Lack of Polymorphism on the Drosophila Fourth Chromosome Resulting From Selection

Andrew J. Berry,* James W. Ajioka† and Martin Kreitman*

**Department of Ecology and Evolutionary Biology, Princeton University, Princeton, New Jersey 08544, and †London School of Hygiene and Tropical Medicine, University of London, London WC1E 7HT, United Kingdom*

Manuscript received May 4, 1991
Accepted for publication August 21, 1991

ABSTRACT

Evolutionary processes can be inferred from comparisons of intraspecific polymorphism and interspecific divergence. We sequenced a 1.1-kb fragment of the *cubitus interruptus Dominant* (ci^D) locus located on the nonrecombining fourth chromosome for ten natural lines of *Drosophila melanogaster* and nine of *Drosophila simulans*. We found no polymorphism within *D. melanogaster* and a single polymorphism within *D. simulans*; divergence between the species was about 5%. Comparison with the alcohol dehydrogenase gene and its two flanking regions in *D. melanogaster*, for which comparable data are available, revealed a statistically significant departure from neutrality in all three tests. This lack of polymorphism in the ci^D locus may reflect recent positive selective sweeps on the fourth chromosome with extreme hitchhiking generated by the lack of recombination. By simulation, we estimate there to be a 50% chance that the selective sweeps occurred within the past 30,000 years in *D. melanogaster* and 75,000 in *D. simulans*.

PATTERNS of genetic variation provide clues about processes affecting evolution, and, ideally, we can use this information to distinguish between adaptive and neutral evolutionary change. However, DNA polymorphism is not only affected by selection and genetic drift: recombination and mutation rates as well as population history and structure have important effects. How can we disentangle the action of selection and/or drift from so complex a framework? Furthermore, how can we distinguish between negative (purifying) selection, in which evolutionary change is constrained, and positive (adaptive) selection, in which favorable mutations are selectively driven to fixation? HUDSON, KREITMAN and AGUADÉ's (1987) approach is based on the neutral expectation of a correspondence between levels of polymorphism within a species and divergence between closely related species. This approach controls for differences in levels of selective constraint between regions: high constraint reduces levels of both intraspecific polymorphism and interspecific divergence. Positive selection, in contrast, can lead to an uncoupling of levels of polymorphism and divergence.

The extent of hitchhiking of neutral mutations in response to evolutionary changes at another part of the genome depends on the recombination distance from the site under selection (MAYNARD-SMITH and HAIGH 1974; KAPLAN, HUDSON and LANGLEY 1989). In the extreme case of no recombination the whole chromosome is a single linkage group. Thus the selective fixation of any mutation leads to the concomitant fixation of all other variants on that chromosome, and the population can only gain polymorphism by the accumulation of new mutations. Periodic selection in *Escherichia coli* chemostats is a well known example of this phenomenon (ATWOOD, SCHNEIDER and RYAN 1951).

Low recombination rates therefore facilitate the indirect detection of positive selection at a distance; the effect of selection in recombining DNA is likely to be much more local. For this reason, there have been a number of attempts to assay natural variation in a genome region of reduced recombination. These studies have concentrated on *D. melanogaster*'s *yellow-achaete-scute* region, which, situated at the distal tip of the *X* chromosome, has reduced levels of recombination. AGUADÉ, MIYASHITA and LANGLEY (1989) report the lowest published estimate of per nucleotide heterozygosity for any studied region of *D. melanogaster* ($\theta = 0.001$; θ is approximately the probability of a nucleotide site being different in two randomly chosen genes), which differs significantly from an assumed value for the species of $\theta = 0.005$ (but see KREITMAN 1991). Three smaller-scale restriction fragment length polymorphism (RFLP) studies of the same region (EANES, LABATE and AJIOKA 1989; BEECH and LEIGH-BROWN 1989; MACPHERSON, WEIR and LEIGH-BROWN 1990) also estimate θ to be low, but nucleotide diversity estimates ($\pi = 0.003$, 0.003 and 0.002, respectively) are markedly higher than that of AGUADÉ, MIYASHITA and LANGLEY (1989) ($\pi = 0.0003$). This discrepancy highlights the first shortcoming of these studies: they are based on restriction enzyme surveys over a large piece of DNA. Because they add an

additional source of variance by selectively sampling only the recognition sites of the enzymes used, RFLP studies employing a small number of enzymes can yield polymorphism estimates with large variances. A second problem is the lack of interspecific comparison. Even if levels of heterozygosity are indeed lower than elsewhere in the genome, other factors, such as strong selective constraint in the region or a lowered mutation rate, could be responsible. Such alternatives to a selective sweep can only be excluded if levels of interspecific divergence in that region are found to be sufficiently high, implying an unexpectedly low level of polymorphism.

