Characterization of the *flamenco* Region of the *Drosophila melanogaster* Genome

Valérie Robert, Nicole Prud'homme, Alexander Kim,[1] Alain Bucheton and Alain Pélisson

CGM/CNRS, 91198 Gif-sur-Yvette, France and IGH/CNRS, 34396 Montpellier, France

Manuscript received August 10, 2000
Accepted for publication March 9, 2001

ABSTRACT

The *flamenco* gene, located at 20A1–3 in the β-heterochromatin of the Drosophila *X* chromosome, is a major regulator of the *gypsy/mdg4* endogenous retrovirus. As a first step to characterize this gene, ~100 kb of genomic DNA flanking a *P*-element-induced mutation of *flamenco* was isolated. This DNA is located in a sequencing gap of the Celera Genomics project, *i.e.*, one of those parts of the genome in which the "shotgun" sequence could not be assembled, probably because it contains long stretches of repetitive DNA, especially on the proximal side of the *P* insertion point. Deficiency mapping indicated that sequences required for the normal *flamenco* function are located >130 kb proximal to the insertion site. The distal part of the cloned DNA does, nevertheless, contain several unique sequences, including at least four different transcription units. *Dip1*, the closest one to the *P*-element insertion point, might be a good candidate for a *gypsy* regulator, since it putatively encodes a nuclear protein containing two double-stranded RNA-binding domains. However, transgenes containing *dip1* genomic DNA were not able to rescue *flamenco* mutant flies. The possible nature of the missing *flamenco* sequences is discussed.

GYPSY/MDG4 is an insect endogenous retrovirus with infectious properties as shown by the integration of new proviruses into the germline chromosomes after horizontal transfer between laboratory strains (Kim *et al.* 1994; Song *et al.* 1994). As with any endogenous retrovirus, *gypsy* is usually transmitted vertically like a Mendelian gene. However, in some genetic backgrounds, vertically inherited *gypsy* proviruses can increase their copy number (Bucheton 1995). Comparison of strains with low and high copy numbers led to the discovery of a gene called *flamenco* that controls *gypsy* proviral multiplication (Prud'homme *et al.* 1995). Deficiencies of *flamenco,* as well as permissive alleles, allow the multiplication of *gypsy* whereas restrictive alleles do not. The latter are either dominant or semidominant over permissive alleles (Pélisson *et al.* 1997; Prud'homme *et al.* 1995). *Gypsy* is allowed to amplify only in the progeny of permissive females whatever the genotype of these progeny (Prud'homme *et al.* 1995). It is necessary and sufficient that somatic tissues of the mother have a permissive genotype for the functional *gypsy* proviruses present in such tissues to amplify in the progeny (Chalvet *et al.* 1999). Consistent with this observation is the fact that the regulation of *gypsy* by *flamenco* is specific to the somatic follicle cells that surround the female germline cells; the expression of a *gypsy-lacZ* reporter transgene is totally abolished in restrictive follicle cells (Pélisson *et al.* 1994). Thus, *gypsy* appears to be a retrovirus able to spread both horizontally and vertically in spite of the fact that its expression and its regulation by the *flamenco* gene occur only in the soma. The soma-to-germline transfer involved in the vertical spread of *gypsy* does not seem to be an infectious process, since a knock-out of the *env* gene, assumed to be reponsible for the infectious properties of the virus (Teysset *et al.* 1998), does not affect its multiplication (Chalvet *et al.* 1999). The respective importance of vertical and horizontal transmission for *gypsy* propagation, the mechanisms underlying both types of germline invasions by *gypsy,* and the mechanism by which *flamenco* represses *gypsy* expression are still unclear. To address these issues, we set out to clone the *flamenco* gene.

The *flamenco* gene is located in 20A1–3, between the complementation groups *extra organs* (*eo*) and *wings apart* (*wap*), a region contained in the β-heterochromatin of the *X* chromosome (Prud'homme *et al.* 1995). Unlike α-heterochromatin, β-heterochromatin is normally replicated in polytene chromosomes but, unlike euchromatin, it exhibits a poorly banded structure that forms the visible chromocenter (Heitz 1934; Yamamoto *et al.* 1990). At the base of the polytene *X* chromosome, the well-banded euchromatin is found distal to section 20 with respect to the centromere (Bridges 1938). Heterochromatin is composed of repeated sequences and is known to be a place for the accumulation of defective transposable elements (Miklos *et al.* 1988; Vaury *et al.* 1989; Miklos and Cotsell 1990; Pimpinelli *et al.* 1995). On the *X* chromosome, only ribosomal genes have been localized in α-heterochromatin whereas complementation groups that could correspond to single copy genes appear to be located within the β-heterochromatin (Perrimon *et al.* 1989). A few

Corresponding author: Alain Pélisson, Institut de Génétique Humaine, CNRS, 141 rue de la Cardonille, 34396 Montpellier Cedex 05, France. E-mail: alain.pelisson@igh.cnrs.fr

[1] *Present address:* Department of Genetics, Lomonosov Moscow State University, 119899 Moscow, Russian Federation.

of these genes such as *S6KII* (Wassarman *et al.* 1994), *fog* (Costa *et al.* 1994), *stn* (Andrews *et al.* 1996), and *su(f)*, which is the most proximal (Mitchelson *et al.* 1993), have been cloned. The *fog* gene was identified in a 65-kb-long gene-poor region; the Drosophila genome consortium reported that "the gene density (. . .) drops abruptly to two genes in 400 kb around *fog*" (Adams *et al.* 2000). The *stn* complementation group is a dicistronic gene that is found as a 12.5-kb island of unique DNA surrounded by repetitive sequences. On the distal side of *su(f)*, unique sequences are interspersed with repeated sequences, whereas all the 30 kb isolated on the proximal side of the gene (*i.e.*, toward the centromere) contain only repeated sequences (Tudor *et al.* 1996). *light*, an autosomal heterochromatic gene, has also been characterized at the molecular level. It is 17 kb long and its 3-kb RNA is expressed from single-copy DNA while its intronic and flanking regions consist of a heterogenous array of middle-repetitive DNA sequences (Devlin *et al.* 1990).

Here we report the *P*-element gene-tagging of *flamenco* and the molecular characterization of a 100-kb walk around the *P*-element insertion. This walk mostly contains middle-repetitive DNA. Its distal part also contains a few unique sequences including some transcription units. A detailed analysis of one of them (*dip1*) is presented here, although it has not yet been possible to demonstrate that this gene corresponds to *flamenco*. The possible nature of the missing *flamenco* sequences is discussed.

MATERIALS AND METHODS

Drosophila strains and fly care: Flies were maintained on standard Drosophila medium (Gans *et al.* 1975) at 25°. Genetic symbols follow Lindsley and Zimm (1992). The strains of the laboratory collection were previously described (Prud'homme *et al.* 1995; Chalvet *et al.* 1999), including the sample of $Df(1)l_x$ deficiencies, which are wap^- X-ray-induced derivatives of the *413(NP)* chromosome, restrictive for *gypsy* expression and mobilization (Pélisson *et al.* 1994). The *y; cn bw sp* restrictive and the oreRw38 permissive stocks were the strains used to sequence the Drosophila genome by the American and European sequencing projects, respectively. The restrictive Rev(R) stock is referred to as RevI in Desset *et al.* (1999).

Ovo^{D1} inactivation assay: The *ovo* gene is located on the *X* chromosome at cytological position 4E2. It is a hot spot for insertion of *gypsy* (Mevel-Ninio *et al.* 1989). Crosses of permissive females containing active *gypsy* proviruses with males carrying the ovo^{D1} dominant female sterile mutation produce fertile daughters at high frequency. This abolition of the dominant female sterile phenotype results from insertions of *gypsy* into the ovo^{D1} gene carried by the paternal chromosome. To estimate the frequency of these insertions, the daughters of the test cross are scored for zero, one, or two ovaries. In most cases fertile females have one ovary. The frequency of females with two ovaries is about the square of that of females with one ovary, suggesting that they result from two independent events. Such females are therefore counted twice to calculate the frequency of ovo^{D1} inactivation.

Collection of embryos homozygous for lethal deficiencies: Females heterozygous for the *X*-linked deficiency and a balancer chromosome were crossed with males carrying the same deficiency and a complementing *Y*-linked duplication (Prud'homme *et al.* 1995). Embryos were collected for 18 hr at 23°. After 1 more day at 23°, the larvae that had nonlethal heterozygous genotypes were manually discarded. The nonhatched embryos, most of which were homozygous lethal, were dechorionated in 3% bleach before being frozen at −80°.

Rescue experiments: Transgenic flies were recovered as colored-eye w^+ individuals after *P*-mediated transformation (Spradling 1986) into the wOR(P) permissive stock. The transgenes were then introduced into different *flamenco* contexts by two successive backcrosses with w^- females. Homozygous transgenic females were crossed with males hemizygous for a eo^-, $flam^-$, wap^- deficiency (l_{11}) and homozygous for the pgyp *gypsy-LacZ* transgene (Pélisson *et al.* 1994). Ovarian *lacZ* expression was studied in the progeny of this cross. *LacZ* staining was performed as described previously (Pélisson *et al.* 1994).

Libraries, screenings, and databases: A λ phage library was constructed with the DNA of $flam^{by+(P)}$ homozygous flies. This DNA, partially digested by *Sau*3A, was ligated with the λ Fix II vector arms (Stratagene, La Jolla, CA) following the protocol provided by the manufacturer.

The NotBamNot-CoSpeR (cosT) genomic cosmid library was kindly provided by J. Tamkun. It was constructed from the restrictive *y; cn bw sp* strain. After transfer to Optitran BA-S85 reinforced nitrocellulose filters (Schleicher & Schuell, Keene, NH), 3×10^4 clones were hybridized overnight at 42° in 5× SSC, 5× Denhardt's solution, 0.1% SDS, 50% formamide with the U2 probe, *i.e.*, the 3-kb *Bbs*I-*Sal*I fragment that flanks the $flam^{by+(P)}$ mutation, radiolabeled by random priming (Megaprime; Amersham, Arlington Heights, IL). Out of three positive clones, two were purified and their DNA was prepared with the plasmid Maxi kit (QIAGEN, Valencia, CA).

