Targeted Gene Replacement in *Drosophila* via P Element–Induced Gap Repair

GREGORY B. GLOOR, NADINE A. NASSIF, DENA M. JOHNSON-SCHLITZ, CHRISTINE R. PRESTON, WILLIAM R. ENGELS

Transposable elements of the P family in *Drosophila* are thought to transpose by a cut-and-paste process that leaves a double-strand gap. The repair of such gaps resulted in the transfer of up to several kilobase pairs of information from a homologous template sequence to the site of P element excision by a process similar to gene conversion. The template was an in vitro–modified sequence that was tested at various genomic positions. Characterization of 123 conversion tracts provided a detailed description of their length and distribution. Most events were continuous conversion tracts that overlapped the P insertion site without concomitant conversion of the template. The average conversion tract was 1379 base pairs, and the distribution of tract lengths fit a simple model of gap enlargement. The conversion events occurred at sufficiently high frequencies to form the basis of an efficient means of directed gene replacement.

ONE LARGE CLASS OF TRANSPOSABLE ELEMENTS IN EUKARYOtes is defined as having short inverted sequence repeats at their termini. The Ac element in maize, Tc1 in nematodes, and P elements in *Drosophila* are examples of this class, sometimes referred to as Ac-like elements (*1, 2*). These elements are thought to transpose by a DNA-only mechanism, but the details are not understood.

P elements are of interest for several reasons. First, their unusual distribution among strains of *D. melanogaster* allows the study of their population genetics and evolutionary biology. P elements are ubiquitous among natural populations worldwide, but are absent in old laboratory stocks. Comparison of the DNA sequences of P elements in related *Drosophila* species suggests that this dichotomy reflects rapid global spread of P elements within natural populations of *D. melanogaster* over the span of just a few decades (*3*). It is not known how P elements were first introduced into the *D. melanogaster* genome, but the expansion of the species' geographical range as a result of human activity during the present century may be involved.

A second reason for the interest in P elements is their usefulness as molecular tools for manipulating the *Drosophila* genome. For example, P elements are commonly used as insertional mutagens for transposon tagging (*4*) and as mobile reporter genes to identify specific enhancer sequences (*5*). Most significantly, P elements provide the means of transferring cloned sequences into the *Drosophila* germline (*6*). This transformation procedure has been a key technique for many areas of *Drosophila* research, but it has one major limitation: the experimenter has no control over where in the genome the introduced DNA is inserted. Therefore, it has not been possible to replace genes in situ. This limitation is important for genes that are too large to insert into a P element vector and genes whose boundaries have not been determined. In addition, the position effects that accompany random insertions make it difficult to study subtle quantitative differences. A gene replacement procedure would circumvent these problems and allow the creation of mutations at loci with no known genetic variability.

Here we provide evidence for a model of P element transposition that helps explain the highly invasive nature of these elements. In addition, the results provide a method for targeted gene replacement that is efficient at regions near P element insertions.

The gap repair model of P transposition. Results on homolog-dependent P element loss (7) suggest a model of P element transposition in which the element is excised, and moves in a nonreplicative manner, usually in the G2 phase of the cell cycle, leaving behind a double-strand gap at the donor site. This gap is then repaired, most often with the sister strand as a template. Because the sister strand still carries a copy of the P element, the result is replacement of the element at the donor site.

According to this model, a homologous sequence other than the sister strand can sometimes be used as the template. This might happen when transposition occurs in G1 instead of G2. If the individual is heterozygous for the P insertion and the corresponding site on the homologous chromosome is used as the template, the result is precise loss of the P element. The template can also be an ectopic site that matches the sequences flanking the P insertion. Such an ectopic site can be constructed in vitro.

Gene replacement strategy. One prediction of the model is that sequence differences in the flanking regions can be copied from the template to the site of P element excision. Standard models of gap repair (*8*) suggest that such a transfer can happen if the gap is widened by exonuclease activity prior to repair.

The present experiments demonstrate this method of gene replacement and provide a further test of the gap repair model of P element transposition (Fig. 1). The *white* gene, an X-linked locus necessary for eye pigmentation, was our target for the replacement. The mutant allele $w^{hd80k17}$ (abbreviated w^{hd}) carries a small P insertion in one of the exons, which results in bleach-white eyes (*9*). This P element cannot produce P transposase, which is the P-encoded function needed for both transposition and excision (*2*). We supplied transposase by crossing in a second P element on chromosome 3. This element, P[ry^+ Δ2-3](99B) [abbreviated Δ2-3(99B)] is not mobile (*10*) but supplies transposase to mobilize other P elements in the genome.

The third component in this gene-replacement system is an ectopic template for repair with an in vitro–modified sequence to be

G. B. Gloor, Faculty of Medicine, Memorial University of Newfoundland, St. John's, Newfoundland, Canada. N. A. Nassif, D. M. Johnson-Schlitz, C. R. Preston, W. R. Engels, Genetics Department, University of Wisconsin, Madison, WI 53706.

transferred to the *white* locus. For this we used P[*w*alter], an altered *white* mini gene carried on a P element transposon (Fig. 2). The *white* gene in this element has 12 substitutions of single base pairs introduced in a region of ~3 kb. These changes do not prevent expression of white, but each adds or removes a restriction enzyme recognition site, thus providing an efficient way for us to characterize the gene conversion tracts. Many of the alterations were placed close to the site corresponding w^{hd} insertion site to provide maximum information about the conversion tracts, which were predicted to originate at the w^{hd} insertion point. To construct P[*w*alter], we used a pair of complementary oligonucleotides for each site to be altered, then filled in the intervening regions by a polymerase chain reaction (PCR) (*11*) with wild-type *white* as a template. Overlapping intermediate products served as complementary primers for DNA synthesis to produce the final 3-kb segment in a method called recombinant PCR (*12*). This segment was then ligated into the P[w^+] vector (CaSpeR) (*13*).

Germline transformation was carried out as described (*10*), and insertions of P[*w*alter] were identified by eye pigmentation (*14*). The presence of pigmented eyes, which ranged from pale yellow to nearly wild-type depending on the genomic position of the insert, confirmed that the modified *white* in P[*w*alter] was functional. We then mobilized two of the P[*w*alter] inserts in the presence of the transposase-producing Δ2-3(99B) element by crosses similar to those described (*10*). The result was a series of lines, each with a P[*w*alter] insertion at a different genomic position and with *white* phenotypes distinguishable from wild-type. These insertion positions (Fig. 3) were determined by in situ hybridization to salivary chromosomes (*15*). We used these lines to set up crosses in which the male parent carries w^{hd}, one of the P[*w*alter] inserts, and the Δ2-3(99B) element. These elements served as the target site, the template, and the transposase source, respectively. Revertants to full wild-type eye color, designated $w^{hd\text{-}R}$, were selected in the next generation. Figure 4 shows the cross for cases with P[*w*alter] on chromosome 2.