Here we examine a region for which there is virtually no genetic evidence of recombination: the Drosophila fourth chromosome. The reasons for this lack of recombination are not understood (HOCHMAN 1976). Under standard laboratory rearing conditions recombination is essentially absent, though it can be induced by heat shock. This "microchromosome" constitutes approximately 4% of the haploid complement of *D. melanogaster* and contains an estimated minimum of 75 major loci (HOCHMAN 1976). To our knowledge, the only cloned single copy locus on the *D. melanogaster* fourth chromosome is the segment polarity gene, *cubitus interruptus Dominant* (ci^D) (ORENIC *et al.* 1990), whose mutation causes defects in the posterior half of every embryonic segment. We sequenced a 1.07-kb genomic DNA fragment encoding part of the ci^D product in ten naturally derived lines of *D. melanogaster* and nine such *D. simulans* lines. To demonstrate that the lines used in general exhibit levels of polymorphism typical of each species, we also present restriction fragment polymorphism data for the alcohol dehydrogenase gene (*Adh*) for *D. melanogaster* lines from the studied population and sequence data for the *white* locus for the *D. simulans* lines used.

MATERIALS AND METHODS

Flies and DNA preparation: All *D. melanogaster* studied were derived from a single population collected at Eckert's Orchard in Belleville, Illinois, in August 1988. Individual wild-caught females were established as isofemale lines and single fourth chromosomes extracted from ten of these lines using a dominantly marked fourth chromosome (ey^D). A balancer is not necessary in view of the lack of recombination. The resulting isochromosomal lines were expanded and DNA prepared by means of sodium dodecyl sulfate lysis, potassium acetate precipitation, phenol extraction and ethanol precipitation. For the *Adh* study, individual males from eight isofemale lines were crossed to virgin A178 females, which carry a deletion in the *Adh* region. DNA was prepared from a single progeny from each of these crosses by means of a scaled down version of the above procedure, though with an isopropanol rather than ethanol precipitation. The *D. simulans* lines used were from a worldwide collection of isofemale lines: single lines from Beltsville, Maryland; Murakata City, Japan; Palmers Island, Australia; Ottawa, Canada; Cairns, Australia; Brazzaville, Congo; Morven, Georgia; Seychelles; and France. No attempt was made to render the fourth chromosome homozygous by inbreeding or other means. To minimize heterozygosity within isofemale lines, we used only a single fly from each line. DNA was prepared as above. The same genomic DNA stock was used for sequencing an intron of the *white* locus in *D. simulans*.

Polymerase chain reaction (PCR) and sequencing template preparation: A fragment 1.2 kb long was PCR-amplified from ci^D using 20-mer oligonucleotide primers starting at positions 1897 ('+' primer 5′ base) and 3003 ('−' primer 5′ base) in the sequence of ORENIC *et al.* (1990). Amplification was carried out in 100-μl reaction volumes with 2 units of *Taq* polymerase (BRL) under standard conditions (SAIKI *et al.* 1988) with a 2 mM magnesium ion concentration. Prior to amplification, one primer was phosphorylated by means of a kinase reaction; this renders the strand extended from that primer in the product susceptible to lambda exonuclease digestion (HIGUCHI and OCHMAN 1989). A 30-min digestion of PCR product with 2 units of λ-exonuclease (BRL) generated single-stranded sequencing template, which was cleaned up with phenol/chloroform/isoamyl alcohol and chloroform extractions. Excess primers from the amplification reaction were selectively removed prior to sequencing by means of a single volume, five minute room temperature ethanol precipitation with ammonium acetate (KREITMAN and LANDWEBER 1989). Each PCR product was used for two sequencing reactions. The entire *Adh* region was amplified under similar standard conditions from the *Adh* deletion heterozygote flies using 20-mer primers flanking the structural locus [starting in positions −654 and 2672 (numbering after KREITMAN and HUDSON 1991)]. Sequencing template for the *white* locus intron in *D. simulans* lines was prepared using appropriate primers in the same way as the ci^D template.

