

A piggyBac Transposon Gene Trap for the Analysis of Gene Expression and Function in *Drosophila*

Christopher P. Bonin and Richard S. Mann¹

Department of Biochemistry and Molecular Biophysics and Center for Neurobiology and Behavior,
Columbia University, New York, New York 10032

Manuscript received February 11, 2004

Accepted for publication May 7, 2004

ABSTRACT

P-element-based gene and enhancer trap strategies have provided a wealth of information on the expression and function of genes in *Drosophila melanogaster*. Here we present a new vector that utilizes the simple insertion requirements of the piggyBac transposon, coupled to a splice acceptor (SA) site fused to the sequence encoding enhanced green fluorescent protein (EGFP) and a transcriptional terminator. Mobilization of the piggyBac splice site gene trap vector (PBss) was accomplished by heat-shock-induced expression of piggyBac transposase (PBbase). We show that insertion of PBss into genes leads to fusions between the gene's mRNA and the PBss-encoded EGFP transcripts. As heterozygotes, these fusions report the normal pattern of expression of the trapped gene. As homozygotes, these fusions can inactivate the gene and lead to lethality. Molecular characterization of PBss insertion events shows that they are single copy, that they always occur at TTAA sequences, and that splicing utilizes the engineered splice site in PBss. In those instances where protein-EGFP fusions are predicted to occur, the subcellular localization of the wild-type protein can be inferred from the localization of the EGFP fusion protein. These experiments highlight the utility of the PBss system for expanding the functional genomics tools that are available in *Drosophila*.

UNDERSTANDING a gene's function usually depends on knowing its pattern of expression, the subcellular localization of its product, and the consequences of knocking out its activity. Various methods have been employed to accomplish these goals in *Drosophila melanogaster*, as well as other model genetic organisms. Enhancer trap strategies have in particular provided an abundance of data on the endogenous expression patterns of both known and unknown genes (O'KANE and GEHRING 1987; BELLEN *et al.* 1989; BIER *et al.* 1989; WILSON *et al.* 1989). Enhancer trap constructs generally rely on a "minimal" promoter fused to a reporter gene (usually *lacZ* or *GFP*) and a transposition system (usually *P* elements in *Drosophila*) to insert the element randomly throughout the genome. Productive insertions typically report a β -galactosidase (β -gal) expression pattern that is dictated by endogenous chromosomal enhancers. Reporter gene expression patterns are generally used to infer the expression patterns of nearby genes that are assumed to be under the control of the same enhancers. For some insertions, however, the enhancer trap element may not respond to chromosomal enhancers due to problems in promoter-enhancer incompatibility. Furthermore, identifying the gene and regulatory sequences that have been trapped by an ele-

ment can be difficult in some cases, especially when the insertion is located many kilobases from the gene.

In addition to providing a readout of a gene's expression pattern, once an enhancer trap insertion is obtained, it is often possible to mutate the gene, allowing for a subsequent genetic analysis. Although in some cases the initial insertion event may cause a mutation, in most cases imprecise excision or other methods must be used to delete DNA surrounding the site of insertion. This step can be very time consuming and in some cases unsuccessful. All current enhancer trap systems are also limited by the insertion preferences of the transposition system being used. For *P* elements, the most commonly used system in *Drosophila*, some work has suggested that they preferentially insert into upstream, or promoter, regions of genes (SPRADLING *et al.* 1995). Furthermore, *P*-element-insertion hotspots and coldspots have left some genes untrapped. For this reason, some researchers have utilized the Hobo transposition system to expand the repertoire of trapped genes (BLACKMAN *et al.* 1989). Another important advance was using the yeast transcriptional activator *Gal4* as a reporter instead of *lacZ* (BRAND and PERRIMON 1993). This approach has the enormous benefit of allowing researchers to express not only *lacZ*, but also any open reading frame (ORF) of interest.

Enhancer trapping methods have been complemented by the development of elements lacking promoters, which therefore rely on an endogenous gene's promoter

¹Corresponding author: Department of Biochemistry and Molecular Biophysics, Columbia University, 701 W. 168th St., HHSC 1104, New York, NY 10032. E-mail: rsm10@columbia.edu

and enhancer(s) to report expression. This method usually uses a splice acceptor (SA) site upstream of a reporter gene. Insertion of the element within a gene's intron "traps" the gene's normal transcription. Some constructs have been designed to terminate transcription in the hope of mutating the gene as a homozygote (GOSSLER *et al.* 1989; SKARNES *et al.* 1992; NIWA *et al.* 1993; SKARNES *et al.* 1995; LUKACSOVICH *et al.* 2001; MORIN *et al.* 2001). Elements for use in mammalian systems have included internal ribosome entry sites (IRESs), which provide the added benefit of driving two different proteins from the same transcript (ZAMBROWICZ *et al.* 1998; LEIGHTON *et al.* 2001). Some of the mammalian vectors depend on "random" insertion events, while others have been used to target specific genes by homologous recombination. Other elements have also included a splice donor site downstream of the reporter gene (NIWA *et al.* 1993; MORIN *et al.* 2001). In those cases in which the frames of both upstream and downstream exons match the frame of the reporter gene, the reporter is seamlessly inserted into an otherwise intact transcript.

The complete *D. melanogaster* genome sequence has made gene identification easier, but genetics is still required to understand a gene's function. Continued improvement of functional genomic strategies, such as those described above, is important for genetics research to proceed in the postgenome era. With this goal in mind, we have developed a new element that is designed to both report a gene's expression and mutate its function in *Drosophila*. The transposition system we use is based on the piggyBac element, originally derived from the cabbage looper moth (CARY *et al.* 1989; O'BROCHTA and ATKINSON 1996). piggyBac elements are known to transpose in *Drosophila* and have a very simple insertion site preference, TTAA. Inserted into a piggyBac backbone is an SA site fused to the *enhanced GFP (EGFP)* gene followed by a transcriptional terminator. We describe our ability to utilize this element for trapping and reporting a gene's transcriptional expression profile. We further demonstrate that in many cases the insertion leads to lethality. Finally, in those cases in which the *EGFP* sequences are spliced in frame to an endogenous ORF we show that information can be obtained about the targeted protein's subcellular localization.

MATERIALS AND METHODS

DNA constructs and transgenic lines: The structure and components of the piggyBac splice site (PBss) gene trap element and the piggyBac transposase (PBase; including their *Drosophila* transformation vector counterparts, C4-PBss and PW8-PBase, respectively) are shown in Figure 1. The piggyBac transposon (5' and 3' sequences) as well as the heat-shock piggyBac transposase sequences derived from the cabbage looper moth (*Tricoplusia ni*) were provided by R. A. Harrell (USDA). In the *EGFP* sequences, which were derived from *Hsp70-3X-flu-EGFP-flu* and obtained from M. Zecca (Columbia

University), the *EGFP* coding sequence is fused to three copies of an epitope from the influenza hemagglutinin protein (CHALFIE *et al.* 1994; ZECCA and STRUHL 2002). The PBss construct contains a 3X-flu-*EGFP* ORF but will be referred to as *EGFP* throughout this manuscript. The splice acceptor site and engineered branch site were designed on the basis of known splice acceptor sequences in *D. melanogaster* (MOUNT *et al.* 1992) and created as an oligo (MWG Biotech, High Point, NC). Primers corresponding to other portions of the constructs were designed to incorporate appropriate restriction sites for efficient cloning into pBluescript SK+ to create pB-PBss. A *KpnI/PvuII* fragment from pB-PBss was inserted into Casper4 (where *P*-element *lacZ* sequences were removed) cut with *KpnI/StuI*, creating C4-PBss. An *EcoRI/NdeI* (T4 polymerase filled) fragment containing the heat-shock piggyBac transposase (hs-PBase) derived from pBdSac (HANDLER and HARRELL 1999) was cloned into *EcoRI/SmaI*-cut pW8 vector, creating pW8-PBase. Molecular cloning was performed using standard procedures (SAMBROOK *et al.* 1989). Transgenic lines were created by co-injecting each construct separately with *P*-element $\Delta 2/3$ helper plasmid. At least four lines for each construct were obtained. None of the C4-PBss transformants expressed *EGFP*.

