MOLECULAR AND CELLULAR BIOLOGY, July 1994, p. 4485–4492
0270-7306/94/$04.00+0

Vol. 14, No. 7

An In Vivo Assay for the Reverse Transcriptase of Human Retrotransposon L1 in *Saccharomyces cerevisiae*

BETH A. DOMBROSKI,[1] QINGHUA FENG,[2] STEPHEN L. MATHIAS,[1] DONNA M. SASSAMAN,[1] ALAN F. SCOTT,[1] HAIG H. KAZAZIAN, JR.,[1] AND JEF D. BOEKE[2]*

Departments of Pediatrics and Medicine,[1] Center for Medical Genetics, and Department of Molecular Biology and Genetics,[2] Johns Hopkins University School of Medicine, Baltimore, Maryland 21205

Received 22 November 1993/Returned for modification 14 January 1994/Accepted 4 April 1994

L1 elements constitute a highly repetitive human DNA family (50,000 to 100,000 copies) lacking long terminal repeats and ending in a poly(A) tail. Some L1 elements are capable of retrotransposition in the human genome (Kazazian, H. H., Jr., C. Wong, H. Youssoufian, A. F. Scott, D. G. Phillips, and S. E. Antonarakis, Nature (London) 332:164–166, 1988). Although most are 5′ truncated, a consensus sequence of complete L1 elements is 6 kb long and contains two open reading frames (ORFs) (Scott, A. F., B. J. Schmeckpeper, M. Abdelrazik, C. T. Comey, B. O'Hara, J. P. Rossiter, T. Cooley, P. Heath, K. D. Smith, and L. Margolet, Genomics 1:113–125, 1987). The protein encoded by ORF2 has reverse transcriptase (RT) activity in vitro (Mathias, S. L., A. F. Scott, H. H. Kazazian, Jr., J. D. Boeke, and A. Gabriel, Science 254:1808–1810, 1991). Because L1 elements are so numerous, efficient methods for identifying active copies are required. We have developed a simple in vivo assay for the activity of L1 RT based on the system developed by Derr et al. (Derr, L. K., J. N. Strathern, and D. J. Garfinkel, Cell 67:355–364, 1991) for yeast *HIS3* pseudogene formation. L1 ORF2 displays an in vivo RT activity similar to that of yeast Ty1 RT in this system and generates pseudogenes with unusual structures. Like the *HIS3* pseudogenes whose formation depends on Ty1 RT, the *HIS3* pseudogenes generated by L1 RT are joined to Ty1 sequences and often are part of complex arrays of Ty1 elements, multiple *HIS3* pseudogenes, and hybrid Ty1/L1 elements. These pseudogenes differ from those previously described in that there are extra base pairs of unknown origin inserted at several of the junctions. In two of three *HIS3* pseudogenes studied, the L1 RT appears to have jumped from the 5′ end of a Ty1/L1 transcript to the poly(A) tract of the *HIS3* RNA.

Retrotransposons are of two basic types, those flanked by long terminal repeat (LTR) sequences, typified by the yeast Ty1 element and those that lack these structures and usually have a 3′ poly(A), oligo(A), or similar structure, typified by the human L1 sequence (2). There is now considerable evidence that both of these elements transpose via reverse transcription; the mechanisms used by the LTR elements are relatively well understood, being similar overall to those of retroviruses, while the mechanisms used by poly(A) elements are less well known.

Previously, a genomic human L1 element, L1.2B, located on chromosome 22, was shown to be the likely precursor of a de novo insertion into the factor VIII gene in a patient with hemophilia A (15, 20). Since the new insertion ended in a poly(A) tail, it was presumed that the transposition process involved an RNA intermediate. That is, the element located on chromosome 22 was transcribed, reverse transcribed, and reintegrated into the genome on the X chromosome. An allele, L1.2A, of the active L1 element was assayed for reverse transcriptase (RT) activity in an expression system in the yeast *Saccharomyces cerevisiae* by taking advantage of an endogenous retrotransposon Ty1. Ty1, which is present at approximately 30 copies in the yeast genome, contains LTRs and has two overlapping open reading frames (ORFs) (7). The first ORF, *TYA*, encodes a Gag-like protein, while *TYB* encodes proteins with protease, integrase, RT, and RNase H activities. When expressed, Ty1 proteins and RNA are coassembled into virus-like particles (VLPs). A construct (pSM2) in which ORF2 of L1.2A replaced most of *TYB* (23) was designed; the expression of this plasmid is regulated by the *GAL1* promoter. After induction with galactose, pSM2-derived VLPs were partially purified and assayed for RT activity. This assay showed that L1 ORF2 encoded an RT activity capable of elongating a DNA strand from homopolymer templates and primers. We have now assayed this RT in vivo on more biologically meaningful templates and established that L1 RT is active in yeast cells.

Derr et al. (14) developed an elegant pseudogene formation system in which an antisense transcript of the *HIS3* gene (produced by an artificial *GAL1-mhis3AI* gene), disrupted by an artificial intron in the sense orientation, served as a reporter gene for rare reverse transcription and integration events in *S. cerevisiae*. The spliced transcript can in principle be used as a substrate for reverse transcription. Upon reverse transcription of the *mhis3* RNA and subsequent integration of the *HIS3* cDNA into host DNA, a His^+ phenotype results. We refer to these stably integrated cDNA copies as *HIS3* pseudogenes. We have shown that the human L1 RT is able to reverse transcribe these sequences and that they become integrated into the genome. These results provide the basis for a general assay for L1 ORF2-encoded RT activity. This assay can be used to functionally define L1 RT through mutational analysis as well as allowing further active L1 element copies in the human genome to be identified.

MATERIALS AND METHODS

Plasmid constructions. The plasmids in Fig. 1B were constructed as described below.

(i) pSM40. The *GAL1*-Ty1 *TRP1* plasmid pSM40 was con-

* Corresponding author. Mailing address: Department of Molecular Biology and Genetics, Johns Hopkins University School of Medicine, 725 N. Wolfe St., Baltimore, MD 21205. Phone: (410) 955-2481. Fax: (410) 955-0831. Electronic mail address: jef_boeke@qmail.bs.jhu.edu.

