

Dual-Tagging Gene Trap of Novel Genes in *Drosophila melanogaster*

Tamas Lukacsovich,^{*,†} Zoltan Asztalos,^{†,1} Wakae Awano,[†] Kotaro Baba,^{‡,2} Shunzo Kondo,[§]
Suguri Niwa[†] and Daisuke Yamamoto^{*,†}

^{*}School of Human Sciences and Advanced Research Institute for Science and Engineering, Waseda University, Saitama 359-1192, Japan, [†]ERATO Yamamoto Behavior Genes Project, Japan Science and Technology Corporation (JST) at Mitsubishi Kasei Institute of Life Sciences, Tokyo 194-8511, Japan, [‡]Department of Physics, University of Tokyo, Tokyo 113-0033, Japan and [§]Mitsubishi Kasei Institute of Life Sciences, Tokyo 194-8511, Japan

Manuscript received August 18, 2000
Accepted for publication November 3, 2000

ABSTRACT

A gene-trap system is established for *Drosophila*. Unlike the conventional enhancer-trap system, the gene-trap system allows the recovery only of fly lines whose genes are inactivated by a *P*-element insertion, *i.e.*, mutants. In the gene-trap system, the reporter gene expression reflects precisely the spatial and temporal expression pattern of the trapped gene. Flies in which gene trap occurred are identified by a two-step screening process using two independent markers, *mini-w* and *Gal4*, each indicating the integration of the vector downstream of the promoter of a gene (dual tagging). *mini-w* has its own promoter but lacks a polyadenylation signal. Therefore, *mini-w* mRNA is transcribed from its own promoter regardless of the vector integration site in the genome. However, the eyes of flies are not orange or red unless the vector is incorporated into a gene enabling *mini-w* to be spliced to a downstream exon of the host gene and polyadenylated at the 3' end. The promoter-less *Gal4* reporter is expressed as a fusion mRNA only when it is integrated downstream of the promoter of a host gene. The exons of trapped genes can be readily cloned by vectorette RT-PCR, followed by RACE and PCR using cDNA libraries. Thus, the dual-tagging gene-trap system provides a means for (i) efficient mutagenesis, (ii) unequivocal identification of genes responsible for mutant phenotypes, (iii) precise detection of expression patterns of trapped genes, and (iv) rapid cloning of trapped genes.

THE enhancer-trap method is widely used in *Drosophila* for identifying new genes on the basis of their expression patterns (O'KANE and GEHRING 1987; BELLEN *et al.* 1989; BIER *et al.* 1989; WILSON *et al.* 1989). It is also used for generating cell- or tissue-specific markers or for manipulating a particular cell type by targeted ectopic expression of transgenes (BRAND and PERRIMON 1993; BELLO *et al.* 1998). Most enhancer-trap insertions result in some reporter gene expression because the *P*-element-based vector constructs tend to become integrated into the 5' regulatory regions of functional transcription units (SPRADLING *et al.* 1995) and thus the promoter of the inserted reporter is activated by a nearby enhancer sequence. In spite of the high incidence of reporter gene expression, many of these enhancer-trap lines do not exhibit detectable mutant phenotypes. That is, the genomic site of integration is under the influence of a particular enhancer sequence, yet with little effect on transcription of the adjacent gene. Thus, the enhancer-trap lines typically do not provide

functional information on nearby genes, although they are quite useful in monitoring the enhancer activity near the vector insertion site. In many cases, this expression pattern resembles that of the transcript of an intrinsic gene. The gene, however, may lie over tens of kilobases away from the insertion point. Under such circumstances, cloning and analyzing the gene under consideration could be a time-consuming process. An even more complicated situation may arise in some enhancer-trap lines in which the reporter gene expression is influenced by multiple enhancer elements of different transcription units (BOLWIG *et al.* 1995).

To circumvent these problems inherent in the enhancer-trap system, we developed a gene-trap method that utilizes a new vector construct. The primary reporter *Gal4* gene in this vector lacks a promoter. Therefore, this reporter gene is expressed only when its mRNA is covalently ligated to an endogenous mRNA either by splicing or by readthrough transcription. The pattern of reporter gene expression exactly matches that of the trapped gene since its transcription starts from the promoter of this gene. All trap lines are mutants because transcripts of the trapped gene are likely to lose their function as a result of fusion with the inserted sequence. The gene-trap system presented here invariably yields fusion products of the reporter and an exon sequence of the inserted gene. Such fusion products in gene-trap

Corresponding author: Daisuke Yamamoto, School of Human Sciences, Waseda University, 2-579-15 Mikajima, Tokorozawa, Saitama 359-1192, Japan. E-mail: daichan@mn.waseda.ac.jp

¹ Present address: Department of Genetics, University of Cambridge, Downing St., Cambridge CB2 3EH, United Kingdom.

² Present address: National Institute of Sericultural and Entomological Science, Tsukuba, Ibaraki 305-8634, Japan.

mutants provide us with a much easier and faster cloning method than any of the currently available enhancer-trap methods.

Few studies have been directed toward developing *Drosophila* gene-trap systems (DUNN *et al.* 1993). This is presumably because of the expectation that gene-trap systems would provide much fewer positive lines exhibiting reporter gene expression than enhancer-trap systems. In the enhancer-trap system, the reporter gene is expressed regardless of its orientation relative to the inserted gene, whereas in the gene-trap system, the vector needs to be integrated in the same orientation as the trapped gene (see Figure 6A and the relevant description in the text for orientation sensitivity of the gene-trap vectors). Therefore, unless a reliable detection method for gene-trap events is available, more studies need to be carried out to obtain a comparable number of positive lines. To overcome this problem, we use two independent markers, *mini-white* (*w*) and *Gal4* genes, as a "dual tag" for the trapped gene. In the first screening step, flies with intense eye color are chosen as the most likely candidates for successful gene-trap events. These lines are then checked for *in vivo* Gal4 expression by crossing them with a fly line harboring a *UAS*-driven *luciferase* gene, the product of which is detectable with a luminometer (*cf.* BRANDES *et al.* 1996).

A double-marker system has been employed in the mouse gene-trap technique (FRIEDRICH and SORIANO 1991), in which the β -galactosidase and neomycin resistance genes are typically coupled. A more recent version of a mouse gene trap uses a vector containing two reporter genes. One of these reporters has a promoter but lacks a polyadenylation signal sequence, while the other has a polyadenylation signal sequence but lacks a promoter (ZAMBROWICZ *et al.* 1998). The promoterless sequence is preceded by an artificial splice acceptor site, while the promoter-controlled sequence is followed by an artificial splice donor site. These two sites are involved in the production of two chimeric mRNAs. This gene-trap system for the mouse, designed for large-scale mutagenesis, is thus constructed on the basis of a principle similar to that of our system for *Drosophila*. Here, we report on the technical details of the dual-tagging gene-trap method together with some analyses of gene-trap fly lines.

MATERIALS AND METHODS

Construction of vectors for the gene-trap system: The *pGT1* vector (Figure 1A) is constructed by assembling three markers (*i.e.*, *mini-w*, *Gal4*, and *hs-neo^r*), splicing sites, and the "stop-start" signal into the pCaSpeR3 vector (PIRROTTA 1988). pCaSpeR3 in its original form harbors an eye-color marker *mini-w*, which is conserved in *pGT1* except for its 3' untranslated region. A *Gal4*-containing DNA fragment is excised from pCaTB (BRAND and PERRIMON 1993) and inserted into the modified polylinker site of pCaSpeR3. This fragment also contains the *hsp70* transcriptional terminator. The third marker

hs-neo^r is inserted between the *Gal4* and the *mini-w* genes. To ensure translation of *Gal4* for the forward reading frames, the stop-start signal (see LE *et al.* 1991) is placed immediately upstream of the *Gal4*-coding region. A splicing donor site, which connects *mini-w* to a downstream exon of the host gene, is inserted in the 3' end of the *mini-w* coding region. The details of vector construction are available upon request. The nucleotide sequence of the entire vector has been submitted to the DDBJ database under the accession no. AB028139. A schematic map of the construct and the principle of its operation in a gene-trap system is illustrated in Figure 1.

We also prepared a variant construct, *pGTD-b*. In this construct, the full-length *Gal4* sequence in *pGT1* is replaced with the sequence of the Gal4-binding domain followed in frame by the coding sequence of the tumor suppressor gene *p53*. The flies carrying *pGTD-b* are unable to express transgenes driven by *UAS*, unless they also carry the second construct, *pGTD-a*, which contains the sequence encoding the activation domain of Gal4 fused to the SV40 large T-antigen, a binding target of p53. Transcription of this fusion gene is controlled by the *hsp70* promoter. These constructs provide a means of temporal control of Gal4 activity in the gene-trap lines. This is because the functional Gal4 protein is produced only when p53 binds to the SV40 large T-antigen, which is inducible by heat shock.

Cloning of trapped genes: cDNAs for the trapped genes are obtained in three consecutive steps (Figure 2). The first step uses the vectorette PCR technique (ARNOLD and HODGSON 1991; ALLEN *et al.* 1994) for a mRNA template. Poly(A)⁺ RNAs extracted from a gene-trap strain are converted into cDNAs from which a chimeric cDNA composed of the *Gal4* or *mini-w* reporter and the exonic part of a host gene is amplified using primers corresponding to the adaptor sequence and the reporter sequence. The exonic sequence, thus obtained, is used in a database search for identical expressed sequence tag (EST) sequences.

When the search does not identify the cloned sequence in the database, we perform rapid amplification of cDNA ends (RACE) with multiple cDNA libraries to obtain a full-length cDNA. The total recombinant λ -phage DNAs prepared from each of several cDNA libraries are used as templates for PCR analysis. In the first round, four parallel PCRs are run with the primer corresponding to one strand of the expressed sequence, paired with either the forward primer or reverse primer with a λ gt10 or λ gt11 (depending on the actual library) sequence. If the library contains a cDNA homologous to the primer sequence, the cDNA appears in two pieces, one corresponding to the 5' part of the cDNA and another to the 3' part. These PCR products can be sequenced either directly by using another nested primer or after insertion into a plasmid. By using this method, many different cDNA libraries can be checked in a single experiment, and a "suitable one," containing the desired cDNA, can be determined easily.

In the third step, the final PCR is performed to obtain the full-length product using the library chosen as the template. The primers used here correspond to the 5' and 3' end sequences of the cDNA obtained as described above.

