P Element Domestication: A Stationary Truncated P Element May Encode a 66-kDa Repressor-like Protein in the Drosophila montium Species Subgroup

Danielle Nouaud and Dominique Anxolabèhère

Département Dynamique du Génome et Evolution, Institut Jacques Monod, CNRS–Université Denis Diderot, Paris, France

Functional P transposable elements can be separated into two distinct classes: mobile elements, which present the canonical structure, with transposase and repressor functions, and immobile P sequences truncated in 5' and 3' by loss of the terminal inverted repeats and exon 3, which retain only the repressor function. This second class was first described in some species of the Drosophila obscura group. Here, we describe a new truncated immobile P sequence cloned from one species of the Drosophila montium subgroup (D. tsacasi) that produces a polyadenylated RNA with a coding capacity for a 66-kDa “repressor-like” protein. The results from a number of different comparisons between P-homologous sequences concerning both coding and noncoding regions strongly suggest that the obscura and montium immobile P sequences as well as the T-type P subfamily derive from the same ancestral mobile P element family. Study of the flanking regions of these immobile P sequences shows that the two immobilizations were produced by two independent events. Our results provide evidence that the molecular domestication of a transposable element family may recur in a species lineage.

Introduction

A growing body of evidence (McDonald 1993) suggests that copies of transposable elements can be captured by host control mechanisms and acquire novel functions. Recently, Best et al. (1996) described a new case of the genetic recruitment of a retroviral sequence. They identified a mouse genomic sequence that originally belonged to a transposable element. This sequence had undergone structural modification and had been brought under the influence of transcriptional control elements from other sources. This mouse retrovirus restriction gene FvI acts on murine leukemia virus (MLV) replication and appears to be derived from the gag region of an endogenous retrovirus unrelated to MLV.

Miller et al. (1992, 1995) described the first example of the conversion of a transposable-element-derived coding sequence into a stable integrated host gene in Drosophila. They described this transition of a genomic parasite to a stationary gene useful to the host as “molecular domestication.” Here, we provide evidence that this kind of evolutionary event has occurred independently in species of another Drosophila lineage with the same mobile DNA element: the P-element family.

The P element is a eukaryotic transposable element with a short inverted terminal repeat which transposes through a DNA intermediate (Engels 1989). Originally discovered in Drosophila melanogaster, it has been found in several species of the Drosophilidae family (Clark, Maddison, and Kidwell 1994; Kidwell 1994) and also in other families of Diptera (Anxolabèhère and Periquet 1987; Perkins and Howells 1992). The 2,907-bp P element initially discovered in D. melanogaster will be referred to as the canonical P element. It has 31-bp inverted repeats and generates an 8-bp direct repeat of genomic DNA at the target site. Its four exons encode an 87-kDa transposase (exon 0 to exon 3) (O’Hare and Rubin 1983) and a 66-kDa repressor protein (exon 0 to exon 2) which acts as a repressor of transposition (Karess and Rubin 1984; Rio, Laski, and Rubin 1986; Misra and Rio 1990). The third intron is spliced exclusively in the germline and thus limits transposase synthesis to these tissues.

The characteristics and properties of the P element in D. melanogaster are paradoxical, because only two nucleotide substitutions have been described between copies found in the species (O’Hare et al. 1992), and no copies can be found in old laboratory strains. These observations, taken together with historical and geographical P-element distributions (Anxolabèhère, Kidwell, and Periquet 1988; Daniels et al. 1990), suggest that P elements may have invaded D. melanogaster populations 40–45 years ago after a horizontal transfer (Kidwell 1983). Drosophila willistoni from the Sophophora subgenus is a likely candidate for the role of donor species, since a P element present in its genome differs from the canonical D. melanogaster sequence by only a single nucleotide substitution (Daniels et al. 1990).

The interspecific distribution of P-homologous sequences in the Drosophilidae family shows some phylogenetic discontinuities. P sequences are quite common in the subgenus Sophophora, where they have been detected by genomic hybridizations with the P element from D. melanogaster as a probe (Daniels et al. 1990). However, according to this assay, P-homologous sequences are apparently absent in the closest relatives of D. melanogaster (Brookfield, Montgomery, and Langley 1984). An explanation for this could be that the P elements have never been present in these lineages or that they have been lost from them by chance through speciation events. A more realistic explanation could be that they have diverged so much in their sequences that they can no longer be detected by hybridization techniques using the canonical P probe.
Based on their nucleotide divergence, several P-element subfamilies can be distinguished in Drosophilidae, all of which have the P canonical structural features (i.e., four exons and terminal inverted repeats) in common. As described above, they have been isolated from *D. melanogaster* and *D. willistoni*, and also from *Scaptomyza pallida* (Simonetig and Anxolabèhère 1991, 1994) and *D. bifasciata*, belonging to the obscura group (Hagemann, Miller, and Pinsker 1992, 1994). In addition, Hagemann, Haring, and Pinsker (1996) isolated and partially sequenced a new subfamily from three closely related species *D. ambigua*, *D. obscura*, and *D. tristis*. Although the structure of the P-elements (T-type) of this lineage is identical to that of the canonical P, these copies are defective since none had intact reading frames.

