Proc. Natl. Acad. Sci. USA
Vol. 84, pp. 5843–5847, August 1987
Genetics

Related polypeptides are encoded by *Drosophila* F elements, I factors, and mammalian L1 sequences

(oligo(A)-terminated sequences/retroposons/mobile elements/reverse transcriptase)

PIER PAOLO DI NOCERA* AND GIORGIO CASARI†

European Molecular Biology Laboratory, Meyerhofstrasse 1, D-6900 Heidelberg, Federal Republic of Germany

Communicated by Maxine Singer, April 29, 1987

ABSTRACT **The structural organization of *Drosophila* F elements closely resembles that of L1 sequences, a major family of repetitive DNA elements dispersed in the genome of all mammals. Members of both families are flanked by target-site duplications of different length, vary in size because of heterogeneity at one end, and invariably terminate at the other end in characteristic adenosine-rich stretches often preceded by polyadenylylation signals. The nucleotide sequence of Fw, an F element found in the white locus of w^{i+A} flies, reveals a large open reading frame upstream of the 3′ adenosine-rich terminus encoding a possible reverse transcriptase homologous to those potentially encoded by functional L1 units and *Drosophila* I factors. A cysteine-rich region within an interrupted frame located at the 5′ terminus of Fw suggests that complete F elements might additionally encode a nucleic acid binding protein. The observation that F elements and I factors encode functionally related polypeptides, and the extensive similarity of their hypothetical reverse transcriptases to L1 open reading frames, favors the hypothesis that all these sequences are evolutionarily related and transpose upon the cDNA conversion of RNA intermediates.**

More than 10% of the *Drosophila melanogaster* genome consists of moderately repetitive DNA, a large fraction of which is accounted for by at least 40 distinct families of transposable elements that can be grouped according to their structure into a few major classes (1–3). Whereas some elements have terminal inverted repeats and presumably transpose by mechanisms similar to those proposed for bacterial transposons (4), the largest group (*copia*-like elements) is made up of sequences that structurally resemble the integrated forms of vertebrate retroviruses (2, 3). The identification of reverse transcriptase-like products in all these elements (5–10) and the presence of *copia* RNA in ribonucleoprotein complexes associated with reverse transcriptase activity in cultured *Drosophila* cells (5) suggest that this type of element might transpose by way of an RNA intermediate using a mechanism similar to that of retroviruses.

Other *Drosophila* transposable elements, markedly different in structure from *copia*-like elements because they lack terminal repeats, may also originate from a reverse transcription process. This group includes I factors, the transposable elements involved in the I-R system of hybrid dysgenesis that encode a reverse transcriptase-like enzyme (11), as well as F (12, 13) and G (14, 15) elements, long, interspersed DNA sequences that terminate at one end with stretches of adenosine residues preceded by polyadenylylation signals (12–15). F and G elements are structurally reminiscent of a variety of dispersed DNA sequences, variously classified as processed pseudogenes (16) or retroposons (17), which are present in mammalian genomes and have been proposed to originate from the reverse transcription of RNA molecules.

While G elements are mostly interspersed with chromocentric repeated DNA sequences and seem to have a relatively stable chromosomal location (15), the nomadic nature of F elements is clearly established by their different location in *Drosophila* stocks (13, 18) and by the isolation of mutant alleles generated by the insertion of F family members at different loci (19–21).

It has been reported (13) that the organization of F sequences is similar to that of LINE-1 or L1 sequences, a family of long interspersed oligo(A)-terminated sequences dispersed throughout all mammalian genomes (reviewed in refs. 22 and 23). In addition to oligo(A) tails at the 3′ end, common features exhibited by F and L1 sequences include size heterogeneity due to different degrees of 5′ truncation and target-site duplications of various lengths flanking individual family members (13, 23).

We show here that F elements encode an open reading frame (ORF) that could encode a protein exhibiting extensive homology to the reverse transcriptase-like domains of the potential products of I factors and L1 sequences. This observation suggests that these DNA elements are closely related and are presumably mobilized within the genome by means of a similar mechanism.

MATERIALS AND METHODS

DNA Sequence Analysis. Restriction fragments derived from pA22.7, a recombinant plasmid carrying the Fw element (21), were subcloned into M13mp18 or M13mp19, and their nucleotide sequence was determined by the dideoxy chain-termination method (24). Part of the Fw DNA sequence has also been determined by the chemical method of Maxam and Gilbert as modified (13). The sequencing strategy adopted is illustrated in Fig. 1. Most of the sequence was determined on both strands.

