Molecular structure of a gypsy element of *Drosophila subobscura* (gypsyDs) constituting a degenerate form of insect retroviruses

Trinidad M. Alberola* and Rosa de Frutos

Departament de Genètica, Universitat de València, Dr. Moliner 50, 46100 Burjassot, Valencia, Spain

Received October 31, 1995; Revised and Accepted January 9, 1996 EMBL accession no. X72390

**ABSTRACT**

We have determined the nucleotide sequence of a 7.5 kb full-size gypsy element from *Drosophila subobscura* strain H-271. Comparative analyses were carried out on the sequence and molecular structure of gypsy elements of *D. subobscura* (gypsyDs), *D. melanogaster* (gypsyDm) and *D. virilis* (gypsyDv). The three elements show a structure that maintains a common mechanism of expression. ORF1 and ORF2 show typical motifs of gag and pol genes respectively in the three gypsy elements and could encode functional proteins necessary for intracellular expansion. In the three ORF1 proteins an arginine-rich region was found which could constitute a RNA binding motif. The main differences among the gypsy elements are found in ORF3 (env-like gene); gypsyDm encodes functional env proteins, whereas gypsyDs and gypsyDv ORF3s lack some motifs essential for functionality of this protein. On the basis of these results, while gypsyDm is the first insect retrovirus described, gypsyDs and gypsyDv could constitute degenerate forms of these retroviruses. In this context, we have found some evidence that gypsyDm could have recently infected some *D. subobscura* strains. Comparative analyses of divergence and phylogenetic relationships of gypsy elements indicate that the gypsy elements belonging to species of different subgenera (gypsyDs and gypsyDv) are closer than gypsy elements of species belonging to the same subgenus (gypsyDs and gypsyDm). These data are congruent with horizontal transfer of gypsy elements among different *Drosophila* spp.

**INTRODUCTION**

Gypsy-like elements constitute a group of retrotransposons which includes elements from heterogeneous species: gypsy, 17,6, 297 and 412, from *Drosophila melanogaster* (1–5), tom from *D. ananassae* (6), Ulysses from *D. virilis* (7), micropia from *D. hydei* (8), TED from the lepidopteran Trichoplusia ni (9,10), del1 and IFG7 from plants (11–13), Ty3 and Tf1 from yeast (14–16), Cfl1 from fungi (17) and SURL from a marine invertebrate (18). According to phylogenetic trees based on reverse transcriptase amino acid sequences, this group of retroelements clusters with vertebrate retroviruses and caulimoviruses (12,13,19,20). They also have a pol gene functional domain order which is similar to that of retroviruses [PR (protease), RT (reverse transcriptase), RH (RNase H) and EN (endonuclease)] and some of the *Drosophila* elements (gypsy, 17,6, tom and 297) exhibit a third ORF of similar size and location to the env protein of retroviruses (2,3,5,6,21). Early functional analyses from ORF3 of gypsy-like elements seemed to indicate that these ORF3 encode non-functional env proteins and consequently gypsy-like sequences lacked the ability to be infectious particles. Nevertheless, it has recently been established that gypsy ORF3 from *D. melanogaster* encodes a fully functional protein which accomplishes the typical functions of a retroviral envelope protein and, consequently, the authors consider the gypsy element as the first insect retrovirus described (22,23). Supporting this hypothesis, it has been found that gypsy sequences can be horizontally transferred by feeding (24). We wonder if this remarkable infectious ability of the gypsy element of *D. melanogaster* (referred to in this paper as gypsyDm) is shared by other members of the gypsy family.

Gypsy sequences homologous to that of *D. melanogaster* are widespread among *Drosophila* species, also occurring in the *copia* and 1731 retrotransposons (25–27). The distribution of gypsy elements is consistent with the hypothesis that these sequences are of ancient origin. They were probably present in a common ancestor before early *Drosophila* radiation and subsequently transmitted vertically. However, the complex phylogenetic relationships of gypsy-like sequences among *Drosophila* and distant, unrelated species suggests that horizontal transfer of gypsy sequences between major taxonomic groups has contributed to the generation of the present gypsy family. As cited above, recent data strongly support consideration of the gypsyDm element as an infectious particle (22–24). If this ability is a general feature of gypsy elements, it is obvious that the spreading of gypsy sequences throughout populations and/or species could be faster than that expected of true retrotransposons. From this point of view it would seem interesting to study the evolutionary history of gypsy sequences amongst *Drosophila* species. To date, only the gypsy element from *D. virilis* (in this paper gypsyDv) has been sequenced, which showed a similar structure to gypsyDm (28). In a previous paper we presented preliminary data on the sequence and molecular structure of gypsy elements from *D. subobscura* (gypsyDs), a species belonging to the *obscura* group. We found that the gypsyDs sequence of *D. subobscura* is closer to that of gypsyDv than that of gypsyDm, which is not consistent with the phylogenetic

