Evolution at the Tip and Base of the *X* Chromosome in an African Population of *Drosophila melanogaster*

David J. Begun and Charles F. Aquadro
Section of Genetics and Development, Cornell University

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 gene 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-

Key words: *Drosophila melanogaster, yellow, achaete, suppressor of forked,* hitchhiking effects, frequency spectrum, DNA polymorphism.

Address for correspondence and reprints: David J. Begun, Section of Evolution and Ecology, Center for Population Biology, University of California, Davis, CA 95616. E-mail: djbegun@ucdavis.edu

Mol. Biol. Evol. 12(3):382–390. 1995.

0737-4038/95/1203-0003$02.00

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* chromosomes (n = 50; Begun and Aquadro 1993). High resolution four-cutter analysis was carried out as previously described (Kreitman and Aguadé 1986; Begun and Aquadro 1993) using 10 four-cutter restriction enzymes: *Alu*I, *Dde*I, *Hae*III, *Hha*I, *Hin*fI, *Msp*I, *Rsa*I, *Sau*3AI, *Scr*FI, and *Taq*I. The probe for *yellow* was a 4.5-kb *Bam*HI/*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 *Bam*HI/*Bgl*II fragment. The probe for *achaete* was a 2.2-kb *Eco*RI fragment from position 1 to 2232 of Villares and Cabrera (1987); the probe for *su(f)* was a 6.4-kb *Bam*HI/*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 described 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. melanogaster* and *D. simulans;* we used a *su[f]* alignment provided by C. Langley). Estimates of divergence (differences/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 *vermilion* gene in Zimbabwe. The number of sites surveyed and segregating sites at *vermilion* were 535 and 30, respectively. 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 $4N\mu$ ($3N\mu$ in the case of *X*-linked genes) can be estimated from the number of segregating sites or from the number of pairwise differences. The difference between these two estimators is expected to be zero under a neutral, equilibrium 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 geographic locations were from a single, panmictic population was tested using permutation-based methods (1,000 trials; Hudson et al. 1992*a;* Roff and Bentzen 1992). Estimates of F_{ST} were carried out as described by Hudson et al. (1992*b*) with intrapopulation heterozygosity weighted by sample size. A cladogram of *su(f)* haplotypes was constructed using PAUP (Swofford 1991).

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)* compared 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 disequilibrium 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 significant. 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*

Site	Line
	0000001111111122222222223333333334444444444555555556 1356790123567801234567890123456901234567890356789 0
y:	
Hha 345	------+--+---------+----------+---?----------
Rsa 665–668	?++++-+++++++++-+++++++++++++++-+++++-+++-+--++
Taq 1247–1250	++++++++++++++++++++++++++++++-++++++++++++++++
Hin 1314–1318	-++++-++-+++++++++-+++++++++-++++++++++++++++++
Hin 1901–1905	++++++-++-+++++++++-+++++++++-++++++++++++++++++
Alu 2124–2127	++-+++++++++
Sau 2256	--+-------
Hae 3097–3100	++-+--++-++---+-+--+-++-+--+++++++---++++++-++--
del 571–806 (84 bp)	-----+--
ins 1547–1639 (400 bp)	--+----------
del 2929–3004 (38 bp)	------------------------+-----------------------
ins 3004–3130 (125 bp)	------------------------+-----------------------
ac:	
Sau 1553	+++++++++++++-++++++-++-++++++-++++++++--++++++++
Alu 2147	+----------+-+-+-+--+--+------+---+---+++-+--++--
del 150–250 (2 bp)	--+----+---+-+-+-++----+-+++++--+++-++---+--++-+++
del 1855–1936 (13 bp)	----+--------------+-------------------------
su(f):	
Hae 3586–3589	++-+++-++++++-+++++--++-+++++?-++++-+++-+-+--++++-
Hha 3791	----------------------+-------------------+--
Dde 5793	++-+++-++-+---+-+++--++-++++++---++--++------++-+-
Hin 7217	-----+---------------+-------------------------
Hae 7538–7541	-++-+++-++++++-+-+++++-+-++--?+++-+++-+++++++++-+++
Hae 7678–7681	-++++++++++++++++++++++++++++?++++++++++++++++-++
ins 1319–2287	---+---+--------+-----+-+----------------------
ins 2398–3076	--------+-------------------------------------
del 3417–3525 (5 bp)	----------------------+----+----------------
del 6016–6051 (2 bp)	--+--++--+-+++-+---+++-+------++--++--+++-+++---+
del 7827–7918 (10 bp)	+--

