

# Spatially regulated expression of retrovirus-like transposons during *Drosophila melanogaster* embryogenesis

DALI DING<sup>1</sup> AND HOWARD D. LIPSHITZ<sup>2</sup>

Division of Biology 156–29, California Institute of Technology, Pasadena, California 91125, U.S.A. Phone: 818–395–6446, Fax: 818–564–8709, E-mail: LipshitzH@Starbase1.caltech.edu

(Received 2 May 1994 and in revised form 26 July 1994)

## Summary

Over twenty distinct families of long terminal direct repeat (LTR)-containing retrotransposons have been identified in *Drosophila melanogaster*. While there have been extensive analyses of retrotransposon transcription in cultured cells, there have been few studies of the spatial expression of retrotransposons during normal development. Here we report a detailed analysis of the spatial expression patterns of fifteen families of retrotransposons during *Drosophila melanogaster* embryogenesis (17.6, 297, 412, 1731, 3S18, blood, copia, gypsy, HMS Beagle, Kermit/flea, mdg1, mdg3, opus, roo/B104 and springer). In each case, analyses were carried out in from two to four wild-type strains. Since the chromosomal insertion sites of any particular family of retrotransposons vary widely among wild-type strains, a spatial expression pattern that is conserved among strains is likely to have been generated through interaction of host transcription factors with *cis*-regulatory elements resident in the retrotransposons themselves. All fifteen families of retrotransposons showed conserved patterns of spatially and temporally regulated expression during embryogenesis. These results suggest that all families of retrotransposons carry *cis*-acting elements that control their spatial and temporal expression patterns. Thus, transposition of a retrotransposon into or near a particular host gene – possibly followed by an excision event leaving behind the retrotransposon's *cis*-regulatory sequences – might impose novel developmental control on such a host gene. Such a mechanism would serve to confer evolutionarily significant alterations in the spatio-temporal control of gene expression.

## 1. Introduction

Retrotransposons are a widely distributed group of eukaryotic mobile genetic elements that transpose via RNA intermediates. Among eukaryotes, the greatest number and diversity of such elements have been described in *Drosophila*. Some ten percent of the *Drosophila melanogaster* genome consists of moderately repeated transposable elements (Finnegan & Fawcett, 1986). There are about fifty families of such elements, and they are divided into different categories according to their structure and mechanism of transposition (Finnegan & Fawcett, 1986). The largest group is comprised of the *copia*-like elements, or LTR-containing retrotransposons, over twenty of which have been described to date. These resemble retroviral proviruses and possess long terminal direct repeats (LTRs) and open reading frames that encode

in particular a polypeptide with sequence similarity to retroviral reverse transcriptase. These elements have been demonstrated to transpose by a mechanism similar to that found in the retroviral life cycle: they are transcribed into RNAs, these RNAs are reverse transcribed into DNA copies, and these DNAs are capable of integration into different chromosomal sites (Heidmann & Heidmann, 1991). Transcription of a retrotransposon provides not only the necessary RNA template for the DNA intermediate, but also the mRNAs that encode the proteins required for transposition.

There has been extensive analysis of the mechanism and control of retrotransposon transcription. Most of these studies have been carried out using cultured cell lines (Arkhipova & Ilyin, 1992, for a review). Developmental and hormone-dependent regulation of retrotransposon transcription have been reported for elements from *Drosophila* (Flavell *et al.*, 1980; Micard *et al.*, 1988; Parkhurst & Corces, 1987; Peronnet *et al.*, 1986; Scherer *et al.*, 1981; Schwartz, Lockett &

<sup>1</sup> Present address: Amgen Inc., 1840 DeHavilland Drive Thousand Oaks, CA 91320–1789.

<sup>2</sup> Author for correspondence.

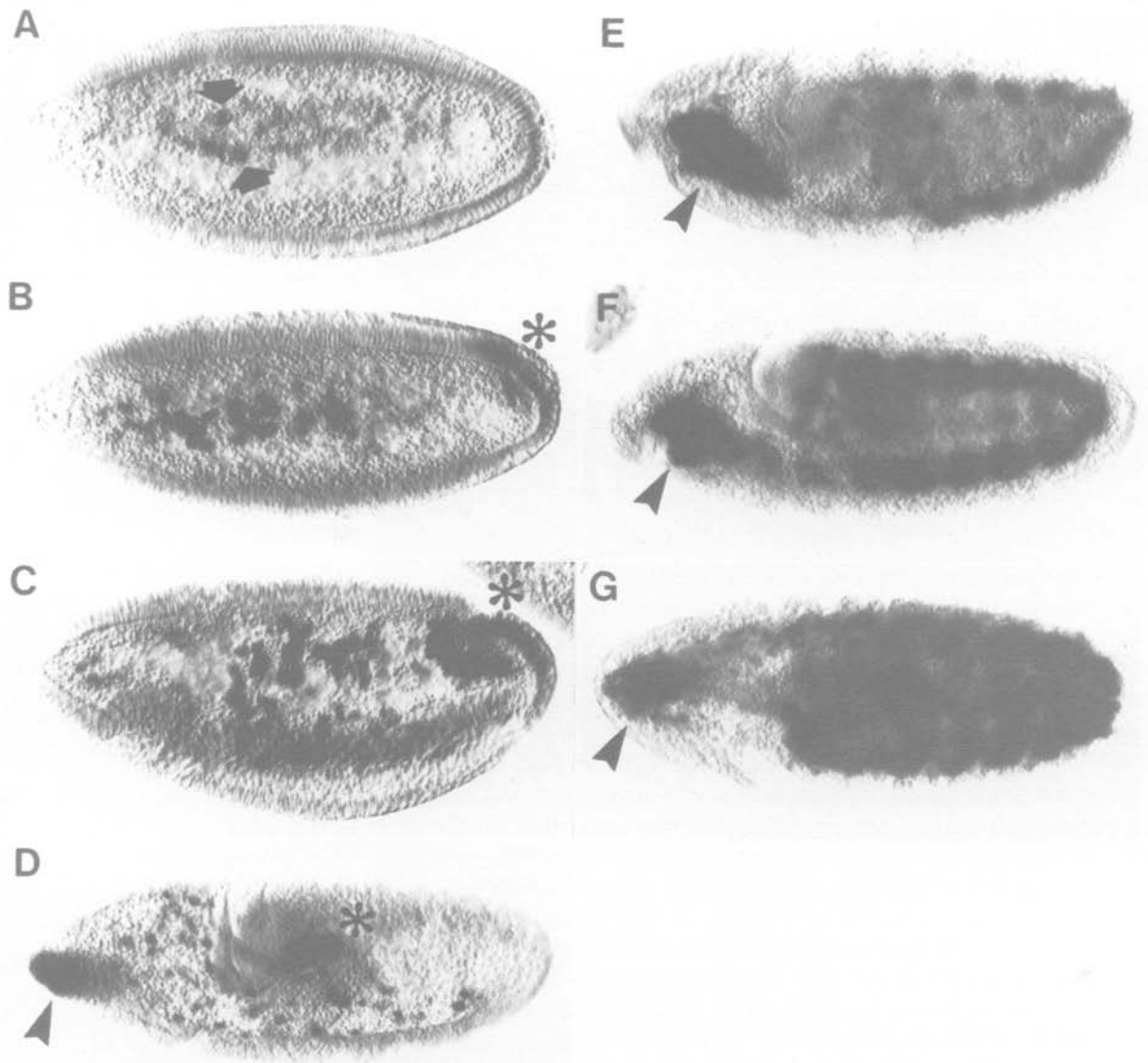


Fig. 1. Spatial expression of *B104* during embryogenesis. (A) Cellular blastoderm stage. Accumulation occurs in the central yolk nuclei (two such nuclei are highlighted with arrows). (B, C) Early gastrulation stages. Accumulation continues in the yolk nuclei and begins in the posterior midgut primordium (asterisk). (D) Late germ band extension stage. Accumulation continues in the yolk nuclei and the posterior midgut primordium (asterisk), and commences at the antero-ventral tip of the embryo (arrowhead). (E–G) Fully extended germ band stages. Accumulation occurs in the somatic mesoderm in a segmentally repeated pattern, and continues in the stomodeum (arrowheads). Anterior is to the left and dorsal up, except in (G), which is viewed from the dorsal side.

Young, 1982; Ting *et al.*, 1992; Young & Schwartz, 1981), mouse (Keshet, Schiff & Itin, 1991, for a review; Ting *et al.*, 1992), plants (Pouteau *et al.*, 1991; Radicella *et al.*, 1992) and yeast (Errede, 1993; Löhning, Rosenbaum & Ciriacy, 1993; Yu & Fassler, 1993). Although tissue-specific transcription of *Drosophila* retrotransposons has appeared likely for some time (Meyerowitz & Hogness, 1982), few such cases have been documented (Brookman *et al.*, 1992; Huijser *et al.*, 1988; Lanckenau, Corces & Lanckenau, 1994; Mozer & Benzer, 1994).

Here we examine the expression patterns of fifteen families of retrotransposons during *Drosophila* embryogenesis. We report dynamic temporal and spatial regulation of transcription of each of these families of elements and demonstrate that each family of elements

exhibits a pattern of expression that is conserved among strains with widely differing sites of insertion. These results suggest that each family of elements carries internal *cis*-regulatory sequences that control its expression pattern. We discuss the biological significance of our findings with particular emphasis on their possible implications for the evolution of developmental gene regulation.

## 2. Materials and methods

### *Fly strains*

Wild-type strains used in this study were Oregon-R (O-R), Canton-S (C-S), Urbana-S (U-S) and Lausanne-S (L-S).

*RNA tissue in situ hybridization to whole-mount embryos and ovaries*

Whole mount RNA tissue in situ hybridization was based on the method of Tautz and Pfeifle (Tautz & Pfeifle, 1989) with the modifications described in (Ding, Parkhurst & Lipshitz, 1993).