Mobilization and screening: For the initial mobilization experiment, C4-PBss flies were crossed to hs-PBase flies and heat-shocked at 37° for 30 min for 3 successive days starting ~48 hr after egg laying (AEL). Screening was accomplished by taking 10 male progeny from this cross carrying the C4-PBss construct (differences in eye color allowed us to follow the original insertion) and crossing them to ~50 *yw* virgin females. Cages were constructed to allow collection of embryos from this cross on apple juice agar plates. These plates were changed every morning and examined periodically during embryonic and larval development under a UV dissecting scope for *EGFP*-positive progeny. *EGFP*-positive embryos and larvae were transferred to standard culture vials to complete development followed by crossing to balancer stocks. The segregation of inserts relative to balancer chromosomes was used to map inserts to chromosomes. Homozygous lethal inserts were maintained over balancers.

Subsequent mobilization experiments were used to determine the approximate frequency of gene trapping events. This was accomplished by crossing hs-PBase flies to flies with one of the initial PBss insertions, *scarface^{PBss}*, and either heat-shocking larvae as described above or heat-shocking adult male hs-PBase; *scarface^{PBss}* flies. For each heat-shock protocol, 10 male flies were crossed to ~50 *yw* virgin females. Progeny from these crosses were screened as described above. Numbers of *EGFP*-positive embryos carrying new *EGFP* expression patterns were counted over a 5- to 6-day period.

Insertion site identification and analysis: Insertion sites were identified by extracting DNA from PBss-trapped flies and performing inverse PCR using previously described methods (<http://www.fruitfly.org/about/methods/inverse.pcr.html#5'pz>) with some modifications. Following extraction, DNA was further purified by phenol/chloroform extraction, precipitation, and resuspension using standard methods (SAMBROOK *et al.* 1989). Modification to the inverse PCR method involved digesting 10 μ l of DNA with *NdeI* or *XbaI*, heat-inactivation of the enzyme, increasing the volume to 400 μ l, and ligating overnight at 4°. Ligations were then precipitated and resuspended in 50 μ l of 10 mM Tris/Cl, pH 8.0. PCR amplification was accomplished by using 2 μ l of ligation reaction, PBss-specific primers (PBssi1 and PBssi2), and Expand Long Template (Roche Applied Science, Indianapolis). Reaction conditions were as follows: 94° for 2 min and 30 cycles of 94° for 15 sec, 55° for 30 sec, 72° for 3.5 min, followed by 72° for 10 min. Single bands were obtained for all of the lines and were

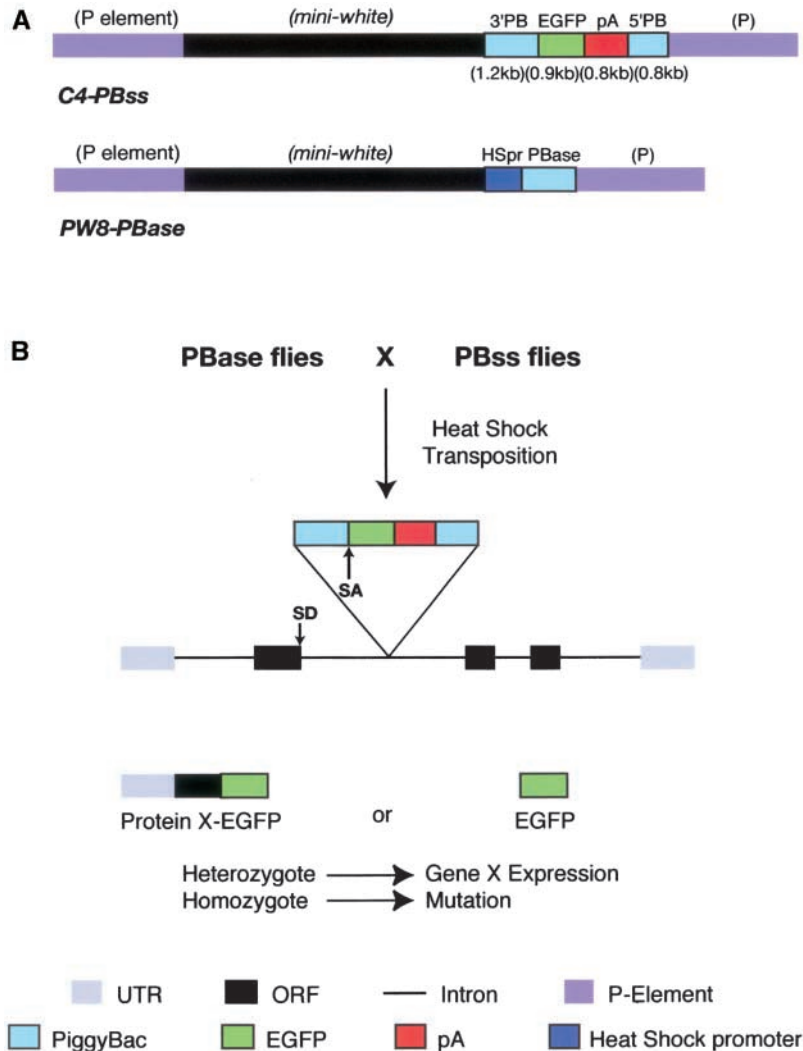


FIGURE 1.—Design and function of the piggyBac transposon gene trap system. (A) 5' and 3' piggyBac transposon sequences flank an engineered 3' intronic splice acceptor site fused to an enhanced green fluorescent protein (*EGFP*) coding region and an α -tubulin 3'-untranslated region containing a poly(A) (pA) site for transcriptional termination. These sequences were integrated into the fly genome by cloning into a Casper4 vector with a *mini-white* selectable marker (C4-PBss). The driver line (PBBase) was created by cloning a heat-shock promoter-piggyBac transposase into a PW8 vector. This was then integrated into the fly genome and selected for using *white*+ selection. (B) Mobilization of the PBss construct is accomplished by crossing flies carrying this construct with those carrying a heat-shock-driven piggyBac transposase (PBBase) and applying heat during germline formation. Reinsertion of a properly oriented PBss sequence into an intron of gene X will result in F₂ heterozygote flies expressing protein X-EGFP or EGFP at the appropriate spatial and temporal parameters dictated by the trapped gene. Flies homozygous for a PBss gene insertion may result in null mutations or EGFP-tagged proteins, allowing the assessment of gene function and/or protein domain localization. *EGFP* sequences contain an in frame 3X-*flu* tag sequence at the 5' end, which adds a 3X-*flu* tag to the N terminus of the translated EGFP sequence. SA, splice acceptor site; SD, splice donor site.

sequenced by the Columbia University DNA Sequencing Facility using the same primers. In addition to the PBss insertions listed in Table 1, we identified the insertion site of our starter insertion, C4-PBss, to be at 88D.