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 Sau* 3AI 2491, *y Hae*III 3097–3100, *y Hae*III 4442–4445, *ac Taq*I 36, *ac Taq*I 2147, *su(f) Dde*I 2242, *su(f) Taq*I 2372 (misidentified as *Taq*I 974 in Langley et al. 1993; C. Langley, personal communication). The location of *su(f) Hinf*I 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 (*y Hae* 3097–3100 and *ac Alu* 2147). 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. 1992*a*). 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 (Martin-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. (1992*a*) 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 Hae*III 4442–4445, *ac Taq*I 36, and *su(f) Dde*I 2242, show nearly fixed differences between the Zimbabwe and Europe/U.S. samples (table 5). Estimates of F_{ST} (Hudson et al. 1992*b*) between Zimbabwe and U.S. samples for *y, ac,* and *su(f)* are 0.56, 0.54, and 0.60, respectively, while F_{ST} 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.

Table 2
Nucleotide Heterozygosity in the *y, ac,* and *su(f)* Gene Regions in *Drosophila melanogaster* Samples from Zimbabwe, the United States, and Europe

	$\hat{\theta}$	$\hat{\pi}$
y:		
Zimbabwe	0.0026	0.0017
U.S.	0.0010	0.0011
Europe	0.0011	0.0002
ac:		
Zimbabwe	0.0010	0.0012
U.S.	0.0011	0.0011
Europe	0.0010	0.0006
su(f):		
Zimbabwe	0.0011	0.0011
U.S.	0.0005	0.0002

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). The size of the probed region for *y* was slightly different for the Zimbabwe versus the U.S./European sample. Estimates of heterozygosity for *y* in the United States and Europe were inferred from the sites that Martin-Campos et al. (1992) could have scored over the same region where the Zimbabwe sample was probed. Heterozygosity for *ac* in Europe is for 50 lines from Barcelona which were surveyed with eight restriction enzymes.

Table 3
Tajima *D* Statistics for Restriction Site Variation in *y, ac,* and *su(f)* Samples from Zimbabwe, the United States, and Europe

	Zimbabwe	U.S.	Europe
y-ac	−0.416	1.069	−1.536
su(f)	0.338	−0.842	...
Pooled	−0.133	0.399	...

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).

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 4
Linkage Disequilibrium (D') in the *y-ac* Region in a Sample of *Drosophila melanogaster* from Zimbabwe

	ac	*y*		
	Alu 2147	*Rsa* 665	*Hin* 1314	*Hae* 3097
ac:				
Sau 1553	−1.00***	0.03	−1.00	−1.00*
Alu 2147	...	−0.40	0.33	0.52
y:				
Rsa 665–668	...	...	−1.00***	−0.67
Hin 1314–1318	...	...	...	−1.00

