Silencing of a gene adjacent to the breakpoint of a widespread Drosophila inversion by a transposon-induced antisense RNA

Marta Puig*, Mario Cáceres†, and Alfredo Ruiz‡‡

*Departament de Genètica i de Microbiologia, Universitat Autònoma de Barcelona, 08193 Bellaterra, Barcelona, Spain; †Department of Human Genetics, Emory University School of Medicine, 615 Michael Street, Atlanta, GA 30322

Communicated by Margaret G. Kidwell, University of Arizona, Tucson, AZ, May 4, 2004 (received for review February 5, 2004)

Adaptive changes in nature occur by a variety of mechanisms, and Drosophila chromosomal inversions was one of the first studied examples. However, the precise genetic causes of the adaptive value of inversions remain uncertain. Here we investigate the impact of the widespread inversion 2j of Drosophila buzzatii on the expression of the CG13617 gene, whose coding region is located only 12 bp away from the inversion proximal breakpoint. This gene is transcribed into a 2.3-kb mRNA present in all D. buzzatii developmental stages. More importantly, the expression level of CG13617 is reduced 5-fold in embryos of lines homozygous for the 2j inversion compared with lines without the inversion. An antisense RNA that originates in the Foldback-like transposon Kepler inserted at the breakpoint junction in all of the 2j lines and that forms duplexes with the CG13617 mRNA in 2j embryos is most likely responsible for the near silencing of the gene. Few examples of RNA interference caused by transposable elements (TEs) have been previously described, but this mechanism might be prevalent in many organisms and illustrates the potential of TEs as a major source of genetic variation. In addition, because chromosomal rearrangements are usually induced by TEs, position effects might be more common than previously recognized and contribute significantly to the evolutionary success of inversions.

In the genus Drosophila, many species are polymorphic for inversions in one or more chromosomes, providing the most extraordinary and best studied example of chromosomal variation in nature (1). Compelling evidence for the adaptive significance of inversion polymorphisms has been drawn from seasonal and long-term frequency changes as well as latitudinal and altitudinal clines (2, 3). However, the molecular mechanisms underlying inversion maintenance in natural populations are still unclear. According to the coadaptation hypothesis, the reduction of recombination in the inverted chromosomal segment of heterokaryotypes keeps together favorable allele combinations (4, 5). Alternatively, the position effect hypothesis proposes that the localization of the inversion breakpoints near or inside genes could affect their function or expression profile (6). Although the available evidence favors coadaptation as the most plausible explanation for the adaptive value of chromosomal inversions (2, 7, 8), little is known of the mutational outcome of breakpoints in natural inversions and their consequences on the expression patterns of nearby genes (9).

A variety of position effects have been observed in spontaneous mutations generated by inversions in diverse organisms. Inversions can disrupt the coding region of a gene, causing the loss of its function, as in the Dpp6 gene in mice (10) or the hemophilia A factor VIII gene in humans (11). The inversion of a chromosomal segment can also remove or exchange the regulatory sequences of a gene and alter its expression pattern, as in the Antp7280 mutation of Drosophila melanogaster (12) or the niveu gene of Antirrhinum (13). Finally, inversions can move genes to distant sites where their expression is silenced by the proximity of centromeric heterochromatin (14). Another factor that could contribute to the position effects of inversions is the presence of transposable elements (TEs) at their breakpoints. TEs constitute a large fraction of most eukaryotic genomes and have been implicated in the generation of inversions in Drosophila and other organisms (15, 16). In addition, TEs have been shown to have diverse effects on gene expression (17, 18), as a result not only of the modification of functional regions (19–21) but also of interference with the transcription of adjacent genes by read-through transcription, antisense transcripts, or insulation (22).