* To whom correspondence should be addressed
relationships between these species (29). In the present paper we report a comparative analysis of the sequence and molecular structures of the gypsyD\textsubscript{s}, gypsyD\textsubscript{m} and gypsyD\textsubscript{v} elements. From this analysis it could be deduced that while gypsyD\textsubscript{m} shows the potential ability to be an infectious particle, gypsyD\textsubscript{s} and gypsyD\textsubscript{v} have lost this ability. The gypsyD\textsubscript{s} and gypsyD\textsubscript{v} elements are probably degenerate forms of ancestral infective particles. From molecular characterization of PCR products we found gypsy elements identical to those of gypsyD\textsubscript{m} in different D. subobscura natural populations, co-existing with gypsyD\textsubscript{m} elements. All these data indicate that D. subobscura natural populations could have been invaded by infectious gypsyD\textsubscript{m} elements.

**MATERIALS AND METHODS**

Isolation, cloning and DNA sequencing of gypsyD\textsubscript{s}

Gypsy-homologous sequences were isolated from a genomic DNA library, kindly provided by R. González-Duarte and G. Marfany, constructed by partial MboI digestion and ligation of D. subobscura strain H271 DNA into the λ EMBL4 phage vector. A 7.0 kb XhoI–XhoI fragment of gypsyD\textsubscript{m} was used as a probe and we followed the procedures of Kaiser and Murray and Benton and Davis (30,31). Twenty eight positive clones were identified, of which at least three contained full-length elements. From these, clone DsE1 was selected for further analyses. Cloning of restriction fragments of this clone was carried out in pUC18/19 by standard procedures (32,33).

The nucleotide sequence of clone DsE1 was determined for both strands by the dideoxynucleotide chain termination method (34,35) using \[^{35}\text{S}\]dATP and the T7 DNA polymerase sequencing kit (Pharmacia). Gaps in the sequence were filled in with nested deletions, using synthetic oligonucleotides as primers for the sequencing reactions.

Chromosomal location of the sequenced gypsyD\textsubscript{s}

In situ hybridization was performed on larval salivary gland chromosomes of D. subobscura strain H271 following the procedures described in Terol et al. (36). As a probe we used a \[^{3}H\] labelled 3 kb PstI–PstI fragment of genomic DNA from the DsE1 clone.

Detection of gypsyD\textsubscript{s} mRNA

mRNA from adult flies of D. subobscura was extracted using the guanidinium thiocyanate method (37). The poly(A)\textsuperscript{+} fraction was purified by affinity chromatography through oligo(dT)–cellulose spin columns (Pharmacia), according to the manufacturer’s protocol. Electrophoresis of a 3 µg RNA sample and Northern blots were carried out as described in Sambrook et al. (33) for nylon filters. Hybridization was performed at 50°C in 7% SDS, 5× SSC, 50 mM phosphate buffer, 50% formamide, 2% blocking reagent, 0.1% laurylsarcosine and 50 µg/ml yeast tRNA. The 6.9 kb XhoI–XhoI fragment of gypsyD\textsubscript{s} was used as the probe. Washing conditions were 1× SSC, 0.1% SDS at 50°C. Detection was performed using the DIG chemiluminescent method, following the manufacturer’s protocol (Boehringer-Mannheim).

Sequence and phylogenetic analyses

Multiple alignments of gypsy sequences from different species were performed using the CLUSTAL program (38). Sequence analyses and phylogenetic constructions were determined using the PHYLIP 3.4 package programs.

Southern blot and PCR analyses of D. subobscura strains genomic DNA

Southern blots of different D. subobscura strains (H-271, Finland; SN, Sweden; TU, Germany; DIE, Switzerland; CJ, USA; PC, Canada; MA, Madeira; RA, Canary Islands; MAR, Morocco) were performed using standard procedures. These strains came from natural populations maintained for some years in the laboratory. A sample of 5 µg genomic DNA of different D. subobscura strains quantified by spectrophotometry were digested with XhoI, EcoRI or KpnI. Hyridizations were performed at 65°C with 0.02% SDS, 5× SSC, 0.5% blocking reagent, 0.1% laurylsarcosine using two probes, the XhoI–XhoI fragments of gypsyD\textsubscript{s} and gypsyDm. Washes were at 65°C in 0.5× SSC, 0.1% SDS when the gypsyD\textsubscript{s} probe was used and at 45°C in 1× SSC, 0.1% SDS when the gypsyD\textsubscript{m} probe was used. DIG colour detection was performed using Boehringer Mannheim standard procedures.