Another type of P sequence which differs in structure from the canonical P element has been identified in the closely related species *D. guanche*, *D. subobscura*, and *D. madeirensis* (obscura group species). The genomes of these species contain tandemly clustered immobile P element derivatives (10 to 50 copies per haploid genome) at a single chromosomal location (Paricio et al. 1991; Miller et al. 1992). The structure of the P-sequence units of these clusters differs from that of the canonical P element by its lack of terminal inverted repeats and its lack of the exon 3 corresponding to the transposase specificity. These repeat units therefore cannot encode the transposase but have retained the coding capacity for a "P repressor-like" 66-kDa protein.

The P-element phylogeny proposed by Clark, Madison, and Kidwell (1994) is highly consistent, but it contradicts the phylogeny of the species. The noncongruence between the phylogenetic tree based on P elements and the tree of their host phylogenetic organisms can be explained by an intraspecies nucleotide divergence of P elements into two or more subfamilies coexisting within descendand lineages until their separation or the extinction of one of them through the speciation events. This explanation requires the existence within a common ancestor genome of different types of active P elements with different rates of evolution (Capy, Anxolabèhère and Lungin 1994). The other, nonexclusive, explanation that is classically proposed is the horizontal transmission of P elements.

Homologous P sequences have been identified outside the Drosophilidae family in three other dipteran lineages: Calliphoridae (Perkins and Howells 1992), Opomyzidae and Trioxocclididae (Anxolabèhère and Periquet 1987). These results favor a long phylogenetic history of P sequences in dipteran lineages but do not exclude horizontal transfers as a recurrent evolutionary event. Under this hypothesis, P sequences would have been present in the ancestor of the Drosophilidae family and could be identified in the genome of modern drosophilids as DNA transposon fossils.

The montium subgroup of species, close to the melanogaster subgroup, has been described as heterogeneous in regard to the distribution of P sequences (Daniels et al. 1990). Of the 26 species tested, 11 showed discernible hybridization to the *D. melanogaster* P element probe. To study the origin of this heterogeneity, we have chosen to investigate several species of this subgroup, and we have identified two types of P sequences. The first is a unique nonmobile P sequence including exon 0 to exon 2, potentially coding a "repressor-like" 66-kDa protein, and located at the same genomic site in all the species of the montium subgroup we studied. In spite of nucleotide distances ranging from 36% to 40% between this immobilized P sequence and the clustered P sequences described in some species of the obscura group (Paricio et al. 1991; Miller et al. 1992), they are more closely related to each other than they are to the mobile P subfamilies. The other type of P sequence identified in the genome of the montium species subgroup is a repeated and dispersed subfamily strongly divergent from the first type of sequence.

Materials and Methods

Fly Stocks Sources

Wild Drosophila stocks of the montium species subgroup were obtained from the National State University, Bowling Green, Ohio, and from the Laboratoire Population, Génétique et Evolution, Gif-sur-Yvette, France.

DNA Amplification

Total genomic DNA was extracted from 30 flies of each species as described previously by Junakovic, Caneva and Balarico (1984). Two oligonucleotide primers were chosen to conserve regions within exon 1 and exon 2 of the P element of *Scaptomyza pallida* (Simonetig and Anxolabèhère 1991). Primers 976-5'-GCA-ACTGCGATTGTCTGGC-3' and 1282-5'GCCGGGTGTCCGAAAAACG-3' were used to amplify the region corresponding to the 306 bp of the P element PS18 of *Scaptomyza pallida*.

Amplification was performed with 2.5 units of AmpliTaq DNA polymerase (Cetus) in 50 µl total reaction volume with 100 ng of genomic DNA and supplier buffer adjusted to 2.5 mM of MgCl₂. The reaction conditions were: 94°C for 5 min and 30 cycles of 94°C for 1 min, 59°C for 1 min, and 72°C for 2 min, followed by 10 min at 72°C.