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

y	*Taq* 1247	*Alu* 2124	*Sau* 2256	*Sau* 2491	*Hae* 3097	*Hae* 4442	Eur.	U.S.	Zim.
1	+	+	−	−	+	+	184	37	0
2	+	+	−	−	+	−	2	4	26
3	+	+	−	−	−	+	1	0	0
4	+	+	−	+	+	−	2	6	0
5	+	+	−	−	−	−	1	0	21
6	+	+	+	−	+	−	0	0	1
7	+	−	−	−	+	−	0	0	1
8	−	+	−	−	+	−	0	0	1

ac	*Taq* 36	*Sau* 1553	*Alu* 2147	Eur.	U.S.	Zim.
9	+	+	+	2	7	0
10	+	−	+	0	0	6
11	+	+	−	2	0	35
12	−	+	−	186	40	0
Sim.	+	−	+			

su(f)	*Dde* 2242	*Taq* 2372	*Hae* 3586	*Hha* 3791	*Dde* 5793	*Hin* 7217	*Hae* 7538	*Hae* 7678	U.S.	Zim.
13	−	−	+	−	+	−	+	+	0	11
14	−	−	+	−	+	−	−	+	0	11
15	−	−	+	−	−	−	+	+	4	0
16	−	−	−	−	−	−	+	+	0	13
17	−	−	+	+	+	−	+	+	0	1
18	−	−	+	+	−	−	+	−	0	1
19	−	−	+	−	+	+	+	+	0	12
20	−	−	+	−	+	−	−	−	0	7
21	+	+	+	−	−	−	+	+	2	0
22	+	−	+	−	−	−	+	+	41	0
Sim.	−	−	+	?	?	?	?	?		

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 indivduals 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-

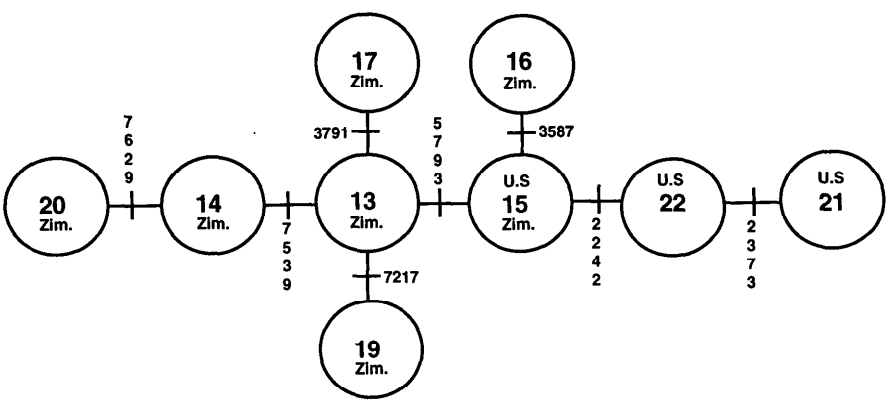

FIG. 1.—Hypothesized evolutionary relationships among *su(f)* restriction site haplotypes from population samples of *Drosophila melanogaster* from Zimbabwe and the United States. Zimbabwe line 58 was excluded from the analysis (see text).

thermore, there are many cases of four gametic types in the data from *y-ac* (and *su[f]*), 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 genetic mapping experiments. This might be a consequence 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* chromosome, we are unaware of any studies demonstrating that gene conversion is similarly reduced. Finally, relatively 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 Zimbabwe and U.S./Europe populations in regions of reduced crossing-over. One explanation for such a pattern is differential selection in different environments (geographically restricted hitchhiking effects; see, e.g., Stephan and Mitchell 1992). An alternative hypothesis is that some unconditionally beneficial mutations are recent 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 example, 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; MacPherson et al. 1990; Begun and Aquadro 1991). Perhaps this heterogeneity is a consequence of transient selection (e.g., fluctuating selection coefficients; cf. Gillespie 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 frequency 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 selection 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 simultaneously. 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 disequilibrium in different types of positive selection models? How important is gene conversion at the tip and base of the *X* chromosome? Could balanced polymorphisms in regions of reduced crossing-over play an important role in the distribution of linked, neutral variation? The role of "negative" (i.e., background) selection (Charlesworth et al. 1993) also requires further exploration. As noted earlier, the observed reduction in heterozygosity 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 deleterious mutations. It will also be interesting to see quantitative predictions of the frequency spectrum, linkage disequilibrium, and population differentiation under background selection.

