

# A *P*-Element Insertion Screen Identified Mutations in 455 Novel Essential Genes in *Drosophila*

Su-Wan Oh, Tracy Kingsley, Hyun-hee Shin, Zhiyu Zheng, Hua-Wei Chen, Xiu Chen, Hong Wang, Peizheng Ruan, Michelle Moody and Steven X. Hou<sup>1</sup>

*The Laboratory of Immunobiology, National Institutes of Health, National Cancer Institute, Frederick, Maryland 21702-1201*

Manuscript received July 24, 2002

Accepted for publication October 15, 2002

## ABSTRACT

With the completion of the nucleotide sequences of several complex eukaryotic genomes, tens of thousands of genes have been predicted. However, this information has to be correlated with the functions of those genes to enhance our understanding of biology and to improve human health care. The *Drosophila* transposon *P*-element-induced mutations are very useful for directly connecting gene products to their biological function. We designed an efficient transposon *P*-element-mediated gene disruption procedure and performed genetic screening for single *P*-element insertion mutations, enabling us to recover 2500 lethal mutations. Among these, 2355 are second chromosome mutations. Sequences flanking >2300 insertions that identify 850 different genes or ESTs (783 genes on the second chromosome and 67 genes on the third chromosome) have been determined. Among these, 455 correspond to genes for which no lethal mutation has yet been reported. The *Drosophila* genome is thought to contain ~3600 vital genes; 1400 are localized on the second chromosome. Our mutation collection represents ~56% of the second chromosome vital genes and ~24% of the total vital *Drosophila* genes.

THE nucleotide sequences of several complex eukaryotic genomes, including those of *Caenorhabditis elegans*, *Drosophila melanogaster*, *Arabidopsis thaliana*, and *Homo sapiens*, have been completed (*C. ELEGANS SEQUENCING CONSORTIUM* 1998; ADAMS *et al.* 2000; ARABIDOPSIS GENOME INITIATIVE 2000; INTERNATIONAL HUMAN GENOME SEQUENCING CONSORTIUM 2001). From genomic sequences of human and model organisms, cDNA, and expressed sequence tags (ESTs), tens of thousands of genes have been identified. Novel informatic tools have been developed to predict transcription units, functional domains, and protein structures. It is now possible to efficiently analyze gene sequence and structure on a genome-wide scale. However, this information has to be correlated with gene function to enhance our understanding of biology and to improve human health care. The most efficient way of identifying gene function is to generate mutations. Genome-wide collections of gene knockouts would provide a vital resource for gene-based approaches to biological research. It is now technically possible, in theory, to mutate virtually any gene that has been molecularly identified in the major multi-cellular model organisms. In practice, obtaining mutants remains a time-consuming and challenging task, and whole-genome gene disruption has been slow to develop, which in turn is the major hindrance to progress in understanding gene function *in*

*vivo*. As a result, linking genes with mutations has become the primary focus in functional genomics and modern biology research.

The combination of powerful genetic manipulations, excellent cytology, and sophisticated molecular biology tools available in *Drosophila* make it a powerful model in the study of conserved gene function. *Drosophila* and humans share many genes whose sequences and functions have been conserved. More importantly, many biological processes between humans and *Drosophila* are remarkably similar. Humans and *Drosophila* share pathways for intercellular signaling, developmental patterning, learning, and behavior, as well as for tumor formation and metastasis (SPRADLING *et al.* 1995). Therefore, it is possible to define human gene function by utilizing *Drosophila*. This can be accomplished by constructing a *Drosophila* mutation library, correlating mutations with genes, and assigning function on the basis of homologs to novel genes in humans and other organisms.

Several large-scale genetic screens have been conducted in *Drosophila* to identify the maternal and zygotic gene products involved in specific events in pattern formation of the embryo. Maternal functions have been identified via screens for female sterility, while zygotic genes have been detected via screens for embryonic lethal mutations (GANS *et al.* 1975; MOHLER 1977; NUSSLEIN-VOLHARD and WIESCHAUS 1980; JURGENS *et al.* 1984; NUSSLEIN-VOLHARD *et al.* 1984, 1987; WIESCHAUS *et al.* 1984; SCHÜPBACH and WIESCHAUS 1986, 1989). These screens have identified ~300 genes with func-

<sup>1</sup>Corresponding author: The Laboratory of Immunobiology, National Institutes of Health, National Cancer Institute, Frederick, MD 21702-1201. E-mail: shou@mail.ncifcrf.gov

tions in oogenesis and 140 patterning genes that are instrumental in controlling specific embryonic decisions. This is a small number of genes, considering that the *Drosophila* genome has been estimated to potentially code for ~13,600 different genes (ADAMS *et al.* 2000).