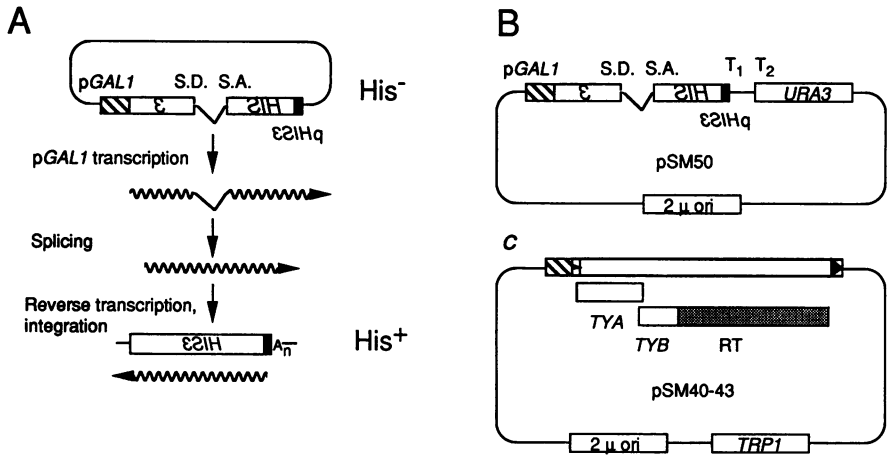

FIG. 1. The *mhis3AI* system for detecting pseudogene formation in *S. cerevisiae*. (A) Outline of the *mhis3AI* system for detecting rare reverse transcription/integration events in yeast cells (14). A plasmid expressing an antisense *his3* (*mhis3*) transcript from the *GAL1* promoter contains an intron inserted in the sense orientation. When this plasmid is introduced into a *his3* strain of yeast, the strain remains His^-. The resultant transcript can be spliced and then reverse transcribed; integration of the product into either the chromosome or the donor plasmid will give rise to His^+ progeny. Hatched box, *GAL1* promoter; open box, *HIS3* coding region (in inverted orientation); p, promoter; black box, *HIS3* promoter; V, intron sequence; S.D., splice donor; S.A., splice acceptor; wavy lines, RNA; A_n, poly(A) tract. (B) Structure of *mhis3AI* plasmid pSM50. The yeast *URA3* gene is the selectable marker; the plasmid contains an origin of replication derived from the yeast 2 μm plasmid. The site of 3′ end formation of the *mhis3* mRNA predicted from earlier work is indicated by T_1; the site of 3′ end formation observed in this study is indicated by T_2. (C) Structure of the series of plasmids (pSM40 to pSM43) used to express the RTs in this study. The yeast *TRP1* gene is used as the selectable marker. Hatched box, *GAL1* promoter; boxed triangles, Ty1 LTRs; shaded box, RT segment. RT fragments are as follows: pSM40, Ty1; pSM41, none (RT deletion); pSM42, L1 ORF2 wild type; pSM43, L1 ORF2 D→Y mutant.

structed by replacing the 4,248-bp *Bst*EII-*Bam*HI fragment of pJEF1114 (25) with the corresponding *Bst*EII-*Bam*HI fragment from the plasmid pJEF1267 (19). This was done to create a *TRP1* plasmid with a unique *Sac*I site between the 3′ end of *TYB* and the 3′ LTR.

(ii) pSM41. The p*GAL1*-TyΔ*Sal-Sac* plasmid pSM41 was constructed by digesting pSM40 with *Sal*I (position 2688) and *Sac*I (position 6079), blunting the overhanging ends with DNA polymerase I Klenow large fragment (PolIk), and recircularizing the 8,580-bp blunt-end fragment.

(iii) pSM42. The p*GAL1*-Ty1/L1 plasmid pSM42 was constructed by replacing the 3,391-bp *Sal*I-*Sac*I fragment of pSM40 with the 4,079-bp *Sal*I-*Sac*I fragment containing L1 ORF2 from the plasmid pORF2S/S (23).

(iv) pSM43. The p*GAL1*-Ty1/L1-D→Y plasmid pSM43 was constructed by replacing the 3,391-bp *Sal*I-*Sac*I fragment of pSM40 with the 4,079-bp *Sal*I-*Sac*I fragment from the plasmid pSM8 (23).

(v) pSM50. The 1,025-bp *Cla*I fragment from pGTy1-H3-m*his3AI* (13) containing the m*his3AI* gene was made blunt ended with PolIk. This fragment was cloned into plasmid pCGE329 (5) at the unique *Xho*I site, after the *Xho*I ends were rendered blunt with PolIk, such that the artificial intron is in the same orientation (5′→3′) as transcription initiated from the *GAL1* promoter and the *HIS3* gene itself is in the opposite transcriptional orientation of the *GAL1* promoter.

Selection and quantitation of His^+ prototrophs. Yeast media were prepared as previously described (6, 30). Glucose (Glu) and galactose (Gal) were added to 2% (wt/vol). Plates were incubated at 30°C overnight unless otherwise indicated. Large (2.5-cm-diameter) patches were grown up on synthetic complete (SC) minus Ura minus Trp plus Glu plates for 2 days. Each patch was derived from an independent colony previously grown on a master SC minus Ura minus Trp plus Glu plate. Three colonies were tested for each strain. Each set of patches was then replica plated to two SC minus Ura minus Trp plus Gal plates, one incubated at 30°C and one incubated at 22°C for 4 days. Each Gal plate was then replica plated to SC minus Ura minus Trp plus Glu. This ensures that the population of survivors from galactose consists only of cells that maintained both plasmids throughout the exposure to galactose (5). Each set of these plates was then replica plated to yeast extract-peptone-dextrose (YPD) and grown at 30°C for 24 h. At this point, the cells were scraped into 10 ml of H_2O. Dilutions of the cells were plated onto SC minus His, SC minus His minus 5-fluoroorotic acid (FOA), and YPD plates. His^+ colonies were counted after incubation at 30°C for 2 days.

PCR primers. Oligonucleotides SM21′ (ATTCTCTTCG GTGGAGCGGG) and SM22 (CTAGGGCTTTCTGCTCT GTC) are derived from the transcribed segment of pSM50 (the sequence of pSM50 is available on request). The Ty1 primers and L1 primers (18 to 20 nucleotides long) are derived from the reference sequences indicated (see Fig. 3); the number refers to the first base of the oligonucleotide, and a prime sign indicates a noncoding strand oligonucleotide.

Genomic DNA blotting and inverted PCR. DNA was prepared, blotted, and hybridized as previously described (5). The *HIS3* hybridization probe used was an internal 474-bp *Msc*I-*Pst*I fragment except as indicated otherwise.

