MOLECULAR AND CELLULAR BIOLOGY, July 1989, p. 3018–3027
0270-7306/89/073018-10$02.00/0

Vol. 9, No. 7

Composite Transposable Elements in the *Xenopus laevis* Genome

JAMES E. GARRETT,† DEBORAH S. KNUTZON,‡ AND DANA CARROLL*
Department of Biochemistry, University of Utah School of Medicine, Salt Lake City, Utah 84132

Received 7 March 1989/Accepted 21 April 1989

Members of two related families of transposable elements, Tx1 and Tx2, were isolated from the genome of *Xenopus laevis* and characterized. In both families, two versions of the elements were found. The smaller version in each family (Tx1d and Tx2d) consisted largely of two types of 400-base-pair tandem internal repeats. These elements had discrete ends and short inverted terminal repeats characteristic of mobile DNAs that are presumed to move via DNA intermediates, e.g., *Drosophila* P and maize Ac elements. The longer versions (Tx1c and Tx2c) differed from Tx1d and Tx2d by the presence of a 6.9-kilobase-pair internal segment that included two long open reading frames (ORFs). ORF1 had one cysteine-plus-histidine-rich sequence of the type found in retroviral *gag* proteins. ORF2 showed more substantial homology to retroviral *pol* genes and particularly to the analogs of *pol* found in a subclass of mobile DNAs that are supposed retrotransposons, such as mammalian long interspersed repetitive sequences, *Drosophila* I factors, silkworm R1 elements, and trypanosome Ingi elements. Thus, the Tx1 elements present a paradox by exhibiting features of two classes of mobile DNAs that are thought to have very different modes of transposition. Two possible resolutions are considered: (i) the composite versions are actually made up of two independent elements, one of the retrotransposon class, which has a high degree of specificity for insertion into a target within the other, P-like element; and (ii) the composite elements are intact, autonomous mobile DNAs, in which the *pol*-like gene product collaborates with the terminal inverted repeats to cause transposition of the entire unit.

Once investigators accepted the concept of mobile DNA segments within otherwise stable genomes (37), the catalog of such transposable elements began to grow steadily (13, 19, 23, 29, 31, 47). No phylogenetic group has been shown to be free of them, and they are presumed to be ubiquitous. Transposable elements can be defined as discrete segments of DNA that are capable of occupying new chromosomal locations by either excisional or replicative mechanisms. In general, all transposable elements encode at least one product necessary to promote their own transposition.

It has been possible to assign the many eucaryotic transposable elements to a few classes on the basis of sequence organization and actual or presumed mode of transposition (19, 23; Table 1). Class I is typified by the Ac elements of maize (12), the P elements of *Drosophila* spp. (42), and the Tc1 elements of nematodes (14). They are characterized by short inverted terminal repeats (less than about 60 base pairs [bp]) and target duplications of reproducible length (in the range of 2 to 9 bp), with some sequence specificity. It is generally believed that class I elements transpose through DNA intermediates. In the case of Ac (also called Mp), it has been shown that excision from a preexisting site accompanies insertion at a new location (25). There is circumstantial evidence for an excision mechanism for P and Tc1 elements, but transposition could also be replicative, as with the bacterial elements Mu or Tn*3*. There is no suggestion of an RNA intermediate in the movement of any class I element.

Class II elements are the retroviruslike elements, such as Ty1 of yeast cells (18), the many families of copialike elements in *Drosophila* spp. (23), and the IAP elements of mammals (23). These elements have all the features of genuine retroviruses except for an extracellular transmissible form. In the genome, they have long terminal (direct) repeats (LTRs) and are flanked by short target duplications that only rarely appear to be specific. They contain coding sequences that are readily identifiable homologs of retroviral *gag* and *pol* genes; the latter include reverse transcriptase, protease, and integrase domains. Viruslike particles containing full-length RNAs and reverse transcriptase in a *gag*-like protein coat have been identified for some class II elements (21, 38, 48), and reliance on an RNA intermediate in transposition has been demonstrated directly in the case of Ty1 (2).

Elements of class III are often called nonviral or non-LTR retrotransposons, reflecting presumptions about their lifestyles. The best characterized examples are mammalian long interspersed repetitive sequences (LINEs) (9, 27, 35, 49), *Drosophila* I factors (15) and F elements (11), silkworm R1 (54) and R2 (3) elements, trypanosome Ingi elements (30), and maize Cin4 elements (46). These elements are characterized by the absence of terminal repeats, relatively large and variable flanking target duplications, and an A-rich region at their 3′ ends. The longest members in each family have two long open reading frames (ORFs), one of which shows modest similarity to genuine reverse transcriptases and the other of which could be a distant relative of viral *gag* genes. For these reasons, it has been suggested that class III elements move via reverse transcription of an RNA intermediate. The 5′-end truncation seen in some families is attributed to incomplete copying of the RNA template (51).

These classes include most but not all of the eucaryotic elements described to date. It is possible that a fourth distinct class exists, characterized by long inverted terminal repeats and including *Drosophila* FB (50) and maize Mu (1) elements.

We have been examining a family of transposable elements, called Tx1, from *Xenopus laevis* (4, 5, 22). There are roughly 1,000 members of this family distributed throughout the *X. laevis* genome. Most Tx1 elements have a structure which suggests that they belong to class I: there is a 19-bp inverted terminal repeat, a specific 4-bp apparent target

* Corresponding author.
† Present address: Institute for Advanced Biomedical Research, Oregon Health Sciences University, Portland, OR 97201.
‡ Present address: Calgene, Inc., Davis, CA 95616.

TABLE 1. Classes of eucaryotic transposable elements

Class	Characteristic				Examples
	Terminal repeats	Target duplication	Genes	Intermediate	
I	Short, inverted	Short	Transposase, regulation	DNA	P, Ac, Tc1
II	Long, direct (LTR)	Short	*gag*, *pol*	RNA	Ty1, copia, IAP
III	None	Longer, variable	*gag*-like, *pol*-like	RNA (?)	L1Md, I, R1Bm

duplication, and good evidence of a mobile lifestyle (22). The sequenced version, however, contains no obvious genes; therefore, we were led to investigate other members of the family. The results presented here demonstrate that a subset of Tx1 elements contains potential protein-coding regions and that these ORFs show considerable homology to those of class III elements. In addition, there is a second family of *Xenopus* elements, Tx2, members of which have very similar structures and are also found in versions both with and without coding sequences. By exhibiting features of both class I and class III elements, the Tx1 and Tx2 elements blur the boundary between these classes and pose questions concerning transposition mechanisms for both classes.

MATERIALS AND METHODS

Two separate libraries of *X. laevis* genomic DNA fragments were obtained from the laboratory of Igor Dawid. The Wahli library consists of mixed *Alu*I and *Hae*III partial digests of the DNA from several different frogs cloned into λCharon 4 (52). The HD-I library consists of *Sau*3AI partial-digest fragments from a single, homozygous frog cloned in the λEMBL4 vector (Z. Jonas and I. B. Dawid, personal communication). The Tx1 clones B4 and B10 from the Wahli library have been described previously (4, 22), as have methods for DNA isolation and analysis.