The above screens have two major drawbacks. First, mutations in these screens were generated by chemically induced mutagenesis. Identification of the open reading frame responsible for a mutant phenotype has varied widely and has frequently been a time-consuming and challenging task. Second, the assumption underlying these screens is that the expression of genes that encode "decisionmaking" functions is tightly restricted to the corresponding developmental stage. Indeed, some of the maternal gene functions could be missed if the gene products were used at multiple times during the development of the animal. Similarly, some zygotic gene functions important for embryonic patterning could be missed if the gene were also expressed maternally, because the maternal product can mask the zygotic requirement. Many genes involved in critical patterning events have not been identified because of their developmental pleiotropy; a screen to analyze the maternal effects of X-linked zygotic lethal mutations has been conducted (PERRIMON *et al.* 1984, 1989). From this analysis, it has been estimated that gene activity of 75% of the essential loci is required for the formation of either a normal egg or a wild-type larva. This represents a significant fraction of the genome because in *Drosophila* it is estimated that 3600 loci are mutable to a visible phenotype and that 95% of these are essential for viability (see PERRIMON *et al.* 1989). From the X-linked studies, a number of zygotic lethal mutations associated with specific maternal-effect phenotypes have been identified. The identification of zygotic lethal mutations with specific maternal-effect phenotypes has been conducted systematically only on the X chromosome (PERRIMON *et al.* 1984, 1989), which represents one-fifth of the *Drosophila* genome.

The *P*-element insertional mutagenesis has many advantages over traditional mutagenesis procedures: It is usually possible to identify rapidly the transcription unit that an insertion has disrupted by sequencing the flanking sequences from one or both ends of the *P*-element insertion; remobilizing the inserted element can generate new alleles, and expression patterns can be characterized by *lacZ* staining of tissues. However, only a small number of existing *P*-element mutations are suitable for analyzing the maternal effects of the zygotic lethal mutations in a germline clonal (GLC) assay (PERRIMON *et al.* 1996). Most of the second chromosome *P*-element mutations (TOROK *et al.* 1993) are not good for GLC due to contamination, which occurred in the initial screen (our unpublished observation). Therefore, we designed an efficient transposon *P*-element-mediated gene disruption procedure and did a genetic screen for a single

*P*-element insertion mutation on the second chromosome of *D. melanogaster*. We recovered 2500 lethal mutations; 2355 are on the second chromosome. Our pilot experiments suggested that the majority of our *P* mutants (>90%) are suitable for GLC assay.

## MATERIALS AND METHODS

**Strains:** A *y w P-lacW* stock (BIER *et al.* 1989) was kindly provided by Yuh Nung Jan. *Sb P(ry<sup>+</sup> Δ2-3)*, referred to as *Sb Δ2-3*, is the standard transposase source (ROBERTSON *et al.* 1988). Both strains were isogenized for the second chromosome before starting the experiment. A *w; HS-hid.Sp/Cyo* stock was used as a dominant temperature-sensitive lethal mutation and was kindly provided by Ruth Lehmann. All flies were maintained and mated on standard yeast-cornmeal-agar medium and all experiments were performed at 25°.

***P*-element mutagenesis:** The basic genetic design involves crossing (*en masse*) males carrying transposase activity on the third chromosome, provided by the *Δ2-3 P*-element, which is inserted at 99B and apparently immobilized, to *white* females that were homozygous for a *P-lacW* insertion on the X chromosome (and therefore had pigmented eyes). Males that carried *P-lacW* on the X chromosome (maternally derived) and transposase on the third were recovered from this cross. The *Δ2-3* transposase-producing chromosome was marked with the dominant *Stubble* bristle marker but could also be scored easily by the *Δ2-3* transposase activity directly. Because the absence of the last intron in the *Δ2-3 P*-element causes expression of transposase in all tissues, rather than exclusively in the germline, flies that carry both *P-lacW* and *Δ2-3* have eyes with patches of color mixed with white areas, a phenomenon known as somatic variegation.