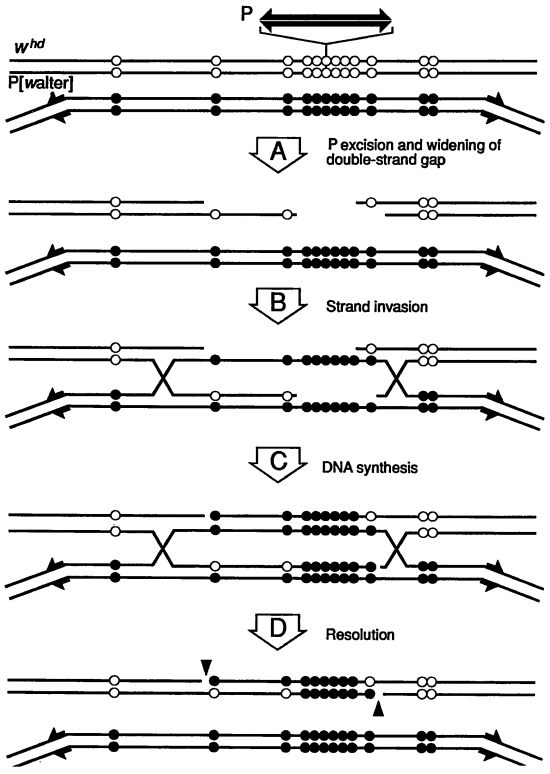


Fig. 1. Experimental design and gap repair model. The double-strand break occurs within the P element insertion allele w^{hd}, and the template is an ectopic *white* gene carried on the transposon P[*w*alter]. The filled and open circles indicate marker sites that differ between the two *white* sequences. (**A**) P element excision leaves a double-strand gap, which is widened to various extents by exonuclease activity. The gap is shown with 3′ overhangs. (**B**) The broken ends and the P[*w*alter] element find each other, and strand invasion is initiated at the overhanging 3′ ends (*8*). (**C**) Polymerization occurs from both broken ends filling in the gap and leaving an intermediate structure with two Holliday junctions. (**D**) The double–Holliday junction intermediate is resolved by the mechanism suggested by Hastings (*29*). The result is a noncrossover product in which the template duplex is unchanged. The repaired duplex has a central region of converted sites flanked by heteroduplex regions and single-strand nicks. The mismatches are later eliminated by mismatch repair or replication, and the nicks (arrowheads) are ligated.

Template-dependent phenotypic reversions. We recovered 1031 males with apparently wild-type eyes among 95,398 progeny scored from 1570 single-male crosses. Approximately 95 percent of the males scored as wild-type were revertants as indicated in the next generation by linkage of eye color to the *white* locus. The rest were presumed to be the result of P[*w*alter] mobilization to produce abnormally strong pigmentation and were not analyzed further. Many of the revertants were recovered as clusters among the progeny of individual males, indicating premeiotic events. The largest cluster was 58 revertants among 74 total males. Results on template-dependent somatic reversions of w^{hd} (7) also show that the reversions are not limited to meiosis.

One of the P[*w*alter] inserts located at cytological position 34C was immobile as indicated by a lack of germline mobility and somatic mosaicism in the presence of Δ2-3(99B). Further analysis by PCR showed that this copy had lost part of its 5′ P element sequence and some of the adjacent 3′ noncoding *white* sequences (*16*). The reversion rate for this line was lower than that of most of the other P[*w*alter] lines. It is not known whether the reduced reversion rate associated with P[*w*alter](34C) was a result of its changed structure or its genomic position.

When the data from P[*w*alter](34C) are excluded, the average reversion rate is estimated at 1.5 percent, which is ~10 times the rate in a similar genotype but without P[*w*alter] present (*7, 17*). Approximately 30 percent of all single-male crosses produced at least one revertant. These values indicate that gaps at w^{hd} are produced at high frequencies and that ectopic P[*w*alter] inserts serve as efficient templates for repair.

The reversion frequencies were dependent on the particular insertion of the P[*w*alter] element (Fig. 3). Some insertions (6F, 100D, 95E1-2) were associated with reversion rates of several times the average, whereas one site (25F) yielded no revertants at all, albeit with a small sample size. The highest reversion rate was associated with a P[*w*alter] insertion on the X chromosome, and other high rates occurred when P[*w*alter] was located in the distal portions of chromosome arms. One possibility is that such insertions can pair more effectively with *white*, which is located distally on the X chromosome. Studies of the three-dimensional arrangement of *Drosophila* polytene chromosomes at interphase (*18*) indicate a "bouquet" pattern in which all centromeres are clustered at one end of the nucleus and all chromosome tips are near the opposite end. With that arrangement the P[*w*alter] inserts near chromosome tips would be physically closer to the *white* locus. In support of this possibility, our data indicate a positive correlation (Kendall's tau = 0.26, $P < 0.05$) between the conversion rates (also in Fig. 3) and proximity to the chromosome tip. However, further data are needed to assess the generality of this correlation.

Conversion tracts. A subset of the revertants was analyzed to determine if any marker sites were copied from P[*w*alter] into the reverted *white*. To do this, we amplified sections of $w^{hd\text{-}R}$ by PCR and digested with the restriction enzymes Hae III, Rsa I, and Nde II (*19*). We performed this analysis on 175 independent revertants and found 123 cases with at least one marker site converted.

Of the 123 conversion tracts, there were 111 (Figs. 5 and 6) with

conversions of at least one of the three central marker sites located near the presumptive origin of the double-strand gap (sites 5, 6, and 7 in Fig. 2), as would be expected from the model (Fig. 1) (*20*). Most of them consisted of a single continuous tract of various lengths (Fig. 5). Such events can be explained by the double-strand gap repair model if we assume the gap is widened to various degrees. No heteroduplex DNA need be postulated to account for the continuous conversion tracts.

In addition to the conversion tracts shown in Figs. 5 and 6, there were 12 revertants that were unconverted for a central segment including sites 3 to 8, but had conversion tracts that flanked this stretch on one or both sides. Examination of the pedigrees showed that each of the 12 exceptional cases came from one of two P[*w*alter] lines with inserts at cytological positions 48D and 95E (Fig. 3). Restriction mapping (*19*) showed that aberrant copies of P[*w*alter] were present at the 48D and 95E insertions sites. We analyzed both the original stocks as well as copies of P[*w*alter] recovered with the reversions and found that marker sites 2 to 8 of the P[*w*alter] insert at 48D had been converted to the normal *white* sequence, and the insert at 95E had conversions at sites 3 to 8. We therefore concluded that these two elements had undergone internal conversion events in a generation prior to the experiment and that the 12 unusual tracts resulted from normal conversion events with these aberrant P[*w*alter] elements as templates. The mechanism of the two internal conversion events is discussed below.

The proportion of reversions with at least one converted site was 69 percent overall, and 82 percent when data from the immobile template P[*w*alter](34C) were excluded from the analysis. Combining this ratio with the total reversion rate for mobile P[*w*alter] inserts (*17*) gives a minimum estimate of 1.3 percent for the conversion frequency.

The estimated conversion rates for the individual P[*w*alter] lines were calculated from the reversion rates (Fig. 3) and the restriction mapping data (*17*). The difference between each reversion rate and the corresponding conversion rate reflects cases in which reversion was not associated with a detectable conversion tract. Some very short conversion tracts might not overlap any of our marker sites, and would go undetected. In addition, a background reversion rate of ~0.1 percent has been observed in a similar genotype but without P[*w*alter] (*7*). Such reversions are apparently template-independent and lack a conversion tract.

We selected 26 of the conversion tracts (Figs. 5 and 6) for confirmation by DNA sequencing (*21*). Among more than 3500 base pairs sequenced, we found only one difference from the sequence predicted by restriction mapping. The exception (Fig. 5) had an A → T substitution at position −25 instead of a T → C at position −24 predicted by the lack of an Nde II site (Fig. 2). This means no P[*w*alter] sequences had been copied to the revertant gene, suggesting that the substitution occurred during a template-independent repair. Thus there is no evidence for misincorporation during template-dependent repair and gene replacement occurred with a high degree of accuracy.

The proportion of conversion tracts that included each of the 12 marker sites is plotted against the marker position (Fig. 7). Site 6, which lies closest to the insertion point of w^{hd}, had the greatest frequency of conversion, as would be expected since the gaps are assumed to originate there. The frequency falls off exponentially and symmetrically in both directions.