RESULTS

Sequence Analysis of Fw. Structural analysis of *Drosophila* F elements showed that family members vary in size and that full-length elements are about 4.7 kilobases (kb) long (13). A 3.6-kb F element has been found within the white locus of w^{i+A} flies (13). w^{i} mutation results from the duplication of a 2.9-kb segment within the white locus (19). w^{i+A} revertants had lost one copy of the 2.9-kb repeat and acquired a DNA segment (21) that was identified by Southern blot and partial nucleotide sequence analysis as Fw, a 5′-truncated F element (ref. 13; Fig. 1).

We have determined the complete nucleotide sequence of Fw (Fig. 2). Fw is 3542 base pairs (bp) long and is flanked by

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: ORF, open reading frame.

*To whom reprint requests should be addressed.
†Present address: Istituto Sieroterapico Milanese, Via Darwin 1, Milan, Italy.

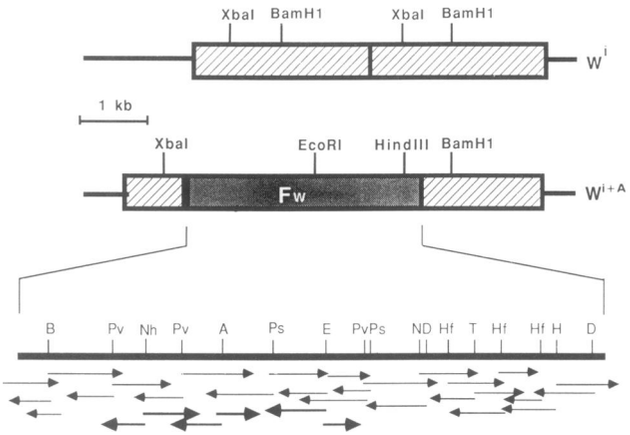

FIG. 1. The white gene DNA segment duplicated in w^i flies is shown together with the same region from w^{i+A} revertants where one repeat is lost and Fw is inserted. An expanded restriction map of the Fw element is reported, and arrows below the map indicate the cloning strategy and the extent of DNA sequence derived from different clones. Thicker arrows denoted DNA regions analyzed by the chemical method. Only relevant restriction sites are shown. A, *Ava* I; B, *Bgl* II; D, *Dra* I; E, *Eco*RI; H, *Hin*dIII; Hf, *Hin*fI; N, *Nru* I; Nh, *Nhe* I; Ps, *Pst* I; Pv, *Pvu* II; T, *Taq* I.

12-bp target-site duplications (13). Like other members of the F family, Fw has no terminal repeats, and one end is marked by a long stretch of adenosine residues preceded by two canonical polyadenylylation signals (Fig. 2). Translation of the Fw DNA sequence reveals a 2577-bp ORF (F-ORF2) that starts at nucleotide 711 and ends 230 bp before the 3′ adenosine-rich terminus (Figs. 2 and 3). The ATG located at the second amino acid residue in F-ORF2 might correspond to an initiating methionine according to the consensus established by Kozak (25). An additional 366-bp ORF (F-ORF1) is present at the truncated 5′ terminus of the element (Fig. 2; see below).

Homology of F-ORF2 to L1 ORFs. Conceptual translation of the DNA sequence of several mammalian L1 sequences established that at least some complete L1 units encode a long ORF (23) that has homology to reverse transcriptases (26, 27). A homologous protein is potentially encoded by F elements, since the introduction of a few gaps permits the alignment of F-ORF2 from residue 378 to the central segments of both mouse and human L1 ORFs (Fig. 4). One hundred and one (21%) and 95 (20%) amino acids are identical between F-ORF2 and human and mouse ORFs, respectively. Taking into account favored amino acid substitutions, the similarity between F and L1 ORFs polypeptides reaches 40%. The homology is higher within the region of F-ORF2 extending from residue 436 to residue 550. In this stretch 40 (35.0%) and 37 (32.1%) amino acid residues are shared by the *Drosophila* sequence and human and mouse sequences, respectively. The overall homology between *Drosophila* and mammalian proteins in this region, including favored amino acid substitutions, is 50%.

Interestingly, this region coincides with CS2, one of the two segments identified by Southern blot analysis as the most evolutionarily conserved portions of L1 sequences throughout mammals (29).

The homology extends further upstream, since we observed that a region similar to the F-ORF2 interval from residue 318 to residue 359 occurs within both L1 ORFs (Fig. 4). Although differently spaced in the two sequences, this upstream segment might correspond to an additional functional domain shared by L1 and F-encoded polypeptides.

