THE ROYAL SOCIETY

What is the impact of transposable elements on host genome variability?

Philip T. J. Emery, Thelma E. Robinson, Robert Duddington and John F. Y. Brookfield*

Institute of Genetics, University of Nottingham, Queen's Medical Centre, Nottingham NG7 2UH, UK

The spread of a transposable element family through a wild population may be of astonishing rapidity. At least three families of transposable genetic elements have recently invaded *Drosophila melanogaster* worldwide, including the *P* element. The mechanism has been a process of effectively replicative transposition, and, for the *P* element, has occurred notwithstanding the sterility induced by unrestricted movement. This element's invasion into *D. melanogaster* has been accompanied by the development of heterogeneity between *P* sequences, most of which now have internal deletions. Increasing evidence suggests that some deleted elements can repress *P* transposition, thereby protecting the host from the harmful effects of complete elements. Such repressing elements may rise to high frequencies in populations as a result of selection at the level of the host. We here investigate selective sweeps invoked by the spread of *P* sequences in *D. melanogaster* populations. Numerous high-frequency sites have been identified on the X chromosome, which differ in frequency between populations, and which are associated with repression of *P*-element transposition. Unexpectedly, sequences adjacent to high-frequency *P*-element sites do not show reduced levels of genetic diversity, and DNA variability is in linkage equilibrium with the presence or absence of a *P* element at the adjacent selected site. This might be explained by multiple insertions or through a selection for recombination analogous to that seen in 'hitchhiking'.

Keywords: *Drosophila*; selective sweeps; *P* elements

1. INTRODUCTION

Population genetics theory traditionally considers evolutionary changes occurring in organisms with free recombination, such that the frequency changes occurring at variable loci are independent, save for epistatic fitness interactions between the genotypes. The observation, in *Drosophila melanogaster*, of a strong positive correlation between the rate of recombination and the level of gene diversity (π) in wild populations (Begun & Aquadro 1992) is, however, clear evidence that linkage between loci may have a major effect on variability. Two types of explanations have been suggested for this correlation. The first is that of selective sweeps, or 'hitchhiking' (Maynard Smith & Haigh 1974; Kaplan *et al.* 1989; Wiehe & Stephan 1993). Selective sweeps are generated by advantageous mutations, whose rise in frequency creates a region of low variability of a length inversely proportional to the local rate of recombination. An alternative is background selection (Charlesworth *et al.* 1993; Charlesworth 1994). In this model, disadvantageous mutations mark chromosomes as being doomed to inevitable extinction, reducing the effective population size of those that remain. The data currently favour the background selection model in that, in low diversity regions, Tajima's *D*-test, which compares the number of variable sites in a sample to the mean heterozygosity, fails to find the strongly negative *D* expected following selective sweeps (Tajima 1989; Braverman *et al.* 1995).

In this manuscript, we use the invasion of a transposable element family as a natural experiment to investigate the impact of selective sweeps on the host genome. At least three families of transposable elements have invaded wild populations of *D. melanogaster* in this century. One of these is the *P* element, which has spread following a horizontal transfer from *D. willistoni*. The effectively replicative transposition mechanism of the *P* transposable element has resulted in a rapid worldwide spread through *D. melanogaster* populations (Kidwell 1994), more than overcoming a reduction in fertility created through movements of elements in the germ line (Kidwell *et al.* 1977). The spread has been associated with a diversification of *P*-element structure. Some elements have come to produce *trans*-acting repressors of the movements of others, and their reduction of the harmful effects of element movements (Gloor *et al.* 1993) has resulted in secondary spread of these sequences through their host populations. Notwithstanding the recentness of their invasion, some chromosomal sites of *P* elements have high frequencies in populations (e.g. Biémont *et al.* 1994), too high for genetic drift to have been the cause. One such site is at 1A on the X chromosome. This site has been shown to repress transposition (Ronsseray *et al.* 1991). The harmful effects of unregulated transposition on fertility imply that repression would increase host fitness if *P* elements are spreading rapidly, and thus that the high

*Author for correspondence.

Received 15 April 1999 *Accepted* 2 June 1999

frequency of this insertion has been created by selection operating on fertility. A number of sites creating *trans*-acting repression have been described and usually, repression requires the combination of a *P*-element structure that generates a repressing polypeptide, coupled with an insertion site that results in a high level of expression in the germ line (Andrews & Gloor 1995). The dependence of the repressing ability of particular deletion derivatives on their insertion sites explains the inconsistencies in results aimed at demonstrating that certain elements (such as the *KP* element (Black *et al.* 1987; Jackson *et al.* 1988; Rasmusson *et al.* 1993)) act as repressors.