Contour-clamped homogeneous electric field (CHEF) pulsed-field gels (11) were run as previously described, with the following parameters: for whole chromosomes, 10 h at 200 V with a 60-s pulse time, followed by 8 h at 200 V with a 90-s pulse time; for the *Mlu*I restriction digests, 17 h at 200 V with a 20-s pulse time. DNA preparation and blotting procedures for chromosome-size DNA were as described previously (4).

For inverted PCR, a modification of the method of Ochman et al. (26) was used. DNA (2 μg) from the yeast His^+ prototrophs was digested with *Hae*II, diluted to a concentration of 2 μg/ml (to promote circularization of the restriction fragments), and self-ligated. The products were then subjected to amplification by PCR with a lower-strand *HIS3* primer

located at the left junction of the insertion adjacent to the *GAL1* promoter (SM21′) and an upper-strand primer located at the right junction of the insertion adjacent to the *HIS3* promoter (SM22). The PCR products were gel purified, electroeluted, and sequenced. All DNA sequencing was performed on an automated sequencer (Applied Biosystems 373) by using cycle sequencing with *Taq* polymerase and dye-labeled dideoxynucleotide terminators.

PCR amplification, cloning, and sequencing of yeast insertions. Conditions for all PCR mixtures were as follows: 100 ng of yeast genomic DNA was amplified in PCR buffer (10 mM Tris-HCl, 50 mM KCl, 1.5 mM $MgCl_2$, 0.01% gelatin, pH 8.3) in the presence of 0.8 mM deoxynucleotide triphosphates. All primers were used at a final concentration of 0.3 μM.

(i) Y5.1. Yeast genomic DNA was amplified with primers SM22 and Ty1′-1760 (30 cycles of 94°C for 30 s, 54°C for 1 min, and 72°C for 2 min 30 s). This 3.0-kb product was cloned into the vector pCR with the TA cloning system (Invitrogen) and sequenced (on both strands).

(ii) Y5.4. Yeast genomic DNA was amplified with SM22 and Ty1′-1581 (30 cycles of 94°C for 1 min, 57°C for 30 s, and 72°C for 2 min), giving a 1.5-kb product, which was cloned and sequenced as described above.

(iii) Y5.5. Yeast genomic DNA was amplified from SM22 and Ty1′-540 (35 cycles of 94°C for 30 s, 54°C for 30 s, and 72°C for 1 min). The 1.7-kb product was cloned into the pCR vector. Also, a number of other yeast genomic PCR products were electroeluted from a 2% NuSieve agarose (FMC) or 5% acrylamide gel and sequenced directly. These samples included a 1.3-kb product, using L1-5461 and Ty1′-540 (35 cycles of 94°C for 30 s, 54°C for 30 s, and 72°C for 1 min); a 2.7-kb product, using L1-6046 and L1′-2347 (30 cycles of 94°C for 30 s, 48°C for 30 s, and 72°C for 1 min 15 s); and a 1.2-kb product, using L1-5647 and SM21′ (30 cycles of 94°C for 30 s, 55°C for 30 s, and 72°C for 1 min).

RESULTS

Assay system. The readout of the pseudogene formation assay (14) is the frequency of His^+ prototrophs from yeast cells bearing the appropriate plasmid(s) (Fig. 1). Previous work (14) showed that the formation of His^+ prototrophs in yeast strains expressing *GAL1-mhis3AI* was dependent on expression of the endogenous yeast retrotransposon Ty1. This was demonstrated in two ways. (i) The frequency of His^+ prototrophs was reduced to an undetectable level in *spt3* mutant strains. In *spt3* mutants, Ty1 transcription (35) and Ty1 transposition (8) are reduced by at least 10- to 20-fold relative to that in the wild type. (ii) The frequency of His^+ prototrophs was stimulated enormously by overexpression of Ty1-encoded proteins from a *GAL1*/Ty1 plasmid.

We have used an *spt3* mutant of *S. cerevisiae* containing both a *GAL1-mhis3AI* plasmid (bearing a *URA3* selection marker) and a *GAL1*/Ty1-derived plasmid as the source of Ty1 RT, L1 RT, or a mutant RT. In these RT plasmids either a native Ty1 element or a chimeric Ty1 element bearing intact human L1.2A ORF2 and lacking Ty1 RT (23) is overexpressed by using the yeast *GAL1* promoter. In this way, we determined that L1 RT, like Ty1 RT, can mediate *HIS3* pseudogene formation.

Ty1 and L1 RTs can generate His^+ prototrophs. Both Ty1 and Ty1/L1 constructs were able to generate His^+ prototrophs at high frequencies at 22°C, with the Ty1/L1 construct giving rise to a higher frequency of total His^+ prototrophs; control constructs lacking RT gave rise to no such prototrophs (Table 1). A segregation analysis of the His^+ prototrophs was carried out by growth on nonselective medium and subsequent replica plating to SC minus His. In accordance with previous work (14), our Ty1 construct (SMY1) gave rise to approximately 80% (19 of 24) His^+ prototrophs in which the His^+ phenotype was unstable and cosegregated with the Ura^+ phenotype, as expected if the *HIS3* gene had recombined into the donor *his3AI* plasmid. The Ty1/L1 construct (SMY5) gave rise to 90% (22 of 24) His^+ prototrophs of this class, and the class of stable chromosomal His^+ prototrophs (*HIS3* pseudogenes) was correspondingly less abundant. The *HIS3* pseudogene class was characterized by a Ura^- phenotype (or unstable Ura^+ phenotype) together with the originally selected His^+ phenotype. In order to select the *HIS3* pseudogene class directly, we selected His^+ prototrophs in the presence of FOA, which selects against Ura^+ cells (6). Under these conditions, the Ty1/L1 construct still gave a very high frequency of His^+ prototrophs. Six randomly chosen His^+ FOA^r prototrophs derived from the Ty1/L1 construct pSM42, designated Y5.1 to Y5.6, were colony purified. The His^+ phenotype in these strains was unlinked to the Ura^+ and Trp^+ plasmid markers and was stable in the absence of selection, indicating a chromosomally integrated state.

TABLE 1. His^+ prototrophy

Strain[a]	Frequency (10^{-7})			
	22°C		30°C	
	His^+	His^+ FOA^r	His^+	His^+ FOA^r
SMY1-pSM40 (Ty1)	29.4	6.44	0.135	<0.153
SMY3-pSM41 (Ty1Δ)	<0.117	<0.117	<0.128	<0.128
SMY5-pSM42 (L1)	244	1.99	31.6	1.77
SMY7-pSM43 (L1-FXDY)	<0.136	<0.136	0.337[b]	<0.133

[a] Strain name is followed by RT plasmid name and type of RT (in parentheses). The host strain used for all these transformants was YH50 (*MATα his3Δ200 ura3-167 trp1Δ1 leu2Δ1 spt3-101*). All four strains also contain plasmid pSM50 (Fig. 1B).

