Reduced Levels of DNA Polymorphism and Fixed Between-Population Differences in the Centromeric Region of *Drosophila ananassae*

Wolfgang Stephan and Samuel J. Mitchell

Department of Zoology, University of Maryland, College Park, Maryland 20742

Manuscript received May 11, 1992
Accepted for publication August 7, 1992

ABSTRACT

We have estimated DNA sequence variation within and between two populations of *Drosophila ananassae*, using six-cutter restriction site variation at *vermilion* (*v*) and *furrowed* (*fw*). These two gene regions are located close to the centromere on the left and right *X* chromosome arms, respectively. In the *fw* region, no DNA polymorphism was detected within each population. In the *v* region, average heterozygosity per nucleotide was very low in both populations ($\pi = 0.0005$ in the Burma population, and 0.0009 in the India population). These estimates are significantly lower than those from loci in more distal gene regions. The distribution of DNA polymorphisms between both populations was also striking. At *fw*, three fixed differences between the Burma and India populations were detected (two restriction site differences and one insertion/deletion of approximately 2 kb). At *v*, each DNA polymorphism in high frequency in the total sample was nearly fixed in one or the other population, although none of them reached complete fixation. The observed pattern of reduced variation within populations and fixed differences between populations appears to correlate with recombination rate. We conclude that recent hitchhiking associated with directional selection is the best explanation for this pattern. The data indicate that different selective sweeps have occurred in the two populations. The possible role of genetic hitchhiking in rapid population differentiation in gene regions of restricted recombination is discussed.

WHEN a selectively favored mutation occurs in a population and is subsequently fixed, the frequencies of polymorphisms at linked loci will be altered. This phenomenon, which is commonly referred to as the hitchhiking effect, was first analyzed by MAYNARD SMITH and HAIGH (1974). STEPHAN and LANGLEY (1989) and AGUADÉ, MIYASHITA and LANGLEY (1989) used this effect as an indirect means of detecting the action of directional selection at the DNA level. They observed reduced levels of molecular variation in chromosomal regions in which the rate of meiotic crossing over is severely restricted, *i.e.*, near centromeres and telomeres. Their results are in qualitative agreement with equilibrium hitchhiking models (KAPLAN, HUDSON and LANGLEY 1989; STEPHAN, WIEHE and LENZ 1992) which have been developed based on MAYNARD SMITH and HAIGH's (1974) suggestion.

The basic observation of reduced polymorphism in regions of restricted recombination has been contested by several authors (BEECH and LEIGH BROWN 1989; EANES, LABATE and AJIOKA 1989; MACPHERSON, WEIR and LEIGH BROWN 1990), mainly because the estimates of average nucleotide heterozygosity vary a great deal between population samples (ranging from zero to almost neutral equilibrium levels). The observed differences in nucleotide heterozygosity between populations may be attributed to some extent to the experimental sampling procedures (KREITMAN 1991), but the stochasticity associated with the evolutionary process appears to be a more important factor in regions of reduced recombination. Analyses of the hitchhiking models show that the reduction of variation is sensitive to levels of recombination, the frequency of selected substitutions and the intensity of selection. Selective sweeps, as postulated by these models, occur very quickly relative to neutral evolution. Thus, the variable levels of nucleotide diversity may just reflect the underlying discrete stochastic nature of hitchhiking scenarios, in which selected substitutions wipe out polymorphism in a population in a collective fashion and after which neutral evolution responds slowly. Under such circumstances, the distribution of nucleotide polymorphisms in regions of reduced recombination may strongly depend on the time elapsed since the last selective sweep and hence may be far away from equilibrium.

Recent population surveys have confirmed the original observation of reduced variation in regions of restricted recombination. Using interspecific comparisons and statistical techniques, such as the HKA test (HUDSON, KREITMAN and AGUADÉ 1987), surveys of several gene regions in *Drosophila simulans* and *Drosophila melanogaster* have found that the reduction of variation below the neutral level is statistically significant (BEGUN and AQUADRO 1991; BERRY, AJIOKA and

KREITMAN 1991; MARTÍN-CAMPOS *et al.* 1992). Furthermore, BEGUN and AQUADRO (1992) have demonstrated that levels of variation correlate with recombination rate among 20 gene regions from across the genome of *D. melanogaster*, suggesting that hitchhiking occurs over a large fraction of the Drosophila genome.