*Probes for in situ hybridization*

Probes were synthesized from purified restriction fragments or from the products of enzymatic amplification, as follows. *B104*: a 4.4 kb HindIII-XhoI fragment (Scherer *et al.*, 1982). *3S18*: a 4 kb BamHI fragment (Bell *et al.*, 1985). *412*: four HindIII-EcoRI fragments with a total length of 4.4 kb, from nucleotides 1621 through 5984 (Yuki *et al.*, 1986). *mdg1*: a 2.8 kb EcoRI fragment from plasmid Dm 225 (Ilyin *et al.*, 1980). *flea*: a 4.9 kb PvuI fragment (Kidd & Young, 1986). *Kermite*: a 4 kb XbaI fragment (Bender, Spierer & Hogness, 1983). *297*: a 3.5 kb XbaI-XhoI fragment from plasmid cDM4006 (Mossie, Young & Varmus, 1985). *1731*: a 3 kb BamHI-SalI fragment from plasmid pFP5c (Peronnet *et al.*, 1986). *mdg3*: A 5.2 kb HindIII fragment (Ilyin *et al.*, 1980). *springer*: 1.1 kb and 3.5 kb EcoRI-SalI fragments (Karlik & Fyrberg, 1985). *copia*: a 2.6 kb fragment from the internal HindIII site to the end of the 3'-LTR (Mount & Rubin, 1985). *HMS Beagle*: a 3.9 kb SalI-XhoI fragment (Snyder *et al.*, 1982). *opus*: a mixture of the 1.6 kb and 3.2 kb SalI-EcoRI fragments (Kidd & Young, 1986). *gypsy*: a 4.5 kb PvuII-XhoI fragment (Marlor, Parkhurst & Corces, 1986). *blood*: two primers, oligo1 and oligo2, were designed based on the LTR sequence (Bingham & Chapman, 1986; Lavorgna *et al.*, 1989). Oligo1: 5'-GGT ACA CTG TAC CTA TAA GTA-3', and Oligo 2: 5'-TGC CAC CAA TTG AGA ACA CG-3'. PCR amplification of O-R genomic DNA with the above primers produced a single DNA fragment of the expected length (361 bp).

**3. Results**

To distinguish between expression of a retrotransposon under the control of its own *cis*-regulatory elements *versus* control of expression exerted by host genomic regulatory sequences located near the insertion site, we examined the embryonic expression pattern of each family of retrotransposons in two to four wild-type strains (see Materials and Methods). Since these strains differ markedly in their transposon insertion sites (Ilyin *et al.*, 1978; Strobel, Dunsmuir & Rubin, 1979), the probability of a spurious result caused by a host-site position effect is low. For 14 of the 15 families of elements studied here, the patterns were consistent among these different strains (the exception was *opus*). Thus, the expression patterns described below are representative of those found in

all strains examined, and are likely to be controlled through sequences within the retrotransposons themselves. It should be noted that the expression patterns presented here for embryogenesis are complete: that is, each family of elements is expressed at high levels during embryogenesis only at the times or places described. Our data cannot, however, exclude the possibility that expression occurs at a low level in other tissues and/or at other stages of embryogenesis.

*B104*

*B104* expression is first detected at the cellular blastoderm stage in the central yolk nuclei (Fig. 1A, arrows). At the onset of germ band extension, intense accumulation of *B104* transcripts is also detected in the posterior midgut primordium and in the yolk nuclei immediately beneath these cells (Figs. 1B and C, asterisk). During germ band extension, transcripts additionally accumulate in cells at the antero-ventral tip of the embryo (Fig. 1D, arrowhead). By the fully extended germ band stage, staining in the yolk nuclei has decreased markedly (Fig. 1E). By this stage,

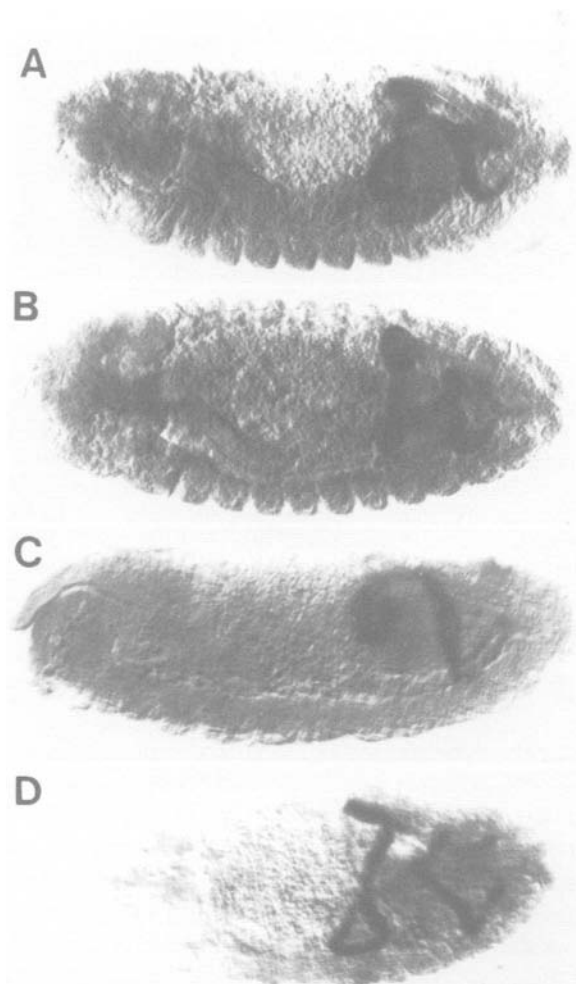


Fig. 2. Spatial expression of *3S18* during embryogenesis. (A–D) *3S18* transcripts accumulate in the malpighian tubules. Anterior is to the left. Lateral (A, C) and dorsal (B, D) views are shown.

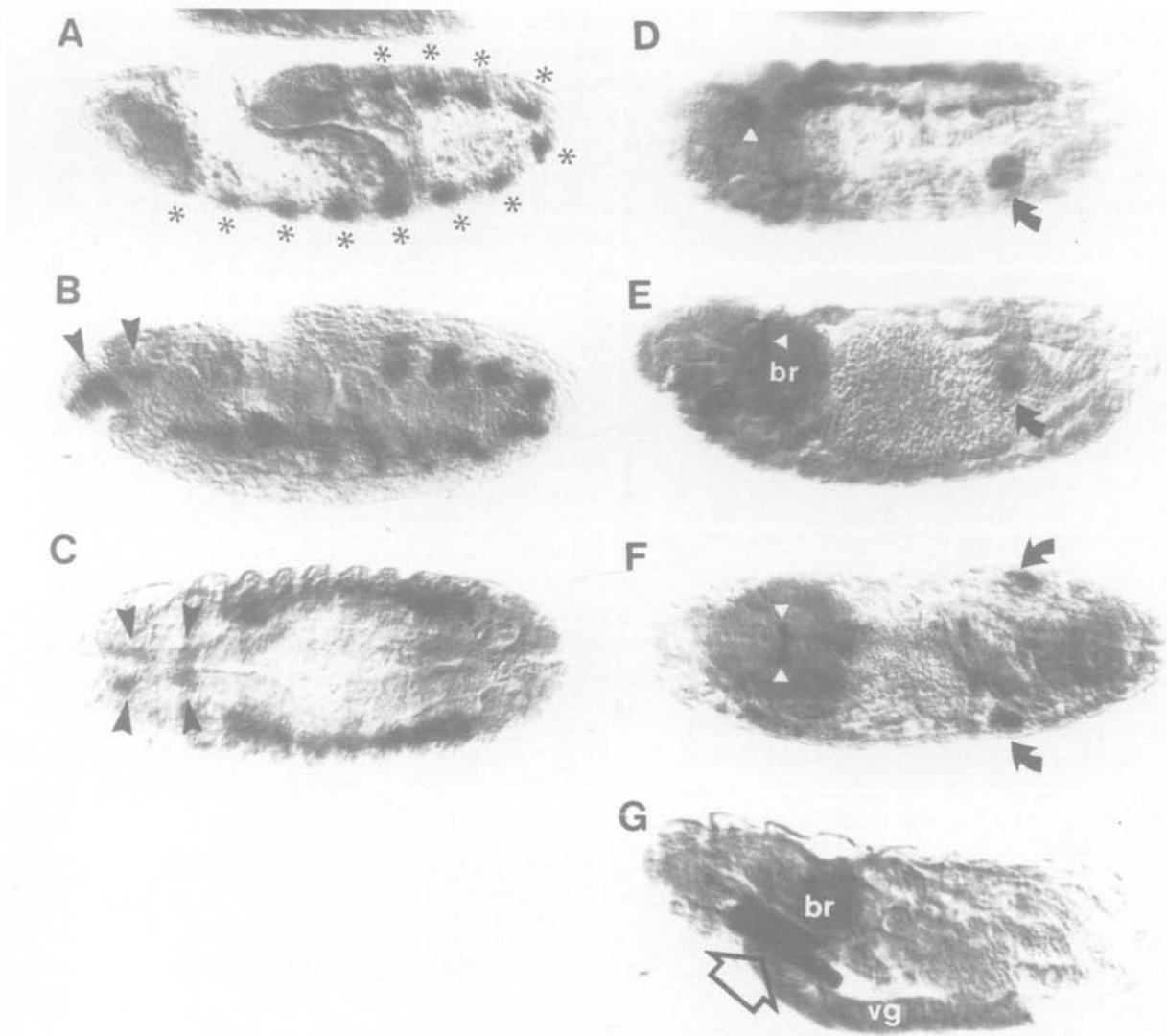


Fig. 3. Spatial expression of *17.6* during embryogenesis. (A) Germ band extension. Accumulation begins in 12 pairs of segmentally repeated patches in the mesoderm on either side of the ventral midline (asterisks). (B) Fully extended germ band stage. Transcripts continue to be detectable in the mesoderm, and accumulation commences in two pairs of bilaterally symmetric groups of cells in the dorsal head region (black arrowheads). (C) Germ band retracted stage. Transcripts are detectable in the somatic mesoderm on both sides of the embryo, and continues in the clusters of cells in the head (black arrowheads). (D) Dorsal closure stage. Intense transcript accumulation occurs in a group of cells in the brain (white arrowhead) and accumulation can also be detected in the gonad (curved arrow), as well as in the cells at the edge of the closing dorsal epidermis. (E, F) Intense staining persists in the brain, particularly the bilaterally symmetric clusters of cells (white arrowheads), and the gonadal staining continues throughout the remainder of embryogenesis (curved arrows). (G) A fully developed embryo. Transcripts can be detected throughout the CNS (br: brain; vg: ventral ganglion). At this stage, there is also intense staining in salivary glands (open arrow). Anterior is to the left. Dorsal is up in (A), (B), (D), (E) and (G). (C) and (F) are views from the dorsal side.

transcript accumulation in the antero-ventral cells has increased and these cells have been brought internally with the stomodeal invagination (Fig. 1E). Novel expression can also be detected at this stage in the somatic mesoderm in a segmentally repeated pattern (Fig. 1E). As development proceeds, this mesodermal accumulation pattern becomes more intense (Fig. 1F) and the mesodermal staining extends laterally to form rings around the embryo (Figs. 1F and G). After germ band retraction, both the intensity of the staining and the number of the cells expressing *B104* are reduced dramatically (data not shown). In the fully developed

embryo, the only detectable accumulation of *B104* transcripts occurs at a low level in the central nervous system (CNS) (data not shown).