RCPI-98 high-density hybridization filters containing the Berkeley Drosophila Genome Project (BDGP) BAC library were purchased from the BACPAC Resource Center, as were 19 clones from the library.

An ovary cDNA library was kindly provided by P. Tolias. The *flamenco* status of the Canton-S strain used to make this library is unknown. A total of 9×10^5 clones were screened as described above.

The databases constructed by BDGP were screened by BLAST (Altschul *et al.* 1990). The BLAST queries were perfomed at the BDGP website with standard parameters. The cDNA clones were purchased from Research Genetics (Birmingham, AL). The whole genome sequence assembly was overviewed at the GenBank site (release 2) (http://www.ncbi.nlm.nih.gov/PMGifs/Genomes/7227).

Restriction mapping of cosmid inserts: The protocol was adapted from Rackwitz *et al.* (1984). A total of 1 μg of cosmid DNA was digested by 5 units of λ terminase (Epicentre, Madison, WI) for 30 min at 25°. After inactivation of the enzyme at 65°, 600 ng of this linearized DNA was preheated at 37° and then incubated with 0.1 unit of *Pst*I (Biolabs). Aliquots of the digestion were sequentially stopped at 2, 8, 20, 45, and 120 min by the addition of 10 mm EDTA and 125 mm NaCl. A total of 5 pmol of dephosphorylated oligonucleotides, either cosL (AGGTCGCCGCCC) or cosR (GGGCGGCGACCT), was [γ-^{32}P]ATP-labeled with polynucleotide kinase (GIBCO BRL, Gaithersburg, MD) as indicated by the manufacturer. One-half of each digested aliquot was incubated with 0.2 pmol of either labeled oligonucleotide for 10 min at 65° and then 45 min at 45°. Samples were diluted 1/3 with loading buffer (1× TAE, 60 mm EDTA, 0.1% bromophenol blue and 50% glycerol), loaded on a 0.5% agarose gel, and fractionated at 4° for 24 hr at 55V in 1× TAE buffer. Gels were vacuum dried

on DE81 paper (Whatman) and exposed to X-Omat AR film (Kodak, Rochester, NY).

Subcloning nonrepetitive sequences: U2 was obtained by subcloning a 3-kb *Sal*I-*Bbs*I fragment from a clone of the λ phage library containing the 5′ segment of the *P* element. U1 and d1 were obtained from the cos7a cosmid (Figure 3A). To obtain U1, a *Hpa*I-*Hin*dIII digestion was first performed to subclone a 0.9-kb fragment. U1 was then subcloned as a 0.2-kb *Hpa*I-*Dra*I fragment of this intermediate clone. d1 is a 1.4-kb *Hin*dIII subclone of a 2.8-kb *Pst*I-*Sal*I intermediate clone.

Pulsed-field gel electrophoresis: High molecular weight DNA was prepared from adult flies essentially as described previously (Birren and Lai 1993). About 250 frozen females were ground to fine powder in liquid nitrogen. The powder was transferred to a 15-ml Dounce homogenizer (Kontes); 15 ml of ice-cold NIB (10 mm Tris pH 8.5, 60 mm NaCl, 10 mm EDTA, 0.15 mm spermine, 0.5% Triton X-100) was added to the powder and the mix was homogenized with 5–10 strokes of pestle A. After transfer to an ice-cold Corex centrifuge tube, the mix was spun 15 min at 580 × *g* at 4° to pellet remaining large body parts. The supernatant was transferred to an ice-cold Corex tube and spun 8 min at 5800 × *g* at 4°. The pellet was resuspended in 250 μl NIB and warmed at 37°. An equal volume of preheated (50°) 1.2% InCert Agarose (FMC, Rockland, ME)/125 mm EDTA was added. Agarose blocks (100 μl) were molded and hardened 5–10 min on ice. The resulting plugs were incubated 20 hr at 50° in 250 μl of digestion buffer (50 mm EDTA, 1% SDS, 1 mg/ml proteinase K) and then dialyzed twice for 1 hr against 1× TE-40 μg/ml phenylmethylsulfonyl fluoride (PMSF) and four additional times each for 20 min in 1× TE. Samples were stored in 1× TE at 4°. For all tested enzymes, except for *Not*I (Promega, Madison, WI), digestion of DNA was performed in the liquid phase. A total of 40 μl of melted plug was incubated overnight with restriction enzyme buffer, BSA, and restriction enzyme in a final volume of 80 μl. For *Not*I digestion, half plugs were incubated overnight in 200 μl of a mix of restriction endonuclease buffer, BSA, and enzyme. Before migration, a second proteinase K digestion was performed either by adding 20 μg of enzyme to the liquid samples or by incubating solid samples in 1× TAE-0.1 mg/ml proteinase K. Pulsed-field gel electrophoresis (PFGE) was performed in 1% SeaKem LE agarose (FMC) gels run at 12°, 160 V, on a 2015 Pulsaphor Plus (LKB, Piscataway, NJ) apparatus, set at either 5- to 15-sec pulses for 20 hr (for restriction mapping) or 25-sec pulses for 24 hr (for analysis of deficiencies). Gels were stained in 0.5 μg/ml ethidium bromide for 30 min and then photographed, blotted, and hybridized as usual.

Poly(A)+ RNA extraction and Northern blot analysis: A total of 100 mg of flies was homogenized in 2 ml of RNA Plus solution (Quantum Biotechnologies, Blaine, WA). Addition of 200 μl of chloroform allowed the extraction of total RNA. Poly(A)+ RNA was extracted from this preparation with the polyATtract mRNA isolation system IV (Promega). Northern blots were performed as previously described (Pélisson *et al.* 1994) except that the running buffer was 2× 3-(*N*-morpholino) propanesulfonic acid (MOPS) instead of 1× MOPS.

PCR analysis: PCR was performed with genomic DNA essentially as described previously (Chalvet *et al.* 1999). The sequences of the primers flanking the *dip1* repetitive region were 5′-GCCTCTTCACTTTGACAG-3′ and 5′-CGGCACCAAT TCACCTACAG-3′.

RESULTS

***P*-induced *flamenco* mutagenesis:** *P[lyB]* is a nonautonomous *P* element carrying the y^+ genetic marker (Busseau and Bucheton 1999). Its mobilization can be induced by the *P* transposase *in trans*. It is located on chromosome *2* of a strain that also contains an *FM7c* balancer carrying a y^- allele and the restrictive allele $flam^{FM7c(R)}$. As shown in Figure 1A, females of this stock were crossed with males carrying the Δ2-3(99B) source of *P* transposase (Robertson *et al.* 1988) to induce mobilization of *P[lyB]*. About 1000 *FM7c* chromosomes acquired a transposed copy of *P[lyB]*, as judged by the linkage of the y^+ and *B* markers in G3. All but 1 of these *FM7c, P[lyB]* chromosomes, when combined in G2 females with a chromosome carrying the permissive $flam^{OR(P)}$ allele, displayed the same ability as the parental *FM7c* chromosome to prevent the transposition of *gypsy* into the female sterile ovo^{D1} allele (0.1% G3 fertile females). The exception was a mutant chromosome that allowed the inactivation of ovo^{D1} in five out of 23 G3 females observed (22% ovo^{D1} inactivation). This chromosome contains a recessive permissive mutation that does not complement the known permissive alleles of *flamenco* and is therefore denoted $flam^{by+(P)}$.

Mobilization of the *P[lyB]* element associated with this permissive chromosome was then induced by crosses with the Δ2-3(99B) transposase source (Figure 1B). Excisions of *P[lyB]* from the *FM7c, P[lyB]* chromosome were identified by looking for the loss of the y^+ marker in G1. As shown in Figure 2, out of 43 independent y^- derivatives of the *FM7c, P[lyB]* permissive chromosome, 33 appeared to carry a restrictive revertant allele. By contrast, the 15 y^+control derivatives all remained as permissive as the parental chromosome. The frequent co-occurrence of excisions with reversions of the $flam^{by+(P)}$ allele strongly suggests that the *P[lyB]* insertion is indeed responsible for this mutation.

Isolation of genomic DNA flanking the *P* insertion site: *The 100-kb walk:* Genomic DNA from flies homozygous for the $flam^{by+(P)}$ allele was used to construct a λ phage library. This library was screened with both ends of the *P* element and 6 kb of genomic DNA spanning the *P[lyB]* insertion point was recovered. Three kilobases of this DNA consist of unique sequences (data not shown). These sequences, in a fragment called U2, flank *P[lyB]* (Figure 3B) and hybridize to the 20A region of the polytene *X* chromosome (data not shown). U2 was used as an entry point to initiate a chromosome walk by screening a cosmid library (see materials and methods). Two positive clones were recovered (cos6a and cos7a, Figure 3A). A restriction map of these clones, established by partial *Pst*I digestion, shows that they encompass a total of 35 kb, 25 kb of which is common to both of them.