The invasion of a new host genome by the *P* family would be expected to create reductions in genetic variability, as repressing transposable elements create selective sweeps at diverse chromosomal locations. Using an estimate of 10^6 for the *D. melanogaster* effective population size and 20 years as the maximum age of a *P*-element site, a prediction can be made of the minimum strength of selection required for a particular *P*-element insertion, initially unique, to reach a given frequency. A *P* element initially occupying a single chromosome, with a selective advantage of s in the heterozygous state, has a probability of increasing to deterministic frequencies of approximately $2s$, and its expected frequency after t generations is $e^{st}/2N$, where N is the population size. This expected frequency is the product of the probability that the element survives and its expected frequency conditional upon its survival, p. Thus $p = e^{st}/4Ns$. Some *P*-element sites, such as that at 1A on the X chromosome (Ronsseray *et al.* 1991) are seen now in some *D. melanogaster* populations at frequencies of around 50%. If the creation of the site was approximately 200 generations ago (corresponding to 20 years with ten generations per year), then this frequency implies a selective coefficient in the heterozygote, s, of 5.8%. The spread of a mutation with such a selective advantage would have the effect of creating a local selective sweep. Consider a locus separated from the site of the *P* element by a recombination frequency r. If $r \ll s$, chromosomes bearing the *P* element would be expected to show recent shared ancestry at this linked locus, and a reduction in molecular variability (Kaplan *et al.* 1989). We seek to demonstrate such a reduction in diversity in chromosomes extracted from English populations, by assessing DNA sequence diversity in adjacent loci.

2. MATERIAL AND METHODS

(a) *Strains*

(i) *C23a*

A pure M strain of genotype In(2R)*Cy sp*^{2}/In(2LR) *Pm*(*bw*^{VI}) *ds*^{33k} *b dp*; In(3LR)*D, cxF ru h*/*Sb*.

(ii) *Muller-5 Birmingham*

An M′ strain, bearing *P* elements with internal deletions and possessing no *P* repression activity.

(iii) *w sn m*

A pure M strain bearing the X chromosomal mutations *white*, *singed* and *miniature*.

(iv) *Harwich*

A wild-type *P* cytotype strain collected at Harwich, Rhode Island.

(v) *RD lines*

Inbred lines with X chromosomes extracted using the Muller-5 technique from a population collected in Lenton, Nottingham, in 1993. The lines fall into two classes. RD♂ lines were started from individual wild-caught males, crossed to Muller-5 females. Individual F_1 females were crossed to Muller-5 males, and wild-type males and females selected and crossed from the F_2 individuals of the latter cross. For the RD♀ lines, wild-caught females were crossed to Muller-5 males, and individual heterozygous Muller-5 daughters crossed with Muller-5 males, and, from the F_2 of this latter cross, wild-type males and females selected and crossed. All lines were then brother–sister mated for at least ten generations. Each RD line is homozygous for a wild-type X chromosome derived from a different wild *D. melanogaster*. The autosomes are a mixture of those from the Muller-5 Birmingham line and those from the original wild-caught fly. The production of the RD♂ lines, but not the RD♀ lines, was potentially *P*–M hybrid dysgenic. However, there is almost no *P* activity in the RD population. Crosses of males from the extracted lines to females from C23a at 29 °C (an A cross (Kidwell 1986)) showed gonadal dysgenesis in only 5 out of 479 F_1 females, notwithstanding the 100% gonadal dysgenesis seen in Harwich ♂ × C23a ♀ controls. *P*-element positions would thus be expected to be stable in the lines, and this was found in re-examination of the lines at intervals.

(vi) *RD extraction lines*

These correspond to a subset of the RD♂ and ♀ lines with their chromosomes II and III replaced by those from the C23a line.

(vii) *North Wootten lines*

These (NW) are a collection of inbred isofemale lines collected at North Wootten, Somerset, kindly supplied by P. Corish of the University of Leicester (Corish *et al.* 1996).

(viii) *Essex lines*

These (E) lines are a collection of strains with extracted X chromosomes from females from a population sampled in Danbury, Essex, in 1995. The methodology was the same as for the RD♀ lines.

(b) *A* test cross*

As in Kidwell (1986), females from 33 RD lines were crossed to males of the strong *P* strain Harwich at 29 °C. The mean level of ovarian dysgenesis in offspring was measured for each strain. These experiments were repeated with ten RD extraction lines.

(c) *Molecular biology*

In situ hybridization to salivary polytene chromosomes was performed, modified from Pardue (1986) using pπ25.1 (O'Hare & Rubin 1983) as a probe labelled by nick translation with biotin. Hybridization was visualized by means of the Vectastain® ABC kit (Vector Laboratories Inc., Burlingame, CA 94010) using diaminobenzidine as a substrate and chromosomes counterstained with Giemsa's Improved Stain. Slides were scored independently by three of us (P.T.J.E., T.E.R. and J.F.Y.B.). DNA was prepared from batches of approximately 20 flies following