A 4.9-kbp *Hin*dIII fragment from the central region of λB10 was subcloned into the *Hin*dIII site of pBR322, and the clone was designated pN3. This was used to screen both libraries by plaque hybridization (36). A 1.3-kbp *Eco*RI-*Bam*HI subfragment of pN3 was cloned between the corresponding sites of pBR322 and designated pN1. Other subfragments of Tx1 used as probes were largely from λB4 (22). Corresponding subclones from Tx2 elements will be described in detail elsewhere. In many experiments, the plasmid inserts or subfragments of them were excised with restriction enzymes, separated by agarose gel electrophoresis, and recovered and concentrated on glass beads (GENECLEAN; Bio 101, Inc., La Jolla, Calif.). Probes were made radioactive by incorporation of α-^{32}P-labeled nucleotides, using either nick translation (36) or random-primer extension (17).

Plaque hybridizations were performed on nitrocellulose (Schleicher & Schuell, Inc., Keene, N.H.) or Biotrans (ICN Pharmaceuticals, Irvine, Calif.) membranes. Restriction enzyme digests of selected clones were screened with the same probes by Southern blot hybridization on nitrocellulose or Transphor (Hoeffer Scientific Instruments) membranes. Most hybridizations were done at 42°C in 5× SSC (SSC is 0.15 NaCl plus 0.015 sodium citrate)–50% formamide, but some were performed at 60°C in modified Church buffer (7) (7% sodium dodecyl sulfate, 0.5 M sodium phosphate [pH 7.2], 0.001 M EDTA), with equivalent results.

For DNA sequencing, subclones of fragments of λB10 (Tx1) and λ10n (Tx2), both from the Wahli library, were made in the plasmid vectors pEMBL18+, pEMBL18−, pEMBL19+ (10), and Bluescript KS+ (Stratagene, La Jolla, Calif.). Deletion series were generated by the method of Henikoff (28). Single strands were produced by superinfection, with f1 helper bacteriophage IR1 (10) or R408 (44), of NM522 cells (24) carrying the desired plasmid clones. To obtain reasonable yields of plasmid-derived single strands, it was important to use fresh transformants for superinfection. Plasmid DNA (usually from a minipreparation) was introduced into competent NM522 cells and plated on ampicillin-containing L agar (36). The following day, a single colony was picked, grown up, and superinfected as described (10, 44). Single-stranded DNA was isolated from the culture supernatant by polyethylene glycol precipitation, phenol and chloroform extraction, and isopropanol precipitation (10). When the ratio of plasmid-derived to helper phage single strands was ≥1, as analyzed by electrophoresis, the sequencing reactions were usually free of artifacts. Sequences were determined by the chain termination method (45), using Klenow DNA polymerase (Boehringer Mannheim Biochemicals, Indianapolis, Ind.) or T7 DNA polymerase (Sequenase; U.S. Biochemical Corp., Cleveland, Ohio). DNA sequences were analyzed by using the Intelligenetics BIONET program package.

RESULTS

Subclasses of Tx1 elements. The Tx1 elements that we characterized previously are represented by that in λ clone B4 (4, 22). Its structure is shown schematically in Fig. 1 as Tx1d. These elements have 19-bp inverted terminal repeats, two types of tandem internal repeats (PTR-1 and PTR-2, each about 400 bp long), and unique segments near each end, designated LCF and RCF. Nucleotide sequence analysis of all of these segments revealed no convincing ORFs (22). Because complete versions of essentially all known transposable elements encode products necessary for their own movement, we sought other versions of Tx1 that might contain such genes.

Among the elements in the eight λ Tx1 clones that we isolated originally, one (B10) differed from the others in having an interruption of about 7 kbp in the PTR-1 cluster (4) (Tx1c in Fig. 1). To determine whether this was a unique occurrence or represented a major subclass of Tx1 structures, we prepared probes from the interruption, including a 1.3-kbp *Eco*RI-*Bam*HI fragment that was subcloned in a plasmid vector as pN1 (Fig. 1). When this was hybridized to digests of *X. laevis* genomic DNA, it detected fragments, the most prominent of which reflected the restriction map predicted from clone B10 (Fig. 2). By comparison with intensities observed with measured amounts of plasmid DNAs in parallel lanes (not shown), we estimated that these fragments were present in the genome at a copy number of roughly 100.

The longer probe, pN3, was used to probe two λ libraries of genomic DNA fragments (see Materials and Methods). With this approach, we not only isolated additional clones of the B10 subclass but discovered a family of related elements that we call Tx2. Only the central regions, represented by pN3, cross-hybridized between Tx1 and Tx2, but further

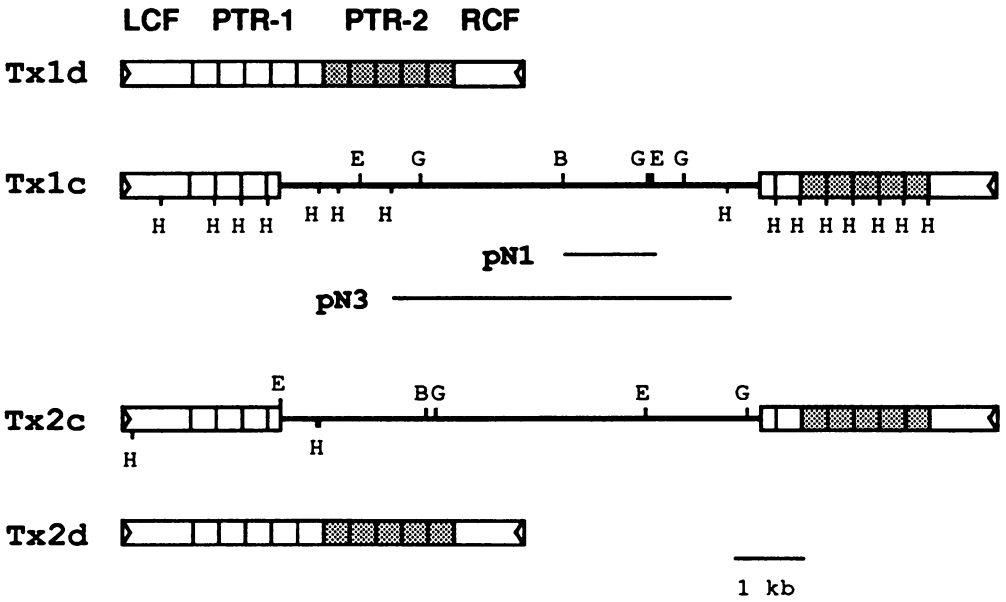

FIG. 1. Diagrams of the structures of elements in the Tx1 and Tx2 families. The map of Tx1d is based on the element in λ clone B4, the sequence of which has been reported (22). Two types of approximately 400-bp internal repeats, PTR-1 and PTR-2 (shaded), are flanked by unique sequences, LCF and RCF. The elements terminate in 19-bp inverted repeats (triangles). The map of Tx1c is that of the element in λ clone B10, which was reported previously (4), the central portion of which was sequenced in this study. Corresponding symbols represent the same components as in Tx1d. The maps for Tx2c and Tx2d elements are composites for a number of different clones; they are drawn to emphasize correspondence to segments of Tx1 elements. In the maps of Tx1c and Tx2c, the locations of sites for restriction enzymes used in the experiment presented in Fig. 2 are shown: B, *Bam*HI; E, *Eco*RI; G, *Bgl*II; and H, *Hin*dIII. Below the Tx1c map are indicated the subcloned fragments used as probes in library screens and Southern blots.

analysis showed that the Tx2 elements contained sequences fully analogous to LCF, PTR-1 repeats, PTR-2 repeats, and RCF. Details of these sequences will be presented elsewhere. The presence of Tx2 sequences in the genome was confirmed by the data in Fig. 2, showing that the secondary bands were those predicted by the restriction maps of Tx2 clones. Their intensities were lower, probably because of imperfect complementarity with the Tx1-derived pN3 probe.

