>del 6016-6051 (2 bp)</td>
<td></td>
</tr>
<tr>
<td>del 7827-7918 (10 bp)</td>
<td></td>
</tr>
</tbody>
</table>

**NOTE.**—Mutations are indicated by an interval of nucleotides for the loss of a site and a single nucleotide for the gain of a site. Size variants are localized to the indicated interval and grouped for each gene region following the restriction sites. A question mark indicates an unscoreable site. The following sites were polymorphic in Europe or the United States and could have been scored in Zimbabwe given the probe-enzyme combinations used: y Hae III 4442-4445, ac TagI 36, ac TaqI 2147, su(f) Ddel 2242, su(f) TaqI 2372 (misidentified as TaqI 974 in Langley et al. 1993; C. Langley, personal communication). The location of su(f) HinfI 7217 is uncertain—it may be at position 2318. su(f) ins 1319-2287 and 2398-3076 are large insertions, the sizes of which cannot be determined.

ing the history of these sequences. For the su(f) region there are also several comparisons with four gametic types; however, omission of line 58 from the analysis eliminates all cases of four gametic types. This haplotype cannot be explained by a single recombination event among other haplotypes present in our sample. Nor would it appear that the line 58 haplotype (number 18 in table 5) results from a single recombination event among intermediates absent from our sample, since the remaining su(f) haplotypes can each be connected by one mutation in a single, most parsimonious tree (fig. 1).

Geographic Variation

Table 5 shows the geographic distribution of four-cutter haplotypes. Ten polymorphic restriction sites in y-ac could have been scored in our survey of Zimbabwe and in previously surveyed samples from Europe and the United States (Martin-Campos et al. 1992) given the enzymes and probes used in the surveys; only 2 of the 10 were observed in both surveys and in previously published results from a U.S. sample (Langley et al. 1993). Similarly, comparison of our data to previously published results from a U.S. sample (Langley et al. 1993) reveals that none of the eight polymorphic restriction sites scoreable at su(f) were segregating in both geographic regions. Most insertion/deletion variation at y, ac, and su(f) is not shared between samples. We used the data from table 5 for tests of population subdivision (Hudson et al. 1992a). The Zimbabwe sample was significantly different from the U.S. samples...
at all three gene regions for all the test statistics (P < 0.001). Previous four-cutter data from U.S. and European samples revealed that these two geographic regions share the same major haplotypes at y-ac but show differences in frequency of minor haplotypes (Martín-Campos et al. 1992). For statistical comparisons involving the European sample, we used the method of Roff and Bentzen (1992) since our computer did not have sufficient memory to execute the Hudson et al. (1992a) program on the large sample. The European sample was significantly different from both the Zimbabwe and U.S. samples (P < 0.01).

Levels of differentiation between Zimbabwe and Europe or U.S. populations are much greater than those observed between the U.S. and Europe groups. Restriction sites y HaeIII 4442–4445, ac TaqI 36, and su(f) Ddel 2242, show nearly fixed differences between the Zimbabwe and Europe/U.S. samples (table 5). Estimates of FST (Hudson et al. 1992b) between Zimbabwe and U.S. samples for y, ac, and su(f) are 0.56, 0.54, and 0.60, respectively, while FST values of four other X-linked loci from regions of "normal" recombination range from 0.25 to 0.32 (Begun and Aquadro 1993).

Heterogeneous Tajima D values across populations at both y-ac and su(f) suggest that the frequency spectra may also be different in different populations, though none of the Tajima D values are significantly different from zero (table 3). As was seen in Zimbabwe, there is no obvious skewness toward rare sites in the y-ac and su(f) sample from the United States.

The “ancestral” Drosophila melanogaster restriction map haplotype for ac and su(f) (as inferred from the D. simulans sequence) occurs at higher frequencies in Zimbabwe than in other samples (table 5; ancestral states could not be inferred for y because there is no sequence from D. simulans). This, along with the higher levels of variability in the Zimbabwe sample (Begun and Aquadro 1993) and the fact that Zimbabwe su(f) haplotypes appear centrally in the phylogenetic network (fig. 1), support the notion that the Zimbabwe population is older and historically larger than the U.S. population.

Discussion

There are four important results from the analysis of four-cutter restriction sites at y-ac and su(f) in Drosophila melanogaster from Zimbabwe (this report; Begun and Aquadro 1993). First, levels of variability are severely reduced compared to levels of variability in genes experiencing higher recombination rates, such as

### Table 3

<table>
<thead>
<tr>
<th></th>
<th>y-ac</th>
<th>su(f)</th>
<th>Pooled</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zimbabwe</td>
<td>-0.416</td>
<td>0.338</td>
<td>-0.133</td>
</tr>
<tr>
<td>U.S.</td>
<td>1.069</td>
<td>-0.842</td>
<td>0.399</td>
</tr>
<tr>
<td>Europe</td>
<td>-1.536</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

Note.—Data from y and ac from Europe and the United States are from Martin-Campos et al. (1992); su(f) data from the United States are from Langley et al. (1993).