DNA Hybridization Analysis and Cloning

Genomic DNAs were digested with restriction enzymes according to the manufacturer's instructions. Restriction fragments were separated by electrophoresis in agarose gels, then transferred onto Nitrocellulose membrane (Schleicher and Schuell) according to standard protocols (Maniatis, Fritsch and Sambrook 1982). Orthologous hybridization condition was 35% formamide; 5 x SSC (1 x SSC is 150 mM NaCl and 15 mM sodium citrate, pH 7.0); 50 mM phosphate buffer, pH 7; 0.1% SDS (sodium dodecyl sulfate); 5 x Denhardt (0.1% each of Ficoll-400, polyvinylpyrrolidone, and glycine); and 250 µg/ml of salmon sperm at 37°C. Washing was done once or twice at 40°C for 15 min in 2 x SSC, 0.1% SDS. Homologous hybridization condition was 6 x SSC, 5 x Denhardt, 0.5% SDS, and 150 µg/ml of salmon sperm at 65°C, and washing was done
twice at 65°C in 2 × SSC and 0.1% SDS, followed by once for 15 min at 65°C in 0.2 × SSC and 0.1% SDS. Probes were labeled with 32P using a random primed kit (Boehringer Mannheim).

Genomic DNA was digested by appropriate restriction enzymes and ligated with the Bluescript KS+ plasmid previously digested by corresponding restriction enzymes. In order to obtain the highest quantity of transformed bacterial colonies allowing high numbers of possible clones, the ligation product was introduced into bacteria using the highly competent cell transformation protocol of Hanahan (1983). The transformed cells were probed by the PCR product from a D. tsacasi template.

Sequence Determination

PCR products were separated by electrophoresis on a low-melting agarose gel. The gel band was cut off, the DNA was denatured in gel for 5 min in a boiling bath, and the sequence was carried out on 7 μl of DNA-gel mixture using the Sequenase version 2.0 kit (USB, Cleveland, Ohio). For each species, the PCR product was sequenced in both strands.

Sequences were made according to the protocols of the Sequenase kit.

In Situ Hybridization

Chromosome spreads were obtained from polytene chromosomes from the salivary glands of third-instar larvae. Hybridization was performed at 37°C in a mixture containing 10 ng/μl of the subcloned restriction fragment xp14 (see fig. 4A), biotin-16-dUTP-labeled by nick translation as probe, 50% formamide, 2 × SSC, 0.5 M NaCl, 0.1% SDS, and 10% Dextran sulfate. Detection was achieved with the extravidin-peroxidase system.

RNA Isolation and Northern Analysis

Total RNA was isolated from adults of D. tsacasi using RNAzol reagent (Bioprobe system). Poly(A)+ RNA was purified through an oligo(T) column and separated by electrophoresis in 1.3% agarose/formaldehyde gel and transferred onto nitrocellulose membrane.

Sequences Analysis

Phylogenetic analyses were carried out on 12 P sequences (table 1), 1 from Drosophila melanogaster (O’Hare and Rubin 1983), 2 from Scaptomyza pallida species (Simonelig and Anxolabéhère 1991), and 9 from the obscura species group, including 2 from Drosophila bifasciata (Hagemann, Miller, and Pinsker 1992), 2 from Drosophila guanche (Miller et al. 1992), 1 from Drosophila madeirensis (Paricio et al. 1996), 3 from Drosophila subtropica (Paricio et al. 1991), and 1 from Drosophila ambigua (Hagemann, Haring, and Pinsker 1996).

Alignments were made using the BESTFIT and PILEUP programs of the GCG (Madison, Wis.) software package (1990) using the ordinary options.

Relationships among nucleotide sequences were inferred from the GCG alignment by using PHYLIP version 3.5c (Felsenstein 1993). Kimura’s two-parameter model for nucleotide differences was used to generate a corrected pairwise distance matrix with the transition/transversion ratio set at 2.0 in PHYLIP. Phylogenetic analysis was performed using neighbor-joining and parsimony methods following the procedures indicated in the text.

Results

Detection and Characterization of P Sequences in Species of the montium Subgroup

Oligonucleotide primers were chosen from conserved regions based on the available published P sequences (fig. 1A). The primers were chosen from the sequence of the P element of Scaptomyza pallida (PS18) at position 976 for the 5' primer and at position 1282 for the 3' primer. They amplify a region of 306 bp characteristic of the functional P element: it crosses the intron 1-2, and its corresponding protein presents a helix-turn-helix motif. This DNA sequence is present in the mobile P sequences found in different groups of species (willistoni, saltans, scaptomiza, and obscura) as well as in the immobilized P sequences so far described in the obscura group (Miller et al. 1992, 1995).

Amplification was performed on 10 species belonging to the montium subgroup: D. bakoue, D. boqueti, D. burlai, D. davi, D. kikkawai, D. malagasia, D. nikananu, D. serrata, D. tsacasi, and D. vulcana. PCR products of the expected 0.3-kb size were obtained with the template DNA from each species examined. All of these bands hybridize with the P-specific probe (PS18 P element), strongly suggesting that they correspond to P sequence amplification products (data not shown).