The analysis of position effects of natural inversions has been hindered by the lack of molecular studies of their breakpoints and flanking genes. In this work we have studied the case of inversion 2j of Drosophila buzzatii, which was originated by a unique event of ectopic recombination between two oppositely oriented copies of the Foldback-like element Galileo and which contains at its breakpoints blocks of different TEs absent from noninverted chromosomes (15, 23). In D. buzzatii chromosome 2, the ancestral 2 standard (2st) arrangement and the derived 2j arrangement are found in most natural populations at high frequencies (24). Both arrangements seem to be maintained as a balanced polymorphism by a fitness tradeoff in which the carriers of the 2j and 2st arrangements are characterized by a larger adult body size and a shorter developmental time, respectively (25), although the causes of these differences are not known. Two genes, Pp10-96-A and nAcrB-96-A (15), were originally found flanking the 2j proximal breakpoint in the 2st arrangement. However, the sequencing of the D. melanogaster genome (26) revealed a putative ORF named CG13617 located outside the inversion and just a few nucleotides away from the breakpoint (Fig. 1). This observation prompted us to ascertain whether the generation of the inversion caused any change in CG13617 expression. The results show that this gene is nearly silenced in embryos of lines homozygous for inversion 2j and that this silencing is not caused by the inversion itself but by the transcription of an antisense RNA from the transposon Kepler inserted at the proximal breakpoint junction.

Materials and Methods

Flies. Eight isogenic D. buzzatii lines homozygous for chromosomal arrangements 2st (n = 4), 2j (n = 3), or 2z3 (n = 1) were used. Arrangement 2z3 is derived from 2j by an additional inversion, 2z3 (27). These lines represent the natural variability within the species’ geographical range (Table 1), and the 2j lines differ in the size and TE content of the insertions at the 2j proximal breakpoint (23).

DNA Sequencing and Sequence Analysis. During the cloning of inversion 2j breakpoints (15), the CG13617 gene was partially sequenced in the D. buzzatii lines st-I and j-I. The CG13617 sequence was completed in both lines by subcloning the Ast9 phage

Abbreviation: TE, transposable element.

Data deposition: The sequences have been deposited in the GenBank database (accession nos. AY551073–AY551076).

To whom correspondence should be addressed. E-mail: alfredo.ruiz@uab.es.

© 2004 by The National Academy of Sciences of the USA
into the pBluescript II SK vector (Stratagene) and by PCR amplification (Fig. 1). PCR and RT-PCR products were gel-purified and sequenced directly with the same primers. GenBank similarity searches were performed by using BLASTX to identify CG13617 coding regions and BLASTP to find homologous proteins in other species. Sequences were aligned with CLUSTALW.

RNA Extraction and Northern Analysis. Total RNA was isolated from embryos, larvae, pupae, and adults by using TRIzol (Invitrogen). Northern blot hybridization was performed as described (28), with 20 μg of total RNA from each developmental stage of lines st-1 and j-1. Loading of the gel was verified by ethidium bromide staining. An antisense [32P]UTP-labeled riboprobe was synthesized by in vitro transcription from a clone containing the 1,128-bp E2/E10 RT-PCR product from st-1 CG13617 cDNA. Hybridization was carried out for 16 h at 68°C.

RT-PCR and PCR Amplification. Total RNA was treated with 1 unit of DNase I (Ambion, Austin, TX) for 30 min at 37°C to eliminate DNA contamination, and cDNA was synthesized from 500 ng of the DNase I-treated RNA by using an oligo(dT) primer (First Strand cDNA Synthesis kit for RT-PCR, Roche Diagnostics). Negative reactions without retrotranscriptase were carried out to control for DNA contamination. PCRs were performed in a total volume of 25 μl, including 1 μl of cDNA or 100–200 ng of genomic DNA, 10 pmol of each primer, 200 μM dNTPs, 1.5 mM MgCl2, and 1.5 units of Taq DNA polymerase. To differentiate the size of amplification products from cDNA and genomic DNA, primer pairs used in RT-PCRs were selected to span an intron of the gene. Primer sequences are available in Table 3, which is published as supporting information on the PNAS web site.

RACE. RACE experiments were done with DNase I-treated RNA from embryos of lines st-1 or j-1; 5′ RACE was carried out by using the 5′/3′ RACE kit (Roche Diagnostics), and 3′ RACE was carried out by using cDNA synthesized with primer poly(T). The gene-specific primers used in each case are listed in Table 3. All of the amplification products spanned one intron of the gene to ensure