Mutagenesis and mutant screening: The jump-start method (COOLEY *et al.* 1988) is used for mutagenesis with the gene-trap (*pGT1* or *pGTD-b*) vector as a mutator and *P(η)⁺ Δ 2-3* transposon as a jump starter, *i.e.*, the source of transposase (Figure 3). All the flies used for mutagenesis have a *w⁻* background. With this background, the original *pGT1* or *pGTD-b* insertion confers only a very faint eye color, presumably because of the instability of the vector-derived *w⁺* mRNA, which lacks polyadenylation signals. When the vector is properly integrated into a gene, the *mini-w* gene in the vector is spliced to an exon 3' to the integration site and is polyadenylated.

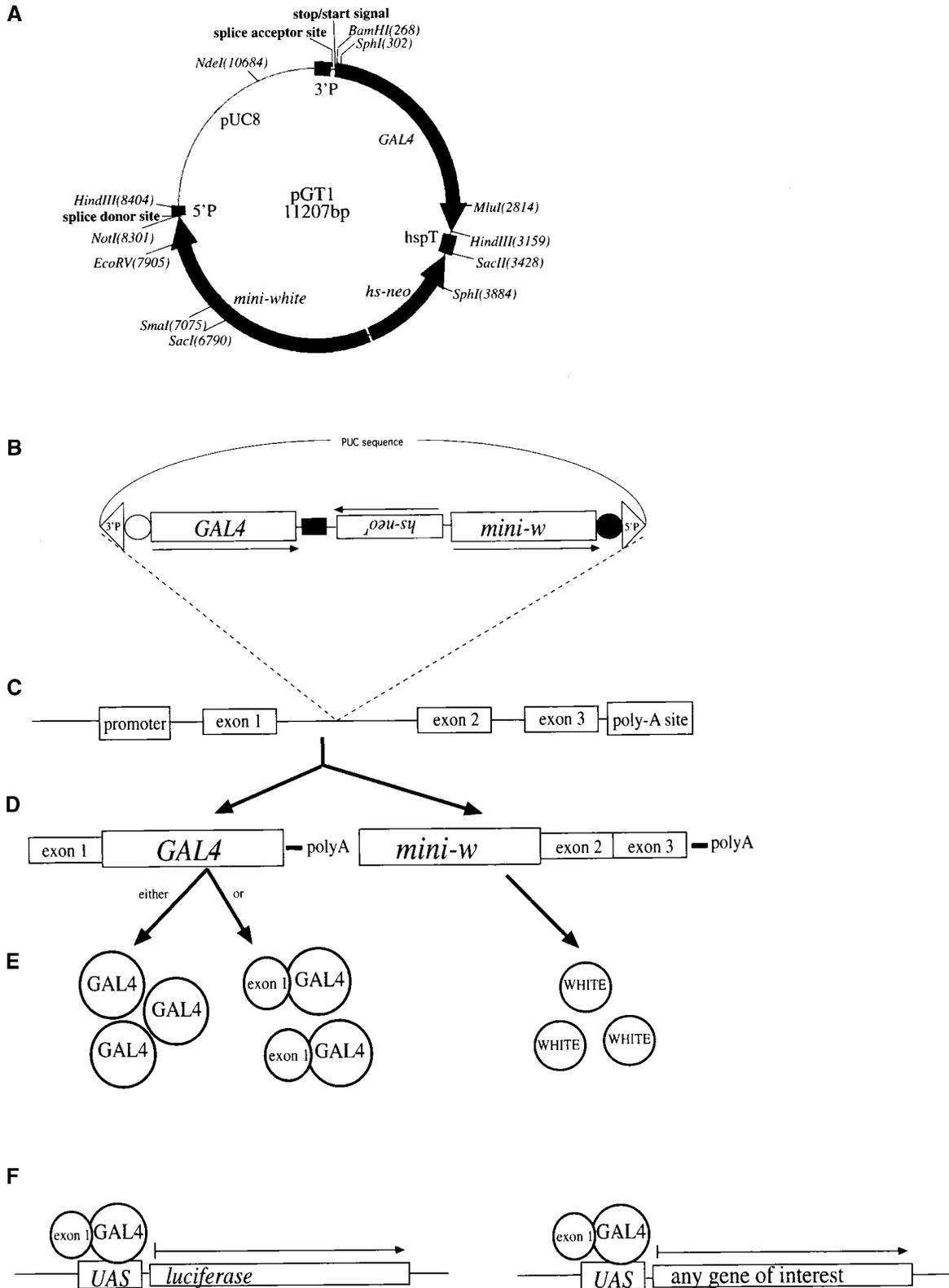


FIGURE 1.—Schematic representation of the principle of trapping genes by the dual-tagging gene-trap method. (A) A map of the *pGT1* vector. The position and direction of *Gal4*, *hs-neo*, and *mini-w* genes are indicated by thick arrows. *hspT* represents the *hsp70* terminator. 3'P and 5'P are the P-element ends. The locations of the stop-start signal, splice donor site, and restriction sites are illustrated. (B) The structure of the gene-trap vector *pGTI*. A solid box represents the *hsp70* terminator; an open circle, the splice acceptor site; and a solid circle, the splice donor site. When integrated into a gene in the *Drosophila* genome (C), *pGTI* operates to produce two fusion transcripts, one encoding the *GAL4* sequence fused to the 5' portion of the inserted gene, and the other encoding *mini-w* fused to the 3' portion of the gene (D). The fusion mRNA is translated into the functional proteins Gal4 and White (E), the former of which can activate transcription of a cDNA placed under the control of *UAS* (F), while the latter confers eye pigmentation in the fly signifying successful gene trapping.

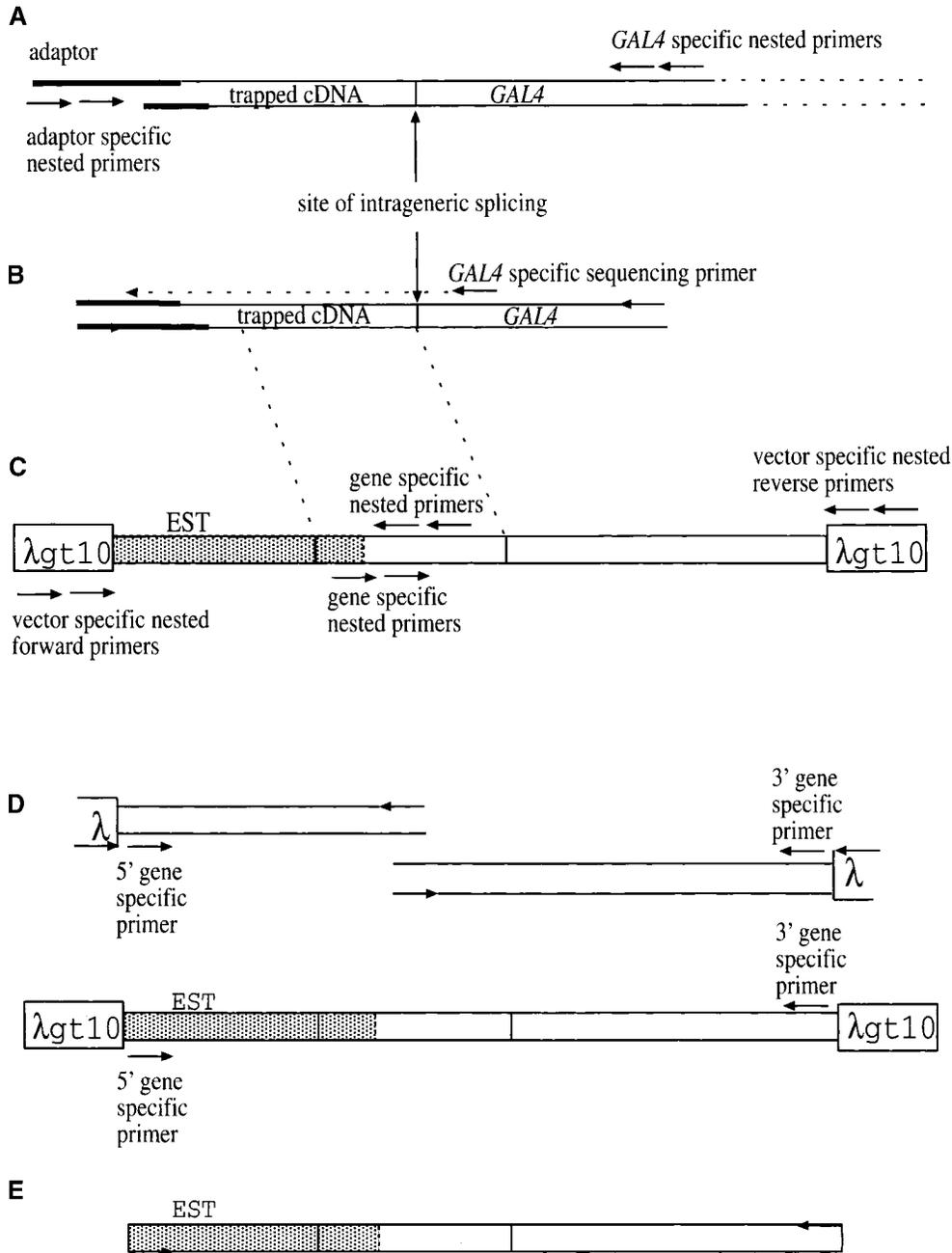


FIGURE 2.—Schematic representation of the steps of Vectorette RT-PCR. The adaptor-ligated, double-strand fusion cDNA is subjected to two rounds of PCR (A), yielding the product illustrated in B. The synthetic adaptor is shown by a thick line (the length not in scale). The primers corresponding to partial sequences of the host gene are used in combination with the vector-specific primers in two subsequent rounds of PCR on the template with the entire cDNA prepared from a whole library (C). The products of these reactions provide the sequence information for the search for corresponding ESTs. The dotted portion represents the sequence matched to an EST. (D) The two PCR products obtained from the previous step, representing the 5' and 3' ends of the trapped cDNA, respectively. After determination of their sequences, another PCR is performed to get the full-length cDNA. (E) The final product.

Therefore, collection of flies with dark eye color after remobilization of *pGT1* or *pGTD-b* allows recovery of chromosomes with the vector inserted into genes.