The PCR products from D. davi, D. kikkawai, and D. tsacasi were sequenced. The sequencing was performed directly on the PCR products from both directions with the same primers used for amplifications. The quality of the sequences was unambiguous, suggesting that in each case the genomic template was present as a unique sequence or as a series of very repetitive sequences identical to each other. The P sequences were compared to the sequence of the corresponding region of the canonical P element of D. melanogaster. The se-
sequences were aligned by PILEUP multiple alignment. Table 2 shows the genetic distances for all pairs of sequences, above the diagonal at the nucleotide level and below the diagonal at the proteic level. A first analysis of the nucleotide sequences of the montium species demonstrates that they diverge from each other (12.9%–19.5%) but less than each of them diverges from the D. melanogaster P element (38.5%–42.5%). The splicing signals for the intron of this region are conserved and located at positions homologous to those in the D. melanogaster canonical P element (fig. 2). The intronic sequence sizes are 58 bp in D. tsacasi and D. kikkawai, 57 bp in D. davidi, and only 54 bp in D. melanogaster. The sequences from the three montium species have one more triplet at the beginning of the corresponding region of exon 2, which does not disrupt the reading frame. As shown in figure 2, the leucine zipper motif as well as the helix-turn-helix motifs, is well conserved. These results suggest that the P sequence detected in each of these three species may be functional.

Genomic Distribution of P Sequences

Genomic hybridization experiments were performed for the three species D. davidi, D. kikkawai, and D. tsacasi, each of them probed by their own 0.3-kb PCR product. As shown in figure 3, after 20 h exposure, single positive signals at 3 kb and 5 kb are detected, respectively, in D. tsacasi and D. kikkawai. In D. davidi, several positive bands are present; the one at 2.8 kb is the strongest. After a longer time of exposure (60 h), several weak positive bands are detected in D. kikkawai and D. tsacasi (fig. 3), suggesting the presence of more divergent repeated P sequences in the genomes of these species. These P repetitive sequences might concern another P family. However, at least in D. tsacasi, such bands might be merely cases of fortuitous weak hybridization with some unrelated genomic sequences.

Cloning and Analysis of the P Sequence of D. tsacasi (P-tsa)

Based on the genomic restriction map built using the 0.3-kb PCR product of D. tsacasi as a probe (data not shown), two clones were successively isolated using the same probe from two genomic libraries, one digested with Pst I and the other digested with Cla I. The P-tSA-P clone is the Pst I restriction fragment of 3 kb.
The pTSA-C clone is the Cla I restriction fragment of 2.5 kb. They overlap (1.5 kb) since the two clones have been picked up by the same probe.

The two clones pTSA-P and pTSA-C were completely sequenced, providing a 4.5 kb genomic sequence. The whole P sequence is included in this 4.5 kb (fig. 4A). The restriction map deduced from this sequence is consistent with the multiple-digestion experiment performed on genomic DNA, which shows that only one P sequence is contained in this 4.5 kb (fig. 4A).

The P-homologous region was aligned to the sequence of the D. melanogaster canonical P element. After alignment by the BESTFIT program, the region of the P sequence is located between positions 1088 and 3046. Upstream and downstream sequences do not correspond to the canonical P sequence (fig. 4B). Therefore, this region does not contain tandemly repeated P sequences unless the intersequence is longer than 2.5 kb.

Chromosomal Localization

In situ hybridization on D. tsacasi polytene chromosomes using the Xho I/Pst I subcloned restriction fragment of pTSA-P as probe gave a single positive signal. This signal was localized at a euchromatic site on the first third of a chromosomal arm (fig. 5). This result shows that the P-homologous sequence under analysis in D. tsacasi does not belong to a dispersed repetitive P-element family.

The P Sequence of D. tsacasi is Transcribed

Transcriptional activity of the P-isa sequence was investigated by a northern blot performed on adult poly(A)^+ RNA with a riboprobe obtained from the subcloned restriction fragment Xp14 as defined in figure 4A. The probe was synthesized using T7 RNA polymerase and labeled with [32P]UTP. As shown in figure 6, a 2.1-kb transcript is detected. The observed size matches the length of a spliced and polyadenylated transcript beginning at about 100 bp upstream of the putative ATG at position 1208, comprising exons 0-2 (1,722 bp), with an untranslated 3´ sequence. In this 3´ sequence, a perfect polyadenylation signal (AATAAA) is located at position 3167, 121 bp downstream from the stop codon at the end of exon 2.