F-ORF2 Is Homologous to ORF2 Encoded by I Factors. Fawcett *et al.* (11) have shown that *Drosophila* I factors, a class of transposable sequences that induce a high rate of mutation in certain *Drosophila* strains (3), contain two ORFs (I-ORF1 and I-ORF2) that are 1278 and 3258 bp long, respectively. I-ORF2 is partly homologous to the ORF encoded by L1 sequences (11), and comparative analysis shows an extensive similarity between I-ORF2 and F-ORF2 (Fig. 5). F-ORF2 can be aligned with few gaps from residue 396 to the end to the interval from residue 251 to residue 724 of I-ORF2. Within the homologous region, 28% of residues are identical in F-ORF2 and I-ORF2; the similarity of the two polypeptides, taking into account favored amino acid substitutions, exceeds 50%. The regions aligned in Fig. 5 correspond to the segments of F-ORF2 and I-ORF2 that are homologous to L1-ORFs (ref. 11; see also Fig. 4). We have not found within I-ORF2, as in the case of L1 (Fig. 4), an additional segment homologous to the F-ORF2 interval from residue 318 to residue 359.

Homology of F-ORF2 to Reverse Transcriptases. Ten of the 13 invariant amino acids present in known and putative reverse transcriptases (30) are in F-ORF2 (Fig. 6). The alignment of Fig. 6 shows that Fw, I, and L1 potentially encoded reverse transcriptases clearly belong to a distinct class, as they are more similar to each other than to retroviral polymerases, both in terms of number and of relative distance of identical and/or similar residues. A noticeable exception (11, 26) is constituted by the remnants of potential gene products encoded by class II mitochondrial introns present in the cytochrome oxidase subunit I gene of *Saccharomyces cerevisiae* (36). Except for residues invariably present in all reverse transcriptases, the Fw polypeptide exhibits a poor similarity to the reverse transcriptase encoded by the 17.6 element, as well as to those encoded by other *Drosophila copia*-like elements (data not shown). Although the diversity of reverse transcriptases encoded by *copia*-like elements is much greater than that in retroviruses (10), this observation supports the notion that F and *copia*-like elements constitute distinct classes of mobile sequences, although they both presumably transpose by way of the cDNA conversion of RNA intermediates.

The homology of F-ORF2 seems to be restricted to the polymerase region of reverse transcriptases. We have not found extensive similarities between segments of F-ORF2 and the protease and ribonuclease domains present in retroviral reverse transcriptases (30).

F-ORF1. A 366-bp ORF (F-ORF1) is found at the 5′ border of Fw and might extend further upstream in complete F family members. F-ORF1 has an unusually cysteine-rich domain containing one copy of the motif $CN_2CN_4HN_4C$, where N stands for any nucleotide, followed by two imperfect ones (Fig. 2). One or more copies of this motif are invariably present in the nuclear binding proteins that originate from the cleavage of retroviral gag proteins (37). These proteins are known to interact specifically with the retroviral genomes within virus particles and might bind the tRNA primer of DNA minus-strand synthesis (37). Notably, one copy of this motif and two imperfect ones are present in the first ORF of I factors (11), and a slightly different cysteine-rich segment is also present in the carboxyl terminus portion of L1 ORFs (38).

DISCUSSION

F elements are DNA sequences quite distinct from most *Drosophila* transposable elements (13). The structural organization of this DNA family, which consists of 60–80 units, is reminiscent of that of L1 sequences, a major repetitive DNA family of dispersed DNA sequences present in 10^4–10^5 copies in several mammalian genomes (22, 23). Members of both families similarly terminate at the 3′ end in a run of adenosine residues preceded by polyadenylylation signals, differ in size because of variable truncation at the 5′ end, and are flanked by target-site duplications of different length (13, 23).

The relationship between these two distant sequence types is further substantiated by the finding that they potentially

FIG. 2. The complete nucleotide sequence of the element Fw is shown. The 12-bp repeats of white gene DNA that flank the element have been reported (13). The amino acid sequence of the two ORFs, F-ORF1 and F-ORF2, is given below the base sequence. Cysteine-rich motifs present in F-ORF1 are underlined, as are the two polyadenylylation signals that precede the oligo(A) end of the element.

encode homologous polypeptides. Fw is an F element that is throughout homologous to other family members analyzed but is 1.1 kb shorter at the 5′ end (13). F-ORF2, a 859-amino acid ORF located upstream of the 3′ oligo(A) terminus of Fw, encodes a polypeptide that exhibits significant homology to the central portion of the potential L1 products (Fig. 4). The presence within the homologous region of amino acid motifs invariably found in most reverse transcriptases suggests that F elements, as proposed for L1 sequences (26, 27), might originate from the self-mediated cDNA conversion of element-specific transcripts. According to this hypothesis the 5′ heterogeneity of individual family members such as Fw would be a consequence of premature stops in the reverse transcription process.