Acknowledgments

We thank G. Karpen and K. O'Hare 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

AGUADÉ, M., W. MEYERS, A. D. LONG, and C. H. LANGLEY. 1994. Single-stranded conformation polymorphism analysis coupled with stratified DNA sequencing reveals reduced sequence variation in the *su(s)* and *su(w^a)* regions of the *Drosophila melanogaster X* chromosome. Proc. Natl. Acad. Sci. USA **91**:4658–4662.

AGUADÉ, M., N. MIYASHITA, and C. H. LANGLEY. 1989. Reduced variation in the *yellow-achaete-scute* region in natural populations of *Drosophila melanogaster*. Genetics **122**:607–615.

AQUADRO, C. F., D. J. BEGUN, and E. C. KINDAHL. 1994. Selection, recombination and DNA polymorphism in Drosophila. Pp. 46–56 *in* B. GOLDING, ed. Non-neutral evolution: theories and molecular data. Chapman & Hall, New York.

ASHBURNER, M. 1989. *Drosophila:* a laboratory handbook. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

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-achaete* region. Genetics **129**:1147–1158.

———. 1993. African and North American populations of *Drosophila melanogaster* are very different at the DNA level. Nature **356**:519–520.

BERRY, A. J., J. W. AJIOKA, and M. KREITMAN. 1991. Lack of polymorphism on the *Drosophila* fourth chromosome resulting from selection. Genetics **129**:1111–1117.

BRAVERMAN, J. M., R. HUDSON, N. L. KAPLAN, C. H. LANGLEY, and W. STEPHAN. 1995. The hitchhiking effect on the site frequency spectrum of DNA polymorphisms. Genetics (in press).

CHARLESWORTH, B. 1994. The effect of background selection against deleterious mutations on weakly selected, linked variants. Genet. Res. Camb. **63**:213–227.

CHARLESWORTH, B., M. T. MORGAN, and D. CHARLESWORTH. 1993. The effect of deleterious mutations on neutral molecular variation. Genetics **134**:1289–1303.

DAVID, J. R., and P. CAPY. 1988. Genetic variation of *Drosophila melanogaster* natural populations. Trends Genet. **4**:106–111.

EANES, W. F., J. LABATE, and J. W. AJIOKA. 1989. Restriction-map variation in the *yellow-achaete-scute* region in five populations of *Drosophila melanogaster*. Mol. Biol. Evol. **6**:492–502.

GEYER, P. K., C. SPANA, and V. G. CORCES. 1986. On the molecular mechanism of gypsy-induced mutations at the *yellow* locus of *Drosophila melanogaster*. EMBO J. **5**:2657–2662.

GILLESPIE, J. H. 1991. The causes of molecular evolution. Cambridge University Press, Cambridge.

———. 1994. Alternatives to neutral models. Pp. 1–17 *in* B. Golding, ed. Non-neutral evolution: theories and molecular data. Chapman & Hall, New York.

HUDSON, R. R. 1982. Estimating genetic variability with restriction endonucleases. Genetics **100**:711–719.

———. 1990. Gene genealogies and the coalescent process. Oxf. Surv. Evol. Biol. **7**:1–44.

HUDSON, R. R., D. D. BOOS, and N. L. KAPLAN. 1992. A statistical test for detecting geographic subdivision. Mol. Biol. Evol **9**:138–151.

HUDSON, R. R., M. KREITMAN, and M. AGUADÉ. 1987. A test of neutral molecular evolution based on nucleotide data. Genetics **116**:153–159.

HUDSON, R. R., M. SLATKIN, and W. P. MADDISON. 1992. Estimation of levels of gene flow from DNA sequence data. Genetics **132**:583–589.

KAPLAN, N. L., R. R. HUDSON, and C. H. LANGLEY. 1989. The "hitchhiking effect" revisited. Genetics **123**:887–899.

KIMURA, M. 1983. The neutral theory of molecular evolution. Oxford University Press, London.