Female flies heterozygous for the insertion of *pGT1* or *pGTD-b* on the X or second chromosome are crossed with male flies carrying *Sb*, $\Delta 2-3$ on the third chromosome over the *TM3*, *Ser* balancer. We use fly lines pTrap1-2 (*pGT1*) or G4-DBD-p53#6-1 (*pGTD-b*) with a second chromosome insertion and exp7-65 (*pGT1*) with an X chromosome insertion as mutators that yield *P*-insertion lines at rates of 20% (pTrap1-2), 15% (G4-DBD-p53 6-1), and 8% (exp7-65) (Table 1). This cross yields F_0 "jump-start" males carrying both the gene-trap vector and the $\Delta 2-3$ element. In these males, the vector gets mobilized and "jumps" into new insertion sites. When the vector is integrated into a transcription unit, the F_0 males should have dark mosaic eyes. Each of these males is chosen as the founder of a strain and is crossed with *w* females having both the *CyO* and *TM3* balancers. When necessary, the *FM7c* balancer is used to maintain the X chromosome insertion. The balancers used are those described by LINDSLEY and ZIMM (1992). We

refer to each fly line using numbers placed after either GT or GTD, depending on the inserted vector in the fly's genome. The numbers of jump-start crosses, established lines, and lines positive for Gal4 expression are summarized in Table 1.

After the establishment of fly lines with a new insertion, they are crossed with another line carrying the *UAS-luciferase* gene (*cf.* BRANDES *et al.* 1996). Females carrying the *UAS-luciferase* reporter gene are mated with males of a putative gene-trap strain and transferred to a vial with fresh food containing 300 μ l of 1 mM luciferin. The females are allowed to lay eggs for 2 hr. The luciferase activity in larvae from these eggs is measured using a Luminescencer JNRAB-2100 (Atto, Tokyo). In the embryonic and pupal stages during which luciferase assay is impossible, a fly line carrying the *UAS-GFP* reporter gene is used for detecting *Gal4* expression by green fluorescent protein fluorescence (CHALFIE *et al.* 1994; YEH *et al.* 1995).

Scanning electron microscopy (SEM): The flies are prepared for critical point drying and coated with a 2-nm layer of gold. Images are obtained using a low-voltage prototype SEM as described by MIYAMOTO *et al.* (1995).

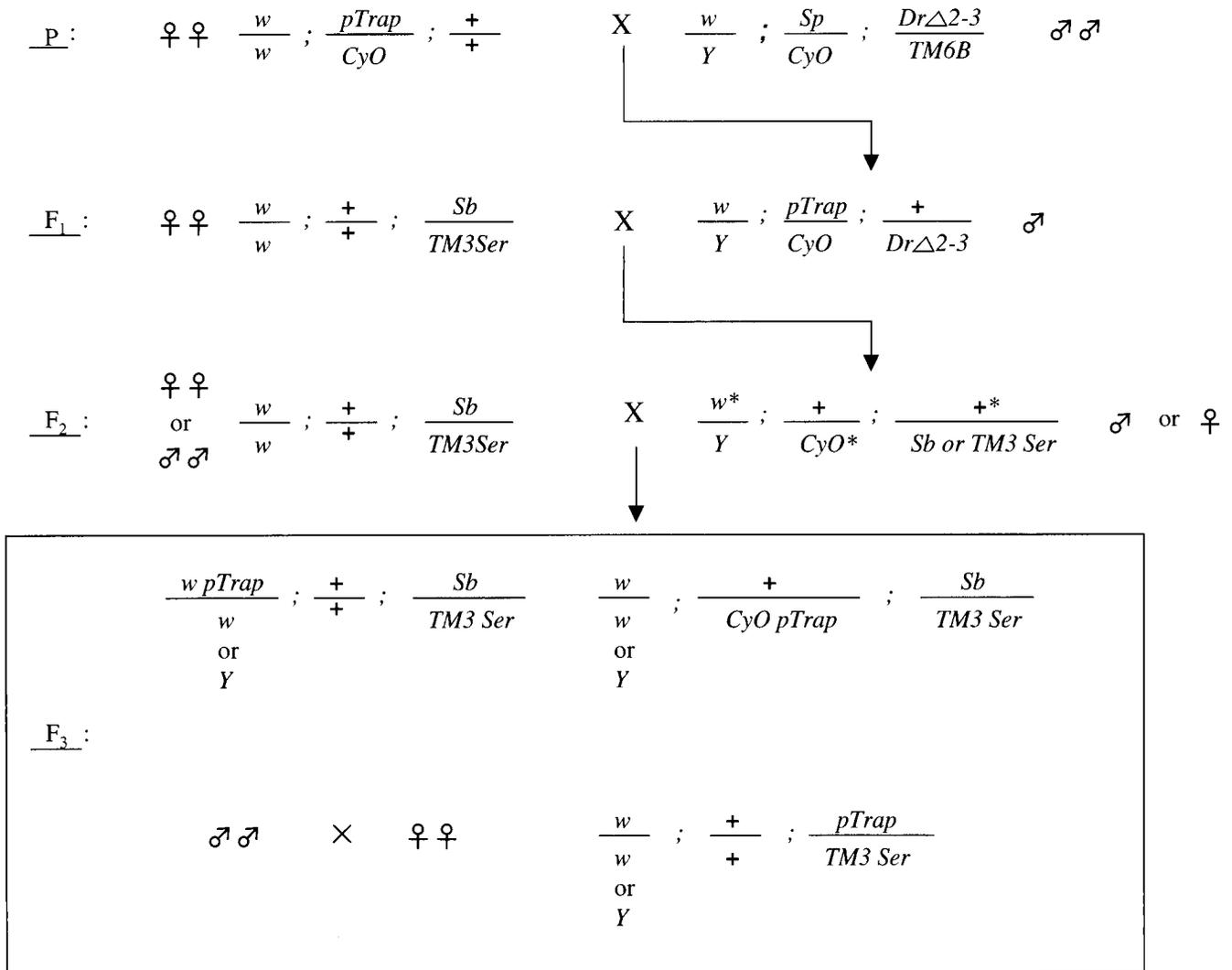


FIGURE 3.—Mating scheme for generating gene-trap fly lines. Mutator females carrying the gene-trap vector (*pTrap*) inserted into the second chromosome are mated with the males carrying the transposase source $\Delta 2-3$ in the third chromosome. All flies used are white-eyed (*w*). Each F₁ male with mosaic eyes is mated with five females, the third chromosome of which is balanced by *Sb/TM3 Ser*. F₂ flies with pigmented eyes bear the *pTrap* insertion in either of the chromosomes indicated by asterisks. The inserted chromosome can be identified in F₃. The flies with pigmented eyes that do not carry *Cy* are discarded to exclude the possibility of recovering the chromosome with the original insertion. Each F₂ male or female with pigmented eyes is mated with females or males having the balanced third chromosome (*Sb/TM3 Ser*). Offspring of this cross carry the mutated chromosome derived from one parent. When eye pigmentation of F₃ is always associated with *CyO*, the insertion has taken place in the *CyO* balancer. Otherwise, the insertion is present in the X or third chromosome. (We exclude the short fourth chromosome in our observation.) When the F₃ flies with *Sb/TM3 Ser* have pigmented eyes, the gene-trap vector must be inserted in the X chromosome. When they have white eyes, the insertion must be in the third chromosome. The gene-trap fly lines are established from the F₃ flies using appropriate balancer chromosomes. The same scheme is used for generating mutant fly lines with the gene-trap vector inserted in the X chromosome.

X-gal and immunochemical staining of eye discs: The tissue-specific expression pattern of the trapped gene is determined after introducing a *UAS-lacZ* construct (instead of *UAS-luciferase*) into the gene-trap fly lines. Staining with X-gal is performed as described previously (O'KANE 1998). Staining with an anti- β -galactosidase antibody is carried out as described by MATSUO *et al.* (1997).

In situ hybridization to polytene chromosomes: The Gal4 coding sequence of the gene-trap vector is labeled with digoxigenin (DIG)-11dUTP (Boehringer Mannheim, Indianapolis) and used as a probe to map the insertion site in the chromosome. In the case of the α -mannosidase II gene, the cDNA

sequence is also used as a probe to confirm the obtained map position, because a paralogous gene has been mapped to a different site. Signal detection is performed on polytene chromosomes using an anti-DIG antibody as described previously (JUNI *et al.* 1996).

RESULTS

Structure and principle of operation of the gene-trap vector *pGT1*: The structure of the gene-trap vector *pGT1*, is illustrated in Figure 1A. *pGT1* is a *P*-element

TABLE 1
The number of integration events of the gene-trap vectors into chromosomes
compared among three different mutators

| | | Mutator strains | | | Total |
|---|-------|---------------------------------|----------------------------------|--|------------------------|
| | | Exp 7-65 (<i>pGT1</i> on X) | pTrap1-2 (<i>pGT1</i> on II) | G4-BDB-p53#6-1 (<i>pGTD-6</i> on II) | |
| Mosaic-eyed males used for generation of founders of strains | | 675 | 460 | 420 | 1555 |
| Established lines with an insertion in the X, second (II), or third (III) chromosome | X | — | 29 | 13 | 42 |
| | II | 23 | 21 ^a | 26 ^a | 70 |
| | III | 31 | 42 | 25 | 98 |
| | total | 54 (8%) ^b | 92 (20%) ^b | 64 (15%) ^b | 210 (14%) ^b |
| Lines with Gal4 activity | | 18 (33%) ^c | 32 (35%) ^c | 24 (38%) ^c | 74 (35%) ^c |

^a The gene-trap vector was integrated into the *CyO* chromosome.

^b The ratio of the number of established lines over that of total jump-started mosaic-eyed males shown in the top row.