PCR analyses of the gypsyD\textsubscript{s} sequences were carried out on the genomic DNA of the TU and MAR strains. Specific oligonucleotides were designed following the canonical gypsyD\textsubscript{m} sequence (3) and selected from the most variable regions in comparison with the

---

Figure 1. In situ hybridization on the polytene chromosomes of D. subobscura using different fragments of DsE1 clone as probes. (A) Signal found at the border of the 35AB puff in strain H-271 with a 3 kb PstI–PstI genomic DNA fragment adjacent to the sequenced gypsyD\textsubscript{s} element as a probe. (B) Signal found inside the 35AB puff in the SN strain with the 6.9 kb XhoI–XhoI fragment of gypsyD\textsubscript{s} as a probe.
Figure 2. (A) Comparisons of main sequences included in the LTRs and adjacent regions among gypsyDm, gypsyDs and gypsyDv elements. In each motif of the LTRs positions corresponding to the 5′ LTR are indicated. (B) Multiple alignment of the 5′ regulatory region corresponding to gypsyDm, gypsyDs and gypsyDv. Poly(A) blocks are indicated by discontinuous boxes, palindromes of gypsyDm and gypsyDs and (GTAAA) 3  of gypsyDv by black boxes and 12 bp repeats by shaded boxes. Two major 109 bp repeats are only present in gypsyDm and they are indicated by lines. Asterisks correspond to identical positions in the three sequences.

Amplifications were in a Perkin Elmer GeneAmp PCR System 2400 cycler with a denaturing step at 94°C for 5 min and the step cycle program set for 40 cycles (with a cycle consisting of denaturation at 94°C for 30 s, annealing at 55°C for 30 s and extension at 72°C for 30 s), followed by an extra extension step at 72°C for 7 min.

RESULTS

Structure of the gypsyDs element and comparisons with gypsyDm and gypsyDv

The genomic library of D. subobscura strain H271 was screened under moderate stringency conditions using the 7.0 kb XhoI–XhoI fragment of the gypsyDm element as probe. Of the 28 positive clones obtained, 16 were used to restrict analysis, which revealed that at least three of them contained putative full-length gypsy elements sequences in the D. subobscura genome. The 5′ position and the sequence of the primers used in the reactions were as follows:

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>GM001</td>
<td>5′-AGGCCAAGCATAGGGTGTCGCCC-3′</td>
</tr>
<tr>
<td>GM002</td>
<td>5′-TGCTGCGGCTAAACTTATTATTGATTGTTG-3′</td>
</tr>
<tr>
<td>GM003</td>
<td>5′-GTAAGCTGGTATGAGGTCTGCAAGG-3′</td>
</tr>
<tr>
<td>GM004</td>
<td>5′-CTAGGCTGGCCAAGAAGAGTGTCGCCC-3′</td>
</tr>
<tr>
<td>GM005</td>
<td>5′-CTAGGCTGGCCAAGAAGAGTGTCGCCC-3′</td>
</tr>
<tr>
<td>GM006</td>
<td>5′-CTCTATAGTGGTTCTGCGGCCC-3′</td>
</tr>
</tbody>
</table>

The pairs of primers used in the reactions and the predicted size of the PCR products were:

- GM001 and GM002, 351 bp;
- GM003 and GM004, 545 bp;
- GM005 and GM006, 313 bp.
Figure 3. Multiple alignments of amino acid sequences encoded by the ORFs of gypsyDm, gypsyDs and gypsyDv elements. The hyphens indicate identical positions in the sequence with respect to the top gypsyDm sequence. The asterisks indicate the following conservative changes: L=V=A, S=T, K=R, E=D, N=Q, Y=F. (A) Amino acid sequences corresponding to gag-like ORF1. The putative arginine-rich RNA-binding motif is highlighted in the three gypsy elements. (B) Amino acid sequences corresponding to pol-like ORF2. The functional domains of protease (PR) reverse transcriptase (RT) RNase H (RH) and endonuclease-integrase (EN) are marked with arrows. (C) Nucleotide sequence of the proposed gypsyDs and gypsyDv splice sites. Consensus sequences and splice sites of gypsyDm are cited in Pelisson et al. (22). (D) Amino acid sequence corresponding to env-like spliced ORF3 products. Characteristic features of envelope proteins are marked. Boxes indicate the peptide signal, tetrapeptide basic motif and transmembrane domain. Arrows represent the putative cleavage sites. Putative glycosylation sites that correspond to the consensus sequence NXT(S/T) are marked with numbers and conserved cysteine residues with letters. Empty circles indicate single insertion events in nucleotide sequence that imply frame recovery.

with a very similar restriction pattern. Clone DsE1 was selected arbitrarily for further study.