To look for additional unique sequences, Southern blots of single and double digestions of the inserts in the two cosmid clones were probed by radiolabeled whole genome DNA. Most of the fragments were readily labeled, indicating that these clones were rich in repetitive sequences. Two fragments that appeared weakly labeled

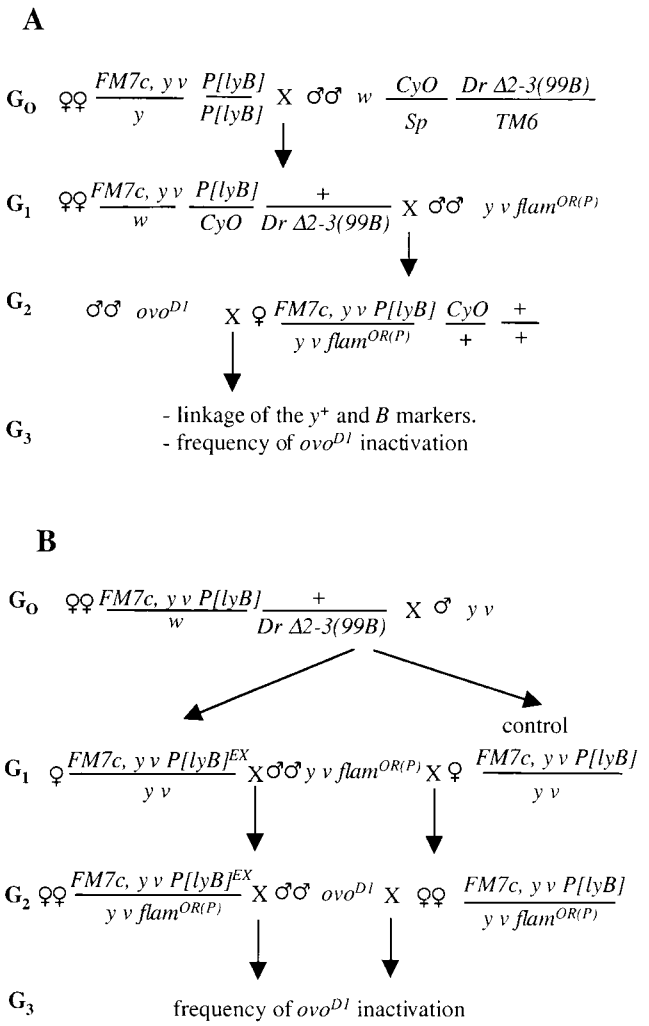


FIGURE 1.—Mating schemes used to select for transpositions of the *P[lyB]* element into and out of the *flamenco* gene. Only the relevant genetic markers are shown. (A) *P*-induced mutagenesis of the $flam^{FM7c(R)}$ allele on the *FM7c* restrictive chromosome. *FM7c, CyO,* and *TM6* are balancer chromosomes that are transmitted through meiosis with little or no recombination with the homologous chromosome. They carry, respectively, the *B, Cy,* and *Ubx* dominant markers, which, like *Sp* and *Dr*, can be recognized to select for the appropriate chromosome combinations in heterozygous flies. G1 females carrying the Δ2-3(99B) transposase source were crossed with males of the permissive *gypsy*-rich OR(520)/FM3 stock carrying the *y, v,* and $flam^{OR(P)}$ markers. *P[lyB]* transpositions in these G1 females were detected by the Cy y^+ phenotype in their daughters. G2 females that had inherited the *y v FM7c* balancer, as judged by their v B phenotype, and that did not inherit the *Dr Δ2-3(99B)* chromosome *3* were individually crossed with ovo^{D1} males. The linkage of the y^+ and *B* markers indicated that *P[lyB]* had inserted into the *FM7c* balancer. Mutation of its $flam^{FM7c(R)}$ restrictive allele into a $flam^{fy+(P)}$ permissive allele was detected by the presence of fertile G3 females (see MATERIALS AND METHODS for the rationale of the ovo^{D1} inactivation assay). (B) Reversion of the $flam^{fy+(P)}$ mutation by excision of the *P[lyB]* element from the *FM7c, P[lyB]* mutant chromosome. G0 females containing the Δ2-3(99B) source of transposase were crossed with males from a *y v* restrictive strain. $P[lyB]^{EX}$ represents independent excision events of the *P[lyB]*-linked y^+ marker, which were selected in G1 as y v B females. These females were individually crossed with males of the permissive *gypsy*-rich OR(520)/FM3 stock. As a control, some of their y^+ v B sisters were also studied. The ability of both the parental and excision chromosomes to control *gypsy* transposition was quantified by the ovo^{D1} inactivation assay.

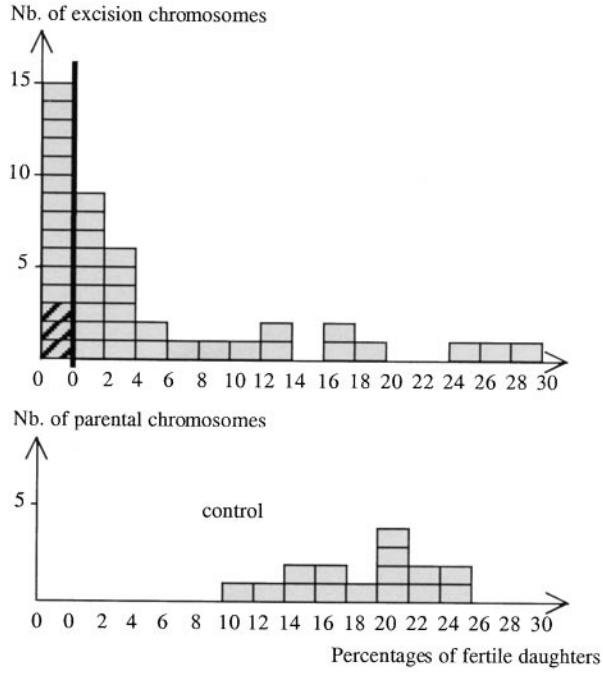


FIGURE 2.—Excision of the y^+ marker from the *FM7c,* $flam^{fy+(P)}$ mutant chromosome is frequently associated with reversion of the permissive mutant phenotype. The numbers of independent excision events recovered in G1 females (denoted *FM7c,* $P[lyB]^{EX}$ in Figure 1B) are plotted on the *y*-axis of the histogram. The *x*-axis indicates their respective levels of permissiveness for *gypsy* mobilization as measured by the percentage of fertile ovo^{D1} daughters of $flam^{OR(P)}$*/FM7c,* $P[lyB]^{EX}$ G2 mothers. The distribution of permissiveness levels of some *FM7c, P[lyB]* parental chromosomes is represented by the control histogram. The three hatched squares correspond to revertants *679-5b, 679-14a,* and *679-16a*, the molecular structure of which was studied (see Figure 5).

by the whole genome probe were further analyzed by several rounds of subcloning and hybridization to genomic blots until single-copy fragments could be identified. Only one such sequence was found by this method, a 0.2-kb *Hpa*I-*Dra*I fragment that is hereafter referred to as U1 (see MATERIALS AND METHODS and Figure 3). The dissection process also revealed a 1.4-kb *Hin*dIII fragment (denoted d1), which often labels more than one fragment when used as a probe on genomic DNA. Further Southern blotting analyses indicated that d1 is part of an ~6-kb tandem duplication. Both copies of the duplication, called D1 and D1′, are contained in a 13-kb *Avr*II-*Sfi*I fragment (Figure 3B). U1 and d1 are both located in the same part of the cos7a cosmid, which also contains the *P[lyB]* insertion point and U2. The

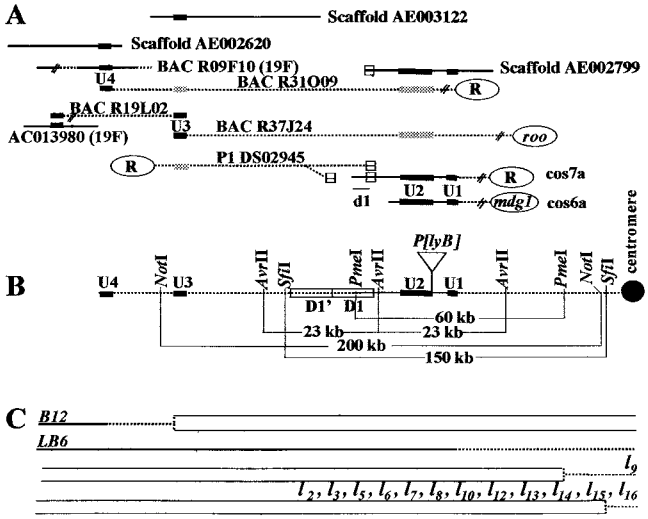


FIGURE 3.—Localization of the walk with respect to genomic landmarks, including the breakpoints of deficiencies in the *flamenco* region. (A) Description of some clones of the walk. BAC R37J24 represents the most proximal group of BACs and BAC R31O09 is the most distal clone of the BAC contig. Positions of the proximal ends of the cosmids and BACs and the distal end of the DS02945 P1 phage are not defined and the figure is not to scale. Some unique sequences used as genomic landmarks are represented by solid boxes, which are drawn on top of each other when they match in BLAST. They are represented by shaded boxes when they hybridize with the indicated clone. Dm1994, the STS of the P1 phage that matches the d1 landmark, is represented by an open square. Since we do not know to which copy of the D1-D1′ tandem duplication this STS belongs, both possible proximal termini of this phage are represented. Ovals indicate STSs containing either unknown repetitive DNA (R) or identified transposable elements. Sequences corresponding to additional repetitive DNA have been identified around the *P[lyB]* insertion (see ovals in Figure 4). Solid lines indicate sequenced regions, including the sequenced part of cos7a (AF182444), and dotted lines indicate regions of unknown sequence. Only a partial sequence of BAC R09F10 is known (draft sequence determined by BDGP); it matches genes located in cytological region 19F. (B) Partial map of the *413(NP)* chromosome showing the known unique sequences, the D1′-D1 duplication, and some restriction sites (same symbols as in A). The position where the *P* element inserted in the *FM7c* chromosome (coordinate 5264 in GenBank accession no. AF182444) is represented by an inverted triangle. (C) Map of 15 deficiencies in the *flamenco* region. *B12* deletes only U4; *LB6* deletes all the landmarks; none of the proximal deficiencies uncover any of the landmarks. Southern blotting of pulsed-field gels (Figure 6) did not show any change in the size of large restriction fragments except for l_9; this alteration of the l_9 restriction map is located between the proximal *Pme*I and *Not*I/*Sfi*I sites. Solid lines define regions clearly established as either intact (double line) or deleted (single line) in the deficiency; the dotted lines define regions of uncertainty.