### Table 4

<table>
<thead>
<tr>
<th></th>
<th>y</th>
<th>ac</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alu 2147</td>
<td>Rsa 665</td>
<td>Hin 1314</td>
</tr>
<tr>
<td>Sau 1553</td>
<td>-1.00***</td>
<td>0.03</td>
</tr>
<tr>
<td>Alu 2147</td>
<td>...</td>
<td>-0.40</td>
</tr>
<tr>
<td>Rsa 665–668</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Hin 1314–1318</td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

Note.—Only sites with frequency >0.1 are included.

* P < 0.05.

*** P < 0.001.
Table 5
Geographic Variation in Haplotype Frequencies between *Drosophila melanogaster* Populations at the *y, ac*, and *su(f)* loci

<table>
<thead>
<tr>
<th>Y</th>
<th>Taq 1247</th>
<th>Alu 2124</th>
<th>Sau 2256</th>
<th>Sau 2491</th>
<th>Hae 3097</th>
<th>Hae 4442</th>
<th>Eur.</th>
<th>U.S.</th>
<th>Zim.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>−</td>
<td>+</td>
<td>+</td>
<td>184</td>
<td>37</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>−</td>
<td>+</td>
<td>−</td>
<td>2</td>
<td>4</td>
<td>26</td>
</tr>
<tr>
<td>3</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>+</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>+</td>
<td>−</td>
<td>−</td>
<td>2</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>1</td>
<td>0</td>
<td>21</td>
</tr>
<tr>
<td>6</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>−</td>
<td>+</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>7</td>
<td>+</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>+</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>8</td>
<td>−</td>
<td>+</td>
<td>−</td>
<td>−</td>
<td>+</td>
<td>−</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>2</td>
<td>7</td>
<td>9</td>
</tr>
<tr>
<td>10</td>
<td>+</td>
<td>−</td>
<td>+</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>2</td>
<td>0</td>
<td>35</td>
</tr>
<tr>
<td>12</td>
<td>−</td>
<td>+</td>
<td>−</td>
<td>186</td>
<td>40</td>
<td>0</td>
</tr>
<tr>
<td>Sim.</td>
<td>+</td>
<td>−</td>
<td>+</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Su(f)</th>
<th>Dde 2242</th>
<th>Taq 2372</th>
<th>Hae 3586</th>
<th>Hha 3791</th>
<th>Dde 5793</th>
<th>Hien 7217</th>
<th>Hae 7538</th>
<th>Hae 7678</th>
<th>U.S.</th>
<th>Zim.</th>
</tr>
</thead>
<tbody>
<tr>
<td>13</td>
<td>−</td>
<td>−</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>+</td>
<td>+</td>
<td>0</td>
<td>11</td>
</tr>
<tr>
<td>14</td>
<td>−</td>
<td>−</td>
<td>+</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>+</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>15</td>
<td>−</td>
<td>−</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>−</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>17</td>
<td>−</td>
<td>−</td>
<td>+</td>
<td>−</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>18</td>
<td>−</td>
<td>−</td>
<td>+</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>+</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>19</td>
<td>−</td>
<td>−</td>
<td>+</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>+</td>
<td>+</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>20</td>
<td>−</td>
<td>−</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>+</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>21</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>−</td>
<td>−</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>22</td>
<td>+</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>−</td>
<td>+</td>
<td>+</td>
<td>41</td>
<td>0</td>
</tr>
</tbody>
</table>

**NOTE.**—Only sites which were scoreable in Zimbabwe, the United States, and Europe are included. For the Zimbabwe sample, the number of individuals does not always sum to 50 because some individuals were not scored for all polymorphisms. Data from *y* and *ac* from Europe and the United States are from Martin-Campos et al. (1992); *su(f)* data from the United States are from Langley et al. (1993). Sim., the ancestral state as inferred from the published *D. simulans* sequence (Martin-Campos et al. 1992; Langley et al. 1993).

**vermilion** and **white** (Begun and Aquadro 1993). Second, the frequency distribution of polymorphic sites is compatible with that expected under neutrality. Third, there is evidence for considerable recombination among polymorphisms, especially at *y-ac*. Finally, there are nearly fixed differences between Zimbabwe and other surveyed populations at both *y-ac* and *su(f)*. We will discuss these observations in turn, asking whether all can be subsumed under a single theoretical model.