To determine the number of PBss insertions in each line, ~2 μ g of DNA derived from each line, including the transposase line, PBBase, and the starter transposon, C4-PBss, was digested with either *EcoRI* and *NdeI* or *SacI* and *NdeI* and probed with a ³²P-labeled *SacI/EcoRI* EGFP probe derived from the PBss plasmid. Autoradiography was performed using Kodak X-ray film and an X-omat developer (Kodak).

Transcript analysis: RNA was extracted from three different PBss-trapped lines, *scarface*^{PBss}, *Akap200*^{PBss}, and *Vha100-2*^{PBss}, using Trizol Reagent (Life Technologies, Frederick, MD); reverse transcribed with MMLV reverse transcriptase (New England Biolabs, Beverly, MA); and amplified using Taq Polymerase (Roche Applied Science) with gene- and PBss-specific primers (sequences available upon request). Sequencing was performed by the Columbia University DNA Sequencing Facility using the same primers.

Immunolocalization experiments: Embryos and discs used for antibody staining were prepared using standard methods. Primary antibodies used included: mouse anti-Wg, guinea pig Anti-Hth, mouse anti-Elav, and Texas red-labeled anti-HRP. Secondary antibodies used included: Texas red-conjugated

goat anti-mouse and CY5-conjugated goat anti-guinea pig antibodies (Jackson ImmunoResearch Laboratories, West Grove, PA).

Microscopy: Embryos and discs stained with antibodies were visualized and photographed on a Bio-Rad confocal imager using Bio-Rad 2000 software (Bio-Rad Laboratories, Hercules, CA).

RESULTS

Design and construction of the PBss gene trap vector: The PBss construct is a small mobile element created for the purpose of trapping a gene's transcriptional expression profile by reporting EGFP, which becomes fused to the gene's native transcript by mRNA splicing (Figure 1). The piggyBac transposon is derived from the cabbage looper moth, *Trichoplusia ni*. This element was chosen because of its loose DNA target sequence preference, TTAA (CARY *et al.* 1989; O'BROCHTA and ATKINSON 1996; HANDLER and HARRELL 1999; LOBO *et al.* 1999; LI *et al.* 2001). Similar to P elements, piggyBac requires transposon-derived sequences, designated as

TABLE 1
Genes identified by PBss

Gene	Protein type/function	Insertion site	Chr./cyt.	Phenotype	RT-PCR	Insertion site
<i>scarface</i>	Serine peptidase-like	Third intron	2R/41F8	Pupal lethal	+	aagccaTTAAggcagc
<i>akap200</i>	A kinase accessory protein	Fifth intron	2L/29C4	Lethal		tagttgTTAAatagta
<i>Vha100-2</i>	V-type ATPase	First exon	3R/91A5	Lethal	+	gaatcaTTAAacaaca
<i>l(2)01424</i>	Translation initiation factor	First intron	2R/49E1	Lethal		aatcaTTAAaccaag
<i>Hsc70-4</i>	Heat-shock cognate	First intron	3R/88E4	Lethal	+	atataTTAAgtatat
<i>Act5c</i>	Actin	First intron	X/5C7	Homozygous viable		tagtgtTTAAagccea
<i>TMII</i>	Tropomyosin	Third intron	3R/88E12-13	Lethal		tctatcTTAAagtat
<i>Idh</i>	Isocitrate dehydrogenase	First intron	3L/66C8	Lethal		tcaataTTAAgattat
<i>desat2/Fad</i>	Fatty acid desaturase	First intron	3R/87B10	ND ^a		agccgaTTAAagcaac
<i>CG16750</i>	Zinc finger	Fourth intron	3R/85D11	ND ^a		tatgcaTTAAgaatat

All insertions reported in this table are derived from the C4-PBss starter insertion at 88E (data not shown). As two of the PBss insertions are also located at 88E, PBss may, like *P* elements, have a preference to insert locally. In all cases, the TTAA sequence present at the site of insertion was present both upstream and downstream of the PBss element. ND, not done; +, an RT-PCR analysis of these inserts was carried out (see Figure 2).

^a Although these two insertions are lethal as homozygotes, they retain the C4 insertion, which is also lethal as a homozygote. Therefore, we have not characterized the phenotype of the PBss insertions on their own.

5' and 3', at each end for transposition to occur. A piggyBac-derived transposase (PBase) catalyzes the transposition of elements containing these sequences (Figure 1A). Between the piggyBac 5' and 3' sequences, a consensus splice site was created upstream of an *EGFP* gene followed by an α -tubulin 3'-untranslated region (pA), for transcriptional termination (Figure 1A). An ATG initiation codon for the *EGFP* gene was placed in the first translated frame since the majority of known introns in *Drosophila* end in this frame (LONG and DEUTSCH 1999; LONG *et al.* 2003). The piggyBac transposase element is composed of a heat-shock (hs) promoter fused to the piggyBac transposase coding sequence (hs-PBase; Figure 1A; HANDLER and HARRELL 1999).

Mobilization, screening, and identification of transgenic flies: To create transgenic flies carrying these constructs, PBss was cloned into the Casper4 vector (C4-PBss) and PBase was cloned into the PW8 vector (PW8-PBase; Figure 1A) to allow *white*⁺ selection of transformed progeny. All of the *w*⁺ C4-PBss transformants failed to express EGFP. Once multiple stable lines were established for each construct, they were crossed to each other and heat-shocked three times at 37° for 30 min for 3 successive days after larval development had begun (~24 hr AEL; Figure 1B and see MATERIALS AND METHODS for details). EGFP-positive embryos and larvae were isolated with the use of a UV light source and a dissecting microscope. From this initial screen we isolated 25 EGFP-expressing lines from ~24,000 embryos and larvae. *EGFP*-expressing lines were balanced and a subset of these lines was used for subsequent analysis (Table 1 and see below).

To determine if PBss insertions can be remobilized, two additional crosses were performed (see MATERIALS AND METHODS). One experiment involved mobilizing PBss during larval development from a previously trapped

gene, CG11066^{PBss}, which we have named *scarface*^{PBss} (see below). A second screen involved mobilizing PBss in adult male flies carrying the same *scarface*^{PBss} insertion. In both screens, only embryos were screened for new EGFP⁺ patterns, thus limiting the number of new insertions obtained. Nevertheless, both experiments generated novel EGFP⁺ insertions at similar frequencies, indicating that PBss elements can be remobilized during both larval and adult stages. The combined number of new EGFP⁺ patterns identified in these two experiments was 4 of ~14,000 embryos. Since *scarface*^{PBss} represents a single insertion as determined by Southern analysis (see MATERIALS AND METHODS), we are confident that new expression patterns represent remobilization events. Many of these stocks have been maintained in the presence of the hs-PBase element without a loss of EGFP expression, suggesting that PBss elements transpose rarely or not at all in the presence of hs-PBase, as long as no heat shock is provided (see Figure 1).

