MOLECULAR AND CELLULAR BIOLOGY, Nov. 1993, p. 7006–7018
0270-7306/93/117006-13$02.00/0
Copyright © 1993, American Society for Microbiology
Vol. 13, No. 11

# P-Element-Induced Interallelic Gene Conversion of Insertions and Deletions in *Drosophila melanogaster*†

DENA M. JOHNSON-SCHLITZ AND WILLIAM R. ENGELS*
*Genetics Department, University of Wisconsin, Madison, Wisconsin 53706*

Received 8 June 1993/Accepted 23 August 1993

**We studied the process by which $w^{hd}$, a P-element insertion allele of the *Drosophila melanogaster white* locus, is replaced by its homolog in the presence of transposase. These events are interpreted as the result of double-strand gap repair following excision of the P transposon in $w^{hd}$. We used a series of alleles derived from $w^{hd}$ through P-element mobility as templates for this repair. One group of alleles, referred to collectively as $w^{hd\text{-}F}$, carried fragments of the P element that had lost some of the sequences needed in *cis* for mobility. The other group, $w^{hd\text{-}D}$, had lost all of the P insert and had some of the flanking DNA from *white* deleted. The average replacement frequencies were 43% for $w^{hd\text{-}F}$ alleles and 7% for the $w^{hd\text{-}D}$ alleles. Some of the former were converted at frequencies exceeding 50%. Our data suggest that the high conversion frequencies for the $w^{hd\text{-}F}$ templates can be attributed at least in part to an elevated efficiency of repair of unexpanded gaps that is possibly caused by the closer match between $w^{hd\text{-}F}$ sequences and the unexpanded gap endpoints. In addition, we found that the gene substitutions were almost exclusively in the direction of $w^{hd}$ being replaced by the $w^{hd\text{-}F}$ or $w^{hd\text{-}D}$ allele rather than the reverse. The template alleles were usually unaltered in the process. This asymmetry implies that the conversion process is unidirectional and that the P fragments are not good substrates for P-element transposase. Our results help elucidate a highly efficient double-strand gap repair mechanism in *D. melanogaster* that can also be used for gene replacement procedures involving insertions and deletions. They also help explain the rapid spread of P elements in populations.**

Repair of double-strand DNA breaks is thought to occur by a gap-filling process in which the missing information is provided by a template with sequences matching those of the broken ends (30, 32, 42). The main steps in this model are as follows: (i) pairing between the broken ends and the template; (ii) strand invasion; (iii) DNA synthesis to fill in the gap; and (iv) resolution of the complex, possibly with crossing over between outside markers.

Transposable elements of the P family (8) have been used to study this process in *Drosophila melanogaster*. Recent evidence indicates that P transposase induces excision of the element, leaving behind a double-strand break (10, 13, 20, 27). According to this model, several alternative templates can be used to repair P-element-induced breaks, including the sister strand, an allele on the other homolog, or a matching ectopic sequence if one is present. A comparison of transposition rates versus reversion frequencies led to an estimate that 85% of the P-element-induced breaks are repaired from the sister strand, thus restoring a copy of the P element to its original site (10). The remaining 15%, which might represent transposition events prior to DNA replication, can result in conversion of the sequences flanking the excision site to a new sequence specified by the template. The lengths of these conversion tracts, which we interpret as the result of gap widening prior to repair, were found to vary from a few base pairs to at least 3 kb (13, 27).

This process has been used as a method for carrying out gene replacement in *D. melanogaster*. A P-element-bearing *white* allele was replaced at high efficiency with an in vitro-modified *white* gene carrying a series of single-base substitutions used as markers (13). We now extend this work to show that the same process can be used for insertions and deletions. We make use of the enhanced frequency of interconversion between homologous *white* genes, as opposed to ectopic sites. Our approach was to screen for events in which the P-element insertion allele, $w^{hd}$, was replaced by one of a series of derivatives of this allele (Fig. 1). We refer to those alleles that carry an immobile P-element fragment as $w^{hd\text{-}F}$ and those alleles with the flanking *white* sequences deleted and no P-element fragment as $w^{hd\text{-}D}$. In most cases, there was no phenotypic difference between $w^{hd}$ and the template allele, necessitating a molecular screen for conversion events. Fortunately, the frequencies were sufficiently great that a large sample could still be obtained for all alleles.

We tested five $w^{hd\text{-}F}$ templates with inserted P fragments of up to 281 bp and three $w^{hd\text{-}D}$ templates with deletions of up to 905 bp of *white* sequences. All were copied in at high frequencies. We observed conversion frequencies in excess of 50% for cases in which there was a perfect match between the template and the presumptive double-strand break. In addition, we found that conversion was unidirectional in the sense that the mobile P element in $w^{hd}$ could be replaced by an immobile P fragment in $w^{hd\text{-}F}$, but the reverse rarely occurred. This result suggests that double-strand breaks are made almost exclusively at the site of the mobile P element.

## MATERIALS AND METHODS

***Drosophila* crosses and stocks.** All *Drosophila* crosses were set up in standard cornmeal-molasses-agar medium at approximately 22°C. Unless otherwise indicated, crosses were done as follows. Virgin females 1 to 7 days after eclosion were mated individually with two or three males and then were transferred to fresh food after 1 week. Progeny emerging between 13 and 28 days after the mating were scored.

The $w^{hd80k17}$ allele (abbreviated $w^{hd}$) is a null allele with a

* Corresponding author. Electronic mail address: wrengels@macc.wisc.edu.
† Paper 3373 from the University of Wisconsin Laboratory of Genetics.

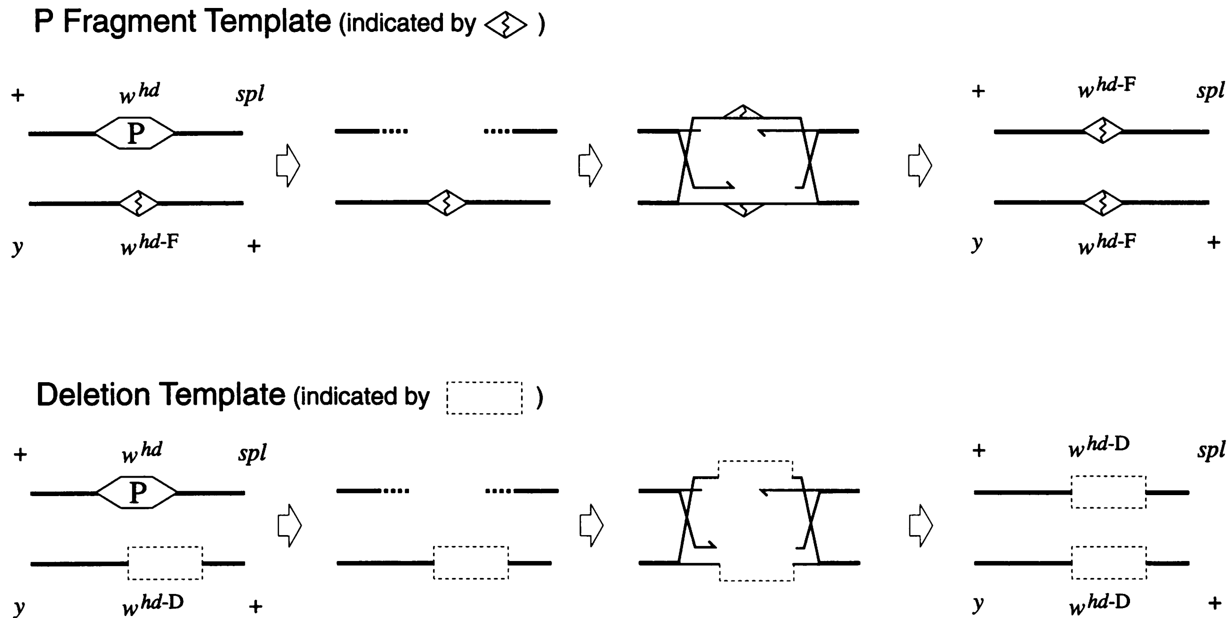


FIG. 1. Mechanism for conversion between *white* alleles. Transposase-induced double-strand break formation occurs with excision of the P element in $w^{hd}$. The break is then repaired with a template allele present on the homolog. The markers *yellow* (*y*) and *split* (*spl*) are linked to *white* by approximately 1.5 centimorgans on each side (23). Conversions are recognized by molecular tests of the $y^+$ *spl* progeny. The template allele was either a P-element fragment generated by internal deletion of $w^{hd}$ (top row) or a flanking deletion (bottom row). These types of $w^{hd}$ derivatives are called $w^{hd\text{-}F}$ and $w^{hd\text{-}D}$, respectively.

629-bp P-element insert flanked by an 8-bp duplication in exon 6 of the *white* gene (34). Other alleles and transposons are either described below or were described by Lindsley and Zimm (23).

**PCR screening.** We extracted DNA from individual flies by macerating each fly in 50 μl of buffer followed by heat treatment as previously described (12). Polymerase chain reaction (PCR) amplification was carried out in 10-μl volumes and amplified over 30 reaction cycles (18). The sizes of the amplified products were determined from agarose gel electrophoresis. We used 4% agarose gels when the expected fragment sizes were less than 250 bp and 2% agarose gels otherwise. Primers p1 through p9 are shown in Fig. 2, and others are defined as they are mentioned.

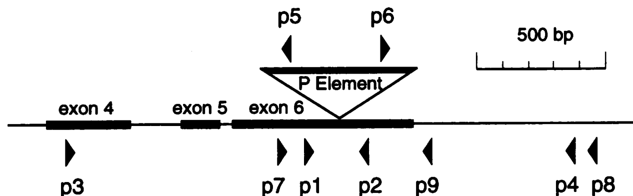


FIG. 2. PCR analysis of $w^{hd}$ derivatives. A portion of the *white* locus is shown with the positions of various PCR primers. Both the *white* gene and the P element are shown in 5'-to-3' orientation, with exons 4, 5, and 6 of the *white* gene shown as heavy lines. The centromere is to the left. The sequences of the primers are as follows: p1, 5' GCACATCGTCGAACACCACG 3'; p2, 5' GTCGG CTACTCCTTGCGTCG 3'; p3, 5' CTGTGGCGGTCCTGGCTGTC 3'; p4, 5' CGAGATCGCGCCAACCAAGTAGCGATAGAT 3'; p5, 5' CGTCCGCACACAACCTTTCC 3'; p6, 5' TCGCTGTCTCACTC AGACTC 3'; p7, 5' GGTTGTCGTACCTCTCATGG 3'; p8, 5' TTA CGAATTCAGGACCTATTTCCGCTGCAC 3'; and p9, 5' GTGTT TTATGTACCGATAAACGAG 3'.