### *3S18*

The only detectable *3S18* transcript accumulation during embryogenesis occurs in the malpighian tubules (Fig. 2).

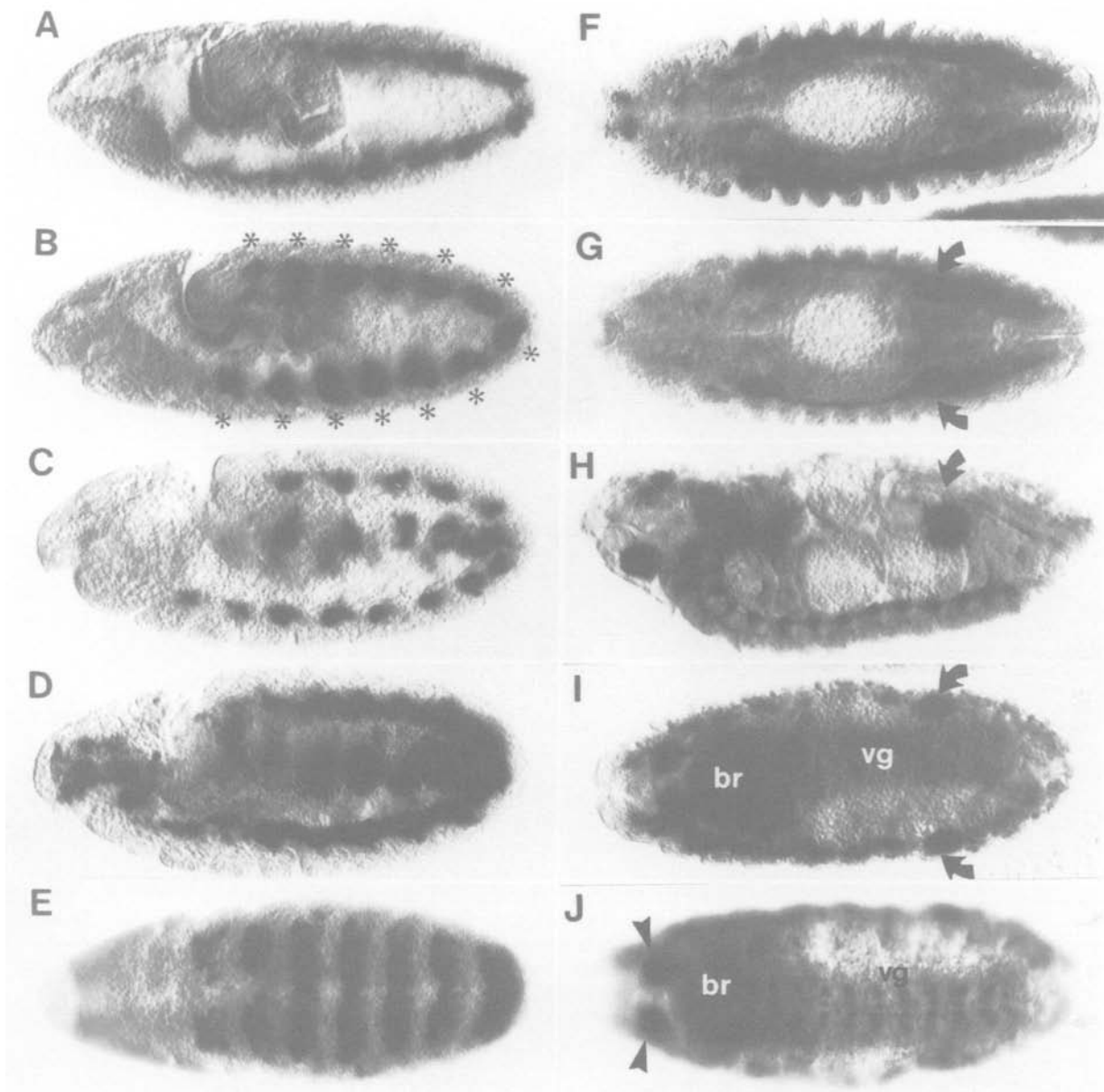


Fig. 4. Spatial expression of *412* during embryogenesis. (A–E) Germ band extension stages. Accumulation occurs in 13 pairs of segmentally-repeated clusters in the mesoderm on either side of the ventral midline (asterisks in B). These patches gradually extend toward the ventral midline and eventually fuse (E). Two additional bilaterally symmetric patches of staining appear in the head region at the fully extended germ band stage (D). (F, G) Following germ band retraction, accumulation is seen throughout the visceral mesoderm. In addition, the gonads are heavily stained (curved arrows in G). (H–J) Late in embryogenesis, the mesodermal accumulation decreases while the head (arrowheads in J) and gonadal (curved arrows in H and I) accumulation intensify. Accumulation can also be detected in segmentally repeated clusters in the CNS (br: brain; vg: ventral ganglion). (I) and (J) show the same embryo in different focal planes. Anterior is to the left and dorsal up, except in (E), (F), (G), (I) and (J), which are ventral views.

### 17.6

*17.6* transcripts first accumulate during germ band extension in 12 pairs of segmentally repeated patches in the mesoderm on either side of the ventral midline (Fig. 3A, asterisks). Shortly thereafter, transcripts accumulate in two pairs of bilaterally symmetric groups of cells in the dorsal head region (Figs. 3B and C, arrowheads). After germ band retraction, transcripts are still present in the somatic mesoderm on both sides of the embryo (Fig. 3C). During dorsal

closure, staining of the head persists and accumulation can also be detected in the gonad (Fig. 3D, curved arrow), as well as in the cells at the edge of the closing dorsal epidermis (Fig. 3D). The gonadal staining persists throughout the remainder of embryogenesis (Figs. 3E and F, curved arrows). In the fully developed embryo, bilaterally symmetric clusters of cells within the brain accumulate high levels of *17.6* transcripts (Fig. 3F, white arrowheads). These may be derived from the same cells that stain in the head region earlier in embryogenesis (Figs. 3D and E, white arrowheads).

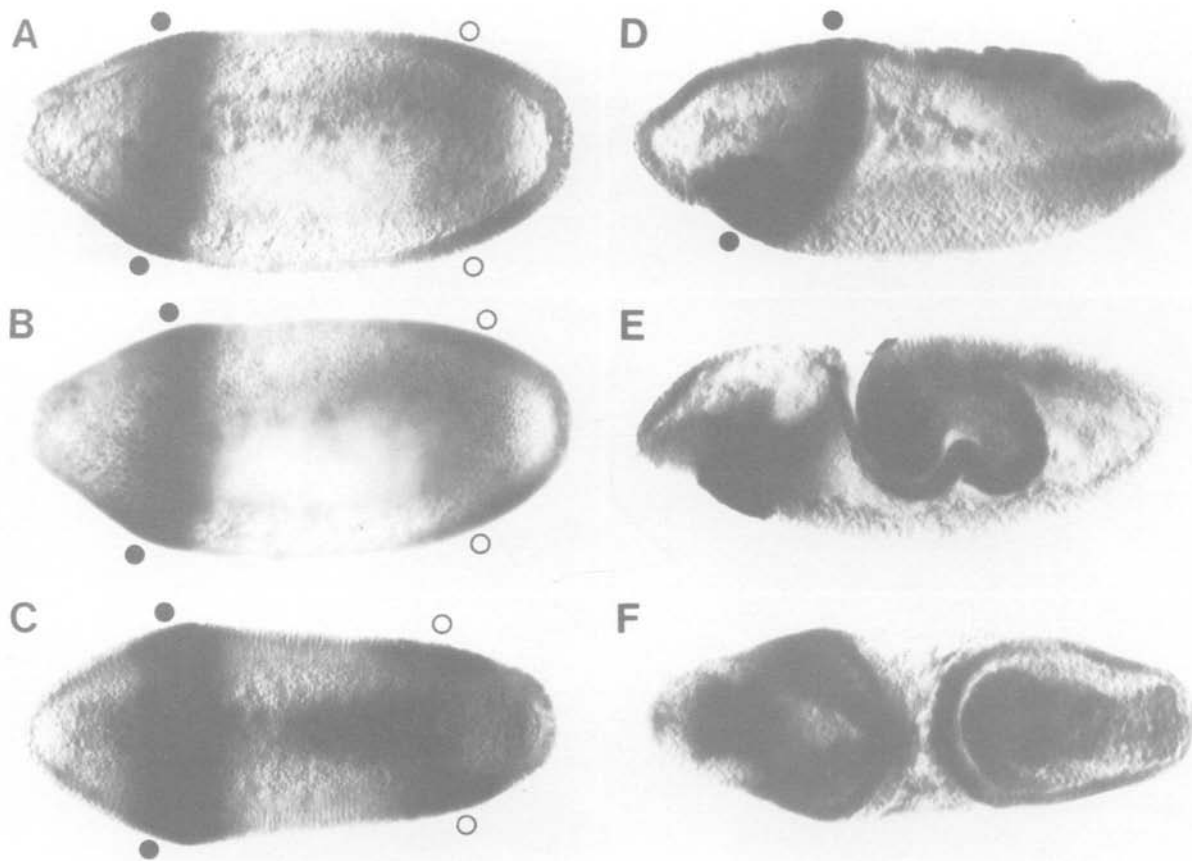


Fig. 5. Spatial expression of *Kermit/flea* during embryogenesis. (A, B) Syncytial blastoderm stage. Accumulation is first detected in a stripe just anterior to where the cephalic furrow will form at about 65% egg length (filled dots). The stripe is wider on the ventral side (15 cells on average) than the dorsal side (4 cells on average). The staining shows a punctate pattern, indicating that the RNA is confined largely to the nuclei. In addition, accumulation occurs in a uniform stripe in the nuclei from 15–30% egg length (open dots). A and B show the same embryo in different focal planes. (C) Cellular blastoderm stage. The anterior (filled dots) and posterior (open dots) stripes persist, and accumulation begins in cells along the dorsal midline from 15–50% egg length. (D–F) Germ band extension stages. The dorsal accumulation persists through germ band extension. The anterior staining intensifies and becomes restricted to the ventral side. Anterior is to the left and dorsal up, except in (C) and (F), which are views from the dorsal side.

Lower levels of *17.6* transcripts can be detected throughout the rest of the CNS (Fig. 3G, brain [br] and ventral ganglion [vg]). At this stage, there is also intense staining in salivary glands (Fig. 3G, open arrow).

