A long interspersed repetitive element—the I factor of Drosophila teissieri—is able to transpose in different Drosophila species

(transposable element/hybrid dysgenesis/evolution)

Pierre Abad*, Chantal Vaury†, Alain Pélisson†, Marie-Christine Chaboissier†, Isabelle Busseau‡, and Alain Bucheton§§

*Station de Recherches de Nématologie et de Génétique Moléculaire des Invertébrés, Institut National de la Recherche Agronomique, B.P. 2078, 06606 Antibes Cedex, France; †Laboratoire de Génétique, Université Blaise Pascal, 63177 Aubiere Cedex, France; and ‡Centre de Génétique Moléculaire, Centre National de la Recherche Scientifique, 9190 Gif-sur-Yvette, France

Communicated by François Jacob, August 7, 1989 (received for review March 30, 1989)

ABSTRACT Long interspersed repetitive elements (LINEs) are transposable elements present in many species. In mammals they are difficult to study because most of them are defective and their transposition frequency is low. The I factor of Drosophila melanogaster is a LINE element that is particularly interesting because its transposition occurs at high frequency during I-R hybrid dysgenesis. This phenomenon occurs when males from the class of inducer strains are crossed with females from the class of reactive strains. Inducer strains contain several complete 5.4-kilobase I factors at various sites on the chromosomal arms. Reactive strains are devoid of complete I factors. Many results indicate that active I factors have invaded the D. melanogaster genome recently. To study the evolutionary history of I elements, we have cloned and sequenced a potentially active I factor from Drosophila teissieri. It is flanked by a target-site duplication and terminates at the 3′ end by tandem repeats of the sequence TAA. When introduced into the germ line of a reactive strain of D. melanogaster by P element-mediated transformation, it is able to transpose and induces hybrid dysgenesis. This strengthens the hypothesis of a recent reinvasion of the D. melanogaster genome by active I factors giving rise to the inducer strains. They could have originated by horizontal transfer from another species. Such events also could occur for other LINE elements and might explain the spread of new variants in mammalian genomes. Moreover, the results give a further insight into I factor functional organization.

Long interspersed repetitive elements (LINEs) are the class of transposable elements known as nonviral retrotransposons. In mammals they are present at 10⁶–10⁸ copies dispersed throughout the genome. These elements are devoid of terminal repeats and are terminated at the 3′ end by an adenosine-rich sequence. They encode putative polypeptides showing similarities with viral reverse transcriptases (1). Recent evidence indicates that in humans they can cause diseases resulting from insertional mutations (2, 3). The mechanisms by which LINEs transpose and spread in the genome remain to be elucidated (4). These problems are difficult to study since their frequencies of transposition are low and most of the members of the family are defective (1).

The I factor is a LINE element responsible for I-R hybrid dysgenesis in Drosophila melanogaster (5). It is particularly interesting because its transposition occurs at high frequencies under certain conditions (6).

There are two categories of strains in D. melanogaster with respect to I-R hybrid dysgenesis: inducer strains, which contain active I factors, and reactive strains which do not. The phenomenon appears in the germ line of the F₁ dysgenic females resulting from crosses between inducer males and reactive females (7) and is characterized by various genetic abnormalities, such as sterility (8) and high rates of mutations (9). All other crosses produce apparently normal progeny. I factors are stable when maintained in inducer strains but transpose with unusually high frequencies when introduced into the cellular environment of reactive strains (6).

I factors are present at about 15 copies dispersed on chromosomal arms in inducer strains (10). Incomplete and defective I elements are present in both inducer and reactive strains and are located in the pericentromeric heterochromatic regions (11).

Active I factors have been cloned, and their ends have been sequenced (12). They are 5.4 kilobases (kb) long and appear to be similar in structure. They do not have terminal repeats, and they are terminated at their 3′ ends by four to seven tandem copies of the sequence (TAA). They are flanked by target-site duplications varying in length from 10 to 14 base pairs (bp). This general organization is similar to that of F, G, and Jockey elements of Drosophila (13–15), L1 elements of mammals (16), Ingi elements of Trypanosoma (16), and Cin-4 elements of maize (17).