The 5′-truncated copies generally exceed full-length L1 sequences, and complete functional units might represent <10% of the family size (39). F elements similarly exhibit 5′ heterogeneity (13). In this context it is noteworthy that the most 3′ segment of F elements is highly homologous to *suffix*, a short interspersed DNA element reiterated ≈300 times per *Drosophila* haploid genome (Fig. 7; see refs. 40 and 41). Remarkably, the homology between F and *suffix* is restricted to the 3′ region immediately following F-ORF2. At present the relationship between F elements is puzzling, since none of the *suffix* elements analyzed is flanked by target site duplications (41). Whether *suffix* sequences represent a specific truncated class of cDNA copies of F elements transcripts or F sequences have a composite origin and derive from the joining of distinct elements cannot yet be determined.

A short ORF (F-ORF1) located at the 5′ terminus of Fw encodes a cysteine-rich polypeptide that is structurally homologous to several nucleic acid binding proteins that orig-

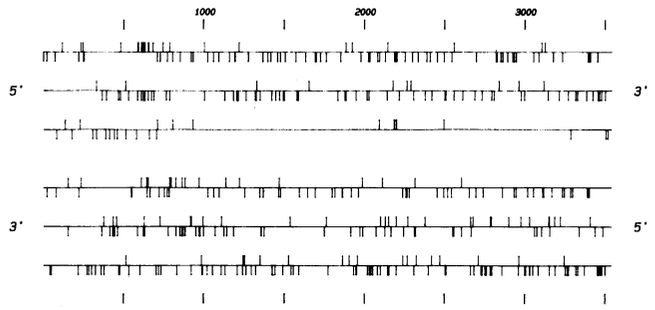

FIG. 3. ATG and stop codons within Fw sequence are shown. The six possible reading frames are represented as horizontal lines. Vertical lines above indicate ATG codons; lines below denote translational stop codons.

inated from cleavage of retroviral gag polyproteins (Fig. 2). The product of F-ORF1 may interact with F DNA or RNA and, therefore, have an important role in the transposition process. Because of the Fw truncation, however, it remains to be established whether this polypeptide might actually be encoded by functional F family members. Interestingly, a cysteine-rich segment located at the carboxyl terminus of L1 ORFs suggests that an analogous domain might be present in the potential L1 product (38).

A remarkable similarity is found in the sequence organization of F elements and I factors, the genetic determinants of the I-R system of hybrid dysgenesis. I factors lack terminal repetitions, generate upon insertion target site duplications of variable length (11), and potentially encode two proteins, the first of which contains a cysteine-rich domain similar to that present in F-ORF1, whereas the second has a reverse transcriptase domain homologous to those encoded by L1 ORFs and I-ORF2 (refs. 11; see *Results* and Fig. 6). The similarity in structure and potential gene products between F elements, I factors, and L1 elements suggests that all these sequences might have a common origin and transpose through a related mechanism via RNA intermediates.

We have not analyzed the relatedness of F elements to I and L1 sequences at the DNA level in great detail. The homology detected by comparing Fw DNA with I and mouse L1 DNA is rather low (41% and 39%, respectively). The homology is slightly higher between the segments encoding the corresponding reverse transcriptase domains (46% and 44% between Fw and I and L1Md, respectively).

```
Fw   ANTFAAHLQNVFTPNQATSTFALPSVPVNRHQQHTPIVFRPKEITKIIKDNLSPKKS-----
I    ANIFAQHFSDLSGDWNFSEEFRNNKYRNNIHLYTPSPIAQTIEENITYLELSSALQTLKGCA
     •• •• •  ++ +  + +  •        • •   ++ +  + •       +  ••  +

     PGYDLITPEMIIQLPHSAVRYITKLFNAITKLGYFPQRWKMMKIIMIPKPGKNHTVASSYRP
     PGLNRISYQMIKNSSHTTKNRITKLFNEIFN-SHIPQAYKTSLIIPILKPNTDKTKTSSYRP
     •• + •+ +•• + ++++   •••••• •   +  •• ••   •• • ••  +++ +•••••

     ISLLSCISKLFEKCLLIRLNQHQTYHNIIPAHQFGFRESHGTIEQVNRITTEIRTAFEYREY
     ISLNCCIAKILDKIIAKRLWWLVTYNNLINDKQFGFKKGKSTSDCLLYVDYLITKSKMHT--
     •••  •••+• •• +  ••    •• •••  +•••• ++++ + +  +   •  +

     CTAVFLDVSQAFDKVWLDGLMF-KIKISLPESTHKLLKSYLYDRKFAVRCNTATS-TVHTIE
     -SLVTLDFSRAFDRVGVHSIIQQLQEWKTGPKIIKYIKNFMSNRKITVRVGPHTSSPLPLFN
      + • •• • •••+• + +++        +    • •• •• +•• •••  + •• ••   +

     AGVPQGSVLGPTLYLI-YTADIPTNSR---LTVSTFADDTAILSRSRSPIQATAQLALYLID
     -GIPQGSPISVILFLIAFNKLSNIISLHKEIKFNAYADDFFLIINFNKNTNTNFNLDNLFDD
      •+••••• ••  •••• +      •    +   ++••••  ••       ••  ••     •

     IKKWLSDWRIKVNEQKCKHVTFTLNRQDCPPLLLNSIPLPKADEVTYLGVHLDRRLTWRRHI
     IENWCSYSGASLSLSKCQHLHICRKRHCTCKISCNNFQIPSVTSLKILGITLNNKYKWNTHI
     •  • •     +   •• ••     •    +  •   +•    +  •••  •• + +  •  ••

     EAKKTQLKLKANNLHWLINSGSPLSLDHKVLLYNSILKPIWTYGSQLWGNASNSNIDIIQRA
     NLLLPKLHNKLNIIKCLSSLKFNCNTHTLLNVAKATIIAKLEYGLFLYGHAPKSILNKIKTP
     +   + •• • • •• •             + +  + + +   ••  •+• •• • •• •  +

     QSKILRTITGAPWYVRSENIQRDLNIPSVTNAITELKEKYL
     FNSAIRLALGAYRSTPINNLLYESNTPPLEMKRDLQTAKLS
         +•   ••      +••  + • •••       •
```