**Mapping conversion tracts.** The restriction site alterations used as markers to determine conversion tracts from the $w^{R812}$ template, one of the $w^{hd\text{-}D}$ alleles, were described previously (13), as was our method for determining them except that primer p4 was substituted for the primer within the deletion of allele $w^{R812}$. Briefly, we extracted DNA from single *Drosophila* males (12), amplified parts of the *white* gene by PCR, and then cut with restriction enzymes (*Nde*II, *Hae*III, and *Rsa*I) corresponding to the marker site changes. The resulting fragment sizes were determined from 4% agarose gels following electrophoresis and ethidium bromide staining.

**Direct DNA sequencing of amplified fragments.** Dideoxy sequencing was performed as previously described (36). For most of the P-fragment derivatives and conversions, we used single-stranded DNA prepared by asymmetric PCR (14) with primers p1 and p2. The sequencing primer was p1. Allele B42.1 was sequenced the same way except that p4 was used instead of p2. For allele A87.2, we used symmetric amplification and double-stranded sequencing after eliminating the primers with PrimeErase (Stratagene). Primers p1 (sequencing) and p2 were used for this allele. Allele B20.3 was sequenced from DNA prepared by amplifying with primers p4 and D+ (5' GATGATTGGACTGCGGGCCG 3') and made single-stranded with the exonuclease gene 6 from bacteriophage T7 (USB) (3). Sequencing was performed from primer p4. The double inserts were amplified asymmetrically with primers p6 and either p5 or 3655 (5' CCTTAGC ATGTCCGTGGGGT 3'). The sequencing primers were p5, p6, or 3654 (5' CGGACGGCTTAATAAGTCCG 3').

## RESULTS

**Isolating derivatives of $w^{hd}$.** The first step was to obtain a series of *white* alleles with lesions in the vicinity of the

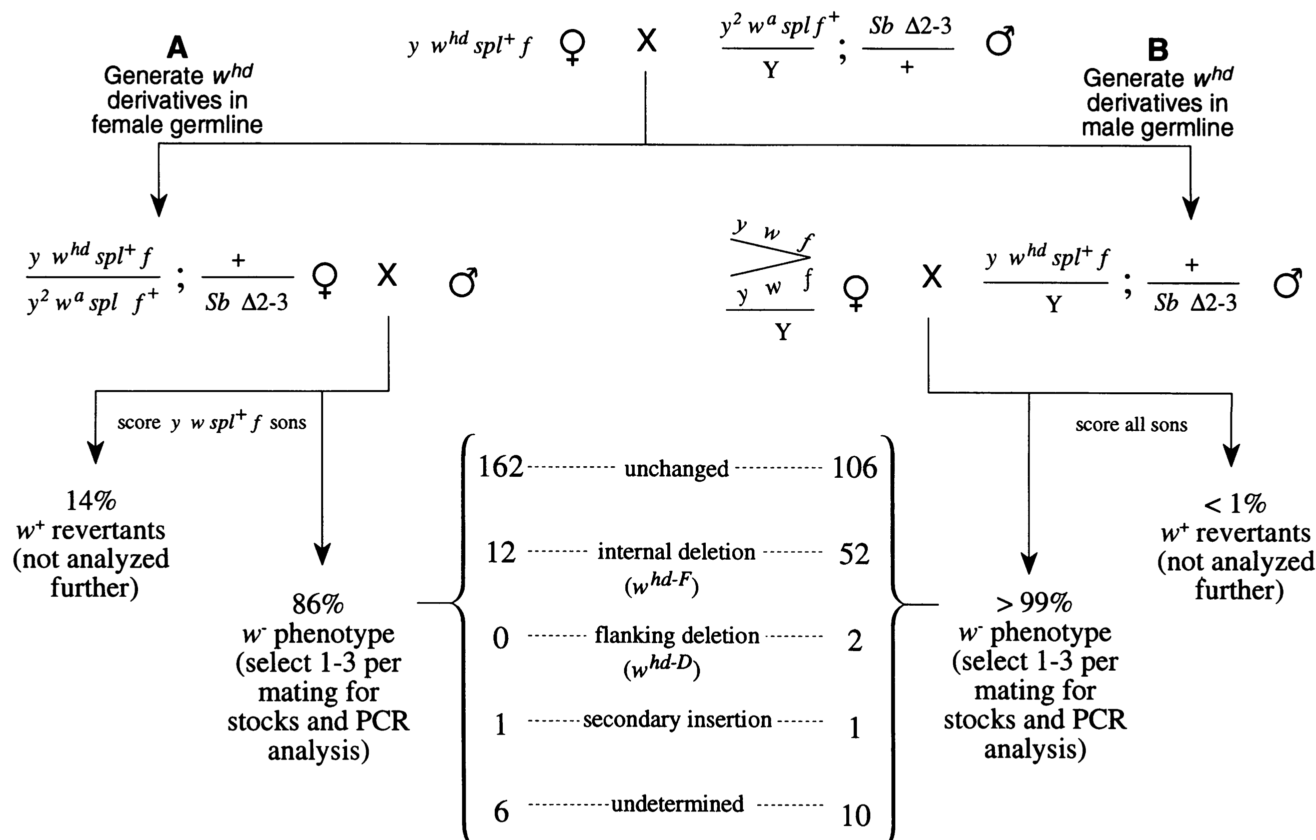


FIG. 3. Crosses to generate $w^{hd\text{-}F}$ and $w^{hd\text{-}D}$ alleles. The $w^{hd}$ allele was combined with the stable source of P transposase, Δ2-3, resulting in a high frequency of P-element excision in the germ line of females (cross A) or males (cross B). Females were mated with their brothers individually in each of 105 crosses of the A type, and 65 males were mated individually in the type B crosses, thus allowing us to distinguish independent events from premeiotic clusters. Heterozygous markers at the *yellow* (*y*), *split* (*spl*), *forked* (*f*), and *Stubble* (*Sb*) loci (23) allowed us to monitor the chromosomes. By analyzing $y\ w\ spl^+ f$ sons, we selected only *white* alleles derived from the parental $w^{hd}$. The $w^+$ revertants were observed and discarded, with their approximate frequencies taken from previous data (10). Compound-X females were used in cross B to produce sons with patroclinous X chromosomes. The same technique was used in the following generation to initiate stocks of $w^{hd}$ derivative alleles from both the A and B crosses. Analysis of the progeny by PCR was done as described in the legend to Fig. 2 and footnote *b* of Table 1.

P-element insertion in $w^{hd}$. We found that a convenient way to do this was to use P transposase to mobilize the $w^{hd}$ element itself and screen the progeny for partial internal excisions of the P element ($w^{hd\text{-}F}$) or deletions of flanking sequences ($w^{hd\text{-}D}$). Both events occur at high frequencies, possibly through aberrant repair following P-element excision. We used the two mating schemes shown in Fig. 3 to mutate $w^{hd}$ in the germ line of females (cross A) and males (cross B). An immobile transposase gene, P[$ry^+$ Δ2-3](99B), abbreviated Δ2-3, was present on chromosome 3 (33). The desired $w^{hd\text{-}F}$ and $w^{hd\text{-}D}$ alleles are expected to have a null phenotype at the *white* locus and hence are not phenotypically distinguishable from the parental $w^{hd}$. It was therefore necessary to use a molecular screen in which each male was tested by PCR with primers p1 and p2 (Fig. 2).

We analyzed 352 progeny to obtain 64 internal deletions of the P element resulting in $w^{hd\text{-}F}$ alleles (Fig. 3). These alleles were identified by amplified fragments whose lengths were reduced by various amounts relative to the $w^{hd}$ allele but were larger than the wild-type fragment. Six independent $w^{hd\text{-}F}$ alleles were selected for further analysis. Sequencing confirmed that each had lost most of the internal P-element sequence but retained at least one of the P termini (Fig. 4). However, only one allele, B41.1, proved to be a simple deletion. Four alleles, B45.2, A87.2, B7.1, and A88.1, had acquired a new sequence of unknown origin at the breakpoint, and one, B2.1, carried sequence alterations in the 8-bp flanking duplication (Fig. 4). Similar structures have also been reported from studies of P-element excision and mutagenesis (37, 43, 45, 46). Note that each of the six P fragments lacks at least some of the *cis*-acting sequences and transposase-binding sites identified from previous work (19, 26). It was expected, therefore, that these elements would be mobilized very rarely if at all by P transposase. This assumption was later verified, as described below.