Table 1. D. buzzatii isogenic lines used in this study

<table>
<thead>
<tr>
<th>Line</th>
<th>Chromosomal arrangement</th>
<th>Geographic origin</th>
<th>Proximal breakpoint insertions</th>
</tr>
</thead>
<tbody>
<tr>
<td>st-1</td>
<td>2st</td>
<td>Carboneras, Spain</td>
<td>0</td>
</tr>
<tr>
<td>st-4</td>
<td>2st</td>
<td>Guaritas, Brazil</td>
<td>0</td>
</tr>
<tr>
<td>st-7</td>
<td>2st</td>
<td>Termas de Río Hondo, Argentina</td>
<td>0</td>
</tr>
<tr>
<td>st-8</td>
<td>2st</td>
<td>Tucucho, Argentina</td>
<td>0</td>
</tr>
<tr>
<td>j-1</td>
<td>2j</td>
<td>Carboneras, Spain</td>
<td>4,313 Galileo, Kepler (2), But1, But3 (2)</td>
</tr>
<tr>
<td>j-13</td>
<td>2j</td>
<td>Guaritas, Brazil</td>
<td>3,214 Galileo, Kepler, But1, But3</td>
</tr>
<tr>
<td>j-19</td>
<td>2j</td>
<td>Tucucho, Argentina</td>
<td>4,724 Galileo, Kepler, But1, But3, Newton</td>
</tr>
<tr>
<td>jz3-4</td>
<td>2jz3</td>
<td>Tilcara, Argentina</td>
<td>6,341 Galileo (2), Kepler, But1, But3 (2)</td>
</tr>
</tbody>
</table>

*The number of TE copies is indicated in parenthesis when more than one copy is present. The exact nature of the breakpoint insertions is described in detail in figure 2 of ref. 23.
that they came from mRNA. In some cases, different amounts of the first amplification were tested as a template for the second one to obtain a specific band. RACE products were cloned into the pGEM-T vector (Promega). A minimum of eight different clones from each reaction were screened by restriction mapping, and two or three clones were finally sequenced. In the 5’ RACE, clones with different inserts were selected for sequencing.

**Real-Time RT-PCR.** Real-time RT-PCR was performed in an ABI PRISM 7900HT Sequence Detection System with the DNA-binding dye SYBR Green (Applied Biosystems). Primers were designed using PRIMER EXPRESS VERSION 1.5 software (Applied Biosystems) in areas conserved between 2st and 2j lines (Table 3). The housekeeping gene Gapdh was used as internal control for differences in cDNA concentration. Previously, 1,017 bp of the D. buzzatii Gapdh gene were amplified and sequenced from st-1 genomic DNA. For each sample, the gene of interest and Gapdh were both amplified in triplicate, and results were analyzed by using SEQUENCE DETECTOR VERSION 1.7 and DISSOCIATION CURVE VERSION 1.0 software (Applied Biosystems). Relative quantification was performed with the standard curve method, and gene amplification levels were normalized by dividing by Gapdh levels in each sample. Expression levels were compared by means of ANOVA.

**Results**

The gene CG13617 and its flanking regions were sequenced in two D. buzzatii lines, with (j-1) and without (st-1) the 2j inversion (Fig. 1). Sequence comparison with the predicted D. melanogaster CG13617 gene (26) allowed us to determine the D. buzzatii CG13617 coding region, which is 2,202 bp long and shares a 65% nucleotide identity with D. melanogaster. The D. buzzatii gene is split into four exons and lacks the second intron of D. melanogaster, but the other three are very similar in size and location (Table 2). The CG13617 sequence presents an overall nucleotide identity of 97.8% between st-1 and j-1. The most remarkable difference was the large TE insertions of the 2j proximal breakpoint located only 12 bp from the stop codon of the gene in line j-1 (Fig. 1). Furthermore, sequence comparison revealed two other small structural changes in line j-1: a 166-bp tandem duplication and a 26-bp insertion located 1,010 and 408 bp upstream of the start codon, respectively (Fig. 1). To further characterize the structural variation between 2st and 2j lines, the entire region comprised by the CG13617 and its flanking regions were sequenced in two D. buzzatii lines along the embryo (Fig. 7, which is published as supporting information on the PNAS web site). Only a few restriction site polymorphisms were detected in three additional lines with each arrangement (Table 1) by PCR amplification and restriction mapping (Table 4, which is published as supporting information on the PNAS web site). A similar real-time RT-PCR analysis was performed with Ppl-a-964, the next closest gene to the proximal breakpoint (Fig. 1). However, no significant differences were detected in this case between 2st and 2j embryos (P = 0.952; Fig. 4 and Table 5), corroborating previous results (15). Finally, in situ hybridization in whole embryos was carried out to study spatial differences in CG13617 expression between 2st and 2j lines. Stage 11–12 embryos showed a similar expression pattern in both lines, which consisted of several signals in the head and in each metamer, forming two lines along the embryo (Fig. 7, which is published as supporting information on the PNAS web site). However, the pattern clearly