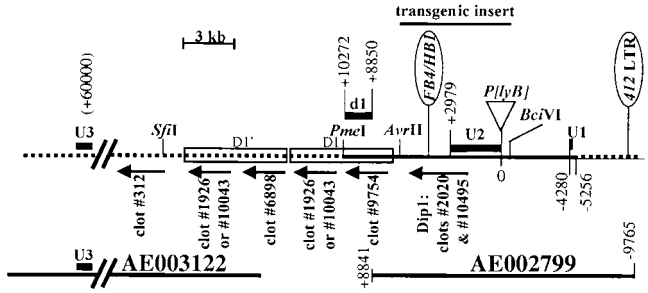


FIGURE 4.—Position of the landmarks, the repeated sequences, and the different transcription units with respect to the *P[lyB]* insertion point. This is a close-up of a portion of Figure 3B with essentially the same symbols; the sequenced regions are represented by continuous lines. Arrows define the 5′–3′ direction of the transcription units. The region introduced into flies for the rescue experiments is indicated above the map.

cosmids and their subclones were used to sequence over 15.8 kb in the region of the insertion (GenBank accession no. AF182444).

The sequence of most of the genome of *Drosophila melanogaster* was reconstructed by assembly of overlapping sequences into contigs, which were then ordered and oriented with respect to one another to obtain large pieces of sequence called "scaffolds" (MYERS *et al.* 2000). To determine whether the *flamenco* region was included in any of these scaffolds, we performed a BLASTN analysis (ALTSCHUL *et al.* 1990) of the Celera/BDGP, release 2, with the sequences of U1, U2, and d1. U1 and U2 both match a 18.4-kb scaffold (AE002799). The first 14 kb of our 15.8-kb AF182444 sequence align with AE002799, which makes the known sequence ~20.3 kb long (Figures 3 and 4). The d1 landmark matches a 55.4-kb scaffold (AE003122), the first 1.5 kb of which can be aligned with the end of AF182444. However, AE003122 does not contain any duplicated sequence. Hence, this match probably corresponds to the ectopic overlap between the part of the second repeat of the duplication (D1′) present at the beginning of AE003122 and the part of the first repeat (D1) present at the end of AF182444 (Figure 4). Neither of the scaffolds AE002799 and AE003122 has yet been mapped in the Drosophila genome. Because of the genetic position of *flamenco* between *dodo* and *S6kII*, we assume that these scaffolds, which contain sequences from the *flamenco* region, are located inside the physical gap between the two scaffolds containing these genes (AE002620 and AE002629, respectively).

To determine whether it was possible to identify repetitive sequences and their location in the sequenced DNA, we compared AE002799, AF182444, and AE003122 with the Drosophila transposon sequences database of the BDGP by performing BLASTN analysis. The 9765 bp located between the *P[lyB]* insertion point and the beginning of AE002799 appeared to be composed mostly of sequences showing low levels of homology to various LTR retroelements. The only nondegenerate sequence corresponds to the 5′ end of the *412* LTR retrotransposon, which can be clearly recognized between −9397 and −9711. By contrast, comparatively very few hits were obtained with the 70 kb of DNA sequenced between

the *P[lyB]* insertion point and the end of AE003122. An *FB4/HB1* DNA transposon is located between positions +3597 and +4929 (Figure 4) and long stretches of the micropia LTR retrotransposon can be recognized between coordinates 12,040 and 15,340 of AE003122 (data not shown). All but one of the six degenerate retroelements found in these 70 kb are clustered in the last 1.5 kb of scaffold AE003122.

Additional clones were obtained from different libraries constructed by the Drosophila sequencing genome projects after BLAST analysis of the BDGP/European Drosophila Genome Project (EDGP) genomic clones and the sequence-tagged site (STS) databases with the U1, U2, and d1 sequences used as queries. The draft sequence of the bacterial artificial chromosome (BAC) clone R32M16 was found to contain U2 and d1. A more detailed analysis reveals that it also contains unique sequences from cytological sections 15–16. We concluded that this clone is rearranged. The STS Dm1994 of the P1 phage DS02945 overlaps d1 in the same orientation. The DS02945 insert is ~70 kb long, which makes the walk at least 100 kb long (Figure 3A). We could not proceed further in this direction because the other STS of DS02945 is repeated. The same is true in the other direction since the STSs of cos6a and cos7a match, respectively, mdg1 and an unknown repetitive sequence.

The BAC contig: To get longer genomic DNA, U2 was used to probe the BDGP BAC library, which contains inserts of 165 kb on average. A contig of 23 positive clones was obtained. The inserts of 19 of them, as well as those of the P1 and cosmid clones, were studied by the fingerprint technique described by Marra *et al.* (1997). The results of this analysis (not shown) can be summarized as follows: (i) Four BACs have very little overlap with the others. They were considered as rearranged and were ignored in this study. (ii) Among the 15 others, two groups of 4 very similar BACs can be formed; one of them, represented by the clone R37J24, seems to be located at one end of the contig (Figure 3A). (iii) The clone R31O09 seems to be located at the other end of the contig.

The STSs of 16 of the 19 nonrearranged clones are in the BDGP database: No clone has 2 unique STSs; 10 clones, including R31O09 and R37J24, have 1 unique STS (the other STS being made of repetitive elements); and 6 clones have 2 repetitive STSs. The unique STSs of clones R31O09 and R37J24 were respectively named U4 and U3. The overall organization of the contig deduced from the fingerprint analysis was confirmed by BLAST analysis performed with the 10 unique STSs. Several clones are indeed very similar since they share at least pieces of STSs. This analysis also gave the orientation of some of the BAC clones with respect to each other and to the centromere, as shown by the following observations (Figure 3A): (i) U4 matches the draft sequence of an additional BAC clone (R09F10) known to be located in cytological section 19F, which is in a more distal position than U2 with respect to the centromere; (ii) U3 is also a more distal STS, since it matches the complementary strand of 1 STS of the BAC R19L02, whose second STS matches GenBank entry AC013980 also originating from 19F. Moreover, this BLAST analysis of unique STSs suggested that they are all on the distal side of the contig whereas most of the repetitive ones are proximal with respect to the centromere (data not shown). The sequence of the 43-kb most distal part of this contig is known, since U4 (but not U3) matches the complementary strand of scaffold AE002620 between positions 43,194 and 42,445.

Orientation of the walk with respect to the centromere: The organization of the BAC contig was further confirmed by using some PCR-amplified STSs to probe the blotted fingerprints. One of them, U3, enabled us to anchor the walk to the BAC contig and therefore to orient it with respect to the centromere. U3 labels the P1 DS02945 but neither of the cosmids cos6a and cos7a, which are therefore proximal to the P1 clone (Figure 3A). The fact that U3 is also complementary to the end of AE003122 is in agreement with the orientation of this scaffold toward the telomere.

Molecular analysis of revertants: As shown in Figure 5A, the *yellow* gene, which was used as a marker to select for excisions, is located in the middle of the *P[lyB]* element. To check for the possible retention of the flanks, *Bam*HI-*Hin*cII and *Bam*HI-*Hin*dIII DNA digests of three revertants were hybridized successively with both ends of this element. In fact, the first probe was a *Bam*HI-*Eco*RI hybrid fragment containing 240 bp of the 3′ end of *P[lyB]* flanked by 160 bp of genomic DNA (Figure 5A). This probe labeled the same fragments in revertants *679-16a* and *679-14a* as in the $flam^{by+(P)}$ mutant flies, indicating that they result from imprecise excisions leaving at least the right-hand side of the element in the footprints (Figure 5B). By contrast, revertant *679-5b* showed the same restriction pattern as the parental *FM7c* chromosome, which, at this level of analysis, may be interpreted as a precise excision event. These conclusions were supported by hybridization with a 274-bp probe (coordinates 239–512 in the *P*-element sequence) specific for the 5′ end of *P[lyB]*. Only *679-16a* and *679-14a* hybridized to this probe (data not shown). In the case of *679-16a*, the sizes of the *Hin*cII-*Hin*cII and *Hin*dIII-*Bam*HI fragments were different from those of the corresponding fragments in $flam^{by+(P)}$ and *679-14a*, indicating that only the *Hin*cII-*Hin*dIII 1-kb fragment flanking the *P[lyB]* element was unaffected by this imprecise excision.

Localization of the walk with respect to the breakpoints of deficiencies: We used the unique sequences previously identified as landmarks to determine the localization of the walk with respect to the breakpoints of the deficiencies affecting the *flamenco* region (Prud'homme *et al.* 1995). All these deficiencies are homozygous lethal because they uncover either the *eo* or the *wap* locus. To