The nucleotide sequence of the gypsy element present in this clone was fully determined in both strands (EMBL accession no. X72390). The chromosomal location of the gypsyDs present in clone DsE1 was determined by in situ hybridization of larval polytene chromosomes. The signal was found at the border of the 35AB puff in the E chromosome using as probes both an adjacent genomic DNA fragment (Fig. 1A) and the 6.9 kb XhoI–XhoI fragment of gypsyDs (Fig. 1B). The target sequence found at the insertion site is TACA-gypsy-TACA, in which the 4 bp repeat generated by the insertion event is underlined. This sequence corresponds to the consensus found in gypsyDm insertions: TA(C/T)A*(C/T)A, in which the asterisk indicates the exact insertion site. The sequenced element is 7522 bp long and contains two identical LTRs of 613 bp and three central ORFs with a general structure very similar to gypsyDm and gypsyDs. Although the LTRs are the most variable regions in length and sequence among gypsy elements, the U3-R-U5 structure characteristic of retroviruses can be observed in the LTRs, as well as the promoter regions and the transcription and polyadenylation signals. Furthermore, motifs involved in the retrotransposition cycle are conserved in the three species near the LTRs (Fig. 2A).

The enhancer region, located between the 5′ LTR and ORF1, shows a heterogeneous sequence but a similar structure in the three elements, in which a conserved repeat of 12 bp is interspersed in A+T-rich regions with variable number and position (Fig. 2B). In D. melanogaster these 12 bp repeats constitute the binding site for a protein with zinc finger domains that acts as a transcription activator (39). The number of repeats varies, being 13 in gypsyDm, six in gypsyDv and nine in gypsyDs. The existence of these repeats in the other gypsy elements could indicate that they have the same function in these species. A poly(A) tract followed by an imperfect palindrome can be found 5′ of this regulatory region in gypsyDm.
and gypsyDs. This sequence in gypsyDm binds a repressor protein, probably encoded by su(fj) (40).

The products of ORF1 and ORF2 are necessary for both intracellular and extracellular retroelement movements. Assuming that the three gypsy elements are transpositionally active (3,28,29), the essential motifs of these proteins must be conserved. This is exactly what happens with the ORF2 products, in which the PR, RT, RH and EN domains can be observed (Fig. 3B). These domains are homologous to those present in the pol gene of retroviruses, being in the same order.

The 5′ region of the ORF1 sequences is the most variable, due to deletions or changes in gypsyDs and gypsyDv that do not produce frame shifting. Proteins with gag functions have RNA binding domains in order to recognize RNA and specifically direct it to the capsid. Although no RNA binding domains could be found in the putative gypsyDm and gypsyDv ORF1 proteins (20), a more detailed analysis looking for different types of motifs (41) revealed that an arginine-rich region near the C-terminus can act as an ARM RNA binding motif in the three gypsy elements (Fig. 3A). Furthermore, the proline content (found at conserved positions) of ORF1 is high in gypsy elements, being 5.1% in gypsyDm and 4.6% in gypsyDs.

ORF3 corresponds in position and size to the env gene of retroviruses and, as has been found in gypsyDm, it acts as a true envelope protein (22,23). As in retroviruses, it has been established that gypsyDm Env is translated from a 2.1 kb mRNA produced by differential splicing. The splicing donor and acceptor sites have been thoroughly determined in gypsyDm (22). Looking for splicing sites in gypsyDs and gypsyDv shows that a donor site located in the 5′ leader region is identical in the three gypsy elements. However, the acceptor site located between ORF2 and ORF3 shows some variation. In gypsyDs we have found an acceptor site consensus sequence, but it was not so evident in gypsyDv. Two potential acceptor sites in these gypsy elements are shown in Figure 3C. As in gypsyDm, in gypsyDs the splicing generates an ATG, underlined in Figure 3C, but translation cannot occur from this ATG because a number of stop codons appear downstream. A second ATG can generate a 5′ truncated homologous envelope protein. In gypsyDv, independent of the splicing acceptor site, a TAA stop codon is present immediately adjacent to the ATG, which also generates a 5′ truncated protein in this element.