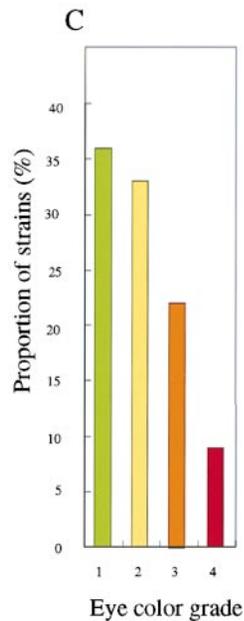
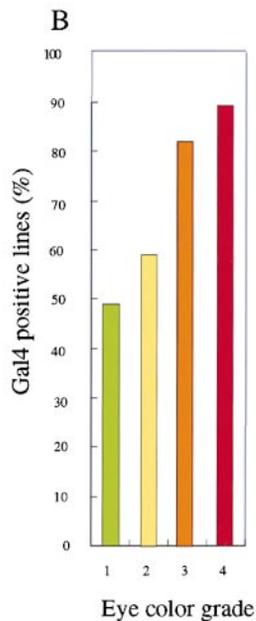
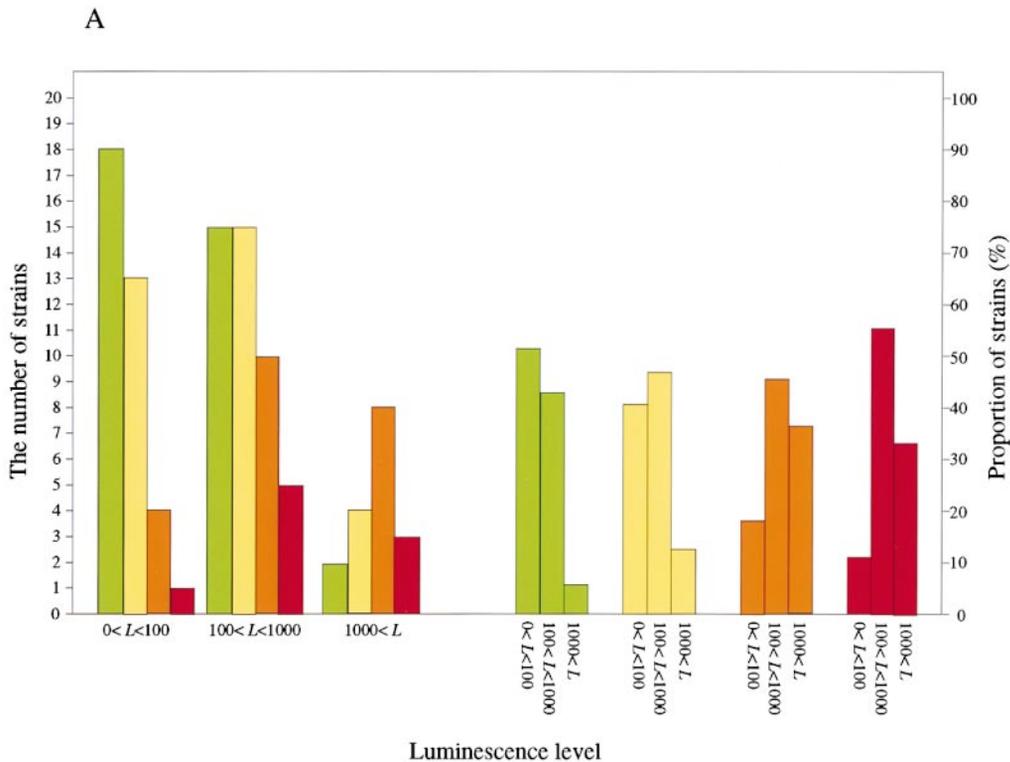
^c The percentage of gene-trap lines with Gal4 expression relative to the total stocks with vector insertions, which are given in the middle row.

vector containing three markers: *Gal4*, *mini-w*, and *hs-neo^r*. The *Gal4* gene lacks a promoter but has a poly(A)⁺ signal sequence. The *mini-w* gene has a promoter but lacks the poly(A)⁺ signal sequence. The *Gal4* gene is preceded by an artificial splice acceptor site, which allows the *Gal4* sequence to be transcribed as a fusion mRNA with an upstream exon sequence (Figure 1, B and C). That is, *Gal4* would never be expressed unless the vector is integrated downstream of the promoter of a host gene. Since the stop-start sequence is placed between the splice acceptor site and the *Gal4* gene, the open reading frame (ORF) for *Gal4* is maintained in any integration event (Figure 1D). It is unequivocal that the Gal4 fusion protein activates transcription from the target sequence *UAS* (Figure 1E). On the other hand, the *mini-w* gene is followed by an artificial splice donor site, which is involved in the production of a chimeric mRNA composed of the *mini-w* sequence and the exonic sequence of a host gene 3' to the vector integration site. This chimeric mRNA is polyadenylated and encodes the functional White protein (Figure 1C). Even when the vector is inserted outside a gene, the *mini-w* gene can be transcribed from its own promoter. In this case, however, the *mini-w* mRNA would not be polyadenylated, because its 3' end cannot be spliced to an exon of the host gene. Such an unpolyadenylated mRNA is likely to be degraded rapidly before being translated, thereby resulting in the absence of eye pigmentation in the fly carrying the vector (the flies have a *w⁻* background). In contrast, the *mini-w* gene is expected to confer a dark eye color on the fly when the vector is integrated into a gene, because the chimeric mRNA is polyadenylated. Thus, the two reporter genes, *Gal4* and *mini-w*, generate functional proteins only when the vector is integrated into a transcription unit. Since the gene trapped in this

way is tagged dually by *Gal4* and *mini-w*, we call this method the “dual-tagging gene trap.”

The vector must be inserted into the genome before any attempt at its intragenic integration. The vector contains *hs-neo^r*, which allows us to recover the flies with the vector, even when the integration site is outside the transcription unit. In fact, the flies with the vector inserted outside the transcription unit had faintly pigmented eyes (pale yellow), which made it possible to recover these flies without using *hs-neo^r* (Figure 4D).

Gene-trap screens: The primary advantage of the dual-tagging gene-trap system is its ability to distinguish the *P*-element insertions occurring downstream of the promoters (*i.e.*, intragenic insertions) from those occurring in other regions of the genome simply by evaluating the eye color intensity of the flies. For procedural convenience, the degree of eye pigmentation is graded into four levels, 1 to 4, by comparing the eye color of individual flies with that of four arbitrarily chosen standard strains; the eye color of each strain represents one of the four pigmentation levels (Figure 4D). In the secondary screening of the flies in which a gene is trapped, we examine the expression of another marker, Gal4. Gal4 activity is detected on the basis of *UAS-luciferase* expression. The females of each strain are allowed to lay eggs on luciferin-containing food, and the offspring are subjected to luminescence measurements at the first, second, and third instar larval stages. Live larvae are subjected to the measurements. Luminescence intensity higher than the background level is detected in some of the strains (Figure 4A). To ascertain that the lines with luciferase activities express Gal4, the putative gene-trap lines are crossed with flies, carrying *UAS-lacZ*, by which histochemistry of β -galactosidase activity can be used to monitor Gal4 expression. In no case have



expression. Approximately 90% of the strains belonging to the grade 4 group with the darkest eye color show Gal4 expression. (C) The percentage of fly lines with different degrees of eye pigmentation. Ten percent of flies generated from mutagenic crosses belong to grade 4 group. (D) The eyes of “representative” strains of the four eye color grade groups.

luciferase and β -galactosidase activities dissociated; thus, the luciferase reporter is proven to be a reliable and efficient tool for mass screening of Gal4-expressing strains. When the trapped gene is tagged dually by *mini-w* and *Gal4*, the presence of Gal4 expression is considered to

be correlated with intense pigmentation of the eye. Note, however, that the expression patterns of *Gal4* and *mini-w* are usually different, because the Gal4 expression is driven by the intrinsic promoter of the trapped gene whereas *mini-w* expression is controlled by its own pro-

FIGURE 4.—Correlation between the degree of eye pigmentation and the level of Gal4 expression in *Drosophila* strains with a gene-trap-vector insertion. (A) Relationships between the degree of eye pigmentation and the level of luminescence that indicates Gal4 expression. Ninety-nine strains with vector insertions are classified into either three groups on the basis of the level of luminescence (L , relative counts for 20s determined using the luminometer; $0 < L < 100$, $100 < L < 1000$, and $1000 < L$) or into four groups on the basis of the eye color grade as illustrated in D. Luminescence reflects luciferase activity as induced by the *UAS-luciferase* reporter gene, which is activated by Gal4 derived from the gene-trap vector inserted into the gene. Results are obtained at the third instar larval stage. The degree of eye pigmentation is indicated by the color (examples are shown in D). In the left part of the graph in A, the number of fly strains that belong to either one of four eye color groups (illustrated by a color code) is shown for three luminescent groups. In the right part of the graph in A, the relative contribution of strains with different luminescence levels in each eye color group is depicted. (B) The percentage of strains with detectable levels of Gal4 expression is shown for four groups of strains on the basis of the degree of eye pigmentation, *i.e.*, grade 1 to grade 4. The strains exhibiting 100 or higher luminescence counts are considered to be positive for Gal4

moter. Thus, the *mini-w* gene is expressed only in the eye, but Gal4 is expressed wherever the trapped gene is expressed.

More than 200 strains with gene-trap-vector insertions are generated from 1550 males having the mosaic eye color due to *P* element mobilization (Table 1). Both the degree of eye pigmentation and the level of luciferase activity are determined in about half of the strains to see whether the expected correlation exists. The fly strains with gene-trap vectors are classified into three classes on the basis of their levels of luminescence, *L* (Figure 4A), and the proportion of strains with four different eye color grades is assessed in each of the classes. The actual number of strains that were grouped according to *L* levels and degree of eye pigmentation is shown in the left half of Figure 4A. The percentage of strains with different *L* levels within the same eye color groups is compared in the histogram shown in the right half of Figure 4A. The percentage of flies that express Gal4 is compared among the four groups classified according to the degree of eye pigmentation in Figure 4B. It is clear that the fly strains with darker eye color are more likely to express Gal4 than those with lighter eye color: 88% of the strains in group 4 express Gal4, whereas 48% of group 1 do (Figure 4B). The proportion of the strains with grade 4 eye color is ~10% of the total *P*-element-insertion stock generated by this method (Figure 4C). Each line carries a single gene-trap vector (Table 2), except for GTDexp1-#31/1-GTDexp1-#2/2, in which the $\Delta 2-3$ element has been reintroduced intentionally to initiate a local jump of the inserted *pGTD-b*. It is noted, however, that there are fly lines carrying a gene-trap-vector insertion without Gal4 expression yet with a dark eye color (Figure 4A). This suggests that a dark coloration of the compound eyes is a reliable indication that a host gene has been trapped even when no Gal4 expression is detected. Thus, the dual-tagging gene trap provides a simple and efficient approach for the generation of mutants.

Sixty-two Gal4-expressing lines are chosen from the 99 lines shown in Figure 4A for further analysis. Among the 62 lines, homozygosity is found to be lethal in 18, semilethal in 7, and associated with sterility in 3. Thus, roughly half of the generated gene-trap lines exhibit a lethal or sterile phenotype. The flies used in the cross have the isogenized Canton-S genetic background that is free of lethal mutations (provided by M. Yoshihara). In 4 other lines, the *P* element is inserted into the *CyO* balancer; therefore, the effect of the insertions on viability and fertility cannot be evaluated. The remaining 30 lines are maintained as homozygous stocks without any difficulty.