An active I factor cloned from the white I-R-induced mutation w¹⁰⁹ has been entirely sequenced (12). It displays two long open reading frames (ORFs), ORF1 and ORF2, of 1278 bp and 3258 bp, respectively, separated by 471 bp. The amino acid sequence of ORF2 shows similarities with viral reverse transcriptases (12, 18). These data indicate that I factors could transpose by reverse transcription of an RNA intermediate.

I elements are present in many Drosophila species (19). Their distribution correlates with the phylogenetic relationships between species. Previous experiments indicate that the structure of the complete I factors of D. melanogaster is very similar to that of the I elements of Drosophila simulans, the species most closely related to D. melanogaster. This suggests that active I factors that spread recently in the latter species (5) probably arose by horizontal transfer from D. simulans (19). The structure of the I homologues in the more distant species Drosophila teissieri differs from that of the complete I factors of D. melanogaster (20).

To study in more detail the evolutionary history of I elements, we have cloned and sequenced a potentially active I factor from D. teissieri. It terminates at the 3′ end by four copies of the sequence TAA and is flanked by a duplication of the target sequence. It contains two long ORFs. When introduced into the germ line of a D. melanogaster reactive

Abbreviations: LINEs, long interspersed repetitive elements; ORFs, open reading frames.

The sequence reported in this paper has been deposited in the GenBank database (accession no. M28878).
strain, it is fully active: it induces hybrid dysgenesis and is able to transpose.

**MATERIALS AND METHODS**

**Drosophila Strains.** *D. teissieri* stocks were kindly provided by the Laboratoire de Génétique Évolutive of the Centre National de la Recherche Scientifique (Gif-sur-Yvette) and were derived from flies recently caught in central Africa.

**Constructions of *D. teissieri* Libraries.** The random library of the *D. teissieri* strain 128.2 was constructed in phage λEMBL4 as described (20). The library of *HindIII* fragments from strain 128.2 was made by digesting 1.5 µg of *Drosophila* DNA and 6 µg of ANM149 DNA with *HindIII*. The libraries of *EcoRI* fragments of strains 128.2 and Tai 81 were made by digesting 1.5 µg of *Drosophila* DNA and 6 µg of agt10 DNA with *EcoRI*. After ligation, the DNAs were packed in vitro (21), and the resulting phages were grown on NM514 (21) and C600 (22).

**P Element-Mediated Transformation.** The 5.4-kb *BamHI* fragment containing the *I* factor of *D. teissieri* (see Fig. 1 Left) was subcloned into the *BamHI* site of the pUCHsneo vector (23). The resulting plasmid, named pUCHsneoAltei, was purified on a CsCl density gradient prior to microinjection. It was then injected (500 µg/ml) together with the *pir25.7* wing-clipped helper plasmid (100 µg/ml) (24) into embryos of the reactive strain Cha of *D. melanogaster* according to the standard procedure (25). Injected embryos were incubated at 18°C. Hatched larvae were transferred onto standard food at 25°C. Go flies resulting from these embryos were individually crossed with flies from the Cha reactive strain, and the selection on G418-containing food was performed as described (23). Transformed lines were established by appropriate crosses and were monitored by *in situ* hybridization of Altei DNA to polytene chromosomes for the presence of *I* elements on chromosomal arms.

**Other Procedures.** All other procedures were carried out as described (10, 11).

**RESULTS**

**Cloning of a Potentially Active I Factor from *D. teissieri*.** To clone a potentially active *I* element from *D. teissieri*, we screened a genomic library of strain 128.2 with *I* factor DNA. Among many *I* elements isolated in this way, only one, named Altei, was flanked by nonrepeated sequences (Fig. 1 Left).