Like Tx1, Tx2 was found in versions both with and without the central region (Fig. 1). We call the longer versions Tx1c and Tx2c (for composite elements) and the shorter versions Tx1d and Tx2d (deleted in comparison with versions c).

The results of screening the two *X. laevis* genomic libraries with probes from various regions of the Tx1 and Tx2 elements are presented in Table 2. There appeared to be roughly equal numbers of the two types of element in the genome. In both cases, approximately 20% of the members of each family contained the central region; that is, they hybridized with pN3. For unknown reasons, the proportion of clones containing any Tx sequences and the fraction of those including the central region were both lower in the HD-I than in the Wahli library. Because both libraries had been amplified before screening, the numbers in Table 2 are not a perfect reflection of genomic abundance; they are, however, in general accord with estimates made from Southern blot hybridization experiments.

The library screens and the genomic blots showed that type c elements are relatively common in both families. Estimates of the abundance of the various elements in the *X. laevis* genome are given in Table 2. In addition, the initial screens yielded a total of 18 clones that hybridized with pN3 but apparently not with probes for other Tx1 or Tx2 segments. Upon rescreening, five of these failed to hybridize with pN3, while seven others showed hybridization with PTR-1 probes. Of 201 clones that hybridized with pN3, only 6 were truly $pN3^+$ $Tx1^-$ $Tx2^-$. Restriction mapping and hybridization with various probes showed that (i) three of these clones were missing one end as a result of interruption by cloning, so deletion of one end (the left end in all cases) could account for their lack of other Tx1 or Tx2 sequences; (ii) two clones appeared to contain insertions that made the elements too large to be contained within a single cloned fragment; and (iii) one clone hybridized only very weakly with probes for a small portion of the central region. Thus, there is no compelling evidence for a repository of central regions outside the Tx elements themselves.

Nucleotide sequence analysis. Bolstered by the knowledge that the central regions of Tx1c and Tx2c were present in many members of the families, we undertook a determination of their complete nucleotide sequences in the hope of finding the genes of the elements. The sequence of the Tx1c central region (from clone B10) is presented in Fig. 3. The entire central region is 6,901 bp long. Novel sequence begins after nucleotide 205 of a typical PTR-1 repeat (4), and nucleotides 183 to 205 of PTR-1 are present as a 23-bp direct duplication around the Tx1c-specific central region. We have sequenced these junctions for a number of Tx1c clones and find that the same 23-bp sequence is duplicated in each case.

The bulk of the central region is occupied by two long ORFs (Fig. 3). Counting the first nucleotide after the 23-bp repeat as position 1, ORF1 extends from positions 400 to 2878 and could encode 826 amino acids. The first methionine codon is at nucleotide 553; from there to the end of ORF1 represents 775 amino acids.

ORF2 begins at position 2871; the fourth codon is an ATG. This frame is open to position 6804, sufficient to encode

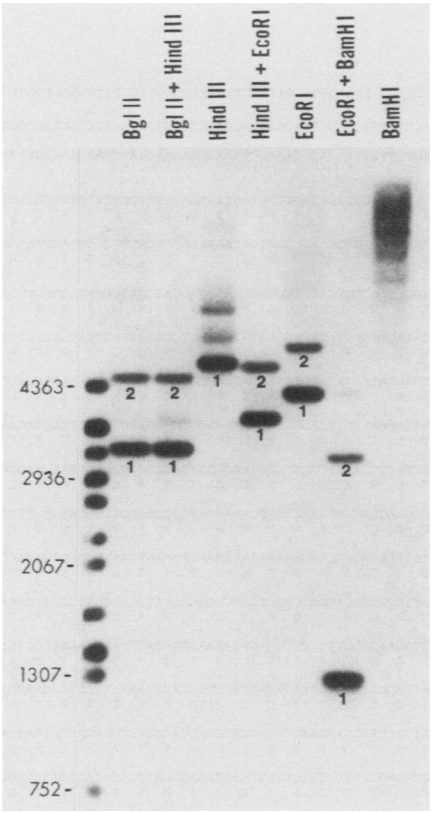

FIG. 2. Southern blot hybridization using the pN1 probe. The leftmost lane contains molecular weight markers, the sizes (in base pairs) of some of which are given on the left. The other lanes contain *X. laevis* chromosomal DNA digested with the enzymes indicated at the top. The numerals below various bands indicate that the bands are expected on the basis of the restriction maps of cloned versions of Tx1c (1) or Tx2c (2) (see Fig. 1). Because there is only a single site in Tx2c for *Hin*dIII, and in both elements for *Bam*HI, no discrete bands are predicted.

1,308 amino acids from the first methionine. ORF2 overlaps ORF1 by 7 bp (Fig. 3).

The corresponding Tx2c sequence is not presented here in detail. It is flanked by 21-bp repeats of identifiably the same region of a PTR-1 repeat as in the case of Tx1. ORF1 and ORF2 are present and correspond to the Tx1 sequences with high fidelity (Fig. 4). Close examination shows that the conservation in ORF2 particularly is that expected of genes under selective pressure on their protein products: there are many silent codon changes, and many of the amino acid substitutions are conservative. All of the sequences in the central regions are 80% or better matched between the families.

Homology with retrotransposons. Both the nucleotide and hypothetically translated sequences of Fig. 3 were compared with other sequences in the National Institutes of Health and EMBL data banks and specifically with sequences of known eucaryotic transposable elements. Because of the structure of the ends of Tx1 elements, we anticipated there might be similarity to genes of class I elements such as P and Ac. Surprisingly, convincing matches were found only with elements of class III, the non-LTR retrotransposons. ORF2 of Tx1c is clearly a relative of the longest ORF in mammalian LINEs, *Drosophila* I factors, *Drosophila* F elements, silkworm R1 and R2 elements, and trypanosome Ingi elements (Fig. 5).

TABLE 2. Screens of *X. laevis* genomic libraries with Tx1 and Tx2 probes

Library	Total no. of plaques screened	Hybridization[a]				Estimated no. in genome			
		Tx1$^+$		Tx2$^+$					
		Total	pN3$^+$	Total	pN3$^+$	Tx1c	Tx1d	Tx2c	Tx2d
HD-I	13,700	110	20	99	13	150	900	100	850
Wahli	15,600	302	95	320	67	630	1,800	450	2,200

[a] Reported are the total number of clones that hybridized with at least one of the probes (LCF, PTR-1, PTR-2, and RCF) specific for the Tx1 or Tx2 family. In most cases, a clone hybridized with all of these probes or a subset of adjacent regions (e.g., LCF + PTR-1), consistent with interruption of the element at one end of the cloned fragment. In each family, the number also hybridizing with a central-region probe (pN3) is given. A few clones hybridized with pN3 only; these are discussed in more detail in the text. The frequency of total clones and pN3$^+$ clones in each family and the average insert size (about 15 kbp) were used to calculate the apparent numbers of various element structures in the genomes that contributed to the HD-I and Wahli libraries.