To carry out the experiment, we set crosses in several vials. In each cross we mated 20 female virgins that were homozygous for a *P-lacW* insertion X chromosome (line A) with 10 *yw; Δ2-3, Sb/TM3, Ser* males (Figure 1). After 2 days, we transferred the flies from vials to bottles. Flies were transferred daily for 5 days to assemble enough bottles with a large number of males that carry *P-lacW* on the X chromosome and transposase on the third. In the next generation, we collected males that carry *P-lacW* on the X chromosome and transposase on the third; flies that carry both *P-lacW* and *Δ2-3, Sb* have orange or red eyes (with patches of color mixed with white areas) and the dominant *Stubble* bristle phenotype. We set up 3000 vials per person [one cycle: 300 vials per day, 300 × 10 days (2 weeks or 10 work days)], each with two *P-lacW; Δ2-3, Sb* males crossed to three *w/w; Sco/Cyo* females. Male progeny from such a cross will inherit the *w* X chromosome from their mothers and will have white eyes unless the *P-lacW* in the father has jumped to an autosome, in which case the *P-lacW* element could segregate into both male and female offspring, resulting in pigmented eyes. One *P-lacW* male (*Cyo* without *Δ2-3, Sb*; the range of eye colors as a result of position effect in these transformed flies may vary from very light yellow to essentially wild-type red, with the bulk of insertions yielding an intermediate orange shade) were crossed to three *w/w; HS-hid SP/Cyo* females. The flies were cultured at room temperature; we discharged the adult flies after 4 days and performed a 60-min incubation in a 37° water bath twice during the early larval stages (days 4 and 5 after cross). Induction of ectopic *hid* expression in this manner presumably causes massive cell death and results in embryonic/larval lethality. Progeny of these flies were checked. If there were white eye flies in the vial, it was a third chromosome insertion. If all flies were orange eye and *Cyo*, we transferred them to a fresh vial to

produce the next generation. Progeny of the above flies then were examined. If *P-lacW* insertions were on the second chromosome and homozygous viable, the observer saw two kinds of flies: *P-lacW/P-lacW* and *P-lacW/Cyo*. Straight-wing flies (*i.e.*, homozygous for *P-lacW*) have darker eyes (usually dull red) than those of *Cyo* flies (*i.e.*, heterozygous for *P-lacW*). If *P-lacW* insertions are on the second chromosome and homozygously lethal, the observer will see one kind of fly: *P-lacW/Cyo* (one eye color and *Cyow*ing). We saved the lethal lines as permanent stocks.

**Germline clonal analysis:** Females carrying GLCs of *P*-lethal mutants were generated using the "FLP-DFS" technique (Hou *et al.* 1995).

**Genomic DNA preparation:** The fresh flies were grown on standard fly food (ASHBURNER 1990) at room temperature until they were collected. About 20–50 anesthetized flies per sample were collected into a 1.5-ml Eppendorf tube and stored in the deep freezer ( $-80^{\circ}$ ) for 2–3 hr until they were completely frozen. The tubes were merged into liquid nitrogen and ground by disposable pestle. The lysis buffer with Protease K was added immediately and incubated overnight at  $55^{\circ}$ . We then added precipitation buffer AL/E (QIAGEN, Chatsworth, CA) with 20  $\mu$ l of RNase A (20 mg/ml) to each sample. The supernatant was transferred into a Dneasy 96-well plate (QIAGEN) and centrifuged. The plate was washed twice by washing buffer and dried out at  $70^{\circ}$  for 15 min. To elute the DNA onto the 96-well plate, 100  $\mu$ l of elution buffer (QIAGEN) was added and centrifuged. The entire next step was performed in the 96-well format plate.

**Digestions, ligations, and PCR:** A total of 20  $\mu$ l of genomic DNA ( $\sim 10$   $\mu$ g/ $\mu$ l) were taken and digested with *Sau3A* I or *HinPI* at  $37^{\circ}$  for 3 hr to overnight separately; then 30  $\mu$ l of ligation mixture was added to the digested solution. The ligation mixture was incubated overnight at  $4^{\circ}$ . Ten microliters of ligates was used as a PCR template for both 5' end and 3' end PCR. PCR was performed on Gene Amp PCR system 9600 and 9700 using the following parameters: 1 time at  $95^{\circ}/5$  min; 35 times at  $95^{\circ}/30$  sec,  $60^{\circ}/2$  min,  $70^{\circ}/2$  min; and 1 time at  $72^{\circ}/10$  min.

**Sequencing and database search:** The PCR products were purified with an enzymatic clean-up method using *SAP/ExoI* and a 96-well PCR purification kit (QIAGEN). Two microliters of purified PCR product was used as a sequencing template. The sequencing reaction was performed using a Gene Amp PCR system 9700 with an ABI Big dye mixture. The sequencing reaction was applied onto an ABI Prism gene analyzer and ABI 3700 automatic sequencer. Vectors were cut from the raw sequence data and the vector-pilled sequences were blasted; the Berkeley *Drosophila* Genome Project (BDGP) database (<http://www.fruitfly.org/>) was used to identify the probable disrupted gene.