Our paper looks at another aspect of genetic hitchhiking, the effect of population subdivision on genetic variation in regions of reduced recombination. Hitchhiking can be expected to cause a great deal of heterogeneity between subpopulations in gene regions of restricted recombination, if different selective sweeps occur in different populations and migration rates are low. If so, hitchhiking may lead to rapid differentiation between subpopulations for genes located in regions of low recombination. To investigate this possibility, we surveyed two gene regions which are located near the centromere of the *X* chromosome in *Drosophila ananassae*, *furrowed* (*fw*) and *vermilion* (*v*). *fw* maps to the junction between heterochromatin and euchromatin on right arm of the *X* chromosome in region 14AB, and *v* to the first major polytene band on the left arm at 13Cp (STEPHAN and LANGLEY 1989; TANDA *et al.* 1989). *D. ananassae* was chosen for study because it shows significant population substructuring. The picture that emerges from analyses of isozyme polymorphism (summarized by JOHNSON 1971) and DNA polymorphism (DA LAGE, CARIOU and DAVID 1989; STEPHAN and LANGLEY 1989; LYNCH and CREASE 1990) is that *D. ananassae* exists in many isolated to semi-isolated populations. It is largely tropical, but has been found on all continents (PATTERSON and STONE 1952; BOCK and WHEELER 1972; MORIWAKI and TOBARI 1975). Its zoogeographical center is thought to be in Southeast Asia (DOBZHANSKY and DREYFUS 1943).

MATERIALS AND METHODS

Strains: Thirty-nine *X* chromosome lines, that originated from two different collections [19 lines from Mandalay (Burma) and 20 lines from Hyderabad (India)], were used in this study. The construction of these lines was described in STEPHAN and LANGLEY (1989). They are isogenic in the centromeric region.

Restriction map analysis: High molecular weight genomic DNA from each of the 39 lines was prepared as described by BINGHAM, LEVIS and RUBIN (1981). For the study of *fw* the DNA was digested with ten hexanucleotide-recognizing restriction enzymes (*Bam*HI, *Bgl*II, *Eco*RI, *Hin*dIII, *Pst*I, *Pvu*II, *Sac*I, *Sal*I, *Xba*I and *Xho*I). Electrophoresis, blotting and nick-translation of probes were carried out according to standard techniques (SAMBROOK, FRITSCH and MANIATIS 1989). The probes used in this experiment were *furrowed Eco*RI subclones derived from the overlapping phage clones λ106, λ1A, λ50A and λ104A. The *fw* subclones span the entire region of approximately 22 kb shown in Figure 2. These subclones were kindly provided by L. LESHKO, S. TANDA and V. CORCES. We used subclones instead of the original phage clones for probing, because the phage clones contained repetitive DNA. For filters which were probed with subclones free of repetitive DNA, washes were done under standard conditions (0.1 × SSC, 0.1% SDS at 50° for twice 30 min). For subclones which contained repetitive DNA, final washes were carried out at higher stringency (0.1 × SSC, 0.5% SDS at 58° for twice 45 min).

The *vermilion* gene region was surveyed with the three enzymes *Bgl*II, *Xba*I and *Xho*I, and the filters were probed using the same subclone as in STEPHAN and LANGLEY (1989). Thus, including the previous *v* data by STEPHAN and LANGLEY (1989), both *v* and *fw* regions were surveyed with the same set of 10 six-cutter enzymes.

Statistical procedures: The levels of nucleotide variation were estimated using NEI and TAJIMA's (1981) measure for average heterozygosity at the nucleotide level (π) and HUDSON's (1982) estimator for $4N_e\mu$ ($3N_e\mu$ for *X*-linked genes), where N_e is the effective population size and μ is the mutation rate per nucleotide site per generation. The latter quantity is expected to be identical with π, if nucleotide variation is neutral and the population is in equilibrium with respect to mutation and drift (NEI 1987, Chap. 10). Differences between these estimates of average nucleotide heterozygosity indicate that forces other than neutral mutation and genetic drift have to be invoked to explain the data. The statistical significance of such differences was examined with TAJIMA's (1989) test. A positive sign of TAJIMA's *D* statistic indicates balancing selection, a negative one purifying or directional selection. To compare levels of heterozygosity between different loci ("heterogeneity"), methods of TAVARÉ (1984) and HUDSON (1990) were used. Differences in the levels of nucleotide heterozygosity can be attributed to differences in mutation rate between loci and/or to the action of natural selection.

RESULTS

Restriction map variants from both populations and each gene region are presented in Table 1. Summary restriction maps for *v* amd *fw* are shown in Figure 1 and 2, respectively. The results of the heterogeneity tests are in Table 2.