Table 1. P-*element positions in all the X chromosomes in the study*

line	*P*-element positions	line	*P*-element positions
RD♂1	3A,5A,6C,10C,11A,16D	RD♀27	16D,19E,19F
RD♂2	1B,1F,3A,12D,16F,19A	RD♀28	2D,6F,7C,10A,11F,12E,12F,17A
RD♂3	1F,3A,9A	RD♀29	1A,6E,7A,9A,10F,12B,13D,19A,19D
RD♂5	1E,3A,6B,6E,9E,12D,12F,17A,19B		
RD♂7	3D,4E,16F,19A,19C	E♀1	1A,3A,3D,10B,11C,13A,13E,18A,18B,19A
RD♂8	10A,18A,19B,19E	E♀2	1B,3A,3F,4B,5D,8E,10E,11A,17A
RD♂13	1A,1C,2A,6E,11A,13A,19A	E♀3	1A,5C,10F,11A,12D,13A,13B,18B,19A
RD♂14	1B,4A,5D,11B,13B	E♀11	1A,1B,4C,5C,6A,9B,10A,12B,12D,12E,13B,13E
RD♂15	1B,3A,5C,7D	E♀12	3A,3C,3F,5A,8B,9B,12B,14B,19A,19C,20A
RD♂16	1B,1E,3A,6B,6F,10A,11A,12A,12D, 18A,19A, 19C	E♀13	2F,3D,4A,8A,8D,10A,11A,19A
RD♂17	1B,2C,9A,17D	E♀14	1C,2C,3A,3D,3E,7D,12D,12F,13E
RD♂19	1B,2C,3C,11B,12A,14A,14B,19A,19C	E♀15	7D,12A,12B,12D,13A,18D,19E,20A
RD♂20	1B,1E,3A,6B,6F,9E,11A,12D,17A,19A, 19D	E♀20	4D,8C,8E,9D,11B,12F,14A,17F,19B
RD♂21	1B,3A,4C,6F,9E,11B,19A,19C	E♀25	6E,9A,9E,12A,12D,14A,20A
RD♂23	1A,1B,3A,5D,6A	E♀26	8A,8B,19E
RD♂25	1B,1F,3A,6F,9B,10A,14B,19A,19C	E♀27	1B,9A,10A,20B
RD♀2	1B,1F,3A,6A,8A,11B,11D,13A,13B	E♀29	2C,6D,9B,10D,10E,12B,12E,13B,18B,19A
RD♀3	6F,10A,19B,19E,20A	E♀30	1A,1B,1F,3A,3C,4F,8A,11A,12F,13A,16D,19A
RD♀4	1B,3A,6A,6D,13E,14A,19A,19D,19F	E♀32	1E,3D,5D,8A,9A,16F
RD♀5	1A,1F,3A,3C,6D,7E,9E,14B,19A,19E,20A	E♀33	4E,8A,11A,11B,12C,19C
RD♀8	6B,9F,10F,12F,14A,16F,19A	E♀35	1A,6E,9B,12C,13A,14B,18A,19A
RD♀9	1F,2C,4D,8B,9C,13D,14B,18C,19B	E♀36	4E,6C,8A,11E,19E
RD♀10	1B,3A,4D,5D,6B,9A,12D,12E,19E		
RD♀11	2E,4B,19A,19D,20A	NW1	1B,5C,5D,6A,9D,10A,10B,12F,13B,18A
RD♀13	1B,3A,7A,8A,9F,11E,12E,19A,19D	NW3	1A,1F,3A,3C,4E,7B,7E,7F,8B,8E,9A, 10A,10B,11A, 14B,17A,18A,19A,19B, 19E,19F
RD♀14	1F,3A,7A,7C,8A,17A	NW4	1A,1F,3A,3B,5D,6A,7C,7E,9A,10A,12A,13D,14A, 15B
RD♀17	1B,1F,2B,3D,5B,8B,12C,12E	NW5	7D,7E,7F,19B
RD♀18	3B,3F,4C,7C,9A,10E,10F,11F,12F,17A	NW6	1A,3B,9A,9B,10A,11B,14B,17A,17F,18B,19A
RD♀24	1B,5D,6C,10D,17F,19A	NW8	1A,1F,3A,3C,5D,6A,6C,6E,9A,10A,10B,11A,11F, 14A,17A,17F,19A,19C,19E
RD♀25	1F	NW10	1B,1C,1F,2C,13B,16A,18B

Ashburner (1989) and PCR amplification carried out using a Perkin Elmer Cetus thermal cycler. PCR products were purified for sequencing using spin columns (Qiagen) and sequenced on both strands using an Applied Biosystems 373A automatic sequencer.

RD extracted lines were tested for the presence of *P* elements on chromosomes II and III by crossing males from the extraction lines to females of the pure M *w sn m* line. PCR was carried out on F_1 males as above using the *P*-element primers JB6 and JB7 (Badge & Brookfield 1998) and the primers for the *period* locus as a positive control. Six extracted RD lines were also tested for *P* elements on the X chromosome using JB6 and JB7.

(d) *Simulations*

A population of effective size N was defined as containing 100 lineages of equal frequency at a locus A. Into one of the N individuals a *P* element was introduced at a locus separated from A by recombination rate r, generating a heterozygous or hemizygous selective advantage of s. Wright–Fisher sampling occurs each generation, and recombination can introduce the *P* element into different lineages at A from that in which it first inserted. If the *P* element rises to a frequency of $N/2$ within 200 generations, the process is stopped. For such runs the diversity is calculated for *P*-bearing chromosomes and for chromosomes lacking *P*. If the frequency of the ith lineage in the chromosomes bearing *P* is $p_{i,p}$ the diversity is $1-\Sigma p_{i,p}^2$ and a corresponding formula exists for chromosomes lacking the *P* element. The relative diversity is the ratio of these two terms. For 1000 runs the distribution of the relative diversities is calculated. A fuller analytical treatment of this problem can be found in Kaplan *et al.* (1989). In all the simulations shown here, N was 10 000 and s was 0.058. This value of N was less than the 10^6 used above in the calculation of s. However, the s-value calculated is a minimum, and is that required to give an expected frequency of 50% within 200 generations. The simulations are more realistic in that runs are only retained in the expected distribution if, in the run, a frequency of 50% is indeed reached in 200 generations or less.