There were also two cases in which no amplification was seen with primers p1 and p2 but a fragment shorter than the wild-type fragment was amplified with primers p3 and p4. These alleles were candidates for flanking deletions, as was confirmed by sequencing (Fig. 4). One of the deletions, B20.3, overlapped the P insertion point, and the other, B42.1, abutted it. Both carried additional sequences of unknown origin at the breakpoints. A third flanking deletion allele was derived from $w^{hd}$ in a previous study (13). This deletion, designated $w^{R812}$ (Fig. 4), does not overlap the $w^{hd}$ P-element insertion point, and its dark brown eye color

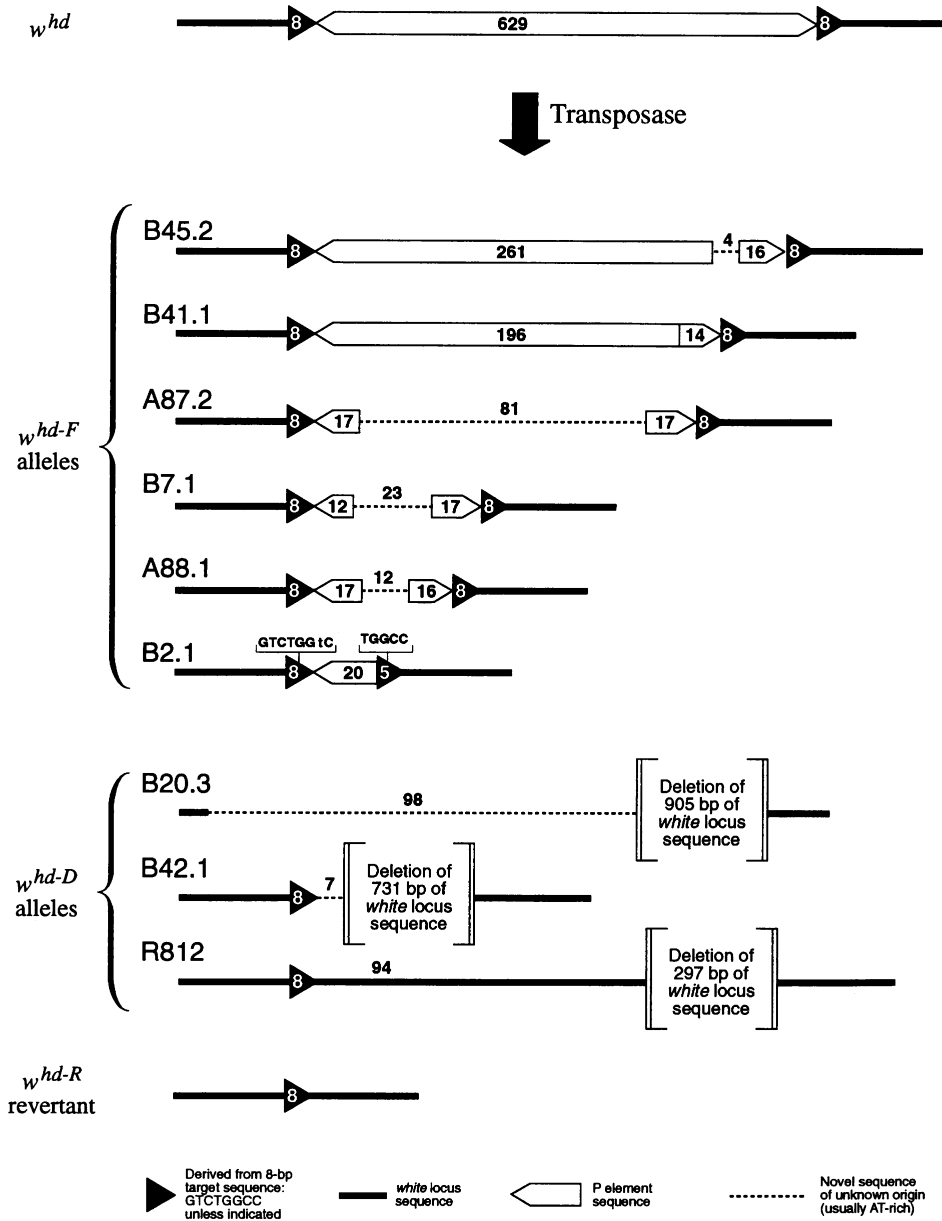


FIG. 4. Structures of $w^{hd}$ and its derivatives. Both the *white* gene and the P element are shown in 5′-to-3′ orientation. Allele B41.1 could also be interpreted as having 197 bp of the 5′ end of the P element and 13 bp of the 3′ end. The ambiguity is due to a common nucleotide at the junction. Similarly, allele B2.1 can be interpreted as having a 23-bp fragment of P and only 2 bp of the right octamer. The deletions in alleles $w^{B20.3}$, $w^{B42.1}$, and $w^{R812}$ remove bp −132 to 773, 5 to 735, and 95 to 391, respectively, from the *white* sequence (coordinates used in reference 13). The novel sequences at the breakpoints of alleles B45.2, B7.1, A88.1, and B42.1 are TTAC, TTGTTATTTGTTGTTATATG TTA, TTGTTATTTGTTGTTATATGTTA, and CAAGATG, respectively. The novel sequences in alleles A87.2 and B20.3 are available (accession numbers L20949 and L20950).

phenotype is easily distinguished from both the wild-type and *white*-null phenotypes.

The remaining 18 alleles listed in Fig. 3 that differed from $w^{hd}$ were not used further in this study. These alleles include two that were later shown to have secondary P-element insertions and 16 whose structures remain undetermined.

Notice from Fig. 3 that $w^{hd\text{-}F}$ and $w^{hd\text{-}D}$ alleles occurred more frequently in the male germ line (cross B) than in the female germ line (cross A). Conversely, revertants were much more frequent in females, where $w^{hd}$ was heterozygous with the $w^a$ (*white-apricot*) allele, than in males where it was hemizygous. These differences can be explained by noting that, in the females, $w^{hd}$ was heterozygous with $w^a$ which has a lesion approximately 2 kb upstream from the $w^{hd}$ insertion point. Previous results (10) confirm that $w^a$ can serve as a template for high-frequency repair of $w^{hd}$ to wild type. Such repair events would explain the high reversion rate in females. These events would also eliminate many double-strand breaks from the pool of possible aberrant repairs that would otherwise result in structural changes at

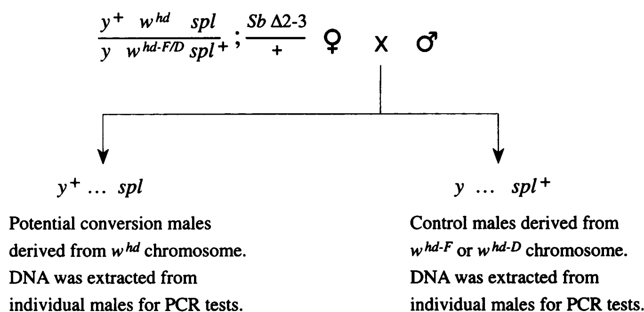


FIG. 5. Cross to screen for conversion events. Conversion occurred in the germ line of $w^{hd}/w^{hd\text{-}F(\text{or } -D)}$ females carrying the Δ2-3 element on chromosome 3 to provide transposase. These females were derived from a cross in which $w^{hd}$ and Δ2-3 came from different parents, thus avoiding mobilization of the $w^{hd}$ element prior to the indicated female. Crosses were set up with individual females mated to their brothers. The sons scored are shown as $y^+$ . . . *spl* or *y* . . . $spl^+$, where the ellipsis designates a *white* allele of unknown structure. Scoring for conversion was done by PCR for all templates except $w^{hd\text{-}R}$ and $w^{R812}$ for which phenotypic screens were available (see text).

$w^{hd}$, thus accounting for the relatively low frequency of $w^{hd\text{-}F}$ and $w^{hd\text{-}D}$ production in these females.

**Method for recovering conversions.** We tested five $w^{hd\text{-}F}$ and three $w^{hd\text{-}D}$ alleles for their ability to serve as a template for conversion. These alleles include all of the derivatives of $w^{hd}$ shown in Fig. 4 with the exception of allele B45.2. We also tested a full revertant of $w^{hd}$, designated $w^{hd\text{-}R}$, which is also shown in Fig. 4. This revertant occurred by a precise excision of $w^{hd}$ described previously (10). Since this reversion occurred in a male, in which there is no X homolog, we can assume that $w^{hd\text{-}R}$ carries only sequences derived from $w^{hd}$, as was confirmed by DNA sequencing for the immediate neighborhood of the $w^{hd}$ P-element insertion site (10).

Each template allele, including the $w^{hd\text{-}R}$ allele used as a control, was tested by the cross shown in Fig. 5, in which the tested template was combined with $w^{hd}$ and a transposase source, Δ2-3, in a heterozygous female. Closely linked markers *yellow* and *split* on either side of *white* were used to identify the chromosomes. All phenotypes were counted, but only the nonrecombinant $y^+$ . . . *spl* sons were scored for conversion. Double recombination in the *yellow-white* and *white-split* intervals is extremely rare, as was subsequently verified in experiments testing the stability of $w^{hd\text{-}F}$ alleles (see below). Note that eliminating the recombinants does not bias the results significantly, since previous work shows that crossing over of flanking markers has no substantial correlation with such conversion events (10, 27). This lack of correlation is also supported by additional data described below. Matings were set up with individual females so that independent events could usually be distinguished from clusters derived from premeiotic events.

We used a phenotypic screen for nonnull alleles, $w^{hd\text{-}R}$ and $w^{R812}$, but for the remaining alleles, it was necessary to use a molecular screen. In those cases, DNA was extracted individually from up to three nonrecombinant sons from each heterozygous female. Amplification with primers on either side of the $w^{hd}$ insertion site produced a fragment to indicate whether that male carried $w^{hd}$, $w^{hd\text{-}F}$, $w^{hd\text{-}D}$, or a different structure, such as a new internal deletion of the P element in $w^{hd}$.

**Conversion frequencies.** We recovered conversions with each of the template alleles. As shown in Table 1, the conversion frequencies fell into three categories according to the structural type of the template. One category consisted solely of the revertant allele, $w^{hd\text{-}R}$, where 18.6% of the $y^+$ . . . *spl* sons were converted, as indicated by reversion to wild-type eye color. This frequency is significantly greater than the rates of 13.2 and 13.6% obtained previously with other *white* templates that were also wild type at the conversion site (10). The difference can be explained by the presence of natural variation between unrelated alleles of *white*, as opposed to $w^{hd\text{-}R}$, which is identical to $w^{hd}$ in the regions surrounding the P insert. Previous work has shown that the conversion frequency between $w^{hd}$ and a template allele is reduced by about 1 percentage point per single-base difference in this region (27).

A second category consisted of the three deletion alleles. They had the lowest conversion rates, ranging from 6.3 to 8.3% of the chromosomes derived from $w^{hd}$. The reduction in frequency relative to $w^{hd\text{-}R}$ can be explained by either assuming that the repair process is not completed properly if one or both endpoints of the gap lies within the deleted portion or by assuming that the deletion allele does not pair as well with its homolog. These possibilities will be discussed below in light of additional data.

The most surprising result was the third category which comprised the five alleles carrying P-element fragments. These templates had remarkably high conversion frequencies. Even when very young females were used, thus reducing the number of cell generations in which conversion can occur, the rate was still nearly twice that of the $w^{hd\text{-}R}$ control. In the remaining cases, the conversion rates exceeded 40%. Possible explanations for these high frequencies will be addressed in experiments described below.