FIG. 5. Homology between F-ORF2 and I factor ORF-2. F-ORF2 is shown from amino acid residue 396 to the end; I-ORF2 (11) is from residue 255 to residue 724. Dots denote identical residues; crosses show favored amino acid substitutions. The single-letter amino acid code is used.

The identification of element specific transcripts, as well as the analysis of extrachromosomal circular copies from cultured cells and embryos that probably correspond to transposition intermediates (42), might partly clarify some of the steps involved in the generation of F elements. Transcripts that give rise to F elements presumably originate from one or few master elements that differ from most family members in that they contain a promoter, although at the moment the possibility that F elements carry an internal promoter cannot be ruled out. The transcription of functional F elements is tightly controlled and/or restricted to early stages of embryo development, since only rare heterogeneous poly(A)$^-$ RNA molecules homologous to F sequences are detectable in total embryos (12). In this context it should be mentioned that polyadenylylated cytoplasmic RNA corresponding in size to complete L1 units has been so far found exclusively in undifferentiated teratocarcinoma cells (43, 44).

The transposition of several *Drosophila* mobile elements can be influenced by changes in environmental conditions (45) and is notably enhanced by outcrossing (46, 47). Given the similarity between F and I factors, it might be interesting

```
Fw   SSRSPTAKQKLKVATRKLANALKQEEDDDQRRYIEQLTPTGT[ 18] VLPIKNSSGG---WARSDEDRANTFAAHLQNVFTPNQATSTFALPSYPVNRHQQHTPIV
L1Hs STLDRSTRQKINKDIQELNSALHQADLIDIYRTLHPKSTEYT[188] IDTIKNDKGDITTDPTEIQTTIREYYKHLYANKLENLEEMDKFLDTYTLPRLNQEEVES
L1Md SSKDRSWKQKLNRDTVKLTEVMKQMDLTDIYRTFYPKTKGYT[198] INKIRNEKGDITTDPEEIQNTIRSFYKRLYSTKLENLDEMDKFLDRYQVPKLNQDQVDH
     •••+  + ••••   • •+++ +•• + •  • +  •++ •     + +•••  •    +  ++   +•  ••       •     • +•+• • ••

FR----PKEITKIIKDNLSPKKSPGYDLITPEMIIQLPHSAVRYITKLFNAITKLGYFPQRWKMMKIIMIPKPGKNHTVASSYRPISLLSCISKLFEKCLLIRLNQHQTYHN
LNRPITSSEIEAIIN-SLPNKKSPGPEGFTAEFYQRYKEELVPFLLKLFQSIEKEGILPNSFYEASIILIPKPGRDTTKKENFRPISLMNIDAKILNKILANQIQQHI--KK
LNSPISPKEIEAVIN-SLPTKKSPGPDGFSAEFYQTFKEDLIPILHHKFHKIEVEGTLPNSFYEATITLIPKPQKDPTKIENFRPISLMNIDAKILNKILANRIQEHI--KA
     •••• ••   •++•••••• •  •••          • ++ •  +++ • •  •+ +    •••••••••+ •    +••••••+   •+• +• •  •+••+   +