gypsyDm ORF3 shows the structural motifs typical of envelope proteins, such as a signal peptide at the N-terminus, a dibasic cleavage site in the middle of the sequence and a transmembrane domain located near the C-terminus. Furthermore, a number of cysteines and three putative N-linked glycosylation sites involved in post-translational modification have been found in gypsyDm (22). Surprisingly, multiple alignments of the putative products of gypsyDm, gypsyDs and gypsyDv ORF3 show that although the sequence is well conserved among them, gypsyDs and gypsyDv lack some of the envelope motifs, while others are conserved (Fig. 3D). Firstly, gypsyDs and gypsyDv lack the signal peptide sequence, because the start codon is located downstream in both cases (Fig. 3D). Furthermore, the sequence that corresponds to the start site in gypsyDs is variable among gypsy elements, gypsyDs and gypsyDv having lost the start codon in such a way that ORF3 in these elements starts from a downstream ATG that is not in phase. This could result in totally different proteins, but the reading frame is recovered in gypsyDs and gypsyDv by independent and single insertion events (indicated in Fig. 3D by empty circles) that have occurred at different locations. Second, the dibasic cleavage site, N-linked glycosylation sites and cysteines are highly conserved among gypsy elements. Finally, gypsyDs lacks the transmembrane domain, because a single deletion in the upstream region has produced a frame-shift in this element that provokes early protein termination. In summary, gypsyDs and gypsyDv lack some essential motifs necessary to produce functional envelope proteins.

Divergences among gypsy elements

Table 1 presents the per cent identity at the nucleic and amino acid levels obtained from general comparisons between gypsy sequences corresponding to the three ORFs. The major differences in the per cent identity between the nucleic and amino acid sequences are found in ORF2, which encodes a putative protein with PR, RT, RH and EN domains. This result is in accord with the important function of this protein in the replicative cycle of retrotransposons. Nevertheless, the most important conclusion from this analysis is that similarity is greater between gypsy elements of species belonging to different subgenera (D.viridis and D.subobscura) than those belonging to the same subgenus (D.melanogaster and D.subobscura). These data are not in accord with a strictly vertical transmission of gypsy sequences in the genus Drosophila.

A more detailed analysis was carried out in order to determine whether gypsyDs is functionally active (Table 2). These studies show that in all cases the number of silent substitutions per effectively synonymous site (D_S) is much higher than the number of replacement substitutions (D_R), indicating that gypsyDs has been subjected to functional constraints. Small differences can be observed among the ORFs. For instance, D_T (number of total substitutions) and D_R are smaller when ORF2s are compared, while D_T is greater in comparison with those of other ORFs. This implies major functional constraints on ORF2. In contrast, ORF1 shows more divergence. Another significant finding obtained from this study is that the divergence between gypsyDs and gypsyDv is less than that expected in all ORFs, suggesting that genetic distance does not correlate with phylogenetic distance between the species.

Table 1. Percent identity at the nucleotide and amino acid levels among the different ORFs of gypsy sequences

<table>
<thead>
<tr>
<th></th>
<th>ORF1</th>
<th>ORF2</th>
<th>ORF3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>DNA</td>
<td>Protein</td>
<td>DNA</td>
</tr>
<tr>
<td>gypsyDm versus gypsyDs</td>
<td>71.46</td>
<td>74.29</td>
<td>75.97</td>
</tr>
<tr>
<td>gypsyDm versus gypsyDv</td>
<td>68.90</td>
<td>71.81</td>
<td>75.18</td>
</tr>
<tr>
<td>gypsyDs versus gypsyDv</td>
<td>90.52</td>
<td>90.35</td>
<td>92.44</td>
</tr>
</tbody>
</table>
Phylogenetic analysis of gypsy sequences

To obtain data about the relationships between gypsy sequences we have constructed trees using different phylogenetic methods (neighbour-joining; Fitch and Margoliash, assuming variable mutation rate; maximum parsimony) using the amino acid sequences corresponding to ORF2 in all cases. As an outgroup we have used the RT-homologous ORF of 412, a D.melanogaster retrotransposon that belongs to the gypsy group. The topology in the trees obtained by different methods agree in those where gypsyDs is grouped with gypsyDv instead of gypsyDm. As an example, Figure 4A shows the results obtained with the neighbour-joining method. In order to obtain a conventional phylogeny we have considered a 250 nt sequence of exon 8 of the Antennapedia (Ant) gene from the same species of Drosophila. As an outgroup we used the homologous sequence of the aquatic crustacean Artemia franciscana. Figure 4B shows that AntDs and AntDm are grouped together rather than with AntDv, which is consistent with the conventional phylogeny and different from the topology obtained for gypsy sequences.

