Title

Ty3, a yeast retrotransposon associated with tRNA genes, has homology to animal retroviruses.

Journal

Molecular and Cellular Biology, 8(12)

ISSN

0270-7306

Authors

Hansen, LJ
Chalker, DL
Sandmeyer, SB

Publication Date

1988-12-01

DOI

10.1128/mcb.8.12.5245

MOLECULAR AND CELLULAR BIOLOGY, Dec. 1988, p. 5245–5256
0270-7306/88/125245-12$02.00/0

Vol. 8, No. 12

Ty3, a Yeast Retrotransposon Associated with tRNA Genes, Has Homology to Animal Retroviruses

LORI J. HANSEN, DOUGLAS L. CHALKER, AND SUZANNE B. SANDMEYER*

Department of Microbiology and Molecular Genetics, College of Medicine, University of California, Irvine, California 92717

Received 11 July 1988/Accepted 22 September 1988

Ty3, a retrotransposon of *Saccharomyces cerevisiae*, is found within 20 base pairs (bp) of the 5′ ends of different tRNA genes. Determination of the complete nucleotide sequence of one Ty3 retrotransposon (Ty3-2) shows that the element is composed of an internal domain 4,748 bp long flanked by long terminal repeats of the 340-bp sigma element. Three open reading frames (ORFs) longer than 100 codons are present in the sense strand. The first ORF, *TYA3*, encodes a protein with a motif found in the nucleic acid-binding protein of retroviruses. The second ORF, *TYB3*, has homology to retroviral *pol* genes. The deduced amino acid sequence of the reverse transcriptase domain shows the greatest similarity to *Drosophila* retrotransposon 17.6, with 43% identical residues. The inferred order of functional domains within *TYB3*—protease, reverse transcriptase, and endonuclease—resembles the order in *Drosophila* element 17.6 and in animal retroviruses but is different from that found in yeast elements Ty1 and Ty2. A second Ty3 element (Ty3-1) from a standard laboratory strain was overexpressed and shown to transpose.

The genome of *Saccharomyces cerevisiae* contains at least three families of retrotransposons: Ty1, Ty2 (reviewed in references 24 and 54), and Ty3 (9). Retrotransposons are mobile genetic elements that transpose through an RNA intermediate and resemble retroviruses except for the apparent absence of an extracellular phase. The first retrotransposon identified in *S. cerevisiae* was named Ty, for transposable element in yeast (5). Two closely related forms of that element have been characterized and designated Ty1 and Ty2 (8, 35, 47, 88). Ty1 and Ty2 are similar in nucleotide and amino acid sequences but have two large regions of heterogeneity which were first demonstrated by heteroduplex analysis (47). Ty1 and Ty2 have an internal domain called epsilon which is 5.3 kilobase pairs (kbp) long. Epsilon is flanked by direct repeats of delta elements, which are 332 to 338 base pairs (bp) long (22, 29). Transcription initiates in the upstream delta long terminal repeat (LTR), and the signal for polyadenylation occurs in the second LTR, downstream of the initiation site sequence (19). Thus, an almost full-length transcript with redundant termini is generated. Ty1 and Ty2 insertions can influence the transcription of neighboring genes positively (21, 89) or negatively (6, 18, 67). The epsilon region contains two overlapping open reading frames (ORFs), which have similarity to the retroviral *gag* and *pol* genes. Boeke and co-workers (2) showed that Ty1 transposition is dependent on Ty1 transcription and that the transposition occurs through an RNA intermediate. These results demonstrated the functional similarity of retrotransposons to retroviruses.

Ty3 is a more recently discovered element (9). It consists of a 4.7-kbp internal domain flanked by direct repeats of sigma elements, each 340 bp long. Characterization of one Ty3 element, now designated Ty3-1, showed that it has the following retroviruslike features: (i) flanking direct repeats of the insertion site sequence, (ii) LTR sequences which terminate in conserved inverted repeats, (iii) a potential primer-binding site for minus-strand DNA synthesis, (iv) a purine-rich region potentially involved in plus-strand DNA synthesis, and (v) an almost full-length transcript with redundant termini. Isolated sigma elements are found exclusively 16 to 19 bp upstream from the 5′ ends of tRNA genes (4, 13, 69, 70). This association is also seen for Ty3-1, which is found 16 bp upstream of a $tRNA^{Cys}$ gene. One possible explanation for the isolated sigma elements next to tRNA genes, then, is that they are the end products of LTR-LTR recombination events which occurred after position-specific Ty3 retrotransposition.

Retrotransposons have organizational and structural features in common with the provirus form of retroviruses (reviewed in references 27 and 85). Retrotransposons encode proteins analogous to the group-specific antigens (*gag*) and the *pol* gene polyprotein of retroviruses. The retroviral *gag* gene is defined by the first ORF and encodes proteins that interact with the viral RNA to form a nucleocapsid. Virus-like particles have been isolated for copia, a *Drosophila* retrotransposon (74), and Ty1 (30, 55). In the case of Ty1, the first ORF, *TYA*, was shown to encode proteins that make up the nucleocapsid (1, 60). The retrovirus *pol* gene encodes a polyprotein with several enzymatic domains, including a reverse transcriptase (polymerase and RNase H) and an endonuclease. A protease-coding sequence occurs in *gag* in avian retroviruses but is found at the beginning of the *pol* gene in most other retroviruses. The reverse transcriptase polymerase and RNase H activities catalyze DNA synthesis from the RNA template and degradation of the RNA portion of the resulting DNA-RNA heteroduplex, respectively. The second ORF of Ty1 and Ty2, *TYB*, encodes a polyprotein which has domains corresponding to the retroviral protease, endonuclease, and reverse transcriptase, in that order. Typically, the first and second ORFs of retroviruses and retrotransposons either overlap or are separated by a stop codon. In systems in which protein expression has been examined, translation of proteins encoded in the *pol* gene or in the second ORF of retrotransposons has been shown to be dependent on readthrough from the first ORF. This readthrough is mediated by nonsense suppression or frameshifting (reviewed in reference 12). A significant difference between retroviruses and retrotransposons is that the latter

* Corresponding author.

do not encode envelope proteins and appear to lack the extracellular phase of the retroviral life cycle. The term retroid element will be used to refer to repeated elements and viruses which replicate through a reverse transcriptase-mediated process (27).

Here we report the complete nucleotide sequence of Ty3-2. Comparison of the deduced amino acid sequences of the ORFs suggests that Ty3 is more closely related to 17.6 and retroviruses than to the other yeast retrotransposons, Ty1 and Ty2. To examine the ability of Ty3 to transpose, Ty3-1 and Ty3-2 were overexpressed in a galactose-inducible system. The results of this experiment showed that whereas Ty3-1 is capable of transposition, Ty3-2 is not.

MATERIALS AND METHODS

Recombinant DNA manipulations. Bacterial culture conditions and recombinant DNA manipulations were as previously described (9), unless otherwise noted. Chromosomal DNA from *S. cerevisiae* AB972 (9) was digested with *Eco*RI and fractionated on a 1% agarose gel. DNA fragments 5.5 to 7.5 kbp long were isolated from low-gelling-temperature agarose (Bio-Rad Laboratories), subcloned into the *Eco*RI site of pIB121 (International Biotechnologies, Inc. [IBI]), and transformed into *Escherichia coli* HB101 (F^- *hsdS20* r_B^- m_B- *recA13 leuB6 ara-14 proA2 lacY1 galK2 rpsL20* [Sm^r] *xyl-5 mtl-1 supE44* λ^-). Plasmid pIBI21 contains a cloning region downstream of the bacteriophage T7 promoter and the IBI primer sequence, which is followed by the M13 primer sequence. The bacteriophage f1 origin of replication is present in pIBI21 and allows production of single-stranded plasmid DNA when bacteria transformed with pIBI21 are superinfected with the helper phage M13K07 (IBI). Transformants containing the Ty3-2 insert were identified by colony hybridization (33) by using a radiolabeled internal restriction fragment as a probe. This plasmid is designated pTy3-2. The Ty3 cloned previously (pSBS12; 9) is contained on a *Hin*dIII-*Eco*RI restriction fragment ligated into the *Hin*dIII and *Eco*RI sites of pIBI20 and is designated pTy3-1.

To facilitate studies of Ty3 transposition, the *GAL1-10* upstream activating sequence (UAS) was fused upstream of the TATAA sequences of Ty3-1 and Ty3-2. In step 1 of the construction, a 276-bp *Alu*I fragment (nucleotide positions 123 to 398 in sigma; 9) containing the Ty3 TATAA sequences and transcription start site was subcloned into the *Sma*I site of the pIBI20 polylinker. This construct was cleaved with *Sal*I in the polylinker and with *Xho*I in the downstream end of the sigma element to produce a sigma promoter fragment with *Xho*I-compatible ends. This fragment was then inserted into the *Xho*I site downstream of the *GAL1-10* UAS in a derivative of pHZ18 (78) to produce pALG28. The *Hin*dIII-*Xho*I fragment from pALG28 containing the yeast *URA3* gene and the *GAL1-10* UAS-sigma promoter fusion was cloned into a site in pTy3-1 or pTy3-2 created by complete *Hin*dIII digestion and partial *Xho*I digestion. Screening by restriction digestion of these constructs identified plasmids pGTy3-1 and pGTy3-2, in which the *GAL1-10* UAS sigma promoter is fused to the Ty3-1 or Ty3-2 internal domain. pGTy3-1 and pGTy3-2 were converted to high-copy yeast vectors, pEGTy3-1 and pEGTy3-2, respectively, by insertion of the 2.2-kbp *Eco*RI fragment from the yeast 2μm episome.

Nucleotide sequencing strategy. The directed-deletion strategy of Henikoff (37) was used to create overlapping subclones for sequence analysis of the noncoding strand of Ty3-2 by the dideoxy-chain termination method (71). Single-stranded templates were made by superinfection of the pTy3-containing *E. coli* NM522 (Δ [*lac-proAB*] *thi* Δ*hsd-5 sup-5 supE* [F′ *proAB* $lacI^qZ$ Δ*M15*]) with the helper phage M13K07 and used for the sequence reactions.