---

**Table 2. Structure of the CG13617 coding region in D. buzzatii and D. melanogaster**

<table>
<thead>
<tr>
<th></th>
<th>D. buzzatii</th>
<th>D. melanogaster</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exon</td>
<td>st-1, bp</td>
<td>j-1, bp</td>
</tr>
<tr>
<td></td>
<td>94</td>
<td>94</td>
</tr>
<tr>
<td>Intron</td>
<td>60</td>
<td>60</td>
</tr>
<tr>
<td>Exon</td>
<td>1,513</td>
<td>1,513</td>
</tr>
<tr>
<td>Intron</td>
<td>56</td>
<td>53</td>
</tr>
<tr>
<td>Exon</td>
<td>157</td>
<td>157</td>
</tr>
<tr>
<td>Intron</td>
<td>83</td>
<td>81</td>
</tr>
<tr>
<td>Exon</td>
<td>438</td>
<td>438</td>
</tr>
</tbody>
</table>

---

**Fig. 2.** Nucleotide sequence of the 3’ end of CG13617 in lines st-1 and j-1. The stop codon, the two putative polyadenylation signals, and the Galileo element inserted in line j-1 are included in shaded boxes. A vertical arrow indicates the 2j inversion proximal breakpoint (BP). Sequences present only in lines with the inversion are shown in italic.
Expression analysis of CG13617 was fainter in 2j than in 2st embryos, possibly reflecting the expression level difference between them.

An unexpected result was that in the E2/E7 RT-PCR, besides the 387-bp band from CG13617 mRNA, an additional band of equal size to the genomic DNA (470 bp) appeared in embryos of the four 2j lines studied (Fig. 3B and data not shown). To confirm this result, two RT-PCR amplifications with primer pairs E2/E9 and I3/E2, in which one primer is placed in an intron sequence, were carried out by using RNA from embryos of the eight lines. Products with the expected size (211 and 398 bp, respectively) were obtained in all 2j lines but not in the 2st lines or any of the negative controls without retrotranscriptase, ruling out genomic DNA contamination (data not shown). Two explanations were possible for this additional transcript containing at least the last two introns of CG13617: an unprocessed pre-mRNA or an antisense RNA transcribed from the other DNA strand, which lacks the appropriate splicing signals. To test the second option, a strand-specific RT-PCR (29) was performed. In the sense-specific cDNAs, a single band of 387 bp, corresponding to the mRNA after the elimination of the third intron, was amplified in all lines, with clearly lower levels in 2j than

![Fig. 3. Expression analysis of CG13617 in four different developmental stages of two lines with (2j) and without (2st) the inversion. (A) Northern blot analysis of CG13617. (Upper) An ~2.3-kb transcript was detected in embryos, pupae, and adults. (Lower) RNA used as a loading control is shown. The RNA amount was higher in embryos, but this fact is not relevant for the comparison between arrangements. (B) Semi-quantitative RT-PCR analysis of CG13617. Primers E2/E7 and H1/H2 were used to amplify, respectively, 387 bp of CG13617 mRNA (Upper) and 438 bp of Gapdh mRNA (Lower) as a reference. In 2j embryos, there is an extra band containing intron 3 that corresponds to the CG13617 antisense transcript. E, embryos; L, larvae; P, pupae; A, adults.](image1)

was fainter in 2j than in 2st embryos, possibly reflecting the expression level difference between them.

An unexpected result was that in the E2/E7 RT-PCR, besides the 387-bp band from CG13617 mRNA, an additional band of equal size to the genomic DNA (470 bp) appeared in embryos of the four 2j lines studied (Fig. 3B and data not shown). To confirm this result, two RT-PCR amplifications with primer pairs E2/E9 and I3/E2, in which one primer is placed in an intron sequence, were carried out by using RNA from embryos of the eight lines. Products with the expected size (211 and 398 bp, respectively) were obtained in all 2j lines but not in the 2st lines or any of the negative controls without retrotranscriptase, ruling out genomic DNA contamination (data not shown). Two explanations were possible for this additional transcript containing at least the last two introns of CG13617: an unprocessed pre-mRNA or an antisense RNA transcribed from the other DNA strand, which lacks the appropriate splicing signals. To test the second option, a strand-specific RT-PCR (29) was performed. In the sense-specific cDNAs, a single band of 387 bp, corresponding to the mRNA after the elimination of the third intron, was amplified in all lines, with clearly lower levels in 2j than