Four-cutter analysis of *Adh* deletion heterozygotes: Amplification product was split eight ways and digested overnight with a single unit of the following enzymes: *Alu*I, *Dde*I with *Bam*HI, *Hae*III, *Hha*I, *Hin*fI, *Msp*I, *Sau*3aI, *Taq*I. Samples were electrophoresed, electroblotted and probed with a 2.7-kb *Sal*I-*Cla*I fragment incorporating the *Adh* structural locus according to KREITMAN and AGUADÉ (1986). The position of variants could be mapped precisely with prior knowledge of the region's complete sequence (KREITMAN 1983).

Sequencing: Both strands were dideoxy-sequenced (SANGER, NICKLEN and COULSON 1977) with modified T7 DNA polymerase (U.S. Biochemical). For ci^D we used three internal primers and the amplification primer for sequencing the plus strand, and four internal primers for the minus strand. Sequence was electrophoresed on short (30 cm × 40 cm × 0.4 mm 6% polyacrylamide/7 M urea) gels with a NaOAc electrolyte gradient (SHEEN and SEED, 1988) and on long (30 cm × 60 cm × 0.4 mm 6% "Long Ranger" (AT Biochem)/7 M urea) gels. *D. simulans* sequence from the *white* intron was obtained using the amplification primers and electrophoresis on short gels.

Analysis: Sequences were manually aligned and divergence between species was calculated in terms of fixed differences; *D. melanogaster* and *D. simulans* are closely related and no correction for multiple hits was made. We restricted tests of neutrality to those regions for which we have full sequence information for both polymorphism within *D. melanogaster* and divergence between *D. melanogaster* and *D. simulans*: *Adh* and its two flanking regions (KREITMAN and HUDSON 1991). No such dataset exists for polymorphism within *D. simulans*. Exclusive use of full sequence data (rather than RFLP information) permits an

exact measure of the number of effectively silent sites under study. We applied the test of HUDSON, KREITMAN and AGUADÉ (1987) with modifications to take into account differences in sample sizes (10 *cubitus interruptus* alleles; 11 *Adh* alleles): the first equation of formula (5) [HUDSON, KREITMAN and AGUADÉ (1987), p. 154] was changed to:

$$c(n_1)\sum_{i=1}^{L}\hat{\theta}_i + c(n_2)\sum_{i=1}^{L}\hat{\theta}_i,$$

where n_1 and n_2 are the number of alleles sampled from each locus.

RESULTS

We sequenced 1075 bp for each line. The *D. melanogaster* sequence for all ten lines is identical, and differs from the published cDNA sequence (ORENIC *et al.* 1990) in three places. There are two putative introns, both of which start and finish with the expected consensus couplets (*i.e.*, GT. . . .AG): the first, 64 bp long, starts after position 1938 (numbering after ORENIC *et al.* 1990); the second, 53 bp long, starts after position 2103. The sequence thus comprises two entire introns, a short intervening exon, and part of a large exon, and includes 330.9 effectively silent sites of which 217.9 are in coding DNA. The two sequences also differ from position 2787 to 2791: ORENIC *et al.* give TATAT whereas we have ATATA.

The results are presented in Table 1. There are 54 substitutions between *D. melanogaster* and *D. simulans* and a single base pair insertion in the second intron of *D. simulans*. Twenty-one of the 44 substitutions in exons lead to amino acid replacements; none of these occurs in the short exon. The only polymorphism is a single difference between *D. simulans* line 1 (from Beltsville, Maryland) and all other *D. simulans* lines. Line 1 is homozygous for this difference.

Table 2 presents comparisons of the data with data from other regions of the Drosophila genome for which comparable divergence/polymorphism data are available. All comparisons indicate a deviation from the neutral expectation of a correspondence between the extent of divergence between species and the amount of polymorphism within a species. Although this survey includes comparison with the *Adh* structural locus which has elevated levels of polymorphism (KREITMAN and HUDSON 1991), the deviation remains statistically significant in comparisons with other, less polymorphic, regions, such as the *Adh-dup* and flanking regions. It is important to note that KREITMAN and HUDSON (1991) were unable to reject neutrality in a comparison of polymorphism and divergence in the *Adh* flanking and *Adh-dup* regions.