The restriction enzymes *Alu*I, *Rsa*I, and *Sau*3A, which cut frequently within the Ty3-2 sequence, were used to make random subclones of suitable length for sequence analysis of the coding strand. Small-scale preparations of this DNA were obtained with the boiling method of Holmes and Quigley (39). Sequences of these inserts were determined from double-stranded plasmids in polymerase reactions primed with the M13 universal or IBI reverse sequencing primer (34) and by using the dideoxy-chain termination method (71). All sequence analyses used the sequenase enzyme (United States Biochemical Corp.) and [^{35}S]dATP (1,000 Ci/mmol; Amersham Corp.). Six synthetic oligonucleotides (Operon Technologies, Inc.) which hybridize to Ty3-2 sequences were used to allow analysis of remaining regions. The nucleotide sequence was compiled, edited, and translated by using the DNA sequence analysis programs of A. Goldin and G. Gutman (University of California, Irvine). Comparisons with the Genbank Nucleic Acid Data Base and the National Biomedical Research Foundation Protein Data Base were made with the use of the University of Wisconsin Genetics Computer Group programs on a VAX computer (14). The amino acid comparisons of the reverse transcriptase and endonuclease were generated with the progressive-alignment programs of Feng and Doolittle (23) on a VAX computer.

Transposition. Yeast strains were cultured by standard methodology (73). Strains yVB109 (*MAT***a** Δ*trp1-901 ura3-52 his3-200 ade2-101 lys2-1 leu1-12 can1-100*) and yVB110 (V. W. Bilanchone, personal communication) were transformed with plasmid pEGTy3-1 or pEGTy3-2 by a modification of the procedure of Ito et al. (42). Strain yVB110 is an isogenic derivative of yVB109 from which the three endogenous Ty3 elements were serially deleted. Three strains from which different Ty3 elements were deleted were obtained by *URA3* disruption of individual Ty3 elements, followed by selection on 5-fluoro-orotic acid-containing medium for colonies with the *ura3* mutation. Strains containing multiple Ty3 deletions were derived by standard genetics (Bilanchone, manuscript in preparation). yVB110 transformants containing the inducible plasmids were selected on synthetic complete medium minus uracil on the basis of the uracil prototrophy conferred by the plasmid. yVB110 transformants were streaked onto the same medium or onto medium with 2% galactose substituted for glucose as a carbon source and incubated at 23 or 30°C for 10 days. At the end of that time, 10 colonies from each condition were streaked for single colonies on rich medium (1% yeast extract, 2% peptone, 2% glucose). Fifty isolates representing each original colony were patched onto 5-fluoro-orotic acid-containing medium to select cells which had lost the *URA3* plasmid marker (3). These cells were streaked onto nitrocellulose filters on rich medium and grown for about 14 h. Filters were processed and hybridized as described previously (73). The probe was a radiolabeled fragment from the internal domain of Ty3. Strain yVB109 was similarly transformed and grown under inducing and noninducing conditions. Single colonies were isolated and cured of the plasmid as described above. Genomic DNA for Southern blot analysis was prepared by the method of Boeke et al. (2). DNA for the analysis displayed in Fig. 5 was prepared from different clonal isolates of a galactose-grown yVB109-

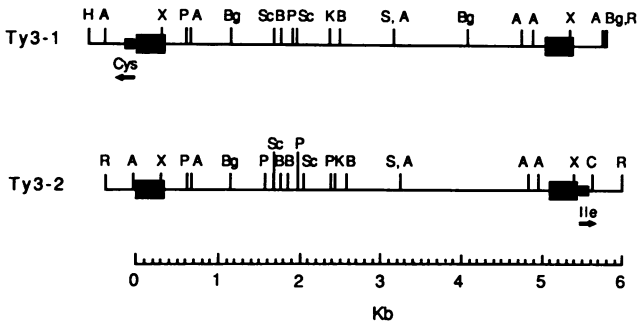


FIG. 1. Restriction site maps of yeast genomic fragments containing Ty3-1 and Ty3-2. Yeast inserts in pTy3-1 and pTy3-2 were mapped. Sigma LTRs are shown as hatched boxes, and tRNA genes are shown as solid boxes. The amino acid accepted by the encoded tRNA is indicated below each tRNA gene. The direction of transcription of the Ty3 elements is left to right, as shown. The directions of transcription of the tRNA genes are indicated by arrows. The scale shown at the bottom (in kilobases [Kb]) is aligned with the left ends of the upstream sigma elements. Abbreviations: A, *Acc*I; B, *Bst*EII; Bg, *Bgl*II; C, *Cla*I; H, *Hin*dIII; K, *Kpn*I; P, *Pst*I; R, *Eco*RI; S, *Sal*I; Sc, *Sca*I; X, *Xho*I.

pEGTy3-1 transformant. We screened 230 colonies in 24 DNA preparations, each representing groups of 8 to 10 colonies. Genomic DNA from these pooled cultures was isolated, digested with *Eco*RI, fractionated by electrophoresis in 0.8% agarose buffered in TBE (2.5 mM EDTA, 45 mM borate, 133 mM Tris hydrochloride, pH 8.3), and transferred to nitrocellulose by the method of Southern. Nitrocellulose-bound DNA was hybridized, and filters were washed as previously described (9) and exposed to Kodak XAR-5 film in the presence of a Cronex Quanta-III Intensifying Screen (Du Pont Co.) at −70°C overnight. DNA was also prepared and analyzed in the same way from individual colonies which contributed to pools that showed evidence of Ty3 rearrangements. In this way, clonal isolates were identified which are homogeneous with respect to the presence of a new Ty3-hybridizing fragment. DNAs from some of these were analyzed by hybridization to Ty3- and sigma-specific probes (see Fig. 5, legend).

RESULTS

Isolation of Ty3-2 from *S. cerevisiae* AB972. Ty3-1 was originally identified in a lambda clone from *S. cerevisiae* AB972 and subcloned on a 6.5-kbp *Hin*dIII-*Eco*RI fragment into the *E. coli* vector pIBI20 (IBI) (9). Southern blot analysis of genomic DNA from AB972 with probes from the internal region of the cloned Ty3-1 element revealed the presence of a second Ty3 element. This element, called Ty3-2, occurs on an *Eco*RI fragment 6.5 kbp long. To clone Ty3-2, *Eco*RI-digested AB972 genomic DNA 5.5 to 7.5 kbp long was inserted into the *Eco*RI site of pIBI21 (IBI) and transformed into *E. coli* HB101. Positive subclones were identified by colony hybridization to a probe from the internal domain of Ty3-1. These plasmids contained the expected 6.5-kbp inserts. The IBI plasmids containing the Ty3-1 and Ty3-2 clones are referred to as pTy3-1 and pTy3-2, respectively.

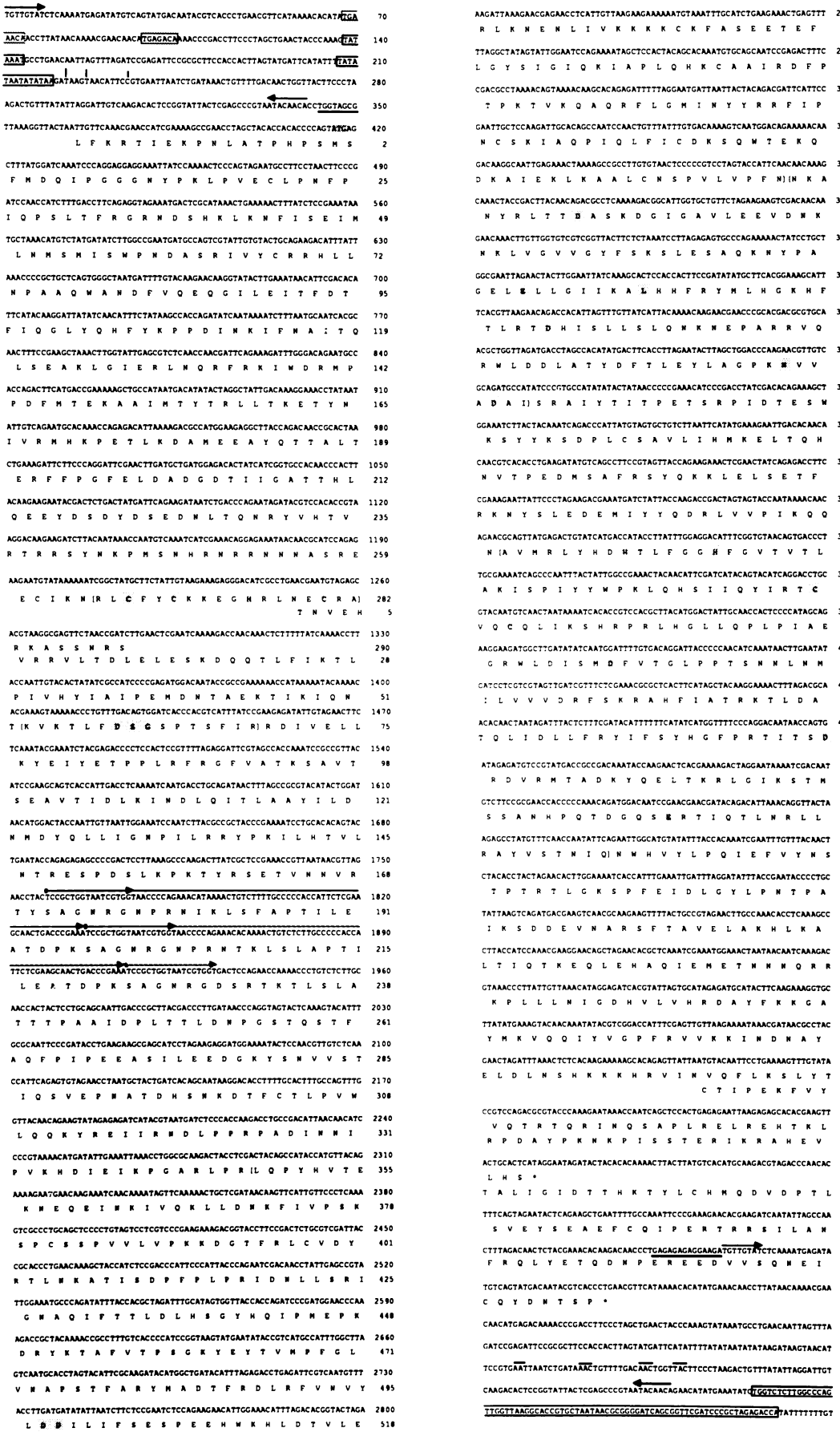
Plasmids pTy3-1 and pTy3-2 were analyzed by restriction enzyme digestion and subsequent fractionation by agarose gel electrophoresis. Figure 1 shows a comparison of the resulting restriction maps of Ty3-1 and Ty3-2. The sigma elements will be referred to as the 5′ or 3′ sigma element to identify each with respect to the direction of Ty3 transcription (9). Overall, Ty3-1 and Ty3-2 have similar restriction patterns for the enzymes used in this analysis. There are some differences, however. Ty3-1 has an additional *Bgl*II site; Ty3-2 contains two additional *Pst*I sites and one extra *Bst*EII site. Ty3-2 is also somewhat longer than Ty3-1, and this difference appears to be accounted for by the presence of a 78-bp repeat which does not occur in Ty3-1 (unpublished data). The regions outside of the sigma elements have different patterns of restriction sites, as expected for Ty3 elements at distinct genomic sites.