The neutral model of molecular evolution predicts a positive correlation between heterozygosity within species and divergence between species (Kimura 1983). Under this model, reduced polymorphism should be accompanied by reduced divergence. We can reject strict neutrality since there is no reduction in DNA sequence divergence between species at *ac* and *su(f)*.

A second model for reduced variation within species is background selection (Charlesworth et al. 1993). Although the model predicts reduced polymorphism in regions of reduced crossing-over, the frequency spectrum is not expected to depart from that expected under neutrality. Thus, estimates of the frequency spectrum at the tip and base of the *X* chromosome in Zimbabwe are consistent with this model. However, as currently formulated, the model predicts that heterozygosity at the tip and base of the *X* chromosome should be reduced below the level predicted under strict neutrality only by about 4% and 24%, respectively. Heterozygosities for *ac* (tip) and *su(f)* (base) are about 90% lower than those observed for **vermilion** and **white** (Begun and Aquadro 1993); the observed heterozygosities are incompatible with background selection over the parameter space deemed plausible by Charlesworth et al. (1993).

A third model for the observed patterns of polymorphism and divergence is the hitchhiking model (Kaplan et al. 1989). Under this model, newly arising,
strongly favored mutations sweep through populations, causing severely reduced heterozygosity at linked neutral sites. Heterozygosity and divergence in regions of low crossing-over can be readily accounted for by this model (Kaplan et al. 1989; Stephan et al. 1992). The simple hitchhiking model leads to another testable prediction: in gene regions with greatly reduced numbers of segregating sites, most polymorphisms should be "new," having occurred subsequent to selective sweeps. These polymorphisms are expected to be rare because new variants in large populations take a very long time to drift to intermediate frequencies. In the language of coalescent models, an excess of rare polymorphisms is expected because the topologies of gene genealogies following recent selective sweeps resemble "star" phylogenies. Mutations in such genealogies are expected to appear as singletons in random samples of genes (Aguadé et al. 1989; Hudson 1990; Langley 1990).

Genes at the tip and base of the X chromosome in Zimbabwe showed no skew toward an excess of rare polymorphisms (i.e., Tajima's $D$ was not significantly different from zero). A similar result was previously obtained from a study of the tip of the X chromosome in a U.S. population of $D. melanogaster$ (Martin-Campos et al. 1992; Aguadé et al. 1994). On the other hand, reduced heterozygosity in some $D. melanogaster$ samples from Spain was accompanied by a significant excess of rare sites (Martin-Campos et al. 1992). A simple explanation for the different results is that the data sets vary in their power to reject the neutral model. This issue has recently been addressed. Simulation studies of a simple, strong selection hitchhiking model provide quantitative support for the notion that gene regions showing reductions of heterozygosity similar to those observed in our data from Zimbabwe (and in data from other low-recombination loci in other populations) should have a significantly negative Tajima's $D$ (Braverman et al., in press).

Could population histories (e.g., expansions, contractions, founder effects) be confounding the predictions of the simple hitchhiking model, vis-à-vis the frequency spectrum? While they cannot be ruled out entirely, such phenomena alone are probably an insufficient explanation for the discordance between theory and data. Bottlenecks should cause a genome-wide loss of rare variants, yet in the Zimbabwe population, gene regions experiencing moderate to high rates of crossing-over show a greater trend (albeit, not significant) toward rare polymorphisms than do genes at the tip or base of the X chromosome (Begun and Aquadro 1993). This suggests that a recent bottleneck in the history of the Zimbabwe population is an unlikely explanation for the lack of excess rare polymorphism at $y$-ac and $su(f)$. The very different Tajima test results in Spain and Raleigh at the tip of the X chromosome (Martin-Campos et al. 1992) are also difficult to explain by simple demographic events, as we would then expect to observe a similar pattern in other gene regions. Data from the white locus in these same two populations showed that the frequency spectra were almost identical and in good agreement with that predicted under neutrality (Miyashita et al. 1993). Thus, the expected skew in areas of low crossing-over is not generally observable, and the lack of agreement between theory and data cannot be readily explained by a lack of power of the tests or by the history of $D. melanogaster$ populations.

The simple arguments which lead to a prediction of a skewed frequency spectrum should also result in a prediction of high linkage disequilibrium. Though there are no quantitative predictions for linkage disequilibrium under a hitchhiking model, $D'$ was significant in only 3 of 10 pairwise comparisons at $y$-ac in Zimbabwe. Fur-