Identification and characterization of trapped genes:

Genes trapped by intronic integration of the vector: We have cloned the partial sequences of 27 trapped genes listed in Table 2. In most cases, the artificial splicing acceptor (SA, marked with open arrows in Figure 5) and donor

(SD) sites work as expected, and the endogenous mRNA is found to be fused to the *Gal4* (and *mini-w*) sequences exactly at the expected nucleotide (Figure 5). Figure 5 also includes examples where upstream, secondary splice sites are used (GTexp16-#8 and GTDexp1-#2/2, Figure 5D). At nucleotides 15 and 24 upstream of the designed splice acceptor site, there are two additional AG dinucleotides (SAs marked with an arrowhead in Figure 5A) that are 100% conserved in the 3' end of *Drosophila* introns (MOUNT *et al.* 1992). This part comes from the 3'-end sequence of the *P* element. In some gene-trap lines, the splicing occurs at either of these sites (Figure 5D, arrowheads). In a few cases, two different splicing products are obtained from the same line (Figure 5D, lower two rows).

In another line, GTexp7-#45, the flanking sequence of the *P*-element insertion point is found to be identical to the Berkeley *Drosophila* Genome Project (BDGP) EST GM04742. Notably, however, the EST sequence corresponds to the first intron of the trapped gene, and not to an exon. The intron in question has been obtained by genomic PCR, after recovering the fusion cDNA containing an exon of the trapped gene. Thus, the intron of the trapped gene is found to contain an independent gene on the complementary strand identical to GM04742. The gene-trap construct is inserted into the first untranslated exon of the gene registered as GM04742, but in the opposite transcriptional orientation. This indicates that the direction of transcription of the *Gal4* reporter gene of the construct is also opposite to that of GM04742 (Figure 6A). Thus, this line could be used to indicate the positional specificity of the reporter gene expression in the gene-trap construct compared with that in the enhancer-trap construct. If an enhancer-trap vector is integrated into the same position, its reporter gene expression would reflect the expression pattern of the gene GM04742 (or maybe both of the genes). On the contrary, in the gene-trap system, Gal4 expression follows the pattern of a gene more distantly located but on the same DNA strand, *i.e.*, the trapped gene, the first exon of which is spliced to the Gal4 mRNA.

The most complicated configuration of trapping events is found in the line GTDexp1-#31/1-GTDexp1-#2/2. In this case, the inserted element is duplicated during local hop with the aid of the $\Delta 2-3$ jump-starter chromosome. As a result, two *P* elements are found in the same intron of the same gene (Figure 6C). Each of the elements traps an exon upstream of it, leading to the identification of two transcription units. One of these two units is contained in the intron of the other. The former gene encodes a putative transcription factor with a ring-finger motif, termed Brain tumor (Brat; EMBL/GenBank database accession nos. AF119332 and AF195870-73).

Genes trapped by exonic integration of the vector: There are examples where the vector is integrated into the first untranslated exon of a gene. It is obvious that the

TABLE 2
Gene-trap lines in which the trapped genes are identified molecularly

| Strain | Gene | Viability of homozygotes | Chromosomal location | Corresponding genomic clones or STSs | Mechanism of Gal 4 expression | Homologous ESTs in fly | Homologs in other organisms |
|----------------|-----------------------------------|--------------------------|----------------------|--------------------------------------|---|------------------------|--|
| GT1pok | <i>anterior open (aop)</i> | Lethal | 22D | BACR01K07 | Intergenic splicing | GH23964 and others | ETS proteins |
| GTDexp74#310-2 | <i>trithorax (trx)</i> | Semilethal | 88B3 | BACR01H04 | Intergenic splicing | GH06495 and others | Mouse Mll human MLL |
| GTDexp74#96 | <i>nebula (nla)</i> | Viable | 89B12-13 | EP(3)3462 | Intergenic splicing | LD15403 and others | Human yy37d03.r1 |
| GTDexp1#31/1 | <i>brain tumor (brat)</i> | Viable | 37C1-D1 | BACR48001 BACR10109 | Intergenic splicing | LD28374 and others | Homologous to Caenorhabditis Ncl-1 zinc finger |
| GTDexp74#310-1 | <i>bancal hmRNP K</i> | Lethal | 57A | AC014285 | Intergenic splicing | LD42648 and others | Human hnRNP K |
| GTexp7#6 | CG17090 | Female sterile | 61C | Dm65A9S | Intergenic splicing | SD05876 and others | Human PKY protein kinase |
| GTexp7#45 | Unknown | Viable | 55C | AE003799 | Intergenic splicing | | |
| GTexp49-2-4 | Unknown | Semilethal | 86D | BACR48E12 BACR06005 | Intergenic splicing | | |
| GTexp16#8 | <i>Fructose 2,6 biphosphatase</i> | Viable | 17C-D | Dm11H6S | Intergenic splicing and readthrough transcription | GH24742 and others | Fructose biphosphatases |
| GTDexp1#2/2 | Unknown | Viable | 37C1-D1 | BACR48001 BACR10109 | Intergenic splicing | SD07237 | |
| GTDexp1#24/1 | <i>α-mannosidase II</i> | Lethal | 89A1-5 | BACR48G17 | Intergenic splicing | LD09921 and others | α-Mannosidases |
| GTDexp1#1/1 | Unknown | Lethal | 79E4-6 | Dm1562 | Intergenic splicing | | |
| GTDexp74#56 | CG1244 | Lethal | 62F1-2 | BACR22P20-T7 | Intergenic splicing | LD26015 and others | Human ye89f06.r1 |
| GTDexp74#94/1 | Unknown | Lethal | 88A-B | BACR10E03 | Intergenic splicing | GM07940 | |
| GTDexp74#159-1 | CG4798 | Viable | 54B-C | BACR21A09 | Intergenic splicing | LP05522 and others | Mouse mg-90f12.r1 |
| GTDexp74#1 | Unknown | Lethal | 56F | AC004290 | Intergenic splicing | | |
| GTDexp74#198 | Unknown | Viable | 89B-C | BACR05P04 | Intergenic splicing | | |

(continued)

TABLE 2
(Continued)

| Strain | Gene | Viability of homozygotes | Chromosomal location | Corresponding genomic clones or STSs | Mechanism of Gal 4 expression | Homologous ESTs in fly | Homologs in other organisms |
|-----------------|--------------------------------------|--------------------------|----------------------|--------------------------------------|-------------------------------|------------------------|--|
| GTDexp74-#135-2 | Unknown | Semilethal | 87D | AC020412 | Intergenic splicing | | |
| GTDexp74-#61-3 | Unknown | Viable | | | Intergenic splicing | | |
| GTDexp7-#49 | Unknown | Not determined (on CyO) | 37A | BACR25C02 | Readthrough transcription | LD32867 and others | Homologous to mouse FAF RNA binding proteins |
| GTDexp7-#11 | <i>mushroom body expressed (mub)</i> | Lethal | 79B1-2 | BACR07J05 EP(3)0783 | Readthrough transcription | | |
| GTDexp74-#159-2 | CG4798 | Viable | 54A1-B5 | EP(3)3638 EP(2)2177 | Readthrough transcription | LD46226 and others | |
| GTDexp74-#95/1 | Unknown | Viable | 91F1-3 | BACR48G05 Dm0333 | Readthrough transcription | GMI0731 GM01567 | |
| GTDexpZ-1 | Unknown | Semilethal | 79B | EP(3)3108 EP(3)3623 | Readthrough transcription | | |
| GTDexp74-#77/3 | Unknown | Viable | 68A8-9 | AC017890 | Readthrough transcription | | |
| GTDexp74-#295-1 | Unknown | Lethal | 61D1-2 | AC019753 | Readthrough transcription | | |
| GTDexp74-#236 | Unknown | Viable | 68C2-4 | AC004438 | Readthrough transcription | | |