We found little or no enhancement of crossing over associated with the conversion process. Table 2 shows that recombination rates between the markers *yellow* and *split* ranged from 1.58 to 2.81% among all progeny scored, which is consistent with the map positions of these loci (23). There was no detectable correlation between recombination rates and conversion rates, thus agreeing with previous results indicating that conversion is not normally accompanied by crossing over in this system (10, 27).

**Fidelity of conversion for P fragment alleles.** One possible explanation for the elevated frequency of conversion from $w^{hd\text{-}F}$ templates lies in the different methods we used to score the events. The $w^{hd\text{-}F}$ alleles were tested for conversion by PCR, whereas $w^{hd\text{-}R}$ conversions were scored phenotypically. Previous work (10) has shown that phenotypic reversion of $w^{hd}$ indicates precise loss of the P insert. In the PCR screen, however, small deviations from precise conversion events could be misclassified as precise. The conversion rates measured by PCR might therefore be inflated as a result of imprecise events. To test this possibility, we chose a sample of the PCR-detected conversion lines from three of the $w^{hd\text{-}F}$ alleles (B7.1, A88.1, and B2.1) plus the $w^{hd\text{-}D}$ allele, B42.1, for a more detailed analysis.

Two methods were used. The first was DNA sequencing of a PCR fragment carrying the conversion site. For the second method, we constructed a specific PCR primer tailored to each template allele. Each primer was designed to yield amplification of its particular $w^{hd\text{-}F}$ or $w^{hd\text{-}D}$ allele when paired with a suitable primer from the flanking *white* sequence. These primers were not expected to produce amplification from $w^{hd}$, $w^+$, or most near-precise conversion alleles. Each such primer was tested to be sure amplification occurred with its corresponding template allele and not with $w^{hd}$. Only two primers, those for alleles B42.1 and B7.1,

TABLE 1. Interconversion results

| Allele | No. of crosses[a] | No. of progeny | | | Conversion rate ± SE (%)[e] |
|---|---|---|---|---|---|
| | | Analyzed[b] | Converted[c] | Other[d] | |
| P fragments ($w^{hd-F}$) | | | | | |
| B41.1 | 60 | 152 | 63 | 10 | 41.4 ± 5.1 |
| A87.2 | 111 | 211 | 85 | 8 | 40.3 ± 3.6 |
| B7.1[f] | | | | | |
| Normal | 110 | 201 | 106 | 3 | 52.7 ± 3.9 |
| Young | 72 | 167 | 57 | 6 | 34.1 ± 4.1 |
| Old | 72 | 203 | 85 | 12 | 41.9 ± 3.1 |
| A88.1 | 55 | 148 | 76 | 2 | 51.4 ± 4.6 |
| B2.1 | 160 | 440 | 180 | 21 | 40.9 ± 2.4 |
| Deletions ($w^{hd-D}$) | | | | | |
| B20.3 | 104 | 224 | 14 | 31 | 6.3 ± 1.8 |
| B42.1 | 95 | 218 | 18 | 7 | 8.3 ± 2.1 |
| R812 | 288 | 5,595 | 370 | NA | 6.6 ± 0.4 |
| Revertant $w^{hd-R}$ | 132 | 1,743 | 325 | NA | 18.6 ± 1.4 |

[a] Each cross represents the progeny of a single female of the genotype shown in Fig. 5.
[b] Only males with $y^+$ and *spl* markers were analyzed. The $w^{hd-F}$ alleles were tested by PCR with primers p1 and p2 (Fig. 2). For allele $w^{B20.3}$, we used primers p7 and p8. Primers p1 and p4 were used for $w^{B42.1}$. Alleles $w^{R812}$ and $w^{hd-R}$ were scored by phenotype.
[c] For alleles analyzed by PCR, those with a fragment characteristic of the corresponding $w^{hd-F}$ or $w^{hd-D}$ allele were classified as converted. For the $w^{R812}$ and $w^{hd-R}$ crosses, we scored the $y^+$ *spl* sons for eye color phenotype. All $y^+$ *spl* sons from the $w^{R812}$ cross with the dark brown eye color characteristic of $w^{R812}$ were counted as converted. There were also 35 $y^+$ *spl* sons with wild-type eye color. These were not scored as converted but were analyzed separately as discussed in the text. In the case of $w^{hd-R}$, only the $y^+$ *spl* sons with wild-type eye color were counted as converted.
[d] This category was used only for the alleles analyzed by PCR. Any amplified fragment different in size from those expected from either $w^{hd}$ or the specific template allele, either $w^{hd-F}$ or $w^{hd-D}$, was placed in this category. NA, not applicable.
[e] Conversion rates are the ratio of converted sons to the total number of $y^+$ *spl* sons scored. The standard error was computed by a method that is not biased by clusters from premeiotic events (7).
[f] We tested whether the age of $w^{hd}/w^{hd-F}$ females had a substantial effect on the conversion rates with two additional sets of crosses with the $w^{B7.1}$ allele. The females in the crosses labeled young and old were mated 3 to 5 days after eclosion. Progeny produced for the young set were from mothers 3 to 12 days after eclosion. Progeny for the old set were from mothers 17 to 25 days after eclosion. For comparison, the normal group for $w^{B7.1}$ plus all the crosses for the other alleles utilized mothers 1 to 22 days after eclosion.

were successful in these tests (Fig. 6); the others failed to amplify at all and were not used further.

We applied the allele-specific primers to 24 of the putative conversions from allele B7.1, and each reaction produced the fragment of the expected size consistent with a precise conversion event. This conclusion was confirmed for five conversions by DNA sequencing. We also tested 14 conversion lines from the deletion allele B42.1, and all but one amplified as expected. DNA sequencing revealed that the one exception had an A→T base substitution at the position indicated in Fig. 6, but the deletion itself had been converted accurately. Interestingly, this A→T substitution results in a 10-bp sequence (overlapping the deletion breakpoint) that matches the first 10 bp of the P-element sequence. The event is consistent with a model discussed below in which P-element excision occurs by a staggered cut.

An additional five putative conversion alleles from A88.1 and six from B2.1 were also sequenced across their conversion sites. All of the B2.1 derivatives and all but one of the A88.1 derivatives carried a precise copy of the correspond-

TABLE 2. Crossover between *yellow* and *split*[a]

| Allele | No. of crosses[b] | No. of progeny | | Recombination rate ± SE[e] (%) |
|---|---|---|---|---|
| | | Nonrecombinant[c] | Recombinant[d] | |
| B41.1 | 62 | 639 | 12 | 1.84 ± 0.527 |
| A87.2 | 124 | 2,521 | 48 | 1.87 ± 0.267 |
| B7.1[f] | 114 | 1,180 | 19 | 1.58 ± 0.361 |
| A88.1 | 40 | 588 | 17 | 2.81 ± 0.672 |
| B2.1 | 157 | 2,738 | 56 | 2.00 ± 0.265 |
| B20.3 | 141 | 2,341 | 56 | 2.34 ± 0.309 |
| B42.1 | 115 | 938 | 18 | 1.88 ± 0.440 |
| R812 | 289 | 11,234 | 278 | 2.41 ± 0.143 |
| Revertant | 132 | 3,201 | 70 | 2.14 ± 0.253 |

[a] Data are from the same crosses as Table 1.
[b] Number of crosses differ from those of Table 1, because some crosses were scored for recombinants but no sons were analyzed for conversion. Other crosses were scored only for conversion.
[c] Progeny of type $y^+$ *spl* or *y* $spl^+$.
[d] Progeny of type $y^+$ $spl^+$ or *y spl*.
[e] Standard errors were computed from the binomial distribution because there was no evidence for clustering.
[f] Only crosses from the normal age group (Table 1) were included.

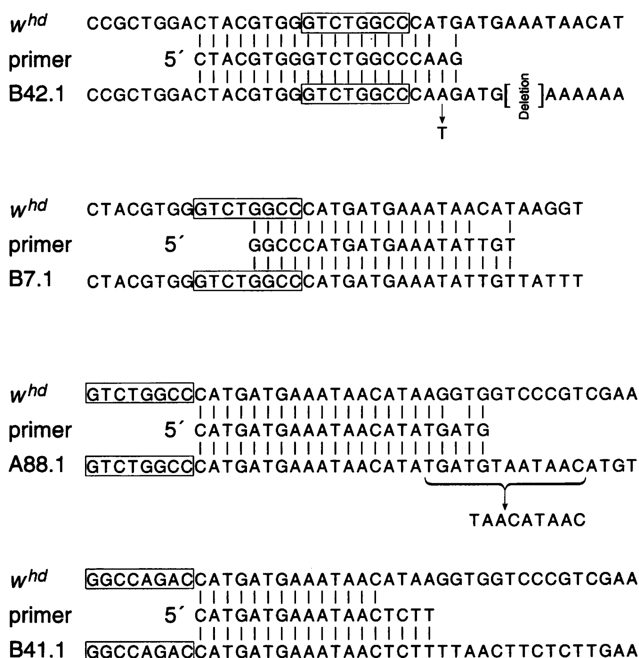


FIG. 6. Allele-specific primers. Four primers were tested for specific amplification. The primers for alleles $w^{B42.1}$ and $w^{B7.1}$ were tested in conjunction with primers p4 and p9, respectively (Fig. 2). We tested the specific primer for allele $w^{A88.1}$ in conjunction with both p4 and p9. The primer for allele $w^{B41.1}$ was tested with p1 and three others not shown in Fig. 2, but whose sequences are 5′ GAAAGGTTGTGTGCGGACGA 3′, 5′ GCCGAAGCTTACCGAA GTAT 3′, and 5′ GTCTGCTGATCAACCAATGG 5′. Each primer is shown in comparison with $w^{hd}$ and its corresponding allele. Changes detected in conversions from alleles $w^{B42.1}$ and $w^{A88.1}$ are indicated by arrows. The orientation of the *white* gene is shown as 5′ to 3′ for $w^{B42.1}$, $w^{B7.1}$, and $w^{A88.1}$, and reversed for $w^{B41.1}$. As noted in the text, only the allele-specific primers for $w^{B42.1}$ and $w^{B7.1}$ gave the expected specific amplification products, and the other two were not used.

ing P fragment. In the one exception, the 12-bp sequence of unknown origin that lies between the ends of the P fragment in allele A88.1 had been changed from TGATGTAATAAC to TAACATAAC (Fig. 4 and 6).