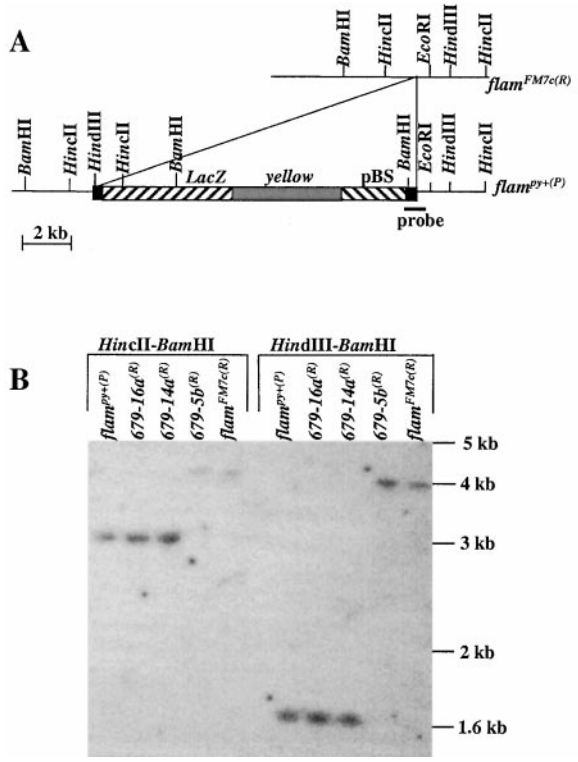


FIGURE 5.—Molecular structure of the *P*-induced *flamenco* mutant and its revertant derivatives. (A) Partial restriction maps of the $flam^{FM7c(R)}$ and $flam^{py+(P)}$ parental and mutant alleles (restriction sites not required for interpretation of the results are not shown). The box represents the *P[lyB]* insertion, bracketed by the *P*-element termini (in black) and flanked by *FM7c* genomic DNA (thin line). (B) Comparison of the hybridization patterns of three revertant alleles (*679-16a*, *679-14a*, and *679-5b*) with those of the $flam^{FM7c(R)}$ and $flam^{py+(P)}$ parental and mutant alleles. The genomic DNA was digested by both *Bam*HI, which cleaves the right-hand end of *P[lyB]*, and by either *Hinc*II or *Hin*dIII, which cut in the DNA flanking the insertion. The blot was hybridized with a 400-bp *Bam*HI-*Eco*RI fragment subcloned from the $flam^{py+(P)}$ genomic DNA. This probe includes the right-hand end of *P[lyB]* and 160 bp of the flanking DNA. Revertant *679-5b* has the same restriction pattern as the parent $flam^{FM7c(R)}$ allele, which suggests that it results from the precise excision of the whole *P* element. The two other revertants, which have the same *Bam*HI-*Hinc*II and *Bam*HI-*Hin*dIII junction fragments as the mutant $flam^{py+(P)}$, retain the right-hand side of the insert.

solve the problem of lethality, several different strategies were used to ask whether or not a unique sequence is uncovered by a deficiency. First, we took advantage of the existence of restriction site polymorphisms to compare various heterozygous combinations. Second, when there was no polymorphism, we quantified the signals obtained on the Southern blots to determine whether they resulted from the presence of one or two allelic doses. Third, since most of the deficiencies to be studied were embryonic lethal, we performed Southern blots with DNA extracted from homozygous embryos selected on the basis of their lethal phenotype (see MATERIALS AND METHODS). Fourth, to resolve the final ambiguities, we took advantage of the observation that the *LB6* deficiency uncovers all the unique sequences tested and combined it with each deficiency studied. The results obtained by these four different strategies are summarized in Table 1 and Figure 3C. They show that none of the $flam^{+}$ deficiencies delete U1 or U2 but that the LB6 $flam^{-}$ deficiency deletes both of these landmarks. This confirmed that the walk was indeed situated in the *flamenco* region. Moreover, U4 but not U3 is deleted by the distal deficiency *B12*, which confirmed the orientation of the walk with respect to the centromere and allowed the mapping of the *B12* deficiency (and therefore at least part of the *eo* sequences) distal to the U3 landmark, which is >60 kb from the *P[lyB]* insertion point (Figures 3 and 4).

None of the unique sequences are uncovered by the proximal deficiencies tested. All the proximal deficiencies denoted l_x were derived from the same *413(NP)* restrictive chromosome (PRUD'HOMME *et al.* 1995). The restriction map of this chromosome was established by digesting its DNA with rare cutting enzymes and probing with U2 and d1 (Figure 3B). We then asked whether any of the restriction fragments with a long proximal extension would be affected by any of these proximal deficiencies. l_9 (Figure 3C) is the only deficiency for which modified fragments were observed; the 200-kb *Not*I fragment (Figure 6) and the 150-kb *Sfi*I fragment (data not shown) present on the parental chromosome were replaced, respectively, by a 250-kb and a 200-kb fragment in l_9. This was not the case for the 23-kb *Avr*II and the 60-kb *Pme*I fragments. We can conclude that: (i) l_9 is the closest rearrangement to U2, although its distal breakpoint is located between the *Pme*I and *Not*I/*Sfi*I sites, that is, between 40 kb and 130 kb away from the *P[lyB]* insertion point; (ii) the breakpoints of the two other $flam^{-}$ proximal deficiencies are located proximal to the *Not*I/*Sfi*I sites, that is, at >130 kb from the *P[lyB]* insertion point. Two hypotheses can explain these results: Either *flamenco* is a gene larger than 130 kb or the region contains more than one gene involved in the regulation of *gypsy* (see DISCUSSION).

At least four different genes are present in the walk: With the aim of determining which gene(s) the *P[lyB]* insertion may affect, we looked for expressed sequence tags (ESTs) by BLASTN analysis. The location of the corresponding cDNAs was confirmed by hybridization to the genomic clones of the walk (Figure 4). At least seven distinct BDGP clots (nos. 312, 1926, 2020, 6898, 9754, 10,043, and 10,495), and several individual cDNAs not classified in clots, were positive. They all map distal to the *P[lyB]* insertion point with respect to the centromere. Four of the clots hybridize with the 6-kb duplication described above. They correspond to two pairs (1926/10,043 and 6898/9754) of very similar clots. Both clots of a pair appear to differ not only by alternative splicing but also by point mutations, which suggests that each originates from a different copy of the duplication.

TABLE 1

Genetic and molecular analysis of deficiencies of the *flamenco* region

	Genetic analysis			Molecular analysis			
	eo	*flam*	*wap*	U4	U3	U2	U1
Distal deficiencies							
B12	−	+[e]	+	−[a]	+[b]	+[a]	NT
LB6	−	−	+	NT	NT	−[a,c]	−[c]
Proximal deficiencies							
R20	+	−	−	NT	NT	+[a,c]	+[c]
R21	+	−	−	NT	NT	+[c]	+[c]
16.2.13	+	−	−	NT	NT	+[a]	NT
DCB1-35c	+	−	−	NT	NT	+[a]	NT
l_3, l_9, l_{15}	+	−	−	NT	NT	+[a,c]	+[c,d]
l_2, l_5, l_6, l_7, l_8, l_{10}, l_{12}, l_{13}, l_{14}, l_{16}	+	+	−	NT	NT	+[a,c]	+[a]
R44	+	+	−	NT	NT	+[c]	NT
l_{11}	−	−	−	NT	NT	−[a,c]	NT

The genetic data are the results of complementation tests between the deficiencies and the three mutant genes; for *flam*, mutant (−) and wild type (+) correspond, respectively, to the permissive and restrictive phenotypes. The − or + molecular data mean that the genomic landmark is or is not deleted by the deficiency. NT, the deficiency was not studied with this particular probe.

[a,b,d] The presence of the four genomic landmarks was investigated by Southern blot analysis of DNA from heterozygous females ([a,b], +/deficiency or [d], *LB6*/deficiency).

[b] The intensity of the band was quantified to compare the number of doses with that of the wild-type control.

[c] Southern blot analysis was performed with DNA of embryos homozygous for the deficiency.

[e] The chromosome bearing the *B12* deficiency is permissive, as a result of a secondary mutation that could be recombined away from the deficiency (Chalvet *et al.* 1999).

For instance, the consensus sequence of clot 9754 is perfectly identical to the sequenced exonic part of the D1 copy whereas up to seven differences are observed when this genomic sequence is compared with the 660 corresponding nucleotides of clot 6898. We infer that the clones composing clot 9754 are encoded by the proximal copy D1 and the clone of clot 6898 by the distal copy D1′ (Figure 4).

The largest member of each clot or pair of clots was sequenced and analyzed by BLASTX. No hit was found for the pair of clots 1926/10,043. Clot 312 shares some common domains with the mammalian glucosidases II. The pair of clots 6898/9754 presents high similarities to *Caenorhabditis elegans* and *Saccharomyces cerevisiae* open reading frames (ORFs) and to murine and human ESTs, as well as lower similarities to prokaryotic genes involved in nitrogen metabolism. The clots 2020 and 10,495 correspond to two groups of cDNAs that differ only by alternative splicing. Three more cDNAs belonging to this class, dip1a (AF175713), UbxBP1 (AF218310), and klett (AJ250866), have been independently described, the products of which respectively interact with the Disco, Ultrabithorax, and SU(VAR)3-9 proteins. Putative proteins encoded by members of this class have a bipartite nuclear localization signal and two double-stranded RNA-binding domains (dsRBD). The dsRBD is present in many proteins (St. Johnston *et al.* 1992; Gibson and Thompson 1994), two of which are human proteins involved in the control of virus propagation (Hovanessian 1989; Gatignol *et al.* 1991). Because of these two domains and because this transcription unit (hereafter called *dip1*) is the closest to the *P[lyB]* insertion point, we decided to characterize it and to test its ability to control *gypsy* mobilization.

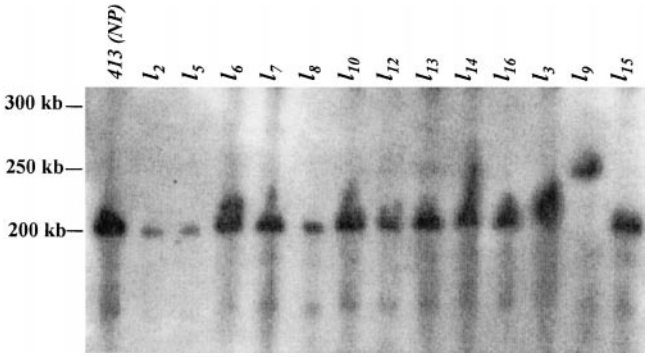


Figure 6.—Size of the *Not*I fragment in the proximal deficiencies. High molecular weight DNA was prepared from females heterozygous for *LB6* and the proximal deficiencies l_x, analyzed by PFGE, and hybridized with U2. 413(NP) is the parental chromosome of the l_x deficiencies.