[b] The finite value of this frequency represents a single His^+ colony; thus, we suspect that this was not a true RT-mediated His^+ prototroph. In a second quantitative experiment with SMY7, no His^+ prototrophs were obtained from 10^8 cells plated at 22 or 30°C, whereas frequencies similar to those reported above were obtained with SMY5. In several independent plating experiments, we have not obtained any additional His^+ prototrophs from SMY7.

In one control construct, pSM41, the L1 RT is deleted; in a second control construct, pSM43, the Ty1/L1 element encodes a FADD→FADY substitution in the most highly conserved amino acid sequence motif (Y/FXDD [36]) of RTs. The mutant elements are reduced at least 100-fold in their abilities to produce His^+ prototrophs (Table 1). The pSM43 mutant expresses normal amounts of RT protein but lacks RT activity in vitro (23).

Genomic blot analysis of *HIS3* pseudogenes. The genomic DNAs of chromosomal His^+ prototrophs Y5.1 to 5.6 were digested with a variety of restriction enzymes, including *Ase*I, *Spe*I, and *Ssp*I, which lack sites in the *HIS3* or flanking transcribed plasmid sequences but have multiple sites in Ty1, and ORF2 of L1.2A and have relatively common sites in yeast DNA. When these digests were probed with *HIS3*, we determined that certain of the strains, such as Y5.4, indeed contained a single *HIS3* band as expected, but surprisingly, several of the prototrophs contained multiple bands, indicating integration of more than one *HIS3* pseudogene copy (Fig. 2A). Consistent with the notion that these are concerted events, strains with multiple *HIS3* bands show hybridization of a *HIS3* probe to a single band on pulsed-field gels of uncut DNA (Fig. 3). Furthermore, hybridization to a single band was observed

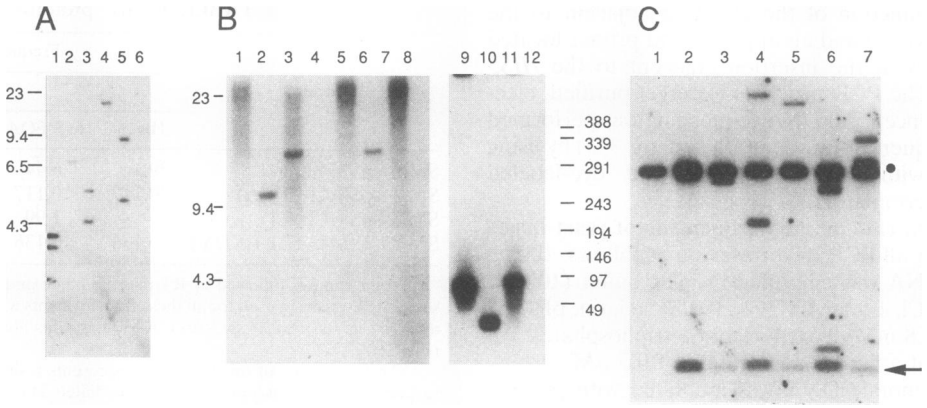

FIG. 2. Genomic DNA blot analyses of His$^+$ strains. (A) There are multiple *HIS3* copies in Y5.1 and Y5.4. Genomic DNA was prepared from His$^+$ prototrophs derived from SMY5 (Table 1) and digested with *Ase*I, which does not cut the *HIS3* pseudogene. The blot was probed with *HIS3*. Samples (left to right) are Y5.1, Y5.2, Y5.3, Y5.4, Y5.5, and SMY5 (parent). M_r markers are in kilobases. (B) The multiple copies of *HIS3* are present on single restriction fragments. DNAs were digested with *Mlu*I (lanes 5 to 12) or *Pvu*I plus *Stu*I (lanes 1 to 4). Lanes 1 to 8 were run on a conventional agarose gel; lanes 9 to 12 were from a CHEF gel. DNAs were from strains SMY5 (lanes 4, 8, and 12), Y5.1 (lanes 3, 7, and 11), Y5.4 (lanes 2, 6, and 10), and Y5.5 (lanes 1, 5, and 9). M_r markers are in kilobases. (C) The intron is precisely removed from the *HIS3* pseudogenes. DNAs were digested with *Pst*I and *Msc*I. The 476-bp band expected from precise intron removal is indicated by an arrow; the band marked with the dot is the *his3Δ200* allele in the chromosome, which cross-hybridizes to the *HIS3* probe used for the blot in this panel only. Strains are SMY5 and Y5.1 to Y5.6 (lanes 1 to 7, respectively).

for each strain when certain enzymes were used (Fig. 2B). Therefore, these multiple *HIS3* gene copies are probably present as tightly linked arrays in the yeast genome, similar to those previously described by Weinstock et al. (34).

A double digest with *Pst*I and *Msc*I was performed to determine whether the *HIS3* pseudogenes contained the restriction fragment diagnostic of proper splicing-mediated intron removal (14); in all six strains tested, the expected *HIS3* DNA fragment was observed (Fig. 2C).

Molecular characterization of inserted DNA. To determine the sequences flanking the *HIS3* pseudogenes in strains Y5.1 to Y5.6, inverted PCR was performed on their DNAs (see Materials and Methods). The yeast DNA was digested with *Hae*II, and the digests were diluted to promote circularization upon ligation. The ligated products were amplified by PCR with primers within *HIS3* oriented toward both ends of the gene. Upon amplification, the product of this reaction consists of sequence flanking *HIS3* to the nearest *Hae*II sites flanking both ends. The left junction of the insertion refers to the sequences originally adjacent to the *GAL1* promoter, while the right junction refers to sequences adjacent to the *HIS3* promoter. Junction products were obtained from three strains: Y5.1, Y5.4, and Y5.5.

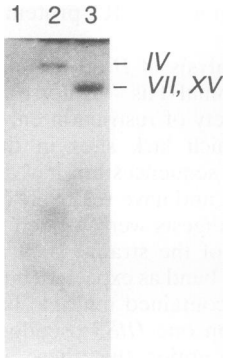

FIG. 3. CHEF gel analysis of yeast chromosomes. The identity of the chromosome(s) in the indicated bands was deduced from the ethidium bromide-stained gel (9). Lanes: 1, SMY5; 2, Y5.1; 3, Y5.5.