The regions of best match between Tx1c ORF2 and the retrotransposon ORFs are those that show a low but consistent degree of similarity to genuine retrovirus *pol* gene products (Fig. 5). The homology extends well beyond these areas, however. In the 325-amino-acid region shown in Fig. 5, there is 29% identity between the long ORFs of Tx1c and L1Md (52% if conservative substitutions are allowed). This match between Tx1c and L1Md is as good as or better than that between any other pair in Fig. 5.

The smaller ORFs of the class III elements are much more divergent, but all have one or two of the cysteine-plus-histidine-rich motifs suggested to be Zn-binding fingers in viral *gag* proteins (8). ORF1 of Tx1c and Tx2c also have one such motif (Fig. 6). In addition to the consistently spaced Cys and His residues, there is a conserved Gly preceding His, and basic amino acids are common in this short segment.

Many but not all class III elements have A-rich tracts at their 3′ ends, which is taken as supporting evidence that they arise by reverse transcription of polyadenylated RNAs. Six of the last nine nucleotides at the right end of the Tx1c central region are A's (Fig. 3), but this is not a compelling poly(A) tract, nor is it preceded by a consensus polyadenylation signal. The Tx1 sequence, AATAATATA, bears some similarity to the $(TAA)_n$ repeats found at the right end of I-factor sequences (15).

DISCUSSION

We have described two closely related families of transposable elements in the genome of *X. laevis*, Tx1 and Tx2. In both families, there are two versions of the basic element (Fig. 1). One consists simply of two types of tandem internal repeats (PTR-1 and PTR-2) flanked by 0.5- to 1-kbp sequences of no evident function (LCF and RCF). The outside ends of the elements have short inverted repeats, and the elements are flanked by 4-bp apparent target duplications. The other version is identical to the first except for the presence of a 6.9-kbp segment containing two long ORFs inserted in the PTR-1 repeat cluster. We call the latter version Tx1c (Tx2c), for composite elements. The former version is Tx1d (Tx2d) because it is deleted in comparison with the longer version. There are approximately 100 to 600 composite versions of each family in the *X. laevis* genome and about 1,000 to 2,000 deleted elements.

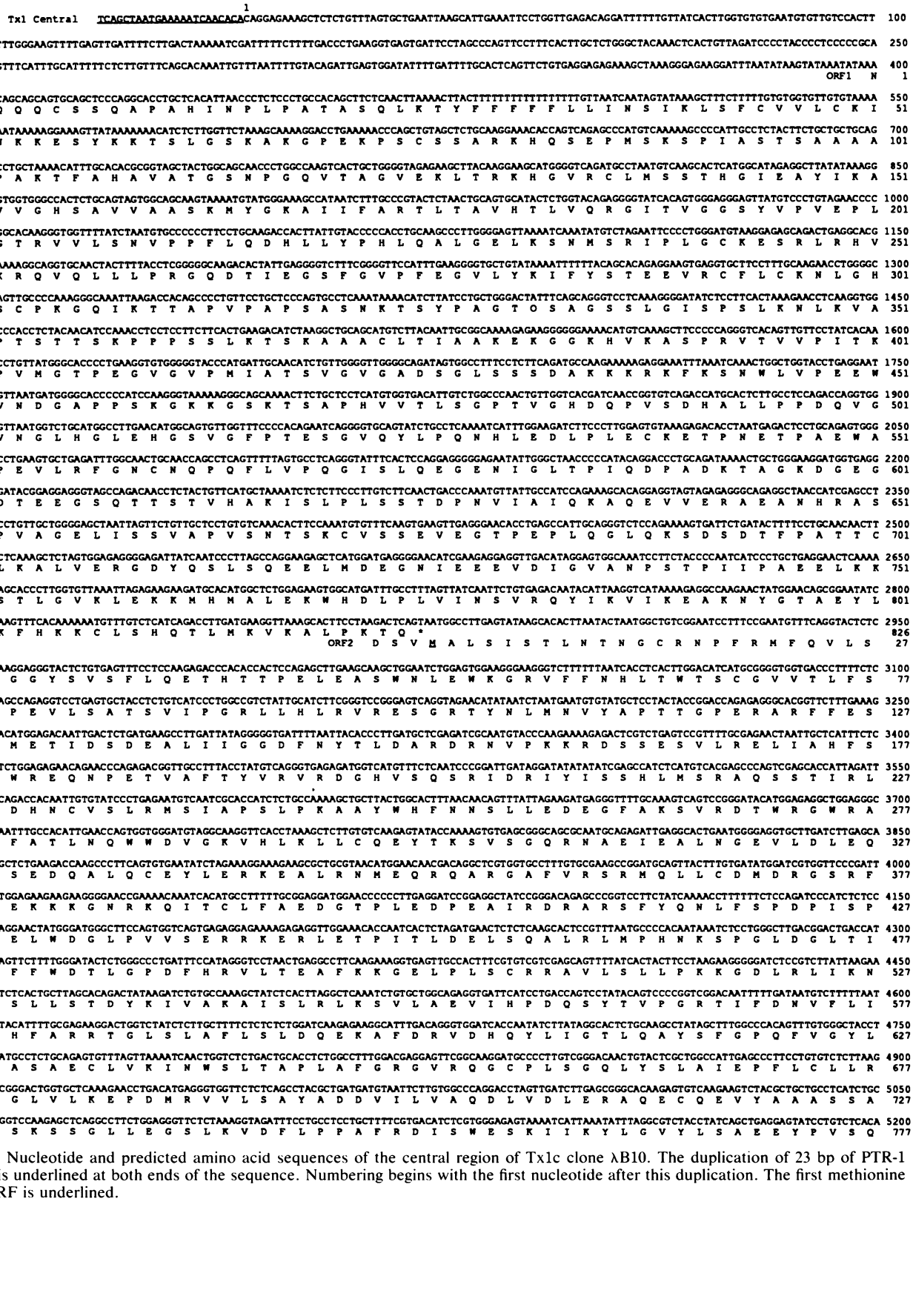

FIG. 3. Nucleotide and predicted amino acid sequences of the central region of Tx1c clone λB10. The duplication of 23 bp of PTR-1 sequence is underlined at both ends of the sequence. Numbering begins with the first nucleotide after this duplication. The first methionine in each ORF is underlined.