Although comparable data are not available to test the significance of the polymorphism observations in *D. simulans*, it is likely that the lack of polymorphism

TABLE 1

Sequence comparison of ten *D. melanogaster* and nine *D. simulans* lines

Region	Bases sequenced	Position[a]	DNA sequence mel	DNA sequence sim	Amino acid sequence mel	Amino acid sequence sim
Intron	64	25	C	A		
		27	A	T		
		28	T	A		
		37	A	G		
		47	A	T		
Exon	165	1980	G	C	Leu	Leu
		1995	G	A	Arg	Arg
		2049	C	T	Ala	Ala
		2076	G	A	Lys	Lys
		2079	T	C	His	His
Intron	53	11	C	T		
		12	A	C		
		17	T	A		
		23	INSERT A			
		24	G	A		
		37	T	G		
Exon	793	2154	G	A	Pro	Pro
		2190	A	T	Gly	Gly
		2193	T	A	Ala	Ala
		2260	A	G	Asn	Asp
		2287	C	A	His	Asn
		2337	G	T	Met	Ile
		2338[b]	T	T/G	Leu	Leu/Val
		2340[b]	A	G	Leu	Leu
		2397	C	G	His	Gln
		2431	T	G	Leu	Val
		2453	T	A	Leu	His
		2478	T	C	Asp	Asp
		2496	T	C	Asp	Asp
		2499	C	T	Asp	Asp
		2508	G	C	Val	Val
		2541	A	G	Val	Val
		2562	A	G	Ala	Ala
		2583	T	G	Gly	Gly
		2584	T	G	Ser	Ala
		2611	G	A	Gly	Ser
		2658	C	T	Cys	Cys
		2667	A	G	Pro	Pro
		2682	C	T	Thr	Thr
		2683	T	G	Phe	Val
		2704	C	G	Gln	Glu
		2737	A	G	Thr	Ala
		2741	A	C	Asp	Ala
		2743[b]	G	C	Ala	Arg
		2744[b]	C	G	Ala	Arg
		2760	T	A	Pro	Pro
		2762	A	T	Lys	Met
		2765	T	C	Leu	Pro
		2772	C	T	Asn	Asn
		2791	A	C	Thr	Pro
		2837	C	T	Ser	Leu
		2847	A	G	Gln	Gln
		2851	A	G	Thr	Ala
		2868	A	G	Ile	Met
		2874	A	G	Gly	Gly
		2879	T	C	Phe	Ser

[a] Position refers to the published sequence of ORENIC *et al.* (1990) except in introns where the numbering starts at the beginning of each intron.

[b] Differences occurring within the same codon.

TABLE 2

Tests of neutrality in *cubitus interruptus* and *Adh* region

Comparison		*Adh* 5′ region		*Adh* locus		*Adh-dup* locus	
		Silent	Total	Silent	Total	Silent	Total
Polymorphism	*Cubitus interruptus*	0 (6.87)	0 (9.8)	0 (9.4)	0 (12.1)	0 (4.4)	0 (6.0)
	Compared locus	30 (23.1)	30 (20.2)	20 (10.6)	21 (8.9)	13 (8.6)	14 (8.0)
Divergence	*Cubitus interruptus*	33 (26.1)	54 (44.2)	33 (23.6)	54 (41.9)	33 (28.6)	54 (48.0)
	Compared locus	78 (84.9)	78 (87.8)	16 (25.4)	18 (30.1)	50 (54.4)	56 (62.0)
Test		5.15*	6.85**	11.14***	15.95***	4.17*	5.81*

Expected values of the number of polymorphic or divergent sites (under neutrality) are given in parentheses after the observed values. The test statistic is approximately chi square distributed with one degree of freedom.
* $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

in this species constitutes an even more extreme departure from the neutral expectation than that in *D. melanogaster* in view of the relatively higher overall levels of nucleotide polymorphism in *D. simulans* (AQUADRO 1991). Given the typical levels of sequence divergence (*i.e.*, between 5 and 10% for silent sites) between the two species, we conclude that ci^D is unusually monomorphic within both these species.