Transcriptional activity of gypsyDs

The presence of mRNA corresponding to gypsyDs was tested by means of the Northern blot technique. In Figure 5 a blot of poly(A)+ mRNA of strain H-271 D.subobscura adults is shown using the XhoI–XhoI fragment of gypsyDs as probe. The same nylon filter was probed with the D.melanogaster actin gene as a control (data not shown) The major band corresponds to the 7.0 kb full-length gypsyDs RNA, which clearly proves that the gypsy element is transcriptionally active in D.subobscura. The other bands could correspond to transcripts of defective elements or splicing products. Interestingly, a weak 2.1 kb band is detected, having the same size as the subgenomic mRNA band that corresponds to spliced env mRNA in gypsyDm.

Detection of gypsyDm elements in D.subobscura strains

Previous data about the distribution of sequences homologous to the gypsyDm element among 25 D.subobscura strains from natural populations over its distribution area indicate that five of the strains analysed showed striking hybridization patterns with strong signal intensities and different banding patterns (29). One of the hypotheses to explain this hybridization pattern is that gypsy elements in these strains are closely related to those of gypsyDm. In order to test this hypothesis we performed Southern blot analysis using the XhoI–XhoI fragment of gypsyDs and gypsyDm sequences as probes. Figure 6 shows the Southern blot of genomic DNA of different D.subobscura strains probed with the XhoI–XhoI fragments of gypsyDm (Fig. 6A) and gypsyDs (Fig. 6B). The results indicate that while strains SN, TU, DIE, PC and MAR (lanes 2, 3, 4, 6 and 9 respectively) give a comparable signal with respect to the D.melanogaster control strain (lane 10), strains H-271, CJ, MA and RA (lanes 1, 5, 7 and 8) show a very weak signal using the gypsyDm probe.

Table 2. Comparative analysis of the gypsy nucleotide sequences corresponding to the three ORFs

<table>
<thead>
<tr>
<th></th>
<th>ORF1 total</th>
<th>silent</th>
<th>replacement</th>
<th>ORF2 total</th>
<th>silent</th>
<th>replacement</th>
<th>ORF3 total</th>
<th>silent</th>
<th>replacement</th>
</tr>
</thead>
<tbody>
<tr>
<td>gypsyDm</td>
<td>1170.2</td>
<td>299.7</td>
<td>917.5</td>
<td>358.0</td>
<td>62.8</td>
<td>295.2</td>
<td>1155.0</td>
<td>252.7</td>
<td>902.3</td>
</tr>
<tr>
<td>gypsyDs</td>
<td>1024.0</td>
<td>290.3</td>
<td>733.7</td>
<td>326.0</td>
<td>60.2</td>
<td>265.8</td>
<td>1045.0</td>
<td>263.7</td>
<td>781.3</td>
</tr>
<tr>
<td>gypsyDv</td>
<td>156.7</td>
<td>54.2</td>
<td>102.5</td>
<td>40.3</td>
<td>13.0</td>
<td>27.3</td>
<td>91.0</td>
<td>33.5</td>
<td>57.5</td>
</tr>
</tbody>
</table>

The number of silent and replacement sites has been calculated according to the method described in Hartl and Clark (64). D are the divergence values and k the corrected percent divergence as estimated as described in Jukes and Cantor (65), k = -3/4(ln(1 – 4D/k)). Dg, Ds and Dr represent the divergence values for total, silent and replacement changes respectively.

Figure 4. Phylogenetic trees obtained with the neighbour-joining method for (A) amino acid sequences corresponding to ORF2 gypsy elements using the 412 sequence as outgroup and (B) nucleotide sequence of a 250 bp fragment corresponding to exon 8 of Drosophila Antennapedia sequences, using the A.franciscana corresponding sequence as outgroup.
Figure 5. Northern blot of the mRNA poly(A)+ fraction of D. subobscura H-271 strain adults using the 6.9 kb XhoI–XhoI fragment of gypsyDs DNA as a probe.

(Fig. 6A). In contrast, strains SN, TÜ, DIE, PC, MAR, and H-271 (lanes 1, 2, 3, 4, 6 and 9), give similar hybridization signals using the gypsyDs probe (Fig. 6B). These data indicate that these five strains probably carry different types of gypsy sequences: one corresponding to the gypsyDs sequenced in H-271 strain and the other more similar to gypsyDm, present only in five strains and absent in the other D. subobscura strains analysed.