#### *412*

*412* RNA accumulation can first be detected during germ band extension in 13 pairs of segmentally repeated clusters in the mesoderm on either side of the ventral midline (Figs. 4 A–C, indicated by asterisks in B). These patches gradually extend toward the ventral midline and eventually fuse (Fig. 4E). Two additional bilaterally symmetric patches of staining appear in the head region at the germ band extended stage (Fig. 4D). Following germ band retraction, accumulation is seen throughout the visceral mesoderm (Figs. 4F and G). In addition, the gonads are heavily stained (Figs. 4G, H [curved arrows] and I). Towards the end of embryogenesis, the mesodermal accumulation decreases while the head and gonadal expression

intensifies (Figs. 4H–J). In addition, novel *412* accumulation is detected in segmentally repeated cell clusters in the CNS (Figs. 4H–J, I and J show the same embryo in different focal planes). A similar pattern of *412* transcripts has been reported previously (Brookman *et al.*, 1992), although those authors did not report accumulation of *412* transcripts in the CNS of the late embryo.

#### *mdg1*

The expression pattern of *mdg1* is virtually identical to that of the *412* element (data not shown). This result is consistent with the fact that these two elements are closely related to each other based on comparative sequence analysis (see Discussion) (Will, Bayev & Finnegan, 1981; Yuki *et al.*, 1986).

#### *Kermit/flea*

Southern blot hybridization analysis indicates that *Kermit* and *flea* cross-hybridize under high-stringency

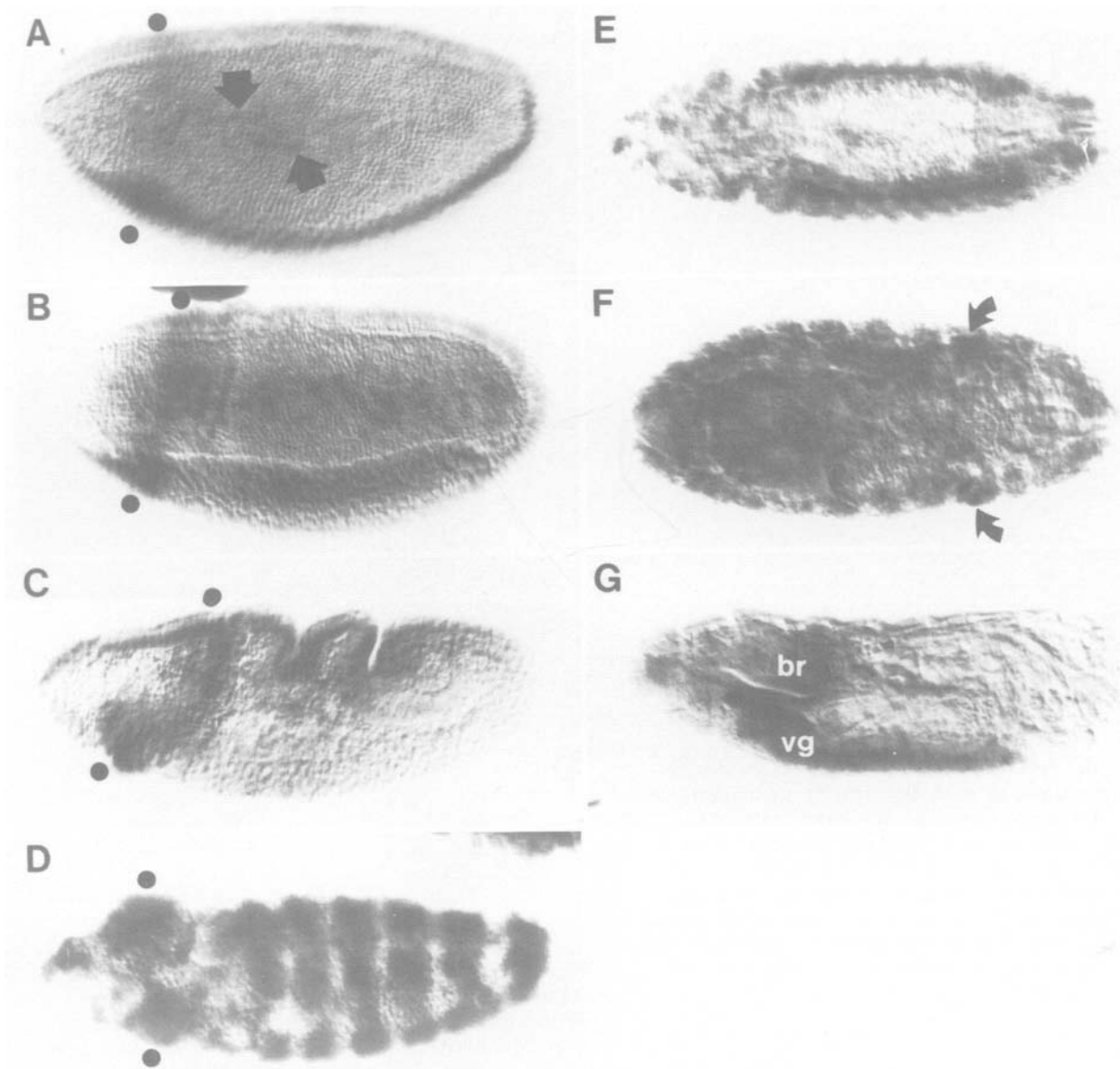


Fig. 6. Spatial expression of 297 during embryogenesis. (A) Cellular blastoderm stage. Accumulation of 297 transcripts can be detected in an anterior stripe (dots), which is most intense and widest on the ventral side, and in yolk nuclei (arrows). (B–D) Germ band extension stages. The anterior stripe persists (dots). Late in germ band extension, transcripts accumulate in a series of segmentally repeated patches (D). (E, F) Germ band retracted embryos. Transcripts can be seen in the visceral mesoderm and in the gonads (curved arrows in F). (G) In fully developed embryos, high levels of 297 transcripts can be detected in the CNS (br: brain; vg: ventral ganglion). Anterior is to the left and dorsal up, except in (E) and (F), which are ventral views.

conditions, suggesting that their sequences are either identical or very closely related (data not shown). *Kermit/flea* transcript accumulation is first detected at the syncytial blastoderm stage in a stripe just anterior to where the cephalic furrow will form at about 65% egg length (Figs. 5A–D, filled dots). The staining shows a punctate pattern, indicating that the RNA is confined largely to the nuclei (most obvious in Fig. 5B). The stripe is wider on the ventral side (15 cells on average) than the dorsal side (4 cells on average). In addition, two regions of more posterior staining emerge: a uniform stripe in the nuclei from 15–30% egg length (Figs. 5A–C, open dots) and a region of staining in cells along the dorsal midline from 15–50%

egg length (Fig. 5C). This dorsal accumulation persists through germ band extension (Fig. 5D–F). By the end of germ band extension (Figs. 5E and F), the anterior staining intensifies and becomes restricted to the ventral side. Commencing at the germ band retraction stage, *Kermit/flea* transcripts accumulate ubiquitously throughout the embryo, gradually diminishing in concentration (data not shown).

#### 297

Accumulation of 297 transcripts can initially be detected at the cellular blastoderm stage in an anterior stripe (Fig. 6A, dots) as well as in yolk nuclei (Fig. 6A,

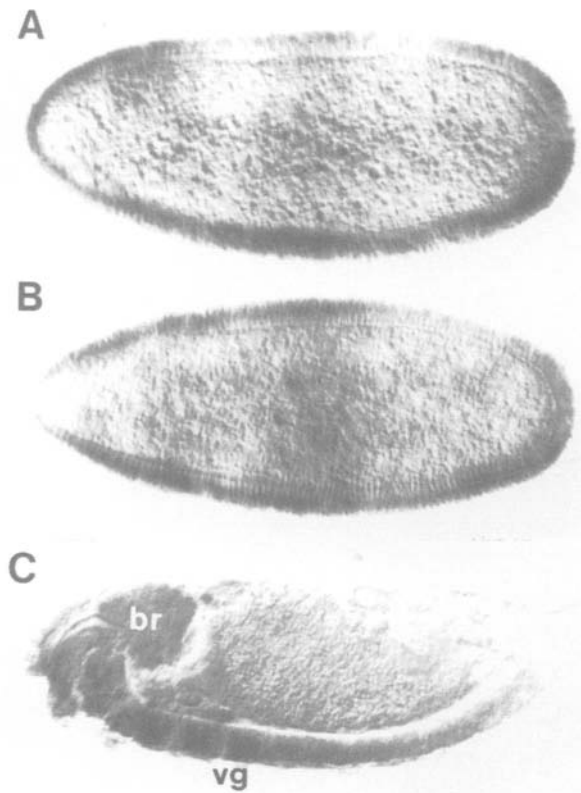


Fig. 7. Spatial expression of *1731* during embryogenesis. (A, B) Late cellular blastoderm stage. Accumulation occurs in an ubiquitous but non-uniformly intense manner. (C) After the germ band has retracted, transcripts can be seen in the CNS in a graded distribution along the anterior-posterior axis, with the highest levels in the brain and anterior part of the ventral ganglion (br: brain; vg: ventral ganglion). Anterior is to the left and dorsal up, except in (B), which is a ventral view.

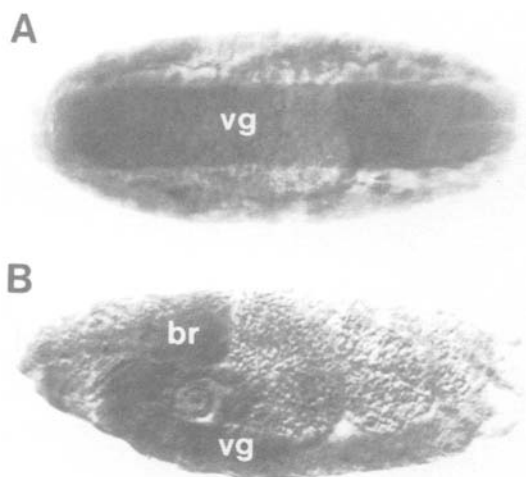


Fig. 8. Spatial expression of *mdg3* during embryogenesis. (A, B) High levels of transcripts can be detected in the CNS after germ band retraction (br: brain; vg: ventral ganglion). Anterior is to the left. (A) is a ventral view. (B) is a lateral view, with dorsal up.