***vermilion* gene region:** Including the data set of the previous study by STEPHAN and LANGLEY (1989), we scored a total of 35 restriction sites in both population samples. Two of these sites were polymorphic in each population. Based on this enlarged data set, we obtained the following estimates for $3N_e\mu$: 0.0017 and 0.0016 for the Burma and India populations, respectively, while estimates for π were 0.0005 and 0.0009. The larger deviation between the estimates of $3N_e\mu$ and π in the Burma population sample is due to the low frequencies of the polymorphic restriction sites. Both polymorphisms occur only once in the sample of 19 lines. Estimates of $3N_e\mu$ and π are closer for the India sample, because the frequencies of both polymorphic sites are higher (2/20). To quantify this, TAJIMA's (1989) *D* statistic was used. Including only restriction sites, we obtained values of $D = -1.51$ and -0.77 for the Burma and India populations, respectively. These results were not significant. Hence, we

TABLE 1

Restriction map variants in the *vermilion* and *furrowed* regions of *D. ananassae*

Line	*vermilion*													*furrowed*		
	Ins (*a*)	*Ins* (*b*)	*Ins* (*c*)	*Ins* (*d*)	*Ins* (*e*)	*Ins* (*f*)	*Pvu*II 5.1	*Pst*I 5.2	*Eco*RI 7.1	*Ins* (*g*)	*Ins* (*h*)–(*l*)	*Bam*HI 10.6	*Del* (*m*)	*Ins*	*Hin*dIII 7.0	*Sal*I 9.8
M61	–	–	–	–	–	–	–	–	–	+	–	–	–	+	–	+
M62	–	–	–	–	–	–	–	–	–	+	–	–	–	+	–	+
M66	–	–	–	–	–	–	–	–	–	+	–	–	–	+	–	+
M68	–	–	–	–	–	–	–	–	–	–	–	–	–	+	–	+
M79	–	–	–	–	–	–	–	–	–	–	–	–	–	+	–	+
M80	–	–	–	–	–	–	–	–	–	–	–	–	–	+	–	+
M86	–	–	–	–	–	–	–	–	–	–	–	–	–	+	–	+
M89	+	–	–	–	–	–	–	–	–	–	–	–	–	+	–	+
M90	–	–	–	–	–	–	–	–	–	–	–	–	–	+	–	+
M91	–	–	–	–	–	–	–	–	–	+	–	–	–	+	–	+
M92	–	–	–	–	–	–	–	–	–	–	–	–	–	+	–	+
M97	–	–	–	–	–	–	–	+	–	–	–	–	–	+	–	+
M99	–	–	–	–	–	–	–	–	–	–	–	–	–	+	–	+
M106	–	–	–	–	–	–	–	–	–	–	–	–	–	+	–	+
M111	–	–	–	–	–	–	–	–	–	–	–	–	–	+	–	+
M114	–	–	–	–	–	–	–	–	–	+	–	–	–	+	–	+
M115	–	–	–	–	–	–	–	–	–	+	–	–	–	+	–	+
M117	–	–	–	–	–	–	–	–	–	–	0.6	–	–	+	–	+
M119	–	–	–	–	–	–	+	–	–	–	1.2	–	–	+	–	+
H3	–	–	–	–	–	+	–	–	–	–	–	+	–	–	+	–
H4	–	–	–	–	–	+	–	–	–	–	0.1	+	–	–	+	–
H9	–	–	–	–	+	–	–	–	–	–	0.1	+	–	–	+	–
H10	–	–	+	–	–	+	–	–	–	–	0.1	+	–	–	+	–
H11	–	–	–	–	–	+	–	–	–	–	–	+	–	–	+	–
H12	–	–	+	–	–	+	–	–	+	–	0.9	+	–	–	+	–
H13	–	–	–	+	–	+	–	–	–	–	0.1	+	–	–	+	–
H15	–	+	–	–	–	+	–	–	–	–	0.1	+	+	–	+	–
H16	–	–	–	–	–	+	–	–	–	–	–	+	–	–	+	–
H21	–	–	–	–	–	+	–	–	–	–	0.1	+	–	–	+	–
H23	–	–	–	–	–	+	–	–	–	–	0.1	+	–	–	+	–
H26	–	–	–	–	–	+	–	–	–	–	–	+	–	–	+	–
H27	–	–	–	–	–	+	–	–	–	–	0.1	+	–	–	+	–
H31	–	–	–	–	–	+	–	–	–	–	1.5	+	–	–	+	–
H36	–	–	+	–	–	+	–	–	+	–	–	+	–	–	+	–
H39	–	–	–	–	–	+	–	–	–	–	–	+	–	–	+	–
H41	–	–	–	–	–	+	–	–	–	–	0.1	+	–	–	+	–
H47	–	–	–	–	–	–	–	–	–	+	–	–	–	–	+	–
H48	–	–	–	–	–	+	–	–	–	–	0.1	+	–	–	+	–
H50	–	–	–	–	–	–	–	–	–	–	–	–	–	–	+	–

Key: "+", present; "–", absent; M, Burma population; H, India population. *Ins*(*h*)–(*l*) in the *vermilion* region was scored as an increase in size (in kb) relative to a variant "–."

have no evidence to reject a neutral model for the *v* locus based on our restriction site data.