In order to demonstrate the presence of putative gypsyDm elements in the genome of some D. subobscura strains, we designed three pairs of primers to selectively amplify gypsyDm sequences (see Materials and Methods). Positive amplifications were subjected to a bi-directional Southern blot using as probes both gypsyDm and gypsyDs XhoI–XhoI fragments. Figure 6C shows positive hybridization fragments of the predicted size using the gypsyDm probe in strains TÜ and MAR. The results with the gypsyDs probe were negative (data not shown). The amplified fragment obtained with primers GM001 and GM002 in strain TÜ (lane 1), which corresponds to the predicted size of 351 bp, was cloned and sequenced and is nearly identical to that of gypsyDm. The differences correspond to two changes of a T to a C at positions 1220 and 1241.

DISCUSSION

Gypsy elements: retroviruses or retrotransposons?

The identity of retroviruses with respect to retrotransposons is based on the product encoded by the env gene, a transmembrane protein necessary for extracellular expansion. The presence of an ORF3 in most of the gypsy-like elements (gypsy, 17.6, 297, TED, tom; 2,3,5,6,9) in the same position as retroviral env could indicate that this group of retrotransposons are infective. However, given the strong divergence between gypsy ORF3 and retroviral env (3,42) it seems that gypsyDm ORF3 encodes a non-functional Env protein, probably derived from an ancestral functional protein. Recent data about gypsyDm env function (22,23), supported by the infective ability of these elements (24), allows gypsyDm to be considered as the first described insect retrovirus. ORF3 of gypsyDm exhibits a number of structural features characteristic of a membrane protein, i.e. a signal peptide, a dibasic cleavage site and transmembrane domains, glycosylation sites and also a number of conserved cysteines (9,22,23,42,43). Interestingly, multiple alignment of ORF3 from gypsyDm, gypsyDv and gypsyDs indicates that gypsyDs and gypsyDv lack essential motifs for potential Env functionality. GypsyDv lacks the signal peptide and gypsyDs lacks...
both the signal peptide and the transmembrane domains. From these data it can be inferred that the sequenced gypsyDs and gypsyDs elements have non-functional Env proteins, which probably evolved from an ancestral gypsyDm-like functional Env. If consideration of gypsyDm as an insect retrovirus is mainly based on Env activity, following this reasoning gypsyDv and gypsyDs have lost this ability and consequently they have evolved to retrotransposons. Possibly gypsyDs and gypsyDv are old parasites of the genomes of these species and the present-day elements are vestiges of infective ones. An alternative hypothesis, that gypsyDm constitutes a retrovirus recently evolved from related retrotransposons, is also consistent with the data. In any case, gypsy elements are a clear example of the fuzzy line that separates retrotransposons and retroviruses, because they have elements representative of both entities.

On the other hand, retroviruses have deleterious effects on the host organisms because of their infective properties. Until now retroviruses had only been described in vertebrates, but it has been established that gypsyDm is an insect retrovirus, constituting the first example described in invertebrates (22,23). Retroviruses have probably been described only recently in invertebrates because they acquired different systems to control retrovirus expansion. Following this reasoning, gypsy elements constitute good examples of the existence of this control: the intercellular expansion of gypsyDm is strongly dependent on the flan host gene, which controls splicing of the mRNA and consequently production of Env proteins. From the comparative analysis of gypsyDs, gypsyDm and gypsyDv it can be deduced that the mechanism of splicing is potentially preserved in the three species. Northern analysis of gypsyDs detects a strong 7.0 kb band that corresponds to the genomic RNA and a very weak band of 2.1 kb that has the same size as that of the spliced product of gypsyDm. Although ORF3 protein has lost its potential functionality in D. subobscura and D. virilis, the splicing mechanism is preserved.

**Horizontal transfer?**

The most simple explanation to account for the higher level of sequence similarity between gypsyDv and gypsyDs than between gypsyDm and gypsyDs and gypsyDm is that horizontal transmission of gypsy elements between D. subobscura and D. virilis genomes occurred 17–24 million years ago. A more detailed study of the distribution of gypsy elements in the obscura and virilis subgroups will have to be done in order to determine the donor and acceptor species. The phylogenetic trees, based on gypsy sequences compared with that of the Antennapedia gene, support a close phylogenetic relationship between gypsyDv and gypsyDs, while Antennapedia sequences cluster according to the phylogenetic relationships proposed for the three species (44,45). Furthermore, gypsyDv and gypsyDs share some LTR and ORF structural characteristics (i.e. variability in the S' region of ORF1 and lack of the potential signal peptide in ORF3) which indicate that these sequences are evolutionarily closer than to that of D. melanogaster.