**Stock distribution:** At present, lines from this collection are maintained at Dr. Steven Hou's laboratory and are available (limited to five lines) upon request to shou@mail.ncicfcrf.gov. The entire collection will be sent to the Bloomington Stock Center later. Return information from further study of any line is welcome.

## RESULTS

**Isolation of *P*-insertion mutants:** A large-scale single *P*-element mutagenesis was performed to saturate the second chromosome with recessive lethal insertion using the high transposition frequency of the *P-lacW* construct. The genetic scheme used to mobilize *P-lacW* to

autosomal sites from the X chromosome is shown in Figure 1.

Approximately 50,000 individual vials were set up, each with two *P-lacW*,  $\Delta 2-3$  males crossed to three *white* females carrying the second chromosome balancer *Cyo*. Male progeny from such a cross will inherit the *white* X chromosome from their mothers and will have white eyes unless the *P-lacW* in the father has jumped to an autosome, in which case, the *P-lacW* element could segregate into both male and female offspring, resulting in pigmented eyes. In 43,000 vials (86%), at least one *w*<sup>+</sup> male was recovered (*i.e.*, a *P-lacW* transposition to an autosome). The range of eye colors as a result of position effect in these transformed flies varied from very light yellow to essentially wild-type red, with the bulk of insertions yielding an intermediate orange shade. One *w*<sup>+</sup> F<sub>1</sub> male from each vial was used for crosses to generate balanced mutant stocks. Of the 17,635 *P* insertions in the second chromosome, 2962 ( $\sim 16.8\%$ ) proved to be lethal or semilethal. To reduce the multiple *P*-insertion lines, we discharged 142 lines that have near wild-type red eyes. We further eliminated 465 semilethal lines. A total of 2355 independently derived second chromosome strains was retained for further characterization. We also analyzed 160 third chromosome lethal strains.

**Characterizing insertions using flanking DNA sequences:** The genomic DNA sequence flanking the insertion sites in the collection lines was needed to associate lines with specific genes. We attempted to recover genomic DNA adjacent to the 5', 3', or both sides of the *P*-element from all collection lines. We used an inverse PCR method that was carried out in a 96-well format (SPRADLING *et al.* 1999). Beginning at the insertion site of all recovered flanks, a single sequencing run was carried out (see MATERIALS AND METHODS). After trying different enzyme digestion and ligation, we were able to obtain usable sequence from 90% of the *P* lines.

**Associating primary collection lines with genes:** Our final collection contains 854 independent strains. To identify as many genes as possible, the *P*-element flank sequences were used to blast the completed *Drosophila* sequences in the BDGP database. Mutation-causing *P* elements are known to preferentially cluster in the 5' region of the affected genes (see SPRADLING *et al.* 1995); we took this consideration into account in linking our *P* element to a gene. In 50% of the lines, the *P* element was located close to the 5' region of a predicted gene, allowing a direct link to a disrupted gene. In 30% of the lines, the *P* element was located within 1 kb of the 5' region of a predicted gene; in 8% of the lines, the *P* element was located within 3 kb of the 5' region of a predicted gene. In these less direct cases, we carefully analyzed the surrounding genomic DNA structures (such as transcriptional control elements, translation initiation consensus sequence, and intron-exon boundary) to figure out the *P*-element disrupted genes. If a *P*



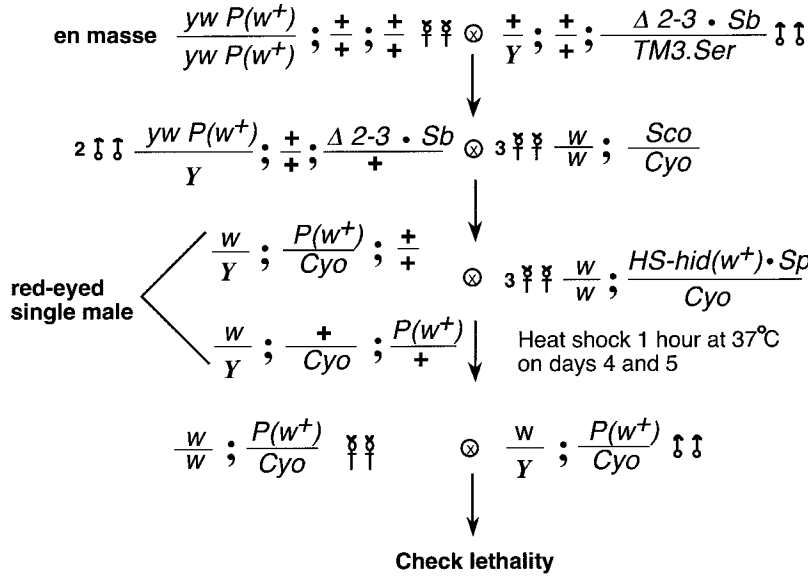


FIGURE 1.—Genetic scheme for mobilization of *P-lacW*. The details of the scheme are provided in MATERIALS AND METHODS.