Insertion site identification: Southern blot analysis was used to confirm that all EGFP⁺ lines contain only a single PBss element (data not shown). BLAST analysis (ALTSCHUL *et al.* 1990) of sequenced inverse PCR products was used to determine the sites of PBss insertion within a subset of the trapped lines. Table 1 lists information on the trapped genes. Analysis of the insertion sites has confirmed that PBss always inserts into a TTAA sequence, with no other obvious sequence preferences (Table 1). This sequence preference for piggyBac elements is consistent with previous observations and supports the idea that PBss will be able to insert widely throughout the genome, with minimal preferences for specific sequences (CARY *et al.* 1989; ELICK *et al.* 1996; HANDLER and HARRELL 1999).

Examination of the insertion sites listed in Table 1 illustrates that PBss transposed to all chromosome arms

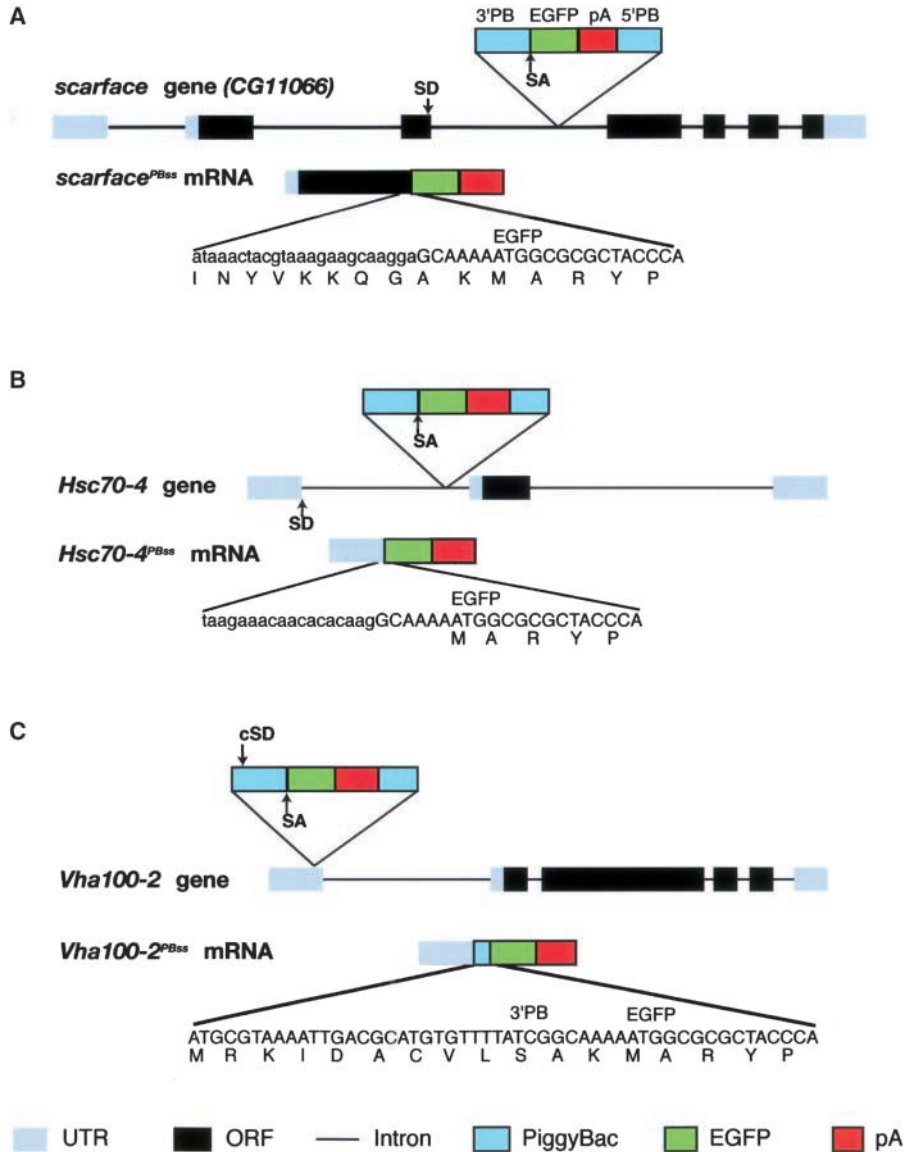


FIGURE 2.—Transcript and predicted protein fusions for (A) *scarface*^{PBss}, (B) *Hsc70-4*^{PBss}, and (C) *Vha100-2*^{PBss}. RT-PCR transcript analyses were performed for all three trapped genes revealing two expected [*scarface*^{PBss} (A) and *Hsc70-4*^{PBss} (B)] and one unexpected [*Vha100-2*^{PBss} (C)] mRNA fusion product with the *EGFP* coding region of PBss. *EGFP* sequences contain an in frame 3X-flu tag sequence at the 5' end, which adds a 3X-flu tag to the N terminus of the translated *EGFP* sequence. SA, splice acceptor site; SD, splice donor site.

and, as expected, usually to a predicted intron. In addition, most of the insertions are lethal as homozygotes. One initially surprising finding from this analysis was the viability of an insertion into the first intron of the *Act5c* gene (Table 1). However, the *Act5c* gene has a second promoter that is downstream of the PBss insertion site (BOND and DAVIDSON 1986; CHUNG and KELLER 1990). The presence of an alternative promoter downstream of the PBss element can account for the viability of this insertion as a homozygote. These results suggest that transcripts initiating from the more distal promoter of the *Act5c* gene are not required for viability.

mRNA splicing utilizes the PBss splice acceptor: To determine the ability of the PBss element to splice as designed we used RT-PCR to analyze the mRNA transcripts from three of the trapped lines (Figure 2; Table 1). Results from this analysis revealed that the engineered splice acceptor site upstream of *EGFP* in PBss

was utilized in all three cases (Figure 2). In two of these three examples, the PBss element inserted into an intron and the splice donor site was derived from an upstream exon as anticipated (Figure 2, A and B). In one case PBss inserted into an exon of the gene *Vha100-2* (Figure 2C). Although the engineered splice acceptor site was used in this case, the splice donor site was a cryptic site present within the 3' piggyBac sequences (Figure 2C). The insertion into *Vha100-2* illustrates that insertions into exons can, in some instances, also lead to a productive gene trap.

Expressional analysis: Some examples of the *EGFP* expression patterns obtained from these screens are shown in Figure 3. Close inspection of embryos from some PBss-trapped lines revealed that *EGFP* was localized to specific cellular compartments (Figure 3, A, D, and F), as opposed to a fairly uniform cellular distribution, which is expected for *EGFP* alone. This suggests