DNAs from four different stocks of *D. teissieri* were digested with *EcoRV* and hybridized with the 2.4-kb *EcoRI/Xho* *I* fragment flanking the *I* element of Altei (see Fig. 1 Left). A 6.1-kb fragment is observed in all stocks. Strain 128.2 contains in addition an 11.5-kb fragment due to insertion of the 5.4-kb-long *I* element cloned in Altei (Fig. 1 Right). This *I* element is the right size to be full-length. The stock 128.2 is polymorphic for the presence of this element at this site, indicating that it has recently transposed. Therefore, it could be transposable. Its restriction map differs from that of active *I* factors from *D. melanogaster*, named *Imel* (Fig. 1 Left).

**The Potentially Functional I Element from *D. teissieri* Is Fully Active in *D. melanogaster*.** As there are no stocks similar to the reactive strains of *D. melanogaster* available in *D. teissieri*, it is not possible to study the genetic properties of *I* elements in this species. Therefore, to study the ability of

---

**Fig. 1.** (Left) Restriction map of phage Altei containing an *I* factor from *D. teissieri*. Below is a restriction map of *I* factor from *D. melanogaster* drawn from the data reported in ref. 10. Solid bars correspond to sequences homologous to *I* factor DNA. Thin lines show the flanking sequences. The *EcoRI* sites at the borders are from AEMBL4 (the arms of the phage are not represented). Phage Altei was isolated by screening a random library of the *D. teissieri* stock 128.2 in AEMBL4 (20) with a 400-bp fragment from the 3' end of *I* factor from *D. melanogaster*. The restriction map is not complete for the flanking sequences in Altei. A, *Ava I*; B, *BamHI*; BgII; E, *EcoRI*; H2, *HindII*; H3, *HindIII*; K, *Kpn I*; P, *Pst I*; X, *Xho I*. (Right) Hybridization of DNAs from four strains of *D. teissieri* with the nonrepeated sequence flanking the *I* element cloned in phage Altei (Left). DNAs from four different *D. teissieri* stocks were digested with *EcoRV*. The resulting fragments were separated on a 1% agarose gel, transferred to a nitrocellulose filter, and hybridized with the 2.4-kb *EcoRI/Xho* *I* fragment flanking the *I* element of Altei (Left). Hybridization and washing conditions were as described (19).
the I factor from D. teissieri to transpose, we subcloned the BamHI fragment containing Itei (Fig. 1 Left) into a P transposon vector for transformation (23). The resulting plasmid has been microinjected together with the helper P element p25.7wc (24) into embryos of the Cha reactive strain of D. melanogaster, which does not contain any P element. Two independent transformed lines were obtained. Both result from insertion of the P transposon containing the I factor from D. teissieri on the second chromosome (results not shown). In situ hybridization of I factor DNA to salivary gland chromosomes of larvae from the two transformed lines indicates that there are multiple copies of this element at various sites on the arms of all chromosomes (Fig. 2). We conclude from these observations that the I factor from D. teissieri transposes at high frequency when introduced into the genome of a reactive strain of D. melanogaster. We have checked that there is no P element transposition in these lines (results not shown).

The I factor from D. teissieri also confers the inducer phenotype in D. melanogaster. This phenotype, characteristic of I-R hybrid dysgenesis, results from I factor activity and corresponds to the induction of a typical sterility of the F1 females resulting from crosses between males from inducer stocks with females from reactive strains. It is due to low hatchability of the eggs laid by these F1 females. All other crosses produce fertile daughters. We have crossed males from both transformed lines with females either from the reactive strain Cha (crosses A) or from the inducer strain Canton S (crosses B). The results (Table 1) show that only crosses A produce sterile F1 females, indicating that both transformed lines behave as inducer strains.

All these data indicate that factors from D. teissieri are fully active in D. melanogaster.