Failure to rescue *flamenco* permissive flies with genomic *dip1* transgenes: As shown in Table 2 and Figure 7A, at least 11 different types of *dip1* transcripts were obtained by using U2 to screen cDNA libraries and EST databases. They vary by alternative splicing and polyadenylation. All contain an ORF with the bipartite

TABLE 2

Description of the *dip1* cDNAs clones

Clone	Genbank accession no.	BDGP clot	Splicing group	poly(A)
dip1a[a]	AF175713		a	A2
UbxBP1[a]	AF218310		b	A1
LD14381[b]	AA439741	2,020	b	A2
	AF182154			
GH15935[b]	AI292905	2,020	b	
LD45242[b]	AI513026		b	A3
LP07878[b]	AI294495	2,020	b	
TO34[c]	AF175711		c	A3
	AF182951			
LD19452[b]	AA540094	2,020	c	
LD08867[b]	AA390793		c	Chimeric[d]
GH20159[b]	AI389075		d	A0
LD36232[b]	AI456315	10,495	d	
LP11783[b]	AI297480	10,495	d	A3
GH26266[b]	AI406089	10,495	d	
LP02946[b]	AI259517	10,495	d	A3
GH26615[b]	AI513497	10,495	d	A0, A1, or A2
LP12356[b]	AI297912	10,495	d	A3
LP07724[b]	AI294366	10,495	e	A0, A1, or A2
GH21996[b]	AI402716	10,495	e	A3
TO67[c]	AF213338		f	A3
GH17590[b]	AI387147		f	A0
GH16142[b]	AI293062	10,495	d, e, or f	
LD23092[b]	AA940836	10,495	e or f	A0, A1, or A2
GH18616[b]	AI387917	10,495	?	

The cDNAs have different 5′ ends. Alternative splicing allows their classification in six different splicing groups, numbered a to f (Figure 7A). The clones truncated at the 5′ end could not be classified. At least four different polyadenylation sites (denoted poly(A) 0–3) were found. Their respective positions in AF182444 are 16,473, 16,478, 16,489, and 16,831. Restriction analysis was used to determine whether a cDNA is polyadenylated in A3 or in the region where A0, A1, and A2 are grouped. The distinction between A0, A1, and A2 was done by sequencing.

[a] The cDNA was selected in a double hybrid screen.

[b] The cDNA was recovered as an EST matching by BLAST with the U2 sequence.

[c] The cDNA was recovered by screening a library with the probe U2.

[d] LD08867 appears to be a chimeric clone: Its 5′ EST is located in the *flamenco* region and its 3′ EST matches with sequences of the cytological section 12. Such rearranged clones are known to be present in the cDNA libraries constructed by the BDGP.

nuclear localization signal and the two dsRBDs. This ORF, however, may not be translated in every clone because of the presence of several short putatively translatable ORFs in the 5′ untranslated region (UTR). For instance, neither GH20159 nor TO67, which respectively correspond to isoforms d and f, could produce any detectable protein when its RNA was incubated in a reticulocyte lysate, whereas the representatives of splicing groups a, b, and c, namely dip1a, TO34, and LD14381, did (data not shown).

Four major bands were observed on a whole female Northern blot (Figure 7B) hybridized with a riboprobe common to all the transcripts (Figure 7A). Their size varies from strain to strain. We demonstrated by genomic PCR that this size polymorphism is linked to differences in the copy number of a 124- to 126-bp repeat located in the 3′ UTR (Figure 8). No correlation between the *flamenco* status of the strain and the number of copies could be established. It is worth noting that the R20 permissive deficiency does not obviously affect the pattern of *dip1* transcription (Figure 7B).

To test whether *dip1* is able to control *gypsy* mobilization, we tried to rescue *flamenco* permissive females with this gene. Because of the high number of different *dip1* transcripts, we constructed a genomic DNA transgene potentially able to produce all of them. An *Avr*II-*Bci*VI 8-kb fragment of the cosmid cos7a, which originates from the restrictive *y; cn bw sp* strain, was subcloned in the *Hpa*I and *Spe*I sites of the pCaSpeR4 vector (Figure 4) and 17 transgenic lines were established. Appropriate crosses (see MATERIALS AND METHODS) yielded females carrying one copy of the transgene and one copy of the *gypsy-lacZ* fusion in different *flamenco* backgrounds. *Gypsy-lacZ* is known to be repressed in restrictive follicle cells (PÉLISSON *et al.* 1994). We looked for the ability of the transgene to control the expression of *gypsy*. Neither in the $flam^{OR(P)}/df(1)l_{11}$ permissive nor in the $flam^{w1118(R)}/df(1)l_{11}$ restrictive background was a differential expression of *gypsy-lacZ* observed between the females that carried the transgene and those that did not (data not shown). Negative results were also obtained with two transgenes that contained the whole genomic sequence of cos7a (data not shown).

DISCUSSION

The *flamenco* gene is located in a region rich in repetitive sequences: With the aim of cloning the *D. melanogaster* gene *flamenco*, which is involved in the control of the mobilization of the *gypsy* retrovirus, we selected a *P*-element-induced permissive mutation of this gene. A unique *P* element was detected in this mutant and excisions of the *P* element often restored the restrictive function of the gene, indicating that the insertion was responsible for the permissive mutant phenotype observed. Since some of the reversions result from imprecise excisions, we infer that *P* is not inserted into a coding region of the gene. About 100 kb of genomic DNA were recovered by gene tagging and chromosome walking and a sequence of 20.3 kb spanning the insertion site was determined.

The *flamenco* gene is located in 20A1–3, between *eo* and *wap,* at the base of the polytenized *X* chromosome (PRUD'HOMME *et al.* 1995). Many of the cloned sequences are repetitive. For instance, a contig of >16 BACs covers this genomic area and 68% of their STSs contain repetitive DNA. This figure is much higher than

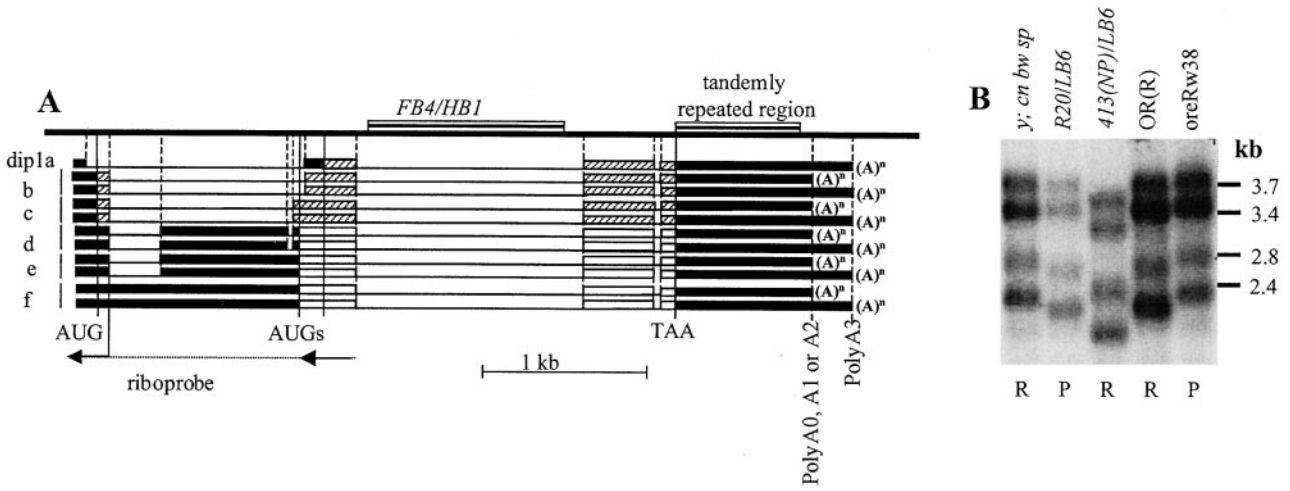


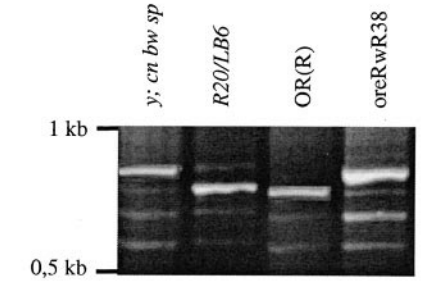
FIGURE 7.—The *dip1* transcription unit. (A) *Dip1* encodes at least 11 different transcripts. Their structure was studied by restriction mapping, PCR analysis, and partial sequencing. Solid boxes indicate the 5′ (left) and the 3′ (right) UTR. Hatched boxes represent the putative translated regions. Open boxes indicate ORFs that are unlikely to be translated because they are downstream of several putatively translatable small ORFs. Thin lines represent introns. The *FB4/HB1* element and the 124- to 126-bp tandem repeats are represented at the top. An antisense riboprobe was synthesized from a cDNA of class c; the sequences present in this probe and their orientation are indicated by two arrows. (B) Northern blot analysis of poly(A)+ RNA extracted from permissive (P) and restrictive (R) whole females and hybridized with the riboprobe shown in A. The RNA size polymorphism between strains is linked to the number of repetitions in the 3′ UTR, which, in this sample of strains (see Figure 8), varies between five and seven. Since the probe is deleted by the distal deficiency *LB6*, the pattern observed with females heterozygous for *LB6* corresponds to the RNA produced by the homologous chromosome. When compared to the restrictive *413(NP)* chromosome, the *R20* permissive proximal deficiency does not obviously affect the *dip1* expression pattern.

the 8.74% of repetitive reads found by MYERS *et al.* (2000) during the sequencing of the Drosophila genome. Some of these repetitive STSs match known transposable elements such as *roo*, *mdg1*, or *gypsy*. Such a wealth of repetitive DNA is characteristic of the β-heterochromatic region in which *flamenco* maps. This might explain why the BAC contig falls into a gap in the Celera sequencing project, since the shotgun sequencing of clustered long repetitive units is difficult to assemble (MYERS *et al.* 2000). Unique sequences and functional transcriptional units were identified only in the distal part of the cloned region.