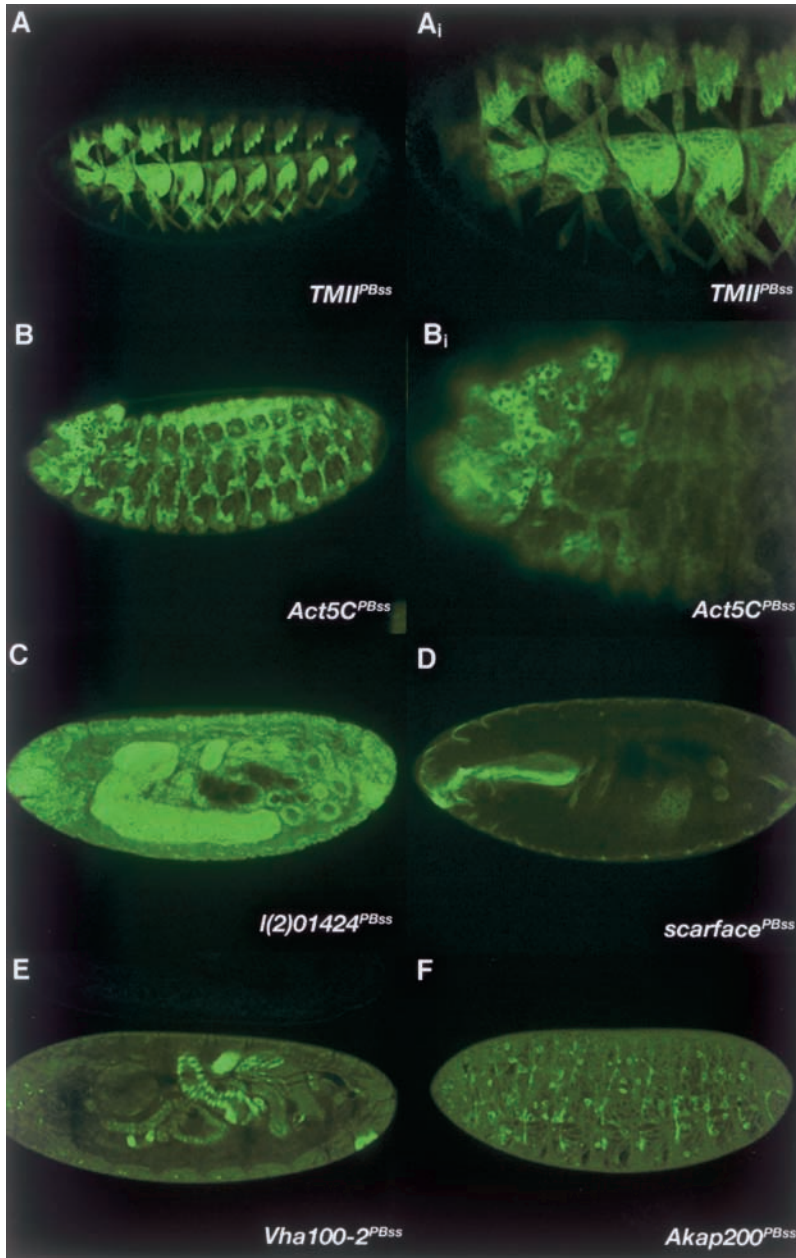


FIGURE 3.—Embryonic EGFP expression patterns of six representative PBss-tagged genes. (A) *TMII*^{PBss}, tropomyosin II protein. (A_i) A close-up of the anterior region of a *TMII*^{PBss} embryo. (B) *Act5C*^{PBss}, actin5c protein. (B_i) A close-up of the anterior region of an *Act5C*^{PBss} embryo. (C) *l(2)01424*^{PBss}, a putative translational initiation factor. (D) *scarface*^{PBss} (*CG11066*), a putative trypsin-like serine protease. (E) *Vha100-2*^{PBss}, a V-type ATPase. (F) *Akap200*^{PBss}, a kinase anchor protein.

that in these cases EGFP is fused to an endogenous protein with specific localization properties. Sequencing of the insertion sites has revealed that this is in fact the case for some lines (Table 1 and see below).

Two lines in particular, *Akap200*^{PBss} and *scarface*^{PBss}, exhibit interesting subcellular localization properties at high magnification (Figures 4 and 5, respectively). The *Akap200* gene encodes an A kinase accessory protein shown to interact with signaling pathway components. The protein is predicted to be membrane bound on the basis of the presence of a myristoylation sequence in its N terminus (LI *et al.* 1999; ROSSI *et al.* 1999). Previous work has shown that the Akap200 protein is localized at the cell surface in developing germline cells (JACKSON and BERG 2002). Analysis of the EGFP expres-

sion in developing germline cells of *Akap200*^{PBss}/+ animals revealed strong expression in the germarium of the developing oocytes as well as in the nervous system (Figures 3 and 4). Closer inspection of EGFP expression in the developing nervous system of *Akap200*^{PBss}/+ embryos suggests that it is localized to the membrane of developing axons of the ventral nerve cord as judged by costaining with a pan-axonal marker, anti-HRP (Figure 4). These results suggest a putative role for this Akap200 in participating in signaling in developing or mature axons, a previously unknown function for this protein.

Description of the PBss insertion into *CG11066/scarface*: On the basis of the phenotype of homozygous escapers of a PBss insertion into *CG11066*, we have

named this gene *scarface* (see below). Here, we describe a developmental expression profile and preliminary analysis of the mutant. On the basis of BLAST comparisons (ALTSCHUL *et al.* 1990), *scarface* is most similar to trypsin-like serine proteases that are secreted or are associated with the outer membrane of cells. The *scarface*^{PBss} insertion was initially identified due to a strong larval EGFP expression pattern in the anterior and posterior termini of third instar larvae. Subsequent analysis showed EGFP expression in many tissues, including wing imaginal discs (Figure 5). Costaining with extracellular and nuclear markers, Wg and Hth, respectively, in the wing discs suggests that *scarface* may be membrane associated or secreted (Figure 5). *scarface*^{PBss/+} adult flies exhibit EGFP expression in the proboscis and in the socket cells of all micro- and macrochaete analyzed (Figure 6, C–F, and data not shown). Finally, most *scarface*^{PBss} homozygotes are pupal lethal. However, some homozygote escapers occasionally eclose. These flies exhibit what appears to be “scarring” around the mouthparts that is indicative of necrotic tissue (Figure 6, G and H).

DISCUSSION

This work describes the design and implementation of a new gene trap system for Drosophila. This system uses a piggyBac transposon to provide a different and, hopefully, broad spectrum of chromosomal insertion sites for functional genomic studies compared to that provided by previous P-element-based systems. We have demonstrated the ability of the PBss element to insert into new genes, splice with endogenous transcripts, and report the expression patterns of trapped genes. Further, a preliminary genetic analysis indicates that many of the insertions are lethal or mutant as homozygotes. Thus, the implementation of this system should significantly aid the analysis of gene function and expression in *D. melanogaster*. Below we focus this discussion on the design of the PBss element, its apparent limitations, and likely improvements.

PBss design: The PBss element was designed to trap a gene's expression pattern using the gene's normal transcriptional and splicing mechanisms. Because PBss has no promoter to initiate *EGFP* transcription, productive (EGFP+) insertions must occur downstream of an endogenous gene's promoter. A second requirement is that PBss must insert in such a manner to allow the in frame translation of EGFP. All of the insertions characterized here fulfill both of these requirements. As expected, most of these insertions were in introns. In these examples, expression of EGFP depended on mRNA splicing between an upstream exon and the engineered splice site 5' to *EGFP*. For those insertions that occurred downstream of a coding exon, EGFP was expressed as a fusion protein. In contrast, for those insertions that occurred downstream of a noncoding exon, translation