The spectrum of DNA polymorphisms which occur in high frequency in the total sample is also very interesting. Table 1 shows that if a insertion/deletion or a restriction site is present (absent) in one line of a sample, it is also present (absent) in almost all of the other lines of this sample (*e.g.*, *Ins*(*f*), *Ins*(*h*)-(*l*), and the *Bam*HI-site at position 10.6). This indicates that all DNA polymorphisms in high frequency in the total sample are nearly fixed in one population or the other. This pattern is even more characteristic of the *fw* region.

***furrowed* gene region:** The results for *fw* can be summarized in a very simple way. Using 10 restriction enzymes, 45 sites were scored scattered over a region of approximately 22 kb. Forty-three of these 45 sites were monomorphic and shared by both populations. The remaining two occurred as fixed differences between the two populations. Only one large insertion/deletion (approximately 2 kb) was found, occurring also as a fixed difference; *i.e.*, it was present in all lines from the Burma population, but absent in the India population sample. Because of the lack of within-population polymorphism, a TAJIMA test can not be done on the *fw* data.

Heterogeneity tests: We found very low levels of nucleotide heterozygosity in the *fw* and *v* regions in both populations. They are roughly a factor ten lower than the estimates for the other loci of *D. ananassae*

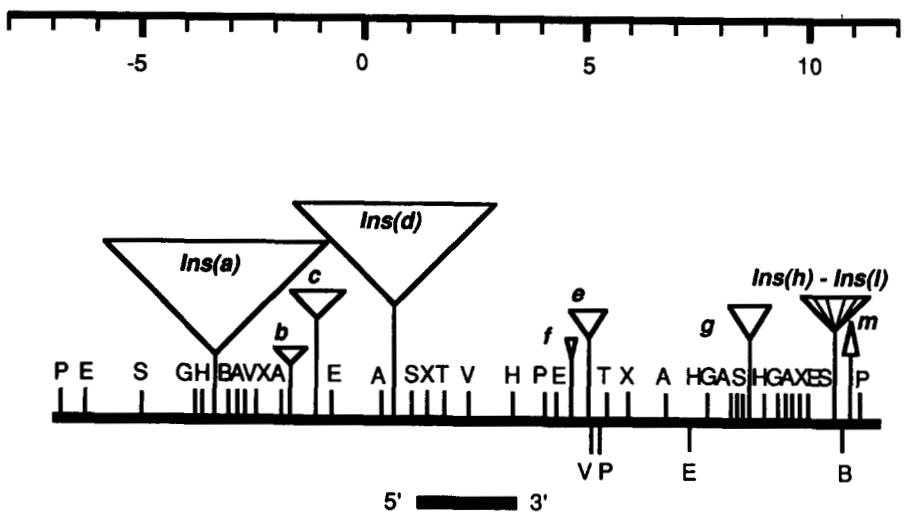

FIGURE 1.—Summary restriction map of DNA sequence variation in the *v* region of *D. ananassae*. This map includes the data by STEPHAN and LANGLEY (1989). The approximate location of the *v* transcriptional unit is indicated by a solid bar below the line. It was determined based on information available from *D. melanogaster* (SEARLES and VOELKER 1986; PASTINK *et al.* 1989; STEPHAN and LANGLEY 1989). The variant and invariant restriction sites are indicated below and above the line, respectively. In the survey, we used the restriction enzymes *Bam*HI (B), *Bgl*II (G), *Eco*RI (E), *Hin*dIII (H), *Pst*I (P), *Pvu*II (V), *Sac*I (T), *Sal*I (S), *Xba*I (A) and *Xho*I (X). Among the polymorphic restriction sites, V and P are unique variants in the Burma population, and E and B vary twice in the India population. Insertion and deletion sizes were estimated as follows: *Ins(a)* = 5 kb, *Ins(b)* = 0.7 kb, *Ins(c)* = 1.2 kb, *Ins(d)* > 4.5 kb, *Ins(e)* = 0.8 kb, *Ins(f)* = 0.2 kb, *Ins(g)* = 0.9 kb, *Ins(h)* = 1.2 kb, *Ins(i)* = 0.6 kb, *Ins(j)* = 0.1 kb, *Ins(k)* = 0.9 kb, *Ins(l)* = 1.5 kb, *Del(m)* = 0.3 kb. The insertions, *Ins(h)* to *Ins(l)*, found in the interval (9.8, 10.9), are lumped together into one symbol. Positions of insertions are identified only to the indicated restriction fragment. The 5′ and 3′ flanking *Sac*I sites and 3′ flanking *Pvu*II site have not been scored because they are too distant to unambiguously interpret variant restriction fragments. The subclone of the previous study (STEPHAN and LANGLEY 1989), ranging from coordinate −3.5 to 9.0, was used for probing. The coordinates adopted from SEARLES and VOELKER (1986) are distances in kilobases.