### Table 1. Test of the inducer phenotype of transformed lines 1 and 2

<table>
<thead>
<tr>
<th>Crosses</th>
<th>Transformed line 1</th>
<th>Transformed line 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0 (401)</td>
<td>0 (454)</td>
</tr>
<tr>
<td>B</td>
<td>70 (590)</td>
<td>87 (413)</td>
</tr>
</tbody>
</table>

Data are the hatching percentages of the eggs laid by F1 females resulting from the crosses described in the text. The numbers of eggs scored are indicated in parentheses. Crosses: A, Cha wIRI (reactive strain) × d♂ from transformed lines 1 or 2; B, Canton S wIRI (inducer strain) × d♂ from transformed lines 1 or 2.

duplication. The 3' end of one strand is identical to that of the I factors of D. melanogaster because it contains several copies of the sequence TAA.

The overall genetic organization of Itei is similar to that of IwIRI (Fig. 4). All of the coding capacity is in one strand. There are two long ORFs with initiation codons near their 5' ends. One of them, ORF1, starts 181 bp from the left-hand end in Fig. 3 and is 1296 bp long. The second ORF, ORF2, is 3750 bp long and terminates 126 bp from the 3' end of the element. The region between the two ORFs is 35 bp long.

The two polypeptides potentially coded by ORF1 in the two species are 82% homologous. However, one part of the polypeptide corresponding to nucleotides 182–1036 is more conserved (94% homologous) than the other part corresponding to nucleotides 1037–1476 (75% homologous). This polypeptide contains a highly conserved motif of the type Cx2C,4x6Hx6C [where C = Cys, H = His, and X = Aaa (for other amino acid) in single-letter code] which is also found in basic nucleic acid binding proteins of retroviral gag polypeptides (Fig. 3).

ORF2 codes for a putative polypeptide that is 86% homologous to that deduced from the ORF2 sequence of IwIRI. The parts of this polypeptide corresponding to the reverse transcriptase and RNase H domains are highly conserved (95–100% homologous).

### Genetic Organization of the I Factor of D. melanogaster Is Identical to That of the I Factor of D. teissieri.

According to the sequence of the I factor associated with the IwIRI mutation previously reported (12), the I factor of D. teissieri shows a substantial difference in the primary structure of ORF2, which is longer in Itei than in IwIRI. This results in a reduction of the region between the two ORFs to 35 bp in D. teissieri. We sequenced the central part of another I factor of D. melanogaster contained in pl407 isolated from the white I-R-induced mutation wIRI (10). This cloned I factor has been shown to be functional when introduced into the genome of a reactive strain (27). Its sequence is exactly the same as that reported for the I factor associated with the wIRI mutation (12) except that it shows an additional cytidine at position 1946 and a deletion of one guanosine at position 1727 in the sequence previously reported (12). This creates a longer second ORF extending toward the 5' end. It is organized in the same way and starts with the same methionine codon as ORF2 in Itei (Fig. 4). Therefore, it is reasonable to conclude that the genetic organization of I factors in the two species is similar. The region between the two ORFs is only 54 bp long in D. melanogaster. Consequently, we assume that ORF2 of the I factors in this species is 3657 bp long and starts at position 1518 in the sequence previously reported (12).

The differences observed for the I factor associated with wIRI could result from two frame-shift mutations produced during cloning of this element.
Fig. 3. Sequence of the I factor of D. teissieri. Lower-case letters indicate the target site duplication flanking the I factor. Asterisks in the amino acid sequences correspond to the stop codons terminating the two large ORFs. The first methionine of each polypeptide is boxed. Typical domains are underlined (see text). GAG is a nucleic acid binding domain of the type Cx2Cx3xHx5x2C. RT1-RT7 correspond to the seven reverse transcriptase domains present in the I factor of D. melanogaster (12) that are highly conserved in D. teissieri. RNase H1 and RNase H2 indicate the positions of the two RNase H domains also present in the I factor of D. melanogaster (26). The black dots indicate identical amino acids and the white circles indicate similar amino acids in GAG, RT, RNase H domains from I factors of D. teissieri and D. melanogaster.