When these products were sequenced, we discovered that in all three cases, the right ends of the *HIS3* pseudogenes were not connected to poly(A) sequence, as expected from earlier work (14). Rather, flanking plasmid sequences derived from pSM50 were found. This suggested that the Ty1/L1-mediated events had utilized RNA templates terminated at a site (T_2 [Fig. 1]) different from those (T_1) mediated by Ty1 RT. Thus, conventional PCR was used to define the end of the *HIS3*/plasmid DNA segment in the insertions.

The left junction sequence proceeds outward through *HIS3* sequence precisely to the transcription start site of the *GAL1-mhis3AI* cassette and is joined to various positions in Ty1 DNA. Although Derr et al. reported *HIS3* pseudogenes embedded in Ty1 sequence (14), we did not anticipate that retrotransposition directed by L1 RT would give similar results. The presence of Ty1 sequence at the left junction suggested that the *HIS3* pseudogene might be flanked by Ty1 sequences on both ends. This correct assumption led to the amplification of right junctions by conventional PCR. Indeed, all of the *HIS3* pseudogenes we obtained had Ty1 sequences flanking the *HIS3* sequences on both sides (Fig. 3).

Right junction sequences. Sequence from the PCR products revealed that the *HIS3* pseudogenes were not flanked directly by poly(A) and Ty1 sequence at their right junctions. To identify the right junction sequence, a PCR primer was chosen from Ty1 sequence approximately 60 bp downstream of the corresponding left junction sequence. This primer was then used with the SM22 primer to amplify the right junction of the *HIS3* insertion. The resulting PCR products were cloned and sequenced. The Y5.1 product was 3.0 kb long, whereas the Y5.5 product was 1.7 kb long. The Y5.4 strain did not yield any products with such primers; therefore, a number of Ty1 primers from both strands were used with SM22 for PCR. One such combination gave a product.

In all cases, the insertion did not terminate within the T_1 sequences but continued into the pSM50 vector sequence for

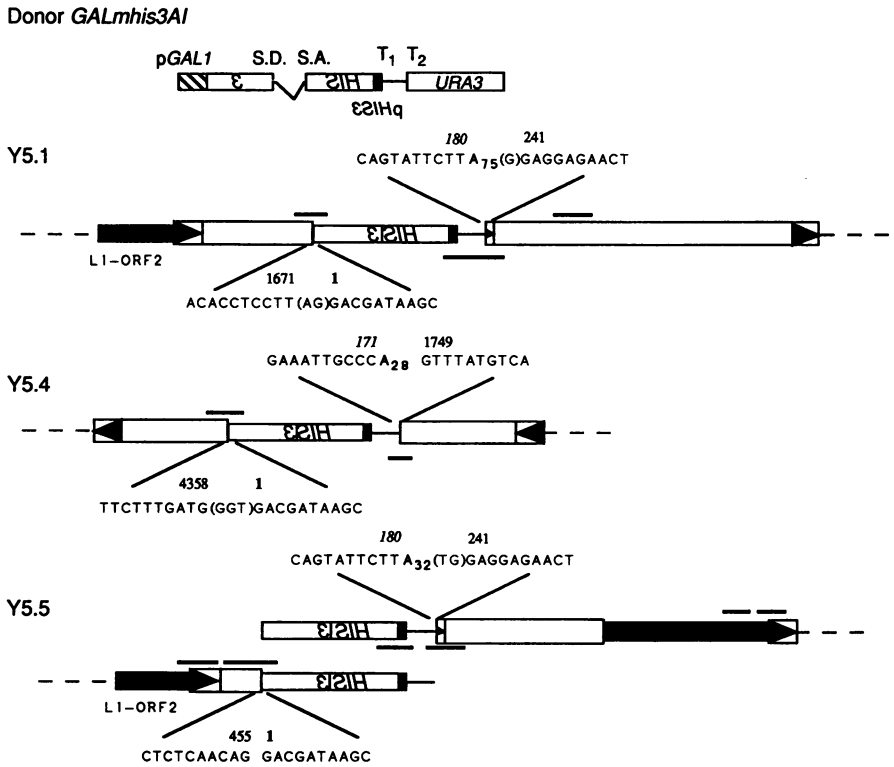

FIG. 4. Structures of *HIS3* pseudogenes and their flanking regions. These structures were deduced from a combination of genomic DNA blots (Fig. 2), PCR assays (not shown), and DNA sequence analyses of PCR products. Junction sequences are shown in the insets; nucleotides in parentheses are of uncertain origin; they separate donor-derived sequences from Ty1 sequences (as do spaces in the right junction of Y5.4 and the left junction of Y5.5, which lack extra bases). Restriction analysis suggests that the Y5.5 junction sequences are derived from two separate copies of *HIS3*; therefore, the left and right junctions are not joined together as in Y5.1 and Y5.4. Symbols are as in Fig. 1; boldface A_n, poly(A) tract; shaded box, L1 ORF2; dashed lines, uncertain flanking sequences; solid line, pSM50-derived sequences lying between T_1 and T_2. Bars indicate sequenced regions. Nucleotide sequence coordinates are from Ty1-H3 (3) (plain text), *URA3* (29) (italic), or *HIS3* (boldface). For *HIS3*, the first transcribed base in pSM50 is designated as 1.

1,240 bp, at which point a poly(A) sequence corresponding to the T_2 site was encountered (within the *URA3* sequence in pSM50 [Fig. 1]). These sites correspond well to the known positions of signals for transcript 3′ end formation previously mapped within the *URA3* 5′ region (37). All three *HIS3* pseudogenes had poly(A) tails of 28 to 75 nucleotides. In both Y5.1 and Y5.5, 3′ of the poly(A) sequences, we encountered Ty1 sequence beginning at position 241 (5′ LTR); in both insertions, the Ty1 sequence reads rightward (Ty1 sequence coordinates are from the Ty1-H3 sequence [3]). Interestingly, this position corresponds precisely to the known transcription start site of the Ty1 element (17). Upon examination of the adjacent Ty1 sequence, we noted a nucleotide sequence diagnostic of the 5′ LTR sequence of the Ty1-H3 element (namely, an A at position 328) used to construct the pSM42 plasmid (3, 5), indicating that this RNA was derived from pSM42. This diagnostic mutation has not been observed in any other sequenced Ty1 elements. The right junction of the Y5.4 pseudogene is connected to *TYB* at position 1749. Using a variety of PCR probes and genomic DNA blots, we made more complete maps of the structures of the inserted *HIS3*/Ty1 sequences (Fig. 4).