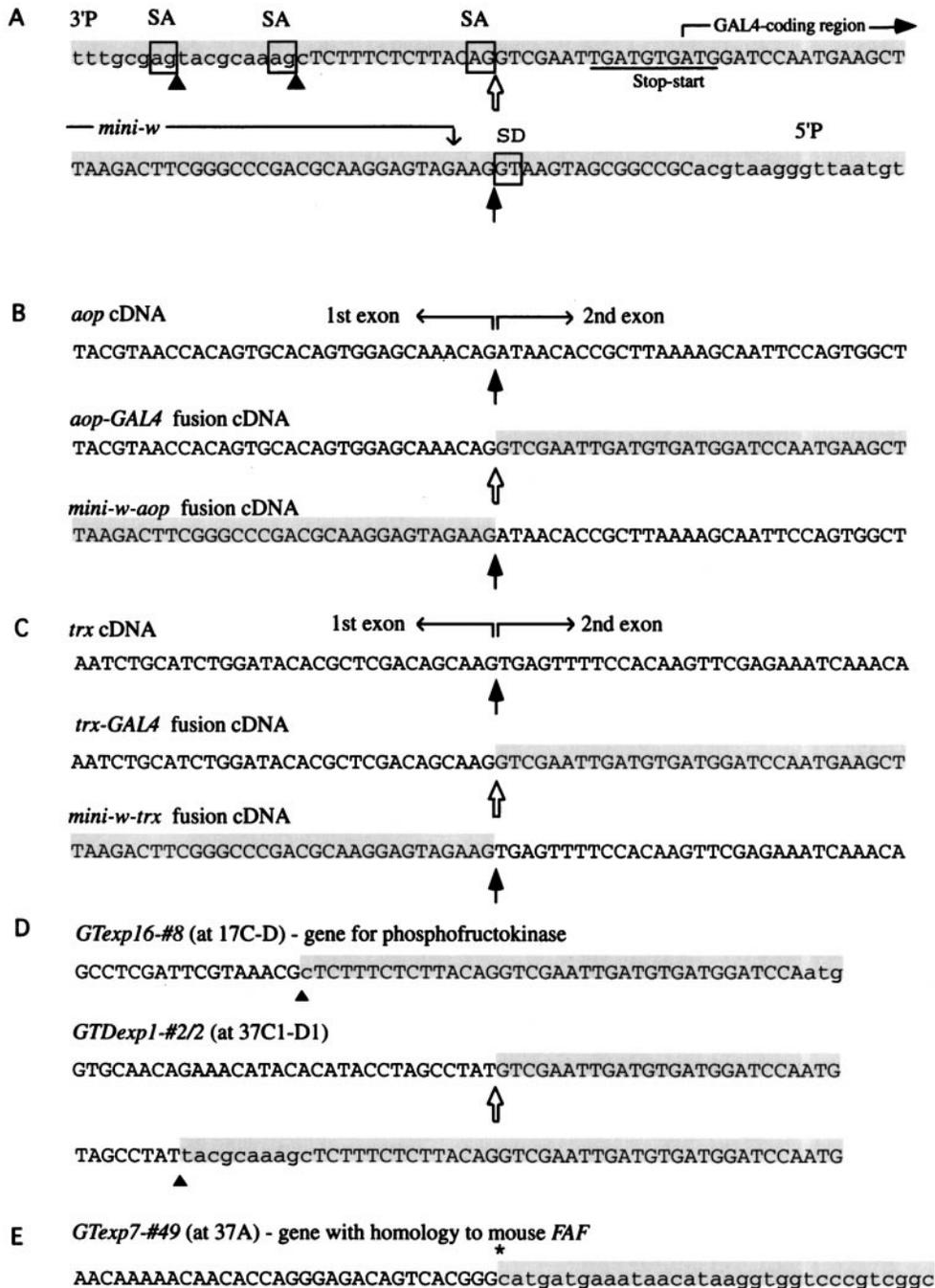


FIGURE 5.—Sequences around the fusion points of chimeric cDNAs obtained in some gene-trap lines. (A) Shown are the sequences of the trap vector surrounding the splice acceptor sites (SA, arrowheads and an open arrow; upper row) or the splice donor site (SD; lower row) following the splicing signal sequence (boxed). The first two SAs (those indicated by arrowheads) are not introduced intentionally into the construct. These two sites are originally contained in the *P*-element sequence and incidentally recognized as SAs by the splicing machinery. The stop-start sequence is underlined. The region indicated by lowercase letters represents the *P*-element sequence. The vector sequences are shaded. (B and C) The fusion junctions of the chimeric mRNAs from two known loci, *anterior open* (*aop*; B) and *trithorax* (*trx*; C). The top row shows the sequence of wild-type cDNA at the junction of the first and second intron. The junction is indicated by an arrow. The second row illustrates the junction of the first exon of the *aop* or *trx* gene and the vector sequence adjacent to the stop-start and Gal4 sequences. The third row shows the junction of the first exon of the *aop* or *trx* gene and the *mini-w* sequence. The sequence of stop-start Gal4 or *mini-w* is shaded. (D) Examples in which either of the first two SAs is responsible for splicing that leads to the fusion events. In the *GTexp16-#8* line, the gene encoding phosphofructokinase is spliced to the vector sequence at the second SA (arrowhead). In the trap event

GTDexp1-#2/2 in the *GTDexp1-#31/1-GTDexp1-#2/2* line, an exon of the host gene is spliced to the vector sequence at the first SA (arrowhead) in addition to that at the third (open arrow) SA. (E) The *GTexp7-#49* line provides an example of readthrough transcription. In this case, the first exon of the host gene is followed by the *P*-element sequence. The junction of the host and *P*-element sequences is indicated with an asterisk.

artificial splice acceptor site is not used in these cases because of the absence of upstream exons. Instead, the *Gal4* gene is expressed by readthrough transcription from the nearby promoter. An example of this is found in *GTexp7-#49*, in which the *Drosophila* homologue of the mouse Fas-associated factor (FAF1, CHU *et al.* 1995) is targeted (Figure 5E; AB013610).

Another line, *GTexp16-#8*, shows readthrough transcription of *Gal4* from a nearby promoter of the putative

Drosophila phosphofructokinase gene (AB023510–11), which harbors the vector in an untranslated exon (Figure 6B). In this line, another *Gal4* fusion mRNA is also detected. This fusion mRNA is a result of the splicing event at the second SA. A distant upstream exon of the *phosphofructokinase* gene (Figures 5D and 6B) is identified by this fusion mRNA. The *phosphofructokinase* gene even has a third untranslated exon (Figure 6B).

Expression and phenotypic analysis: The gene-trap lines

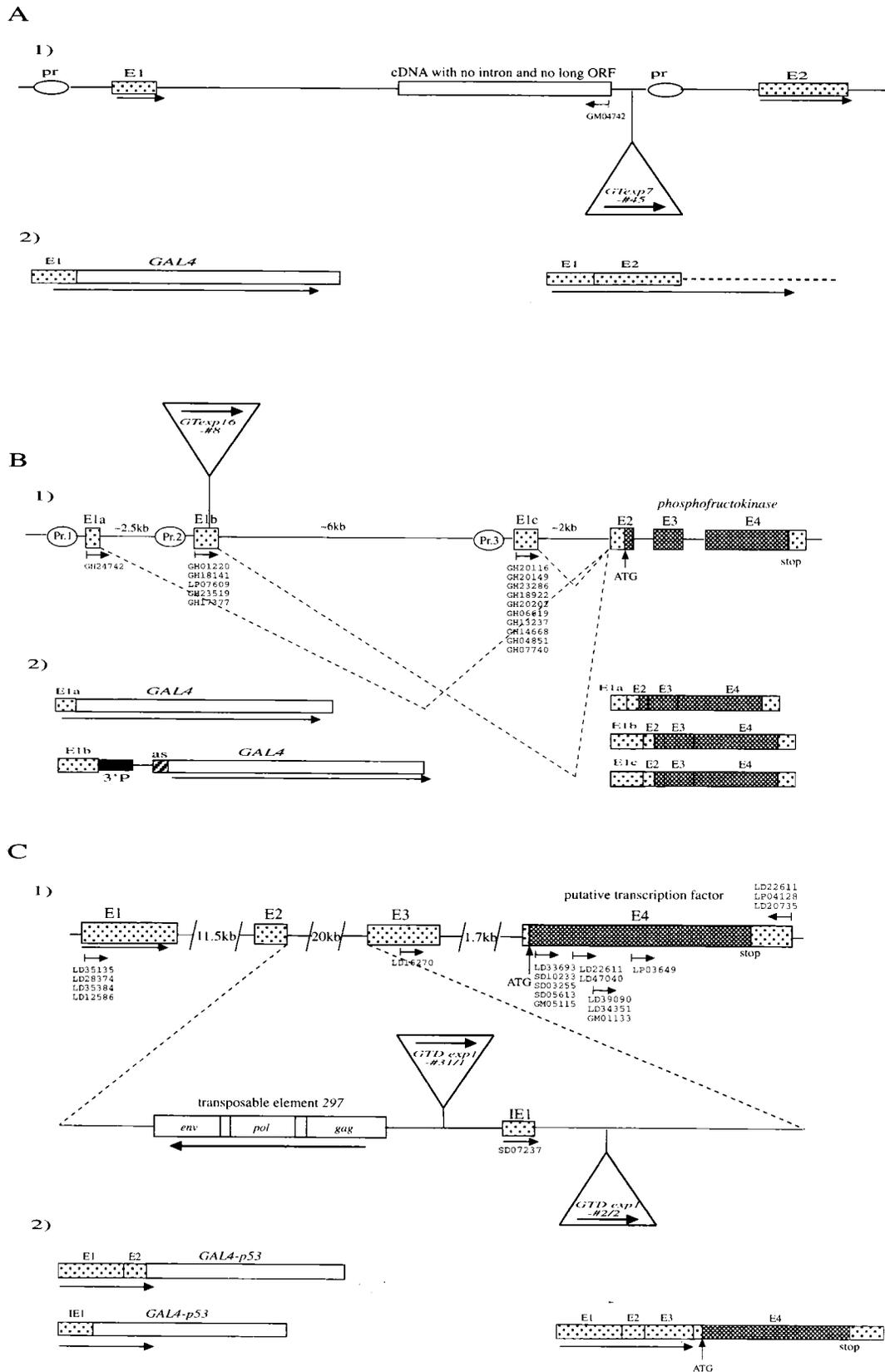


FIGURE 6.—Complex genomic arrangements observed in three gene-trap lines. (A) Gene-trap line GT exp7-#45. The *P* element is integrated into an intron of a novel gene, which contains a second gene with a sequence identical to GM04742 in the complementary strand. (B) Gene-trap line GTexp16-#8. The *P* element impinges upon the second noncoding exon of the gene encoding phosphofructokinase. (C) Gene-trap line GTDexp1-#31/1-GTDexp1-#2/2. This line contains *pGTD-b*, which produces a Gal4-binding domain followed by p53 (Gal4-p53) instead of full-length Gal4 (see MATERIALS AND METHODS). As a result of the local hop of the *P* element induced by the $\Delta 2-3$ chromosome, two insertions occur in tandem in the same intron of the

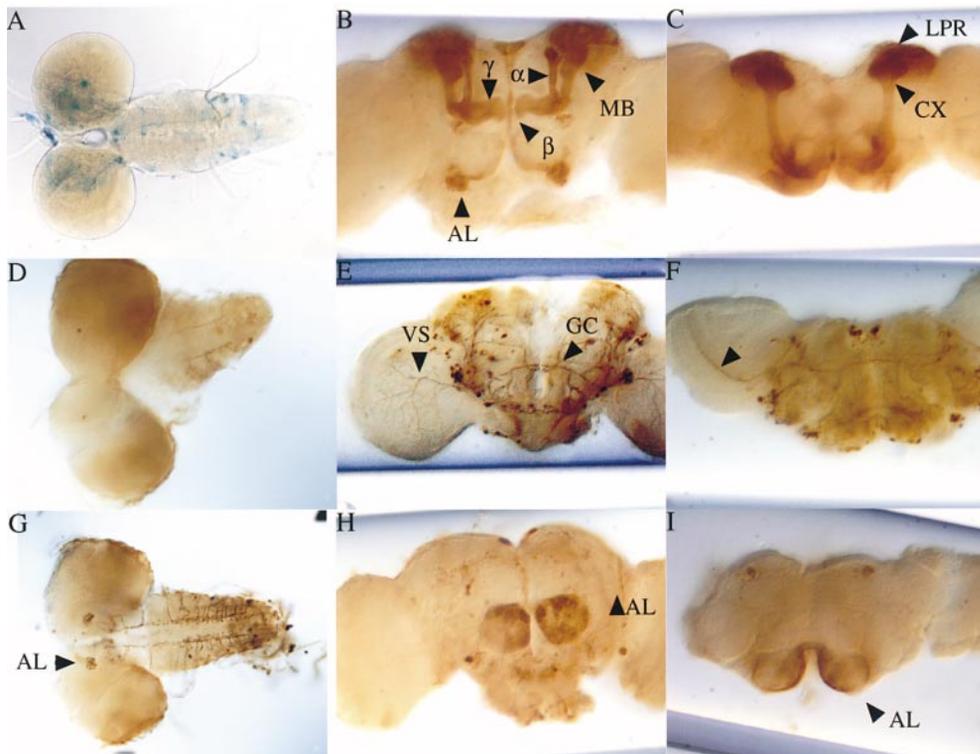


FIGURE 7.—Monitoring of the expression of trapped genes by the *Gal4* reporter gene. *Gal4* reporter gene expression as detected by *UAS-lacZ* is shown in three gene-trap lines, GTexp16-#8 (A–C), GTexp7-#77 (D–F), and GTexp49-#2 (G–I), which revealed expression in defined neural structures with unique geometry. The *lacZ* expression is detected by anti- β -galactosidase antibody immunocytochemistry except for A, which represents X-gal staining. A, D, and G represent dorsal views of the brain-ventral ganglia complex of the third instar larvae. B, E, and H are frontal views and C, F, and I are dorsal views of the adult brain. In GTexp16-#8, *Gal4* is preferentially expressed in the mushroom body and the antennal lobe. In GTexp7-#77, *Gal4* is expressed in different classes of neurons, including the giant visual interneurons (arrow) that extend arborizations horizontally