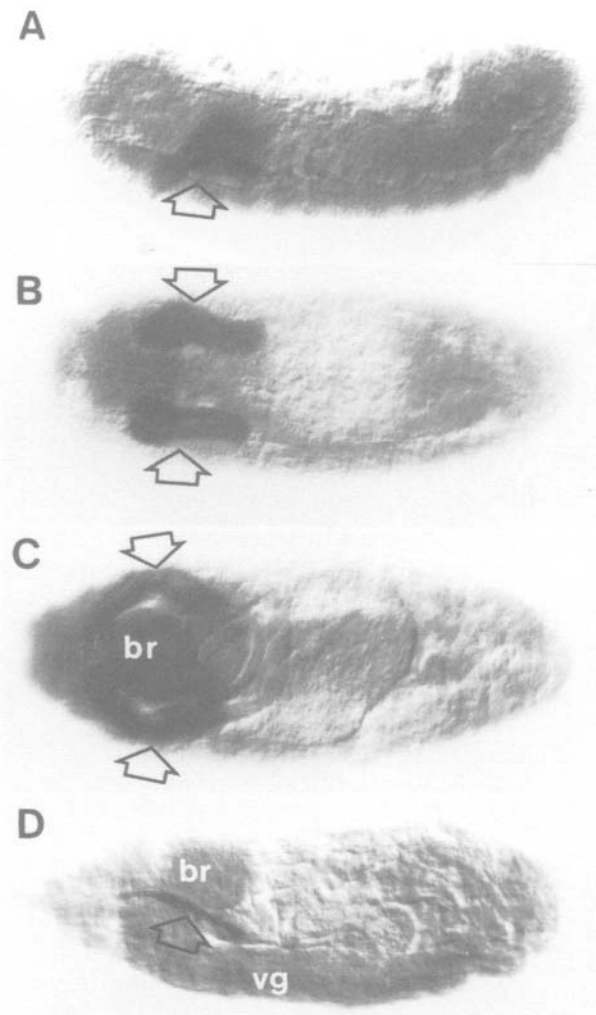


Fig. 9. Spatial expression of *springer* during embryogenesis. (A) Germ band retracting embryo. Transcripts accumulate at high levels in the salivary glands (unfilled arrow). (B–D) Post-germ band retraction stages. Salivary gland staining persists (unfilled arrows). Transcripts also accumulate, although to a lower level, in the CNS (C and D) (br: brain; vg: ventral ganglion).

arrows). The stripe is most intense and widest on the ventral side, and persists through germ band extension (Fig. 6B–D, dots). In late germ band extension, *297* transcripts accumulate in a series of segmentally repeated patches (Fig. 6D). After germ band retraction, transcripts can be seen in the visceral mesoderm (Fig. 6E) and gonads (Fig. 6F, curved arrows). In fully developed embryos, high levels of *297* transcripts can be detected in the CNS (Fig. 6G, brain [br] and ventral ganglion [vg]).

#### *1731*

Accumulation of *1731* RNA is first detected in the late cellular blastoderm stage in an ubiquitous but non-uniformly intense manner (Figs. 7A and B). This pattern disappears gradually during germ band extension. After the germ band has retracted, *1731* transcripts can be seen in the CNS in a graded



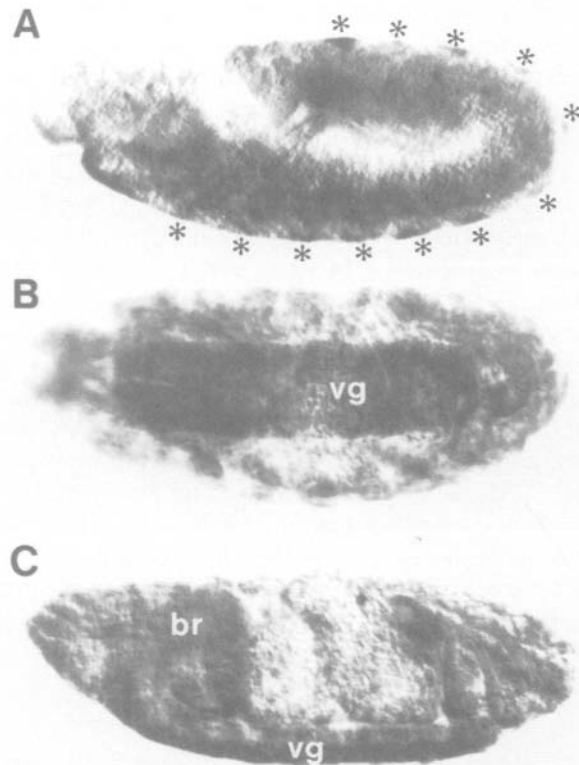


Fig. 10. Spatial expression of *copia* during embryogenesis. (A) Germ band extended embryo. Accumulation of transcripts is detected in segmentally repeated patches of cells along the ventral midline (asterisks). (B, C) After germ band retraction, transcripts accumulate at high levels in the CNS (br: brain; vg: ventral ganglion). Anterior is to the left and dorsal up except in (B), which is viewed from the ventral side.

distribution along the anterior posterior axis, with the highest levels in the brain and anterior ventral ganglion (Fig. 7C).

#### *mdg3*

High levels of *mdg3* transcripts can be detected in the CNS after germ band retraction (Figs. 8A and B).

#### *springer*

High levels of *springer* transcripts initially accumulate in the salivary glands starting during germ band retraction and persisting through the rest of embryogenesis (Fig. 9A–C, unfilled arrows). Transcripts also accumulate, although to a lower level, in the CNS (Figs. 9C and D; brain [br] and ventral ganglion [vg]).

#### *copia*

Accumulation of *copia* transcripts is first detected in segmentally repeated patches of cells along the ventral

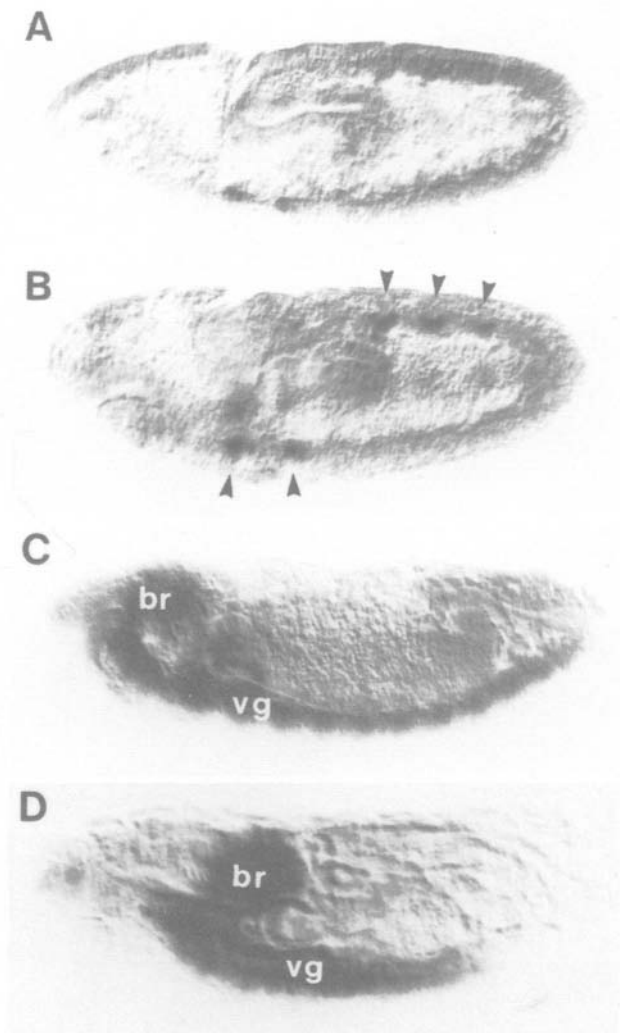


Fig. 11. Spatial expression of *HMS Beagle* during embryogenesis. (A) Germ band extension. Transcripts are first detectable in the neuroblasts. (B) Fully germ band extended embryo. Transcripts accumulate at high levels in five pairs of bilaterally symmetric clusters of cells on the ventral side of the embryo (arrowheads). (C, D) After germ band retraction, transcripts accumulate in the CNS (br: brain; vg: ventral ganglion). Anterior is to the left and dorsal up.

midline during germ band extension (Fig. 10A, asterisks). After germ band retraction, *copia* accumulates at high levels in the CNS (Figs. 10B and C; brain [br] and ventral ganglion [vg]).

#### *HMS Beagle*

There is no detectable accumulation of *HMS Beagle* RNA prior to germ band extension. During germ band extension, *HMS Beagle* transcripts are first detectable in the neuroblasts (Fig. 11A), and they subsequently accumulate at high levels in five pairs of bilateral clusters of cells on the ventral side of the embryo (Fig. 11B, arrowheads). After germ band retraction, *HMS Beagle* accumulates in the CNS

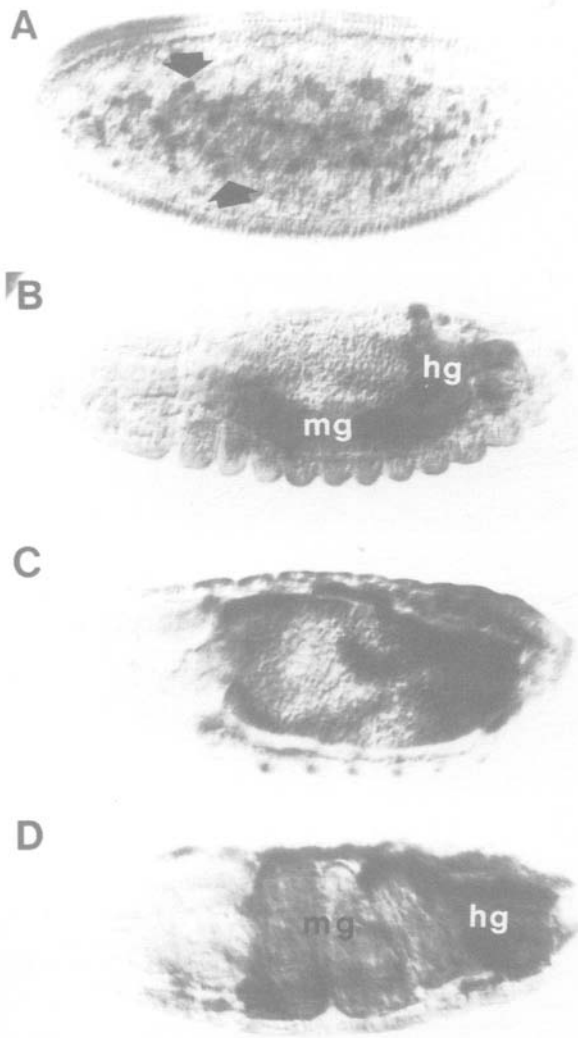


Fig. 12. Spatial expression of *blood* during embryogenesis. (A) Cellular blastoderm stage. RNA can be detected in the yolk nuclei (arrows). (C–D) Post-germ band retraction stages. Transcripts accumulate in the midgut (mg) and hindgut (hg). Anterior is to the left and dorsal up.

(Figs. 11C and D; brain [br] and ventral ganglion [vg]).