3. RESULTS

Table 1 shows the positions of all *P*-element sites detected by *in situ* hybridization on the X chromosome in the three populations. For the largest sample, that for RD, the frequency spectrum is compared in figure 1 to that expected from the infinite alleles distribution (Langley *et al.* 1983), with $\theta = 4.98$ (a value of θ chosen to make the expected mean site frequency the same as in the data). The three sites of high frequency (16 out of 33 chromosomes occupied) is more than the 0.57 sites expected with this frequency or higher under this model. The values of θ for the Essex and North Wootten populations are 4.66 and

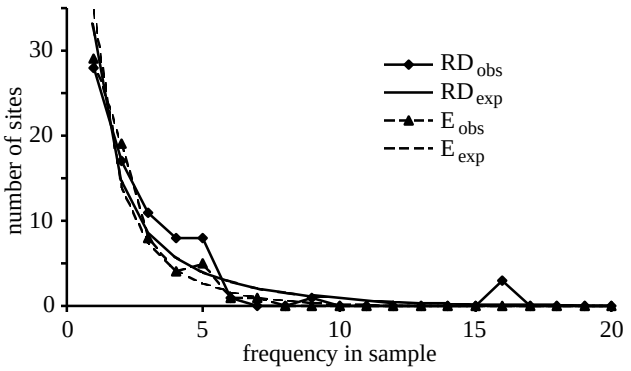

Figure 1. Frequency spectra for the *P*-element sites on the X chromosomes of the RD and E populations ('RD_{obs}' and 'E_{obs}', respectively), along with their expectations according to the infinite alleles model (Langley *et al*. 1983). The three sites in the RD population with a frequency of 16 out of 33 chromosomes sampled (sites 1B, 3A and 19A) are obvious.

2.04 respectively, and the observed and expected data for Essex are also shown on figure 1. Sites with frequencies greater than 4 out of 33 chromosomes occupied in the RD data were examined for linkage disequilibria. Seven significant linkage disequilibria were detected in Fisher's exact tests: 1B × 3A(+ve*), 3A × 2D(+ve*), 6B × 12D(+ve*), 6F × 10A(+ve**), 6F × 19C(+ve*), 19A × 9C(+ve*), 19A × 19D(+ve*). (Here, * is 5% and ** is 1% significance.) Based on the site frequencies and the number of comparisons, one would expect 1.41 significantly positive associations, and 0.72 significantly negative associations. Sites involved in significant linkage disequilibria were more tightly linked genetically (mean map distance 19 cM) than those not involved in significant linkage disequilibria (mean map distance 31.9 cM), but not significantly so.

Table 2 shows the degree of hybrid dysgenesis in A* crosses. The hybrid dysgenesis values are highly non-normal. Thus, Mann–Whitney *U*-tests are used to compare the rank order of strains with or without specific *P* insertions. Previous data (Ronsseray *et al*. 1991) revealed an association between the presence of *P* elements at 1A and repression. We also find such an association ($p = 0.010$). Of the three abundant sites, *P*-element occupancy at each site is associated with a reduced level of hybrid dysgenesis. The corresponding probabilities are 0.171, 0.051, and 0.029 for sites 1B, 3A and 19A, respectively.

No significant correlations between *P* sites on the X are found in the greatly reduced data set using the extracted RD chromosomes. F_1 males from each of these strains failed to amplify any *P* elements in the PCR, notwithstanding amplification of part of the *period* locus, demonstrating the absence of *P* elements on chromosomes II, III and IV in each strain. *P* elements were amplified from the X chromosomes of all strains tested, RD♂1, RD♂2, RD♂7, RD♂8, RD♂13, and RD♂16. In each case an amplification product consistent with the *KP* element was produced, plus, for all strains except RD♂13, between one and four shorter *P*-element sequences. No longer *P* elements were amplified.

The spread of an advantageous allele is expected to have an impact on genetic variability in the surrounding chromosome, provided that the selective coefficient greatly exceeds the recombination rate (Wiehe & Stefan 1993). To test if high-frequency *P* sites have reduced DNA sequence variation in flanking regions of the chromosomes, parts of introns and, in one case (*period* 2), an exon, from genes very close to these *P*-element sites were PCR amplified and sequenced. Figure 2 shows the results of sequencing DNA samples tightly linked to the *P*-element sites at 1A (*yellow*), 3A (*white* and *period*) and 19A (*annexin-X*). The genes *period* and *white* are, respectively, 0.2 and 0.5 map units from the site at polytene position 3A. While the level of variability at the *period* locus is less than at *white*, the variation that exists is not in linkage disequilibrium with the presence or absence of the *P* element at 3A. For the sites at 19A and 1A, no variation was found in the introns of the adjacent genes *annexin-X* and *yellow*, respectively, regardless of the presence or absence of the *P* element. This result is consistent with these genes' locations in regions of low recombination (Begun & Aquadro 1992). Table 3 shows the summary of the *white* and *period* data in terms of the gene diversity among chromosomes with and without the *P* element respectively, in each of the three samples. No consistent reduction in π in *P*-bearing chromosomes is seen, even for *period*.