Nucleotides of uncertain origin at the left and right junctions. Examination of the junction sequences revealed that for four of six junctions sequenced, there were one to three nucleotides at the junctions that could not be attributed to either template. These nucleotides may have been added in a template-independent manner (as many polymerases appear to do when a 3′ end is reached [12]). In no case were there short stretches of sequence identity between the Ty1 and *HIS3* sequences at the junctions.

Y5.1. Y5.1 has three *HIS3* pseudogenes (Fig. 2A). By genomic blot and pulsed-field gel analyses, we demonstrated that these are tightly linked in a large array of Ty1 and Ty1/L1 sequences on chromosome IV (Fig. 2B and 3). Similar arrays of Ty1 elements and marked Ty1 elements have previously been reported to occur during high-frequency Ty1 transposition (34).

When PCRs with primers derived from pSM50 and L1 ORF2 were attempted, no amplification products were observed. This indicated that no L1 ORF2 sequence existed right of the *HIS3* pseudogenes. Also, no PCR product was obtained when primers L1-2566 and L1′-4305 were used. When primers L1-4903, L1-5264, and L1-5461 were used with SM21′, PCR products were detected. However, primers L1-4686 and SM21′ gave no PCR product. This approach allowed the left junction to be analyzed; we conclude that the Ty1/L1 element in this strain is truncated between positions 4700 and 4900 (Fig. 4; sequence coordinates from Scott et al. [31]).

Y5.4. The structure of the single *HIS3* pseudogene in Y5.4 was simpler to interpret than that of Y5.1 and Y5.5 because there is only one copy of *HIS3* in this strain (Fig. 2). This *HIS3* pseudogene, like the rest, had both the left and the right junctions embedded in Ty1 sequences (Fig. 4). Y5.4 differs

from the other pseudogenes in that the orientation of *HIS3* relative to Ty1 sequences is reversed; also, both junctions are in the *TYB* region. This suggests that in this pseudogene, *HIS3* sequences were joined to Ty1 DNA during synthesis of Ty1 plus-strand DNA.

Y5.5. Y5.5 has two *HIS3* pseudogenes; a *HIS3* probe hybridizes to a single pulsed-field gel band including chromosomes XV and VII in this strain (Fig. 3). Restriction digestion and genomic DNA blotting suggest that these two copies of *HIS3* are tightly linked (Fig. 2B). PCR products were obtained from primers within pSM50 and L1-ORF2, revealing the presence of at least one full Ty1/L1 element at the right junction of one of the *HIS3* pseudogenes (Fig. 4).

Further investigation of the structure of Y5.5 by PCR and sequence analysis indicated that (i) there is a tandem array of Ty1/L1 elements and (ii) these are joined via a single LTR to an intact Ty1 element at the right junction of the Ty1/L1 array. This was demonstrated by amplification of a 2.6-kb PCR product with primers L1-6046 and L1′-2347. Sequencing of this product demonstrated that it consisted of a solo LTR separating two copies of Ty1/L1. This conclusion was further supported by PCR with a variety of pairs of L1-derived primers for PCR that amplified products of the expected size. Similar methods were used to establish the existence of a Ty1/L1 element upstream of a native Ty1 element; these were also separated by a solo LTR. Assuming that all of these structures are a part of the same array, its minimum size would be 25 kb. We do not know whether these tandem arrays of Ty1/L1 and Ty1 elements are upstream of, downstream of, or interspersed with the *HIS3* pseudogenes. The single LTRs at the junctions resemble the Ty1 concatemers studied by Weinstock et al. (34).

DISCUSSION

The RT of the human L1.2A element is clearly able to function in vivo in the yeast cell, because it can reverse transcribe an *mhis3AI*-derived mRNA. However, as is the case when Ty1 provides the RT activity, many of the molecular details of how *HIS3* cDNAs are primed and integrated into the yeast genome remain uncertain. To our surprise, the DNA sequences immediately flanking the *HIS3* pseudogenes are all Ty1 derived, as when Ty1 is the source of RT.

It is likely that the reverse transcription steps involved in *HIS3* pseudogene formation occur within Ty1 VLPs, because the L1 RT is assembled into these VLPs (23). A particulate location for this reverse transcription process is also supported by the higher frequency of L1-mediated pseudogene formation at 22 than at 30°C, which resembles the temperature dependence of Ty1-mediated pseudogene formation as well as Ty1 retrotransposition (8, 27, 28). Recent results suggest that this temperature dependence is correlated with the assembly and/or stability of VLPs in the yeast cell (20a). Interestingly, L1-mediated pseudogene formation is somewhat more active than Ty1-mediated pseudogene formation. The latter may reflect the innate thermal stabilities of these two RTs.

The fact that the *HIS3* pseudogenes isolated were imbedded in Ty1 sequences suggests that their insertion into host DNA is mediated by homologous recombination with preexisting Ty1 or LTR elements in the yeast genome. Because our experiments were performed in an *spt3* strain in which Ty1 is neither expressed (35) nor transpositionally active (8) and because our constructs do not express Ty1 integrase, it seems unlikely that the insertion of the Ty1-pseudogene arrays is actively Ty1 mediated. Previous experiments have shown that the rate of gene conversion of Ty1 elements can be greatly increased by overexpressing Ty1 proteins (24).

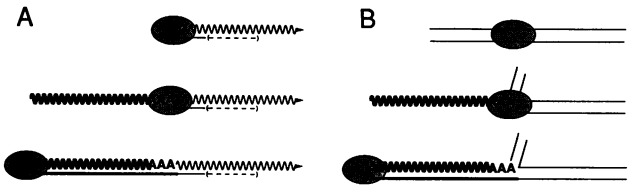

FIG. 5. Mechanism of formation of the right junctions of Y5.1 and Y5.5 mediated by L1 RT resembles proposed R2Bm RT mechanism. Wavy lines, RNA; straight lines, DNA; shaded oval, RT; AAA, 3′ poly(A) tract. (A) A model for the mechanism of formation of the right junctions of Y5.1 and Y5.5. The sequences of these junctions suggest that L1 RT reaches the 5′ end of Ty1/L1 RNA template (top) and jumps templates (middle), allowing copying of the poly(A) 3′ terminus of the *mHIS3* transcript and synthesis of hybrid reverse transcript (bottom). Thin lines, Ty1/L1 RNA or DNA; thick lines, *mHIS3* RNA or DNA. Dashed lines in parentheses indicate uncertainty as to position of initial priming event. (B) Proposed R2Bm cleavage/reverse transcription model (adapted from Luan et al. [22]). The R2Bm retrotransposon, like L1, is characterized by a 3′ poly(A) tract. R2Bm RT nicks the target DNA (top) and then uses the 3′ OH of the nick as a primer on an R2Bm RNA 3′ terminus, requiring a template switch to the 3′ poly(A) tract (middle); extension continues to generate the R2Bm reverse transcript (bottom). Thin lines, target DNA strands; thick lines, R2Bm RNA or DNA.