Sequence Data

Interspecific comparisons between P homologous sequence included in the 4.5-kb genomic region cloned from D. tsacasi and five transcriptionally active P sequences from D. melanogaster, S. pallida, D. guanche, and D. bifasciata are shown in table 3. Only the region including exon 0 to exon 2 (positions 1208-3049 in fig. 4B) significantly matches the corresponding regions of the other P sequences. The highest identity values are obtained with the A1 P element from D. guanche. Despite the presence of 1,207 bp of sequence upstream of the presumed ATG that begins exon 0 and 1,446 bp downstream of the end of exon 2, no significant similarity was detected that could delimit homologous P sequences in these regions. In the canonical P element, the 31-bp perfect inverted terminal repeats lie within 200 bp of the start and the end of the transcription unit. In P-isa, the apparent existence of a "skeleton" of the 5´ inverted terminal repeat can be detected when the first 152 bp of the P element of D. melanogaster are aligned with the sequence upstream of the ATG of exon 0 (fig. 4C). Ten base pairs of the 16 outer 31-bp terminal inverted repeats are similar. This 16 bp corresponds to the nucleotide site which interacts with a cellular non-P-element-encoded DNA-binding protein (IRBP) (Rio and Rubin 1988; Beall and Rio 1996). Furthermore, with this alignment, the 5' copy of the subterminal inverted repeats (Engels 1989) is found to have a good similarity (10 out of 11 bp). In D. melanogaster, these repeats, although not strictly necessary, appear to facilitate transposition (Mullins, Rio, and Rubin 1989). However, neither the P promoter nor the transposase-binding site as defined by Kaufman, Doll, and Rio (1989) can be iden-
Fig. 4.—Molecular analysis of the genomic region of D. tsacasi including P-tsa. A. Cloned and subcloned fragments and partial restriction map. B. Molecular structure of P-tsa. C. BESTFIT alignment of P upstream noncoding region between P sequences of D. tsacasi and Drosophila melanogaster. Underlined arrow: subterminal inverted repeat; light gray box: IRBP; dashed line: transposase binding site; box: promoter region; broken arrow: start of transcription; underline: internal inverted repeat (see text).

Fig. 5.—Picture of in situ hybridization of biotinylated subcloned xp14 DNA (see fig. 4A) to salivary gland polytene chromosomes of D. tsacasi. Only one site per genome is detected.
Table 3
Interspecific P Sequence Comparisons at the DNA Level

<table>
<thead>
<tr>
<th>NO. OF BP</th>
<th>Identity with the P Sequence in D. tsacasi (P-tsa)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>pπ25.1</td>
</tr>
<tr>
<td>IR . . . . .</td>
<td>a</td>
</tr>
<tr>
<td>5’ NCR . .</td>
<td>a</td>
</tr>
<tr>
<td>Exon 0 . .</td>
<td>289</td>
</tr>
<tr>
<td>Intron 1 . .</td>
<td>55</td>
</tr>
<tr>
<td>Exon 1 . .</td>
<td>679</td>
</tr>
<tr>
<td>Intron 2 . .</td>
<td>57</td>
</tr>
<tr>
<td>Exon 2 . .</td>
<td>754</td>
</tr>
<tr>
<td>Intron 3 . .</td>
<td>ab</td>
</tr>
<tr>
<td>Exon 3 . .</td>
<td>ab</td>
</tr>
<tr>
<td>3’ NCR . .</td>
<td>ab</td>
</tr>
<tr>
<td>IR . . . . .</td>
<td>ab</td>
</tr>
</tbody>
</table>

Note:—Identities for various subregions between the P homolog of D. tsacasi (P-tsa) and five different P elements: pπ25.1, the canonical P element of D. melanogaster, the PS18 element of S. pallida, the A1 element of D. suauca, and the M-type and O-type elements of D. bifasciata.

*See text and figure 4C.

The absence of any significant identity means that these P regions cannot be distinguished.

mains are found to be highly conserved. The helix-turn-helix motif, near the ORF1-ORF2 boundary, which corresponds to a potential DNA-binding domain (Rio, Laski, and Rubin 1986), is found in P-tsa with 75% identity when compared with the D. melanogaster P element.

Three leucine zipper motifs, which could be involved in transposase dimerization, have been described in the canonical P element (Rio 1990). In the P-tsa putative protein (fig. 7), two of these are well conserved (amino acids 287–315 in ORF1 and 502–530 in ORF2). The third is located at the beginning of ORF1 (amino acids 105–126), but its accordance with the leucine zipper motif is weaker, because three leucine positions are replaced by other hydrophobic amino acids. However, the conservation of the helix-turn-helix and leucine zipper motifs in P-tsa reinforces the hypothesis of that this sequence is functional.