The orientation of the Ty3-1 and Ty3-2 sigma elements and positions of the associated tRNA genes were first determined by Southern blot analysis with sigma-specific and total-yeast tRNA probes (9; unpublished data) and confirmed by sequence analysis. The Ty3-1 and Ty3-2 sigma elements occur as terminal direct repeats with the same orientation relative to the internal domain. Ty3-1 and Ty3-2 are inserted, however, in opposite orientations with respect to the 5′ ends of the flanking tRNA genes (Fig. 1). Ty3-1 is associated with a $tRNA^{Cys}$ gene (9), and Ty3-2 is associated with a $tRNA^{Ile}$ gene (this work). One result of these different orientations is that the direction of Ty3-2 transcription is toward that of the associated tRNA gene and on the same strand, while the direction of Ty3-1 transcription is away from that of the tRNA gene and on the opposite strand. Of 17 sigma elements and associated tRNA genes cloned from strain AB972, 11 show the same orientation as Ty3-1 with respect to the tRNA gene and 6 show the same orientation as Ty3-2 (69). Thus, although individual sigma elements are primarily in one orientation relative to the associated tRNA genes, both orientations are represented by these two Ty3 elements.

DNA sequence analysis of Ty3-2. The complete nucleotide sequence of Ty3-2 was determined for both strands (Fig. 2). The entire Ty3-2 retrotransposon is 5,428 bp long and consists of two 340-bp sigma elements flanking an internal domain of 4,748 bp. There is a 5-bp (GAACA) direct repeat flanking Ty3-2, which is different from that flanking Ty3-1 (GAAAG) and reflects the uniqueness of these insertion sites. The $tRNA^{Ile}$ gene associated with Ty3-2 is located 18 bp downstream of the 3′ sigma element (nucleotides 5446 to 5519), and its coding sequence is identical to the $tRNA^{Ile}$ sequence determined by Pixa et al. (64).

The positions of the 5′ ends and polyadenylation sites of the Ty3 5.2-kb transcripts are heterogeneous, with major endpoints within the sigma element (Fig. 2; 9). These transcript endpoints were determined with poly(A) RNA obtained from the same strain background which was the source of the Ty3-1 and Ty3-2 elements. The 5′ sigma element contains a major transcriptional start site at nucleotide 223, which is downstream of two potential TATA sequences, TATAAA and TATAATATATAA. The 3′ end of the Ty3 transcript is heterogeneous but mapped to the 3′ sigma element downstream of the initiation sequences. This results in a terminally redundant RNA, which is characteristic of retroid elements.

The Ty3-1 and Ty3-2 transcripts contain a potential primer-binding site for minus-strand DNA synthesis and a purine-rich region implicated in plus-strand DNA synthesis (80, 85). The nucleotide sequences of Ty3-1 and Ty3-2 have identical sigma elements and plus- and minus-strand primer sites. The deduced primer-binding site is located 2 bp inside the 5′ end of the internal domain and consists of eight nucleotides that are complementary to the 3′ end of initiator $tRNA^{Met}$ (86). If G-U pairing could occur, the complementary region would



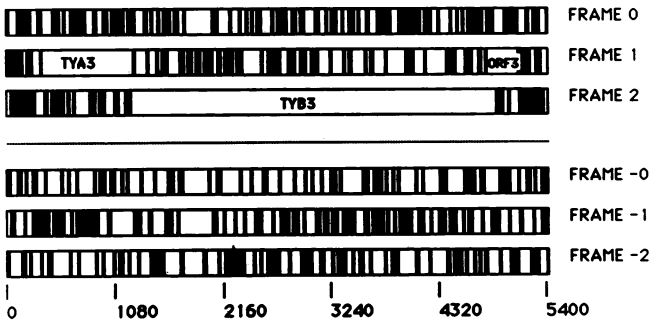


FIG. 3. ORF analysis of the Ty3-2 sequence. The nucleotide sequence of Ty3-2 was analyzed with the ORF display program of MacGene on an Apple Macintosh Plus computer. The top panel shows the three reading frames of the strand which is represented by the 5.2-kb transcript; the bottom panel shows the three reading frames of the opposite strand. Vertical lines indicate the presence of stop codons (thickness is not to scale). The scale at the bottom is in nucleotides, starting with the left end of the upstream sigma LTR.

be 10 nucleotides long. A 10-nucleotide sequence immediately inside the internal domain of Ty1 which is complementary to $tRNA^{Met}$ was previously pointed out (17). In retroviruses, the tRNA-binding sequence occurs 2 bp inside the internal domain and extends for 16 to 18 nucleotides. Thus, if initiator $tRNA^{Met}$ is the Ty primer, then the length of the primer-binding site is shorter for yeast retrotransposons than for retroviruses. The purine-rich region of Ty3 occurs in the 3′ end of the internal domain and consists of 14 consecutive purines in the transcribed strand. The length of the purine tract is 15 to 21 nucleotides for retroviruses. Retrotransposons have shorter purine-rich regions, e.g., 8 of 13 positions in Ty1 (8, 35), 7 of 13 positions in Ty2 (88), 8 of 9 positions in copia (59), and 11 of 11 positions in 17.6 (68).

ORF analysis of Ty3-2. The nucleotide sequence of Ty3-2 was translated in all three reading frames by using the DNA programs of A. Goldin and G. Gutman (University of California, Irvine). Three ORFs longer than 100 codons occur in the 5.2-kb RNA strand. These will be referred to as *TYA3-2*, *TYB3-2*, and ORF3. These gene names were chosen to be consistent with yeast nomenclature and to indicate functional similarity with the *TYA* and *TYB* genes from Ty1 and Ty2 but to distinguish the source. The protein sequences predicted from these ORFs are shown in Fig. 2, below the nucleotide sequence. A computer-generated ORF analysis of Ty3-2 is displayed in Fig. 3, indicating the positions of all potential termination codons. The Ty3-2 ORFs *TYA3-2*, *TYB3-2*, and ORF3 are 307, 1,221, and 110 codons long, respectively. The first ORF, *TYA3-2*, begins at nucleotide 365 and ends at nucleotide 1285. The first methionine codon is at nucleotide positions 416 to 418, starting 76 bp inside the internal domain. The sequence flanking the ATG, C-C-A-G-T-A-T-G-A is consistent with the consensus context for eucaryotic translation initiation sites (48). *TYA3-2* extends to the first terminator, 290 codons downstream. The second ORF, *TYB3-2*, extends from nucleotides 1248 to 4910 and overlaps *TYA3-2* by 13 codons in the plus-one frame. The first methionine codon in *TYB3-2* does not occur until nucleotide positions 1362 to 1364, 39 codons downstream of the beginning of the ORF. This ATG does occur in an acceptable context for translation initiation, but its position in the ORF suggests that initiation may not occur here. It is also possible that the initiator codon is supplied on a spliced 5′ terminus. Nevertheless, no yeast consensus splice sequences (GTAPyGT;TACTAAC...AG; 50, 63, 77) are found in Ty3-2. These observations and comparisons with other retroid elements (see Discussion) suggest that a plus-one translational frameshift is required for expression of *TYB3-2*.

ORF3 begins at nucleotide 4805 and ends at nucleotide 5134, 15 codons inside the 3′ sigma element. The first methionine is encoded 49 codons into ORF3, at nucleotide positions 4949 to 4951, and does not conform to the consensus context for initiation. *TYB3-2* and ORF3 overlap by 36 codons, with ORF3 in the minus-one frame (the same frame as *TYA3*) with respect to *TYB3-2*. The significance of ORF3 is not known. Sequence analysis of Ty3-1 in the region of the *TYB3-2*–ORF3 overlap shows that Ty3-1 contains a single-base insertion compared with Ty3-2 (unpublished data). If other reading frame differences do not occur between Ty3-1 and Ty3-2, then *TYB3-1* would extend through the sequence which constitutes Ty3-2 ORF3. Further sequence analysis is required to explore this possibility.

Comparison of the amino acid sequences of Ty3-2 ORFs with those of proteins encoded by other retroid elements. Protein sequences predicted from *TYA3-2*, *TYB3-2*, and ORF3 nucleotide sequences were used to search the National Biomedical Research Foundation Protein Data Base (version 13) to identify sequence similarities between predicted Ty3-2 proteins and previously described proteins. The portion of retroid elements that encodes reverse transcriptase is characteristically the most conserved domain (52). The data base search revealed that a region between codons 348 and 632 of *TYB3-2* has 43% identity with the polymerase domain of reverse transcriptase encoded by 17.6, a *Drosophila* retrotransposon. Inspection of *TYA3-2* and *TYB3-2* for conserved motifs found in retroid elements resulted in identification of domains with similarities to previously described nucleic acid-binding proteins, proteases, and endonucleases, in addition to reverse transcriptases. To determine the relatedness of Ty3-encoded proteins to those encoded by other retroid elements, the Ty3 ORFs were partitioned into putative functional domains based on previous comparisons (43, 52, 82) and compared with proteins from nine other retroid elements. The results of these comparisons are displayed in Fig. 4.