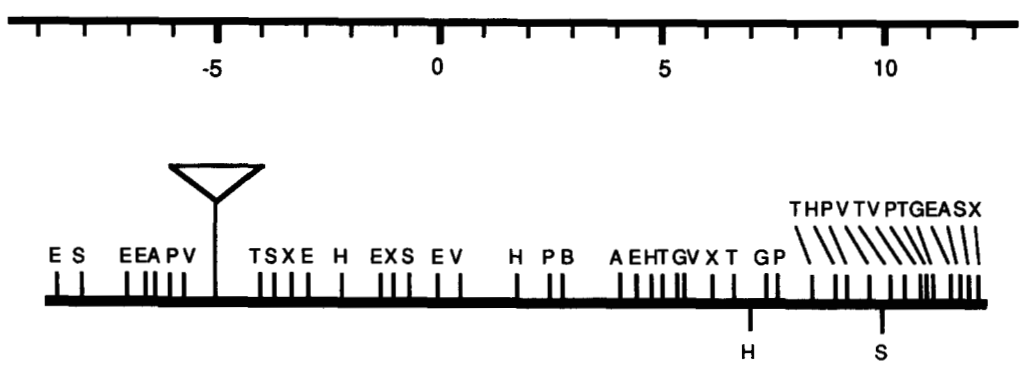

FIGURE 2.—Summary restriction map of DNA sequence variation in the *fw* region of *D. ananassae*. The location of the *fw* transcriptional unit is not exactly known. It appears to overlap with the region from −10.0 to 0 (L. LESHKO, S. TANDA and V. CORCES, personal communication). The size of the insertion was estimated as 2.0 kb. This insertion is "fixed" in the Burma population, but absent in the India population. Similarly, the restriction sites H and S (below the line) are alternatively "fixed" or absent in the India and Burma populations. The subclones used for probing (see MATERIALS AND METHODS) covered the entire region.

which have been surveyed, *forked* and *Om(1D)* (STEPHAN and LANGLEY 1989; STEPHAN 1989). In order to examine whether these between-loci differences are significant, we used a method proposed by HUDSON (1990) which is based on the known distribution of polymorphic sites under a neutral infinite-site model with zero-recombination (TAVARÉ 1984). For instance, we would like to know whether the *v* region is incompatible with the estimate of $3N_e\mu$ at *forked* (or *Om(1D)*). Therefore, we ask what is the probability of finding two or fewer polymorphic sites in our sample in the *v* gene region, given the estimate of $3N_e\mu$ at *forked* (or *Om(1D)*). Similarly, we ask what is the probability of detecting no polymorphic site at *fw* in the Burma (or India) population sample. Table 2 shows that all such probabilities are very low. Thus, the estimates of average nucleotide heterozygosity at *fw* and *v* are significantly lower than those at *forked* or *Om(1D)*.

DISCUSSION

Patterns of variation appear to correlate with recombination rate: Our six-cutter survey of the *fw* and *v* regions in two *D. ananassae* populations found three salient results. (1) Average nucleotide hetero-

TABLE 2

Heterogeneity in estimates of $3N_e\mu$ between gene regions in two *D. ananassae* populations

	Burma	India
$P(S_v \leq 2 \mid forked)$	0.00097	0.02134
$P(S_v \leq 2 \mid Om(1D))$	0.00275	0.00877
$P(S_{fw} = 0 \mid forked)$	0.00001	0.00027
$P(S_{fw} = 0 \mid Om(1D))$	0.00002	0.00008

Our surveys in the *v* region revealed two polymorphic nucleotide sites in 35 × 12 = 420 sites examined in 19 lines of the Burma population sample and in 20 of the India sample. Using TAVARÉ's formula (9.5), we calculated the probability $P(S_v \leq 2 \mid forked)$ of observing two or fewer polymorphic sites at *v*, given the known estimate of $3N_e\mu$ at *forked* (see also HUDSON 1990). Similarly, we calculated the probability $P(S_v \leq 2 \mid Om(1D))$ of observing two or fewer segregating sites at *v*, given the known estimate of $3N_e\mu$ at *Om(1D)*. In the *fw* region, 44 × 12 = 528 sites were surveyed. All were monomorphic. The probabilities of observing no segregating sites at *fw* (given the estimate of $3N_e\mu$ at *forked* and *Om(1D)*) are denoted by $P(S_{fw} = 0 \mid forked)$ and $P(S_{fw} = 0 \mid Om(1D))$, respectively.