![Fig. 4. Relative quantification of CG13617 mRNA and antisense RNA and Pp1α-96A mRNA by real-time RT-PCR. Average expression levels for four 2st and four 2j lines are represented in the graphs. Standard deviation within each arrangement is indicated by error bars. Expression levels are shown in relation to those of lines st-1 (CG13617 and Pp1α-96A mRNA) and j-1 (antisense RNA). A primer spanning the third and fourth exons (E2S) and a primer located in the second intron (A51) were used to ensure that, respectively, only the sense or the antisense transcripts of CG13617 were amplified.](image2)

in 2st embryos (Fig. 5A). No extra band was observed in any case. In the antisense-specific cDNAs, a 470-bp product was obtained in 2j lines, but only a weak signal appeared in some of the 2st lines (Fig. 5A). Thus, an antisense transcript of CG13617 is expressed in all 2j embryos but is not found or found only in very low amounts in 2st embryos. This antisense RNA is apparently absent in other developmental stages, except for a barely visible band in 2j larvae and pupae (Fig. 5B).

The level of the CG13617 antisense RNA in embryos was quantified by real-time RT-PCR by using a primer located in the second intron. The same eight lines were analyzed, and no differences were found among those with the same arrangement. However, the average antisense RNA expression level was 5 times higher in 2j than in 2st lines (Fig. 4), a difference that is highly significant (P = 0.016; Table 5). In addition, sense and antisense RNA expression was analyzed by real-time RT-PCR in embryos heterozygous for 2 inversion produced by the two reciprocal crosses of lines st-1 and j-1. In both analyses, expression levels were again significantly different between the parental st-1 and j-1 lines, even though a new primer pair from the third intron region was used to quantify the antisense RNA, and no significant differences were found between the two reciprocal heterokaryotypes (Table 6, which is published as supporting information on the PNAS web site). In the sense RNA, the expression level in both heterokaryotypes was similar to that of line j-1 and was significantly lower compared with the average of st-1 and j-1 homoygotes (Table 6). Conversely, the average antisense RNA level in the heterokaryotypes was not significantly different from the average of the two homozygous parental lines (Table 6), pointing to an intermediate expression of this transcript in heteroygotes.

The proximity of CG13617 to the TEs inserted at the proximal breakpoint in the 2j lines (Fig. 1) suggests that the antisense RNA is transcribed predominantly from these elements. In the lines without the inversion there may be a low level of antisense transcription in the absence of TEs, but the origin of this RNA is still unclear. The 5’ end of the antisense RNA in the 2j lines was characterized by RT-PCR from the strand-specific cDNAs using eight primers (G17 to G11) located in the part of the TE insertions shared by most 2j lines, in combination with three primers located in the last exons of CG13617 (E7, D1, and D2) (Fig. 1). RT-PCR products of the expected size were obtained with five primers closest to the coding region (G17 to K1) in all four 2j lines. Primers T5 and K2 yielded some amplification only in lines j-13 and j-14. Finally, no amplification was observed in any 2j line with the

![Fig. 5. Strand-specific RT-PCR of CG13617. Two different cDNAs were synthesized for each sample by using primers specific for the sense (E5) and antisense (E6) transcript and were then amplified with primer pair E2/E7 (RT + ). Negative controls for each sample without retrotranscriptase (RT - ) are also shown. (A) Amplification of sense and antisense RNAs in embryos from eight different lines. The 387-bp band in the sense mRNA is smaller than that of the DNA and the antisense transcript (470 bp) because of the splicing of the 83-bp intron. (B) Amplification of antisense RNA in four developmental stages of lines st-1 and j-1. E, embryos; L, larvae; P, pupae; A, adults.](image3)
outermost primer (G11), placing the origin of the antisense RNA inside the Kepler element inserted within the Galileo copy that generated the 2j inversion. Repeated attempts to clone the 5’ end of the transcript by 5’ RACE failed, perhaps because of the complex secondary structures formed by these Foldback-like TE s. RT-PCR with j-13 and j-19 oligo(dT)-synthesized cDNA was used to determine the approximate extension of the antisense RNA. The expected amplification product was obtained with primers I1, located in the first intron, and E11, but not with primers E13 and U1, which is located in the upstream region (Fig. 1). A 3’ RACE carried out to characterize precisely the 3’ end of the antisense RNA revealed that it coincides with an A-rich region located in the CG13617 5’ UTR of the four 2j lines. Unfortunately, this A-rich region may not provide a binding site for the oligo(dT) primer used in the cDNA synthesis, and we cannot be certain that the antisense transcript really ends there. The antisense RNA thus has an estimated length of ~3 kb and includes the complete coding region of CG13617, as well as all its introns (Fig. 1). Conceptual translation of the 2j antisense RNA sequence revealed 23 small ORFs of 53–143 aa, but none shared significant homology with any known protein, suggesting that this transcript does not have coding capability.