One of the genes of the walk contains two double-stranded RNA-binding domains also present in some regulators of viral expression: The closest gene to the insertion point of the *P* element was assumed to be the best candidate for *flamenco*, because of the following circumstantial evidence: (i) As expected for a *P*-induced mutation, the insertion occurred 5′ of this gene. In this case, *P[lyB]* is inserted ~2 kb upstream of the beginning of the coding region. (ii) This gene, which was discovered independently (AF175713) in a double hybrid screen for interactors of the Disco protein [hence its name *disco interacting protein 1* (*dip1*)], might be involved in the tissue-specific autoregulation of the *disco* gene (LEE *et al.* 1999). (iii) As expected for a transcription factor, the Dip1 protein contains a putative nuclear localization signal and accumulates in the nuclei of many tissues (M. MUKHOPADHYAY, D. DESOUSA, P. PELKA and A. REGINA CAMPOS, personal communication), including the follicle cells where *flamenco* downregulates *gypsy* expression (PÉLISSON *et al.* 1994). (iv) Moreover, Dip1 also contains dsRBDs, which is reminiscent of two dsRNA-binding proteins that behave as regulators of viral proliferation. The first one is PKR, a serine/threonine kinase. During viral infection this protein is activated by the presence of dsRNA to become competent to block translation (HOVANESSIAN 1989; CLEMENS 1996). PKR can also be induced by interferon treatment and mediates the antiviral and antiproliferative effects of interferon (DUBOIS *et al.* 1989; CHONG *et al.* 1992). The second dsRBD-containing protein is the TAR-binding protein (TRBP). It was first identified because of its ability to bind the TAR stem-loop structure of human immunodeficiency virus (HIV; GATIGNOL *et al.* 1991). TRBP represses HTLV (DONZEAU *et al.* 1997) and activates HIV by antagonizing the dsRNA-mediated translational inhibition by the PKR (PARK *et al.* 1994; BENKIRANE *et al.* 1997).

However, none of the attempts to show that *dip1* is *flamenco* have succeeded. First, the proximal deficiencies affecting *flamenco* do not show any obvious effect on the expression of *dip1*, as shown by Northern blots of whole females. Their effect, if any, must be subtle, for instance, affecting *dip1* expression only in a few tissues. Second, a genomic fragment cloned from a restrictive strain and encompassing the *dip1* gene and ~2.5 kb of upstream DNA was unable to repress *gypsy* expression in an otherwise permissive background. Moreover, chromosomal rearrangements that knock out the *flamenco* function

A



B

strain, chromosome or cDNA	*flamenco* status	number of copies	order of repeats
TO34 (CantonS)	?	5	r1, r7, r6, r5, r4
413(NP)	R	5	
R20	P	6	
Rev	R	6	
OR(R)	R	6	
UbxBP1	?	6.5	r'1, r1, - , - , - , - , r4
oreRwR38	P	7	
wOR(P)	P	7	
y; cn bw sp	R	7	r1, r2, r3, r4, r5, r5, r4
w1118	R	7	
dip1a	?	8	r1, - , - , - , - , - , - , r4

C

```
r1   AGCGACACCACTTAAGACACCAACTCCAACCGCAGACTCCAACTGCCGACGACAACACACATAACGGACACTCTTAACGGACAGCAATCAGAAGATTGCGACCCTGACGTCAATTGCGCAACTTTA
r2   AGCGACACCACTTAAGACACCAACTCCAACCGCAGACTCCAACTGCAGACGACAACACACCTAT..GCCTCTCTTAACGGACAGCAATCAGAAGATTGCGACCCTGACGTCAATGGCGCAACTTTA
r3   AGCGACACCACTTAAGACACCAACTCCAACCGCAGACTCCAACTGCAGACGACAACACACCTAT..GCCTCTCTTAACGGACAGCAATCGGAAGATTGCGACCCTGACGTCAATTGCGCAACTTTA
r4   AGCGACACCACTTAAGACACCAACTCCAACCGCAGACTCCAACTGCCGACGACAACACACCTAT..GCCTCTCTTAACGGACAGCAATCAGAAGATTGCGACCCTGACGTCAATGGCGCAACTTTA
r5   AGCGACACCACTTAAGACACCAACTCCAACCGCAGACTCCAACTGCCGACGACAACACACCTAT..GCCTCTCTTAACGGACAGCAATCGGAAGATTGCGACCCTGACGTCAATGGCGCAACTTTA
r6   AGCGACACCACTTAAGACACCAACTCCAACCGCAGACTCCAACTGCCGACAACAACACACCTAT..GCCTCTCTTAACGGACAGCAATCAGAAGATTGCGACCCTGACGTCAATGGCGCAACTTTA
r7   AGCGACACCACTTAAGACACCAACTCCAACCGCAGACTCCAACTGCAGACGACAACACACCTAT..GCCTCTCTTAACGGACAGCAATCGGAAGATTGCGACCCTGACGTCAATGGCGCAACTTTA
r1'  AGCGACACCACTTAAGACACCAACTCCAACCGCAGACTCCAACTGCCGACGACAACACACATAA
```

FIGURE 8.—Study of the tandemly repeated region of *dip1*. (A) Example of a PCR experiment performed to determine the number of repeats in four different genotypes. The products obtained by genomic PCR with oligonucleotides flanking the repetitive region (see MATERIALS AND METHODS) were separated on an agarose gel and stained with ethidium bromide. The size of the major bands corresponds to either six, for *R20* and OR(R), or seven times, for *y; cn bw sp* and oreRw38, the size of the 124- to 126-bp repeated unit. (B) Structure of the tandemly repeated region in all the strains studied by genomic PCR and/or sequencing of genomic or cDNA clones. The number of repeats varies between five and eight irrespective of the permissive/restrictive *flamenco* status of the strains. (C) Alignment of the seven different copies sequenced in the *y;cn bw sp* genomic DNA and the TO34, UbxBP, and dip1a cDNAs. BLASTN analysis and Southern blotting demonstrated that these sequences are not present elsewhere in the genome. In r1, the most divergent variant, a sequence of 8 nucleotides seems to have been replaced by an almost completely different sequence of 10 nucleotides (underlined). The six other copies differ only by a few point mutations (boldface characters). As shown in B, the repeated region always starts with the r1 copy (or with its r1′ truncated version in UbxBP1) and ends with r4.

affect DNA located >130 kb upstream of *dip1*, suggesting that, if *dip1* is *flamenco,* it has regulatory regions located much farther away.

Some deficiencies, permissive for *gypsy* mobilization, are located >130 kb away from the *P*-element insertion point, proximal to the chromocenter: None of the unique sequences flanking the *P*-element insertion point are deleted by the deficiencies known to uncover both the *flam* and *wap* complementation groups. Southern blot analysis of pulsed-field gel electrophoresis showed that the breakpoint of the l_9 permissive deficiency is located >40 kb away from the *P*-element insertion point and that the breakpoints of all the other deficiencies (2 permissive and 10 restrictive were studied in these experiments) map >130 kb away. It seems unlikely that the permissive *flamenco* phenotype of all three X-ray-induced deficiencies is due to secondary mutations, which would have arisen simultaneously and independently from the DNA rearrangements themselves. This suggests that sequences responsible for the *flamenco* function lie >130 kb away from each other.

The search for the proximal limit of such sequences will require the mapping of the distal breakpoints of all these 13 deficiencies. Provided that these breakpoints are located randomly enough, this limit should be located in a short interval corresponding to the "overhang" between the shortest of the permissive and the longest of the restrictive deficiencies. *A priori*, different kinds of sequences could be found there: (i) some long-range *cis* regulator(s) of one of the short transcription units (like *dip1*) located in the vicinity of the *P*-element insertion site; (ii) some sequence belonging to an unusually long, yet unidentified, transcription unit, which would start close to this site; (iii) a second hypothetical short gene, which, in addition to another one located near the insertion site, would also be involved in the repression of *gypsy*; or (iv) some repetitive DNA. Indeed, some of the repetitive sequences interspersed in the β-heterochromatin seem to play a role in the expression and the control of the genes located in their neighborhood (WAKIMOTO and HEARN 1990). Moreover, sequences deriving from retroelements were reported to be involved in retrovirus control. For example, Fv1 (BEST *et al.* 1996) and Fv4 (KAI *et al.* 1986) are sequences affecting the susceptibility of mice to the Friend virus. They respectively encode a Gag-related protein that interferes with the retroviral life cycle by an unknown mechanism and an endogenous retroviral envelope protein that blocks cell surface receptors. Degenerate retroviruses were found in the neighborhood of the

P-element insertion point. For instance, a 1.4-kb *Pst*I fragment located 20 kb proximal to this site was found to hybridize to the full-length *gypsy* probe (data not shown). It might also be informative to investigate the possible implication of such sequences in the regulation of *gypsy*.

We thank Geneviève Payen-Groschene for her excellent technical assistance and Léonie Ringrose for reading the manuscript. V.R. was the recipient of fellowships from the MESR and La Ligue Nationale Contre le Cancer. This work was supported by grants from the Association pour la Recherche sur le Cancer (ARC) and the FRM.

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.

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.

Andrews, J., M. Smith, J. Merakovsky, M. Coulson, F. Hannan *et al.*, 1996 The *stoned* locus of *Drosophila melanogaster* produces a dicistronic transcript and encodes two distinct polypeptides. Genetics **143:** 1699–1711.

Benkirane, M., C. Neuveut, R. F. Chun, S. M. Smith, C. E. Samuel *et al.*, 1997 Oncogenic potential of TAR RNA binding protein TRBP and its regulatory interaction with RNA-dependent protein kinase PKR. EMBO J. **16:** 611–624.

Best, S., P. Le Tissier, G. Towers and J. P. Stoye, 1996 Positional cloning of the mouse retrovirus restriction gene *Fv1*. Nature **382:** 826–829.

Birren, B., and E. Lai, 1993 *Pulse-Field Gel Electrophoresis. A Practical Guide*. Academic Press, London.