#### *blood*

At the cellular blastoderm stage, *blood* RNA can be detected in the yolk nuclei (Fig. 12A, arrows). The only other embryonic accumulation of *blood* transcripts occurs after germ band retraction in the midgut and hindgut (Figs. 12B–D; midgut [mg] and hindgut [hg]).

#### *opus*

The accumulation pattern of *opus* in L-S embryos (Figs. 13A–G) differs markedly from that in U-S embryos (Figs. 13H–N). In both strains, *opus* is found in the yolk nuclei during early embryogenesis (Figs. 13A, H, I; filled arrows) and at high levels in the CNS

from germ band retraction on (Figs. 13D, F, G, K, M and N; brain [br] and ventral ganglion [vg]). The conservation of these aspects of the pattern suggests that they are regulated by sequences within the *opus* element itself. In addition, in L-S embryos, *opus* RNA can also be detected in seven stripes during the cellular blastoderm and early gastrulation stages (Figs. 13A and B, arrowheads). During germ band extension, this pattern evolves into a 14-stripe pattern characteristic of some of the segment polarity genes (e.g. *engrailed* (Fjose, McGinnis & Gehring, 1985; Kornberg *et al.*, 1985)) (Fig. 13C, asterisks). In contrast, U-S embryos exhibit no striped pattern of accumulation; instead, *opus* transcripts accumulate at high levels in the salivary glands from germ band retraction onwards (Figs. 13K–N, unfilled arrows). The strain-specific expression patterns of *opus* are likely to be caused by transcriptional regulatory elements located nearby one or more of the *opus* insertion sites, which are specific to each strain.

#### *gypsy*

Gypsy transcripts first accumulate during germ band retraction in a small group of the outermost ventral cells in the stomodeal invagination (Fig. 14A) as well as in the yolk nuclei (Fig. 14A and B; filled arrows). In addition to persisting in the stomodeal derivatives, *gypsy* transcripts can be detected after dorsal closure in the visceral mesoderm (Fig. 14C) and the gonads (Fig. 14C, curved arrows). In a fully developed embryo, the only detectable transcript accumulation is in the pharyngeal opening and esophagus (Fig. 14D, white arrowheads).

## 4. Discussion

*Conserved transcript accumulation patterns suggest that retrotransposons carry cis-regulatory sequences that interact with host transcription factors*

The patterns of transcript accumulation for each of the fifteen families of retrotransposons investigated here are almost perfectly conserved among the different wild-type *Drosophila melanogaster* strains examined. Since the chromosomal insertion sites for each family of elements are highly polymorphic among different strains (Strobel, Dunsmuir & Rubin, 1979), the conserved transcript accumulation pattern for each family of retrotransposons is likely to be a consequence of *cis*-regulation through sequences contained within the elements themselves. Only in the case of the *opus* element did we observe marked position effects likely to have been caused by the chromosomal sequences adjacent to the insertion site of one or more of the elements resident in each strain. Significantly, however, these position effects were overlaid on a conserved pattern of transcript accumulation that is likely to have been mediated

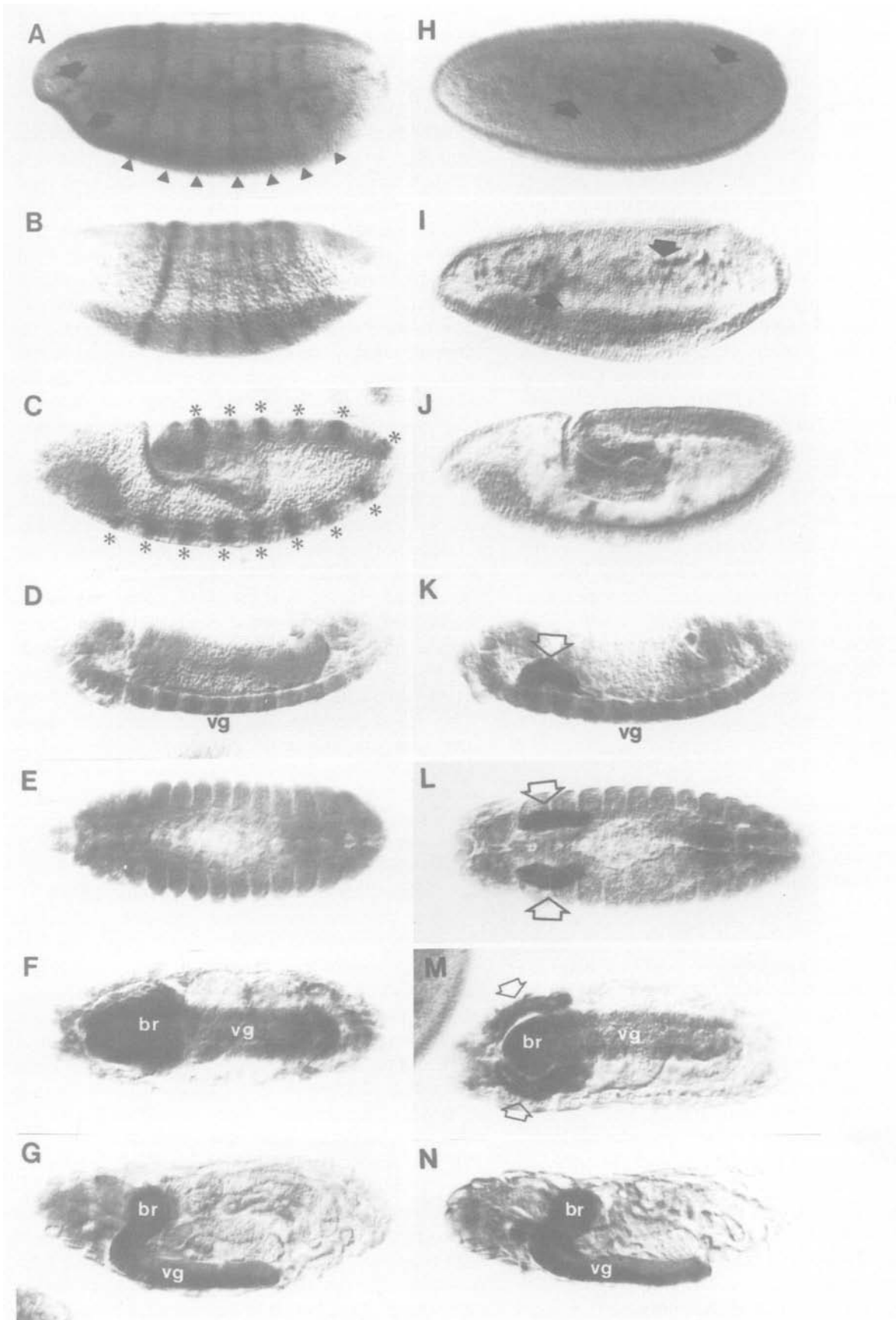


Fig. 13. Spatial expression of *opus* during embryogenesis. (A–G) Accumulation in the Lausanne-S strain; (H–N) accumulation in the Urbana-S strain. (A, B, C, H, I, J) Cellular blastoderm/germ band extension stages. Conserved accumulation occurs in the yolk nuclei (filled arrows). In addition, in L-S embryos (A, B), RNA can be detected in seven stripes during these stages (arrowheads). This pattern in L-S evolves into a 14-striped pattern during germ band extension (asterisks). (D, K) Germ band retraction stages. Conserved transcript accumulation occurs at high levels in the

through sequences within the *opus* element itself. Additional support for the idea that retrotransposons carry their own regulatory sequences comes from our identification of indistinguishable patterns of accumulation of transcripts encoded by the *412* and *mdg1* retrotransposons. These share significant sequence similarity, particularly within their LTRs and at the junction of their 5'-LTRs and their internal regions, suggesting that these sequence identities might mediate the conserved expression patterns.

To identify the specific internal sequences that are responsible for the retrotransposon expression patterns, detailed transgenic analyses will be required. For one of these elements, the *412* retrotransposon, such evidence has recently been provided. *412* has been shown to carry sequences within the LTR that are bound by the *ULTRABITHORAX (UBX)* homeodomain protein and capable of conferring appropriate spatial expression upon reporter sequences in transgenes (P. Feinstein and R. Mann, personal communication). Thus, expression of *412* in the visceral mesoderm of parasegments 6 through 14 (Fig. 4) is likely to be a direct consequence of transcriptional activation by the *UBX* transcription factor. In addition, it has recently been shown that *17.6* expression occurs specifically in the lamina anlage of the developing adult visual system, and is controlled by the combined action of an enhancer in the 5'-LTR and a repressor region in the body of the retrotransposon (Mozer & Benzer, 1994). It is not yet known what *trans*-acting factors interact with these regions.

*Comparisons of conserved transcript patterns suggest that some cis-regulatory sequences are shared by more than one family of retrotransposons while other regulatory sequences may be unique to a particular family*

Although developmental regulation of transcription of retrotransposons has been well documented, the extent of temporal and spatial regulation revealed by this study is striking—all fifteen families of retrotransposons examined here show dynamic spatio-temporal expression patterns during embryogenesis. Expression in several tissues appears to be conserved among a number of the retrotransposon families. For example, *B104* and *17.6* are expressed in the somatic mesoderm; *412*, *297* and *gypsy* in the visceral mesoderm; *17.6*, *412*, *mdg1*, *297* and *gypsy* in the gonads (likely the mesodermally derived cells rather than the germline (Brookman *et al.*, 1992)); *297*, *1731*, *mdg3*, *copia*, *HMS Beagle* and *opus* in the central nervous system late in embryogenesis; and *B104*, *blood*, *gypsy*, *297* and *opus* in yolk nuclei of early embryos. It will be

interesting to determine whether these conserved accumulation patterns correlate with conserved sequences within the elements, possibly within their LTRs. We have already mentioned the highly conserved expression patterns of *412* and *mdg1*, possibly a consequence of their extensive overall sequence similarity. In particular, 14 of the 18 bases of their putative primer-binding sites are identical, and there are 27 bases immediately internal to their 5'-LTRs that are identical (Will, Bayev & Finnegan, 1981). In contrast, although the *297* and *17.6* elements share extensive sequence homology over a 1.7 kb 3'-region, they exhibit almost completely non-overlapping patterns of expression in the embryo, suggesting that the internal regions that confer these patterns might reside in the more 5'-region of these elements. This has recently been shown to be the case for the control of *17.6* expression in the lamina anlage of the visual system (Mozer & Benzer, 1994).

Available sequence similarity and transcription factor binding-motif computer search programs are unable to reveal sequence motifs that are common to elements with conserved expression patterns (data not shown). Convincing proof of shared *trans*-acting factors and related internal *cis*-regulatory sequences will require direct molecular evidence for binding of specific host factors to particular retrotransposon sequences in transgenes as well as *in vitro* (see above for *412* and *UBX*).