Discussion

CG13617 was described as a potential ORF in the genome of D. melanogaster (26). We have shown that the sequence and structure of CG13617 are conserved in D. buzzatii and that CG13617 is a fully functional gene expressed through the entire life cycle. D. buzzatii CG13617 is transcribed into a 2.3-kb mRNA that encodes a 734-aa protein predicted in D. melanogaster. Both proteins share several domains and structural characteristics typical of transcription factors (30): (i) a C2H2 zinc finger, (ii) four regions of 27–44 aa able to form coiled coils (one of the principal protein oligomerization motifs), (iii) a putative nuclear localization signal, and (iv) two PEST sequences (a motif involved in targeting proteins for rapid destruction) (Fig. 8), which is published as supporting information on the PNAS web site. An exhaustive search of homologous proteins in other species outside Drosophila detected similar proteins in Anopheles gambiae, Rattus norvegicus, Mus musculus, and Homo sapiens (Fig. 8). However, none of these proteins has a known function yet.

In D. buzzatii, CG13617 is found adjacent to the proximal breakpoint of the widespread 2j inversion, providing a unique opportunity to investigate possible position effects. Our results show that the stop codon of this gene is only 12 bp from the breakpoint and that the TE insertions responsible for the generation of the inversion (15, 23) took place inside the 3’ UTR of the gene and altered the end of the CG13617 transcriptional unit in the 2j lines (Fig. 2). More importantly, the expression level of CG13617 was compared between 2st and 2j homozygous lines in embryos, larvae, pupae, and adults. Despite the proximity of the inversion breakpoint, no differences in CG13617 expression were detected in any developmental stage other than embryos, in which the average expression level was 5 times lower in 2j than in 2st lines (Fig. 4).

Several causes could explain the decrease in CG13617 expression, such as the change of position of a downstream enhancer (31), a silencing effect of the repetitive DNA blocks at the breakpoint similar to that of heterochromatin (14), or problems in the mRNA processing due to the modification of the 3’ UTR (32). However, the specific down-regulation affecting only CG13617 expression in embryos (but not Pplα-964, which is also located very close to the breakpoints, or in any other developmental stage) makes most of these explanations unlikely. In addition, other evidences indicate that the silencing is caused by an antisense transcript overlapping the whole CG13617 coding region. First, the reduction of CG13617 expression levels occurs only in 2j embryos, which have the highest amounts of antisense RNA. No differences in the expression level of this gene were observed in other developmental stages, in which the antisense transcript was not found or was found in only very small quantities (Fig. 5B), probably insufficient to cause gene silencing (33). Second, the expression levels of the sense and antisense transcripts are negatively correlated in embryos (Fig. 4). In 2j lines, a 5-fold increase in the level of antisense transcript is accompanied by a decrease in CG13617 mRNA level compared with 2st lines, where the antisense RNA is almost undetectable. Third, the intermediate expression of the antisense RNA in heterozygotes for 2j inversion, together with the low levels of CG13617 expression similar to those of 2j lines, are consistent with a dominant effect acting in trans to silence both copies of the gene. Finally, the higher expression of the antisense RNA in 2j embryos is, apart from the TE insertions at the breakpoints and the inversion, the only characteristic common to all 2j lines that differentiates them from 2st ones.