element was inserted between two CGs, we chose the CG downstream of the *P* insertion as the disrupted gene. Altogether, we identified genes for 93% (796/854) among the 854 sequenced *P* strains (Table S1 at <http://www.genetics.org/supplemental/> and Table 1). A total of 1% (12/854) of the *P* strains were linked to an EST (Table S1 at <http://www.genetics.org/supplemental/> and Table 1). Four of the remaining lines were linked to a transposon element (Table S1 at <http://www.genetics.org/supplemental/> and Table 1). We could not find a gene or an EST for the remaining 42 lines.

TABLE 1  
Probable function and numbers of *P*-element lines from this screen

Probable function	No. of <i>P</i> -element line
Apoptosis (A)	12
Cell adhesion (CA)	12
Cell cycle	8
Cytoskeleton (CS)	24
Enzyme (E)	114
Replication, DNA repair, and other chromosomal proteins (CP)	31
Novel (N)	274
Transcription and gene regulation (TGR)	61
Translation and RNA processing (TRP)	64
Signal transduction (ST)	91
Channel and transporter (TP)	28
Other cellular process (O)	77
EST	12
No gene in 3 kb	42
Retrotransposon and others	4
Total numbers	854

The assignment of a probable function class to each gene is based on FlyBase information, biochemical functions, or known structural homologs in other eukaryotes.

The collection provides an opportunity to link 850 *Drosophila* genes or ESTs with a genetic phenotype. Among them, 783 genes are on the second chromosome and 67 genes are on the third chromosome. Lethal mutations of 455 genes have not yet been reported, according to the information from FlyBase.

***P*-element selectivity:** This screen identified most genes that are hotspots for *P*-element insertion on the autosomes (Table S1 at <http://www.genetics.org/supplemental/> and Table 2). However, in comparison with the previously published BDGP primary collection (SPRADLING *et al.* 1999), our collection has several unique features. For example, we isolated 45 alleles of the gene *for*; the BDGP collection has only 5 alleles of *for*; for the *S* gene, the BDGP has 10 alleles and we isolated 2 alleles. Some genes are hotspots in the BDGP collection and are hit only one or two times in our collection; some genes have many alleles in our collection and have one or two alleles in the BDGP collection. On the one hand, these data confirm that the hotspots are easily mutated by the *P*-element insertion; on the other hand, they also imply that the *P*-element selectivity is somehow dependent on the specific variable screen conditions (such as temperature, humidity, fly age, etc.).

We also considered whether strong preferences exist for insertion within certain classes of genes among all those disrupted in our collection. In the BDGP collection (SPRADLING *et al.* 1999), the genes involved in signal transduction were usually well represented (the collection mutates ~50% of all autosomal genes known to be involved in the EGFR, *dpp*, *ras*, *wg*, *hh*, or *N* signaling pathways). The posterior group genes are also well represented (46% of autosomal posterior group genes are disrupted), but only 14% of the ribosomal protein genes were disrupted. In our collection, all classes of genes are generally evenly represented (Table S1 at <http://www.genetics.org/supplemental/> and Table 1).