The distribution in Fig. 7 can be explained by a simple model of gap enlargement in which exonuclease activity proceeds independently in both directions. After each nucleotide is removed, the enzyme continues to the next nucleotide with probability x or stops with probability $1 - x$. Each step is assumed to be an independent trial with a constant value of x. This model results in the final gap length, and therefore the conversion tract length, having a negative binomial distribution (*22*) whose parameter, x, can be estimated from our data by maximum likelihood (*23*). This model results in an excellent fit to the data with $x = 0.99855$ (Fig. 7).

The same model provides an estimate of the average conversion tract length, $2/(1 - x) = 1379$ bp, which is comparable to that for meiotic conversion in yeast and *Drosophila* (*24, 25*). Our estimation method (*23*) avoids a bias leading to overestimation of the length that was inherent in the previous estimates in which the ends of the conversion tracts were assumed to lie at the midpoints between marker sites.

The homology between P[*w*alter] and w^{hd} is interrupted on one side of the markers by an intron deletion and on the other side by the P element sequence (Fig. 2). Marker site 1 is just 550 bp from the boundary on the left, and site 10 is only 115 bp from the right boundary. Despite such close proximity to nonhomologous sequences, there were 12 conversion tracts that extended to site 1, and 16 that extended to site 10 (*26*). These numbers suggest that very little homology is required for the repair and conversion process.

Complex events and phenotypic changes. We noted that most of the conversion tracts consisted of single contiguous runs of converted sites. However, there were exceptions. We found five discontinuous conversion tracts, consisting in each case of two runs of converted sites (Fig. 6A). Discontinuous conversion tracts can be explained by mismatch repair of heteroduplex DNA (Fig. 1) such that some mismatched sites are repaired to restore the original sequence. They can also arise through multiple gap repair events. The presence of mismatches is thought to promote subsequent rounds of chromosome breakage and repair even when the density of heterozygous sites is much lower than in our system (*27*).

We also recovered two cases of aberrant conversion events detected as incomplete phenotypic reversion of w^{hd} (*28*). One of

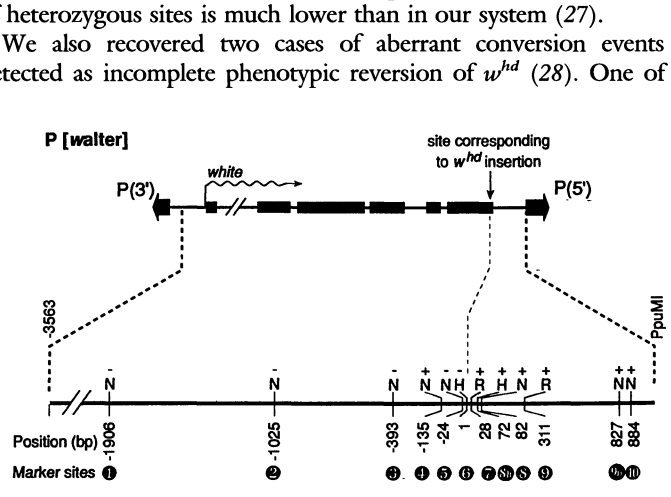


Fig. 2. P element carrying an altered *white* gene used for detecting conversion tracts. Only restriction sites that differ between w^{hd} and P[*w*alter] are shown. N, Nde II; R, Rsa I; H, Hae III. ±, addition or removal of a recognition site. Black boxes indicate exons. We have numbered the base pairs with the origin at marker site 6, which lies within the 8 bp that compose the w^{hd} insertion site. These coordinates can be converted to the published sequence (*41*) by subtracting our positive numbers from −2028 and subtracting our negative numbers from −2029. The substitutions are (from left to right): A → G, A → G, C → T, T → C, T → C, G → A, G → A, A → G, T → G, C → G, T → C, C → A. Two of the substitutions designated 8h and 9n are not within the oligonucleotides used for the recombinant PCR construction. These are probably construction artifacts, as they do not appear in the original sequences. Substitution 8h converts the *white* stop codon to a Trp codon. This change, along with the change in marker site 8, adds three amino acids to the implied polypeptide sequence but has no apparent phenotypic consequence in hemizygous males. The CaSpeR transposon (*13*) was used as the template for constructing the 5′ portion of *white* including the partial deletion of the first intron. The 3′ end, which extends 159 bp beyond what is contained in CaSpeR, was derived from a laboratory stock with a wild-type allele of *white*. The termini (arrows) consist of 503 bp of the P element 5′ end and 223 bp of the P element 3′ end.

these had a deletion of 297 bp just downstream of the coding region and abutting the right end of the conversion tract (Fig. 6B). The other also had a conversion tract, but we found no other sequence differences to explain its phenotype (Fig. 6B).

In addition, we found several duplications and one triplication of part of the *white* sequence. Each extra copy included a 5′ P element end sequence that matched the P[*w*alter] terminus (Fig. 6, C and D). We postulate that the duplications were generated when the left endpoint of a gap extended arbitrarily into the *white* sequences of w^{hd}, but the right end of the gap was not extended and retained some or all of the 31-bp terminal duplication at the 3′ end of the P element originally present in w^{hd}. If the right endpoint then invades the matching sequence at the 5′ P element end of P[*w*alter], the ensuing gap repair process will yield the observed structures. The triplication can also come about by the same process followed by a second round of breakage and repair in which the sister strand serves as the template. If duplications at the other end of P[*w*alter] occurred as well, they were eliminated by our screening method (*19*).

Aside from the two partial revertants (Fig. 6B), all revertant males had fully wild-type eye color. We noticed, however, that some of the revertant alleles displayed a mutant phenotype in females when heterozygous with a null allele of *white*. Specifically, a red-orange eye color with a reduced pseudopupil appeared in revertants in which site 8h (Fig. 2) was converted. Those with no conversion at site 8h had normal eye color even in heterozygous females. The modification in site 8h altered the implied *white* gene product by the addition of three amino acids to the carboxy-terminal end (Fig. 2). Our interpretation is that this altered polypeptide is mildly hypomorphic, but the effect is not visible in males because of dosage compensation. This case illustrates the utility of gene replacement as a technique for studying subtle phenotypic changes that would be masked by position effects if the gene were carried in a transposon.

We also tested the possibility that some of the P[*w*alter] elements had acquired sequences from w^{hd} (backward conversion). This was done by analyzing the restriction sites of P[*w*alter] elements recovered along with the $w^{hd\text{-}R}$ revertants. When P[*w*alter] was autosomal, approximately half the $w^{hd\text{-}R}$ revertants also carried the P[*w*alter] element (Fig. 4). If the gap repair occurred in G1 of the cell cycle, then this copy of P[*w*alter] will always represent the template used for conversion, provided there was no transposition of P[*w*alter]. If repair happened in G2, the P[*w*alter] element will still represent the utilized template in approximately half the cases.

We tested 48 independent P[*w*alter] elements in this way. Six of them came from lines 48D and 95E, and carried the aberrant copies of P[*w*alter] described above. The rest were associated with simple, complex, or partial reversion events (Figs. 5 and 6). In all cases, P[*w*alter] was identical in its restriction map to the original construct, thus giving no evidence of backward conversion. We conclude that the transfer of information was unidirectional in all tested cases.

Gap repair mechanisms. Current models of double-strand gap repair (*8*) involve an intermediate structure with two Holliday junctions flanking the filled-in gap (Fig. 1C). Resolution of each junction is thought to occur by strand breakage and rejoining in one of the two equivalent orientations to yield either a noncrossover or a crossover product, depending on whether the orientations were the same or different in the two junctions. An alternative mechanism (the Hastings model) (*29*), does not require endonucleolytic strand breakage. Instead, the two Holliday junctions are brought together by topoisomerase activity, eventually canceling each other. The resulting products are invariably noncrossover chromosomes with no backward conversion (Fig. 1D).