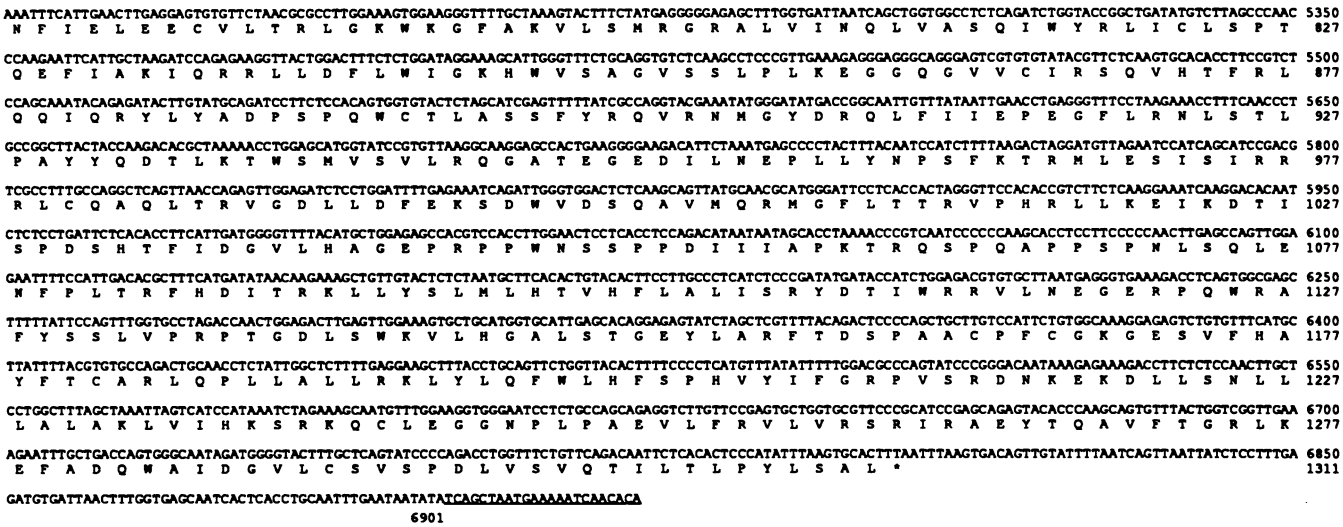

FIG. 3—*Continued*

In general terms, this organization of the Tx1 family is not unusual. Many mobile DNAs occur in autonomous and defective forms, the latter being mobilized in trans by transposase produced by the former. This dichotomy was recognized in the Ac-Ds family by McClintock (37); molecular characterization has shown that many of the dependent Ds elements are related to autonomous Ac by deletions of coding sequences (12, 16). Because the coding region in the Tx families is flanked on both sides by PTR-1 repeats, it is easy to imagine that deletion could occur by homologous recombination between those flanking repeats.

What is surprising about the Tx elements is that they have features of two previously distinct classes of transposable elements. Their termini suggest membership in class I (Table 1), and therefore mobility via a DNA intermediate, whereas their genes are like those of elements in class III, which are supposed to move through RNA intermediates.

We can imagine two general explanations for this paradox. First, the class I-like and class III-like sequences may belong to entirely separate elements. That is, Tx1d and Tx2d may be independent class I elements, whereas the additional sequences in Tx1c and Tx2c are class III elements. This would require that there be four distinct elements in the sequences that we have identified. Each of the two retrotransposons must have an extremely high degree of target specificity for the corresponding class I element. The R1Bm and R2Bm elements of *Bombyx mori* also show a high degree of apparent target specificity (3, 54). Each of these is most commonly found at a specific site in the 28S rRNA gene, although there may also be some dispersed copies. Occupation of specific targets is one way for the damage done by element insertion to be minimized. Repeated sequences, like rDNA and other transposable elements, are plausibly among the most harmless targets in a complex genome. The putative retrotransposons in the Tx elements are flanked by long direct repeats of PTR-1 sequence (23 bp in Tx1c), which could be a target duplication created upon insertion. Large duplications are common for class III elements, although the

	LCF	PTR-1	NT	ORF1	ORF2	NT	PTR-2	RCF
bp	808	394	555	2328	3924	133	394	82
% match	69.8	73.1	81.6	83.6	87.2	84.2	73.4	76.8
aa				775	1308			
% match				73.9 (84.1)	87.1 (94.2)			
sil/sub				0.88	1.98			