Nucleic acid-binding protein. The retroviral *gag* gene encodes the structural proteins of the nucleocapsid. The se-

FIG. 2. Nucleotide sequence of Ty3-2 and associated tRNA gene and amino acid sequence deduced from ORFs. The Ty3-2 nucleotide sequence is shown above the predicted amino acid sequence. Numbering of the nucleotide sequence begins at the first base of the upstream sigma element. Orientation is the same as in Fig. 1. Arrows above the ends of the sigma elements indicate the 8-bp perfect inverted repeats. Potential TATAA sequences (solid lines) and upstream pheromone control sequences (49) (broken lines) are boxed. The 5′ ends of the 5.2-kb transcript are indicated by vertical bars over the sequence; the 3′ ends are indicated by horizontal bars. Predicted minus- and plus-strand primer regions are underlined. Arrows over the internal domain indicate the position of the 78-bp direct repeat which occurs in Ty3-2. The copy of this repeat, which is missing in Ty3-1, is indicated by the dashed arrow (unpublished data). The presumed initiator ATG codon in *TYA3-2* is shaded. The sequence of the $tRNA^{Ile}$ gene downstream of the Ty3-2 3′ sigma element is boxed. The deduced amino acid sequences of the three ORFs longer than 100 codons are shown below the nucleotide sequence in the single-letter code. Numbering of the amino acids begins with the putative initiation methionine for *TYA3-2* and with the first amino acid encoded in *TYB3-2* and ORF3. Brackets indicate the amino acid sequences displayed in the alignments in Fig. 4 and also the RNase H domain shown by Johnson et al. (43), which is not shown in an alignment in Fig. 4. Amino acid residues which are conserved among the sequences compared in Fig. 4 are shaded, as are amino acid residues of RNase H which are in *TYB3-2* and found to be conserved in retroid RNase H domains (43).

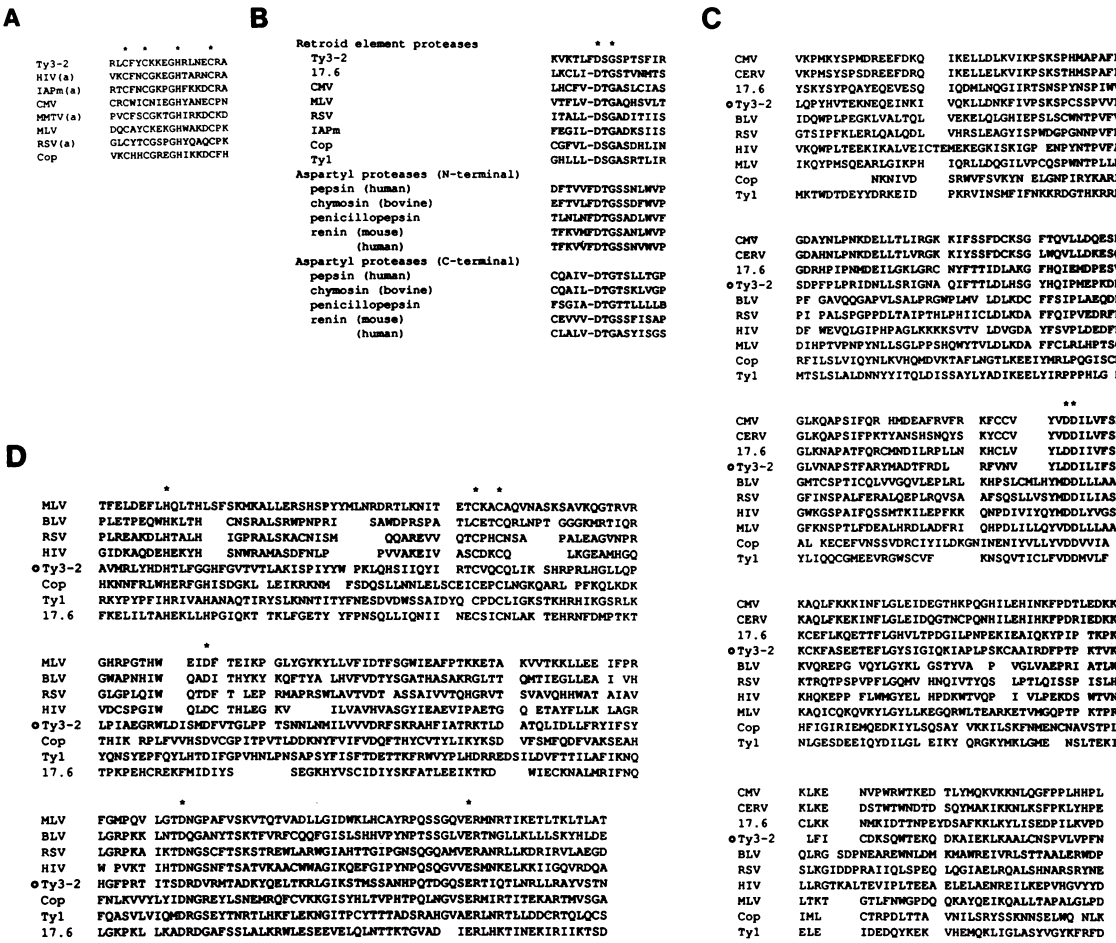


FIG. 4. Alignment of deduced amino acid sequences from Ty3 with conserved domains of retroid element proteins. (A) Nucleic acid-binding motifs of seven different retroid elements are compared with amino acids 265 to 282 predicted from *TYA3-2*. A similar alignment of non-Ty3 nucleic acid-binding sequences shown was previously published by Covey (11). The (a) indicates the first of two metal finger motifs which occur in most retroviruses. (B) Conserved residues from retroid element proteases and cellular aspartyl proteases are compared with amino acids 53 to 68 predicted from the *TYB3-2* sequence. Similar alignments of retrovirus proteases have been shown previously (46, 62, 90). Alignments of predicted retrotransposon protease sequences with aspartyl proteases were shown previously (59, 82, 83). (C) The polymerase domains of the reverse transcriptases from nine retroid elements are compared with amino acids specified by codons 348 to 632 of *TYB3-2*. Alignments of retroid and nonretroid polymerase sequences have been described previously (43, 45, 52, 59, 81, 82; R. F. Doolittle, D.-F. Feng, M. S. Johnson, and M. A. McClure, Quart. Rev. Biol., in press). (D) The endonuclease domains from seven retroid elements are compared with amino acids specified by codons 847 to 1040 of *TYB3-2*. Alignments of retroid element endonuclease sequences have also been described previously (43, 52, 59; Doolittle et al., in press). Polymerase and endonuclease sequences were displayed and printed by using the MULPUB program (23). Boldface circles indicate lines of the Ty3 sequence. Asterisks indicate residues conserved among all of the protein sequences in the comparison. The cellular protease sequences are human pepsin (76), bovine chymosin (25), penicillopepsin (40), and mouse (61) and human (38) renins. Abbreviations and sources of predicted retroid element protein sequences: HIV, human immunodeficiency virus (65); IAPm, mouse intracisternal type A particle (56); MMTV, mouse mammary tumor virus (58); MLV (75); RSV, Rous sarcoma virus (72); CMV, cauliflower mosaic virus (26); Cop, copia (59); 17.6 (68); CERV, carnation etched-ring virus (41); BLV, bovine leukemia virus (66); Ty1 (8).

quences of these proteins are not highly conserved among retroviruses, except for a small domain in the nucleocapsid protein which has been suggested to mediate nucleic acid binding (10, 11, 15, 36). This region is the carboxy-terminal portion of the *gag* polyprotein and consists of cysteine and histidine residues arranged in a $C\text{-}X_2\text{-}C\text{-}X_4\text{-}H\text{-}X_4\text{-}C$ motif. This sequence is repeated in retroviruses, with the exception of Moloney murine leukemia virus (MLV) (75), but occurs once, if at all, in retrotransposons. The sequences of Ty3-1 and Ty3-2 are identical in the downstream end of the first ORF and predict that the protein made from *TY3A* contains one copy of the nucleic acid-binding motif. The protein predicted from the copia DNA sequence also contains this short sequence, but Ty1, Ty2, and 17.6 proteins do not. A translated sequence from *TYA3* containing this motif is aligned in Fig. 4A with similar regions of other retroid elements.

Protease. The protease encoded by retroviruses is required for processing the polyprotein precursors to mature proteins (87, 91). More recently, a region in *TYB* with similarity to this protease was demonstrated to be required for processing Ty1 polyproteins (1, 60, 92). Retroid element proteases contain a highly conserved hexapeptide, $(\text{hydrophobic residue})_2\text{-D-T/S-G-A/S}$, which is also found at the active sites of aspartyl proteases (46, 62, 82, 90). Thus, the retrovirus protease is hypothesized to be an aspartyl protease

distantly related to these cellular proteases. The predicted protein sequence of *TYB3-2* contains the conserved active-site hexapeptide close to its amino terminus. This region from the Ty3-2 protein is compared with proteases from other retroid elements in Fig. 4B. In the *TYB3* protease sequence, unlike the protease sequences of most other elements, phenylalanine occurs immediately before D-S-G and serine follows it. These positions in the predicted Ty3-2 protease sequence resemble conserved positions in the cellular aspartyl proteases.

Reverse transcriptase: polymerase and RNase H. Retroviral reverse transcriptase is encoded downstream of the protease. Johnson et al. (43) showed that domains with homology to *E. coli* polymerase and RNase H proteins occur, in that order, within retroviral reverse transcriptases and are likely to be responsible for the synthetic and nucleolytic activities of reverse transcriptase. Domains with similarity to the polymerase and RNase H domains of other retroid elements are predicted in the *TYB3* protein. Figure 4C shows a comparison of the polymerase domain of the Ty3-2 reverse transcriptase with the polymerases encoded by other retroid elements. As noted above, in this region, Ty3-2 is most similar to 17.6. Within the polymerase domain, Ty3-2 and 17.6 have 43% identical residues in an alignment which requires introduction of only four gaps. The next highest similarities are with two plant DNA viruses which replicate through RNA intermediates, cauliflower mosaic virus (26) and carnation etched-ring virus (41), and with MLV (75). The polymerase domains of these retroid elements have 36, 33, and 26% identity, respectively, with the polymerase of Ty3-2. Weaker similarity is found between the Ty3-2 polymerase and those encoded by Ty1 and copia, 12 and 11%, respectively.

The protein sequence predicted from *TYB3-2* immediately downstream of the region that codes for the polymerase has similarity to conserved positions of RNase H from different retroid elements (43). Strictly conserved positions noted by that group are shaded in the *TYB3-2* protein sequence presented in Fig. 2. The tether region separating polymerase and RNase H domains in the retroviral protein sequence (43) is absent in the protein sequence predicted for Ty3. Comparison of the RNase H region from the Ty3 protein with other retroid elements showed less conservation of the protein sequence overall than did the polymerase comparison (unpublished data).