The test for *D. melanogaster* compared *cubitus interruptus* data derived from flies from a single population with *Adh* and *Adh*-flanking data from a worldwide sample. Is the greater than expected homozygosity merely a product of comparing variation within a local population to that within a global sample? This is unlikely because we find from four-cutter studies of *Adh* that all silent allelic variation detected in KREITMAN's (1983) worldwide sequencing study occurs within individual populations in North America (KREITMAN and AGUADÉ 1986; SIMMONS *et al.* 1989; A. J. BERRY and M. KREITMAN, unpublished results). To test the possibility that a local bottleneck has caused the lack in variation in the Eckert's Orchard population, we looked at polymorphism at another locus, *Adh*. This four-cutter study revealed four polymorphisms segregating among eight lines: *Dde*I site loss at position 1527 (numbering after KREITMAN 1983) in three lines; *Hae*III site losses at positions 686 (two lines) and 817 (three lines); *Hha*I site loss at position 573 (two lines). This level of variation is consonant with that observed for a much larger sample of 1533 North American individuals from 25 populations (A. J. BERRY and M. KREITMAN, unpublished results), implying that the nucleotide variation within the Eckert's Orchard population is not unusually low. A similar check on polymorphism was carried out on the *D. simulans* lines studied: sequence of a 250-bp region of a *white* intron from 13 individuals, including the nine used in this study, revealed 24 segregating polymorphic substitutions (J. COYNE, personal communication).

DISCUSSION

The ci^D region is unusually monomorphic in both *D. melanogaster* and *D. simulans*. This lack of polymorphism is not caused by selective constraint or by a reduced mutation rate because the divergence between the two species is similar to the other test loci. What therefore accounts for this homozygosity? Either these flies come from highly inbred populations or a deterministic process such as selection must be responsible. We can reject the first alternative on the basis of our survey of other loci. If population structure is the cause of the monomorphism at ci^D, then we expect to see a corresponding low heterozygosity throughout the genome. This is not the case. Although we have made no systematic attempt to survey other parts of the genome, the polymorphism data presented for *Adh* in *D. melanogaster* and for *white* in *D. simulans* indicates that we are not dealing with peculiar populations; rather, it is the level of polymorphism at ci^D that is lower than expected under neutrality.

The direction of the departure from neutrality in both species suggests two alternative models of the action of selection. The first invokes purifying selection. The argument is as follows: original population sizes in both species were small, permitting effectively neutral drift of mildly deleterious alleles; this accounts for the accumulation of fixed differences between the two species. Subsequent population expansion has increased the efficacy of selection against such mildly deleterious mutations, and what we see, within species, is the wholesale removal of variation by purifying selection. Given the monomorphism of both silent and replacement sites, a slightly deleterious alleles model requires that selection coefficients against both silent and replacement sites fall within the same range of values between $1/2N_{e1}$ and $1/2N_{e2}$, where N_{e1} and N_{e2} are the pre- and postexpansion effective population sizes, respectively (EWENS 1979). It is unlikely that these two types of mutation, which have entirely different functional consequences, would have similar selection coefficients. In order for a purifying selection model to work, the preexpansion population size would have to have been small. While there is little doubt that both species have recently expanded world-

wide, it is likely that their original African population sizes were nevertheless relatively large. *Drosophila yakuba*, a member of the melanogaster subgroup that has not expanded its range beyond the group's native Africa, exhibits high levels of polymorphism, which imply a relatively large population size (McDONALD and KREITMAN 1991). In short, purifying selection is unlikely to account for the observed levels of polymorphism at *cubitus interruptus*.

The second selection model invokes a selective sweep to account for the elimination of variation in this region in both species. Possible mechanisms for such a selective sweep include natural selection of a favored mutation, segregation distortion, and biased gene conversion. We cannot distinguish between these alternatives but point out that there is no evidence of either drive or biased gene conversion occurring on the fourth chromosome; we will therefore couch the following mechanistic discussion in terms of natural selection. The same arguments, however, may be applied to drive and biased gene conversion. If, anywhere on the fourth chromosome, a favored mutation arises that is selectively driven to fixation, then the population will become fixed for the entire chromosome carrying the mutation. The standing crop of variation on the fourth chromosome is eliminated because variants cannot recombine onto the favored chromosome as it is driven to fixation. Polymorphism will only accumulate through mutation of the favored chromosome. If such a selective sweep has occurred in the recent history of a species, then we would expect to see monomorphism. Presumably such sweeps have occurred in both *D. melanogaster* and *D. simulans*. That we see the same pattern of monomorphism in both species hints that such positively selected evolutionary events are not rare. The possibility of recombination on the fourth chromosome does not affect our conclusion of a selective sweep incorporating *cubitus interruptus*. Recombination would only affect a calculation of the rate of selective substitution per locus (see below).