The juxtaposition of the *HIS3* pseudogene poly(A) sequence with the 5′ transcription start site of Ty1 suggests a mechanism for pseudogene formation. Priming of reverse transcription probably occurred in adjacent Ty1 sequences, and the L1 RT jumped templates from Ty1 RNA (or Ty1 minus-strand DNA in the case of Y5.4) to the *mhis3* mRNA. However, priming does not necessarily occur at the native Ty1 primer binding site, because mutation of this site in the Ty1/L1 construct (10) has little effect on the frequency of His^+ prototroph formation (data not shown). In two cases (Y5.1 and Y5.5), a template jump appears to have occurred from the 5′ end of *GAL1*-driven Ty1/L1 mRNA directly to the poly(A) tail of the *mhis3* mRNA (Fig. 5). Similarly, template jumps appear to have occurred precisely at the 5′ terminus of the *mHIS3* RNA in all three cases, suggesting that a template terminus is a signal for jumping.

The L1-mediated template jumps differ from Ty1-mediated events in that extra bases are inserted at the junctions. Incorporation of extra bases may be a consequence of the RT reaching a template terminus. We propose that addition of nontemplated nucleotides occurs at template termini, as has been described for many DNA polymerases (12). The untemplated bases would form a 3′ overhang, and this could include a region of fortuitous complementarity to a second RNA template; this complementarity could mediate the jump to a second template. If the complementarity is distal to a few noncomplementary bases, these would appear as extra bases in the final pseudogene junction. In principle, all of the jumps could have occurred in this way, and the junctions that lack extra bases may simply represent those molecules that fortuitously added a few untemplated bases of the appropriate sequence to allow pairing with the target template, without intervening extra bases.

This template-jumping hypothesis is of interest in light of recent results obtained for the R2Bm element (an L1-like element from the silkworm). Reverse transcription of R2Bm RNA can be primed directly from a nick in cellular DNA in vitro (22). In both R2Bm and the case described here, the RT jumps templates, utilizing the 3′ poly(A) terminus of a second template (Fig. 5). The mechanism of L1 integration is un-

known, but these results are consistent with the possibility that L1 and R2Bm normally use similar integration and reverse transcription mechanisms (16). In yet a third variation on this template-jumping theme, the *Drosophila* poly(A) retrotransposons TART and Het-A are proposed to use chromosomal telomeres as primers and TART RNA as a template for reverse transcription in a manner similar to the reaction carried out by R2Bm (1, 21).

Another unexpected outcome of these studies is the location of the *HIS3* pseudogenes within a complex array of Ty1 elements. Such arrays have been reported previously, usually when there is some selection for their formation (34). Arrays do not appear to be necessary for expression of the His^+ phenotype in this case, as single-copy *HIS3* pseudogene isolates are readily obtained (14) (Fig. 2A). Rather, the arrays may be a consequence of the unusual and unpredictable end structures of the cDNAs that are actually made during these aberrant retrotranscription events. Perhaps the presence of these complex cDNAs induces recombination enzymes, or they may be more likely to integrate into preexisting Ty1 arrays residing in the genome.

This study demonstrates that the yeast Ty1 system in combination with the *HIS3* pseudogene formation assay provides a straightforward method for examining the in vivo RT functionality of L1 elements. This approach can also be applied to RT genes from a variety of other sources, such as other retrotransposons and hepadnaviruses (18, 32). The *HIS3* pseudogene formation assay will be of great value in determining which L1 elements contain active RTs and are potentially functional. Unlike current in vitro assays, this in vivo assay is easily quantifiable and very sensitive, which should allow for large-scale screening of L1 elements for activity as well as the identification of inactive or partially inactive mutants. This type of assay can also provide important clues as to the replication mechanisms used by diverse retroelements when the structures of the products generated are analyzed.

It may be possible to isolate additional L1-mediated insertion events in which *HIS3* is not joined to Ty1 sequences by using this system. The detection of such events could provide further evidence for the mechanism of pseudogene formation in mammalian cells, which is probably mediated by L1 or related elements (33).

ACKNOWLEDGMENTS

The first three authors contributed equally to this work.

We thank Roxann Ashworth for assistance with DNA sequencing. We thank Leslie Derr for helpful discussions and for communicating unpublished data.

This work was supported in part by National Institutes of Health and American Cancer Society grants to J.D.B. and H.H.K.