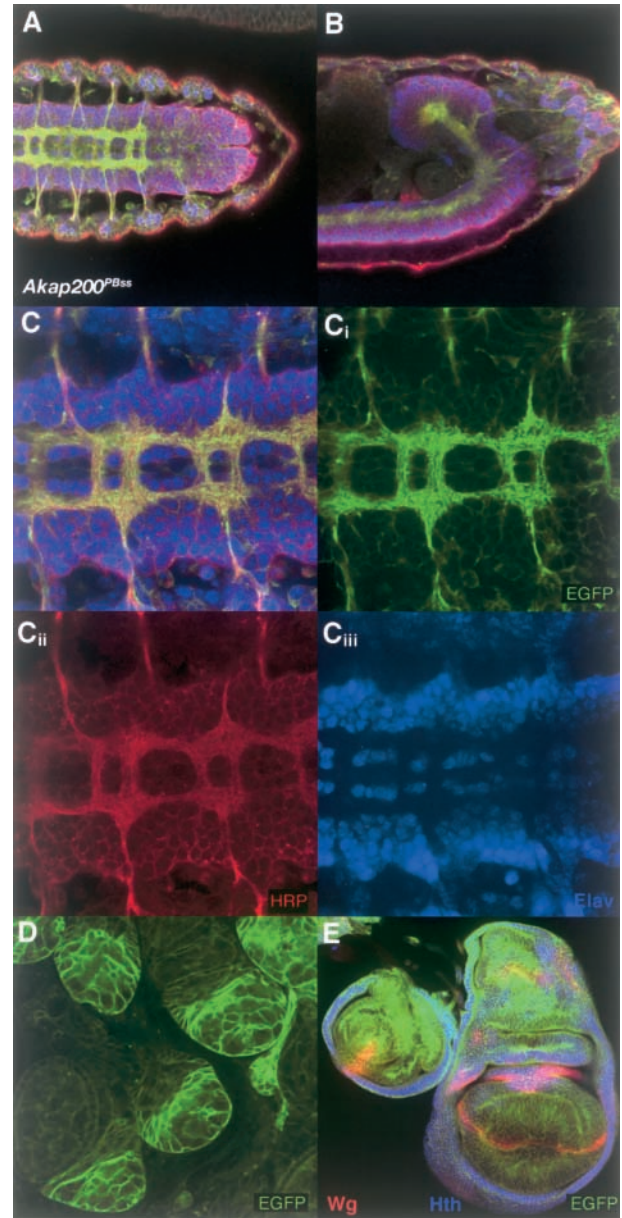


FIGURE 4.—Immunofluorescence staining of *Akap200*^{PBss} embryos, germarium cells, and third instar larval wing and leg discs. *Akap200*^{PBss} EGFP-expressing embryos were stained with HRP (red) to label axons and Elav (blue), a pan-neuronal nuclear marker, to label neurons. (A) Ventral view of *Akap200*^{PBss} embryos. (B) Side view of *Akap200*^{PBss} embryos. (C) High magnification view of the developing ventral nerve cord. (C_i) *Akap200*^{PBss} EGFP alone. (C_{ii}) HRP alone. (C_{iii}) Elav alone. (D) *Akap200*^{PBss} germline cells exhibiting EGFP expression in germarium cells. (E) Third instar wing and leg discs expressing *Akap200*^{PBss} EGFP and stained with an extracellular marker, Wg (red), and a nuclear marker, Hth (blue).

of EGFP was initiated from its own ATG. In one unusual case, we found an insertion into an exon. In this case, splicing still occurred and EGFP was expressed as a fusion protein. This example illustrates that some PBss insertions into exons can also be productive.

We also engineered a transcriptional stop sequence

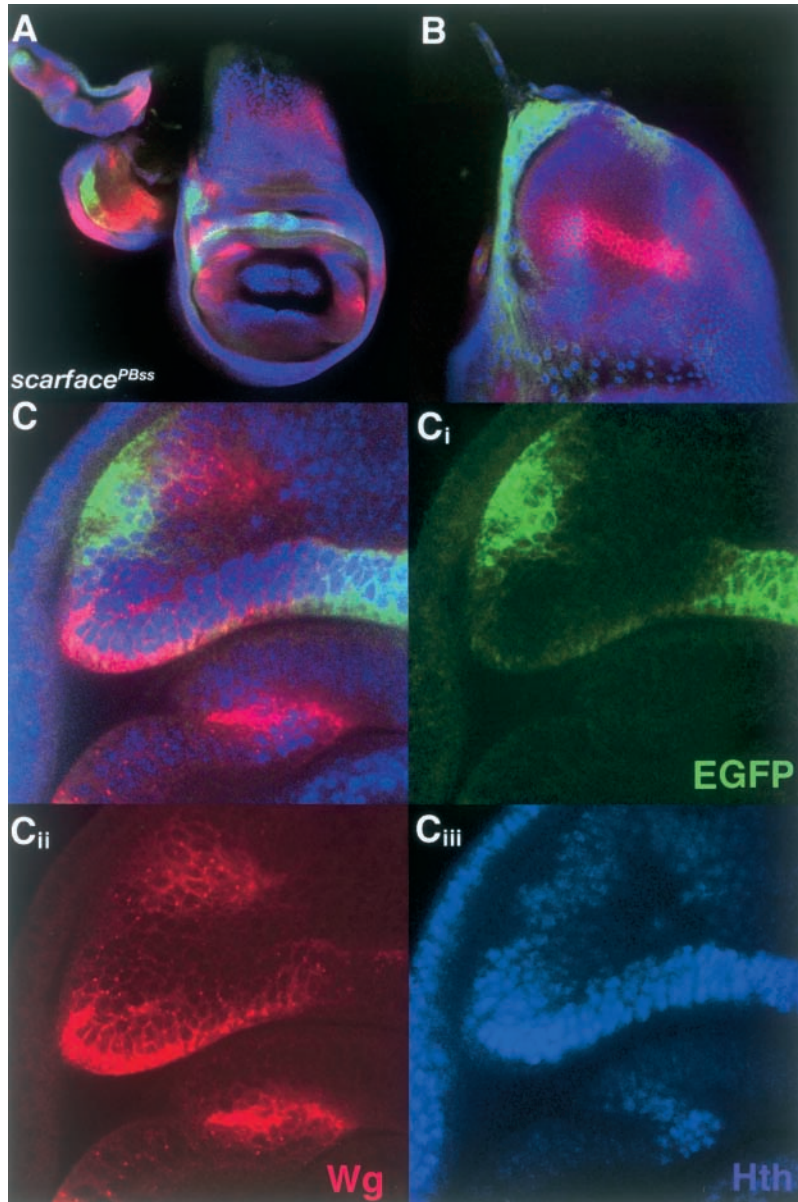


FIGURE 5.—Immunofluorescence staining of *scarface*^{PBss} third instar larval wing discs with markers labeling nuclear (Homothorax, Hth) and extracellular (Wingless, Wg) compartments. (A) Wing, haltere, and leg discs. (B) Close-up of notum section of wing disc. (C) High magnification view of anterior section of wing disc. (C_i) *scarface*^{PBss} EGFP alone. (C_{ii}) Wg alone. (C_{iii}) Hth alone.

downstream of the *EGFP* sequences to terminate transcription of trapped genes. The reasoning behind this aspect of the design was the hope that most productive insertions would also cause a mutation, allowing a subsequent genetic analysis of the gene to be carried out. On the basis of the frequency of lethal and mutant insertions, it is likely that transcription termination is occurring due to the PBss insertion. This aspect of the PBss design contrasts with other gene trap vectors that have splice sites on both sides of the reporter gene, allowing the incorporation of the reporter sequences into an otherwise intact transcript (ZAMBROWICZ *et al.* 1998; LUKACSOVICH *et al.* 2001; MORIN *et al.* 2001). In the latter case, a mutation of the endogenous gene is much less likely. Thus, in these systems subsequent

genetic analysis would require a second step to delete the element and neighboring chromosomal DNA to generate a mutation.