The Hastings model can account for several of our observations that are difficult to explain otherwise; namely, (i) crossover products were rarely, if ever, produced in the conversion events. Depending on the orientation of P[*w*alter], crossing over often yields inversions or translocations detectable in our experiment as pseudo-linkage or crossover suppressors in generations after F3 (Fig. 4). None was seen. Furthermore, experiments (7) in which w^{hd} is reverted with a template allele on the homologous chromosome indicate only a slight increase in crossing over of outside markers among the revertants as compared to nonrevertants (4.2 compared to 2.6 percent). (ii) The lack of backward conversion in our data is predictable from the Hastings model in which information transfer occurs in one direction only. (iii) Only 4 percent of the conversion tracts in our data were complex events as opposed to 20 percent seen for meiotic gene conversion in yeast (*25*). The rarity of discontinuous tracts can be explained by the nicks on each strand of the repaired chromosome predicted by the Hastings model (Fig. 1D). These nicks could cause a strong strand bias for mismatch repair (*30*) such that each mismatch will be repaired to the original w^{hd} sequence. The result is a single continuous conversion tract. (iv) Internal P element deletions, which occur at high frequencies in the presence of transposase (*2*), can be explained if resolution by the Hastings route happens before the 3′ ends are fully extended. The resulting nonoverlapping ends can then be rejoined preferentially at sites of short fortuitous homology, as has been observed (*2*). (v)

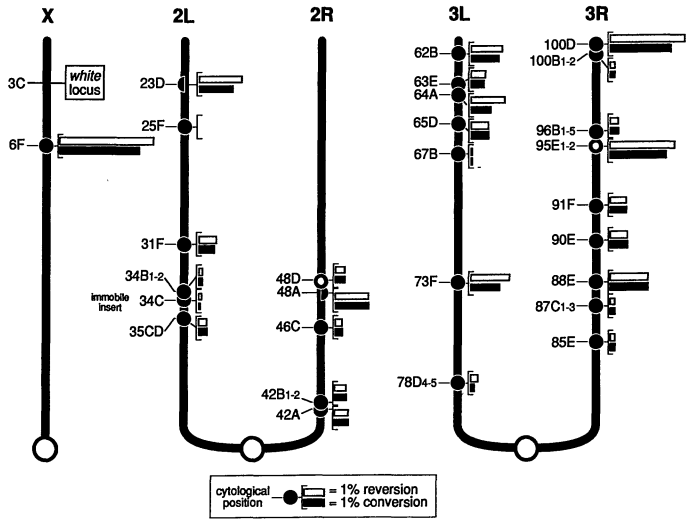


Fig. 3. Cytological positions of P[*w*alter] inserts. The genomic position of each P[*w*alter] insertion was determined by in situ hybridization to salivary chromosomes (*15*). Chromosome arms are labeled at the top of the figure and the centromeres are shown as open circles at the bottom. The reversion frequency indicates the proportion of true phenotypic revertants among the F2 males (see Fig. 4). The conversion frequency is the proportion of F2 males with at least one of the 12 marker sites converted in the revertant *white* allele. The difference between the two values indicates cases of revertants where no conversion tract was detected. The number of progeny scored for each line averaged 3181, and ranged from 313 for line 100B1-2 to 35,800 for line 34C (*42*). An additional 9504 progeny were also obtained from lines whose P[*w*alter] position was not determined. The two P[*w*alter] insertions with internal conversion tracts are shown as open circles. Two of the sites, indicated as semicircles, were present in the same line and were therefore tested together. It cannot be determined which of these two sites was involved in any given conversion event. In addition, it is possible that other lines had sites not detected by in situ hybridization. The insertions at 6F and 34C came from the original transformation procedure (*14*). Lines 25F, 31F, 62B, 64A, 65D, 73F, 78D4-5, 85E, 88E, 90E, 91F, 96B1-5, 100B1-2, and 100D were derived by transposition from the 6F site. The rest were derived by transposition from the site at 3C which was not itself tested due to its close proximity to *white* (*14*).

Finally, a similar explanation can account for the two internal conversions we observed for P[*w*alter] elements at positions 48D and 95E. The first step is transposase-induced excision of the P[*w*alter] element, leaving a double-strand gap to be repaired with the sister strand as the template. The repair complex is then separated when the 3′ ends have been only partially extended, resulting in a P[*w*alter] element with an internal double-strand gap. Repair of this gap with *white* as the template would result in the structures we observed. Several observations (*31*) have been interpreted as internal conversions between two P elements, which could have occurred by the same route. It is not known whether the close proximity between the *white* locus and the progenitor element of the 48D and 95E insertions (*14*) (Fig. 3) was important in generating these events.

In summary, the Hastings model best explains our data. It is possible, however, that resolution by breakage and rejoining might also occur in a minority of cases, thus explaining the five discontinuous conversion tracts (Fig. 6A) as well as P-induced crossing over in males (*32*).

Gene replacement and directed mutagenesis. We have shown that an efficient gene replacement technique is possible for *Drosophila*. It requires the presence of a target site, an ectopic template, and a transposase source in the same individual. In our system, more than 1 percent of all progeny from such flies had conversion tracts from a few base pairs to at least 2790. The probability of a conversion tract including a site n bp in either direction from the P insertion site was $(0.99855)^n$.

Although we used this method to copy base substitutions, it should also be useful for copying insertions or deletions. The size of any converted deletion is limited by the gap size, and the distribution of the gap size is presumed to follow our theoretical curve (Fig. 7). There is no corresponding limit to the length of additional sequences that can be copied into a gap. The only requirement is that the insertion point lies within the boundaries of the gap.

This gene replacement procedure should be applicable to a wide range of genetic problems. However, several complications can arise. First, not all target sites provide a convenient phenotypic selection such as eye color. Fortunately, the high frequency of conversion means that even a much less efficient screen, such as by PCR, should be adequate for recovery of at least a small number of conversion events. Second, some P inserts are less mobile than the one at w^{hd}, and the frequency of conversion will be correspondingly reduced. Such cases will require screening on a larger scale. Finally, when the target site is autosomal, unlike *white*, which is X-linked, steps should be taken to prevent the homologous chromosome from competing with the ectopic copy by serving as the template. This problem can be eliminated in some cases by the use of a deletion heterozygote. When that is not possible, the use of a multiply

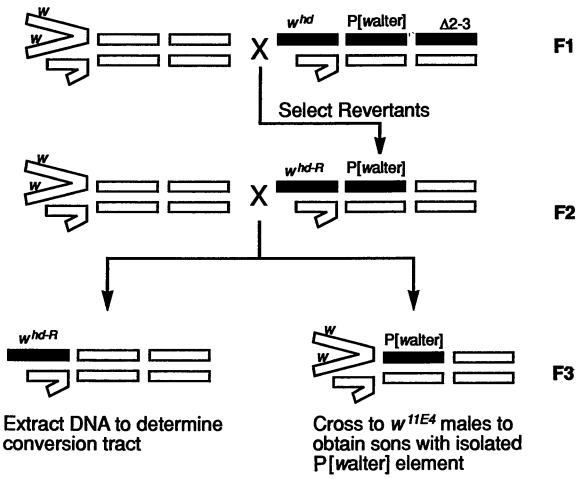


Fig. 4. Screen for revertants. Each of 1570 F1 males was mated individually to several compound-X females, and the sons were scored for revertants to normal eye color, designated $w^{hd\text{-}R}$. Genotypes are indicated in the figure showing the X and Y chromosomes and the two major autosomes. The use of compound-X females ensures that the free X chromosome is passed from father to son. Full phenotypic revertants were usually distinguishable from the partial pigmentation produced by P[*w*alter]. The F2 male is depicted carrying P[*w*alter], which is true for approximately half of such males. The presence of P[*w*alter] in these males was ascertained only in generation F3 by the appearance of some pigmented daughters. DNA was extracted from the F3 sons as described (*43*) for analysis by PCR. Each sample was first screened to ensure that no copy of P[*w*alter] was present and then analyzed to determine the presence or absence of each of the restriction sites shown in Fig. 2 (*19*). To check for transfer of information from w^{hd} to P[*w*alter], we mated pigmented F3 females to w^{11E4} males. The w^{11E4} chromosome carries a deletion for the entire *white* gene (*44*), so that sons from this cross have no *white* sequences other than P[*w*alter]. These sons were then analyzed by the same procedure used for the $w^{hd\text{-}R}$ males of F3.