In our collection, only four *P* elements are not associ-

TABLE 2  
Comparison of hotspot alleles from this screen and the BDGP collection

Gene	Site	Probable function	Alleles in this study	Alleles in SPRADLING <i>et al.</i> (1990)
<i>kis</i>	21 B3-4	Myb DNA-binding domain	22	27
<i>S</i>	21 E2	Component of the plasma membrane	2	10
<i>for</i>	24 A2-3	cGMP-dependent protein kinase	45	5
<i>vkq</i>	25 C1	Collagen type IV	3	13
<i>Cg25C</i>	25 C1-2	Collagen type IV	16	6
<i>eIF-4a</i>	26 A9-B1	Translation initiation factor	25	11
<i>Chi</i>	26 A9-B1	Actin binding	4	13
<i>mts</i>	28 D1-3	Protein serine/threonine phosphatase	12	4
<i>Pka-C1</i>	30 C5-7	Protein serine/threonine kinase	6	13
<i>sop</i>	30 E1-7	Structural protein of ribosome	12	3
<i>bun</i>	33 E1-7	Transcription factor	17	20
<i>esg</i>	35 C5-D1	Transcription factor	37	14
<i>crp</i>	35 F6-7	Transcription factor	20	11
<i>neb</i>	38 B4-6	Chromatin binding	13	5
<i>EcR</i>	42 A1-3	Transcription factor	11	3
<i>Uba1</i>	46 A1-3	Ubiquitin-activating enzyme	16	9
<i>lola</i>	47 A9-13	Transcription factor	47	24
<i>Shn</i>	47 D5-7	Transcription factor	19	7
<i>mam</i>	50 C23-D3	Mesoderm determination	5	11
<i>apt</i>	59 F1	Transcription factor	14	3
<i>tsr</i>	60 B1-2	Actin depolymerization	10	9

For each strain the table gives the gene name, cytogenetic location of the gene (Site), the probable function, the number of alleles from this screen, and the number of alleles in the BDGP collection.

ated with a protein-coding gene. In the BDGP collection, a number of *P* elements are inserted into locations that are not associated with protein-coding genes. It remains unclear whether these differences are due to the BDGP search using an incomplete *Drosophila* sequence database or whether our search has pushed to link each *P* element with a gene.

**Association of newly induced recessive lethality with the new *P* insertion:** To ensure that the newly induced recessive lethality is directly attributable to the new *P* insertion, we performed the following experiments. First, the *P* element in line *l(2)2276*, which inserted into a gene (CG8902) for which no mutation had previously been described, was remobilized by the  $\Delta 2-3$  transposase. Among 200 white-eye *P*jump-out lines, 158 lines were homozygously viable, indicating that no second lethality exists in the *l(2)2276* line. Second, we analyzed seven new alleles of *kis* in a germline clonal assay; all of them developed the *kis*-type segmentation defects (data not shown), suggesting that the *P*'s indeed disrupted only the *kis* gene. Finally, we isolated nine new *zip* alleles; all of them developed the "dorsal open" phenotype as expected, implying that no second lethality exists in these lines. In another experiment, we recombined 700 new *P* lines onto FRT chromosomes; 20 lines (~3%) lost lethality after recombination, suggesting that the lethality in these 20 lines is not associated with the *P* elements. From these data we conclude that most lethal mutations (>90%) are attributed to the *P* insertions.

**Suitability of the *P* elements for GLC analysis:** We

tested 133 of our *P*-element mutants for performing germline clonal analysis, which also served as means to verify the quality of the *P*-element mutants, because our previous experiences taught us that only one mutant on each chromosome arm is suitable for performing a GLC assay. Most of the Kiss lines (TOROK *et al.* 1993) cannot be used for a GLC assay (S. HOU, unpublished observation).

We recombined the 133 *P* lines onto the FRT chromosome and performed a GLC assay (see Table 3). Among the 133 *P* lines, 18 (14%) did not lay eggs, 67 (50%) laid eggs and the embryos hatched, and 48 (36%) laid abnormal eggs or had embryos that died before hatching and showed cuticles with patterning defects. This result is better than the one obtained in Perrimon's

TABLE 3  
Results from 133 *P* lines in a germline clone test

Total	No egg	Egg hatch	Abnormal eggs and patterning defects
133	18 (14%)	67 (50%)	48 (36%)

The results were classified into three groups. Group 1: No eggs ("No egg") were laid and no obviously developed eggs could be found following brief inspection of the adult abdomen. Group 2: Most eggs hatched ("Egg hatch") and larvae developed normal cuticle. Group 3: Abnormal eggs were laid or a substantial fraction of the eggs failed to hatch and cuticles had pattern defects ("Abnormal eggs and patterning defects").

study (PERRIMON *et al.* 1996) that used mostly Spradling and Jan's *P*-element collections (BIER *et al.* 1989; KARPEN and SPRADLING 1992). Therefore, the quality of our *P*-element collection is at least compatible with Spradling and Jan's collections.