To summarize, a total of 49 putative conversion events were analyzed by DNA sequencing, allele-specific PCR, or both. All but two proved to be precise conversion events by these criteria. We conclude that misclassification of imprecise conversion events can be ruled out as a potential explanation for the high conversion rate observed for the P-fragment alleles.

**Excess frequency of unexpanded gaps in conversions from a $w^{hd\text{-}F}$ template.** Another potential explanation for the observed high conversion frequencies for $w^{hd\text{-}F}$ templates is that they present a better match to the broken ends than the wild-type allele. The P insertion in $w^{hd}$ is flanked by an 8-bp duplication of genomic sequence that is characteristic of P-element insertions (8, 29). This octamer, GTCTGGCC, is also present on either side of the P fragment in each of the $w^{hd\text{-}F}$ alleles, except B2.1 where the octamer is modified (Fig. 4 and 7B). In addition, each $w^{hd\text{-}F}$ allele has a sequence matching the first few base pairs of one or both of the P-element ends (Fig. 4). Therefore, as shown in Fig. 7A, the ends of the gap correspond more closely to the $w^{hd\text{-}F}$ sequence than to the wild-type sequence. If gap expansion occurs, however, the $w^{hd\text{-}F}$ and wild-type templates should match the broken ends equally well.

One prediction from this model is that conversion of the $w^{hd\text{-}F}$ alleles often occurs without enlargement of the double-strand gap, thus leaving both copies of the octamer at the 3′ ends. Such events can be detected in the case of allele B2.1 in which the duplicated octamers are altered relative to $w^{hd}$. In agreement with this model, we found that in four of the six B2.1 conversions sequenced in the experiment described previously, the C→T substitution was not present in the left octamer (Fig. 4 and 7B). Thus, the conversion tracts did not extend more than 1 bp leftward. They did extend further to the right, as indicated by the 3 missing bp in the right octamer. We then tested an additional sample of conversions of B2.1 to determine whether the C→T substitution was present. Instead of sequencing each one, we made use of a *Hae*III restriction site that is disrupted by this substitution. Each of 114 conversion alleles was amplified with primers p7 and p9 (Fig. 2), cut with *Hae*III, and examined on a 4% agarose gel. We found that 35 alleles had the 209-bp fragment indicating absence of the restriction site and implying that the conversion tract did not extend to the C→T substitution which lies only 2 bp from the P insertion. The rest had the 179-bp fragment indicating that the conversion tract included the C→T substitution site (Fig. 7B).

This result contrasts with previous data in which only 2 of 81 conversion tracts from a wild-type template on the homolog failed to include a marker site within the target octamer (Fig. 3 of reference 27). The difference between the current data and those results is significant at $P = 7 \times 10^{-6}$ by Fisher's exact test. (Note that in order to apply Fisher's exact test, it was necessary to use only independent events. We selected at random one conversion line from each of the crosses used to produce our initial set of 114 conversion lines. The reduced set had 22 without and 62 with the *Hae*III site, which was compared with 2 and 79 from the previous work.)

We conclude that approximately one-third of the conversion tracts for B2.1 failed to extend beyond the 8-bp duplicated site, as opposed to a much smaller proportion when the template was wild type. One way to explain this difference is to postulate that the presence of the P fragment and the 8-bp duplication in $w^{hd\text{-}F}$ alleles makes these alleles better templates for an unexpanded gap. Therefore, at least some of the elevated rates we observed for the $w^{hd\text{-}F}$ alleles might be due to more efficient repair of unexpanded gaps. However, this model probably does not provide a complete explanation for the observed rates. Note that the B2.1 allele does not match the postulated sequence of unexpanded gaps as well as the other $w^{hd\text{-}F}$ alleles, yet it has a high conversion frequency (Table 1).

**Conversion tracts for a deletion template.** One of our deletion alleles, $w^{R812}$, carried 7 of the marker sites that had been copied in from the template, P[walter], at the time of the deletion formation (13). These sites allowed us to analyze the conversion tracts in more detail. As indicated in Table 1 for the $w^{R812}$ crosses, we recovered 370 offspring with the dark brown eye color phenotype indicating that the $w^{R812}$ deletion had been copied onto the $w^{hd}$ chromosome. From these offspring, we selected 39 independent lines to determine which of the seven marker sites were present. The results are shown in the Fig. 8A. All 39 tracts included the marker sites at nucleotide positions 1, 28, 72, and 82 as expected, since these sites lie between the P insertion of $w^{hd}$ and the left end of the deletion of $w^{R812}$. The remaining three sites, −393, −135, and −24, can be interpreted as indicating

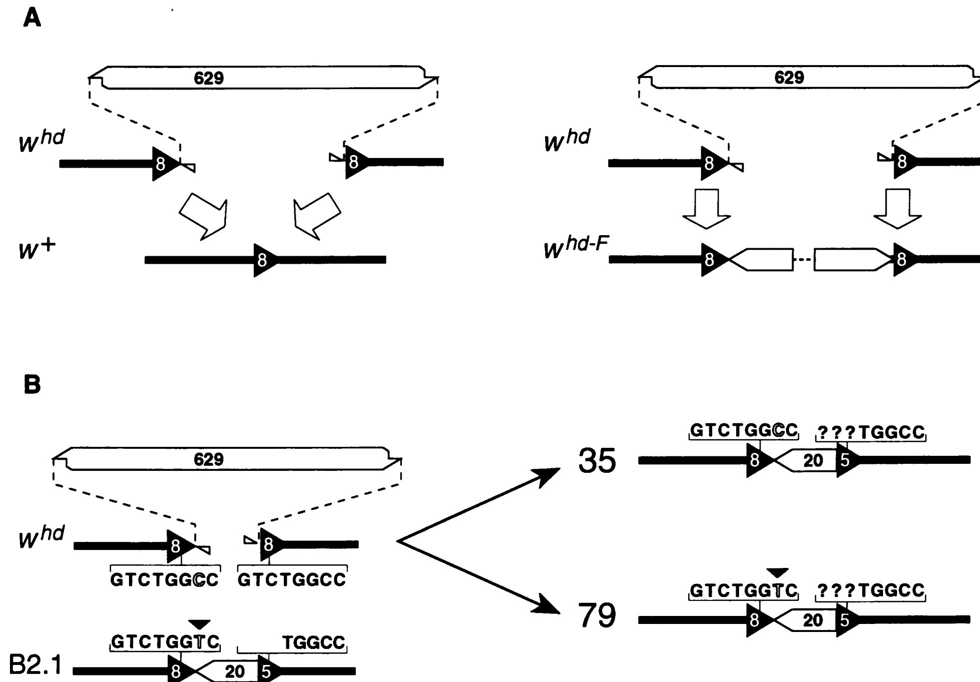


FIG. 7. Repair of unexpanded gaps. (A) If excision of the P insert of $w^{hd}$ leaves both copies of the duplicated octamer, only one end of the break can pair with a $w^{+}$ allele, but both can pair with $w^{hd\text{-}F}$. The excision is drawn to suggest a staggered cut within the P element, but the exact nature of the cut is not known. Such a staggered cut would give an additional advantage to the $w^{hd\text{-}F}$ templates relative to the wild-type allele. (B) The C→T base substitution in allele $w^{B2.1}$ (▼) provides a means of testing whether a substantial fraction of conversion tracts fail to extend through the left octamer when both copies of the duplicated sequence is present. We did not test whether the 3-bp deletion in the right octamer of $w^{B2.1}$ was present in the conversion tracts. The 35 conversion tracts without the C→T substitution and the 79 with it were determined by *Hae*III restriction mapping.

the left end of the conversion tract, which appears to have a distribution similar to that observed when a wild-type *white* gene was used as a template (13, 27).

In addition to the 370 flies with dark brown eye color, we also recovered 35 with wild-type eye color. We interpreted these as conversion tracts which did not extend to the deletion endpoint, thus restoring a fully functional *white* gene. Analysis of the marker sites (Fig. 8B) confirms this interpretation. All but four of the lines had at least one of the marker sites converted, and none carried the deletion. Significantly, most of the conversion tracts did not extend to the marker site at nucleotide position 82. We can interpret these events as conversion tracts that have been selected for short extension on the right side. The one tract that did include position 82 apparently ended somewhere within the 13 bp that separate this site from the endpoint of the deletion. Previous work indicates that a short length of homology is sufficient to allow conversion (27). It is also possible that the double-strand gap did not extend beyond position 82, but heteroduplex formation resulted in conversion of that site.

**The $w^{hd\text{-}D}$ and $w^{hd\text{-}F}$ alleles are relatively stable.** In order to determine whether the template alleles were themselves altered during the conversion process, we selected $y$ . . . $spl^{+}$ sons from the cross shown in Fig. 5 and analyzed their *white* alleles by the same PCR procedure as was used for their $y^{+}$ . . . *spl* brothers. We examined a total of 461 such progeny from five $w^{hd\text{-}F}$ alleles and two $w^{hd\text{-}D}$ alleles (Table 3) but found only one son with an allele that differed from the parental template allele. Therefore, changes at the $w^{hd\text{-}F}$ and $w^{hd\text{-}D}$ alleles are much less frequent than changes at $w^{hd}$.

A related question is whether the $w^{hd\text{-}F}$ alleles can serve as sites for transposase-induced chromosome breakage in the absence of $w^{hd}$. The previous experiment suggests that such events will be rare if they occur at all. Therefore, in order to generate a large sample size, we used a phenotypic screen for events in which a $w^{hd\text{-}F}$ allele, rather than $w^{hd}$, could be converted to its homolog. The potential template was $w^{a}$, a nonnull allele which, as mentioned above, has the wild-type sequence in the vicinity of the $w^{hd}$ insertion and can serve as a template for conversion to the wild type. We scored the eye color of $y$ . . . $spl^{+}$ sons from the cross

$$\frac{y \quad w^{hd\text{-}F} \quad spl^{+}}{y^{2} \quad w^{a} \quad spl};\ \frac{\Delta 2\text{-}3}{+}\ ♀ \times ♂$$

in which individual females were mated to their brothers. If a gap were made at $w^{hd\text{-}F}$ and repaired with the $w^{a}$ homolog, the result would be a wild-type or apricot phenotype, depending on the extent of the conversion tract. Because this is a phenotypic screen, it presumably requires a more specific event than could be detected in the previous experiment, but it allows for a much larger sample size. We screened a total of 6,785 sons with the $y$ and $spl^{+}$ markers and found only one with eye pigmentation (Table 4). The sole exception could have been either a conversion at $w^{B45.2}$ or a double recombinant.