(STRAUSFELD 1976). In the GTexp49-#2 line, prominent expression is observed in the antennal lobe in both larval (G) and adult (H and I) brains. In the larval ventral ganglia, longitudinal axon bundles and segmental motor neurons are clearly stained. MB, mushroom body; α , α -lobe; β , β -lobe; γ , γ -lobe; LPR, lateral protocerebrum; CX, calyx; GC, giant commissure; AL, antennal lobe.

are crossed with a strain carrying *UAS-lacZ*, which expresses β -galactosidase detectable either using the chromogenic substrate X-gal or by immunohistochemistry. The patterns of reporter gene expression are unique in each gene-trap line but are similar among the specimens in the line. In cells exhibiting reporter gene expression, uniform cytoplasmic staining is observed, making it possible to visualize even cellular structures with complex geometry, such as axons and dendrites of neurons, in addition to cell bodies (Figure 7). Dynamic changes in the expression of trapped genes during development are monitored clearly (Figure 7).

The expression pattern detected by the *Gal4* reporter gene matches the localization of the protein encoded by the trapped gene. This is demonstrated for *aop*^{GT1} by comparing the protein localization, as determined using a specific antibody, and *Gal4* reporter expression. The *aop* gene encodes an E-26-specific-domain transcriptional repressor (LAI and RUBIN 1992; TEI *et al.* 1992) that prevents cells from differentiating prematurely (REBAY

and RUBIN 1995; ROGGE *et al.* 1995). The expression of *Aop* in developing eye discs has been documented in detail (LAI and RUBIN 1992; REBAY and RUBIN 1995; YAMAMOTO 1996): In the third instar eye disc, almost all undifferentiated cells located basally expressed the *Aop* protein, whereas none of the differentiated cone cells and photoreceptors expressed this protein. Our results of staining experiments on eye discs using an anti-*Aop* antibody coincide with those of previous observations (Figure 8E). The reporter gene expression in the *aop*^{GT1} gene-trap line is found in these undifferentiated cells (Figure 8D). The eye discs of *aop*^{GT1} heterozygotes are subjected to double staining with an anti- β -galactosidase antibody for the detection of the reporter expression and with mAb22C, a marker for differentiated neurons (FUJITA *et al.* 1982). It is found that each antibody labels a distinct group of cells (Figure 8F). These results indicate that the *Gal4* reporter in the gene-trap vector labels the cells in which the trapped gene is normally expressed.

novel gene. The second insertion traps an exon of a different gene contained in an intron of the former gene. The insertion sites of the gene-trap vectors (triangle) and the exon-intron organization of the first trapped gene are schematized in 1. Shown in 2 are the spliced mRNA structures in the gene-trap lines (left) in comparison with those in the wild type (right). The dotted boxes indicate noncoding exonic sequences; the striped boxes show artificial splice sites. In each case, the direction of transcription is indicated by arrows. pr, promoter; E1, E2, E3, and E4, exons; IE1, an exon of the intraintronic gene; as, artificial splicing acceptor site; ds, artificial splicing donor site; 5'P, 3'P, P-element end sequences; EST, sequences that match with those in the BDGP EST library; *env*, *pol*, and *gag* are the genes in an inserted retroviral sequence; ATG, translation start codon.

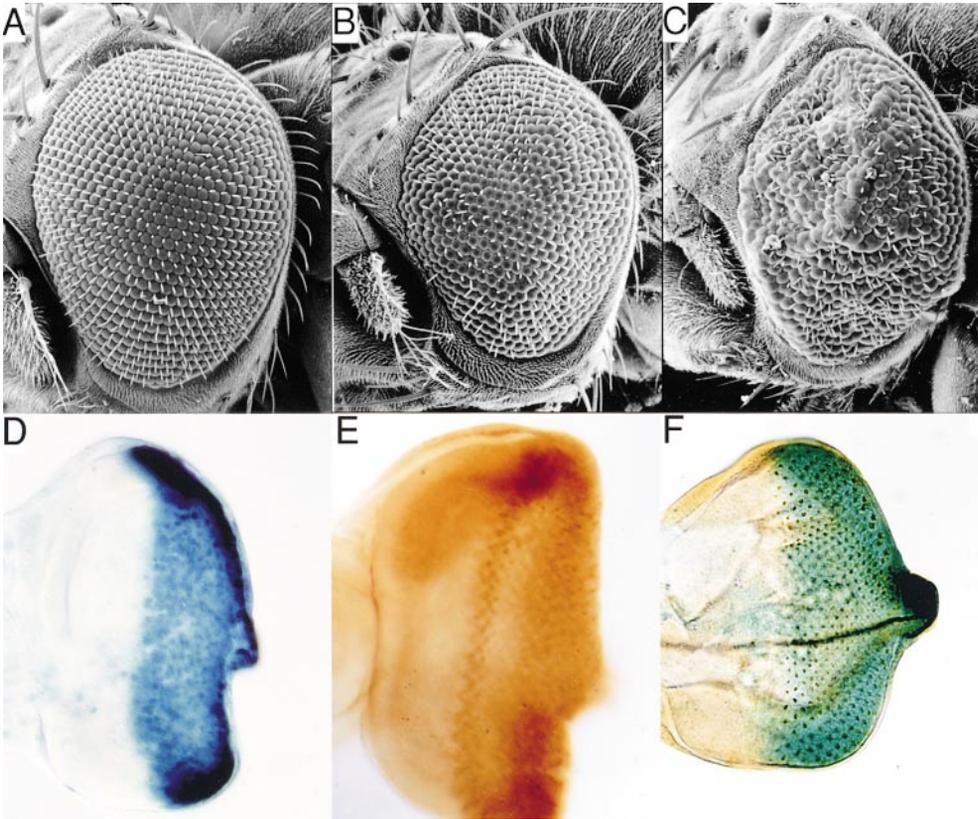


FIGURE 8.—The mutant phenotype associated with a gene-trap allele of *aop*, *aop*^{GT1}. (A–C) Scanning electron microscopic observations of the compound eyes of Canton-S wild type (A), *aop*^{pok3r5} (B), and *aop*^{GT1}/*aop*^{pok3r5} (C). (D–F) The expression pattern of *aop* in a third instar eye disc of a *aop*^{GT1}/+ larva determined by X-gal staining with the Gal4 reporter (D and F) or by an anti-Aop antibody (E). (F) An eye disc doubly stained for β -galactosidase reporter activity with X-gal and for a neuronal antigen recognized by mAb22C10. The anti-Aop antibody and X-gal staining labels basally located undifferentiated cells, whereas mAb22C10 marks apical projections of differentiated photoreceptors. Anterior to the right.

The null mutations in the *aop* locus are embryonic lethal (NÜSSLEIN-VOLHARD *et al.* 1984), as is the *aop*^{GT1} allele. The hypomorphic *aop* mutants exhibit the rough-eye phenotype in varying degrees (YAMAMOTO 1996), and *aop*^{pok1} has a strong eye phenotype among such viable alleles (TEI *et al.* 1992). No *aop*^{GT1}/*aop*^{pok1} adults are obtained, indicating that this allelic combination is lethal. When placed over a chromosome carrying a weak allele, *aop*^{pok3r5}, the *aop*^{GT1} mutation yields a few escaper adults having the rough-eye phenotype (Figure 8B). The roughness of the eyes of *aop*^{pok3r5}/*aop*^{GT1} adults is more severe than that of any of the known *aop* mutants that develop into adults (Figure 8C). These results suggest that *aop*^{GT1} represents a very strong, likely null, mutant in the *aop* locus.

DISCUSSION

The advantages of the newly developed gene-trap system over the conventional enhancer-trap system are threefold. First, all Gal4 gene-trap lines obtained are, in principle, mutants, because the trapped gene produces aberrant chimeric transcripts. This is in contrast to the *Gal4* enhancer trap in which only ~10% of the lines have some discernible phenotype. In the case of enhancer-trap lines, phenotypical silence may result simply from insufficient disruption of the gene by the insert. With the gene-trap construct, the functions of the trapped gene are vitally impaired, even if the gene is

phenotypically silent when mutated. Second, the expression pattern of the reporter gene in gene-trap lines is the same as that of the trapped gene. This is in contrast to the case in the enhancer-trap system where the reporter gene expression merely reflects the activity of the nearest enhancer. Third, the gene-trap method allows us to identify unequivocally the affected gene and to readily clone it, because the transcripts (as with cDNAs) have reporter tags that provide specific primer sequences for PCR.

While the gene-trap system has the above-mentioned advantages, all the important features of the conventional enhancer-trap system are also preserved. Due to the presence of the *Gal4*-coding sequence, any gene of interest can be expressed when introduced into the gene-trap line in the form of a *UAS* fusion (BRAND and PERRIMON 1993). The gene-trap insert can be remobilized by crossing the $\Delta 2-3$ chromosome into the gene-trap lines to generate revertants by precise excision or new alleles by imprecise excision (COOLEY *et al.* 1988).

The crucial step in efficient mutagenesis using the gene-trap vector is the collection of founder males from which the putative gene-trap strains are established by subsequent genetic crosses. The collection of the founder males relies primarily upon the dark eye pigmentation, followed by luciferase assays for the detection of Gal4 expression. Of 27 strains analyzed molecularly, all have a gene interrupted by the gene-trap vector, which results in the production of a fusion mRNA composed of the

host gene and *Gal4*. We did not study in detail the fly strains with dark eye color but without detectable *Gal4* expression. On the basis of the results of PCR, in two of such strains *pGT1* is found to be inserted into the 5' untranslated region of the gene without producing a chimeric mRNA of the host sequence and *Gal4* (M. UMEDA, personal communication). However, the *mini-w* gene in *pGT1* is transcribed even in these fly lines. It is likely that the *mini-w* gene is spliced to the downstream exon of the trapped gene and is polyadenylated, thereby resulting in the dark eye color.