The same interspecific nucleotide comparisons presented in table 3 are presented in table 4 with the putative amino acid sequences. For each P-tsa exon the similarity value with the corresponding exon of mobile or immobile P elements is high (64.9%–80%). Exon 2 presents the highest similarity and exon 0 the lowest, except when these comparisons are made with the D. bifasciata O-type element. However, on their own, the similarity values corresponding to the entire putative proteins do not allow us to say that the P-tsa product is closer to any one of the five others. It should be noted that all the nucleotide identity values (table 3) are systematically higher than the amino acid identities (values in parentheses in table 4). This could be due to a long-term evolution of sequences that have undergone similar selective constraints to conserve the same function.

The question of the phylogenetic relationship of P-tsa was investigated among the clustered P sequences of the obscura group and among the corresponding region of mobile P elements. Table 5 shows the nucleotide and protein matrices generated by Kimura’s correspond-
**FIG. 7**—PILEUP multiple alignment of the repressor-deduced protein from the mobile and immobilized *Drosophila*.

<table>
<thead>
<tr>
<th></th>
<th>Spa12</th>
<th>Dmel</th>
<th>DbifO</th>
<th>DbifM</th>
<th>DsubG2</th>
<th>DguaAl</th>
<th>DsubA2</th>
<th>Dtsa</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>mRNA</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Cons</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Consensus Line (Cons) are shown by large dots, and hydrophobic amino acids corresponding to leucine zipper motifs, as defined by Rio (1990), noted by code name (see table 1). Identities with the top sequence are shown by small dots, gaps are shown by dashes, differences in the alignments are shown by down arrows. The helix-turn-helix motif is boxed.**

---

<table>
<thead>
<tr>
<th><strong>Legend</strong></th>
<th><strong>Description</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Cons</td>
<td>Consensus Line</td>
</tr>
<tr>
<td>Spa12</td>
<td>Spa12 mRNA</td>
</tr>
<tr>
<td>Dmel</td>
<td>Dmel mRNA</td>
</tr>
<tr>
<td>DbifO</td>
<td>DbifO mRNA</td>
</tr>
<tr>
<td>DbifM</td>
<td>DbifM mRNA</td>
</tr>
<tr>
<td>DsubG2</td>
<td>DsubG2 mRNA</td>
</tr>
<tr>
<td>DguaAl</td>
<td>DguaAl mRNA</td>
</tr>
<tr>
<td>DsubA2</td>
<td>DsubA2 mRNA</td>
</tr>
<tr>
<td>Dtsa</td>
<td>Dtsa mRNA</td>
</tr>
</tbody>
</table>

---

**Note:**
- **Consensus Line (Cons):** Consensus line in the alignment.
- **Spa12, Dmel, DbifO, DbifM, DsubG2, DguaAl, DsubA2, Dtsa:** Name of the mRNA sequences.
- **Identities with the top sequence are shown by small dots:** Identity points in the alignment.
- **Gaps are shown by dashes:** Gaps in the alignment.
- **Differences in the alignments are shown by down arrows:** Differences in the alignment.
- **The helix-turn-helix motif is boxed:** Highlighted motif in the alignment.
The Immobilized P Sequence of D. tsacasi Occurs at Orthologous Chromosomal Sites in Other Species of the montium Subgroup

Southern blot experiments were performed with genomic DNA from eight other species belonging to the montium subgroup (D. bakoue, D. bocqueti, D. burlai, D. davidi, D. kikkawai, D. malagasta, D. serrata, and D. vulkana). DNA samples were digested with PstI endonuclease, and the restriction fragments were blotted on nitrocellulose membranes. One filter was probed with the inner XP14 restriction fragment of D. tsacasi as a probe, the filter corresponding to the figure 8B experiment was rehybridized under orthologous conditions. The same picture is found with D. bocqueti, since the two positive restriction fragments seen with the xp14 probe are also detected with the c09 probe; a third 0.5-kb restriction fragment is detected only with the c09 probe (not visible here on the shortened autoradiography shown in fig. 8B). They correspond to an incomplete hydrolysis and the presence of a PstI restriction site inside the genomic sequence covered by the c09 probe. For D. malagasta, D. bakoue, and D. vulkana, the restriction fragments that are positive with the xp14 probe are not positive with the c09 probe, but these two distinct fragments were detected simultaneously on the same filter when the two were rehybridized with the PvuII/ClaI restriction fragment as probed (pc05 in fig. 4A) (data not shown). These results mean that a PstI restriction site occurs in these species inside the genomic sequence corresponding to the pc05 probe. Thus, in each of these three species, a K restriction homologous sequence is also detected at the orthologous genomic site of D. tsacasi.