Endonuclease. The endonuclease domain is less conserved among retroid elements than the reverse transcriptase domain. Nevertheless, a protein sequence is predicted from *TYB3-2* which has distinct similarity to these endonucleases. Figure 4D shows an alignment of sequences from known retrovirus endonuclease domains with similar regions predicted from Ty3 and other retroid elements. Six residues are conserved among the eight retroid elements compared in this alignment. As noted in a similar alignment previously described by Johnson et al. (43), the occurrence in these endonuclease sequences of a pair of histidines, followed after 20 to 30 residues by a conserved pair of cysteines, is reminiscent of the metal fingers of some DNA-binding proteins. Johnson et al. (43) suggested that this domain mediates interactions between the endonuclease and substrate DNA. The occurrence of these conserved residues in the Ty3-2 protein sequence suggests that some aspects of the Ty3 insertion mechanism are in common with those of other integrating retroid elements.

De novo transposition of Ty3. Nucleotide sequence analysis showed that Ty3 has the structural properties of a retrotransposon, including a sequence which encodes a protein with similarity to reverse transcriptase. Nevertheless, it was not known whether Ty3-1 or Ty3-2 (or both) could transpose in standard strains of *S. cerevisiae*. Boeke et al. (2) demonstrated that Ty1 transposition occurred at high

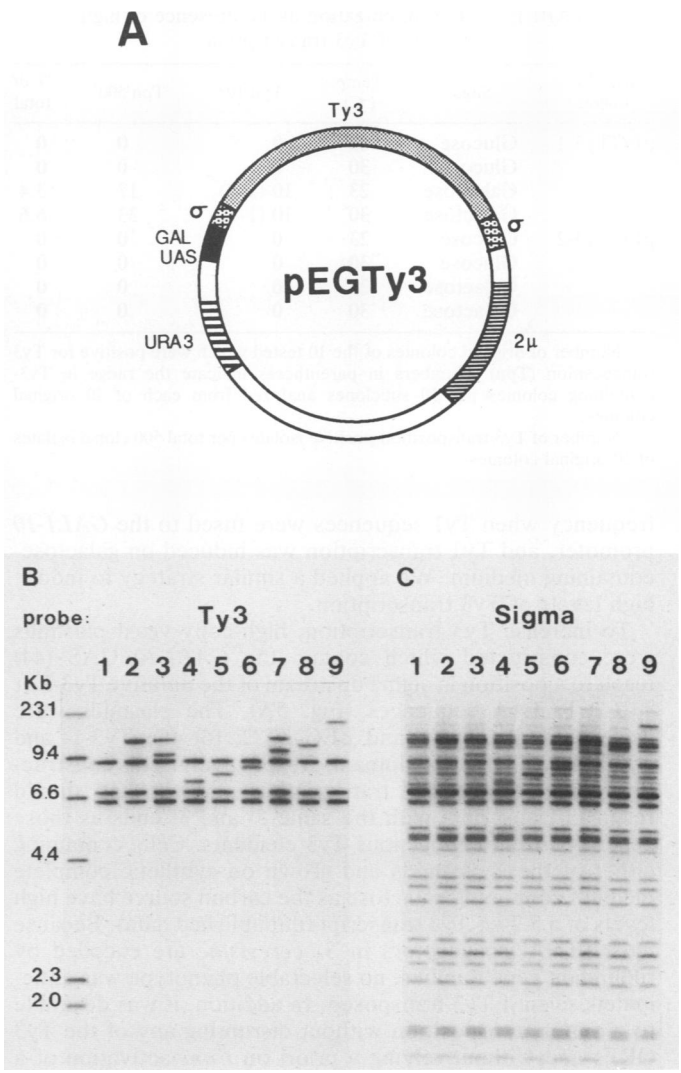


FIG. 5. Transposition of Ty3. (A) Donor plasmids containing galactose-inducible Ty3-1 and Ty3-2 sequences were constructed as described in Materials and Methods and are shown generalized as pEGTy3. The yeast sequences marked are as follows: *GAL1-10* UAS, solid; sigma elements, lines with stippling; Ty3 internal domain, dark stippling; 2μm plasmid replication sequences, thin lines; and *URA3* sequences, thick lines. *E. coli* sequences, *ori* and *AmpR*, are shown as open spaces. The scale is approximate. (B and C) Autoradiograph of Southern blot analysis of genomic DNA from yVB109 clonal isolates which underwent Ty3 rearrangements. The leftmost lane shows migration of lambda DNA digested with *Hin*dIII and ^{32}P labeled with T4 polynucleotide kinase. The numbers on the left indicate molecular sizes in kilobases (Kb). Panel B shows hybridization with a Ty3 internal domain-specific probe; panel C shows hybridization with a sigma-specific probe (see Materials and Methods). Lanes: 1, hybridization of genomic DNA from the pVB109 host strain; 2 to 9, hybridization of genomic DNAs from eight clonal isolates with Ty3 rearrangements. The samples shown in B and C were fractionated on the same agarose gel. This autoradiograph was produced by exposure of the hybridized blot to XAR-5 film overnight at room temperature.

TABLE 1. Ty3 mobilization in the presence of high levels of Ty3 transcription

GAL-Ty3 fusion	Sugar	Temp (°C)	Tpn/10[a]	Tpn/500[b]	% of total
pEGTy3-1	Glucose	23	0	0	0
	Glucose	30	0	0	0
	Galactose	23	10 (1–3)	17	3.4
	Galactose	30	10 (1–10)	33	6.6
pEGTy3-2	Glucose	23	0	0	0
	Glucose	30	0	0	0
	Galactose	23	0	0	0
	Galactose	30	0	0	0

[a] Number of original colonies of the 10 tested which were positive for Ty3 transposition (Tpn). Numbers in parentheses indicate the range in Ty3-containing colonies per 50 subclones analyzed from each of 10 original colonies.

[b] Number of Ty3 transposition-positive isolates per total 500 clonal isolates of 10 original colonies.

frequency when Ty1 sequences were fused to the *GAL1-10* promoter, and Ty1 transcription was induced on galactose-containing medium. We applied a similar strategy to induce high levels of Ty3 transcription.

To increase Ty3 transcription, high-copy yeast plasmids were constructed which contain the *GAL1-10* UAS (44) fused to a position in sigma upstream of the putative Ty3-1 or Ty3-2 TATAA sequences (Fig. 5A). The plasmids were designated pEGTy3-1 and pEGTy3-2, for the Ty3-1- and Ty3-2-derived internal domains, respectively. This construction preserves the Ty3 transcription start site and should result in transcripts with the same 5′ and 3′ ends as those generated from endogenous Ty3 elements. Cells containing either of these plasmids and grown on synthetic complete medium containing galactose as the carbon source have high levels of a 5.2-kb Ty3 transcript (unpublished data). Because most tRNA isoacceptors in *S. cerevisiae* are encoded by redundant gene families, no selectable phenotype was anticipated, even if Ty3 transposed. In addition, it was desirable to monitor transposition without disrupting any of the Ty3 ORFs and without relying a priori on *trans*-activation of a marked Ty3. Evidence for de novo Ty3 transposition was therefore sought by use of a hybridization screen rather than by selection.

To determine the effect of Ty3 transcription on Ty3-related rearrangements, yeast cells containing the inducible Ty3 plasmid pEGTy3-1 or pEGTy3-2 were maintained on synthetic complete medium minus uracil for 10 days with either glucose or galactose as the carbon source. The yeast strain used in these experiments, yVB110, has no endogenous Ty3 elements (Bilanchone, personal communication). Parallel sets of cultures were maintained at 23 and 30°C to investigate the effect of temperature on Ty3 rearrangements. Ten colonies from each experimental condition were streaked for single colonies on rich medium. Fifty clonal isolates representing each of the original 10 colonies were patched onto 5-fluoro-orotic acid-containing medium to select cells which no longer contained the plasmid. The results of this analysis are presented in Table 1. Genomic DNA from pEGTy3-1-transformed cells grown on galactose, but not on glucose, showed hybridization to Ty3. Each of the original 10 colonies grown at 30°C showed evidence of Ty3 rearrangements, with an average frequency of 6.6% over the total of 500 colonies screened. Genomic DNA from 3.4% of the pEGTy3-1-transformed colonies grown on galactose at 23°C showed Ty3 hybridization. The pEGTy3-1-transformed cells grown on glucose did not show evidence of Ty3-related rearrangements. The pEGTy3-2-transformed cells showed no evidence of Ty3 rearrangements during growth on glucose or galactose after 10 days (Table 1) or 20 days (data not shown) at either temperature. These data are consistent with the predicted dependence of Ty3 transposition on transcription but suggest that of the two elements tested only Ty3-1 is capable of independent high-frequency transposition.

The acquisition of Ty3 sequences by host genomes could also be mediated by recombination unrelated to transposition. We consider this an unlikely explanation of the results described above, because the frequency of transposition was dependent on transcription of the Ty3 sequences. Rearrangements were not detected in the glucose-grown cells in which Ty3 was not transcribed. It could be argued that this recombination was transcriptionally activated. This explanation, however, is not consistent with the absence of detectable Ty3 rearrangements in the pEGTy3-2-transformed, galactose-grown cells. Therefore the results presented in Table 1 are not easily explainable by a simple recombination mechanism.

Recombination between the plasmid and host chromosomal sequences could not alone explain the Ty3 rearrangements. Nevertheless, these data do not rule out transcription-dependent synthesis of a full-length Ty3 DNA and subsequent integration by homologous recombination of this intermediate with endogenous sigma elements. Southern blot analysis of genomic DNA from cells in which Ty3-related rearrangements occurred was performed to investigate this possibility. Figure 5B and C shows the results of that analysis. The host strain yVB109, transformed with pEGTy3-1 and grown on galactose, contains three endogenous elements (Bilanchone, personal communication). Hybridization with a Ty3-specific probe (Fig. 5B) showed that there are four or five Ty3-hybridizing fragments in each of these genomes. In the 230 colonies screened by pooled DNA preparations for this analysis, at least 35 novel bands were observed. Comparison of the Ty3 hybridization pattern with the sigma-specific hybridization pattern (Fig. 5C) showed that there are additional sigma-hybridizing fragments in each genome which acquired a Ty3 element(s) but that in no case is a sigma-hybridizing fragment present in the control genome absent in a genome which has acquired Ty3 elements. These results suggest that Ty3 rearrangements are not commonly mediated by integration of Ty3 at endogenous sigma loci. Although our experiments do not specifically demonstrate an RNA intermediate, both sequence analysis and transposition data are consistent with a retroviruslike mechanism of transposition for Ty3.