FIG. 4. Comparison between Tx1 and Tx2 sequences in various regions. Designations are as given in the legends to Fig. 1 and 3. NT designates two apparently nontranslated regions (i.e., no long ORF) between PTR-1 sequences and the two obvious ORFs. The PTR-1 and PTR-2 comparisons are between consensus sequences derived on the basis of several examples of each type. All other comparisons are based on single examples from each family. In the case of RCF, the region sequenced in each clone was hundreds of base pairs long, but only the terminal 82 bp showed clear homology. It is not known whether the Tx2 RCF sequenced is representative of the family. Values in parentheses for ORFs are matches achieved when conservative amino acid substitutions are allowed. Sil/sub is the ratio of translationally silent nucleotide changes to those that cause an amino acid substitution.

```
Tx2      442      K     N       H              V           H       N  L  T  M
Tx1      442  ERLET PITLDELSQA LRLMPHNKSP GLDGLTIEFF QFFWDTLGPD FHRVLTEAFK KGELPLSCRR
L1Md     472  DHLNS PISPKEIEAV INSLPTKKSP GPDGFSAEFY QTFKEDLIPI LHKLFHKIEV EGTLPNSFYE
I Fact   293  QTIEE NITYLELSSA LQTLK-GCAP GLNRISYGMI KNSSHTTKNR ITKLFNEIFN -SHIPQAYRT
F Ele    434  QHTPI VFRPKEITKI IKNLSPKKSP GYDLITPEMI IQLPHSAVRY ITKLFNAITK LGYFPQRWKM
R2Bm     421  TQLWK PISVEEIKAS RFDWR--TSP GPDGIRSGQW RAVPVHLKA- --EMFNAWMA RGEIPEILRQ
R1Bm     444  GRDAD PLSKDSLRAI IGSLK-NTAP GIDGLTARII KKALPAAEAE FVAVYARCVV EGTFPPVWKD
Ingi     128  ASEFS PITMAELRRS IKLLPSGSAA GPDCLYNEAL QHLGITALNV VLRLFNESLR TGVVPPAWKT
Cons              pit  EL    i  l    sP G Dgl  e                lfn      G  P   k
RTase                                                                          N

Tx2                                                                V                  V
Tx1      AVLSLLPK-KG DL-RLIKNWRP VSLLSTDYKI VAKAISLRLK SVLAEV-IHPD QSYTVPGRTI FDNVFLIRDL
L1Md     ATITLIPKPQK DP-TKIENFRP ISLMNIDAKI LNKILANRIQ EHIKAI-IHPD QVGFIPGMQG WFNIRKSINV
I Fact   SLIIPILKPNT DK-TKTSSYRP ISLNCCIAKI LDKIIAKRLW WLVTYNNLINDKQFGFKKGKST SDCLLYVDYL
F Ele    MKIIMIPKPGK NH-TVASSYRP ISLLSCISKL FEKCLLIRLN QHQTYHNIIPAHQFGFRESHGT IEQVNRITTE
R2Bm     CRTVFVPKVE- RP-GGPGEYRP ISIASIPLRH FHSILARRLL ACCP---PDAR QRGFICADGT LENSAVLDAV
R1Bm     GRLLVLPKGNG RPLTDPKAYRP VTLLPVLGKI LEKVLLQCA- SGLTHS-ISPR QHGFSPGRST ---VTALRTL
Ingi     GVIIPILKAGK KA-EDLDSYRP VTLTSCLCKV MERIIAARPR DTVESQ-LTPQ QSGFRPGCST LEQLLHVRAA
Cons       i  ipK       t    YRP ISL s   Ki   kiLa Rl         i p  Q GF pG  t
RTase    tPvf v K            R  v D R  N                     a

Tx2           KA                            L              S    L  T
Tx1      LHFARRTGLS LA-FLSLDQEK AFDRVDHQYL IGTLQAYSFG PQFVGYLKTM YASAECLVKI NWSLTAPLAF
L1Md     IHYINKLKDK NHMIISLDAEK AFDKIQHPFM IKVLERSGIQ GPYLNMIKAI YSKPVANIKV NGEKLEAIPL
I Fact   ITKSKMHTS- ---LVTLDFSR AFDRVGVHSI IQQLQEWKTG PKIIKYIKNF MSNRKITVRV GPHTSSPLPL
F Ele    IRTAFEYREY CT-AVFLDVSQ AFDKVWLDGL MFKIKIS-LP ESTHKLLKSY LYDRKFAVRC NTATSTVHTI
R2Bm     LGDSRKKLRE CH-VAVLDFAR AFDTVSHEAL VELLRLRGMP EQFCGYIAHL YDTASTTLAV NNEMSSPVKV
R1Bm     LDVSRASEQR YVMAIFLDISG AFDNAWWPMI MVKAKRN-CP PNIYRMLTDY FRGRRIAV-V AGECAEWKVS
Ingi     LCHHTHQYRT G--AVFVDYEK AFDTVDHDKI AREMHRMKVS PHIVKWCVSF LSNRTGRVRF KEKLFRSRTF
Cons     L               LD  k AFD v h  l    l       p      k       r   v v n
RTase          l          iDl   af  i l         aFt                              y

Tx2      K               A                    R
Tx1      GRGVRQGCPL SGQLYSLAIE PFLCLLRKRL TGLVLKEPDM RVVLSAYADD VILVAQDLVD LERAQ-ECQEV
L1Md     KSGTRQGCPL SPYLFNIVLE VLARAIRQQK EIKGIQIGKE EVKISLLADD MIVYISDPKN STRELLNLINS
I Fact   FNGIPQGSTI SVILFLIAFN KLSNIISLHK ---------- EIKFNAYADD FFLIINFNKN TNTNFNLDNLF
F Ele    EAGVPQGSVL GPTLYLIYTA DIPTNS---- ---------- RLTVSTFADD TAILSRSRSP IQATAQLALYL
R2Bm     GRGVRQGDPL SPILFNVVMD LILA--SLPE -RVGYRLEME LVSALAYADD LVLLAGSKV- GMQESISAVDC
R1Bm     TMGCPQGSVL GPTLWNVLMD DLLALPQGIE ---------- GTEMVAYADD VTVLVRGDS- --RAQLERRAH
Ingi     ERGVPQGTVP GSIMFIIVMN SLSQ--RLAE VPLLQHG--- -----FFADD LTLLARHTER DVINHTLQCGL
Cons       GvpQG  l Sp LF I     l                        ayADD   ll
RTase    w  lP G    sP l    l    l                         YmDD ili s

Tx2                        T         P H   S  S         T V
Tx1      -------YAAASSARIN WSKSSGLLEG SLKV-DFLPPA FRDISWESKI IKYLGVYLSA
L1Md     -------FGEVVGYKIN SNKSMAFLYT KNKQAEKEIRE TTPFSIVTNN IKYLGVTLTK
I Fact   --DDIENWCSYSGASLS LSKCQHLHIC RKRHCTCKISC NNFQIPSVTS LKILGITLNN
F Ele    --IDIKKWLSDWRIKVN EQKCKHVTFT LNRQDCPPLLL NSIPLPKADE VTYLGVHLDR
R2Bm     -------VGRQMGLRLN CRKSAVLSMI          (29)        ER WRYLGVDFEA
R1Bm     -----AVLGLAEGWA-S RNKLDFAPAK          (26)        NQ VTVLGVSSTI
Ingi     NVVLQWSKEYFMSVNVA KTKCTLFGCT -ERH-PLTLQL DGERIGADRT PKLLGVTFQC
Cons                 g   n   K                                kyLGV l
RTase        l   l   G       K                                 fLG  i
```

FIG. 5. Comparisons of predicted amino acid sequences of a portion ORF2 from Tx1c, Tx2c, and various class III elements. Shown is a 325-amino-acid region of Tx1c ORF2, which includes but is somewhat larger than that compared among elements by Xiong and Eickbush (53). Tx2c sequences are shown only where they differ from those of Tx1c. The number of amino acids in ORF2 prior to the region compared is given at the beginning of the sequence for each element. Gaps have been introduced to maximize the matches among pairs of sequences. Cons indicates positions for which there is a good consensus among these elements; capital letters are used when at least six of seven sequences agree or at least four of seven agree and the others have conservative substitutions, and lowercase letters are used when at least four of seven agree or three agree and three others have conservative substitutions. The line labeled RTase shows consensus sites shared by a majority of reverse transcriptase genes in genuine retroviruses and class II elements, as identified by Xiong and Eickbush (53). Capital letters indicate essentially perfect conservation; lowercase letters indicate lesser degrees of match. Sources of the sequences: mouse LINE L1Md (35), *Drosophila* I factor (15), *Drosophila* F element (11), silkworm R2Bm (3), silkworm R1Bm (54), and trypanosome Ingi (30).

```
Tx1,2       rCflCknlGHtrqsCpk
I Fact      rCkkClrfGHptpiCks
RlBm        cCnkCqqyGHpekfCra
            tCgrCgedGHrmeaCka
F Ele       qCtnCqeyGHtrsyCtl
Cin4         CynClspdHlafrCs
             CwqClhfGHraraCp

RSV         lCytCgspGHyqaqCpk
            rCqlCngmGHnakqCrk
CaMV        rCwiCnieGHyaneCpn
MoMSV        CtyCeeqGHwakdCp
HIV          CfnCgkeGHiarnC
             CwkCgkeGHqmknC
Copia       kChhCgreGHikkdCfh

Consensus     C..C...GH....C
```

FIG. 6. Comparisons of fingerlike sequences in ORF1 of Tx1c, Tx2c, and several class III elements with each other and with *gag* sequences of selected retroviruses and a class II element. Sources of sequences not given in the legend to Fig. 5: maize Cin4 (46), Rous sarcoma virus, Moloney murine sarcoma virus, human immunodeficiency virus, and copia (40), and cauliflower mosaic virus (15).

consistency of the size and sequence of the duplication is not.

There are some complications in this interpretation, however. The target specificity must have coevolved with the divergence of the Tx1 and Tx2 families, since a Tx1 central region is never found associated with Tx2 ends and vice versa (our unpublished results). Furthermore, there are essentially no central regions of either class that are not associated with Tx ends, so evidence for their independence is lacking. Finally, where are the genes of the class I (version d) Tx elements? Essentially all transposable elements code for at least one product required for some aspect of their mobility. It remains possible that there are undetected minor versions of Tx1d and Tx2d which carry coding sequences.

The second possibility is that Tx1c and Tx2c are the complete, integral elements of the two families. This simplifies the picture to some extent, but it raises the issue of what the transposition mechanism might be. One possibility is that the *gag*- and *pol*-like products of ORF1 and ORF2 mediate transposition through a DNA intermediate, using the short inverted terminal repeats. The similarity of the ORF2 sequence to viral reverse transcriptase may reflect nucleic acid-polymerizing activity and not use of an RNA template per se. Alternatively, the composite element might move via an RNA intermediate in which the inverted terminal repeats identify the discrete ends of the element.