Finally, the PBss vector described here uses a piggyBac-based transposition system. The reason for choosing this system was to allow for a broad and unbiased set of insertion sites within the *Drosophila* genome. Consistent with previous analysis of piggyBac transposition preferences, all of the insertions identified here inserted into a TTAA sequence (Table 1; HANDLER and HARRELL 1999). Since the *D. melanogaster* genome is very AT rich, the chances that a TTAA sequence will be found within most introns is high. Although it is possible that piggyBac elements may also exhibit insertional hotspots, our preliminary findings suggest that this system will provide

a larger and less-biased set of insertion sites than that provided by *P*-element-based vectors (SPRADLING *et al.* 1995).

The design of the PBss vector contrasts with another recently described gene trap vector, pGT1 (LUKACSOVICH *et al.* 2001). This construct is similar to ours in its ability to splice to upstream exons. However, it reports expression through a fusion protein with the yeast transcriptional activator, Gal4, whereas ours reports expression via EGFP. The choice of EGFP allows PBss insertions to be identified in live animals without any UAS-reporter construct. In addition, we have shown that some insertions generate fusion proteins that are localized to distinct cellular compartments. This feature of

the PBss system can potentially provide information about the endogenous protein's subcellular localization. This observation also has implications for Gal4-based vectors. Because Gal4 must be nuclear to be functional, some Gal4 fusion proteins may not be active if they are localized to nonnuclear compartments.

PBss limitations and improvements: On the basis of these results we suggest that, in its current state, the PBss element is a valuable functional genomics tool. However, our analysis also suggests some limitations and potential improvements that can be implemented in the future. One potential drawback of the current element is the apparently low frequency of transposition events measured in the remobilization experiments. In the three miniscreens described here we obtained a productive transposition frequency that ranged from ~ 0.3 to ~ 1 event per 1000 animals examined. This frequency in part reflects the fact that productive (*i.e.*, EGFP-expressing) insertions are inherently low-frequency events. For a productive insertion to occur, the PBss element must insert in the correct orientation downstream of either a noncoding exon (and thus utilize the in frame ATG 5' to EGFP) or a coding exon whose frame matches the EGFP frame (in this case, +1). Other nonproductive insertions would have been undetected in our screens. We emphasize, however, that despite the low frequency, identifying EGFP-positive embryos or larvae is very easy. This is especially true for the system described here because the starting elements (C4-PBss) do not express EGFP. Thus, following mobilization, any EGFP-expressing animal stands out and is easy to identify. Given this absence of background, the process of identifying productive insertions can also be automated using whole-embryo sorters, thus allowing the quick isolation of thousands of EGFP-expressing lines (FURLONG *et al.* 2001).

Our observations also suggest some improvements that can be introduced to the PBss system. We were surprised to observe a cryptic splice donor site in the 3' piggyBac sequences present in PBss. In the example

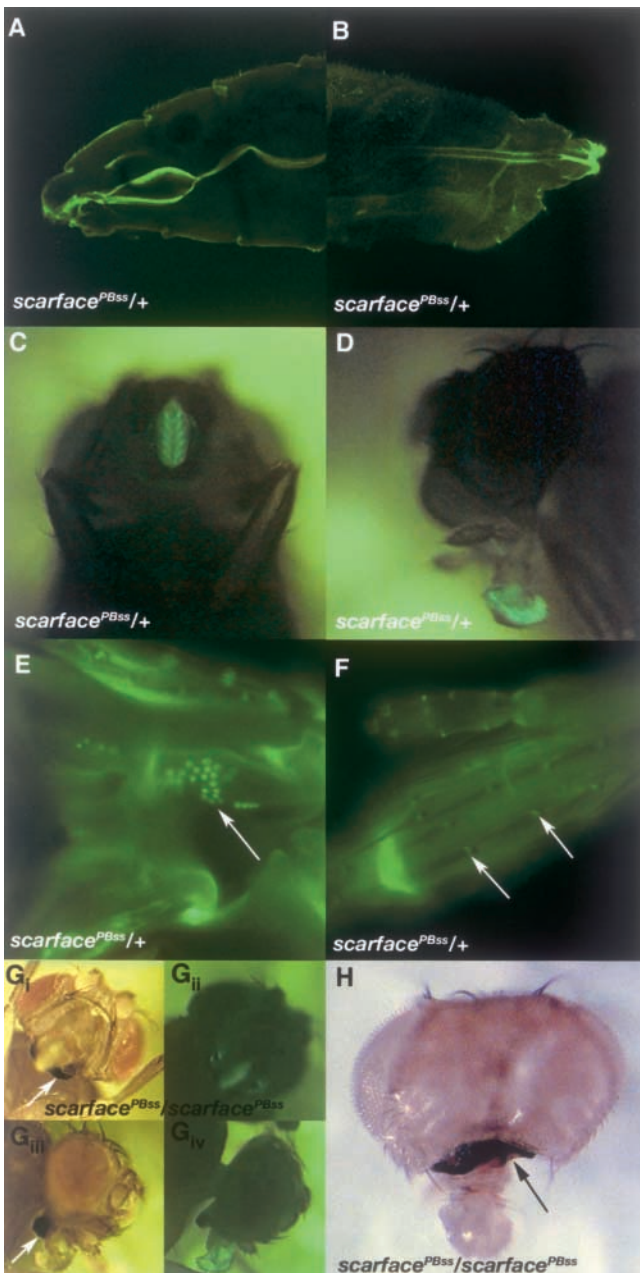


FIGURE 6.—EGFP expression and mutant phenotype of larval and adult *scarface*^{PBss} flies. (A) Anterior side view of a heterozygote *scarface*^{PBss}/*+* larva. (B) Posterior side view of a heterozygote *scarface*^{PBss}/*+* larva. (C) Ventral view of a heterozygote *scarface*^{PBss}/*+* fly expressing EGFP in the adult proboscis. (D) Side view of a heterozygote *scarface*^{PBss} fly. (E) Close-up of an adult wing hinge showing expression in patches of specific cells in the hinge. (F) Close-up of an adult *scarface*^{PBss} leg indicating expression in the socket cells of the macrochaete. (G and H) Brightfield and EGFP expression images of an adult homozygote *scarface*^{PBss}/*scarface*^{PBss} fly exhibiting black, visibly necrotic tissue near the proboscis (arrow). (G_i and G_{iii}) Brightfield. (G_{ii} and G_{iv}) EGFP expression of the fly in G_i and G_{iii}, respectively. (H) Brightfield ventral view close-up of *scarface*^{PBss} head with arrow indicating necrotic tissue close to proboscis (arrow).

analyzed here, the utilization of this splice site may have been required for EGFP to be translated in frame. However, there may be some cases when the utilization of this splice site could interfere with EGFP expression. To prevent this splice from occurring, it may be possible to mutate the essential "GT" dinucleotide at this cryptic splice donor site.