The third method was similar to the above except that the $w^{918}$ allele was used instead of $w^{a}$. As described previously (13, 27), $w^{918}$ is a wild-type allele derived by reversion of $w^{hd}$ in the presence of the P[walter] template. It carries 12 single-bp substitutions that allow mapping of any conversion

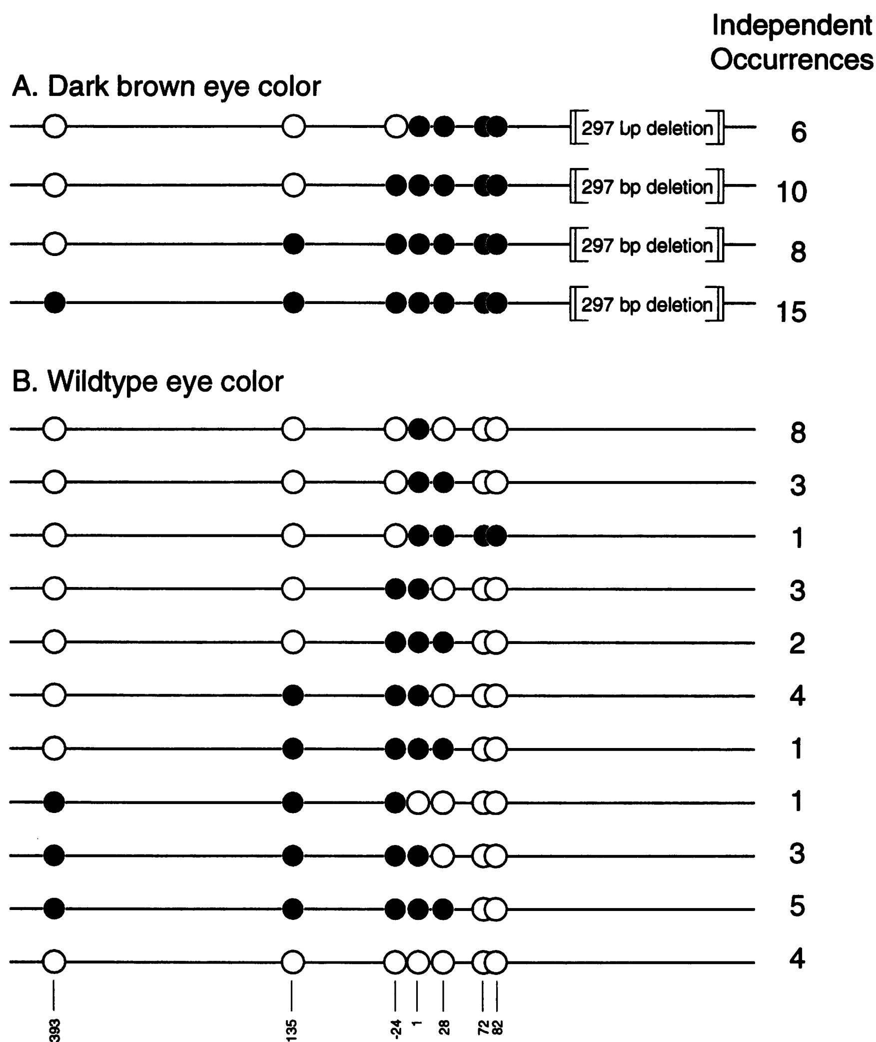


FIG. 8. Conversion tracts from the $w^{R812}$ template. The 1-bp marker sites that differentiate $w^{hd}$ (○) and $w^{R812}$ (●). They were determined by PCR amplification and restriction mapping as described above and elsewhere (13). (A) 39 independent conversion tracts from flies having the dark brown eye color characteristic of the $w^{R812}$ deletion. (B) 35 full revertants.

TABLE 3. PCR tests of template chromosomes[a]

| Allele | No. of crosses[b] | No. of *y . . . spl*$^+$ sons analyzed[c] | No. of sons different from parental template allele |
|---|---|---|---|
| B41.1 | 29 | 52 | 0 |
| A87.2 | 36 | 69 | 0 |
| B7.1 | 42 | 76 | 0 |
| A88.1 | 31 | 60 | 1[d] |
| B2.1 | 31 | 67 | 0 |
| B20.3 | 38 | 63 | 0 |
| B42.1 | 40 | 74 | 0 |

[a] Crosses were the same as in Table 1, but a different class of progeny was analyzed: *y . . . spl*$^+$ rather than $y^+$ *. . spl*.
[b] The numbers of crosses differ from Table 1, because *y . . . spl*$^+$ sons were not tested from every cross.
[c] Individual males were tested by PCR as described in footnote *b* of Table 1.
[d] In the one exception, the $w^{A88.1}$ allele had been replaced by a larger P fragment, presumably derived from $w^{hd}$. Sequencing showed that the new fragment carried 71 bp on the P 3′ side and 18 bp on the 5′ side, with 4 bp of unknown origin between the two P ends.

tracts. We found only four revertants (from two independent events) among 30,147 sons with the *y* and $spl^+$ markers (Table 4). In one cluster of three, the ***white*** allele was found to carry all 12 of the markers from $w^{918}$, as would be

TABLE 4. Progeny of $w^{hd\text{-}F}/w^a$; Δ2-3/+ or $w^{hd\text{-}F}/w^{918}$; Δ2-3/ + females

| Template | Allele | No. of crosses | No. of progeny | |
|---|---|---|---|---|
| | | | *y* $w^{hd\text{-}F}$ $spl^+$ | *y* $w^+$ $spl^+$ |
| $w^a$ | B45.2 | 148 | 3,868 | 1 |
| | B41.1 | 65 | 1,081 | 0 |
| | A88.1 | 119 | 1,835 | 0 |
| $w^{918}$ | B45.2 | 530 | 13,708 | 0 |
| | B7.1 | 374 | 10,046 | 3[a], 1[b] |
| | B2.1 | 288 | 6,389 | 0 |

[a] This event could be a premeiotic double crossover, since all 12 marker sites of $w^{918}$ were present. Two of the three members of the cluster were analyzed in this way.
[b] Only five of the marker sites from $w^{918}$ were present, indicating that this individual represented a true conversion event.

expected for a double recombinant. The other had a true conversion tract with a run of five converted sites (nucleotide positions −135 through 72 in Fig. 8). The most likely explanation is that this conversion tract occurred by repair of a double-strand break produced at the $w^{B7.1}$ allele. Presence of the five marker sites implies that $w^{918}$ served as the template.

We conclude that the conversion events between $w^{hd}$ and either $w^{hd\text{-}F}$ or $w^{hd\text{-}D}$ are primarily unidirectional, with $w^{hd}$ being replaced by its homolog. Moreover, transposase-induced breakage at the $w^{hd\text{-}F}$ alleles is extremely rare, as indicated by only one apparent conversion tract from such a break in a screen of more than 30,000.

## DISCUSSION

**Base substitutions versus insertions and deletions.** The present results show that insertions and deletions can be copied into P-element-induced gaps, thus expanding on previous work (2, 13, 27) in which the template differed from $w^{hd}$ by single-base substitutions only. Prior to this work, however, there were already some suggestions that extra DNA could be copied into the site of a P excision. When we compared the frequency of new P-element insertion sites to the rate of precise loss of the donor element, we found that no more than about 15% of the P-element insertion events could be associated with excision at the donor site (10). We interpreted this result to mean that the majority of P excisions were repaired with the sister chromatid serving as the template, which would result in the donor element being replaced by a newly synthesized copy. This interpretation is equivalent to postulating that extra DNA is routinely copied into the P-element-induced gaps. In addition, Whitely et al. (50) found that an allele of the *escargot* locus which consisted of a 3-kb deletion closely linked to a P-element insertion could revert to the wild type in the presence of P transposase. They interpreted this reversion as a P-induced gap repair event, with the homolog serving as the template. Restoration of the deleted material implies that extra DNA was copied into the gap. More recently, we have shown that an ectopic template can be used to copy both insertions and deletions into the $w^{hd}$ site (27b).

**Gap widening model.** In previous work, the extent of the tracts were considered to be determined by widening of the P-excision-induced gap prior to repair (10, 13, 27). An alternative interpretation is that the conversion tracts are determined by heteroduplex formation, as opposed to gap widening, and uninterrupted runs of converted sites come about through strand-biased mismatch repair of the heteroduplex. The present data favor the gap-widening hypothesis, since heteroduplex formation alone cannot easily account for deletions being copied into the $w^{hd}$ site. However, it is still likely that heteroduplex formation occurs at least in some cases to explain the occasional discontinuous conversion tracts observed (13, 27).

A simple model of gap enlargement can be used to explain the distribution of conversion tract lengths. We assume that each successive nucleotide is cleaved independently with probability $x$, and the endpoint of the gap occurs at a given nucleotide site with probability $1 - x$. The value of $x$ has been estimated at $0.9987 \pm 0.0001$, which leads to a good fit to the data (13, 27). This model can also be applied to the case where extra DNA from the template is being copied into the gap. Accordingly, the proportion of conversion tracts that include the extra sequence is $x^n$, where $n$ is the number of base pairs separating the P-element insertion site from the site of the extra sequence. That is, the gap need only widen by $n$ bases in the given direction to allow the extra sequence to be copied into the gap, regardless of the length of the extra DNA sequence. In the present case, the extra DNA is actually a P-element fragment derived from $w^{hd}$. Therefore, $n = 0$.

However, an additional complication arises when the template contains a deletion relative to the gapped strand. We must then consider what happens when the gap enlarges to an extent such that one of its ends lies within the deleted portion of the template. That end would not match the template sequence. Such an enlargement should occur with probability $x^m - x^k$, where $m$ and $k$ are the numbers of nucleotides between the P-element insertion site and the deletion endpoint proximal and distal to it, respectively. For example, we compute this probability to be 28% when the $w^{R812}$ allele, which has $m = 95$ and $k = 392$, was used as the template, and with $x$ estimated at 0.9987 from previous data (13, 27).