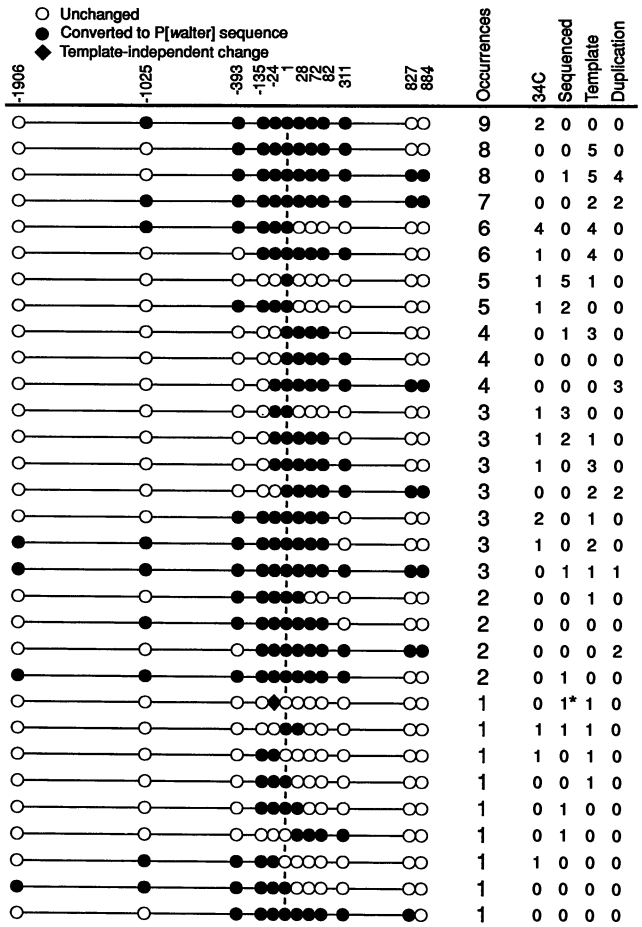


Fig. 5. Continuous conversion tracts. We classified 104 of the 123 conversion tracts as simple, meaning that they had a single continuous run of converted sites including at least one of the three central sites (bp −24, 1, and 28 in Fig. 2). These tracts fell into 31 distinguishable categories, with the number of cases in each category indicated as occurrences. The column labeled 34C shows how many of those occurrences were associated with the immobile insert, P[*w*alter](34C). The rightmost two marker sites are not informative in those cases as the corresponding portion of P[*w*alter](34C) was deleted (*16*). Also indicated are the numbers of occurrences that were partially sequenced, those whose P[*w*alter] templates were analyzed, and the ones with duplications as diagramed in Fig. 6, C and D. DNA sequencing showed that one of the tracts (asterisk) had a base substitution different from the one predicted from the restriction mapping (see text). The twelve conversion tracts derived from the two aberrant P[*w*alter] inserts, 48D and 95E, (not shown) account for two of the sequenced tracts and six of the analyzed templates.

inverted balancer chromosome can reduce pairing between the homologs (7), thus allowing the ectopic site to compete more effectively.

The most important limitation of this gene replacement technique is its requirement for a P element insertion near the target site. For many genetic loci, P element insertion mutations are easily selected. For others, large collections of well-defined P element insertion sites or PCR screening procedures (*33*) can provide the needed insertion. Given the large number of P element insertion mutations available, this technique may prove applicable to a wide range of specific problems. It is also possible that the distance between the P insertion point and the target site can be increased beyond the distribution in Fig. 7. Altering the genetic background of the F1 males (Fig. 4), such as with DNA repair mutations, might change the shape of the distribution to include longer conversion tracts. Increasing the length of homology between the target and template sequences might also yield longer conversion tracts.

Implications for transposon biology. P elements are one example of a large class of transposable elements, and the gap repair model of transposition might apply to other elements of this class. One important difference among Ac-like elements is the type of empty donor sites (footprints) left behind at the site of imprecise excision. For example, the Tam1 element in plants removes all of its own sequence plus a few base pairs of the host duplicated sequence, often with an inversion of a few base pairs at the junction (*34*). Tc1 in nematodes usually leaves a few base pairs of its own terminal sequence (*35*), whereas P elements often leave much larger segments of the donor element (*2*). It is possible that all of these elements produce double-strand gaps in a similar fashion and the different footprints reflect species-specific gap repair mechanisms. In fact, the homolog dependence of precise excision of Tc1 (*36*) suggests an excision process similar to that of P elements.

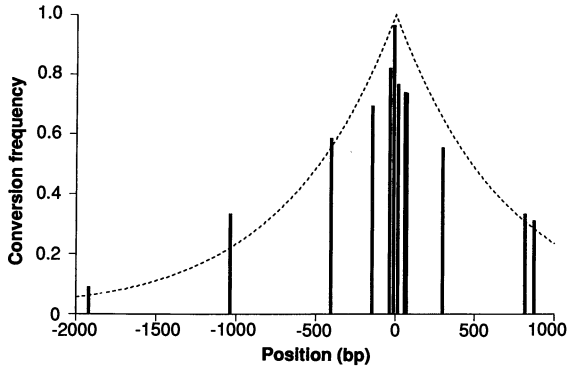


Fig. 7. The distribution of conversion frequencies. The plot includes 111 conversion tracts, excluding the 12 cases derived from internally converted P[*w*alter] elements. The proportion of conversion tracts that include each of the 12 marker sites is plotted against the distance from that site to the presumptive origin of the double-strand gap. The theoretical curve (dotted line) is based on the negative binomial waiting-time model, with its parameter estimated from the data by maximum likelihood (*23*). The average tract length computed with this model is 1379 bp estimated with a standard error of 113 bp. Only the 90 continuous conversion tracts without a duplication were used for this estimate, as the model does not apply to the complex conversion events. The standard error of the estimated average tract length is also based on maximum likelihood analysis.