Is it possible to estimate the time back in generations, t, to the selective sweep? We will assume an instantaneous fixation of the favored allele. Then, if the sweep occurred recently enough, the time back to the sweep is also the time back to the common ancestor of all the alleles in the sample, the coalescent of the sample. Although it may be possible to estimate the sweep times analytically with coalescence theory (J. HEY, personal communication), the simplest calculation assumes a star configuration for the genealogy (or phylogeny) of alleles. This assumption would be valid for a sufficiently recent common ancestry (discussed below). We can then calculate t as follows. First note that the total time, T_{tot}, on the genealogy of the alleles is $T_{tot} = nt$, where n is the sample size. For any genealogy the expected number of polymorphisms, S, is $S = T_{tot}\mu k$, where μ is the mutation rate per nucleotide site per generation and k is the number of silent sites. For *D. melanogaster*, with no observed polymorphism, the estimate of t is 0. For *D. simulans*, substituting 1 for S, 331 for k, 9 for n, and taking μ to be 1×10^{-9} (ROWAN and HUNT 1991; MORIYAMA 1987) (we assume 10 generations per year) we estimate t to be 3.3×10^5 generations ago. With 10 generations per year, the selective fixation occurred 3.3×10^4 years ago.

Is the assumption of a star phylogeny valid? To answer this question, we can calculate the probability that none of the nine alleles in the sample share a common ancestor more recently than 3.3×10^5 generations ago. Under the infinite alleles neutral model, coalescent times are exponentially distributed. The probability, $P(n)$, of n alleles having n distinct ancestors for each of the past t generations in a population of size N is,

$$P(n)^t \approx e^{-\binom{n}{2}t/2N}.$$

For $n = 9$, $t = 3.3 \times 10^5$ and setting $2n = 5 \times 10^6$ (see AQUADRO 1991), $P(n)^t \approx 0.03$. Therefore, it is likely that two or more of the alleles share a common ancestor subsequent to the selective sweep. This result is somewhat surprising in view of the large effective population size used in the calculation and the recentness of the sweep. With a more recent coalescent of alleles in the sample likely, the star phylogeny method must be considered a minimum estimate of the time back to the selective sweep.

To avoid the questionable assumption of a star phylogeny, we have simulated neutral samples to investigate the probability of observing 1 or fewer polymorphisms for *D. simulans* and 0 for *D. melanogaster* for instantaneous selective fixations occurring at different times in the past. We generated a neutral genealogy of alleles (either 9 or 10) under an infinite alleles model with no recombination (HUDSON 1983, 1990) and imposed an absolute bottleneck (reducing the population size to 1) on the population at a given time. The absolute bottleneck, equivalent to an instantaneous selective sweep, implies that the selection coefficient on the favored mutant is much larger than the reciprocal of the population size. Such a scheme removes any assumption about the topology of the tree (*i.e.*, we need no longer assume a star phylogeny). The distribution of mutations on the gene trees is governed by the mutation rate, μ, the population size, N, and the random sampling process. For the simulations we used $\theta = 4N\mu$ values of 1.5 and 3.0 for *D. melanogaster* and *D. simulans*, respectively. These values are averages of many polymorphism studies at several gene loci (AQUADRO 1991) and correspond to

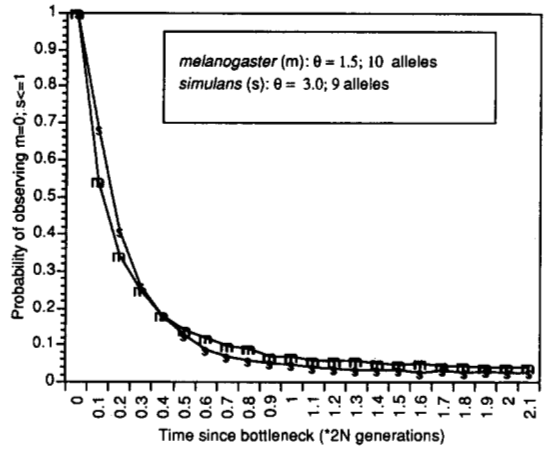

FIGURE 1.—The probability of observing 0 (*D. melanogaster*) or 1 or fewer (*D. simulans*) polymorphisms in samples of 10 or 9 alleles, respectively, subsequent to an instantaneous and absolute bottleneck. The distribution is an estimate obtained by generating 10,000 replicates by a Monte Carlo simulation of a neutral tree genealogy described in the text. N is effective population size.

per nucleotide θ estimates of 0.005 and 0.01 for the two species.