REFERENCES

1. **Biessman, H., L. E. Champion, M. O'Hair, K. Ikenaga, B. Kasravi, and J. M. Mason.** 1992. Frequent transpositions of *Drosophila melanogaster* Het-A transposable elements to receding chromosome ends. EMBO J. **11**:4459–4469.
2. **Boeke, J. D., and V. G. Corces.** 1989. Transcription and reverse transcription in retrotransposons. Annu. Rev. Microbiol. **43**:403–433.
3. **Boeke, J. D., D. J. Eichinger, D. Castrillon, and G. R. Fink.** 1988. The *Saccharomyces cerevisiae* genome contains functional and nonfunctional copies of transposon Ty1. Mol. Cell. Biol. **8**:1432–1442.
4. **Boeke, J. D., D. J. Eichinger, and G. Natsoulis.** 1991. Doubling Ty1 element copy number in *Saccharomyces cerevisiae*: host genome stability and phenotypic effects. Genetics **129**:1043–1052.
5. **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.
6. **Boeke, J. D., F. Lacroute, and G. R. Fink.** 1984. A positive selection for mutants lacking orotidine 5′-phosphate decarboxylase activity in yeast: 5-fluoro-orotic acid resistance. Mol. Gen. Genet. **197**:345–346.
7. **Boeke, J. D., and S. B. Sandmeyer.** 1991. Yeast transposable elements, p. 193–261. *In* J. Broach, E. Jones, and J. Pringle (ed.), The molecular and cellular biology of the yeast *Saccharomyces*. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
8. **Boeke, J. D., C. A. Styles, and G. R. Fink.** 1986. *Saccharomyces cerevisiae SPT3* gene is required for transposition and transpositional recombination of chromosomal Ty elements. Mol. Cell. Biol. **6**:3575–3581.
9. **Carle, G. F., and M. V. Olson.** 1985. An electrophoretic karyotype for yeast. Proc. Natl. Acad. Sci. USA **82**:3756–3760.
10. **Chapman, K. B., A. S. Byström, and J. D. Boeke.** 1992. Initiator methionine tRNA is essential for Ty1 transposition in yeast. Proc. Natl. Acad. Sci. USA **89**:3236–3240.
11. **Chu, G., D. Vollrath, and R. Davis.** 1986. Separation of large DNA molecules by contour-clamped homogeneous electric fields. Science **234**:1582–1585.
12. **Clark, J. M.** 1988. Novel non-templated nucleotide addition reactions catalyzed by prokaryotic and eukaryotic DNA polymerases. Nucleic Acids Res. **16**:9677–9686.
13. **Curcio, M. J., and D. J. Garfinkel.** 1991. Single-step selection for Ty1 element retrotransposition. Proc. Natl. Acad. Sci. USA **88**: 936–940.
14. **Derr, L. K., J. N. Strathern, and D. J. Garfinkel.** 1991. RNA-mediated recombination in *S. cerevisiae*. Cell **67**:355–364.
15. **Dombroski, B. A., S. L. Mathias, E. Nanthakumar, A. F. Scott, and H. H. Kazazian, Jr.** 1991. Isolation of an active human transposable element. Science **254**:1805–1808.
16. **Eickbush, T. H.** 1992. Transposing without ends: the non-LTR retrotransposable elements. New Biol. **4**:430–440.
17. **Elder, R. T., E. Y. Loh, and R. W. Davis.** 1983. RNA from the yeast transposable element Ty1 has both ends in the direct repeats, a structure similar to retrovirus RNA. Proc. Natl. Acad. Sci. USA **80**:2432–2436.
18. **Gabriel, A., and J. D. Boeke.** 1991. Reverse transcriptase encoded by a retrotransposon from the trypanosomatid *Crithidia fasciculata*. Proc. Natl. Acad. Sci. USA **88**:9794–9798.
19. **Jacobs, E., M. Dewerchin, and J. D. Boeke.** 1988. Retrovirus-like vectors for *Saccharomyces cerevisiae*: integration of foreign genes controlled by efficient promoters into yeast chromosomal DNA. Gene **67**:259–269.
20. **Kazazian, H. H., Jr., C. Wong, H. Youssoufian, A. F. Scott, D. G. Phillips, and S. E. Antonarakis.** 1988. Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man. Nature (London) **332**:164–166.

20a. **Keeney, J., C. Baker-Brachmann, and J. D. Boeke.** Unpublished data.

21. **Levis, R. W., R. Ganesan, K. Houchens, L. A. Tolar, and F.-M. Sheen.** 1993. Transposons in place of telomeric repeats at a Drosophila telomere. Cell **75**:1083–1093.
22. **Luan, D. D., M. K. Korman, J. L. Jakubczak, and T. H. Eickbush.** 1993. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. Cell **72**:595–605.
23. **Mathias, S. L., A. F. Scott, H. H. Kazazian, Jr., J. D. Boeke, and A. Gabriel.** 1991. Reverse transcriptase encoded by a human retrotransposon. Science **254**:1808–1810.
24. **Melamed, C., Y. Nevo, and M. Kupiec.** 1992. Involvement of cDNA in homologous recombination between Ty elements in yeast. Mol. Cell. Biol. **12**:1613–1620.
25. **Natsoulis, G., W. Thomas, M. C. Roghmann, F. Winston, and J. D. Boeke.** 1989. Ty1 transposition in *Saccharomyces cerevisiae* is nonrandom. Genetics **123**:269–279.
26. **Ochman, H., A. S. Gerber, and D. L. Hartl.** 1988. Genetic applications of an inverse polymerase chain reaction. Genomics **120**:621–623.
27. **Paquin, C. E., and V. M. Williamson.** 1984. Temperature effects on the rate of Ty transposition. Science **226**:53–55.

28. **Paquin, C. E., and V. M. Williamson.** 1986. Ty insertions at two loci account for most of the spontaneous antimycin A resistance mutations during growth at 15°C of *Saccharomyces cerevisiae* strains lacking *ADH1*. Mol. Cell. Biol. **6:**70–79.
29. **Rose, M., P. Grisafi, and D. Botstein.** 1984. Structure and function of the yeast *URA3* gene: expression in *Escherichia coli*. Gene **29:**113–124.
30. **Rose, M. D., F. Winston, and P. Hieter.** 1990. Methods in yeast genetics: a laboratory course manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
31. **Scott, A. F., B. J. Schmeckpeper, M. Abdelrazik, C. T. Comey, B. O'Hara, J. P. Rossiter, T. Cooley, P. Heath, K. D. Smith, and L. Margolet.** 1987. Origin of the human L1 elements: proposed progenitor genes deduced from a consensus DNA sequence. Genomics **1:**113–125.
32. **Tavis, J. E., and D. Ganem.** 1993. Expression of functional hepatitis B virus polymerase in yeast reveals it to be the sole viral protein required for correct initiation of reverse transcription. Proc. Natl. Acad. Sci. USA **90:**4107–4111.
33. **Tchénio, T., E. Segal-Bendirdjian, and T. Heidmann.** 1993. Generation of processed pseudogenes in murine cells. EMBO J. **12:**1487–1497.
34. **Weinstock, K., M. Mastrangelo, T. Burkett, D. Garfinkel, and J. Strathern.** 1990. Multimeric arrays of the yeast retrotransposon Ty. Mol. Cell. Biol. **10:**2882–2892.
35. **Winston, F., K. J. Durbin, and G. R. Fink.** 1984. The *SPT3* gene is required for normal transcription of Ty elements in *S. cerevisiae*. Cell **39:**675–682.
36. **Xiong, Y., and T. H. Eickbush.** 1990. Origin and evolution of retroelements based on their reverse transcriptase sequences. EMBO J. **9:**3353–3362.
37. **Yarger, J. G., G. Armilei, and M. C. Gorman.** 1986. Transcription terminator-like element within a *Saccharomyces cerevisiae* promoter region. Mol. Cell. Biol. **6:**1095–1101.