The use of antisense transcripts as a mechanism of expression down-regulation acting at the posttranscriptional level has been reported in an increasing number of genes in many species ranging from prokayotes to humans, in which the occurrence of sense–antisense transcriptional units is a more common phenomenon than previously thought (29, 34). Antisense RNAs may control gene expression at various levels (35) and have been implicated in diverse processes including genomic imprinting (36), DNA methylation (37), X-inactivation (38), RNA editing (39), and transposon silencing (40), aside from being part of the natural regulation system of some genes (41). The mechanism of action of antisense RNAs is triggered by the formation of double-stranded RNA (dsRNA) duplexes with the sense mRNA. These duplexes are then cleaved into small pieces of 21–23 nucleotides that target the degradation of the complementary transcripts in a process known as RNA interference (42, 43). According to this model, the increased expression of the antisense RNA could be responsible for the elimination of part of the CG13617 mRNA in embryos carrying the 2j inversion. We investigated this possibility by testing the presence of CG13617 dsRNA in embryos (44). As expected, after the dsRNA isolation, a denaturation-dependent amplification product was obtained in 2j, but not in 2st, embryos (Fig. 6), suggesting the existence of sense–antisense duplexes only in embryos carrying the 2j inversion.

In 2j lines, the antisense transcript extends from the TEs inserted at the inversion breakpoint (Fig. 1). Although the exact localization of its 5’ end could not be determined, transcription appears to start inside the Kepler element inserted at the proximal breakpoint in all 2j lines (23). This Kepler element could be contributing a promoter that drives the synthesis of the antisense transcript and causes the silencing of CG13617 in 2j embryos. A similar situation has been recently described in wheat, where retrotransposons inserted throughout the genome generate sense or antisense transcripts of the adjacent genes that increase or silence, respectively, the expression of these genes (33). Previously, the Drosophila TE Hoppel had been shown to silence the Stellate gene through a similar mechanism (44). In the 2j inversion, Kepler is a DNA transposon with structural similarity to the Foldback family (23). Although no ORF encoding a putative transposase has been found and the element is probably
defective, a functional promoter might still be present. Besides, TE activity can be tightly regulated, and it is not unusual that TE promoters are active only in some specific tissues or developmental stages (45, 46).

Our results provide a clear example of a position effect associated with the breakpoints of a Drosophila natural inversion. The 2j inversion of D. buzzatii is widespread in natural populations (24) and can be considered quite successful evolutionarily. Two events may have contributed to the formation of an antisense RNA overlapping the CG13617 gene: the Kepler insertion and the 2j inversion. If the TE insertion happened first, the ectopic recombination, which probably took place in the fragment of Galileo adjacent to CG13617, situated Kepler in its current position and orientation, allowing the antisense transcription of the gene. Alternatively, Kepler could have inserted in an already inverted chromosome assisted by the reduction of recombination at the breakpoints (47). The generation of antisense RNA and consequent silencing of CG13617 might then have resulted in a favorable mutation that swept all other 2j chromosomes in the population, causing perhaps an increase in inversion frequency. This would explain the observed discrepancy between the younger age of the different 2j alleles at the breakpoints compared with the inversion 2j itself, which is much older (23, 48). The consequences of the 5-fold decrease in CG13617 expression level in 2j embryos are hard to guess, because the function of the protein encoded by this gene remains unknown. However, it seems legitimate to speculate that the change in CG13617 expression could be related to the pheno-
typic differences observed in adult body size and developmental time between individuals with the 2ε and 2j arrangements (25) and the adaptive effect of the inversion. The CG13617 expression pattern in embryos and the presence of motifs characteristic of transcription factors suggest a role in development consistent with significant fitness effects.

An important aspect in this case is that the silencing effect is not caused by the inversion itself, but by one of the TEs inserted at the breakpoint junctions. TEs have been largely recognized as an important source of genetic variation in the shaping of genomes and the adaptation of organisms to the environment (18, 49). Here, we show that they were not only involved in the origin of the 2j inversion, but also in the regulation of the expression of a gene adjacent to its breakpoints by a mechanism of transcriptional interference (22). Because most inversions in Drosophila and other organisms are probably generated by TEs (15, 16), this kind of position effect may be much more common than previously thought.

We thank J. Hidalgo and J. Carrasco for assistance with the Northern blotting, J. Casanova and M. Llimargas for help with in situ hybridization in embryos, and M. Ashburner, S. Campuzano, J. Modolell, M. L. Pardue, and J. M. Ranz for valuable comments. This work was supported by Dirección General de Investigación, Ministerio de Ciencia y Tecnología (Spain) Grant BMC2002-01706 (to A.R.) and a doctoral fellowship from the Departament d’Universitats, Rererca i Societat de la Informació (Generalitat de Catalunya, Spain; to M.P.).