Bridges, C. B., 1938 A revised map of the salivary gland X-chromosome of *Drosophila melanogaster*. J. Hered. **29:** 11–13.

Bucheton, A., 1995 The relationship between the *flamenco* gene and *gypsy* in *Drosophila*: how to tame a retrovirus. Trends Genet. **11:** 349–353.

Busseau, I., and A. Bucheton, 1999 A Drosophila enhancer detector transposon marked with the *yellow* gene. Dros. Inf. Serv. **82:** 94–95.

Chalvet, F., L. Teysset, C. Terzian, N. Prud'homme, P. Santamaria *et al.*, 1999 Proviral amplification of the *gypsy* endogenous retrovirus of *Drosophila melanogaster* involves *env*-independent invasion of the female germline. EMBO J. **18:** 2659–2669.

Chong, K. L., L. Feng, K. Schappert, E. Meurs, T. F. Donahue *et al.*, 1992 Human p68 kinase exhibits growth suppression in yeast and homology to the translational regulator GCN2. EMBO J. **11:** 1553–1562.

Clemens, M. J., 1996 Protein kinases that phosphorylate eIF-2α and eIF-2β and their role in eukaryotic cell translational control, pp. 139–172 in *Translational Control*, edited by M. J. Clemens. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Costa, M., E. T. Wilson and E. Wieschaus, 1994 A putative cell signal encoded by the *folded gastrulation* gene coordinates cell shape changes during Drosophila gastrulation. Cell **76:** 1075–1089.

Desset, S., C. Conte, P. Dimitri, V. Calco, B. Dastugue *et al.*, 1999 Mobilization of two retroelements, *ZAM* and *Idefix*, in a novel unstable line of *Drosophila melanogaster*. Mol. Biol. Evol. **16:** 54–66.

Devlin, R. H., B. Bingham and B. T. Wakimoto, 1990 The organization and expression of the *light* gene, a heterochromatic gene of *Drosophila melanogaster*. Genetics **125:** 129–140.

Donzeau, M., E. L. Winnacker and M. Meisterernst, 1997 Specific repression of Tax trans-activation by TAR RNA-binding protein TRBP. J. Virol. **71:** 2628–2635.

Dubois, M. F., J. Galabru, P. Lebon, B. Safer and A. G. Hovanessian, 1989 Reduced activity of the interferon-induced double-stranded RNA-dependent protein kinase during a heat shock stress. J. Biol. Chem. **264:** 12165–12171.

Gans, M., C. Audit and M. Masson, 1975 Isolation and characterization of sex-linked female-sterile mutants in *Drosophila melanogaster*. Genetics **81:** 683–704.

Gatignol, A., A. Buckler-White, B. Berkhout and K. T. Jeang, 1991 Characterization of a human TAR RNA-binding protein that activates the HIV-1 LTR. Science **251:** 1597–1600.

Gibson, T. J., and J. D. Thompson, 1994 Detection of dsRNA-binding domains in RNA helicase A and Drosophila maleless: implications for monomeric RNA helicases. Nucleic Acids Res. **22:** 2552–2556.

Heitz, E., 1934 Uber α- and β-heterochromatin sowie konstanz und bau der chromomeren bei *Drosophila*. Biologisches Zentralblatt **54:** 588–609.

Hovanessian, A. G., 1989 The double stranded RNA-activated protein kinase induced by interferon: dsRNA-PK. J. Interferon Res. **9:** 641–647.

Kai, K., H. Sato and T. Odaka, 1986 Relationship between the cellular resistance to Friend murine leukemia virus infection and the expression of murine leukemia virus-gp70-related glycoprotein on cell surface of BALB/c-Fv-4wr mice. Virology **150:** 509–512.

Kim, A., C. Terzian, P. Santamaria, A. Pélisson, N. Prud'homme *et al.*, 1994 Retroviruses in invertebrates: the *gypsy* retrotransposon is apparently an infectious retrovirus of *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA **91:** 1285–1289.

Lee, K. J., M. Mukhopadhyay, P. Pelka, A. R. Campos and H. Steller, 1999 Autoregulation of the Drosophila *disconnected* gene in the developing visual system. Dev. Biol. **214:** 385–398.

Lindsley, D. L., and G. G. Zimm, 1992 *The Genome of Drosophila melanogaster*. Academic Press, London.

Marra, M. A., T. A. Kucaba, N. L. Dietrich, E. D. Green, B. Brownstein *et al.*, 1997 High throughput fingerprint analysis of large-insert clones. Genome Res. **7:** 1072–1084.

Mevel-Ninio, M., M. C. Mariol and M. Gans, 1989 Mobilization of the *gypsy* and *copia* retrotransposons in *Drosophila melanogaster* induces reversion of the *ovoD* dominant female-sterile mutations: molecular analysis of revertant alleles. EMBO J. **8:** 1549–1558.

Miklos, G. L., and J. N. Cotsell, 1990 Chromosome structure at interfaces between major chromatin types: alpha- and beta-heterochromatin. Bioessays **12:** 1–6.

Miklos, G. L., M. T. Yamamoto, J. Davies and V. Pirrotta, 1988 Microcloning reveals a high frequency of repetitive sequences characteristic of chromosome 4 and the beta-heterochromatin of *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA **85:** 2051–2055.

Mitchelson, A., M. Simonelig, C. Williams and K. O'Hare, 1993 Homology with *Saccharomyces cerevisiae* RNA14 suggests that phenotypic suppression in *Drosophila melanogaster* by *suppressor of forked* occurs at the level of RNA stability. Genes Dev. **7:** 241–249.

Myers, E. W., G. G. Sutton, A. L. Delcher, I. M. Dew, D. P. Fasulo *et al.*, 2000 A whole-genome assembly of Drosophila. Science **287:** 2196–2204.

Park, H., M. V. Davies, J. O. Langland, H. W. Chang, Y. S. Nam *et al.*, 1994 TAR RNA-binding protein is an inhibitor of the interferon-induced protein kinase PKR. Proc. Natl. Acad. Sci. USA **91:** 4713–4717.

Pélisson, A., S. U. Song, N. Prud'homme, P. A. Smith, A. Bucheton *et al.*, 1994 *Gypsy* transposition correlates with the production of a retroviral envelope-like protein under the tissue-specific control of the Drosophila *flamenco* gene. EMBO J. **13:** 4401–4411.

Pélisson, A., L. Teysset, F. Chalvet, A. Kim, N. Prud'homme *et al.*, 1997 About the origin of retroviruses and the co-evolution of the *gypsy* retrovirus with the Drosophila *flamenco* host gene. Genetica **100:** 29–37.

Perrimon, N., D. Smouse and G. L. Miklos, 1989 Developmental genetics of loci at the base of the X chromosome of *Drosophila melanogaster*. Genetics **121:** 313–331.

Pimpinelli, S., M. Berloco, L. Fanti, P. Dimitri, S. Bonaccorsi *et al.*, 1995 Transposable elements are stable structural components of *Drosophila melanogaster* heterochromatin. Proc. Natl. Acad. Sci. USA **92:** 3804–3808.

Prud'homme, N., M. Gans, M. Masson, C. Terzian and A. Bucheton, 1995 *Flamenco*, a gene controlling the *gypsy* retrovirus of *Drosophila melanogaster*. Genetics **139:** 697–711.

Rackwitz, H. R., G. Zehetner, A. M. Frischauf and H. Lehrach, 1984 Rapid restriction mapping of DNA cloned in lambda phage vectors. Gene **30:** 195–200.

Robertson, H. M., C. R. Preston, R. W. Phillis, D. M. Johnson-

Schlitz, W. K. Benz *et al.*, 1988 A stable genomic source of P element transposase in *Drosophila melanogaster*. Genetics **118:** 461–470.

Song, S. U., T. Gerasimova, M. Kurkulos, J. D. Boeke and V. G. Corces, 1994 An *env*-like protein encoded by a Drosophila retroelement: evidence that *gypsy* is an infectious retrovirus. Genes Dev. **8:** 2046–2057.

Spradling, A. C., 1986 *P* element-mediated transformation, pp. 175–197 in *Drosophila: A Practical Approach*, edited by D. B. Roberts. IRL Press, Oxford.

St. Johnston, D., N. H. Brown, J. G. Gall and M. Jantsch, 1992 A conserved double-stranded RNA-binding domain. Proc. Natl. Acad. Sci. USA **89:** 10979–10983.

Teysset, L., J. C. Burns, H. Shike, B. L. Sullivan, A. Bucheton *et al.*, 1998 A Moloney murine leukemia virus-based retroviral vector pseudotyped by the insect retroviral *gypsy* envelope can infect Drosophila cells. J. Virol. **72:** 853–856.

Tudor, M., A. Mitchelson and K. O'Hare, 1996 A 1.5 kb repeat sequence flanks the *suppressor of forked* gene at the euchromatin-heterochromatin boundary of the *Drosophila melanogaster* X chromosome. Genet. Res. **68:** 191–202.

Vaury, C., A. Bucheton and A. Pelisson, 1989 The beta-heterochromatic sequences flanking the I elements are themselves defective transposable elements. Chromosoma **98:** 215–224.

Wakimoto, B. T., and M. G. Hearn, 1990 The effects of chromosome rearrangements on the expression of heterochromatic genes in chromosome 2L of *Drosophila melanogaster*. Genetics **125:** 141–154.

Wassarman, D. A., N. M. Solomon and G. M. Rubin, 1994 The *Drosophila melanogaster* ribosomal S6 kinase II-encoding sequence. Gene **144:** 309–310.

Yamamoto, M. T., A. Mitchelson, M. Tudor, K. O'Hare, J. A. Davies *et al.*, 1990 Molecular and cytogenetic analysis of the heterochromatin-euchromatin junction region of the *Drosophila melanogaster X* chromosome using cloned DNA sequences. Genetics **125:** 821–832.

Communicating editor: M. J. Simmons