IIPAHQFGFRESHGTIEQVNRITTEIRTAFEYREYC-TAVFLDVSQAFDKVWLDGLMF-KIKISLPESTHKLLKSYLYDRKFA-VRCNTATSTVHTIEAGVPQGSVLGPTLY
LIHHDQVGFIPAMQGWFNIRKSINIIQHINRTKDTNHMIISIDAEKAFDKIQQPFMLKPLNKLGIDGTYLKIIRA-IYDKPTANIILNGQKLEAFPLKTGTRQGCPLSPLLF
IIHPDQVGFIPGMQGWFNIRKSINVIHYINKLKDKNHMIISLDAEKAFDKIQHPFMIKVLERSGIQGPYLNMIKA-IYSKPVANIKVNGEKLEAIPLKSGTRQGCPLSPYLF
••   • ••  +  +  ++ +    ••     ++     + •   ••••+    ++    •+++ ++  •++•+ +•••  • ++ ••       ++ +•  ••  •+• ••

LIYTADIPTNSR--------------LTVSTFADDTAILSRSRSPIQATAQLALYLIDIKKWLSDWRIKVNEQKCKHVTFTLNRQDCPPLLLNSIPLPKADEVTYLGVHLDR
NIVLEVLARAIRQEKEIKGIQLGKEEVKLSLFADDMIVYLENPIV---SAQNLLKL--ISNFSKVSGYKINVQKSQAFLYTNNRQTESQIMSELPFTIASKRIKYLGIQLTR
NIVLEVLARAIRQQKEIKGIQIGKEEVKISLLADDMIVYSDPKN----STRELLNL--INSFGEVVGYKINSNKSMAFLYTKNKQAEKEIRETTPFSIVTNNIKYLGVTLTK
 •    ++   •               + +• •••• + • ++     +••  • •  •  +      •+• •:     +• •••  + ++ ++     ++++ •••• • •

RLTWRRHIEAKKTQLKLKANNLHWLINSGSPLSLDHKVLLYNSILKPIWTYGSQ-LWGNASNSNIDIIQRAQSKILRTITGAPWYVRSENIQRDLNIPSVTNAITELKEKYL
DVKDLFKENYKPLLNEIKEDTNKWKNIPCS--WVGRINIVKMAIL-PKVIYRFNAIPIKLPMTFFTELEKTTLKFIWNQKRAHIAKATISQKNKAGGITLPDFKLYYKATVT
FVKDLYDKNFKSLKKEIKEDLRRWKDLPCS--WIGRINIVKMAIL-PKAIYRFNAIPIKIPTQFFNELEGAICKFVWNNKKPRIAKSLLKDKRTSGGITMPDLKLYYRAIVI
 +    + + •     +• +  ••   + •   + +  ++  +•• •   •  • + +    + +  + ++ •  • +      •     +     •      ++++      •   +
```

FIG. 4. F-ORF2 is aligned to human and mouse L1 ORFs. L1Hs (*Homo sapiens*) is the consensus sequence derived by Hattori *et al.* (26), L1Md (*Mus domesticus*) corresponds to the sequence of the clone L1Md-A2 (27). Numbers in brackets refer to the residues that separate homologous segments in the three ORFs. F-ORF2 is shown from amino acid residue 318 to the end; L1Hs and L1Md ORFs are from amino acid residues 156 and 176, respectively. Filled circles below sequence lines denote identical residues in F-ORF2 and any of the two mammalian sequences. Crosses indicate favored amino acid substitutions, grouped as follows: (alanine, serine, threonine, proline, and glycine), (an unspecified amino acid, aspartic acid, glutamic acid, and glutamine), (histidine, arginine, and lysine), (methionine, leucine, isoleucine, and valine), and (phenylalanine, tyrosine, and tryptophan) (28). Identical residues and/or favored substitutions between L1 sequences are not highlighted. Gaps introduced to maximize homology are indicated by dashes. The single-letter amino acid code is used.