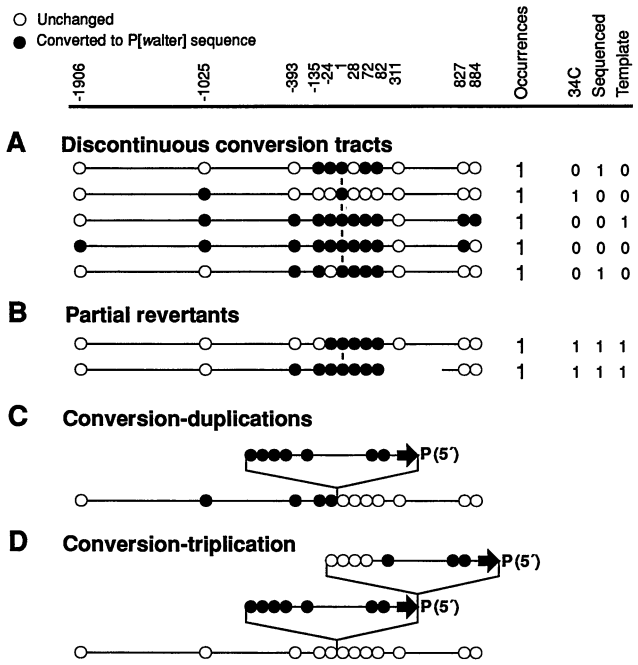


Fig. 6. Exceptional conversion tracts. (**A**) We identified five discontinuous conversion tracts. See Fig. 5, legend for description of symbols. (**B**) Conversion tracts of two partial revertants are shown. For the first partial revertant, DNA sequencing of ~600 bp centered around site 6 revealed no differences from the sequence expected from the restriction map of this revertant. The second partial revertant carried a deletion of bp 95 to 391 in the coordinates of Fig. 2. (**C**) There were a total of 14 conversion-duplications whose structure was determined by PCR analysis with primers described previously (*19*). One of the duplications was derived from one of the internally converted copies of P[*w*alter], and the rest are indicated with their conversion tracts in Fig. 5. The seven sites carried in the second copy of the duplicated segment were unconverted in all but one case, the exception having at least two of the marker sites converted. (**D**) There was one conversion-triplication whose structure and pattern of converted marker sites was determined by PCR and restriction mapping as described (*19*) except that different primer combinations were used to amplify specific copies of the triplicated segment.

Another difference among Ac-like transposons is the type of chromosome rearrangements produced. For example, the hobo and P elements both make frequent chromosome rearrangements in which the breakpoints lie at element insertion sites (*37*). However, rearrangements produced by hobo resemble homologous recombination events and are dependent on the orientation of the elements, whereas P-induced rearrangements resemble random breakage and rejoining. Species-specific repair mechanisms cannot be invoked in this case, as both elements occur in *D. melanogaster*. One possibility is that the timing of double-strand breaks within the cell cycle differs between the two elements, resulting in different kinds of chromosome rearrangements.

We suggest that some or all of the transposable elements with structures similar to P elements might transpose by an analogous mechanism. It is therefore possible that our gene replacement technique will be applicable with other transposons. In addition, transposable elements that lack the ability to make double-strand breaks could be modified to do so by the addition of a specific endonuclease recognition site such as for HO in yeast (*38*). The resulting element could then be used in the same way we have used the P element in w^{hd}.

There are various ways by which transposable elements can spread through populations (*39*). One way is to provide a beneficial function to their host organism and thus be favored by natural selection. Another is to eliminate element-free chromosomes, such as by producing dominant lethal chromosome breaks at the donor sites after nonreplicative transposition. The most efficient strategy is that of replicative transposition. Even a harmful element can spread if its transposition mechanism produces new copies faster than natural selection can eliminate them.

Nonreplicative transposable elements can make use of the DNA synthesis machinery of the host organism to achieve the same effect as replicative transposition. Some plant transposons are thought to jump during the S phase of the cell cycle and do so preferentially

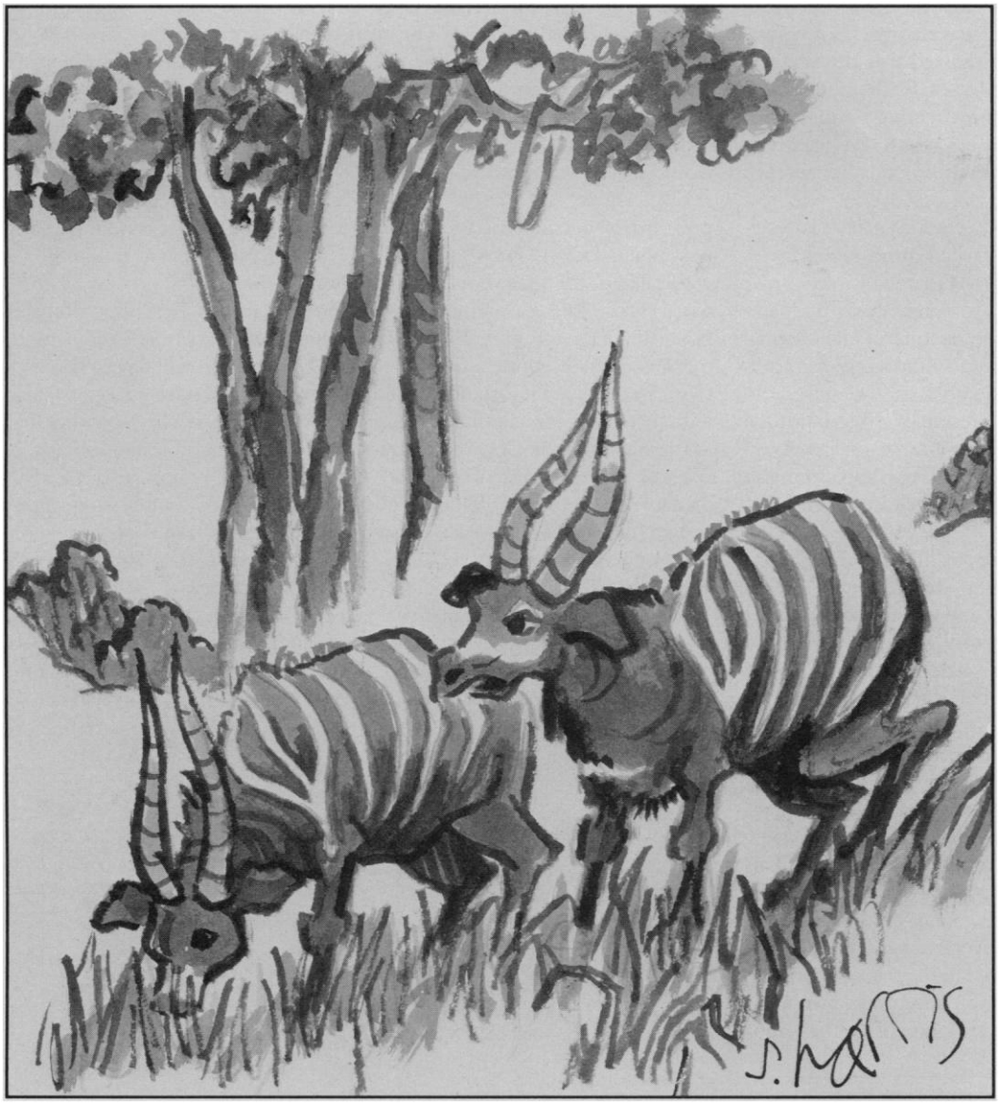
from replicated parts of the genome to nonreplicated parts (*40*). Thus, after DNA synthesis is completed, the element copy number has increased even though there are no element-encoded replication functions. Our data show that gap repair provides another way by which a nonreplicative transposon can increase its copy number. There is a net gain of one P element every time the gap left at the donor site by a transposition is repaired with the sister strand as a template. For homozygotes, the same effect is achieved when the homologous chromosome is used. We suggest that this mechanism was essential for P elements to colonize the *D. melanogaster* genome so rapidly. A similar mechanism is probably employed by Tc1 in nematodes (*36*). It thus seems likely that gap repair is an essential part of the ability of many transposable elements to establish themselves in eukaryotic genomes.