The basis of the difference in transposition activity between the pEGTy3-1- and pEGTy3-2-transformed cells is not clear. Although Ty3-1 and Ty3-2 are highly similar over long portions of the sequences we compared (unpublished data), there are some differences. These include a 78-bp repeat within *TYB3-2* and, potentially, a difference in end point between *TYB3-2* and *TYB3-1*. Alternatively, there may be important differences in the levels or structures of transcripts produced from these two plasmids. The pEGTy3-2 construct retains the tRNA gene downstream of the element. If termination of Ty3 transcription is affected by the tRNA gene, a less active Ty3 transcript may result. These possibilities can be readily distinguished by the activities of chimeric elements.

DISCUSSION

Ty3 is a yeast retrotransposon which contains coding information for proteins with similarity to previously de-

scribed retroviral nucleic acid-binding proteins, aspartyl proteases, reverse transcriptases, and endonucleases. The two Ty3 elements characterized, like previously described sigma elements, occur close to the 5′ ends of tRNA genes. One Ty3 element isolated from a common laboratory strain was shown to be capable of high-frequency transposition. The results presented here suggest that Ty3 transposition is responsible for the unusual position specificity observed for isolated sigma elements. A surprising finding of this study is that Ty3 is apparently more closely related to a family of *Drosophila* retrotransposons and to animal retroviruses than to previously described yeast retrotransposons.

Relatedness of Ty3 to Ty1 and Ty2. DNA sequences involved in the regulation of expression of Ty3, Ty1, and Ty2 have some features in common. The sigma LTRs of Ty3 and the delta LTRs of Ty1 and Ty2 have a number of short regions of identity (31). Ty1 and Ty3 transcripts have been shown to start in the LTR about 100 bp upstream of the beginning of the internal domain in the respective elements (9, 19). Transcription of Ty elements is under mating type control and is down regulated in diploids (20; V. W. Bilanchone, K. Y. Sato, and S. B. Sandmeyer, manuscript in preparation). The primer-binding site in the transcripts of all three types of elements has complementarity to initiator $tRNA^{Met}$ (9, 17, 88). As in retroviruses, the first and second ORFs of each of the Ty classes overlap. This overlap is 38, 44, and 38 nucleotides long for Ty3, Ty1, and Ty2, respectively (this work; 8, 28, 35, 88). Nine nucleotides within the overlap are conserved in all three elements. A plus-one translational frameshift within the overlap is required for expression of proteins from the second ORFs of Ty1 and Ty2 RNAs (8, 53). These similar features of Ty3, Ty1, and Ty2 could be the result of divergent evolution from some ancestral element or could result from common adaptation of cellular regulatory mechanisms to limit transposition.

In contrast to the similarities noted above, the proteins encoded by Ty3 are not highly similar to the analogous proteins encoded by Ty1 and Ty2. The DNA sequence of *TYA3* predicts a domain containing the $C\text{-}X_2\text{-}C\text{-}X_4\text{-}H\text{-}X_4\text{-}C$ motif which is also found in the nucleocapsid proteins of animal retroviruses. This motif does not occur in the *gag*-like proteins of Ty1 and Ty2 (1, 8, 35, 88). The sequences of the predicted polymerase domains and the endonuclease domains from Ty3 and Ty1 were compared by using the progressive-alignment programs of Feng and Doolittle (23). This comparison showed a minimum number of identical positions. The overall identity between Ty3 and Ty1 reverse transcriptases in the polymerase domains is 12%. The significance of the reverse transcriptase and endonuclease alignments was scored (23) to reflect positions at which equivalent, although not identical, amino acids are conserved and to reflect the number of gaps introduced with optimal alignment of the protein sequences. Despite the low percentage of identical residues, the score of the Ty3 and Ty1 reverse transcriptase alignment (Fig. 4C) was 8 standard deviations away from the mean of the scores generated by 50 comparisons between random jumbles of the same two amino acid sequences. Therefore, the similarity of these sequences is still greater than for random sequences of this composition. The comparison of endonucleases showed about the same level of identity as the comparison of reverse transcriptases but scored closer to the mean of the random jumbles. Thus, the proteins encoded in Ty3 have similarity to those encoded in Ty1 but are clearly distinct.

In addition to protein sequence differences between Ty3 and the other yeast retrotransposons, there are also organizational differences. In the predicted polyprotein sequence of Ty3, the reverse transcriptase domain precedes the endonuclease domain—the reverse of the order in the Ty1 and Ty2 polyproteins. On the basis of these comparisons, Ty3 seems distantly related, if at all, to the other yeast elements. Thus, Ty3 could have diverged long ago from a common ancestor with Ty1 and Ty2 or might have been assembled independently from cellular genes (79).

Relatedness of Ty3 and *Drosophila* retrotransposons. A comparison of the predicted *TYB3* protein sequence to the National Biomedical Research Foundation Protein Data Base showed, as mentioned above, that this sequence has the greatest similarity to the polyprotein encoded by the second ORF of 17.6 (68, 82). The Ty3 polymerase domain showed 43% sequence identity with the analogous region of the 17.6 polyprotein. The alignment score of the Ty3 and 17.6 polymerase comparison was 19.3 standard deviations above the mean score of the random-jumble comparisons. This score shows that the Ty3 polymerase resembles the 17.6 polymerase much more closely than it resembles the Ty1 polymerase. The relatively high level of similarity between the polymerases encoded by Ty3 and 17.6 argues strongly that they are homologous.

Our data are consistent with a common ancestor for Ty3 and 17.6, and the data of others suggest that Ty1 and copia also stem from a common lineage. If these pairs of elements are related, then multiple Tys might have existed before the divergence of single-celled fungi from other eucaryotes. It might also be expected that these yeast elements would have diverged for similar periods of time from their respective *Drosophila* homologs and possibly under similar selection. Why, then, are the polymerase sequences of Ty3 and 17.6 so much more similar than the polymerase sequences of Ty1 and copia? There are at least three possible explanations. (i) It is possible that the lineages by which the genera *Drosphila* and *Saccharomyces* might be related are much more complex than can be represented by the limited set of characterized elements. The present or past existence of additional elements, for instance, yeast elements more closely related to copia, could resolve the apparent discrepancies in the rates of divergence of these retrotransposons. (ii) The assumption that different retrotransposons followed through the same eucaryotic lineage of organisms should be subject to similar rates of change could be incorrect. Different functions could have evolved for the two types of elements, resulting in differences in the rates of permissible change. Alternatively, the two reverse transcriptases could have significantly different error rates, also leading to different rates of change between the lineages. (iii) The similarity between Ty3 and 17.6 could be explained by interspecies horizontal transmission of some ancestral retrotransposon sequence after the divergence of fungi from other eucaryotic lineages.

It is difficult to choose among these possibilities on the basis of the data available. Neither of the first two explanations—incomplete catalogs of retrotransposon lineages and different rates of change for the Ty1-copia and Ty3-17.6 lineages—can be easily verified. Several observations may, however, shed light on the likelihood of explanation iii, interspecies horizontal transmission of some ancestral element. Because Ty3 (sigma) and 17.6 are quite different, any interspecies transmission would presumably have occurred long ago. Although it is not known how widespread Ty3 is in different yeasts, the sigma element is found in at least two species closely related to *S. cerevisiae*, *S. norbensis* and *S. carlsbergensis* (13). Interestingly, it is known that retrotrans-

posons in different *Drosophila* species vary in both occurrence and copy number. In particular, Martin et al. (51) suggested that the widespread but variable occurrence of 412 and 297, two elements considered related to 17.6, is consistent with horizontal transmission. An appealing aspect of the hypothesis that horizontal transmission could involve completely different organisms is the existence of fruit flies and yeasts in the same habitat.

In an effort to evaluate the probability of horizontal transmission by additional criteria, the codon usage of Ty3 and 17.6 second ORFs was compared with codon usage tables generated from *Drosophila* and *Saccharomyces* DNA sequences. Our comparison, performed with the University of Wisconsin Genetics Computer Group CORRESPOND program, showed that usage in both the copia ORF and the 17.6 second ORF resembles *Saccharomyces* usage more than *Drosophila* usage (unpublished data). This codon bias is consistent with, although it does not prove, horizontal transmission.

Relatedness of Ty3 to retroviruses. Ty3 encodes proteins and has a gene organization which resembles that found in animal retroviruses. The organizational similarities have been noted above. We aligned reverse transcriptase and endonuclease sequences from human immunodeficiency virus, bovine leukemia virus, Rous sarcoma virus, and MLV with the Ty3 reverse transcriptase (polymerase) and endonuclease sequences. These retroviruses have been compared to each other previously and are relatively distantly related (7, 43, 65, 81). The MLV lineage is considered to have diverged relatively early from the lineages of other modern retroviruses. The similarity between Ty3 and animal retrovirus proteins is most clear in sequences from the polymerase domains of reverse transcriptase, in which the alignments show from 19% identity with Rous sarcoma virus to 26% identity with MLV and bovine leukemia virus (Fig. 4C). These similarities are not as compelling as the similarity between the Ty3 and 17.6 reverse transcriptases; nevertheless, the alignment scores of these comparisons were 11 to 17 standard deviations above the mean of the scores of the random-jumble comparisons. Similarity was also present, although to a lesser extent, in the endonuclease domains we compared. Among these retrovirus endonucleases, the sequence from Ty3 was least similar to human immunodeficiency virus, with 17% identity, and most similar to MLV, with 23% identity (Fig. 4D). Of the retroviral polymerase and endonuclease sequences we compared to the Ty3 proteins, similarity was greatest with those from MLV. The organizational and protein sequence similarities between Ty3 and these animal retroviruses are consistent with derivation of Ty3 and animal retroviruses from a common ancestral species.

Position specificity of Ty3. A striking feature of isolated sigma elements and the two Ty3 retrotransposon insertions which we characterized is close association with the 5′ ends of different tRNA genes. Although we did not analyze the insertion sites of the newly transposed Ty3 elements for the presence of tRNA genes, one simple explanation of the existing insertion site data is that the Ty3 element transposes with position specificity for tRNA genes. If this hypothesis is correct, then one contributor to this position specificity could be the endonuclease. The endonuclease encoded by the retroid element is presumed to be responsible for cleaving the DNA transposition intermediate to produce the mature termini of the inserted element (reviewed in reference 84). Although it is not known whether this intermediate is circular or linear, there is in vitro evidence that the avian retrovirus endonuclease pp32 binds and cleaves at the ends of the LTRs (16, 32, 57). Because the length of the target site repeats corresponds to the retroid element rather than the "host," the retroid element endonuclease is inferred to cut the genomic target, as well as the transposition intermediate (85). We examined the sequence of the *TYB3* protein in the region most similar to the sequences of retrovirus endonucleases. The *TYB3* sequence contained the conserved residues at six of seven positions which were conserved among all of the other endonucleases we compared. This suggests that Ty3 and the other retroid elements we compared integrate through fundamentally similar mechanisms.