Table 2. *Levels of ovarian dysgenesis (proportion of dysgenic ovaries) for A* crosses using the RD lines*

strain	ovarian dysgenesis in cross A* with Harwich ♂s	ovarian dysgenesis of extracted lines in cross A* with Harwich ♂s
RD♂1	0.850	0.982
RD♂2	0.050	0.667
RD♂3	0.667	—[a]
RD♂5	0.333	—
RD♂7	0.167	0.908
RD♂8	0.217	0.880
RD♂13	0.017	0.870
RD♂14	0.483	—
RD♂15	0.133	—
RD♂16	0.200	0.923
RD♂17	0.950	—
RD♂19	0.367	—
RD♂20	0.483	—
RD♂21	0.000	—
RD♂23	0.033	0.000
RD♂25	0.000	0.013
RD♀2	0.517	—
RD♀3	0.200	—
RD♀4	0.300	—
RD♀5	0.000	—
RD♀8	0.750	—
RD♀9	0.800	—
RD♀10	0.100	—
RD♀11	0.567	0.889
RD♀13	0.000	—
RD♀14	0.650	—
RD♀17	0.617	—
RD♀18	0.167	—
RD♀24	0.583	—
RD♀25	0.933	—
RD♀27	0.617	—
RD♀28	0.367	—
RD♀29	0.050	0.026

[a] Not tested.

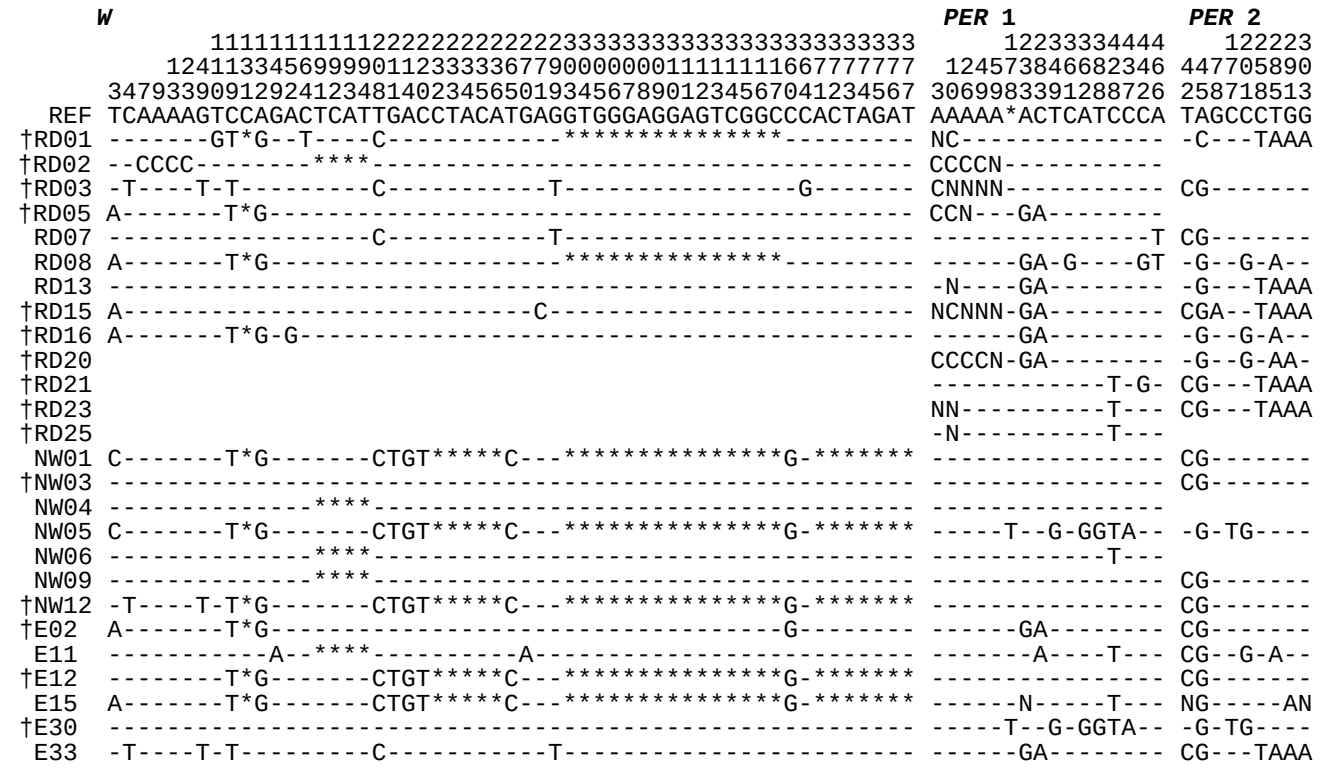

Figure 2. Variable sites at *white* and *period* in samples of X chromosomes from three populations of *D. melanogaster*. The first three rows indicate the positions of variable sites within the sequenced fragment. For *white*, the fragment was bases 7869–8346 of GenBank sequence X02974 (O'Hare *et al*. 1984). For *period* 1 and *period* 2, bases 608–1110 and 3300–3700, respectively, were studied from GenBank sequence M11969 (Jackson *et al*. 1986). Nucleotides identical to the reference sequences (shown in row 4) are indicated by a hyphen. Deleted bases are indicated by an asterisk. Unresolved bases are indicated by N. Strains which carry a *P*-element insertion at cytological position 3A are marked †. Four hundred and eighty-one bases corresponding to positions 2897–3378 of GenBank sequence X04427 (Geyer *et al*. 1986), were sequenced in nine alleles of *yellow* (six from RD, two from NW and one from E). Six hundred and two bases were sequenced in five alleles of *annexin-X* (all from the RD population) corresponding to positions 1180–1782 of GenBank sequence X78323 (Johnston *et al*. 1990). No variation was found at either locus.