![Fig. 1. —Hypothesized evolutionary relationships among $su(f)$ restriction site haplotypes from population samples of $Drosophila melanogaster$ from Zimbabwe and the United States. Zimbabwe line 58 was excluded from the analysis (see text).](https://academic.oup.com/mbe/article-abstract/12/3/382/979927/.../07February2019)
thermore, there are many cases of four gametic types in
the data from y-ac (and su/fj), observations which are
incompatible with only one or two cross-overs. There
has been considerable exchange of mutations among
chromosomes in these regions of very low crossing-over.
How can we explain this pattern? First, crossing-over at
the tip (or base) of the X chromosome in Zimbabwe
might be considerably higher than in stocks used for
generic mapping experiments. This might be a conse-
quency of inversion heterozygosity in Africa (Lemeunier
and Aulard 1992) leading to increased crossing-over at
the tip of the X chromosome (the Schultz-Redfield
[1951] or interchromosomal effect). Other genetic or
environmental factors affecting recombination rates
could also be involved (summarized in Ashburner
1989). Second, the decay of linkage disequilibrium by
gene conversion could be important; although crossing-
over is suppressed at the tip and base of the X chro-
mosome, we are unaware of any studies demonstrating
that gene conversion is similarly reduced. Finally, rel-
atively low levels of linkage disequilibrium might reflect
the fact that the polymorphisms in Zimbabwe are not
very recent (consistent with the frequency spectrum
data). One possible interpretation is that small regions
of neutral polymorphism escape the hitchhiking effect
via conversion rather than crossing-over. So while overall
levels of variability are severely reduced, “patches” of
old polymorphisms could persist in the face of selective
events at linked sites. Theoretical and empirical studies
of the role of gene conversion in regions of low crossing-
over are required to address these issues.

The remaining unusual aspect of the data are the
major differences in allele frequencies between Zim-
babwe and U.S./Europe populations in regions of re-
duced crossing-over. One explanation for such a pattern
differential selection in different environments (geo-
graphically restricted hitchhiking effects; see, e.g., Ste-
phan and Mitchell 1992). An alternative hypothesis is
that some unconditionally beneficial mutations are re-
cent enough so as to have spread only through part of
the species range. However, the chance of observing such
a phenomenon must be very small given the rapid spread
of such mutants. Differentiation between the United
States and Europe is significant but not so great as that
seen between Zimbabwe and other populations. For ex-
ample, table 5 shows that most of the difference between
U.S. and European populations at y can be explained
by a haplotype which occurs at very low frequency in
Europe but at intermediate frequency in the United
States (this also accounts for the greater skew toward
rare sites in the European populations). Interestingly,
there is some evidence for heterogeneity at the tip of the
X chromosome even between different U.S. samples of
D. melanogaster (Aguadé et al. 1989; Eanes et al. 1989;
Perhaps this heterogeneity is a consequence of transient
selection (e.g., fluctuating selection coefficients; cf. Gil-
lespie 1991). Gillespie (1994) has simulated one model
with fluctuating selection (the TIM model) and found
that heterozygosity can be reduced well below the neutral
level without a concomitant significant skew in the fre-
quency spectrum.

In summary, we can say with some certainty that
neither the simplest neutral model nor the simplest
hitchhiking model can accommodate the data from the
tip and base of the X chromosome. That leaves us with
interesting data for which we have no good explanation
(see also Charlesworth 1994). What type of positive se-
lection models might serve as alternatives to simple
hitchhiking? The simple model assumes that favorable
mutants start at very low frequency and that a neutral
locus cannot be influenced by multiple sweeps simulta-
aneously. Perhaps models relaxing these assumptions
could explain the data. Models with fluctuating selection
coefficients (Gillespie 1994) are also worth further study.
What are the theoretical predictions for linkage dis-
equilibrium in different types of positive selection mod-
els? How important is gene conversion at the tip and
base of the X chromosome? Could balanced polymor-
phisms in regions of reduced crossing-over play an im-
portant role in the distribution of linked, neutral vari-
ation? The role of “negative” (i.e., background) selection
(Charlesworth et al. 1993) also requires further explora-
tion. As noted earlier, the observed reduction in het-
erozygosity appears to be greater than that predicted by
the background selection model. It is unclear, however,
how sensitive this prediction is to assumptions about the
number and distribution of selection coefficients of de-
leterious mutations. It will also be interesting to see
quantitative predictions of the frequency spectrum,
lkage disequilibrium, and population differentiation
under background selection.

Acknowledgments

We thank G. Karpen and K. O’Harc for clones, M.
Aguadé and C. Langley for information allowing data
to be directly compared between laboratories, and R.
Hudson for population subdivision programs. We thank
M. Nachman, J. Gillespie, and especially J. Braverman
and C. Langley for useful discussion. This work was
supported by a grant from the National Institutes of
Health (GM 36431) to C.F.A., a National Science
Foundation Doctoral Dissertation Improvement Grant
(BSR-9112480) to D.J.B. and C.F.A., and a National
Institutes of Health predoctoral traineeship to D.J.B.
LITERATURE CITED


THOMAS H. EICKBUSH, reviewing editor

Received June 20, 1994

Accepted November 23, 1994