Therefore, we can use our analysis of the conversion events involving the $w^{R812}$ allele to examine several hypotheses concerning the fate of such gaps. First, we can eliminate the possibility that they are repaired with only part of the deletion being copied in. Figure 8 shows that all of the 79 conversion tracts examined carried either the entire 297-bp deletion or no deletion, as opposed to 28% of them having a partial deletion. A second possibility is that these gaps are not repaired, resulting in a dominant lethal event which we would fail to recover in our screen. To test this idea, note that 35 of our 405 total conversion events did not carry the deletion. According to the dominant lethal hypothesis, the expected number is given by multiplying 405 by the conditional probability of a gap endpoint being less than 95 bp from the P-element insertion point, given that it does not lie between 95 and 392 bp from the P insertion point. This conditional probability is $(1 - x^{95})/(1 - x^{95} + x^{392})$, implying an expected number of 66, which is substantially larger than the observed 35. A better fit to the data can be obtained by assuming that gaps ending within the deleted region are trimmed back to a point beyond the distal deletion endpoint where they are homologous to the template. In that case, the proportion of conversions expected to lack the $w^{R812}$ deletion is simply $(1 - x^{95})$, implying an expected number of 47.

For this reason, we favor the idea that gap extension continues until both ends are homologous to the template. A similar process has been observed in *Saccharomyces cerevisiae* (11). However, this hypothesis does not explain why the three deletion templates in Table 1 all had substantially lower absolute frequencies of conversion than that of the revertant. It seems likely, therefore, that some part of the process is inhibited by the presence of a deletion. For example, perhaps the deletion allele does not pair with its nondeletion homolog as readily as a nondeletion allele would.

**There were few changes in the template alleles.** We found that nearly all the changes in $w^{hd}/w^{hd\text{-}F}$ or $w^{hd}/w^{hd\text{-}D}$ heterozygotes occurred on the $w^{hd}$ chromosome rather than the homolog. One trivial conclusion from this result is that double recombination in the *yellow-white* and *white-split* regions is sufficiently rare to allow us to use *yellow* and *split* as reliable markers to distinguish between the two chromosomes.

The result also provides information on whether the P fragments in the $w^{hd\text{-}F}$ alleles can serve as the substrate for P transposase, resulting in chromosome breakage. Previous data (19, 26) suggest that the *cis*-acting requirements for

P-element mobility include their 31-base inverted terminal repeats plus the transposase-binding site, which lies internal to the repeats. It also includes some other sequences near the termini. None of the P fragments tested in Tables 3 and 4 has these requirements. Two of the P fragments, B45.2 and B41.1, lack the requirements on one side of the element, and the other four lack them on both sides.

Recent data, however, bring these requirements into question. Two cases have been reported in which a P element appears to have "captured" a segment of flanking DNA resulting in a "chimeric" transposon. In one case (47), a 703-bp P element transposed in conjunction with a 5.4-kb segment of single-copy genomic DNA terminating in 8 bp identical to the P terminus. Thus, the functional 3′ end of the transposed sequence had only 8 bp of the required *cis* sequences. The second case (16) is similar, except that 15 rather than 8 bp of P-element sequence lies at the terminus and it is the 5′ functional end that lacks the putative P-terminal requirements. Our data in Table 4 suggest a resolution to the apparent contradiction. We find that P fragments lacking the full *cis*-acting sequences can serve as transposase substrate but only at a frequency that is well below what would be needed for detection in the initial experiments by Mullins et al. (26). It should be noted that both of the chimeric elements were reported to excise at high frequencies (16, 47). However, we favor the interpretation that these events come about by excision of only the fully functional P element that lies internal to the chimeric element. Gap widening followed by repair with the homolog as the template would result in a high frequency of events that would appear to be excisions of the entire chimeric sequence. Therefore, these reports are consistent with our suggestion that incomplete P ends can serve as transposase substrate only at very low frequencies.

Finally, our search for changes at the $w^{hd\text{-}F}$ and $w^{hd\text{-}D}$ alleles can address the question of whether the template tends to be altered following gap repair. Previous studies (13) indicate that the template is usually left intact when single-base changes are copied into the site of $w^{hd}$ excision. The present data extend this result to include conversions involving insertions and deletions. However, we cannot rule out the possibility that changes in the template can occur in a small proportion of the repair events. One reason for retaining the hypothesis of a low frequency of template alterations is the one case of an aberrant derivative of allele $w^{A88.1}$ shown in Table 3. Further experiments in which a phenotypic screen allows for a larger sample size are in progress (27a).

**High frequencies of conversion from P-fragment templates.** The most surprising result in these studies was the extremely high frequency of conversion events for the $w^{hd\text{-}F}$ alleles. The frequencies ranged from 34 to 53%, as opposed to the wild-type allele, which was converted at a rate less than 19% (Table 1). We showed that this difference cannot be explained by our selection method. Moreover, it is clear from the above discussion that it is not the result of breakage at the P-fragment sites.

We suggest that at least part of this effect can be explained by a greater propensity of $w^{hd\text{-}F}$ alleles to serve as templates for repair of unexpanded gaps. As indicated in Fig. 7, a double-strand break following P excision is expected to match the $w^{hd\text{-}F}$ allele better than a wild-type homolog. We found that conversions from a $w^{hd\text{-}F}$ template, B2.1, were unexpanded in the leftward direction much more frequently than conversion tracts from an analogous experiment with a wild-type template.

This explanation requires several assumptions. The first is that P-element excision leaves both copies of the duplicated octamer intact. There are several lines of evidence that P elements are excised by double-strand cuts at both ends (5, 10, 13, 20, 27), but the precise nature of the cut is unknown. A study of the excision of plasmid-borne P elements (28) tends to support the assumption that P excision does not remove either copy of the octamer. The authors found that the most common excision products included both copies of the octamer flanking the tetramer CATG, which matches the P-terminal sequence at either end. Thus, the excision could come about by a 4-bp staggered cut within the P ends. Note that the only aberrant conversion from allele B42.1 carried an A→T substitution (Fig. 6) changing the CAAG sequence to CATG. This event is consistent with the invading broken end terminating with CATG, as would be expected if excision occurs by a staggered cut.

A second assumption is that there is opportunity for repair prior to gap enlargement. Results of our analysis of conversions involving allele B2.1 are consistent with approximately one-third of gaps having no leftward expansion. Since the left flanking sequence of B2.1 carries a base substitution, the corresponding frequency for a perfectly matching octamer could be greater than one-third. Similarly, the fact that all of the six sequenced conversions of B2.1 had tracts extending rightward beyond the immediate flanking bases can be explained by the 3-bp deletion from the right octamer of B2.1 (Fig. 4).

Finally, our interpretation requires an assumption that the ends of the gap must match the template for maximum efficiency of repair. The match need not be perfect, as indicated by the B2.1 allele, which has a conversion rate greater than 40% (Table 1) despite altered sequences in both octamers and one missing P-element end (Fig. 4). It is possible that the search for homology is based upon sequences located at some distance from the breakpoint, but successful completion of the repair requires a close match in its immediate vicinity. Recent data from *S. cerevisiae* (11) indicate that normal double-strand gap repair requires prior removal of nonhomologous ends by the *RAD1* gene product. An alternative pathway is used in the absence of *RAD1*. It is not known whether *D. melanogaster* has a similar requirement.

**Implications for homologous gap repair and recombination.** Much has been learned recently about the mechanisms of recombination and double-strand break repair, mostly from studies of bacteria, yeast, and mammalian cells. Current reviews are available (e.g., 15). One difference between the present work and previous approaches is that in our system there is only one template site and it is located on the homolog, as opposed to many previous studies where the two interacting sequences were closely linked within a few kilobases (1, 24, 25, 35, 41, 48) or at least one of the sequences was present in multiple copies on a plasmid (6, 25, 39, 40, 44, 48, 49). This means that the cell's ability to search the genome for a homologous template is likely to be more important in determining the efficiency of repair in our system than in the previous studies. In addition, the effect we postulate for the 8-bp flanking duplication is relevant only to transposon-induced breaks, which have not been used extensively in the previous studies.

**Implications for gene replacement techniques.** Transposon-induced gene replacement has been suggested as a means for manipulating genetic loci in *D. melanogaster* (2, 13) and other species (31), as recently reviewed (38). The present results imply that this method will prove useful for placing

not only base substitutions but also specific insertions and deletions into targeted genomic sites. Furthermore, our data suggest that the most efficient way to insert new sequences into the site of a P insertion is to design the template similar to the $w^{hd-F}$ alleles. In such a template, the sequence to be inserted would be flanked by an 8-bp duplication identical to that surrounding the P element at the target site, and the tetramer, CATG, which matches the P termini, would lie just inside each copy of the duplicated octamer.

**Implications for the spread of P elements in populations.** How did P elements establish themselves in the genome of *D. melanogaster*? They confer no known selective advantage on the organism, and their transposition mechanism is thought to be nonreplicative. One suggestion is that the gap repair process tends to restore a P element to the donor site following transposition, resulting in a net gain of one P-element copy per transposition event (10). However, if an individual is heterozygous for a P element, the template can be the non-P-bearing homologous site, resulting in no net gain in copy number. If transposition occurs after chromosome replication, there will be two potential templates: the sister chromatid and the non-P-bearing homolog. The rate of P-element spread would then depend on the relative utilization of these two potential templates. The results presented above suggest that the sister chromatid would have a significant advantage relative to the non-P homolog. It is analogous to the advantage in utilization we observe for the $w^{hd-F}$ alleles relative to the wild-type *white* allele (Table 1). Consistent with this interpretation, previous results imply that the sister strand is utilized in approximately 85% of the P-element-induced gap repair events, with most of the rest being repaired from the homolog (10). In fact, it is possible that the sister chromosome advantage is actually greater than these estimates suggest, since some of the transposition events might occur prior to replication, when the only suitable template is the homolog.

Kidwell (21, 22) has suggested that P elements may have been introduced into *D. melanogaster* very recently and spread throughout the species only in the last few decades. Several lines of evidence support this view (4, 17), as recently reviewed (9). The extraordinary efficiency with which the $w^{hd-F}$ alleles were used as gap repair templates implies a strong tendency for a P-bearing chromatid, such as the sister strand, to be utilized preferentially, and can therefore help explain this rapid spread.