zygosity in these proximal gene regions was low within each population relative to the more distal regions *forked* and *Om(1D)*. In the *fw* region, no DNA polymorphism was detected within each population. (2) The frequency spectrum of restriction site polymorphisms in the *v* region was skewed in both populations. The frequency of site polymorphisms in each sample was low, in particular, in the sample from Burma. TAJIMA's *D*, a measure of skewness, was negative. However, the latter result was not statistically significant. (3) The distribution of DNA polymorphisms between the two populations shows an interesting pattern. At *fw*, three fixed differences between the Burma and India populations were detected [two restriction site differences and one insertion/deletion (approximately 2 kb) difference]. At *v*, all three DNA polymorphisms with high frequencies in the (total) sample (*Ins(f)*, *Ins(h)–(l)*, and the *Bam*HI-site at position 10.6) occurred as nearly fixed differences, although 100% fixations were not detected. This pattern is completely different from the distribution of polymorphisms at *forked* and *Om(1D)*. Although the frequencies of many polymorphisms at the latter two loci were significantly different between the two populations, only a small minority of the differences was nearly fixed between populations, and none was completely fixed. Most polymorphisms were in intermediate frequencies in both populations (STEPHAN and LANGLEY 1989; STEPHAN 1989).

This pattern of within-population variation and between-population differentiation seems to be associated with recombination rate. Cytological studies (STEPHAN and LANGLEY 1989; TANDA *et al.* 1989), combined with genetical mapping (HINTON 1988), indicate that recombination rates in *v* and *fw* are much lower than in *forked* and *Om(1D)*. Furthermore, it is likely that recombination rate in the *fw* gene region is lower than in the *v* region. *fw* maps to the junction between centric heterochromatin and euchromatin on the right *X* chromosome arm, where no clear polytene banding pattern exists, whereas *v* is located in a region of the left arm, where polytene bands are still clearly visible. In fact, *v* maps to the last major band, but there are two more minor bands recognizable between this band and the centromere (TANDA *et al.* 1989). This evidence suggests that recombination rate is lowest in the *fw* region, increases in the *v* region and is much higher in *forked* and *Om(1D)*. In parallel, our population surveys indicate an increase of levels of within-population variation from *fw* to *v*, and to *forked* and *Om(1D)*. Fixed differences between populations disappear in that same order and are replaced by a more transient pattern of molecular evolution. Rather than observing a bimodal distribution of fixations, a broad range of frequencies of polymorphisms is found.

Distinguishing between hypotheses: The heterogeneity tests indicate that average nucleotide heterozygosity at *fw* is low relative to the levels of polymorphism at *forked* and *Om(1D)*. There are several explanations for reduced variation in this gene region. It may be explained by the neutral theory by assuming a reduced mutation rate. In principle, this explanation could be tested by the method of TAJIMA (1989) or that of HUDSON, KREITMAN and AGUADÉ (1987). However, because no within-population polymorphisms were detected in the *fw* region, the TAJIMA test can not be applied. The HKA test requires interspecific data, which are not available yet for the *D. ananassae* system. Nevertheless, between-population data can be used to argue against the hypothesis of a reduced neutral mutation rate at *fw*. We have observed that all sites segregating in the total sample of 39 chromosomes from Burma and India are fixed in one population or the other. HEY (1991) has demonstrated that, under neutrality, the mere presence of fixed differences between samples randomly drawn from the same population is statistically significant, even for small samples. Whether our result deviates significantly from predictions of the neutral theory depends on the amount of gene flow between these two populations. Analyses of the *v*, *forked* and *Om(1D)* regions suggest that the rate of migration between the Burma and India populations is to some extent restricted, but is certainly not very low. For instance, Figure 4 of STEPHAN and LANGLEY (1989) shows that several *v* haplotypes are shared between the two populations. If these haplotypes are shared because of interpopulation migration, then our observation that all segregating sites are fixed in one population or the other seems to contradict a neutral model. To test this conclusion, an accurate measurement of gene flow between the Burma and India populations is required.

Reduced levels of variation may also be explained

by models of natural selection. The first model invokes purifying selection. This model assumes that most mutations are slightly deleterious and are therefore eliminated from the population. The explanation of the fixed differences at *fw* further requires that the ancestral populations were smaller than the present-day populations, so that slightly deleterious mutations in the ancestral populations behaved essentially neutrally and were able to go to fixation quickly. There is evidence for a size expansion of *D. ananassae* populations. The recent colonization of the Americas by *D. ananassae* and its high abundance in today's Brazil is perhaps the best known example of population expansion in this species (FREIRE-MAIA 1961). However, the distribution of selection coefficients in a purifying selection model requires a very fine tuning to explain the data, *i.e.*, strong selection against variation at *fw*, somewhat weaker selection at *v*, and about 10-fold lower selection coefficients at *forked* and *Om(1D)*.