## DISCUSSION

**Features of this screen:** This *P*-element-mediated gene disruption screen has several distinct advantages. First, screen efficiency is widely variable in the generation of single *P*-element-induced mutations (SPRADLING *et al.* 1999). This variation can be attributed to the overall rate of *P* transposition and other unidentified factors in the genetic backgrounds used for *P*-element mutagenesis. Practically, high screen efficiency can be achieved by strict selection of *P*-element transposition to eliminate multi-transposition events at the early stages of screening as described in BIER *et al.* (1989). We followed the BIER *et al.* (1989) strategy in the early steps of our screen. Second, we used a *HS-hid* transgene to eliminate unwanted genotypes in the screen, which saved us from picking a large number of virgins and enabled us to screen a large number of mutant lines in a relatively short time. Finally, we took advantage of the entire available *Drosophila* genome sequence and directly sequenced our selected *P* lines. This makes polytene localization of *P* insertions and genetic complementation unnecessary, because that information can be obtained from the sequences and *P*-disrupted genes through a BLAST search of the complete genomic database. Direct sequencing also automatically eliminated multiple insertion lines. Sequence can be obtained only from single *P*-insertion lines. We retained lines only from which a flank sequence had been obtained. Our GLC assay of 133 randomly selected lines indicated that our screen efficiency is at least compatible to that of previous screens conducted in Spradling and Jan's laboratories.

**All remaining genes are within 5 kb:** This screen generated *P*-element insertion mutations for 850 autosomal vital genes. Among them, 783 are on the second chromosome and 67 are on the third chromosome. The *Drosophila* genome is thought to contain ~3600 vital genes. Among them, 1400 are localized on the second chromosome. Our mutation collection represents 56% of second chromosome vital genes and 24% of total vital *Drosophila* genes. In combination with the previously published BDGP collection (SPRADLING *et al.* 1999), the total *P*-element collection disrupted ~80% of second chromosome vital genes and ~30% of total vital *Drosophila* genes. Furthermore, we checked the remaining second chromosome genes and found that most of them are within 5 kb of a nearby *P*-element insertion. The deletion mutations for these genes can be easily generated by imprecise excision of the nearby *P* element. Theoretically, this makes it possible to generate an en-

tire chromosome's gene mutation, which will be the first time this has been done in any of the widely used models of multi-cellular eukaryotes, including *Arabidopsis*, *C. elegans*, zebrafish, and mice.

**Genes:** This screen identified all classes of genes involved in important biological processes (Table 1): 12 genes may regulate apoptosis; 12 belong to cell adhesion molecules; 8 are cell cycle regulators; 24 are cytoskeleton molecules; 114 are enzymes; 31 are involved in DNA replication, DNA repair, and other chromosomal functions; 61 are involved in transcription or gene regulation; 64 are involved in RNA processing or translation regulation; 91 are signal transduction components; 28 are channel and transporter molecules; 77 are involved in other cellular processes; and 274 are novel molecules.

The signal transduction molecules are richly represented in this collection. Our *P*-element collection disrupted 91 genes involved in signal transduction. Among them, 32 correspond to genes for which no lethal mutation had yet been reported.

**Some predicted genes may be inaccurate:** In linking a *P*-element insertion to a gene, we followed the rule that most *P* elements insert into the 5' portion of the transcription units (SPRADLING *et al.* 1995). Genes disrupted in most lines were identified by comparing the DNA sequence flanking the insertion to the published complete genomic sequence data (ADAMS *et al.* 2000). It was recently reported that the algorithms used to predict genes from genomic databases might have missed a significant number of genes (MORIN *et al.* 2001), so the prediction of the *P*-element-disrupted genes might also have some errors. For each *P* line in Table S1 at <http://www.genetics.org/supplemental/>, a GenBank accession number of the nucleotide sequence was provided; readers are encouraged to verify our prediction through researching the genomic database. Return information on any line is welcome. Final confirmation of this prediction has to depend on studies of the effects of the mutation on the expression of gene products at the levels of RNA and protein.

We are grateful to Jasmine Young, Cristina Chuang, Robin Deniker, Gregory Szeto, Katie Renn, Amy Lin, Kelly Jacobs, and Jim Wang for their help in the experimental work. We thank Yuh Nung Jan and Ruth Lehmann for providing fly strains. This work is supported in part by a grant from the U.S. Army (award number: DAMD17-00-1-0356) to S.X.H.

## LITERATURE CITED

- ADAMS, M. D., S. E. CELNIKER, R. A. HOLT, C. A. EVANS, J. D. GOCAYNE *et al.*, 2000 The genome sequence of *Drosophila melanogaster*. *Science* **287**: 2185–2195.
- ARABIDOPSIS GENOME INITIATIVE, 2000 Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**: 796–815.
- ASHBURNER, M., 1990 *Drosophila: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- 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.