These considerations illuminate some of the deficiencies in our understanding of class III elements in general. Although circumstantial evidence favors an RNA-mediated transposition mechanism, this has not been proven. Clearly, they do not use the same mechanism as do the genuine retroviruslike class II elements to regenerate their ends, and therefore their promoters, upon transposition. An alternative mechanism is suggested by the discovery of an internal RNA polymerase II promoter in a *Drosophila* jockey element (39). In the cases of the mouse (35, 43) and rat (20, 41) LINEs, there are repeated sequences at the element 5′ ends that could serve as essentially external but renewable promoters for RNA polymerase II. The PTR-1 repeats in Tx1c and Tx2c elements occupy a similar position with respect to the class III-like ORFs.

Two other recent observations have reinforced the argument that class III elements move via RNA intermediates. First, mitochondrial plasmids of *Neurospora crassa*, which have *pol*-like genes related to those of LINEs, etc. (53), have been shown to be associated with a reverse transcriptase activity that is specific for plasmid transcripts as templates (32). Second, expression of the *pol*-like product of R2Bm in bacteria indicates that it has an endonuclease activity that is specific for the natural target site of the element in the 28S rRNA gene (55); this is analogous to the integrase function of retroviral *pol* products.

To summarize, the most likely possibilities are as follows: (i) The class I-like (version d) elements and the central class III-like sequences are independent elements; the first move through DNA-mediated events, catalyzed by unknown gene products, whereas the second are retrotransposons with enormous target specificity for PTR-1 sequences and unusually consistent termini and target duplications. (ii) Tx1c represents a single element that moves through DNA-mediated events, catalyzed by the products of ORF1, ORF2, or both. (iii) Tx1c is a single element that moves through an RNA intermediate, supported in part by reverse transcriptase activity of the ORF2 product. In cases ii and iii, movement of Tx1d elements would be stimulated in *trans* by the products of ORF1, ORF2, or both.

The question of the mode of transposition of Tx elements will be resolved only by more direct experimentation. Searches for functionally relevant RNA copies of class III elements have been largely unrewarding. Discrete copies may be obscured by adventitious transcripts, and it is difficult to guess in what tissue(s) they might be produced. We know that Tx1 elements are not transposing at a great rate in frog populations (22), and our initial searches for RNAs corresponding to both external segments of Tx1c and Tx1d and the central region of Tx1c have yielded negative results (33; unpublished results). We are hopeful that injection of appropriately designed Tx1- or Tx2-derived materials into *Xenopus* oocytes will allow transposition events to occur in an experimentally manipulable setting. Oocytes are known to support other types of DNA metabolism, including transcription (26), repair (34), and homologous recombination (6), so there is reason to be optimistic that they will provide a milieu favorable for transposition as well.

ACKNOWLEDGMENTS

We are grateful to Rajesh Sharma for technical assistance and to Zsoka Jonas and Igor Dawid for providing the genomic libraries. We thank members of our laboratory for continuing discussions, Barry Condron for assistance with computing, and Glenn Herrick for thoughtful comments on the manuscript.

This work was supported in part by Public Health Service grant GM22232 from the National Institute of General Medical Sciences.

LITERATURE CITED

1. **Barker, R. F., D. V. Thompson, D. R. Talbot, J. A. Swanson, and J. L. Bennetzen.** 1984. Nucleotide sequence of the maize transposable element Mu1. Nucleic Acids Res. **12**:5955–5967.
2. **Boeke, J. D., D. J. Garfinkel, C. A. Styles, and G. R. Fink.** 1985. Ty elements transpose through an RNA intermediate. Cell **40**:491–500.
3. **Burke, W. D., C. C. Calalang, and T. H. Eickbush.** 1987. The site-specific ribosomal insertion element type II of *Bombyx mori* (R2Bm) contains the coding sequence for a reverse transcriptase-like enzyme. Mol. Cell. Biol. 7:2221–2230.