At present a model of genetic hitchhiking associated with directional selection appears to be the more parsimonious explanation of the data. As outlined in the Introduction, a selected sweep can account for a collective elimination of variation in chromosomal regions of low recombination. It can also explain the skewness of the frequency spectrum toward rare variants and the tendency toward fixed differences between populations. Crucial questions concern the selection coefficient and the frequency with which selective sweeps have to be postulated in order to explain the observations. These questions are hard to answer for *D. ananassae*, because only four gene regions have been surveyed. However, for *D. melanogaster* for which levels of variation have been quantified for some 20 loci, we were able to obtain estimates of the parameters of a simple equilibrium hitchhiking model (STEPHAN, WIEHE and LENZ 1992). Selection coefficients in the order of 1% and frequencies of selected fixations of 1 in 5000 can explain the data (T. WIEHE and W. STEPHAN, unpublished results). AQUADRO and BEGUN (1992) obtained similar values using a different approach. These frequencies are much less than HALDANE's (1957) maximum rate of adaptive fixations (one every 300 generations) which was calculated on the basis of genetic deaths that a species can tolerate without going extinct. The mechanisms for selective sweeps are unknown for all gene regions for which reduced variation has been found. It may be that selection at the level of the individual or meiotic drive is responsible for the sweep. The evidence compiled by BEGUN and AQUADRO (1992) suggests that some form of zygotic selection leading to allelic fixations has been operating in the Drosophila genome (CHARLESWORTH 1992).

Evolutionary significance of genetic hitchhiking: MAYNARD SMITH and HAIGH (1974) suggest that hitchhiking may explain why the extent of polymorphism in natural populations does not vary as much as one would expect on the basis of the neutral theory of molecular evolution, an observation made by LEWONTIN (1974). The neutral theory (KIMURA 1983) predicts that levels of variation scale with effective population size N_e, whereas hitchhiking models show only a weak dependency on N_e (KAPLAN, HUDSON and LANGLEY 1989; STEPHAN, WIEHE and LENZ 1992). The data compiled by BEGUN and AQUADRO (1992) suggest that hitchhiking seems to occur over a large portion of the *D. melanogaster* genome and may constitute a major factor in constraining levels of variation. However, data from other species which could be used for examining the population size effect are unfortunately not available at present.

Another important consequence of hitchhiking is suggested by this study. As the accumulation of fixed differences between subpopulations in the *fw* and, to a lesser extent, in the *v* region indicates, hitchhiking can lead to rapid genetic differentiation between subpopulations in regions of low recombination. Evidence for considerable genetic differentiation in regions of low recombination has also been found in *D. melanogaster* (MARTÍN-CAMPOS *et al.* 1992; D. J. BEGUN and C. F. AQUADRO, personal communication). The present study provides also some evidence that the mean number of fixed differences between populations is not changed in regions undergoing intense hitchhiking [relative to neutral loci such *forked* and *Om(1D)*], but that the variance of the number of differences between populations is increased. If so, this observation is consistent with an argument by BIRKY and WALSH (1988) who pointed out that the rate of (linked) neutral molecular evolution (*i.e.*, the mean number of neutral substitutions per time interval linked to a selected substitution) is not altered by selective fixations. More data are desirable to put these conclusions on a firmer basis.

We thank LEE ANN LESHKO, SOICHI TANDA and VICTOR CORCES (The Johns Hopkins University) for the *fw* subclones; ELAINE KASPER (deceased), DEREK REID and GABRIELLA RYAN for help with plasmid and genomic DNA preparation; ALBA GONZALES and JAROSLAW KALKA for technical assistance; and CHARLES AQUADRO, ANDREW CLARK, DAVID KIRBY, CHARLES LANGLEY, LEE ANN LESHKO, GAIL SIMMONS, SOICHI TANDA and an anonymous reviewer for comments on earlier versions of this manuscript. This work was supported by an award from the General Research Board of the University of Maryland to W.S.

LITERATURE CITED

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., and D. J. BEGUN, 1992 Evidence for and implications of genetic hitchhiking in the Drosophila genome (or, a hitchhiker's guide to the Drosophila genome), pp. xxx, in

Molecular Paleo-population Biology, edited by N. TAKAHATA and A. G. CLARK. Springer Verlag, Berlin (in press).