- C. *ELEGANS* SEQUENCING CONSORTIUM, 1998 Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* **282**: 2012–2018.
- GANS, M., C. AUDIT and M. MASSON, 1975 Isolation and characterization of sex-linked female sterile mutations in *Drosophila melanogaster*. *Genetics* **81**: 683–704.
- HOU, X. S., T.-B. CHOU, M. MELNICK and N. PERRIMON, 1995 The torso receptor tyrosine kinase can activate Raf-independent pathway. *Cell* **81**: 63–67.
- INTERNATIONAL HUMAN GENOME SEQUENCING CONSORTIUM, 2001 Initial sequencing and analysis of the human genome. *Nature* **409**: 860–921.
- JURGENS, G., E. WIESCHAUS, C. NUSSLEIN-VOLHARD and H. KLUDING, 1984 Mutations affecting the pattern of the larval cuticle in *Drosophila melanogaster*. 2. Zygotic loci on the third chromosome. *Roux's Arch. Dev. Biol.* **193**: 283–295.
- KARPEN, G. H., and A. C. SPRADLING, 1992 Analysis of subtelomeric heterochromatin in the *Drosophila* minichromosome Dp1187 by single P element. *Genetics* **132**: 737–753.
- MOHLER, J. D., 1977 Developmental genetics of the *Drosophila* egg. I. Identification of 50 sex-linked cistrons with maternal effects on embryonic development. *Genetics* **85**: 259–272.
- 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.
- NUSSLEIN-VOLHARD, C., and E. WIESCHAUS, 1980 Mutations affecting segment number and polarity in *Drosophila*. *Nature* **287**: 795–801.
- NUSSLEIN-VOLHARD, C., E. WIESCHAUS and H. KLUDING, 1984 Mutations affecting the pattern of the larval cuticle in *Drosophila melanogaster*. 1. Zygotic loci on the second chromosome. *Roux's Arch. Dev. Biol.* **193**: 267–282.
- NUSSLEIN-VOLHARD, C., H. G. FROHNHOFER and R. LEHMANN, 1987 Determination of anteroposterior polarity in *Drosophila*. *Science* **238**: 1675–1681.
- PERRIMON, N., L. ENGSTROM and A. P. MAHOWALD, 1984 The effects of zygotic lethal mutations on female germ-line functions in *Drosophila*. *Dev. Biol.* **105**: 404–414.
- PERRIMON, N., L. ENGSTROM and A. P. MAHOWALD, 1989 Zygotic lethals with specific maternal effect phenotypes in *Drosophila melanogaster*. I. Loci on the X-chromosome. *Genetics* **121**: 333–352.
- PERRIMON, N., A. LANJUN, C. ARNOLD and E. NOLL, 1996 Zygotic lethal mutations with maternal effect phenotypes in *Drosophila melanogaster*. II. Loci on the second and third chromosomes identified by P-element-induced mutations. *Genetics* **144**: 1681–1692.
- ROBERTSON, H. M., C. R. PRESTON, R. W. PHILLIS, D. M. JOHNSON-SCHLITZ, W. K. BENZ *et al.*, 1988 A stable source of P-element transposase in *Drosophila melanogaster*. *Genetics* **118**: 461–470.
- SCHÜPBACH, T., and E. WIESCHAUS, 1986 Maternal effect mutations affecting the segmental pattern of *Drosophila*. *Roux's Arch. Dev. Biol.* **195**: 302–307.
- SCHÜPBACH, T., and E. WIESCHAUS, 1989 Female sterile mutations on the second chromosome of *Drosophila melanogaster*. I. Maternal effect mutations. *Genetics* **121**: 101–117.
- SPRADLING, A., C. DIANNE, M. STERN, I. KISS, J. ROOTE *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.
- SPRADLING, A. C., D. STERN, A. BEATON, E. J. REHM, T. LAVERTY *et al.*, 1999 The Berkeley *Drosophila* Genome Project gene disruption project: single P-element insertions mutating 25% of vital *Drosophila* genes. *Genetics* **153**: 135–177.
- TOROK, T., G. TICK, M. ALVARADO and I. KISS, 1993 P-lacW insertional mutagenesis on the second chromosome of *Drosophila melanogaster*: isolation of lethals with different overgrowth phenotypes. *Genetics* **135**: 71–80.
- WIESCHAUS, E., C. NUSSLEIN-VOLHARD and G. JURGENS, 1984 Mutations affecting the pattern of the larval cuticle in *Drosophila melanogaster*. 3. Zygotic loci on the X-chromosome and 4th chromosome. *Roux's Arch. Dev. Biol.* **193**: 296–307.

Communicating editor: T. SCHÜPBACH