The apparent position specificity of Ty3 is one of its most intriguing features. Site-directed mutagenesis of the endonuclease and tRNA gene target plasmids can now be used to examine the molecular basis of Ty3 integration. The similarity which exists between Ty3 and other retroid elements suggests that these studies may also offer insights into the mechanics of retrotransposition in systems in which position specificity has not been observed.

ACKNOWLEDGMENTS

We thank B. Rymond in M. Rosbash's laboratory for the construct pHZ18. We thank D.-F. Feng and R. F. Doolittle for making their progressive-alignment computer programs and formatted retroid element protein sequences available. We are grateful to M. A. McClure, M. S. Johnson, D.-F. Feng, R. F. Doolittle, and J. D. Boeke for making unpublished results available and for helpful discussions.

This work was supported by a Public Health Service grant from the National Institutes of Health (GM33281). L.J.H. and D.L.C. were supported by a predoctoral training grant from the National Institutes of Health (GM07134).

LITERATURE CITED

1. **Adams, S. E., J. Mellor, K. Gull, R. B. Sim, M. F. Tuite, S. M. Kingsman, and A. J. Kingsman.** 1987. The functions and relationships of Ty-VLP proteins in yeast reflect those of mammalian retroviral proteins. Cell **49**:111–119.
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. **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.
4. **Brodeur, G. M., S. B. Sandmeyer, and M. V. Olson.** 1983. Consistent association between sigma elements and tRNA genes in yeast. Proc. Natl. Acad. Sci. USA **80**:3292–3296.
5. **Cameron, J. R., E. Y. Loh, and R. W. Davis.** 1979. Evidence for transposition of dispersed repetitive DNA families in yeast. Cell **16**:739–751.
6. **Chaleff, D. T., and G. R. Fink.** 1980. Genetic events associated with an insertion mutation in yeast. Cell **21**:227–237.
7. **Chiu, I.-M., R. Callahan, S. R. Tronick, J. Schlom, and S. A. Aaronson.** 1984. Major *pol* gene progenitors in the evolution of oncoviruses. Science **223**:364–370.
8. **Clare, J., and P. Farabaugh.** 1985. Nucleotide sequence of a yeast Ty element: evidence for an unusual mechanism of gene expression. Proc. Natl. Acad. Sci. USA **82**:2829–2833.
9. **Clark, D. J., V. W. Bilanchone, L. J. Haywood, S. L. Dildine, and S. B. Sandmeyer.** 1988. A yeast sigma composite element, Ty3, has properties of a retrotransposon. J. Biol. Chem. **263**: 1413–1423.
10. **Copeland, T. D., M. A. Morgan, and S. Oroszlan.** 1984. Complete amino acid sequence of the basic nucleic acid binding protein of feline leukemia virus. Virology **133**:137–145.
11. **Covey, S. N.** 1986. Amino acid sequence homology in *gag* region of reverse transcribing elements and the coat protein gene of cauliflower mosaic virus. Nucleic Acids Res. **14**:623–633.

12. **Craigen, W. J., and C. T. Caskey.** 1987. Translational frameshifting: where will it stop? Cell **50**:1–2.
13. **Del Rey, F. J., T. F. Donahue, and G. R. Fink.** 1982. Sigma, a repetitive element found adjacent to tRNA genes of yeast. Proc. Natl. Acad. Sci. USA **79**:4138–4142.
14. **Devereux, J., P. Haeberli, and O. Smithies.** 1984. A comprehensive set of sequence programs for the VAX. Nucleic Acids Res. **12**:387–395.
15. **Dickson, C., R. Eisenman, and H. Fan.** 1985. Protein biosynthesis and assembly, p. 135–145. *In* R. Weiss, N. Teich, H. Varmus, and J. Coffin, (ed.), RNA tumor viruses. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
16. **Duyk, G., M. Longiaru, D. Cobrinik, R. Kowal, P. deHaseth, A. M. Skalka, and J. Leis.** 1985. Circles with two tandem long terminal repeats are specifically cleaved by *pol* gene-associated endonuclease from avian sarcoma and leukosis viruses: nucleotide sequences required for site-specific cleavage. J. Virol. **56**: 589–599.
17. **Eibel, H., J. Gafner, A. Stotz, and P. Philippsen.** 1980. Characterization of the yeast mobile element Ty1. Cold Spring Harbor Symp. Quant. Biol. **45**:609–617.
18. **Eibel, H., and P. Philippsen.** 1984. Preferential integration of yeast transposable element Ty into a promoter region. Nature (London) **307**:386–388.
19. **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.
20. **Elder, R. T., T. P. St. John, D. T. Stinchcomb, and R. W. Davis.** 1980. Studies on the transposable element Ty1 of yeast. I. RNA homologous to Ty1. Cold Spring Harbor Symp. Quant. Biol. **45**: 581–584.
21. **Errede, B., T. S. Cardillo, and F. Sherman.** 1980. Mating signals control expression of mutations resulting from insertion of a transposable repetitive element adjacent to diverse yeast genes. Cell **25**:427–436.
22. **Farabaugh, P. J., and G. R. Fink.** 1980. Insertion of the eukaryotic transposable element Ty1 creates a 5-bp duplication. Nature (London) **286**:352–356.
23. **Feng, D.-F., and R. F. Doolittle.** 1987. Progressive sequence alignment as a prerequisite to correct phylogenetic trees. J. Mol. Evol. **25**:351–360.
24. **Fink, G. R., J. D. Boeke, and D. J. Garfinkel.** 1986. The mechanism and consequences of retrotransposition. Trends Genet. **2**:118–123.
25. **Foltmann, B., V. B. Pedersen, H. Jacobsen, D. Kauffman, and G. Wybrandt.** 1977. The complete amino acid sequence of prochymosin. Proc. Natl. Acad. Sci. USA **74**:2321–2324.
26. **Franck, A., H. Guilley, G. Jonard, K. Richards, and L. Hirth.** 1980. Nucleotide sequence of cauliflower mosaic virus DNA. Cell **21**:285–294.
27. **Fuetterer, J., and T. Hohn.** 1987. Involvement of nucleocapsids in reverse transcription: a general phenomenon? Trends Biochem Res. **12**:92–95.
28. **Fulton, A. M., J. Mellor, M. J. Dobson, J. Chester, J. R. Warmington, K. J. Indge, S. G. Oliver, P. de la Paz, W. Wilson, A. J. Kingsman, and S. M. Kingsman.** 1985. Variants within the yeast Ty sequence family encode a class of structurally conserved proteins. Nucleic Acids Res. **13**:4097–4112.
29. **Gafner, J., and P. Philippsen.** 1980. The yeast transposon Ty1 generates duplications of target DNA on insertion. Nature (London) **286**:414–418.
30. **Garfinkel, D. J., J. D. Boeke, and G. R. Fink.** 1985. Ty element transposition: reverse transcriptase and virus-like particles. Cell **42**:507–517.
31. **Genbauffe, F. S., G. E. Chisholm, and T. G. Cooper.** 1984. Tau, sigma and delta: a family of repeated elements in yeast. J. Biol. Chem. **259**:10518–10525.
32. **Grandgenett, D. P., A. C. Vora, R. Swanstrom, and J. C. Olsen.** 1986. Nuclease mechanism of the avian retrovirus pp32 endonuclease. J. Virol. **58**:970–974.
33. **Grunstein, M., and D. S. Hogness.** 1975. Colony hybridization: a method for the isolation of cloned DNAs that contain a specific gene. Proc. Natl. Acad. Sci. USA **72**:3961–3965.
34. **Haltiner, M., T. Kempe, and R. Tjian.** 1985. A novel strategy for constructing clustered point mutations. Nucleic Acids Res. **13**: 1015–1025.
35. **Hauber, J., P. Nelböck-Hochstetter, and H. Feldmann.** 1985. Nucleotide sequence and characteristics of a Ty element from yeast. Nucleic Acids Res. **13**:2745–2758.
36. **Henderson, L. E., T. D. Copeland, R. C. Sowder, G. W. Smythers, and S. Oroszlan.** 1981. Primary structure of the low molecular weight nucleic acid-binding proteins of murine leukemia viruses. J. Biol. Chem. **256**:8400–8403.
37. **Henikoff, S.** 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene **28**:351–359.
38. **Hobart, P. M., M. Fogliano, B. A. O'Connor, I. M. Schaefer, and J. M. Chirgwin.** 1984. Human renin gene: structure and sequence analysis. Proc. Natl. Acad. Sci. USA **81**:5026–5030.
39. **Holmes, D. S., and M. Quigley.** 1981. A rapid boiling method for the preparation of bacterial plasmids. Anal. Biochem. **114**:193–197.
40. **Hsu, I.-N., L. T. J. Delbaere, M. N. G. James, and T. Hofmann.** 1977. Penicillopepsin from *Penicillium janthinellum* crystal structure at 2.8 Å and sequence homology with porcine pepsin. Nature (London) **266**:140–145.
41. **Hull, R., J. Sadler, and M. Longstaff.** 1986. The sequence of carnation etched ring virus DNA: comparison with cauliflower mosaic virus and retroviruses. EMBO J. **5**:3083–3090.
42. **Ito, H., Y. Fukuda, K. Murata, and A. Kimura.** 1983. Transformation of intact yeast cells treated with alkali cations. J. Bacteriol. **153**:163–168.
43. **Johnson, M. S., M. A. McClure, D.-F. Feng, J. Gray, and R. F. Doolittle.** 1986. Computer analysis of retroviral *pol* genes: assignment of enzymatic functions to specific sequences and homologies with non-viral enzymes. Proc. Natl. Acad. Sci. USA **83**:7648–7652.
44. **Johnston, M., and R. W. Davis.** 1984. Sequences that regulate the divergent *GAL1-GAL10* promoter in *Saccharomyces cerevisiae*. Mol. Cell. Biol. **4**:1440–1448.
45. **Kamer, G., and P. Argos.** 1984. Primary structural comparison of RNA-dependent polymerases from plant, animal, and bacterial viruses. Nucleic Acids Res. **12**:7269–7282.
46. **Katoh, I., T. Yasunaga, Y. Ikawa, and Y. Yoshinaka.** 1987. Inhibition of retroviral protease activity by an aspartyl proteinase inhibitor. Nature (London) **329**:654–656.
47. **Kingsman, A. J., R. L. Gimlich, L. Clarke, A. C. Chinault, and J. Carbon.** 1981. Sequence variation in dispersed repetitive sequences in *S. cerevisiae*. J. Mol. Biol. **145**:619–632.
48. **Kozak, M.** 1984. Compilation and analysis of sequences upstream from the translational start size in eukaryotic mRNAs. Nucleic Acids Res. **12**:857–872.
49. **Kronstad, J. W., J. A. Holly, and V. L. MacKay.** 1987. A yeast operator overlaps an upstream activation site. Cell **50**:369–377.
50. **Langford, C. J., and D. Gallwitz.** 1983. Evidence for an intron-contained sequence required for the splicing of yeast RNA polymerase II transcripts. Cell **33**:519–527.
51. **Martin, G., D. Wiernasz, and P. Schedl.** 1983. Evolution of *Drosophila* repetitive-dispersed DNA. J. Mol. Evol. **19**:203–213.
52. **McClure, M. A., M. S. Johnson, D.-F. Feng, and R. F. Doolittle.** 1988. Sequence comparisons of retroviral proteins: relative rates of change and general phylogeny. Proc. Natl. Acad. Sci. USA **85**:2469–2473.
53. **Mellor, J., S. M. Fulton, M. J. Dobson, W. Wilson, S. M. Kingsman, and A. J. Kingsman.** 1985. A retrovirus-like strategy for expression of a fusion protein encoded by yeast transposon Ty1. Nature (London) **313**:243–246.
54. **Mellor, J., A. J. Kingsman, and S. M. Kingsman.** 1986. Ty, an endogenous retrovirus of yeast? Yeast **2**:145–152.
55. **Mellor, J., M. H. Malim, K. Gull, M. F. Tuite, S. M. 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.
56. **Mietz, J. A., Z. Grossman, K. K. Lueders, and E. L. Kuff.** 1987.