REFERENCES AND NOTES

1. D. E. Berg and M. M. Howe, *Mobile DNA* (American Society of Microbiology, Washington, DC, 1989), chapters 6–8, 14–16, 20–22, and 24.
2. W. R. Engels, in *Mobile DNA*, D. E. Berg and M. M. Howe, Eds. (American Society of Microbiology, Washington, DC, 1989) pp. 437–484.
3. M. G. Kidwell, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 1655 (1983); S. Daniels, K. Peterson, L. Strausbaugh, M. Kidwell, A. Chovnick, *Genetics* **124**, 339 (1990) M. A. Houck, J. B. Clark, K. R. Peterson, M. G. Kidwell, *Science* **253**, 1125 (1991).
4. P. M. Bingham, M. G. Kidwell, G. M. Rubin, *Cell* **29**, 995 (1982).
5. C. J. O'Kane and W. J. Gehring, *Proc. Natl. Acad. Sci. U.S.A.* **84**, 9123 (1987).
6. G. M. Rubin and A. C. Spradling, *Science* **218**, 348 (1982); A. C. Spradling and G. M. Rubin, *ibid.*, p. 341.
7. W. R. Engels, D. M. Johnson-Schlitz, W. B. Eggleston, J. Sved, *Cell* **62**, 515 (1990).
8. J. W. Szostak, T. L. Orr-Weaver, R. J. Rothstein, F. W. Stahl, *ibid.* **33**, 25 (1983); T. L. Orr-Weaver and J. W. Szostak, *Microbiol. Rev.* **49**, 33 (1985); M. A. Resnick, *J. Theor. Biol.* **59**, 97 (1976).
9. K. O'Hare and G. M. Rubin, *Cell* **34**, 25 (1983).
10. H. M. Robertson *et al.*, *Genetics* **118**, 461 (1988).
11. M. A. Innis, D. H. Gelfand, J. J. Sinsky, T. J. White, *PCR Protocols: A Guide to Methods and Applications* (Academic Press, New York, 1990).
12. R. Higuchi, in *PCR Protocols: A Guide to Methods and Applications*, M. A. Innis, D. H. Gelfand, J. J. Sinsky, T. J. White, Eds. (Academic Press, New York, 1990), pp. 177–183. These procedures were used with some modifications: Substitution-bearing primers were used to amplify nine overlapping fragments constituting the 3-kb altered region of *white*. We minimized the number of PCR cycles by starting with plasmid templates whenever possible. To remove the primers, we performed gel electrophoresis with NuSieve agarose (FMC BioProducts, Rockland, ME) and excised our PCR product band. We then merged the 9 fragments in a two-stage PCR procedure generating three overlapping partial combinations as intermediates. The yield of overlapping product was greatly increased when we performed the first several annealing steps for at least 5 min. We used Vent DNA polymerase (New England Biolabs) instead of Taq polymerase for some of the amplifications in order to minimize the chance of misincorporation.
13. V. Pirrotta, in *Vectors*, R. L. Rodriguez and D. T. Denhardt, Eds. (Butterworth, Stoneham, MA, 1988), pp. 437–456.
14. We obtained three independent transformants of P[*w*alter] with insertions at cytological positions 3C, 6F, and 34C. The one at 3C was indistinguishable from the *white* locus at the level of in situ hybridization. In addition, a mapping experiment in which 10694 progeny were scored failed to generate any recombinants between the insertion and another *white* allele. We confirmed this very close linkage to *white* by PCR analysis and DNA sequencing to show that P[*w*alter] was inserted into the 8-bp target sequence corresponding to positions 959 to 966 (Fig. 2). The orientation of the inserted *white* was opposite that of the indigenous gene. It is not known whether the presence of part of *white* within the transposon was involved in this possible site specificity of insertion. Other cases of insertional site specificity have been reported [C. Hama, Z. Ali, T. B. Kornberg, *Genes Dev.* **4**, 1079 (1990); J. A. Kassis, E. Noll, E. VanSickle, W. Odenwald, N. Perrimon, personal communication].
15. W. R. Engels, C. R. Preston, P. Thompson, W. B. Eggleston, *Focus* **8**, 6 (1986); G. Lefevre, in *Genetics and Biology of Drosophila*, M. Ashburner and E. Novitski, Eds. (Academic Press, London, 1976), vol. 1a, pp. 32–66.
16. The insertion at 34C was unusual for its lack of mobility in the presence of Δ2-3(99B). The PCR failed to amplify any fragment of this insertion when primers located in the 5′ end of the P element were combined with a primer overlapping site 10 (Fig. 2) directed to the right. (Right and left are defined as in Fig. 2.) It also failed to amplify with a left-directed primer overlapping the same site in combination with a right-directed primer overlapping site 9. Other insertions of P[*w*alter] amplified normally with these combinations. The element at 34C did, however, amplify with a left-directed primer overlapping site 9 paired with right-directed primers near sites 4 and 5. The amplified fragment sizes were identical to those produced by normal P[*w*alter] inserts. Our interpretation is that the breakpoint of a deletion lies between sites 9 and 10. The other breakpoint probably removes some or all of the P sequences at the 5′ end required in cis for transposition. Consistent with this interpretation is the finding that some of the conversion tracts associated with P[*w*alter](34C) were converted for site 9, but there were no conversions of sites 9n or 10 (see Figs. 5 and 6). The 34C insert was used in 569 of our 1570 F1 matings (Fig. 4) and accounts for 35,800 of the F2 males scored.
17. In order to estimate the reversion and conversion rates it was necessary to correct for cases that were not analyzed for phenotypic transmission or conversion tracts. Each F2 male with wild-type eye color was first classified as revertant, nonrevertant, or reversion-undetermined based on whether the phenotype was transmitted to the next generation and whether it showed linkage to the *white* locus. The reversion-undetermined category, representing 55 percent of the total, was used for males that were not successfully test-crossed, usually because of sterility or because they were part of a large cluster. The revertants were then classified as conversions, nonconversions or conversion-undetermined based on restriction mapping data. The conversion-undetermined category, representing 58 percent of the reversions, was used for males who were not successfully analyzed for conversion tracts, usually because they were part of a large cluster or because we were unable to recover a segregant free of any P[*w*alter] elements. We then obtained reversion and conversion estimates for each of the 1570 F1 crosses as follows: The numbers in each of the two undetermined categories were multiplied by the proportion of reversions or conversions among the tested males within the same sibship. When there were none tested in the sibship, the total ratio among tested individuals for that P[*w*alter] insertion site was used. Thus, we obtained estimated numbers of reversions and conversions for each cross which were then totalled to determine the frequencies in Fig. 3.
18. D. Mathog, M. Hochstrasser, Y. Gruenbaum, H. Saumweber, J. Sedat, *Nature* **308**, 414 (1984).
19. The following primers were used in this analysis: p1 = 5′-GAGTGTCGTATTGAGTCTGAG-3′; p2 = 5′-CGCAGCGGCGAAAGAGACGG-3′; p3 = 5′-GGTTGGCGCGATCTCGCGCTCT-3′; p4 = 5′-GGCTATACCAGTGGGAGTAC-3′; p5 = 5′-CGCAGTCGGCTGATCTGTGT-3′; p6 = 5′-AGGTGTTCCCTGGCCGTTAG-3′; p7 = 5′-AGCCCACCTCCGGACTGGAC-3′; p8 = 5′-CCATTGGTTGATCAGCAGAC-3′; p9 = 5′-GGTTGTCGTACCTCTCATGG-3′; p10 = 5′-GTGTTTTATGTACCGATAAACGAG-3′; p11 = 5′-GCCGACATAGATCCGAAATAACTG-3′; p12 = 5′-TTACGAATTCAGGACCTATTTCCGCTGCAC-3′; p13 = 5′-GATTAGCCAGGCTGGGCTAG-3′; p14 = 5′-CATCTGCCGAGCATCTGAAC-3′. To screen for absence of P[*w*alter] we used the combinations p1-p2 and p3-p4. These pairs amplify fragments that span the junctions between P element and *white* sequences on the left and right sides, respectively, of P[*w*alter]. They do not amplify *white* at its normal position because the P sequences are not present. If primer pair p1-p2 yielded an amplified fragment, we concluded that at least a part of P[*w*alter] was present, and that sample was not used. We did, however, analyze several revertants that had amplification with p3-p4 but not p1-p2. Most such revertants were found to have duplication or triplication structures (Figs. 6, C and D). Restriction mapping was performed by amplifying with primer pairs p5-p6, p7-p8, p9-p10, and p11-p12 in four separate reactions. The restriction enzyme Nde II was used to cut aliquots from each amplification; Rsa I was used to cut a portion of the third and fourth amplifications, and a portion of the third amplification was cut with Hae III. The fragment sizes in all seven DNA samples were determined by electrophoresis in 4 percent agarose gels. The results indicated the presence or absence of each of the restriction sites shown in Fig. 2. To check the reliability of this analysis, we repeated the procedure for nine of the conversion tracts independently. The results of both replicates were identical in each case. For some of the conversion tracts, an additional test using primer pairs p6-p13 and p6-p14 was performed to determine whether the intron deletion carried by P[*w*alter] was copied to the revertant. This was only done for the 12 cases (ten of them indicated in Figs. 5 and 6) in which the leftmost marker site was converted. Primer p13 lies within the deleted portion of the intron, and p14 is in exon 1 of *white*. Therefore, if the deletion was included in the conversion tract we would see amplification with the p6-p14 pair but not with p6-p13. Instead, we observed amplification with p6-p13 and not p6-p14 for all 12 cases, indicating that the intron deletion was not converted.
20. The P insertion in w^{hd} is flanked by an 8-bp duplication, as is typical of P element insertions (*9*). Site 6 (Fig. 2) lies within the sequence corresponding to this duplication. A simple conversion tract that does not include site 6 can occur if the double-strand gap is widened only in one direction, leaving a copy of the duplicated sequence to invade the template.
21. F. Sanger, S. Nicklen, A. R. Coulson, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5463 (1977); U. Gyllensten and H. Erlich, *ibid.* **85**, 7652 (1988).
22. W. Feller, *An Introduction to Probability Theory and Its Applications* (Wiley, New York, 1950), vol. I.
23. According to the model, the probability of a simple conversion tract having an endpoint between two sites located j and k base pairs from the origin is $(x^j - x^k)$. The probability of an endpoint lying between the origin and the most proximal marker j base pairs from the origin (sites 5 and 7 of Fig. 2) is $1 - x^j$. The probability of the endpoint lying beyond the most distal marker k base pairs from the origin (sites 1 and 10 of Fig. 2, or site 9 in cases involving the immobile insert at 34C) is x^k. These probabilities were used in a multinomial distribution in which the left and right endpoints of each of the 90 conversion tracts classified as "simple" conversions (see Fig. 5) were treated as independent trials. The estimate of x was found by numerically maximizing this likelihood [M. G. Kendall and A. Stuart, *The Advanced Theory of Statistics* (Griffin, London, 1973), vol. 2]. The precision of the estimate was determined from the negative reciprocal of the second derivative of the log-likelihood evaluated at the estimate. The length of the conversion tract, according to our model, is distributed as a negative binomial with success probability x and the number of successes equal to two. Equivalently, it is the sum of two geometric distributions, one for each half of the gap. Either way, its expectation is $2/(1 - x)$ as given by Feller (*22*). The standard error of our estimate of the average tract length was obtained by maximum likelihood in the same way as the standard error of x.