The Genomic Sites of Immobilized P Sequences of D. tsacasi and D. guanche Are Not Orthologous

Using the 2.5-kb 5' genomic flanking sequence of the D. guanche P cluster (kindly provided by W. Miller) as a probe, the filter corresponding to the figure 8B experiment was rehybridized under orthologous conditions. Despite the evolutionary distance between the species of Drosophila and Scaptomyza, four restriction elements of the P-tsacasi P 'repressor-like' protein and the five corresponding subregions of the pr25.1 element of D. melanogaster, the PS18 element of Scaptomyza pallida, the M-type and O-type elements of D. bifasciata, and the DguaA1 element of D. guanche. The values in parentheses are percent identity.

<table>
<thead>
<tr>
<th>Exons</th>
<th>Similarity and Identity (%)</th>
<th>MOBILES</th>
<th>IMMUMBLES</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exon 0</td>
<td>67.0 (56.7)</td>
<td>64.9 (57.7)</td>
<td>69.1 (61.9)</td>
</tr>
<tr>
<td>Exon 1</td>
<td>73.1 (57.7)</td>
<td>73.1 (58.7)</td>
<td>72.5 (59.5)</td>
</tr>
<tr>
<td>Exon 2</td>
<td>80.0 (63.6)</td>
<td>77.9 (62.2)</td>
<td>79.1 (65.1)</td>
</tr>
</tbody>
</table>

NOTE.—Identities for subcoding regions between D. tsacasi P 'repressor-like' protein and the five corresponding subregions of the pr25.1 element of D. melanogaster, the PS18 element of Scaptomyza pallida, the M-type and O-type elements of D. bifasciata, and the DguaA1 element of D. guanche. The values in parentheses are percent identity.
montium subgroup of species and D. guanche, a positive signal was detected after a long exposure time (7 days). When these hybridization signals (fig. 8C) were compared with those of the previous hybridization with xpi4 as a probe (fig. 8A) and the mirror filter hybridized with c09 (fig. 8B), no common positive signal was found. This result strongly suggests that the immobilized P sequences in both species do not occur at the same genomic site.

Discussion

A new type of P sequence was discovered in the genomes of different Drosophila species belonging to the montium subgroup. It is present as a single copy and is located at the same genomic site in the nine species investigated, indicating that this sequence was already present in the ancestor of the montium subgroup of species. In contrast to the canonical P element of D. melanogaster, the cloned sequence of D. tsacasi is devoid of inverted repeats and appears to be truncated at the 3' region resulting in the absence of exon 3. Nevertheless, it appears to code for a putative 66-kDa repressor-like protein. This P sequence thus presents the same structural characteristics as the clustered truncated P sequences previously described by Miller et al. (1992) and by Paricio et al. (1991) in the genomes of some species of the obscura group. The pairwise comparisons at the DNA and amino acid levels (table 5) exhibit a significant divergence of P-tsa with other previously described P elements, indicating that this sequence is a new type of P sequence. To determine its relationships with the other P sequences, figure 9 shows a dendrogram generated by the neighbor-joining method (Saitou and Nei 1987) and based on the amino acid sequence of the repressor-like proteins encoded by the mobile and immobile P sequences studied here. The dendrogram shows two clusters, one corresponding to the P repressor-like proteins encoded by the mobile P elements, the other corresponding to the repressor-like proteins encoded by immobile, tandemly repeated P sequences in the obscura group. The divergences between the repressor-like proteins encoded by P-tsa and each of these two clusters are approximately the same, implying that the immobilized P sequence in D. tsacasi represents a new type of P sequence.

This stationary P sequence, behaving as a functional gene and identified in several Drosophila species as belonging to the montium subgroup, may represent an example of a sequence that originally belonged to a
transposable-element family but has lost mobility, perhaps by integration of an intact element followed by deletion or degeneration of sequences on either side. An alternative hypothesis for the origin of the P-tsa might be that it is an ancient genomic sequence which was at the origin of mobile P elements, but this hypothesis is less parsimonious than the immobilized P sequence hypothesis, since it requires the addition of the fourth exon and the terminal inverted repeats in mobile elements. Moreover, the "ancient-gene" hypothesis also supposes that the transposase would have been generated despite the presence of its repressor.