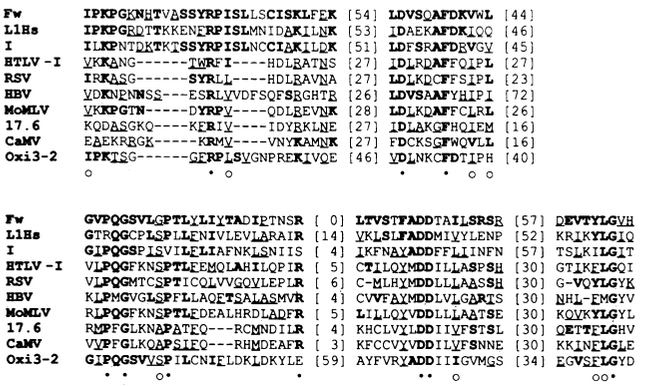

FIG. 6. F-ORF2 segments are aligned (using the single-letter amino acid code) with homologous regions from several known or hypothetical reverse transcriptases. Bold-face characters denote identical residues; underlined characters denote chemically related amino acids grouped as described in the legend to Fig. 4. Dots and open circles indicate positions, respectively, occupied by identical or chemically similar amino acids among a large group of reverse transcriptases (30). L1Hs, *Homo sapiens* L1 consensus sequence (ref. 26, see also Fig. 4); I, *Drosophila* I factor ORF-2 (11); HTLV-I, human adult T-cell leukemia virus type I (31); RSV, Rous sarcoma virus (32); HBV, hepatitis B virus (33); MoMLV, Moloney murine leukemia virus (34); 17.6, *Drosophila copia*-like transposable element (6); CaMV, cauliflower mosaic virus (35); Oxi3-2, a2 intron of *S. cerevisiae* mitochondrial cytochrome oxidase subunit I gene (36). Numbers in brackets refer to amino acid residues present between the reported regions. Gaps introduced to maximize homology are represented by dashes.

to investigate whether crosses that mobilize I factors could trigger the transposition of F elements. This type of approach might furnish an experimental system for more detailed analysis of the processes that control the dispersion and the maintenance of F sequences within the *Drosophila* genome.

We thank Paolo Monaci for contributing to the initial steps of this work, Edoardo Boncinelli and Michael Etzerodt for helpful discussions, and Graham Tebb for careful revision of the manuscript. We are indebted to Giovanna Grimaldi for suggestions and critical reading of the manuscript. G.C. was a recipient of a short-term fellowship granted by the Italian Ministry of Education.

1. Spradling, A. C. & Rubin, G. M. (1981) *Annu. Rev. Genet.* **15,** 219–264.
2. Rubin, G. M. (1983) in *Mobile Genetic Elements*, ed. Shapiro, J. A. (Academic, New York), pp. 329–361.
3. Finnegan, D. J. (1985) *Int. Rev. Cytol.* **93,** 281–326.
4. Kleckner, N. (1981) *Annu. Rev. Genet.* **15,** 341–404.
5. Shiba, T. & Saigo, K. (1983) *Nature (London)* **302,** 119–124.
6. Saigo, K., Kugimiya, W., Matsuo, Y., Inouye, S., Yoshioka, K. & Yuki, S. C. (1984) *Nature (London)* **312,** 659–661.
7. Emori, Y., Shiba, T., Kanaya, S., Inouye, S., Yuki, S. & Saigo, K. (1983) *Nature (London)* **315,** 773–776.
8. Inouye, S., Yuki, S. & Saigo, K. (1986) *Eur. J. Biochem.* **154,** 417–425.
9. Yuki, S., Inouye, S., Ishimaru, S. & Saigo, K. (1986) *Eur. J. Biochem.* **158,** 403–410.
10. Yuki, S., Ishimaru, S., Inouye, S. & Saigo, K. (1986) *Nucleic Acids Res.* **7,** 3017–3030.
11. Fawcett, D. H., Lister, C. K., Kellett, E. & Finnegan, D. J. (1986) *Cell* **47,** 1007–1015.
12. Dawid, I. B., Long, E. O., Di Nocera, P. P. & Pardue, M. L. (1981) *Cell* **25,** 399–408.
13. Di Nocera, P. P., Digan, M. E. & Dawid, I. B. (1983) *J. Mol. Biol.* **168,** 715–728.
14. Di Nocera, P. P. & Dawid, I. B. (1983) *Nucleic Acids Res.* **11,** 5475–5482.
15. Di Nocera, P. P., Graziani, F. & Lavorgna, G. (1986) *Nucleic Acids Res.* **14,** 675–690.
16. Wilde, C. D. (1986) *Crit. Rev. Biochem.* **19,** 323–352.

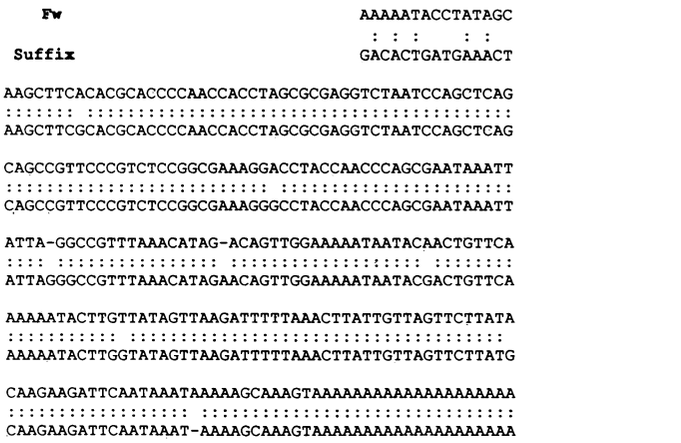

FIG. 7. Homology between the 3′ end of Fw and *suffix* element. The sequence of *suffix* is from Tchurikov *et al.* (41). Colons, identical nucleotides.