## ACKNOWLEDGMENTS

We thank James Crow, Carlos Flores, Greg Gloor, Nadine Nassif, and Christine Preston for helpful comments on the manuscript.

This work was supported by public service grant GM30948.

## REFERENCES

1. **Ahn, B.-Y., K. J. Dornfeld, T. J. Fagrelius, and D. M. Livingston.** 1988. Effect of limited homology on gene conversion in a *Saccharomyces cerevisiae* plasmid recombination system. Mol. Cell. Biol. **8:**2442–2448.
2. **Banga, S. S., and J. B. Boyd.** 1992. Oligonucleotide-directed site-specific mutagenesis of *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA **89:**1735–1739.
3. **Bastia, D., J. Germino, J. Grosa, and J. Ram.** 1981. The nucleotide sequence surrounding the replication terminus of R6K. Proc. Natl. Acad. Sci. USA **78:**2095–2099.
4. **Daniels, S., K. Peterson, L. Strausbaugh, M. Kidwell, and A. Chovnick.** 1990. Evidence for horizontal transmission of the P transposable element between Drosophila species. Genetics **124:**339–355.
5. **Daniels, S. B., and A. Chovnick.** 1993. P element transposition in *Drosophila melanogaster*: an analysis of sister-chromatid pairs and the formation of intragenic secondary insertions during meiosis. Genetics **133:**623–636.
6. **Deng, C., and M. R. Capecchi.** 1992. Reexamination of gene targeting frequency as a function of the extent of homology between the targeting vector and the target locus. Mol. Cell. Biol. **12:**3365–3371.
7. **Engels, W. R.** 1979. The estimation of mutation rates when premeiotic events are involved. Environ. Mutagen. **1:**37–43.
8. **Engels, W. R.** 1989. P elements in *Drosophila melanogaster*, p. 437–484. *In* D. Berg and M. Howe (ed.), Mobile DNA. American Society of Microbiology, Washington, D.C.
9. **Engels, W. R.** 1992. The origin of P elements in *Drosophila melanogaster*. Bioessays **14:**681–686.
10. **Engels, W. R., D. M. Johnson-Schlitz, W. B. Eggleston, and J. Sved.** 1990. High-frequency P element loss in Drosophila is homolog-dependent. Cell **62:**515–525.
11. **Fishman-Lobell, J., and J. E. Haber.** 1992. Removal of nonhomologous DNA ends in double-strand break recombination: the role of the yeast ultraviolet repair gene RAD1. Science **258:**480–484.
12. **Gloor, G. B., and W. R. Engels.** 1992. Single fly DNA preps for PCR. Dros. Inform. Ser. **71:**148–149.
13. **Gloor, G. B., N. A. Nassif, D. M. Johnson-Schlitz, C. R. Preston, and W. R. Engels.** 1991. Targeted gene replacement in Drosophila via P element-induced gap repair. Science **253:**1110–1117.
14. **Gyllensten, U., and H. Erlich.** 1988. Generation of single-stranded DNA by the polymerase chain reaction and its application to direct sequencing of the *HLA-DQA* locus. Proc. Natl. Acad. Sci. USA **85:**7652–7656.
15. **Haber, J. E.** 1992. Exploring the pathways of homologous recombination. Curr. Opin. Cell Biol. **4:**401–412.
16. **Heslip, T. R., J. A. Williams, J. B. Bell, and R. B. Hodgetts.** 1992. A *P* element chimera containing captured genomic sequences was recovered at the *vestigial* locus in Drosophila following targeted transposition. Genetics **131:**917–927.
17. **Houck, M. A., J. B. Clark, K. R. Peterson, and M. G. Kidwell.** 1991. Possible horizontal transfer of Drosophila genes by the mite *Proctolaelaps regalis*. Science **253:**1125–1128.
18. **Innis, M. A., D. H. Gelfand, J. J. Sninsky, and T. J. White.** 1990. PCR protocols: a guide to methods and applications. Academic Press, New York.
19. **Kaufman, P. D., R. F. Doll, and D. C. Rio.** 1989. Drosophila P element transposase recognizes internal P element DNA sequences. Cell **59:**359–371.
20. **Kaufman, P. D., and D. C. Rio.** 1992. P element transposition in vitro proceeds by a cut-and-paste mechanism and uses GTP as a cofactor. Cell **69:**27–39.
21. **Kidwell, M. G.** 1979. Hybrid dysgenesis in *Drosophila melanogaster*: the relationship between the P-M and I-R interaction systems. Genet. Res. **33:**105–117.
22. **Kidwell, M. G.** 1983. Evolution of hybrid dysgenesis determinants in *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA **80:**1655–1659.
23. **Lindsley, D. L., and G. G. Zimm.** 1992. The genome of *Drosophila melanogaster*. Academic Press, San Diego, Calif.
24. **Liskay, R. M., A. Letsou, and J. L. Stachelek.** 1987. Homology requirement for efficient gene conversion between duplicated chromosomal sequences in mammalian cells. Genetics **115:**161–167.
25. **Mézard, C., D. Pompon, and A. Nicolas.** 1992. Recombination between similar but not identical DNA sequences during yeast transformation occurs within short stretches of identity. Cell **70:**659–670.
26. **Mullins, M. C., D. C. Rio, and G. M. Rubin.** 1989. *cis*-acting DNA sequence requirements for P element transposition. Genes Dev. **3:**729–738.
27. **Nassif, N. A., and W. R. Engels.** 1993. DNA homology requirements for mitotic gap repair in Drosophila. Proc. Natl. Acad. Sci. USA **90:**1262–1266.

27a. **Nassif, N. A., and W. R. Engels.** Unpublished data.

27b. **Nassif, N., J. Penney, S. Pul, W. R. Engels, and G. B. Gloor.**

Efficient copying of nonhomologous sequences from ectopic sites via P element-induced gap repair. Submitted for publication.
28. **O'Brochta, D. A., S. P. Gomez, and A. M. Handler.** 1991. P element excision in Drosophila melanogaster and related drosophilids. Mol. Gen. Genet. **225:**387–394.
29. **O'Hare, K., and G. M. Rubin.** 1983. Structure of P transposable elements and their sites of insertion and excision in the *Drosophila melanogaster* genome. Cell **34:**25–35.
30. **Orr-Weaver, T. L., and J. W. Szostak.** 1985. Fungal recombination. Microbiol. Rev. **49:**33–58.
31. **Plasterk, R. H., and J. T. Groenen.** 1992. Targeted alterations of the *Caenorhabditis elegans* genome by transgene instructed DNA double strand break repair following Tc1 excision. EMBO J. **11:**287–290.
32. **Resnick, M. A.** 1976. The repair of double-strand breaks in DNA: a model involving recombination. J. Theor. Biol. **59:**97–106.
33. **Robertson, H. M., C. R. Preston, R. W. Phillis, D. Johnson-Schlitz, W. K. Benz, and W. R. Engels.** 1988. A stable genomic source of P element transposase in *Drosophila melanogaster*. Genetics **118:**461–470.
34. **Rubin, G. M., M. G. Kidwell, and P. M. Bingham.** 1982. The molecular basis of P-M hybrid dysgenesis: the nature of induced mutations. Cell **29:**987–994.
35. **Rubnitz, J., and S. Subramani.** 1984. The minimum amount of homology required for homologous recombination in mammalian cells. Mol. Cell. Biol. **4:**2253–2258.
36. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
37. **Searles, L. L., A. L. Greenleaf, W. E. Kemp, and R. A. Voelker.** 1986. Sites of P element insertion and structures of P element deletions in the 5′ region of *Drosophila melanogaster RpII215*. Mol. Cell. Biol. **6:**3312–3319.
38. **Sentry, J. W., and K. Kaiser.** 1992. P element transposition and targeted manipulation of the Drosophila genome. Trends Genet. **8:**329–331.
39. **Shen, P., and H. V. Huang.** 1986. Homologous recombination in *Escherichia coli*: dependence on substrate length and homology. Genetics **112:**441–457.
40. **Singer, B. S., L. Gold, P. Gauss, and D. H. Doherty.** 1982. Determination of the amount of homology required for recombination in bacteriophage T4. Cell **31:**25–33.
41. **Sugawara, N., and J. E. Haber.** 1992. Characterization of double-strand break-induced recombination: homology requirements and single-stranded DNA formation. Mol. Cell. Biol. **12:**563–575.
42. **Szostak, J. W., T. L. Orr-Weaver, R. J. Rothstein, and F. W. Stahl.** 1983. The double-strand-break repair model for recombination. Cell **33:**25–35.
43. **Takasu-Ishikawa, E., M. Yoshihara, and Y. Hotta.** 1992. Extra sequences found at P element excision sites in Drosophila melanogaster. Mol. Gen. Genet. **232:**17–23.
44. **Thomas, K. R., C. Deng, and M. R. Capecchi.** 1992. High-fidelity gene targeting in embryonic stem cells by using sequence replacement vectors. Mol. Cell. Biol. **12:**2919–2923.
45. **Tsubota, S., M. Ashburner, and P. Schedl.** 1985. P-element-induced control mutations at the *r* gene of *Drosophila melanogaster*. Mol. Cell. Biol. **5:**2567–2574.
46. **Tsubota, S., and P. Schedl.** 1986. Hybrid dysgenesis-induced revertants of insertions at the 5′ end of the *rudimentary* gene in *Drosophila melanogaster*: transposon-induced control mutations. Genetics **114:**165–182.
47. **Tsubota, S. I., and H. Dang-Vu.** 1991. Capture of flanking DNA by a P element in *Drosophila melanogaster*: creation of a transposable element. Proc. Natl. Acad. Sci. USA **88:**693–697.
48. **Waldman, A. S., and R. M. Liskay.** 1987. Differential effects of base-pair mismatch on intrachromosomal versus extrachromosomal recombination in mouse cells. Proc. Natl. Acad. Sci. USA **84:**5340–5344.
49. **Watt, V. M., C. J. Ingles, M. S. Urdea, and W. J. Rutter.** 1985. Homology requirements for recombination in *Escherichia coli*. Proc. Natl. Acad. Sci. USA **82:**4768–4772.
50. **Whitely, M., P. Noguchi, S. Sensabaugh, W. Odenwald, and J. Kassis.** 1992. The Drosophila gene *escargot* encodes a zinc finger motif found in *snail*-related genes. Mech. Dev. **36:**117–127.