Several of the transcript accumulation patterns we have seen are unique to particular tissues and/or times, as well as to particular families of elements. For example, *3S18* transcripts are the only ones found in the malpighian tubules, and *blood* transcripts are the only ones found in the hindgut. Again, it will be of some interest to determine which sequences within these elements mediate activation in these tissues.

Since the retrotransposons do not themselves encode tissue-specific transcription factors, the observed spatio-temporal transcript accumulation patterns are likely to be a consequence of the interaction of host factors with sequences within the elements themselves. This has been shown to be the case for *UBX* and the *412* element (see above) as well as *copia* (Cavarec & Heidmann, 1993), and it will be of interest to identify additional host factors responsible for the expression patterns of the other elements.

It is worth noting that the spatially restricted expression of some of these retrotransposons makes them potentially useful as markers to follow the formation of specific embryonic structures in both wild-type and mutant embryogenesis, as has been demonstrated for the *412* element in the somatic component of the gonad (Brookman *et al.*, 1992). For

---

CNS (vg: ventral ganglion). In addition, in U-S embryos, transcripts accumulate in the salivary glands (open arrow). (E, F, G, L, M and N) Post-germ band retraction stages. Conserved accumulation continues in the brain (br) and ventral ganglion (vg). In U-S embryos, transcripts continue to accumulate at high levels in the salivary glands (unfilled arrows). Anterior is to the left and dorsal up, except in (E), (F), (L) and (M), which are ventral views.

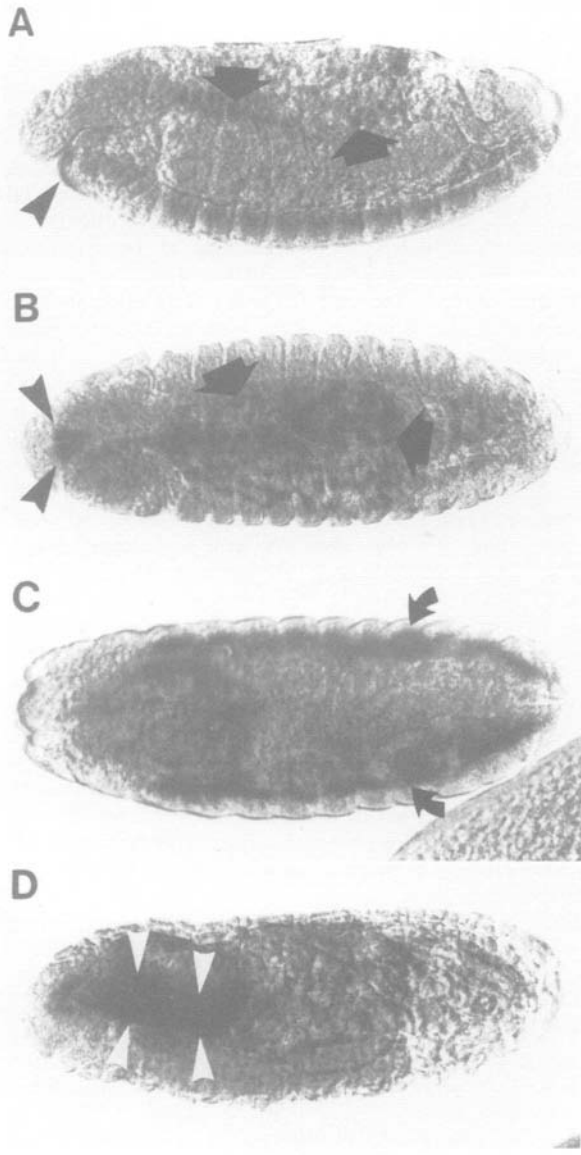


Fig. 14. Spatial expression of *gypsy* during embryogenesis. (A, B) Germ band retraction stages. Transcripts first accumulate in a small group of the outermost ventral cells in the stomodeal invagination (arrowhead) as well as in the yolk nuclei (filled arrows). (C) Post-dorsal closure stage. In addition to continued staining of the stomadeal derivatives, transcripts can be detected in the visceral mesoderm, and the gonads (curved arrows). (D) In a fully developed embryo, the only detectable staining is in the pharyngeal opening and esophagus (white arrowheads). Anterior is to the left. (A), (D): lateral views with dorsal up; (B), (C): ventral views.

example, *springer* could serve as a marker for the salivary glands and *3S18* for the malpighian tubules.

#### *Evolutionary implications of intrinsic spatio-temporal control of retrotransposon expression patterns*

It is over forty years since the discovery of transposable elements led to the suggestion that they might play a role in the control of gene expression (McClintock, 1951). Subsequently, the fact that insertion of transposable elements could cause alterations in devel-

opmental gene expression led to the suggestion that such transposition events could also be responsible for changes in gene expression during evolution (Fincham & Sastry, 1974). This type of change might be expected to be of particular importance since alterations in the regulation of gene expression are likely to be more important for morphological evolution than alterations in the coding regions of genes (King & Wilson, 1975; Li & Noll, 1994).

It was suggested some time ago that retrotransposon insertion could cause mutant phenotypes by affecting the transcription of a nearby gene through an influence of the element's regulatory sequences on the nearby endogenous promoter (Parkhurst & Corces, 1985). Some more recent analyses of spontaneous mutations caused by retrotransposon insertion illustrate how this type of change can affect morphological phenotype. In chickens, the presence of a retroviral LTR-derived promoter upstream of the gonad-specific aromatase gene causes extragonadal expression of aromatase and gives rise to the *Henny-feathering* trait (Matsumine *et al.*, 1991). Similarly, tissue-specific transcriptional regulatory sequences within the retrotransposon *Tom* are believed to be responsible for the *Om(1D)* rough-eyed phenotype of *Drosophila ananassae* by induction of eye-imaginal disc expression of a homeodomain protein (Tanda & Corces, 1991).

While the previous two examples result in mutant defects, there are also examples of normal genes whose transcriptional regulatory elements contain LTRs derived from retrotransposons [reviewed in (Robins & Samuelson, 1992)]. First, the LTR sequence of an ancient provirus is thought to impose androgen regulation on the mouse *Sex-limited protein (Slp)* (Adler *et al.*, 1991; Stravenhagen & Robins, 1988). Second, a retrotransposon has been found to be responsible for regulation of salivary gland expression of the human amylase gene (Matsumine *et al.*, 1991). Third, retrotransposons might program tissue specific expression of anthocyanins in maize (Radicella *et al.*, 1992).

Our results suggest that the retrotransposons in the *Drosophila* genome contain a rich repertoire of transcriptional regulatory elements that could confer novel spatial and temporal patterns of expression on host genes into, or near, which they insert (Robins & Samuelson, 1992). Should such an insertion occur in the germline, the altered expression pattern would be heritable and thus could ultimately become material for evolutionary change. The hypothesis that gene duplication followed by divergence forms the basis for much evolutionary novelty was elaborated some time ago (Lewis, 1951). It has, however, only recently been shown that divergence in the *cis*-regulatory regions is likely to serve as an important basis for the evolution of distinct developmental functions (Li & Noll, 1994). On the basis of the data presented here, insertion of retrotransposons into regulatory regions might serve as one of the mechanisms promoting the acquisition

of novel spatio-temporal control functions by these regions. Elucidation of the exact regulatory sequences within the retrotransposons and their comparison with the *cis*-regulatory regions of developmentally important genes expressed in similar, or overlapping, patterns might ultimately allow the elucidation of evolutionary alterations caused by specific retrotransposon insertion events.

We are very grateful to Dr. Susan Parkhurst for her invaluable encouragement and help in initiating and conducting this research project. We thank S. Parkhurst, W. Bender, M. Best-Belpomme, S. Kidd, and N. Junakovic for gifts of DNA; S. Celniker and E. B. Lewis for the L-S and U-S wild-type strains; R. Mann and S. Benzer for communicating unpublished results; S. Carstairs for conducting the computer database searches; and D. Mathog for assistance with sequence analysis. We are grateful to D. Finnegan, E. B. Lewis, E. Meyerowitz, W. Dreyer, M. Lamka, S. Halsell, R. Yip, A. Bashirullah and K. Whittaker for providing critical comments on the manuscript. D. D. was supported, in part, by graduate fellowships from the California Foundation for Biochemical Research and the Howard Hughes Medical Institute. This research was supported by a grant from the Gustavus and Louise Pfeiffer Research Foundation and a gift for the support of genetics research from Millard and Muriel Jacobs (both awards to H. D. L.).