Table 3. *Calculated* π *values for* white *and for two regions within* period, *in the three populations RD, NW and E, for chromosomes with and without P-element insertions at 3A*

strain	locus	*P*	π
RD	*period 1*	*P*-bearing strains ($n=10$)	0.0056
		P-free strains ($n=3$)	0.0067
	period 2	*P*-bearing strains ($n=7$)	0.0088
		P-free strains ($n=3$)	0.0092
	white	*P*-bearing strains ($n=6$)	0.0275
		P-free strains ($n=3$)	0.0293
NW	*period 1*	*P*-bearing strains ($n=2$)	0.0000
		P-free strains ($n=5$)	0.0052
	period 2	*P*-bearing strains ($n=2$)	0.0000
		P-free strains ($n=3$)	0.0058
	white	*P*-bearing strains ($n=2$)	0.0795
		P-free strains ($n=5$)	0.0515
Essex	*period 1*	*P*-bearing strains ($n=3$)	0.0093
		P-free strains ($n=3$)	0.0030
	period 2	*P*-bearing strains ($n=3$)	0.0050
		P-free strains ($n=3$)	0.0075
	white	*P*-bearing strains ($n=3$)	0.0509
		P-free strains ($n=3$)	0.0642

This result is unexpected, since figure 3 shows the distributions of expected relative gene diversities for *P*-bearing and non-*P*-bearing chromosomes predicted to be induced by a selectively advantageous allele ($s=0.058$) introduced into a finite population for recombination rate, r, of 0.2 map units, corresponding to the genetic map distance between 3A and *period*. The vast majority of simulations show a very great reduction in π in the *P*-bearing chromosomes relative to those lacking *P* elements.

4. DISCUSSION

Paradoxically, there is thus no evidence that the presence of high-frequency *P*-element sites has reduced flanking sequence variability. One possible explanation is that high-frequency sites result not from selection but from insertion hotspots. This is not consistent with the repression associated with some such sites. Furthermore, significant variation between populations argues that high frequencies do not result from multiple insertions into 'hot spots', which would be expected to show similar frequencies in all populations (figure 4). In addition, while there is evidence for hot spots for *P*-element insertion, such as the X-chromosomal *singed* locus, which has a mutation rate of between 0.1 and 1% in hybrid dysgenic conditions, this rate, measured at the level of the chromosomal division, does not, in itself, create a particularly high rate of insertion in the division, 7A. The rate of insertion into the X chromosome as a whole is around 0.5–1.0 insertions per generation, corresponding to a mean rate of 0.4–0.8% of

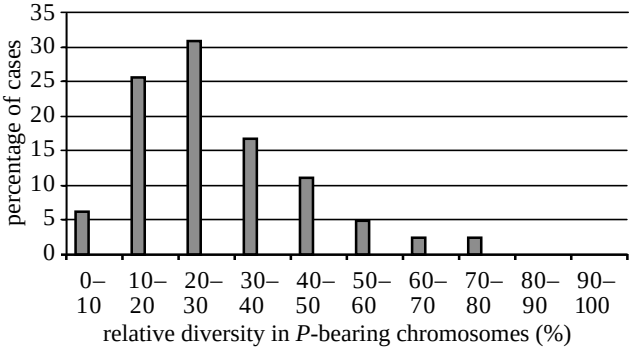

Figure 3. Expected values of the sequence diversity flanking a *P* insertion if there is constant selection of strength 0.058 and recombination of 0.002 in a population of size 10 000. The values are based on 2000 simulations, in 209 of which the *P* element rose to a frequency of 50%. (One case in which this did not happen within 200 generations is not included here.) The calculation assumes that the *P* element is initially introduced into a single chromosome, and can be spread by recombination into other lineages.

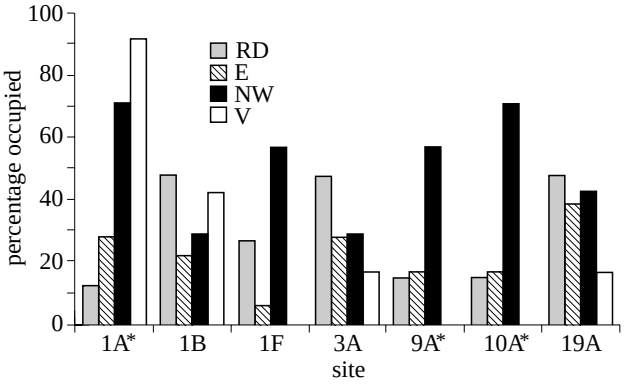

Figure 4. The frequencies of the seven most frequently occupied sites in the X chromosome in the data from RD, E, NW populations, plus the site at Valance (France) studied by Biémont *et al.* (1994). 1A, 9A and 10A show significant variation in frequency between samples.

an insertion per generation per chromosomal division (Engels 1989).