Nucleotide sequence of a complete mouse intracisternal A-particle genome: relationship to known aspects of particle assembly and function. J. Virol. **61**:3020–3029.

57. **Misra, T. K., D. P. Grandgenett, and J. T. Parsons.** 1982. Avian retrovirus pp32 DNA-binding protein. I. Recognition of specific sequences on retrovirus DNA terminal repeats. J. Virol. **44**:330–343.
58. **Moore, R., M. Dixon, R. Smith, G. Peters, and C. Dickson.** 1987. Complete nucleotide sequence of a milk-transmitted mouse mammary tumor virus: two frameshift suppression events are required for translation of *gag* and *pol*. J. Virol. **61**:480–490.
59. **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.
60. **Müller, F., K.-H. Brühl, K. Freidel, K. V. Kowallik, and M. Ciriacy.** 1987. Processing of Ty1 proteins and formation of Ty1 virus-like particles in *Saccharomyces cerevisiae*. Mol. Gen. Genet. **207**:421–429.
61. **Panthier, J.-J., S. Foote, B. Chambraud, A. D. Strosberg, P. Corvol, and F. Rougeon.** 1982. Complete amino acid sequence and maturation of the mouse submaxillary gland renin precursor. Nature (London) **298**:90–92.
62. **Pearl, L. H., and W. R. Taylor.** 1987. A structural model for the retroviral proteases. Nature (London) **329**:351–354.
63. **Pikielny, C. W., J. L. Teem, and M. Rosbash.** 1983. Evidence for the biochemical role of an internal sequence in yeast nuclear mRNA introns: implications for U1 RNA and metazoan mRNA splicing. Cell **34**:395–403.
64. **Pixa, G., G. Dirheimer, and G. Keith.** 1984. Sequence of $tRNA^{Ile}$ from brewer's yeast. Biochem. Biophys. Res. Commun. **119**:905–912.
65. **Ratner, L., W. Haseltine, R. Patarca, K. J. Livak, B. Starcich, S. F. Josephs, E. R. Doran, J. A. Rafalski, E. A. Whitehorn, K. Baumeister, L. Ivanoff, S. R. Petteway, Jr., M. L. Pearson, J. A. Lautenberger, T. S. Papas, J. Ghrayeb, N. T. Chang, R. C. Gallo, and F. Wong-Staal.** 1985. Complete nucleotide sequence of the AIDS virus, HTLV-III. Nature (London) **313**:277–284.
66. **Rice, N. R., R. M. Stephens, A. Burny, and R. V. Gilden.** 1985. The *gag* and *pol* genes of bovine leukemia virus: nucleotide sequence and analysis. Virology **142**:357–377.
67. **Roeder, G. S., and G. R. Fink.** 1980. DNA rearrangements associated with a transposable element in yeast. Cell **21**:239–249.
68. **Saigo, K., W. Kugimiya, Y. Matsuo, S. Inouye, K. Yoshioka, and S. Yuki.** 1984. Identification of the coding sequence for a reverse transcriptase-like enzyme in a transposable genetic element in *D. melanogaster*. Nature (London) **312**:659–661.
69. **Sandmeyer, S. B., V. W. Bilanchone, D. J. Clark, P. Morcos, G. F. Carle, and G. M. Brodeur.** 1988. Sigma elements are position-specific for many different yeast tRNA genes. Nucleic Acids Res. **16**:1499–1515.
70. **Sandmeyer, S. B., and M. V. Olson.** 1982. Insertion of a repetitive element at the same position in the 5′-flanking regions of two dissimilar tRNA genes. Proc. Natl. Acad. Sci. USA **79**: 7674–7678.
71. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
72. **Schwartz, D. E., R. Tizard, and W. Gilbert.** 1983. Nucleotide sequence of Rous sarcoma virus. Cell **32**:853–869.
73. **Sherman, F., G. R. Fink, and J. B. Hicks.** 1986. Methods in yeast genetics. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
74. **Shiba, T., and K. Saigo.** 1983. Retrovirus-like particles containing RNA homologous to the transposable element copia in *D. melanogaster*. Nature (London) **302**:119–124.
75. **Shinnick, T. M., R. A. Lerner, and J. G. Sutcliffe.** 1981. Nucleotide sequence of Moloney murine leukemia virus. Nature (London) **293**:543–548.
76. **Sogawa, K., Y. Fujii-Kuriyama, Y. Mizukami, Y. Ichihara, and K. Takahashi.** 1983. Primary structure of human pepsinogen gene. J. Biol. Chem. **258**:5306–5311.
77. **Teem, J. L., N. Abovich, N. F. Kaufer, W. F. Schwindinger, J. R. Warner, A. Levy, J. Woolford, R. J. Leer, M. M. van Raamsdonk-Duin, W. H. Mager, R. J. Planta, L. Schultz, L. Friesen, H. Fried, and M. Rosbash.** 1984. A comparison of yeast ribosomal protein gene DNA sequences. Nucleic Acids Res. **12**: 8295–8312.
78. **Teem, J. L., and M. Rosbash.** 1983. Expression of a β-galactosidase gene containing the ribosomal protein 51 intron is sensitive to the *rna2* mutation of yeast. Proc. Natl. Acad. Sci. USA **80**:4403–4407.
79. **Temin, H. M.** 1970. Malignant transformation of cells by viruses. Perspect. Biol. Med. **14**:11–26.
80. **Temin, H. M.** 1982. Function of the retrovirus long terminal repeat. Cell **28**:3–5.
81. **Toh, H., H. Hayashida, and T. Miyata.** 1983. Sequence homology between retroviral reverse transcriptase and putative polymerases of hepatitis B virus and cauliflower mosaic virus. Nature (London) **305**:827–829.
82. **Toh, H., R. Kikuno, H. Hayashida, T. Miyata, W. Kugimiya, S. Inouye, S. Yuki, and K. Saigo.** 1985. Close structural resemblance between putative polymerase of a *Drosophila* transposable genetic element 17.6 and *pol* gene product of Moloney murine leukemia virus. EMBO J. **4**:1267–1272.
83. **Toh, H., M. Ono, K. Saigo, and T. Miyata.** 1985. Retroviral protease-like sequence in the yeast transposon Ty1. Nature (London) **315**:691.
84. **Varmus, H., and R. Swanstrom.** 1984. Replication of retroviruses, p. 369–512. *In* R. Weiss, N. Teich, H. Varmus, and J. Coffin (ed.), RNA tumor viruses, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
85. **Varmus, H. E.** 1983. Retroviruses, p. 411–503. *In* Shapiro, J. A., (ed.), Mobile genetic elements. Academic Press, Inc., New York.
86. **Venegas, A., E. Gonzalez, P. Bull, and P. Valenzuela.** 1982. Isolation and structure of a yeast initiator $tRNA^{Met}$ gene. Nucleic Acids Res. **10**:1093–1096.
87. **Von der Helm, K.** 1977. Cleavage of Rous sarcoma viral polypeptide precursor into internal structural proteins *in vitro* involves viral protein p15. Proc. Natl. Acad. Sci. USA **74**:911–915.
88. **Warmington, J. R., R. B. Waring, C. S. Newlon, K. J. Indge, and S. G. Oliver.** 1985. Nucleotide sequence characterization of Ty1-17, a class II transposon from yeast. Nucleic Acids Res. **13**: 6679–6693.
89. **Williamson, V. M., E. T. Young, and M. Ciriacy.** 1981. Transposable elements associated with constitutive expression of yeast alcohol dehydrogenase II. Cell **23**:605–614.
90. **Yasunaga, T., N. Sagata, and Y. Ikawa.** 1986. Protease gene structure and *env* gene variability of the AIDS virus. FEBS Lett. **199**:145–150.
91. **Yoshinaka, Y., and R. B. Luftig.** 1977. Properties of a P70 proteolytic factor of murine leukemia viruses. Cell **12**:709–719.
92. **Youngren, S. D., J. D. Boeke, N. J. Sanders, and D. J. Garfinkel.** 1988. Functional organization of the retrotransposon Ty from *Saccharomyces cerevisiae*: Ty protease is required for transposition. Mol. Cell. Biol. **8**:1421–1431.