## References

- Adler, A., Scheller, A., Hoffman, Y. & Robins, D. M. (1991). Multiple components of a complex androgen-dependent enhancer. *Molecular Endocrinology* **5**, 1587–1596.
- Arkhipova, I. R. & Ilyin, Y. V. (1992). Control of transcription of *Drosophila* retrotransposons. *BioEssays* **14**, 161–168.
- Bell, J. R., Bogardus, A. M., Schmidt, T. & Pellegrini, M. (1985). A new *copia*-like transposable element found in a *Drosophila* rDNA gene unit. *Nucleic Acids Research* **13**, 3861–3871.
- Bender, W., Spierer, P. & Hogness, D. S. (1983). Chromosomal walking and jumping to isolate DNA from the *Ace* and *rosy* loci and the *bithorax* complex in *Drosophila melanogaster*. *Journal of Molecular Biology* **168**, 17–33.
- Bingham, P. M. & Chapman, C. H. (1986). Evidence that *white-blood* is a novel type of temperature-sensitive mutation resulting from temperature-dependent effects of a transposon insertion on formation of *white* transcripts. *EMBO Journal* **5**, 3343–3351.
- Brookman, J. J., Toosy, A. T., Shashidhara, L. S. & White, R. A. (1992). The *412* retrotransposon and the development of gonadal mesoderm in *Drosophila*. *Development* **116**, 1185–1192.
- Cavarec, L. & Heidmann, T. (1993). The *Drosophila copia* retrotransposon contains binding sites for transcriptional regulation by homeoproteins. *Nucleic Acids Research* **21**, 5041–5049.
- Ding, D., Parkhurst, S. M. & Lipshitz, H. D. (1993). Different genetic requirements for anterior RNA localization revealed by the distribution of *Adducin-like* transcripts during *Drosophila* oogenesis. *Proceedings of the National Academy of Sciences USA* **90**, 2512–2516.
- Errede, B. (1993). MCM1 binds to a transcriptional control element in *Ty1*. *Molecular and Cellular Biology* **13**, 57–62.
- Fincham, J. R. S. & Sastry, G. R. K. (1974). Controlling elements in maize. *Annual Review of Genetics* **8**, 15–50.
- Finnegan, D. J. & Fawcett, D. H. (1986). Transposable elements in *Drosophila melanogaster*. *Oxford Surveys on Eukaryotic Genes* **3**, 1–62.
- Fjose, A., McGinnis, W. J. & Gehring, W. J. (1985). Isolation of a homeobox-containing gene from the *engrailed* region of *Drosophila* and the spatial distribution of its transcripts. *Nature* **313**, 284–289.
- Flavell, A. J., Ruby, S. W., Toole, J. J., Roberts, B. E. & Rubin, G. M. (1980). Translation and developmental regulation of RNA encoded by the eukaryotic transposable element *copia*. *Proceedings of the National Academy of Sciences USA* **77**, 7107–7111.
- Heidmann, O. & Heidmann, T. (1991). Retrotransposition of a mouse IAP sequence tagged with an indicator gene. *Cell* **64**, 159–170.
- Huijser, P., Kirchhoff, C., Lankenau, D.-H. & Hennig, W. (1988). Retrotransposon-like sequences are expressed in Y-chromosomal lampbrush loops of *Drosophila hydei*. *Journal of Molecular Biology* **203**, 689–697.
- Ilyin, Y. V., Chmeliauskaite, V. G., Ananiev, E. V. & Georgiev, G. P. (1980). Isolation and characterization of a new family of mobile dispersed genetic elements, *mdg3*, in *Drosophila melanogaster*. *Chromosoma* **81**, 27–53.
- Ilyin, Y. V., Chmeliauskaite, V. G., Ananiev, E. V., Lyubomirskaya, N. V., Kulguskin, V. V., Bayev, A. A. & Georgiev, G. P. (1980). Mobile dispersed genetic element *mdg1* of *Drosophila melanogaster*: structural organization. *Nucleic Acids Research* **8**, 5333–5346.
- Ilyin, Y. V., Tchurikov, N. A., Ananiev, E. V., Ryskov, A. P., Yenikoplov, G. N., Limborska, S. A., Maleeva, N. E., Gvozdev, V. A. & Georgiev, G. P. (1978). Studies on the DNA fragments of mammals and *Drosophila* containing structural genes and adjacent sequences. *Cold Spring Harbor Symposia on Quantitative Biology* **42**, 959–969.
- Karlik, C. C. & Fyrberg, E. A. (1985). An insertion within a variably spliced *Drosophila* tropomyosin gene blocks accumulation of only one encoded isoform. *Cell* **41**, 57–66.
- Keshet, E., Schiff, R. & Itin, A. (1991). Mouse retrotransposons: A cellular reservoir of long terminal repeat (LTR) elements with diverse transcriptional specificities. *Advances in Cancer Research* **56**, 215–251.
- Kidd, S. & Young, M. W. (1986). Transposon-dependent mutant phenotypes at the *Notch* locus of *Drosophila*. *Nature* **323**, 89–91.
- King, M. C. & Wilson, A. C. (1975). Evolution at two levels in human and chimpanzees. *Science* **188**, 107–116.
- Kornberg, T., Siden, I., O'Farrell, P. & Simon, M. (1985). The *engrailed* locus of *Drosophila*: *in situ* localization of transcripts reveals compartment-specific expression. *Cell* **40**, 45–53.
- Lankenau, S., Corces, V. & Lankenau, D.-H. (1994). The *Drosophila micropia* retrotransposon encodes a testis-specific antisense RNA complementary to reverse transcriptase. *Molecular and Cellular Biology* **14**, 1764–1775.
- Lavorgna, G., Malva, C., Manzi, A., Gigliotti, S. & Graziani, F. (1989). The abnormal oocyte phenotype is correlated with the presence of *blood* transposon in *Drosophila melanogaster*. *Genetics* **123**, 485–494.
- Lewis, E. B. (1951). Pseudoallelism and gene evolution. *Cold Spring Harbor Symposia on Quantitative Biology* **16**, 159–174.
- Li, X. & Noll, M. (1994). Evolution of distinct developmental functions of three *Drosophila* genes by acquisition of different *cis*-regulatory regions. *Nature* **367**, 83–87.
- Löhning, C., Rosenbaum, C. & Ciriacy, M. (1993). Isolation of the *TYE2* gene reveals its identity to *SWI3* encoding a general transcription factor in *Saccharomyces cerevisiae*. *Current Genetics* **24**, 193–199.

- Marlor, R. L., Parkhurst, S. M. & Corces, V. G. (1986). The *Drosophila melanogaster gypsy* transposable element encodes putative gene products homologous to retroviral proteins. *Molecular and Cellular Biology* **6**, 1129–1134.
- Matsumine, H., Herbst, M. A., Ou, S.-H. I., Wilson, J. D. & McPhaul, M. J. (1991). Aromatase mRNA in the extragonadal tissue of chickens with the henney-feathering trait is derived from a distinctive promoter structure that contains a segment of a retroviral long terminal repeat. *Journal of Biological Chemistry* **266**, 19900–19907.
- McClintock, B. (1951). Chromosome organization and genic expression. *Cold Spring Harbor Symposia on Quantitative Biology* **16**, 13–47.
- Meyerowitz, E. M. & Hogness, D. S. (1982). Molecular organization of a *Drosophila* puff site that responds to ecdysone. *Cell* **28**, 165–176.
- Micard, D., Couderc, J. L., Sobrier, M. L., Giraud, G. & Dastugue, B. (1988). Molecular study of the retrovirus-like transposable element 412, a 20-OH ecdysone responsive repetitive sequence in *Drosophila* cultured cells. *Nucleic Acids Research* **16**, 455–470.
- Mossie, K. G., Young, M. W. & Varmus, H. E. (1985). Extrachromosomal DNA forms of  *copia*-like transposable elements, *F* elements and middle repetitive DNA sequences in *Drosophila melanogaster*. Variation in cultured cells and embryos. *Journal of Molecular Biology* **182**, 31–43.
- Mount, S. M. & Rubin, G. M. (1985). Complete nucleotide sequence of the *Drosophila* transposable element  *copia*: homology between  *copia* and retroviral proteins. *Molecular and Cellular Biology* **5**, 1630–1638.
- Mozer, B. A. & Benzer, S. (1994). Ingrowth by photoreceptor axons induces transcription of a retrotransposon in the developing *Drosophila* brain. *Development* **120**, 1049–1058.
- Parkhurst, S. M. & Corces, S. M. (1985). *forked*, *gypsies*, and suppressors in *Drosophila*. *Cell* **41**, 429–437.
- Parkhurst, S. M. & Corces, V. G. (1987). Developmental expression of *Drosophila melanogaster* retrovirus-like transposable elements. *EMBO Journal* **6**, 419–424.
- Peronnet, F., Becker, J. L., Becker, J., d'Auriol, L., Galibert, F. & Best-Belpomme, M. (1986). 1731, a new retrotransposon with hormone modulated expression. *Nucleic Acids Research* **14**, 9017–9033.
- Pouteau, S., Huttner, E., Grandbastien, M. A. & Caboche, M. (1991). Specific expression of the tobacco Tnt1 retrotransposon in protoplasts. *EMBO Journal* **10**, 1911–1918.
- Radicella, J. P., Brown, D., Tolar, L. A. & Chandler, V. L. (1992). Allelic diversity of the maize *B* regulatory gene: different leader and promoter sequences of two *B* alleles determine distinct tissue specificities of anthocyanin production. *Genes and Development* **6**, 2152–2164.
- Robins, D. M. & Samuelson, L. C. (1992). Retrotransposons and the evolution of mammalian gene expression. *Genetica* **86**, 191–201.
- Scherer, G., Telford, J., Baldari, C. & Pirrotta, V. (1981). Isolation of cloned genes differentially expressed at early and late stages of *Drosophila* embryonic development. *Developmental Biology* **86**, 438–447.
- Scherer, G., Tschudi, C., Perera, J., Delius, H. & Pirrotta, V. (1982). *B104*, a new dispersed repeated gene family in *Drosophila melanogaster* and its analogies with retroviruses. *J. Mol. Biol.* **157**, 435–451.
- Schwartz, H. E., Lockett, T. J. & Young, M. W. (1982). Analysis of transcripts from two families of nomadic DNA. *Journal of Molecular Biology* **157**, 49–68.
- Snyder, M. P., Kimbrell, D., Hunkapiller, M., Hill, R., Fristrom, J. & Davidson, N. (1982). A transposable element that splits the promoter region inactivates a *Drosophila* cuticle protein gene. *Proceedings of the National Academy of Sciences USA* **79**, 7430–7434.
- Stravnenhagen, J. B. & Robins, D. M. (1988). An ancient provirus has imposed androgen regulation on the adjacent mouse sex-limited protein gene. *Cell* **55**, 247–254.
- Strobel, E., Dunsmuir, P. & Rubin, G. M. (1979). Polymorphisms in the chromosomal locations of elements of the 412,  *copia* and 297 dispersed repeated families in *Drosophila*. *Cell* **17**, 429–439.
- Tanda, S. & Corces, V. G. (1991). Retrotransposon-induced overexpression of a homeobox gene causes defects in eye morphogenesis in *Drosophila*. *EMBO Journal* **10**, 407–417.
- Tautz, D. & Pfeifle, C. (1989). A non-radioactive *in situ* hybridization method for the localization of specific RNAs in *Drosophila* embryos reveals translational control of the segmentation gene *hunchback*. *Chromosoma* **98**, 81–85.
- Ting, C.-N., Rosenberg, M. P., Snow, C. M., Samuelson, L. C. & Meisler, M. H. (1992). Endogenous retroviral sequences are required for tissue-specific expression of a human salivary amylase gene. *Genes and Development* **6**, 1457–1465.
- Will, B. M., Bayev, A. A. & Finnegan, D. J. (1981). Nucleotide sequence of terminal repeats of 412 transposable elements of *Drosophila melanogaster*. *Journal of Molecular Biology* **153**, 897–915.
- Young, M. W. & Schwartz, H. E. (1981). Nomadic gene families in *Drosophila*. *Cold Spring Harbor Symposia on Quantitative Biology* **45**, 629–640.
- Yu, G. & Fassler, J. S. (1993). SPT13 (GAL11) of *Saccharomyces cerevisiae* negatively regulates activity of the MCM1 transcription factor in Ty1 elements. *Molecular and Cellular Biology* **13**, 63–71.
- Yuki, S., Inouye, S., Ishimaru, S. & Saigo, K. (1986). Nucleotide sequence characterization of a *Drosophila* retrotransposon, 412. *European Journal of Biochemistry* **158**, 403–410.