17. Rogers, J. H. (1985) *Int. Rev. Cytol.* **93,** 187–279.
18. Dawid, I. B. & Pardue, M. L. (1981) *Chromosoma* **83,** 29–43.
19. Levis, R., O'Hare, K. & Rubin, G. M. (1984) *Cell* **38,** 471–481.
20. Scott, M. P., Weiner, A. J., Hazelrigg, T. I., Polisky, B. A., Pirrotta, V., Scalenghe, F. & Kaufman, T. C. (1983) *Cell* **35,** 763–776.
21. Karesse, R. E. & Rubin, G. M. (1982) *Cell* **30,** 63–69.
22. Singer, M. F. (1982) *Int. Rev. Cytol.* **76,** 67–112.
23. Singer, M. F. & Skowronski, J. (1985) *Trends Biol. Sci.* **10,** 119–122.
24. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
25. Kozak, M. (1984) *Nucleic Acids Res.* **12,** 857–872.
26. Hattori, M., Kuhara, S., Takenaka, O. & Sakaki, Y. (1986) *Nature (London)* **321,** 625–628.
27. Loeb, D. D., Padgett, R. W., Hardies, S. C., Shehee, W. R., Comer, M. B., Edgell, M. H. & Hutchison, C. A. (1986) *Mol. Cell. Biol.* **6,** 168–182.
28. Schwartz, R. M. & Dayhoff, M. O. (1978) in *Atlas of Protein Sequence and Structure*, ed. Dayhoff, M. O. (Natl. Biomed. Res. Found., Washington, DC), Vol. 5, pp. 353–358.
29. Burton, F. H., Loeb, D. D., Voliva, C. F., Martin, S. M., Edgell, M. H. & Hutchison, C. A. (1986) *J. Mol. Biol.* **187,** 291–304.
30. Toh, H., Kikuno, R., Hayashida, H., Miyata, T., Kugimiya, W., Inouye, S., Yuki, S. & Saigo, K. (1985) *EMBO J.* **4,** 1267–1272.
31. Seiki, M., Hattori, S., Hirayama, Y. & Yoshida, M. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 3618–3622.
32. Schwartz, D. E., Tizard, R. & Gilbert, W. (1983) *Cell* **32,** 853–869.
33. Galibert, F., Mandart, E., Fitoussi, F., Tiollais, P. & Charnay, P. (1979) *Nature (London)* **281,** 646–650.
34. Shinnick, T. M., Lerner, R. A. & Sutcliff, J. G. (1981) *Nature (London)* **293,** 543–548.
35. Gardner, R. C., Howarth, A. J., Hahn, P., Brown-Luedi, M., Shepherd, R. J. & Messing, J. (1981) *Nucleic Acids Res.* **9,** 2871–2887.
36. Bonitz, S. G., Coruzzi, G., Thalenfeld, B. E., Tzagaloff, A. & Macino, G. (1980) *J. Biol. Chem.* **255,** 11927–11941.
37. Covey, S. N. (1986) *Nucleic Acids Res.* **14,** 623–633.
38. Fanning, T. & Singer, M. F. (1987) *Nucleic Acids Res.* **15,** 2251–2260.
39. Grimaldi, G., Skowronski, J. & Singer, M. F. (1984) *EMBO J.* **3,** 1753–1759.
40. Tchurikov, N. A., Naumova, A. K., Zelentsova, E. S. & Georgiev, G. P. (1982) *Cell* **28,** 365–373.
41. Tchurikov, N. A., Ebralidze, A. K. & Georgiev, G. P. (1986) *EMBO J.* **5,** 2341–2347.
42. Mossie, K. G., Young, M. W. & Varmus, H. E. (1985) *J. Mol. Biol.* **182,** 31–43.
43. Skowronski, J. & Singer, M. F. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 6050–6054.
44. Skowronski, J. & Singer, M. F. (1986) *Cold Spring Harbor Symp. Quant. Biol.* **51,** 457–464.
45. Junakovic, N. & Angelucci, V. (1986) *J. Mol. Evol.* **24,** 83–88.
46. Rubin, G. M. & Spradling, A. C. (1982) *Science* **218,** 348–353.
47. Gerasimova, T. I., Mizrokhi, L. J. & Georgiev, G. P. (1984) *Nature (London)* **309,** 714–716.