4. **Carroll, D., J. E. Garrett, and B. S. Lam.** 1984. Isolated clusters of paired tandemly repeated sequences in the *Xenopus laevis* genome. Mol. Cell. Biol. **4**:254–259.
5. **Carroll, D., D. S. Knutzon, and J. E. Garrett.** 1989. Transposable elements in *Xenopus* species, p. 567–574. *In* M. M. Howe and D. E. Berg (ed.), Mobile DNA. American Society for Microbiology, Washington, D.C.
6. **Carroll, D., S. H. Wright, R. K. Wolff, E. Grzesiuk, and E. B. Maryon.** 1986. Efficient homologous recombination of linear DNA substrates after injection into *Xenopus laevis* oocytes. Mol. Cell. Biol. **6**:2053–2061.
7. **Church, G. M., and W. Gilbert.** 1986. Genomic sequencing. Proc. Natl. Acad. Sci. USA **81**:1991–1995.
8. **Covey, S. N.** 1986. Amino acid sequence homology in *gag* region of reverse transcribing elements and the coat protein of cauliflower mosaic virus. Nucleic Acids Res. **14**:623–633.
9. **D'Ambrosio, E., S. D. Waitzkin, F. R. Witney, A. Salemme, and A. V. Furano.** 1986. Structure of the highly repeated, long interspersed DNA family (LINE or L1Rn) of the rat. Mol. Cell. Biol. **6**:411–424.
10. **Dente, L., M. Sollazzo, C. Baldari, G. Cesareni, and R. Cortese.** 1985. The pEMBL family of single-stranded vectors, p. 101–107. In D. M. Glover (ed.), DNA cloning, vol. 1. A practical approach. IRL Press, Washington, D.C.
11. **Di Nocera, P. P., and G. Casari.** 1987. Related polypeptides are encoded by *Drosophila* F elements, I factors and mammalian L1 sequences. Proc. Natl. Acad. Sci. USA **84**:5843–5847.
12. **Doring, H.-P., and P. Starlinger.** 1984. Barbara McClintock's controlling elements: now at the DNA level. Cell **39**:253–259.
13. **Doring, H.-P., and P. Starlinger.** 1986. Molecular genetics of transposable elements in plants. Annu. Rev. Genet. **20**:175–200.
14. **Emmons, S. W., and L. Yesner.** 1984. High-frequency excision of transposable element Tc1 in the nematode *Caenorhabditis elegans* is limited to somatic cells. Cell **36**:599–605.
15. **Fawcett, D. H., C. K. Lister, E. Kellett, and D. J. Finnegan.** 1986. Transposable elements controlling I-R hybrid dysgenesis in *D. melanogaster* are similar to mammalian LINEs. Cell **47**:1007–1015.
16. **Fedoroff, N. V.** 1989. About maize transposable elements and development. Cell **56**:181–191.
17. **Feinberg, A. P., and B. Vogelstein.** 1983. A technique for radiolabeling DNA restriction fragments to high specific activity. Anal. Biochem. **132**:6–13.
18. **Fink, G. R., J. D. Boeke, and D. J. Garfinkel.** 1986. The mechanism and consequences of retrotransposition. Trends Genet. **2**:118–122.
19. **Finnegan, D. J.** 1985. Transposable elements in eukaryotes. Int. Rev. Cytol. **93**:281–326.
20. **Furano, A. V., S. M. Robb, and F. T. Robb.** 1988. The structure of the regulatory region of the rat L1 (L1Rn, long interspersed repeated) DNA family of transposable elements. Nucleic Acids Res. **16**:9215–9231.
21. **Garfinkel, D. J., J. D. Boeke, and G. R. Fink.** 1985. Ty element transposition: reverse transcriptase and virus-like particles. Cell **42**:507–517.
22. **Garrett, J. E., and D. Carroll.** 1986. Tx1: a transposable element from *Xenopus laevis* with some unusual properties. Mol. Cell. Biol. **6**:933–941.
23. **Georgiev, G. P.** 1984. Mobile genetic elements in animal cells and their biological significance. Eur. J. Biochem. **145**:203–220.
24. **Gough, J. A., and N. E. Murray.** 1983. Sequence diversity among related genes for recognition of specific targets in DNA molecules. J. Mol. Biol. **166**:1–19.
25. **Greenblatt, I. M.** 1974. Movement of *Modulator* in maize: a test of an hypothesis. Genetics **77**:671–678.
26. **Gurdon, J. B., and M. P. Wickens.** 1983. The use of *Xenopus* oocytes for the expression of cloned genes. Methods Enzymol. **101**:370–386.
27. **Hattori, M., S. Kuhara, O. Takenaka, and Y. Sakaki.** 1986. L1 family of repetitive DNA sequences in primates may be derived from a sequence encoding a reverse transcriptase-related protein. Nature (London) **321**:625–628.
28. **Henikoff, S.** 1984. One-way exonucleolytic digestion to target deletion end-points for DNA sequencing. Gene **28**:331–359.
29. **Howe, M. M., and D. E. Berg (ed.).** 1989. Mobile DNA. American Society for Microbiology, Washington, D.C.
30. **Kimmel, B. E., O. K. ole-Moiyoi, and J. R. Young.** 1987. Ingi, a 5.2-kb dispersed sequence element from *Trypanosoma brucei* that carries half of a smaller mobile element at either end and has homology with mammalian LINEs. Mol. Cell. Biol. **7**:1465–1475.
31. **Kingsman, A. J., K. F. Chater, and S. M. Kingsman (ed.).** 1988. Transposition. Cambridge University Press, Cambridge.
32. **Kuiper, M. T. R., and A. M. Lambowitz.** 1988. A novel reverse transcriptase activity associated with mitochondrial plasmids of *Neurospora*. Cell **55**:693–704.
33. **Lam, B. S., and D. Carroll.** 1983. Tandemly repeated DNA sequences from *Xenopus laevis*. II. Dispersed clusters of a 388 bp repeating unit. J. Mol. Biol. **165**:587–597.
34. **Legerski, R. J., J. E. Penkala, C. A. Peterson, and D. A. Wright.** 1987. Repair of UV-induced lesions in *Xenopus laevis* oocytes. Mol. Cell. Biol. **7**:4317–4323.
35. **Loeb, D. D., R. W. Padgett, S. C. Hardies, W. R. Shehee, M. B. Comer, M. H. Edgell, and C. A. Hutchison III.** 1986. The sequence of a large L1Md element reveals a tandemly repeated 5′ end and several features found in retrotransposons. Mol. Cell. Biol. **6**:168–182.
36. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
37. **McClintock, B.** 1984. The significance of responses of the genome to challenge. Science **226**:792–801.
38. **Mellor, J., M. Malim, K. Gull, M. F. Tuite, S. McCready, T. Dibbayawan, S. M. Kingsman, and A. J. Kingsman.** 1985. Reverse transcriptase activity and Ty RNA are associated with virus-like particles in yeast. Nature (London) **318**:583–586.
39. **Mizrokhi, L. J., S. G. Georgieva, and Y. V. Ilyin.** 1988. Jockey, a mobile *Drosophila* element similar to mammalian LINEs, is transcribed from the internal promoter by RNA polymerase II. Cell **54**:685–691.
40. **Mount, S. M., and G. M. Rubin.** 1985. Complete nucleotide sequence of the *Drosophila* transposable element copia: homology between copia and retroviral proteins. Mol. Cell. Biol. **5**:1630–1638.
41. **Nur, I., E. Pascale, and A. V. Furano.** 1988. The left end of rat L1 (L1Rn, long interspersed repeated) DNA which is a CpG island can function as a promoter. Nucleic Acids Res. **16**:9233–9251.
42. **O'Hare, K., and G. M. Rubin.** 1983. Structure of P transposable elements and their sites of insertion and excision in the *Drosophila melanogaster* genome. Cell **34**:25–35.
43. **Padgett, R. W., C. A. Hutchison III, and M. H. Edgell.** 1988. The F-type 5′ motif of mouse L1 elements: a major class of L1 termini similar to the A-type in organization but unrelated in sequence. Nucleic Acids Res. **16**:739–749.
44. **Russel, M., S. Kidd, and M. R. Kelley.** 1986. An improved filamentous helper phage for generating single-stranded plasmid DNA. Gene **45**:333–338.
45. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
46. **Schwarz-Sommer, Z., L. Leclercq, E. Gobel, and H. Saedler.** 1987. Cin4, an insert altering the structure of the A1 gene in *Zea mays*, exhibits properties of nonviral retrotransposons. EMBO J. **6**:3873–3880.
47. **Shapiro, J. A. (ed.).** 1983. Mobile genetic elements. Academic Press, Inc., New York.
48. **Shiba, T., and K. Saigo.** 1983. Retrovirus-like particles containing RNA homologous to the transposable element copia in *Drosophila melanogaster*. Nature (London) **302**:119–124.
49. **Singer, M. F., and J. Skowronski.** 1985. Making sense out of LINEs: long interspersed repeat sequences in mammalian genomes. Trends Biochem. Sci. **10**:119–122.
50. **Truett, M. A., R. S. Jones, and S. S. Potter.** 1981. Unusual structure of the FB family of transposable elements in *Drosophila*. Cell **24**:753–763.

51. **Voliva, C. F., C. L. Jahn, M. B. Comer, C. A. Hutchison III, and M. H. Edgell.** 1983. The L1Md long interspersed repeat family in the mouse: almost all examples are truncated at one end. Nucleic Acids Res. **11**:8847–8859.
52. **Wahli, W., and I. B. Dawid.** 1980. Isolation of two closely related vitellogenin genes, including their flanking regions, from a *Xenopus laevis* gene library. Proc. Natl. Acad. Sci. USA **77**:1437–1441.
53. **Xiong, Y., and T. H. Eickbush.** 1988. Similarity of reverse transcriptase-like sequences of viruses, transposable elements, and mitochondrial introns. Mol. Biol. Evol. **5**:675–690.
54. **Xiong, Y., and T. H. Eickbush.** 1988. The site-specific ribosomal DNA insertion element R1Bm belongs to a class of non-long-terminal-repeat retrotransposons. Mol. Cell. Biol. **8**:114–123.
55. **Xiong, Y., and T. H. Eickbush.** 1988. Functional expression of a sequence-specific endonuclease encoded by the retrotransposon R2Bm. Cell **55**:235–246.