It is likely that the timing and/or length of heat shocks to induce expression of the piggyBac transposase can also be improved over the initial conditions tried here. It is possible that the PBss insertion we used as the starting point for two of the miniscreens (*scarface*^{PBss}) may, for some reason, be less able to excise than other insertions. A survey of additional PBss insertions may reveal a more prolific starting point for initiating new transposition events. The PBss element used here has the *EGFP* sequences in the +1 translational frame (relative to the engineered splice acceptor site). Thus, another potential improvement would be to simultaneously use three PBss elements that have the *EGFP* sequences in each of the three translational frames. Finally, it may be possible to include an *IRES-Gal4* cassette downstream of *EGFP*. This would allow both the tagging of a gene and the subsequent ability to mis-express other gene products in the same pattern under Gal4 control. In preliminary experiments, however, we have been unable to find an IRES that functions in an unbiased manner in *Drosophila* embryos and imaginal discs (data not shown). Thus, additional surveys of IRES sequences must be carried out before this feature can be added to the PBss system.

We emphasize, however, that the PBss system as it stands is able to both tag and mutate endogenous *Drosophila* genes as designed. Thus, PBss is a functional genomics tool that is ready to be used by the *Drosophila* community.

We thank William Cigich for testing some of the PBss lines and IRES constructs, Rokhaya Cisse for injecting the C4-PBss and PW8-PBss constructs, Myriam Zecca for the flu-*EGFP* DNA, and R. A. Harrel for piggyBac DNAs. This work was supported by National Institutes of Health (NIH) grants to R.S.M. and an NIH National Research Service Award to C.P.B.

LITERATURE CITED

- ALTSCHUL, S. F., W. GISH, W. MILLER, E. W. MYERS and D. J. LIPMAN, 1990 Basic local alignment search tool. *J. Mol. Biol.* **215**: 403–410.
- 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.
- BIER, E., H. VAESSIN, 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.
- BLACKMAN, R. K., M. M. KOEHLER, R. GRIMAILA and W. M. GELBART, 1989 Identification of a fully-functional hobo transposable element and its use for germ-line transformation of *Drosophila*. *EMBO J.* **8**: 211–217.
- BOND, B. J., and N. DAVIDSON, 1986 The *Drosophila melanogaster* actin 5C gene uses two transcription initiation sites and three polyadenylation sites to express multiple mRNA species. *Mol. Cell. Biol.* **6**: 2080–2088.
- 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.
- CARY, L. C., M. GOEBEL, H. H. CORSARO, H. H. WANG, E. ROSEN *et al.*, 1989 Transposon mutagenesis of baculoviruses: analysis of *Trichoplusia ni* transposon IFP2 insertions within the FP-locus of nuclear polyhedrosis viruses. *Virology* **172**: 156–169.
- 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.
- CHUNG, Y.-T., and E. B. KELLER, 1990 Regulatory elements mediating transcription from the *Drosophila melanogaster* actin 5C proximal promoter. *Mol. Cell. Biol.* **10**: 206–216.
- ELICK, T. A., C. A. BAUSER, N. M. PRINCIPE and M. J. FRASER, JR., 1996 PCR analysis of insertion site specificity, transcription, and structural uniformity of the Lepidopteran transposable element IFP2 in the TN-368 cell genome. *Genetica* **97**: 127–139.
- FURLONG, E. E., D. PROFITT and M. P. SCOTT, 2001 Automated sorting of live transgenic embryos. *Nat. Biotechnol.* **19**: 153–156.
- GOSSLER, A., A. L. JOYNER, J. ROSSANT and W. C. SKARNES, 1989 Mouse embryonic stem cells and reporter constructs to detect developmentally regulated genes. *Science* **244**: 463–465.
- HANDLER, A. M., and R. A. HARRELL, 1999 Germline transformation of *Drosophila melanogaster* with the piggyBac transposon vector. *Insect Mol. Biol.* **8**: 449–457.
- JACKSON, S. M., and C. A. BERG, 2002 An A-kinase anchoring protein is required for protein kinase A regulatory subunit localization and morphology of actin structures during oogenesis in *Drosophila*. *Dev. Biol.* **129**: 4423–4433.
- LEIGHTON, P. A., K. J. MITCHELL, L. V. GOODRICH, X. LU, K. PINSON *et al.*, 2001 Defining brain wiring patterns and mechanisms through gene trapping in mice. *Nature* **410**: 174–179.
- LI, X., J. C. HEINRICH and M. J. SCOTT, 2001 piggyBac-mediated transposition in *Drosophila melanogaster*: an evaluation of the use of constitutive promoters to control transposase gene expression. *Insect Mol. Biol.* **10**: 447–455.
- LI, Z., E. A. ROSSI, J. D. HOHEISEL, D. KALDERON and C. S. RUBIN, 1999 Generation of a novel A kinase anchor protein and a myristoylated alanine-rich C kinase substrate-like analog from a single gene. *J. Biol. Chem.* **274**: 27191–27200.
- LOBO, N., X. LI and M. J. FRASER, JR., 1999 Transposition of the piggyBac element in embryos of *Drosophila melanogaster*, *Aedes aegypti* and *Trichoplusia ni*. *Mol. Gen. Genet.* **261**: 803–810.
- LONG, M., and M. DEUTSCH, 1999 Association of intron phases with conservation at splice site sequences and evolution of spliceosomal introns. *Mol. Biol. Evol.* **16**: 1528–1534.
- LONG, M., M. DEUTSCH, W. WANG, E. BETRAN, F. G. BRUNET *et al.*, 2003 Origin of new genes: evidence from experimental and computational analyses. *Genetica* **118**: 171–182.
- LUKACSOVICH, T., Z. ASZTALOS, W. AWANO, K. BABA, S. KONDO *et al.*, 2001 Dual-tagging gene trap of novel genes in *Drosophila melanogaster*. *Genetics* **157**: 727–742.
- MORIN, X., R. DANEMAN, M. ZAVORTINK and W. CHIA, 2001 A protein trap strategy to detect GFP-tagged proteins expressed from their endogenous loci in *Drosophila*. *Proc. Natl. Acad. Sci. USA* **98**: 15050–15055.
- 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.
- NIWA, H., K. ARAKI, S. KIMURA, S. TANIGUCHI, S. WAKASUGI *et al.*, 1993 An efficient gene-trap method using poly A trap vectors and characterization of gene-trap events. *J. Biochem.* **113**: 343–349.
- O'BROCHTA, D. A., and P. W. ATKINSON, 1996 Transposable elements and gene transformation in non-drosophilid insects. *Insect Biochem. Mol. Biol.* **26**: 739–753.
- 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.
- ROSSI, E. A., Z. LI, H. FENG and C. S. RUBIN, 1999 Characterization of the targeting, binding, and phosphorylation site domains of

- an A kinase anchor protein and a myristoylated alanine-rich C kinase substrate-like analog that are encoded by a single gene. *J. Biol. Chem.* **274**: 27201–27210.
- SAMBROOK, J., E. F. FRISTCH and T. MANIATIS, 1989 *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- SKARNES, W. C., B. A. AUERBACH and A. L. JOYNER, 1992 A gene trap approach in mouse embryonic stem cells: the lacZ reported is activated by splicing, reflects endogenous gene expression, and is mutagenic in mice. *Genes Dev.* **6**: 903–918.
- SKARNES, W. C., J. E. MOSS, S. M. HURTLEY and R. S. BEDDINGTON, 1995 Capturing genes encoding membrane and secreted proteins important for mouse development. *Proc. Natl. Acad. Sci. USA* **92**: 6592–6596.
- 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.
- 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.
- 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.
- ZECCA, M., and G. STRUHL, 2002 Subdivision of the *Drosophila* wing imaginal disc by EGFR-mediated signaling. *Development* **129**: 1357–1368.

Communicating editor: A. J. LOPEZ