The results are presented in Figure 1. As expected, if the bottleneck occurred in the recent past, then there is a high probability of obtaining the observed low levels of polymorphism in the two species. At the other extreme of an ancient bottleneck, there is still a reasonable chance (4% and 3% for *D. melanogaster* and *D. simulans*) of observing 0 or 1 or fewer polymorphisms in the samples. Indeed, this is the limiting probability for observing the same number of polymorphisms in a neutral sample of appropriate θ when there is no bottleneck at all [expressions for this calculation have been obtained by TAVARÉ (1984) and HUDSON (1990)]. However, there is a 50% probability that an instantaneous selective sweep occurred within the last $0.28N$ and $0.36N$ generations for *D. melanogaster* and *D. simulans*, respectively. If N is 1×10^6 for *D. melanogaster* (KREITMAN 1983) and 2×10^6 for *D. simulans*, this yields estimates of 2.8×10^5 generations ago, or 2.8×10^4 years ago for *D. melanogaster* and 7.2×10^4 years ago for *D. simulans*.

These calculations draw attention to the difficulty of accurately estimating the time of a selective sweep. Even with larger sample sizes the confidence interval remains wide for an estimate of the time of a selective event. For selective sweeps which occur in recombining chromosomal regions, sufficiently slow selective sweeps (*i.e.*, weak selection) and/or high recombination rates may make it impossible to detect even a recent selective sweep. On the other hand, if the reduction of polymorphism is sufficient to detect a selective substitution, an estimate of the size of the region exhibiting reduced polymorphism may provide useful information about the time of the event.

Although our calculations suggest frequent selectively driven substitutions on the fourth chromosome, the rate per locus must be substantially lower. If selective substitutions can only occur at one of the estimated 75 major gene loci on the chromosome then the average rate per gene will be 1/75th the rates calculated above.

Our conclusion that selective sweeps are occurring with some regularity on the fourth chromosome is not dependent on whether or not there is actually a low level of recombination in nature. Recombination will reduce the size of the linkage group affected by the selective sweep, and therefore implies even higher per locus rates of adaptive substitution. Ideally, we would like to conduct a similar study on a locus on another part of the fourth chromosome to see if we find long range linkage disequilibrium as evidence of a lack of recombination. However, this is currently impractical because *cubitus interruptus* is the only molecularly characterized locus on the chromosome. It also lacks polymorphic markers for such a study. The assumption of no recombination merely makes our estimate of the per locus rate of adaptive evolution a conservative one.

How does our estimate of the rate of adaptive fixation compare with theoretical expectations? HALDANE (1957), on the basis of the number of "genetic deaths" that a species can sustain, calculated the maximum rate of adaptive fixation per genome to be one per 300 generations. Do our observations fall within this limit? If the fourth chromosome constitutes about 1% of the *Drosophila* coding genome (HOCHMAN 1976), we expect to see no more than one adaptive substitution per fourth chromosome per 300/0.01 = 30,000 generations. This is well within the minimal rate calculated from our data above.

Although the result can be explained by the joint actions of selection and no recombination, it is not necessarily expected. Selection could equally act to enhance levels of variation on the fourth chromosome through heterozygote advantage. This could either be through classical balanced polymorphism at one locus or through overdominance of the entire linkage group. The latter is intuitively appealing because such a region of no recombination could provide an ideal location for coevolved groups of alleles, DOBZHANSKY's (1943, 1970) "coadapted gene complexes." Two or more such arrangements could be maintained in the population over a long period of time, allowing the build-up of extensive neutral polymorphism. The lack of variation on the fourth chromosome in Drosophila, however, implies that neither form of heterozygote advantage is operating at any one of its 75 loci (in the case of a balanced polymorphism) or across any groups of loci on all or part of the chromosome. This suggests that KREITMAN and HUDSON's (1991) identification of balancing selection at *Adh* in *D. mel-*

anogaster may prove to be exceptional when other loci are studied in detail, and that "coadapted gene complexes" may, at best, be rare.