KREITMAN, M., and M. AGUADÉ. 1986. Genetic uniformity in two populations of *Drosophila melanogaster* revealed by filter hybridization of four-nucleotide-recognizing restriction enzyme digests. Proc. Natl. Acad. Sci. USA. **83**:3562–3566.

LANGLEY, C. H. 1990. The molecular population genetics of Drosophila. Pp. 75–91 *in* N. TAKAHATA and J. F. CROW, eds. Population biology of genes and molecules. Baifukan, Tokyo.

LANGLEY, C. H., J. MACDONALD, N. MIYASHITA, and M. AGUADÉ. 1993. Lack of correlation between interspecific divergence and intraspecific polymorphism at the suppressor of forked region in *Drosophila melanogaster* and *Drosophila simulans*. Proc. Natl. Acad. Sci. USA **90**:1800–1803.

LEMEUNIER, F., and S. AULARD. 1992. Inversion polymorphism in *Drosophila melanogaster*. Pp. 339–405 *in* C. B. KRIMBAS and J. R. POWELL, eds. Drosophila inversion polymorphism. CRC Press, Boca Raton, Fla.

LEWONTIN, R. C. 1964. The interaction of selection and linkage. I. General considerations; heterotic models. Genetics **49**:49–67.

MACPHERSON, J. N., B. S. WEIR, and A. J. LEIGH BROWN. 1990. Extensive linkage disequilibrium in the *achaete-scute* complex of *Drosophila melanogaster*. Genetics **126**:121–129.

MARTIN-CAMPOS, J. M., J. M. CAMERON, N. MIYASHITA, and M. AGUADÉ. 1992. Intraspecific and interspecific variation at the *y-ac-sc* region of *Drosophila simulans* and *Drosophila melanogaster*. Genetics **130**:805–816.

MAYNARD SMITH, J., and J. HAIGH. 1974. The hitch-hiking effect of a favorable gene. Genet. Res. **23**:23–35.

MITCHELSON, A., M. SIMONELIG, C. WILLIAMS, and K. O'HARE. 1993. Homology with *Saccharomyces cerevisiae* RNA14 suggests that phenotypic suppression in *Drosophila melanogaster* by *suppressor of forked* occurs at the level of RNA stability. Genes Devel. **7**:241–249.

MIYASHITA, N. T., M. AGUADÉ, and C. H. LANGLEY. 1993. Linkage disequilibrium in the *white* locus region of *Drosophila melanogaster*. Genet. Res. Camb. **62**:101–109.

NEI, M., and W.-H. LI. 1979. Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc. Natl. Acad. Sci. USA **76**:5269–5273.

ROFF, D. A., and P. BENTZEN. 1989. The statistical analysis of mitochondrial DNA polymorphisms: χ^2 and the problem of small samples. Mol. Biol. Evol. **6**:539–545.

SCHULTZ, J., and H. REDFIELD. 1951. Interchromosomal effects on crossing over in Drosophila, Cold Spring Harbor Symp. Quant. Biol. **16**:175–197.

STEPHAN, W., and S. J. MITCHELL. 1992. Reduced levels of DNA polymorphism and fixed between-population differences in the centromeric region of *Drosophila ananassae.* Genetics **132**:1039–1045.

STEPHAN, W., T. H. E. WIEHE, and M. W. LENZ. 1992. The effects of strongly selected substitutions on neutral polymorphism: analytical results based on diffusion theory. Theor. Pop. Biol. **41**:237–254.

SWOFFORD, D. L. 1991. PAUP: phylogenetic analysis using parsimony, version 3.0s. Illinois Natural History Survey, Champaign.

TAJIMA, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics **123**:585–595.

VILLARES, R., and C. V. CABRERA. 1987. The *achaete-scute* complex of *D. melanogaster:* conserved domains in a subset of gene required for neurogenesis and their homology to myc. Cell **50**:415–424.

THOMAS H. EICKBUSH, reviewing editor

Received June 20, 1994

Accepted November 23, 1994