The success of the new gene-trap system depends upon several technical factors. The use of artificial splice sites provides a means to generate two fusion mRNAs, *mini-w* and *Gal4*, each of which can function as an independent marker for integrating a transgene into a transcription unit (*i.e.*, gene-trap event). Another technical factor that is crucial for the success of the present system is the development of the new RACE technology. The triad of vectorette RT-PCR, RACE using multiple cDNA libraries, and PCR using a whole library proves to be a collection of extremely powerful methods for cloning the exon of the trapped gene that could be at a location physically quite remote from the vector insertion point on the chromosome. ESTs instead of sequence-tagged sites can be determined in the screening when the gene-trap system is used. Because the entire *Drosophila* genome has been sequenced and made available, coding sequences of a trapped gene can be identified easily using some ORF-detecting software, such as Genie.

Screening for genomic sequences that produce dominant phenotypes when misexpressed has been employed in several laboratories (RØRTH 1996; RØRTH *et al.* 1998; TOBA *et al.* 1999). The misexpression strategy could be useful for deducing the potential roles of the overexpressed sequence, while the gene-trap system could be used to define the role of the trapped gene on the basis of its loss-of-function phenotype, as in classical genetics. Thus, these two systems are complementary to each other for functional analysis of the genome.

We thank M. Yoshihara, G. M. Rubin, S. Kondo, and Y. Kai for materials and assistance. This study was supported in part by Special Coordination Funds for Promoting Science and Technology from the Science and Technology Agency of Japan and by Waseda University grant No. 200B-029 to D.Y.

LITERATURE CITED

- ALLEN, M. J., A. COLLICK and A. J. JEFFREYS, 1994 Use of vectorette and subvectorette PCR to isolate transgene flanking DNA. *PCR Methods Appl.* **4**: 71–75.
- ARNOLD, C., and I. J. HODGSON, 1991 Vectorette PCR: a novel approach to genomic walking. *PCR Methods Appl.* **1**: 39–42.
- BELLEN, H. J., C. J. O'KANE, C. WILSON, U. GROSSNIKLAUS, R. K. PEARSON *et al.*, 1989 P-element-mediated enhancer detection: a versatile method to study development in *Drosophila*. *Genes Dev.* **3**: 1288–1300.
- BELLO, B., D. RESENDEZ-PEREZ and W. J. GEHRING, 1998 Spatial and temporal targeting of gene expression in *Drosophila* by means of a tetracycline-dependent transactivator system. *Development* **125**: 2193–2202.
- BIER, E., H. VASSIN, S. SHEPHERD, K. LEE, K. MCCALL *et al.*, 1989 Searching for pattern and mutation in the *Drosophila* genome with a *P-lacZ* vector. *Genes Dev.* **3**: 1273–1287.
- BOLWIG, G. M., M. DEL VECCHIO, G. HANNON and T. TULLY, 1995 Molecular cloning of *linotte* in *Drosophila*: a novel gene that functions in adults during associative learning. *Neuron* **15**: 829–842.
- BRAND, A. H., and N. PERRIMON, 1993 Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* **118**: 401–415.
- BRANDES, C., J. D. PLAUTZ, R. STANEWSKY, C. F. JAMISON, M. STRAUME *et al.*, 1996 Novel features of *Drosophila period* transcription revealed by real-time luciferase reporting. *Neuron* **16**: 687–692.
- CHALFIE, M., Y. TU, G. EUSKIRCHEN, W. W. WARD and D. C. PRASHER, 1994 Green fluorescent protein as a marker for gene expression. *Science* **263**: 802–805.
- CHU, K., X. NIU and L. T. WILLIAMS, 1995 A Fas-associated protein factor, FAF1, potentiates Fas-mediated apoptosis. *Proc. Natl. Acad. Sci. USA* **92**: 11894–11898.
- COOLEY, L., R. KELLEY and A. C. SPRADLING, 1988 Insertional mutagenesis of the *Drosophila* genome with single P-elements. *Science* **239**: 1121–1128.
- DUNN, J. M., K. E. ZINSMAIER, B. A. MOZER and S. BENZER, 1993 A new 'gene trap' system for *Drosophila*. *Ann. Rep. Biol. Calif. Inst. Technol.* **1993**: 206.
- FRIEDRICH, G., and P. SORIANO, 1991 Promoter traps in embryonic stem cells: a genetic screen to identify and mutate developmental genes in mice. *Genes Dev.* **5**: 1513–1523.
- FUJITA, S. C., S. L. ZIPURSKY, S. BENZER, A. FERRUS and S. L. SHOTWELL, 1982 Monoclonal antibodies against the *Drosophila* nervous system. *Proc. Natl. Acad. Sci. USA* **79**: 7929–7933.
- JUNI, N., T. AWASAKI, K. YOSHIDA and S. H. HORI, 1996 The *Om(1E)* mutation in *Drosophila ananassae* causes compound eye overgrowth due to *Tom* retrotransposon-driven overexpression of a novel gene. *Genetics* **143**: 1257–1270.
- LAI, Z.-C., and G. M. RUBIN, 1992 Negative control of photoreceptor development in *Drosophila* by the product of the *yan* gene, an ETS domain protein. *Cell* **70**: 609–620.
- LE, S. Y., B. A. SHAPIRO, J. H. CHEN, R. NUSSINOV and J. V. MAIZEL, 1991 RNA pseudoknots downstream of the frameshift sites of retroviruses. *Genet. Anal. Tech. Appl.* **8**: 91–205.
- LINDSLEY, D. L., and G. G. ZIMM, 1992 *The Genome of Drosophila melanogaster*. Academic Press, San Diego.
- MATSUO, T., K. TAKAHASHI, S. KONDO, K. KAIBUCHI and D. YAMAMOTO, 1997 Regulation of cone cell formation by *Canoe* and *Ras* in the developing *Drosophila* eye. *Development* **124**: 2671–2680.
- MIYAMOTO, H., I. NIHONMATSU, S. KONDO, R. UEDA, S. TOGASHI *et al.*, 1995 *canoe* encodes a novel protein containing a GLGF/DHR motif and functions with *Notch* and *scabrous* in common developmental pathways in *Drosophila*. *Genes Dev.* **9**: 612–625.
- MOUNT, S. M., C. BURKS, G. HERTZ, G. D. STORMO, O. WHITE *et al.*, 1992 Splicing signals in *Drosophila*: intron size, information content, and consensus sequences. *Nucleic Acids Res.* **20**: 4255–4262.
- NÜSSLEIN-VOLHARD, C., E. WIESCHAUS and H. KLUDING, 1984 Mutations affecting the pattern of the larval cuticle in *Drosophila melanogaster*. I. Zygotic loci on the second chromosome. *Roux's Arch. Dev. Biol.* **193**: 267–282.
- O'KANE, C. J., 1998 Enhancer-traps, pp. 131–178 in *Drosophila: A practical approach*, Ed. 2, edited by D. B. ROBERTS. IRL Press, Oxford.
- O'KANE, C. J., and W. J. GEHRING, 1987 Detection *in situ* of genomic regulatory elements in *Drosophila*. *Proc. Natl. Acad. Sci. USA* **84**: 9123–9127.
- PIRROTTA, V., 1988 Vectors for P-mediated transformation in *Drosophila*, pp. 437–456 in *Vectors: A Survey of Molecular Cloning Vectors and Their Uses*, edited by R. L. RODRIGUEZ and D. T. DENHARDT. Butterworths, Boston.
- REBAY, I., and G. M. RUBIN, 1995 *Yan* functions as a general inhibitor of differentiation and is negatively regulated by activation of the *Ras1*/MAPK pathway. *Cell* **81**: 857–866.
- ROGGE, R., P. J. GREEN, J. URANO, S. HORN-SABAN, M. MLODZIK *et al.*, 1995 The role of *Yan* in mediating the choice between cell division and differentiation. *Development* **121**: 3947–3958.
- RØRTH, P., 1996 A modular misexpression screen in *Drosophila* de-

- tecting tissue-specific phenotypes. *Proc. Natl. Acad. Sci. USA* **93**: 12418–12422.
- RØRTH, P., K. SZABO, A. BAILEY, T. LAVERTY, J. REHM *et al.*, 1998 Systematic gain-of-function genetics in *Drosophila*. *Development* **125**: 1049–1057.
- SPRADLING, A. C., D. M. STERN, I. KISS, J. ROOTE, T. LAVERTY *et al.*, 1995 Gene disruptions using P transposable elements: an integral component of the *Drosophila* genome project. *Proc. Natl. Acad. Sci. USA* **92**: 10824–10830.
- STRAUSFELD, N. J., 1976 *Atlas of an Insect Brain*. Springer-Verlag, Berlin.
- TEI, H., I. NIHONMATSU, T. YOKOKURA, R. UEDA, Y. SANO *et al.*, 1992 *pokkuri*, a *Drosophila* gene encoding an E-26-specific (Ets) domain protein, prevents overproduction of the R7 photoreceptor. *Proc. Natl. Acad. Sci. USA* **89**: 6856–6860.
- TOBA, G., T. OHSAKO, N. MIYATA, T. OHTSUKA, K.-H. SEONG *et al.*, 1999 The gene search system: a method for efficient detection and rapid molecular identification of genes in *Drosophila melanogaster*. *Genetics* **151**: 725–737.
- WILSON, C., R. K. PEARSON, H. J. BELLEN, C. J. O'KANE, U. GROSSNIKLAUS *et al.*, 1989 P-element-mediated enhancer detection: an efficient method for isolating and characterizing developmentally regulated genes in *Drosophila*. *Genes Dev.* **3**: 1301–1313.
- YAMAMOTO, D., 1996 *Molecular Dynamics in the Developing Drosophila Eye*. R. G. Landes Co., Austin, TX.
- YEH, E., K. GUSTAFSON and G. L. BOULIANNE, 1995 Green fluorescent protein as a vital marker and reporter of gene expression in *Drosophila*. *Proc. Natl. Acad. Sci. USA* **92**: 7036–7040.
- ZAMBROWICZ, B. P., G. A. FRIEDRICH, E. C. BUXTON, S. L. LILLEBERG, C. PERSON *et al.*, 1998 Disruption and sequence identification of 2,000 genes in mouse embryonic stem cells. *Nature* **392**: 608–611.

Communicating editor: N. TAKAHATA