BEECH, R. N., and A. J. LEIGH BROWN, 1989 Insertion-deletion variation at the *yellow-achaete-scute* region in two natural populations of *Drosophila melanogaster*. Genet. Res. **53:** 7–15.

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.

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

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.

BINGHAM, P. M., R. LEVIS and G. M. RUBIN, 1981 Cloning of DNA sequences from the *white* locus of *D. melanogaster* by a novel and general method. Cell **25:** 693–704.

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

BOCK, I. R., and M. R. WHEELER, 1972 The *Drosophila melanogaster* species group. Univ. Texas Publ. **7213:** 1–102.

CHARLESWORTH, B., 1992 New genes sweep clean. Nature **356:** 475–476.

DA LAGE, J.-L., M.-L. CARIOU and J.-R. DAVID, 1989 Geographical polymorphism of amylase in *Drosophila ananassae* and its relatives. Heredity **63:** 67–72.

DOBZHANSKY, T., and A. DREYFUS, 1943 Chromosomal aberrations in Brazilian *Drosophila ananassae*. Proc. Natl. Acad. Sci. USA **29:** 301–305.

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.

FREIRE-MAIA, N., 1961 Peculiar gene arrangements in Brazilian natural populations of *Drosophila ananassae*. Evolution **15:** 486–495.

HALDANE, J. B. S., 1957 The cost of natural selection. J. Genet. **55:** 511–524.

HEY, J., 1991 The structure of genealogies and the distribution of fixed differences between DNA sequence samples from natural populations. Genetics **128:** 831–840.

HINTON, C. W., 1988 Formal relations between *Om* mutants and their suppressors in *Drosophila ananassae*. Genetics **120:** 1035–1042.

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

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

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

JOHNSON, F. M., 1971 Isozyme polymorphisms in *Drosophila ananassae*: Genetic diversity among island populations in the South Pacific. Genetics **68:** 77–95.

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*. Cambridge University Press, Cambridge, England.

KREITMAN, M., 1991 Detecting selection at the level of DNA, pp. 204–221 in *Evolution at the Molecular Level*, edited by R. K. SELANDER, A. G. CLARK and T. S. WHITTAM. Sinauer, Sunderland, Mass.

LEWONTIN, R. C., 1974 *The Genetic Basis of Evolutionary Change*. Columbia University Press, New York.

LYNCH, M., and T. J. CREASE, 1990 The analysis of population survey data on DNA sequence variation. Mol. Biol Evol. **7:** 377–394

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.

MARTÍN-CAMPOS, J. M., J. M. COMERÓN, 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 hitchhiking effect of a favorable gene. Genet. Res. **23:** 23–35.

MORIWAKI, D., and Y. N. TOBARI, 1975 *Drosophila ananassae*, pp. 513–535 in *Handbook of Genetics*, Vol. 3, edited by R. C. KING. Plenum, New York.

NEI, M., 1987 *Molecular Evolutionary Genetics*. Columbia University Press, New York.

NEI, M., and F. TAJIMA, 1981 DNA polymorphism detectable by restriction endonucleases. Genetics **97:** 145–163.

PASTINK, A., C. VREEKEN, M. J. M. NIVARD, L. L. SEARLES and E. W. VOGEL, 1989 Sequence analysis of N-ethyl-N-nitrosourea-induced *vermilion* mutations in *Drosophila melanogaster*. Genetics **123:** 123–129.

PATTERSON, J. T., and W. S. STONE, 1952 *Evolution in the Genus Drosophila*. Macmillan, New York.

SAMBROOK, J., E. F. FRITSCH and T. MANIATIS, 1989 *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

SEARLES, L. L., and R. A. VOELKER, 1986 Molecular characterization of the *Drosophila vermilion* locus and its suppressible alleles. Proc. Natl. Acad. Sci. USA **83:** 404–408.

STEPHAN, W., 1989 Molecular genetic variation in the centromeric region of the *X* chromosome in three *Drosophila ananassae* populations. II. The *Om(1D)* locus. Mol. Biol. Evol. **6:** 624–635.

STEPHAN, W., and C. H. LANGLEY, 1989 Molecular genetic variation in the centromeric region of the *X* chromosome in three *Drosophila ananassae* populations. I. Contrasts between the *vermilion* and *forked* loci. Genetics **121:** 89–99.

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

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

TANDA, S., A. E. SHRIMPTON, C. W. HINTON and C. H. LANGLEY, 1989 Analysis of the *Om(1D)* locus in *Drosophila ananassae*. Genetics **123:** 495–502.

TAVARÉ, S., 1984 Line-of-descent and genealogical processes and their applications in population genetics models. Theor. Popul. Biol. **26:** 119–164.

Communicating editor: A. G. CLARK