It is also unlikely that high-frequency sites have arisen through a founder effect in expanding local populations, since that would predict a reduction in gene diversity in these samples, which is not seen.

An alternative explanation is that selective sweeps occur in the context of chromosomes undergoing a process of background selection. Linked deleterious sites might be particularly abundant in these populations due to harmful *P*-element insertions. Such weakly deleterious sites are likely to create selection for recombination between the advantageous site and flanking DNA sequences. To test this, the simulations were repeated with the assumption that the differing chromosomal lineages of locus A differed in fitness. Twenty per cent had a fitness of 0.96, 20% had 0.97, 20% had 0.98, 20% had 0.99, and the remaining 20% had a fitness of 1.00. Figure 5 shows the distribution of relative diversity values when advantageous *P* elements ($s = 0.058$) are initially introduced into chromosomes with these differing fitnesses.

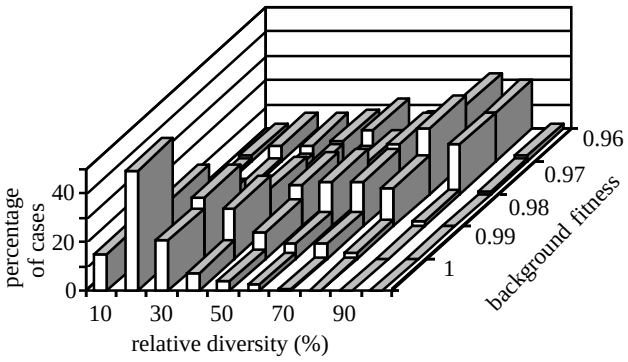

Figure 5. A repeat of the simulation shown in figure 3, but now allowing the advantageous *P* element to initially fall into a chromosome of fitness 0.96, 0.97, 0.98, 0.99 or 1.00. The last two are based on 2000 simulations, the 0.98 and 0.97 cases are based on 3000 simulations, and the 0.96 case is based on 4000. The proportions of simulations in which the element rose to high frequencies are, respectively, 6.4, 7.4, 9.3, 13.3 and 13.3%, but in the first four cases a frequency of 50% in the population was not always attained within 200 generations, and the figure only includes simulations in which it was. (The proportions of cases of spread in which this frequency was reached in 200 generations were 53, 79, 91 and 98% for the 0.96, 0.97, 0.98 and 0.99 classes, respectively.)

When *P* elements insert into low fitness chromosomes and subsequently rise to 50% frequency, their diversity relative to chromosomes lacking *P* is typically close to 1.00, as is seen in the data. Selection has favoured the recombinant chromosomes in which the *P* element has been moved into a more favourable background. Selective accumulation of recombinants between two selected loci recalls the hitchhiking model for recombination (Strobeck *et al.* 1976). It is surprising that all of the three high-frequency sites seen here (plus the 1A site with high occupancy in French populations), are in low recombination regions. The requirement, modelled here, of recombinational 'escape' from deleterious flanking mutations would predict that high-frequency sites should be commoner in high recombination regions. While a general accumulation of transposable elements in low recombination regions has been seen (Langley *et al.* 1989), the force that is thought to explain this, which is the reduction in ectopic recombination in these regions, would be insufficiently powerful to explain our observation.

Many extracted RD chromosomes fail to repress transposition, including some bearing the 'repressor site' 19A. All significant changes in repression between an RD line and its extraction line were in the direction of the RD line repressing less, presumably due to the removal of repressing *P* elements from the autosomes. Some sites, such as 19A, may only function as part of a synergistic repression system, as shown by Badge & Brookfield (1998). X chromosomes would have been found in the wild in the genetic background of repressing autosomal sites and selection in favour of repression at 19A could thus have been effective. The production of extracted lines should be followed by the recombinational isolation of individual repressing sites, which will allow the identification of the flanking sequences of individual insertions. These will permit the demonstration of whether sites apparently shared by

different chromosomes truly are identical, and thus of shared descent.

The authors thank the UK Natural Environment Research Council for support (Research Studentship 94/387 to R.D., Research Grant GR3/9158 to J.F.Y.B.).

REFERENCES

Andrews, J. D. & Gloor, G. B. 1995 A role for the *KP* leucine zipper in regulating *P* element transposition in *Drosophila melanogaster*. *Genetics* **141**, 587–594.

Ashburner, M. 1989 In Drosophila: *a laboratory manual*, pp. 108–109. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Badge, R. M. & Brookfield, J. F. Y. 1998 A novel repressor of *P* element transposition in *Drosophila melanogaster*. *Genet. Res. Camb*. **71**, 21–30.

Begun, D. J. & Aquadro, C. F. 1992 Levels of naturally occurring DNA polymorphism correlate with recombination rates in *Drosophila melanogaster*. *Nature* **356**, 519–520.

Biémont, C., Lemeunier, F., Garcia Guerriero, M. P., Brookfield, J. F., Gautier, C., Aulard, S. & Pasyukova, E. G. 1994 Population dynamics of the copia, mdgl, mdg3, gypsy, and *P* transposable elements in a natural population of *Drosophila melanogaster*. *Genet. Res. Camb.* **63**, 197–212.