24. D. Curtis, S. H. Clark, A. Chovnick, W. Bender, *Genetics* **122**, 653 (1989); D. Curtis and W. Bender, *ibid.* **127**, 739 (1991).
25. R. H. Borts and J. E. Haber, *Genetics* **123**, 69 (1989); S. S. Symington and T. D. Petes, *Mol. Cell. Biol.* **8**, 595 (1988).
26. These totals include the discontinuous conversion tracts (Fig. 6A) and the cases of conversion tracts derived from internally converted P[*w*alter] elements. However, the second total did not include the duplications (Fig. 6C) or triplication (Fig. 6D), as such events are thought to occur without extension of the gap to include site 10.
27. R. H. Borts and J. E. Haber, *Science* **237**, 1459 (1987).
28. Our screen for revertants did not normally pick up partial revertants because such events are indistinguishable from some transpositions of P[*w*alter]. However, partial revertants were distinguishable when we used the immobile element P[*w*alter](34C), since no transposition occurred. We cannot make a reliable estimate of the frequency of partial phenotypic reversions, as only a subset of our data could be used.
29. P. J. Hastings, *BioEssays* **9**, 61 (1988); C. McGill, B. Shafer, J. Strathern, *Cell* **57**, 459 (1989).
30. J. J. Holmes, S. Clark, P. Modrich, *Proc. Natl. Acad. Sci. U.S.A.* **87**, 5837 (1990).
31. P. K. Geyer, K. L. Richardson, V. G. Corces, M. M. Green, *ibid.* **85**, 6455 (1988); H. K. Salz, T. W. Cline, P. Schedl, *Genetics* **117**, 221 (1987); J. A. Williams, S. S. Pappu, J. B. Bell, *Mol. Cell. Biol.* **8**, 1489 (1988).
32. J. A. Sved, W. B. Eggleston, W. R. Engels, *Genetics* **124**, 331 (1990); J. A. Sved, L. M. Blackman, A. S. Gilchrist, W. R. Engels, *Mol. Gen. Genet.* **225**, 443 (1991).
33. L. Cooley, R. Kelley, A. Spradling, *Science* **239**, 1121 (1988); D. G. Ballinger and S. Benzer, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 9402 (1989); K. Kaiser and S. Goodwin, *ibid.* **87**, 1686 (1990).
34. E. S. Coen, T. P. Robbins, J. Almeida, A. Hudson, R. Carpenter, in *Mobile DNA*, D. Berg and M. Howe, Eds. (American Society of Microbiology, Washington, DC, 1989), pp. 413–436.
35. D. G. Moerman and R. H. Waterston, in *Mobile DNA*, D. Berg and M. Howe, Eds. (American Society of Microbiology, Washington, DC, 1989) pp. 537–556.
36. R. H. A. Plasterk, *EMBO J.* **10**, 1919 (1991).
37. J. K. Lim, *Proc. Natl. Acad. Sci. U.S.A.* **85**, 9153 (1988); W. R. Engels and C. R. Preston, *Cell* **26**, 421 (1981); W. R. Engels and C. R. Preston, *Genetics* **107**, 657 (1984).
38. R. Kostriken and F. Heffron, *Cold Spring Harb. Symp. Quant. Biol.* **49**, 89 (1984).
39. J. W. Ajioka and D. L. Hartl, in *Mobile DNA*, D. Berg and M. Howe, Eds. (American Society of Microbiology, Washington, DC, 1989), pp. 939–958.
40. N. Fedoroff, in *Mobile DNA*, D. Berg and M. Howe, Eds. (American Society of Microbiology, Washington, DC, 1989), pp. 375–412.
41. K. O'Hare, C. Murphy, R. Levis, G. Rubin, *J. Mol. Biol.* **180**, 437 (1984).
42. The numbers of F2 males scored for each site are: 7989, 2384, 463, 1216, 2153, 35800, 703, 1162, 3131, 1362, 2384, 970, 2852, 3480, 344, 642, 2020, 2265, 876, 416, 4990, 1623, 1635, 484, 2371, 722, 313, and 1143 in the cytological order shown (Fig. 3).
43. G. B. Gloor and W. R. Engels, *Dros. Inform. Ser.* , in press.
44. Z. Zachar and P. M. Bingham, *Cell* **30**, 529 (1982).
45. This is paper number 3200 from the Laboratory of Genetics, University of Wisconsin, supported by NIH grant GM30948. We thank S. P. Howard for use of equipment.

10 May 1991; accepted 12 July 1991



"The environment people know we're an endangered species, the hunters know we're an endangered species . . . if only the *lions* knew we're an endangered species."