The occurrence of horizontal transfer of eukaryotic transposable elements among phylogenetically related or distant species could be an extended phenomenon (46–49). From the inconsistencies found in the phylogenetic analysis horizontal gene transfer may be suspected, but various additional supporting lines of evidence are usually necessary to firmly establish this hypothesis, as occurs in the evolution of the P transposable element (47,49–52). In the case of gypsy elements horizontal transfer was earlier proposed to explain the close relationship between gypsy elements from D. melanogaster and Ty3 from yeast (53). A more extended distribution of gypsy group elements among different taxa strongly supports lateral transfer events in their evolutionary history. Gypsy group elements have been detected in insects (gypsy, 412, 17.6, 297, micropia and Ulysses from Drosophila and mag and TED from the lepidoptera Bombyx mori and Trichoplusia ni respectively), yeast (Ty3 from Saccharomyces cerevisiae and Ty1 from Schizosaccharomyces pombe), fungi (Cfl1 from Cladosporium fulvum), marine invertebrate (SURL from the sea urchin Tripneustes gratilla) and plants (del from Lilium henryi and IFG7 from Pinus spp.) (7–9,15–18,54,55). Phylogenetic trees based on RT and RH amino acid sequence domains show clear inconsistencies, i.e. Ulysses and micropia are closer to the plant SURL family than other Drosophila gypsy elements.

Assuming that some horizontal transfer events could have occurred in gypsy group evolution, it can be implied for gypsy elements between D. virilis and D. subobscura on the basis of structural analogues and sequence similarity. However, on considering, as cited above, that gypsyDs and gypsyDv seem to be old non-infective parasites of the genome of both species, it could alternatively be proposed that gypsyDs and gypsyDv derive from a common ancestor present in some ancestral species before Sophophora radiation and consequently derived sequences are found among Sophophora and Drosophila subgenera. The potentially infective gypsyDm sequences could be considered recent parasites of the D. melanogaster genome. Nevertheless, this hypothesis does not explain the 90% similarity between gypsyDs and gypsyDv when genomic sequences from these species are more divergent.

Horizontal transfer requires special conditions to be successful. Among these has been proposed the existence of a transmitting vector (47). If we admit that gypsyDs and gypsyDv were probably retroviruses in the past, the horizontal transfer event proposed to explain the similarity existing between these elements could constitute a simple infection event.

**Two types of gypsyDs elements co-exist in natural populations of D. subobscura**

Assuming that gypsyDs and gypsyDs represent old sequences probably dispersed in the genome of Drosophila spp. and that gypsyDm represents a new potentially infective element, it is possible that both old and new sequences co-exist in the genome of a given species. In support of this proposition highly diverged gypsyDm elements have been found in the j1 suffix strain (56). Two subfamilies of gypsyDm elements belonging to SS and MS strains of D. melanogaster have been described (57). The MS strain is characterized by a high frequency of spontaneous mutation, which has been correlated with gypsy transposition (58,59). Moreover, co-existence of different subtypes correlated with their activity and/or time of persistence in a given genome has been described in other transposable element families, i.e. mariner, I and P (60–62).

In the D. subobscura genome the gypsy element analysed was obtained from a genomic library constructed with DNA from strain H271, which is a laboratory strain homozygous for inversions in all of their five chromosomes (63). A survey for the presence of gypsy elements in natural populations representative of the dispersion area (North Africa, Europe and North and South America) had already been carried out (29) by Southern analysis. In five of the populations analysed a differential hybridization pattern was obtained. In the present paper we have sequenced a PCR-amplified fragment from D. subobscura strain TU that was nearly identical to the canonical gypsyDm element. From these data we can infer that similar elements could be present in the other four strains. It can be deduced that in these strains sequences co-exist which are homologous to
gypsyDs (old gypsy elements) and to gypsyDm (new gypsy elements). In this context, it is possible that gypsyDm is invading D. subobscura populations by means of its infective properties. The five strains cited came from natural populations, but have been maintained in the laboratory for some years. Laboratory conditions could possibly increase the likelihood of infection by gypsy viral particles originating from D. melanogaster strains.

ACKNOWLEDGEMENTS
This work was supported by grants from the Spanish CICYT (no. PB90-426) and FPI programmes. The nucleotide sequence was obtained at ‘Servicio de Bioinformática’, both at the University of Valencia. We would like to thank R. González-Duarte and M. Garfany for the D. subobscura genomic library and N. Paricio for kindly providing the mRNA and for her help with the Northern blots.

REFERENCES