Black, D. M., Jackson, M. S., Kidwell, M. G. & Dover, G. A. 1987 *KP* elements repress *P*-induced hybrid dysgenesis in *Drosophila melanogaster*. *EMBO J.* **6**, 4125–4135.

Braverman, J. M., Hudson, R. R., Kaplan, N. L., Langley, C. H. & Stephan, W. 1995 The hitchhiking effect on the site frequency spectrum of DNA polymorphisms. *Genetics* **140**, 783–796.

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

Charlesworth, B., Morgan, M. T. & Charlesworth, D. 1993 The effect of deleterious mutations on neutral molecular variation. *Genetics* **134**, 1289–1303.

Corish, P., Black, D. M., Featherstone, D. W., Merriam, J. & Dover, G. A. 1996 Natural repressors of *P*-induced hybrid dysgenesis in *Drosophila melanogaster*: a model for repressor evolution. *Genet. Res. Camb.* **67**, 109–121.

Engels, W. R. 1989 *P* elements in *Drosophila melanogaster*. In *Mobile DNA* (ed. D. E. Berg & M. M. Howe), pp. 437–484. Washington, DC: American Society for Microbiology.

Geyer, P. K., Spana, C. & Corces, V. G. 1986 On the molecular mechanism of *gypsy*-induced mutations at the *yellow* locus of *Drosophila melanogaster*. *EMBO J.* **5**, 2657–2662.

Gloor, G. B., Preston, C. R., Johnson-Schlitz, D. M., Nassif, N. A., Phyllis, R. W., Benz, W. K., Robertson, H. M. & Engels, W. R. 1993 Type I repressors of *P* element mobility. *Genetics* **135**, 81–89.

Jackson, F. R., Bargiello, T. A., Yun, S.-H. & Young, M. W. 1986 Product of *per* locus of *Drosophila* shares homology with proteoglycans. *Nature* **320**, 185–188.

Jackson, M. S., Black, D. M. & Dover, G. A. 1988 Amplification of *KP* elements associated with the repression of hybrid dysgenesis in *Drosophila melanogaster*. *Genetics* **109**, 459–480.

Johnston, P. A., Perin, M. S., Reynolds, G. A., Wasserman, S. A. & Sudhof, T. C. 1990 Two novel annexins from *Drosophila melanogaster*. Cloning, characterization, and differential expression in development. *J. Biol. Chem.* **265**, 11 382–11 388.

Kaplan, N. L., Hudson, R. R. & Langley, C. H. 1989 The 'hitchhiking effect' revisited. *Genetics* **123**, 887–899.

Kidwell, M. G. 1986 *P*-M mutagenesis. In Drosophila: *a practical approach* (ed. D. B. Roberts), pp. 59–81. Oxford, UK: IRL Press.

Kidwell, M. G. 1994 The evolutionary history of the *P* family of transposable elements. *J. Hered.* **85**, 339–346.

Kidwell, M. G., Kidwell, J. F. & Sved, J. A. 1977 Hybrid dysgenesis in *Drosophila melanogaster*: a syndrome of aberrant traits including mutation, sterility and male recombination. *Genetics* **86**, 813–833.

Langley, C. H., Brookfield, J. F. Y. & Kaplan, N. L. 1983 Transposable elements in Mendelian populations: a theory. *Genetics* **104**, 457–472.

Langley, C. H., Montgomery, E. A., Hudson, R., Kaplan, N. & Charlesworth, B. 1989 On the role of unequal exchange in the containment of transposable element copy number. *Genet. Res. Camb.* **52**, 223–235.

Maynard Smith, J. & Haigh, J. 1974 The hitch-hiking effect of a favourable gene. *Genet. Res. Camb.* **23**, 23–35.

O'Hare, K. & Rubin, G. M. 1983 Structures of transposable elements and their sites of insertion and excision in the *Drosophila melanogaster* genome. *Cell* **34**, 25–35.

O'Hare, K., Murphey, C., Levis, R. & Rubin, G. M. 1984 DNA sequence of the *white* locus of *Drosophila melanogaster*. *J. Mol. Biol.* **180**, 437–455.

Pardue, M. L. 1986 *In situ* hybridization to DNA of chromosomes and nuclei. In Drosophila: *a practical approach* (ed. D. B. Roberts). Oxford, UK: IRL Press.

Rasmusson, K. E., Raymond, J. D. & Simmons, M. J. 1993 Repression of hybrid dysgenesis in *Drosophila melanogaster* by individual naturally occurring *P* elements. *Genetics* **133**, 605–622.

Ronsseray, S., Lehmann, M. & Anxolabéhерère, D. 1991 The maternally inherited regulation of *P* elements can be elicited by two *P* copies at cytological site 1A on the X chromosome. *Genetics* **129**, 501–512.

Strobeck, C., Maynard Smith, J. & Charlesworth, B. 1976 The effects of hitch-hiking on a gene for recombination. *Genetics* **82**, 547–558.

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

Wiehe, T. H. E. & Stephan, W. 1993 Analysis of a genetic hitchhiking model and its application to DNA polymorphism data from *Drosophila melanogaster*. *Mol. Biol. Evol*. **10**, 842–854.

As this paper exceeds the maximum length normally permitted, the authors have agreed to contribute to production costs.