**DISCUSSION**

The present results indicate that functional I factors occur in species other than D. melanogaster. Previous in situ hybridization experiments to salivary gland chromosomes of D. teissieri showed that this species contains I elements on chromosomal arms. Their locations differ from strain to strain, as would be expected for I transposable elements (20). In this paper, we describe the isolation and the characterization of one of these active I factors. It results from recent insertion in strain 128.2 and is inserted in a nonrepeated sequence located on a chromosomal arm. Its organization is similar to that of functional I factors from D. melanogaster. It is flanked by a target site duplication and terminates at the 3' end by TAA repeats. It contains two long ORFs.

The most conserved regions between the I factors from the two species correspond to the reverse transcriptase and RNase H domains of ORF2 previously defined (12, 26), supporting the hypothesis that I factors transpose by reverse transcription of an RNA intermediate. The overall homology of the nucleotide sequences is 85%. It appears to be weaker than that estimated for other sequences between D. teissieri and D. melanogaster.

Fig. 4. Comparison of the structure of the I factors of D. melanogaster and D. teissieri. The organization of the I factor from D. melanogaster has been drawn by compiling the data reported in ref. 12 for the I factor associated with the w^{12} mutation and from our data for the I factor associated with the w^{12} mutation. Boxes represent ORFs. Black boxes represent highly conserved domains: (1) nucleic acid binding domain; (2) reverse transcriptase domains; (3) RNase H domains.
and the sibling species *D. melanogaster* and *D. simulans* from which the *I* factors of the inducer strains appear to have originated (28, 29). This difference could be due to the fact that the spread of DNA sequences through the process of reverse transcription of an RNA intermediate is more prone to error than DNA-dependent DNA polymerase activity.

The first 37 nucleotides at the 5′ ends of the *I* factors of *D. melanogaster* and *D. teissieri* are identical with the exception of the nucleotide 7 (Fig. 3), suggesting that this region is important for *I* factor activity. It has been suggested (12) that *I* factors transpose by reverse transcription of a full-length RNA synthesized from an internal RNA polymerase II promoter. This has been proved to be the case for another transposable element named *jockey*, which is structurally related to LINEs (15). The similarity of the 5′ ends of the two *I* factors suggests that such an internal promoter could be located at the very beginning of the element.

The *I* factor from *D. teissieri*, when introduced into the genome of a *D. melanogaster* reactive strain by *P*-mediated transformation, is able to transpose and to induce typical *I-R* hybrid dysgenesis. This strengthens the hypothesis, assuming that *I* factors have been able to reinvade the *D. melanogaster* genome after horizontal transfer from another species as we have suggested (19, 20).

Similar events could occur for other LINE elements such as *L1* in mammals. Various results suggest that in some cases, new LINE variants have recently invaded mammalian genomes. In mouse the comparative study of the structure of the 5′ end of the *L1* family in a series of related species indicates that two types of LINE elements (A and F) have invaded the mouse genome at different times. Type A was introduced more recently than type F (4). In the rat genome, the LINE family is relatively homogeneous, which contrasts with the heterogenous LINE families in mice (30). This homogeneity of the rat LINE family could be explained by the fact that only one invasion occurred recently in this species. In mice, several distinct but related LINE families were introduced during evolution (4). The mechanisms by which these LINE sequences have been generated and have invaded mammalian genomes remain to be elucidated. This can be achieved by successive amplifications of new sequence variants (as suggested in refs. 4 and 30). However, this could also result from transposition of a new LINE element arising by horizontal transfer from another species in the same way as this appears to have occurred for *I* factors in *D. melanogaster*.

We thank D. Finnegan for criticisms of the manuscript, A. Lenoir for excellent technical assistance, and J. C. Bregliano for discussions. This work has been financed by grants from the Centre National de la Recherche Scientifique, the Université Blaise Pascal, and the North Atlantic Treaty Organization.