To infer the phylogenetic relationships of the P-tsa sequence under the immobilized P sequence hypothesis, a parsimony analysis was performed with the mobile and immobilized nucleotide sequences (fig. 10). We included the clustered immobilized P sequences previously described in the obscura group and the four full-sized P-element subfamilies currently known: the P element from D. melanogaster or D. willistoni, the M-type from S. pallida or D. bifasciata, the O-type from D. bifasciata, and the T-type, a recently described P element subfamily from three closely related species: D. ambigua, D. obscura, and D. tristis (Hagemann, Haring, and Pinsker 1996). The parsimony analysis was restricted to exon 0 because the complete T-type nucleotide sequences of exons 1 and 2 are not yet available. In the dendogram, the D. ambigua T-type element appears to be the closest relative of the truncated P homolog of D. tsacasi, and these two sequences cluster with the truncated P sequence present in D. guanche and D. subobscura. The T-type subfamily might represent a P element with a long evolutionary history, not only in the obscura group, as previously proposed by Hagemann, Haring, and Pinsker (1996), but also at the very least in the radiation leading to the obscura and melanogaster groups. It could be the common mobile ancestor of the immobilized and transcriptionally active P sequences in the obscura species group and in the montium species subgroup. This result, together with the fact that the genomic sites of immobilized P sequences of D. tsacasi and D. guanche are nonorthologous, strongly suggests that the T-type P subfamily underwent two independent events of domestication. However, based on the present data, it is not possible to deduce the chronology of the immobilization events at the origin of these two P-element domestication events. One event has been identified in the obscura lineage and the other in the montium lineage, but both could have occurred before the split of the obscura and melanogaster groups. The analysis of the 5' and 3' noncoding regions of P-tsa as well as the homologous sections of exon 3 shows a more degenerated structure than that of D. guanche, indicating that the stationary montium P element might be the older one. But the similarity in the 3' region is so low that a deletion event in this region cannot be excluded.

The analysis of the cis-acting sequence required for germline-specific splicing of the D. melanogaster P element ORF2-ORF3 intron (Chain et al. 1991) enable us to investigate the relative dating of these two immobilization events. This sequence is located in exon 2, 12 to 31 bases from the 5' splice site. They match strongly in DsubA1, DsubA5, and DguA1 (18 out of 20 bp) but cannot be identified in DsubG2 (8 out of 20 bp) and can be identified weakly or not at all in P-tsa (13 out of 20 bp). Two hypotheses could explain these observations: either this cis-acting regulatory sequence must have been functional in some immobilized copies but not in others, or these differences are due to the time of immobilization. According to the latter hypothesis, the P-tsa immobilization might predate that of obscura. Nevertheless, the sequence analyses obtained from the repressor-like coding region suggest that the D. guanche P elements as well as the D. tsacasi element belong to the same ancient subfamily of previously active P elements.

On the other hand, the P-tsa sequence could be derived from a type of P subfamily other than the T-type, and its position inside the immobilized P sequence group could result from convergent evolution between sequences undergoing specific mutation rates and specific selective constraints distinct from those undergone by mobile P elements. Whatever the origin of the immobilized P sequences, what selective constraints are responsible for the independent selective conservation of these "repressor" genes in those two branches during evolution? A possible explanation of this conservation can be proposed based on the present situation found in D. melanogaster. It is known that P elements with different structures are involved in P regulation and that the regulatory properties of an element are strongly dependent on its insertion site. For example, Ronsseray, Lehmann, and Anxolabèhère (1991) have shown that a pair of full-length P elements inserted at the telomere of the X chromosome have a complete P repression capacity. More recently, Ronsseray et al. (1996) found evidence that a single full-length element at this chromosomal location is sufficient to repress many autonomous P elements dispersed through the genome. If there is a mutation in the 5' and 3' regions implicated in the transposition mechanism, it will lead to the immobilization of the regulatory element. Natural selection will retain this type of event, which counteracts
the possible loss of the regulatory site. Thus, the regions not involved in repressor production will be not constrained, and the accumulation of mutations will decrease the similarity to the corresponding regions of the autonomous element. This would mean that a single P sequence producing the repressor protein stops the transposition of other P elements even if these were introduced by horizontal transfer. The P elements, the activity of which is repressed by the regulatory element, might lose their functionality, since they do not transpose and no selective pressure controls the accumulation of mutations. Constraint due to the effect of transposition events thus no longer acts on the regulatory element. Moreover, the probable low frequency of horizontal transfers does not represent a selective pressure sufficient to maintain the repressor gene in a functional state. It should thus degenerate at a rate similar to that of the immobi-

The evolutionary life cycle of a distinct P subfamily can therefore be hypothesized. The P-tsa gene might have been present in D. tsacast, and at least in the montium lineage, for a very long time. Its former function was to repress the transposition of mobile elements belonging to a P-element subfamily such as the one which corresponds to the repeated P sequences detected in the genomic hybridization experiment presented in figure 3. Secondarily, it has undergone a molecular domestication.

Acknowledgments

We thank H. Quesneville and S. Ronsseray for helpful discussions and comments on the manuscript. We are particularly grateful to W. Miller for stimulating discussions. This research was supported by the Centre National de la Recherche Scientifique, by a grant from the Groupement de la Recherche sur les Genomes and by ACC SV3 no. 9503079 from MESR.

References


MILLER, W. J., S. HAGEMANN, E. REITER, and W. PINSKER. 1992. P element homologous sequences are tandemly re-


Daniel L. Hartl, reviewing editor

Accepted August 5, 1997