We thank ALISON COLWELL for collecting flies, TERESA ORENIC for supplying the *cubitus interruptus* cDNA sequence prior to publication, and R. HUDSON, M. TAYLOR, J. HEY and J. H. MCDONALD for critical comments. The work is supported by National Institutes of Health grant GM 39355 and National Science Foundation grant DCB-8646262 to M. K.

LITERATURE CITED

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

AQUADRO, C., 1991 Molecular population genetics of *Drosophila*, in *Molecular Approaches to Pure and Applied Entomology* (Series in Experimental Entomology), edited by J. G. OAKESHOTT and M. WHITTEN. Springer-Verlag, Berlin.

ATWOOD, K. C., L. K. SCHNEIDER and F. J. RYAN, 1951 Selective mechanisms in bacteria. Cold Spring Harbor Symp. Quant. Biol. **16:** 345–354.

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.

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

EWENS, W. J., 1979 *Mathematical population genetics*. Springer-Verlag, Berlin.

DOBZHANSKY, TH., 1943 Genetics of natural populations. IX. Temporal changes in the composition of populations of *Drosophila pseudoobscura*. Genetics **28:** 162–186.

DOBZHANSKY, TH., 1970 *Genetics of the Evolutionary Process*. Columbia University Press, New York.

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

HIGUCHI, R. G., and H. OCHMAN, 1989 Production of single-stranded DNA templates by exonuclease digestion following the polymerase chain reaction. Nucleic Acids Res. **17:** 5865.

HOCHMAN, B., 1976 The fourth chromosome of *Drosophila melanogaster*, pp. 903–928 in *The Genetics and Biology of Drosophila*, Vol. 1b, edited by M. ASHBURNER and E. NOVITSKI. Academic Press, New York.

HUDSON, R. R., 1983 Properties of a neutral tree model with intragenic recombination. Theor. Popul. Biol. **23:** 183–201.

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

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

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

KREITMAN, M., 1983 Nucleotide polymorphism at the *Alcohol Dehydrogenase* locus of *Drosophila melanogaster*. Nature **304:** 412–417.

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. WHITTMAN. Sinauer, Sunderland, Mass.

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

KREITMAN, M., and R. R. HUDSON, 1991 Inferring the evolutionary histories of the *Adh* and *Adh-dup* loci in *Drosophila melanogaster* from patterns of polymorphism and divergence. Genetics **127:** 565–582.

KREITMAN, M., and L. LANDWEBER, 1989 A strategy for producing single-stranded DNA in the polymerase chain reaction: a direct method for genomic sequencing. Genet. Anal. Tech. **6:** 84–88.

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.

MCDONALD, J. H., and M. E. KREITMAN, 1991 Adaptive protein evolution at the *Adh* locus in *Drosophila*. Nature **351:** 652–654.

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

MORIYAMA, E. N., 1987 Higher rates of nucleotide substitution in Drosophila than in mammals. Jpn. J. Genet. **62:** 139–147.

ORENIC, T. V., D. C. SLUSARSKI, K. L. KROLL and R. A. HOLMGREN, 1990 Cloning and characterization of the segment polarity gene *cubitus interruptus Dominant* of *Drosophila*. Genes Dev. **4:** 1053–1067.

ROWAN, R. G., and J. A. HUNT, 1991 Rates of DNA change and phylogeny from the DNA sequences of the alcohol dehydrogenase gene for five closely related species of Hawaiian *Drosophila*. Mol. Biol. Evol. **8:** 49–70.

SAIKI, R. K., D. H. GELFAND, S. STOFFEL, S. J. SCHARF, R. G. HIGUCHI, G. T. HORN, K. B. MULLIS and H. A. ERLICH, 1988 Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science **239:** 487–491.

SANGER, F., S. NICKLEN and A. R. COULSON, 1977 DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:** 5463–5467.

SHEEN, J., and B. SEED, 1988 Electrolyte gradient gels for DNA sequencing. BioTechniques **6:** 942–944.

SIMMONS, G. M., M. E. KREITMAN, W. F. QUATTLEBAUM and N. MIYASHITA, 1989 Molecular analysis of the alleles of alcohol dehydrogenase along a cline in *Drosophila Melanogaster*. I. Maine, North Carolina, and Florida. Evolution **43:** 393–409.

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

Communicating editor: A. G. CLARK