JOURNAL OF VIROLOGY, May 1997, p. 4005–4015 Vol. 71, No. 5
0022-538X/97/$04.00+0

Retrovirus-Like End Processing of the Tobacco Tnt1 Retrotransposon Linear Intermediates of Replication

FRANK FEUERBACH,[1] JAN DROUAUD,[2] AND HÉLÈNE LUCAS[1]*

Laboratoire de Biologie Cellulaire[1] and Laboratoire de Génétique et Amélioration des Plantes,[2] Institut National de la Recherche Agronomique, 78026 Versailles Cedex, France

Received 6 September 1996/Accepted 13 February 1997

The tobacco retrotransposon Tnt1 can transpose through an RNA intermediate in the heterologous host *Arabidopsis thaliana*. We report here the identification and characterization of extrachromosomal linear and circular DNA forms of Tnt1 in this heterologous host. Our results demonstrate that Tnt1 linear intermediates possess two extra base pairs at each end compared with Tnt1's integrated forms. Prior to integration into the host genome, the two terminal nucleotides at the 3′ end of these linear intermediates are removed, as in the case of the yeast Ty3 retrotransposon and of retroviruses. Our data, together with those from recent studies of Ty3, reinforce the idea that 3′ dinucleotide cleavage is not restricted to retroviral integrases and is probably a feature shared by many different retrotransposons' enzymes.

Reverse transcription of retroviruses and long terminal repeat (LTR) retrotransposons is a complex process generating a full-length double-stranded DNA copy from a transcript that initiates and ends within the LTRs of the element. Following reverse transcription, the linear double-stranded molecules are transported to the nucleus, where they are integrated into the host genome by the element-encoded integrase (IN) (5, 20). Circular molecules generated after ligation of linear molecules (two-LTR head-to-tail circles) and circular molecules derived from recombination between LTRs (one-LTR circles) or from autointegration events are also found in the nucleus (53) and are generally considered to be dead-end by-products of reverse transcription.

Reverse transcription is initiated near the 5′ end of the element at the primer binding site (PBS), generally by using a host-encoded tRNA or a tRNA degradation product as a primer. A short DNA molecule termed the minus-strand strong-stop DNA is generated, the 3′ extremity of which corresponds to the R region located at the 5′ end of the mRNA. Due to its complementarity to the R region at the 3′ end of the mRNA, this DNA fragment is subsequently transferred to the 3′ end of another mRNA molecule, where it is elongated for extensive minus-strand DNA synthesis. Degradation of the genomic RNA associated with the newly synthesized DNA by the RNase H activity of the reverse transcriptase creates a polypurine tract (PPT) oligoribonucleotide that serves as the plus-strand primer.

In most retroviruses, the PBS and the PPT are separated from the 5′ and 3′ LTRs by 2 bp. The 5′ end of the minus-strand strong-stop DNA primed from the tRNA begins with the two nucleotides located between the PBS and the 5′ LTR. After reverse transcription, these two nucleotides are found at the 3′ end of the element. Similarly, priming of the plus strand at the PPT results in addition of two nucleotides at its 5′ end that are copied upon minus-strand completion. Thus, extrachromosomal linear forms of retroviruses possess two extra base pairs at each extremity compared with proviruses (53). In vivo studies of the Moloney murine leukemia virus have shown that prior to integration, the terminal two bases from each 3′ end of these blunt-ended molecules are removed, generating linear intermediates of replication characterized by recessed 3′ termini. Cleavage of the two bases is strictly dependent on IN function (5, 47). The cleavage activity of IN was further demonstrated in vitro with purified IN from avian myeloblastosis virus (30, 55), Moloney murine leukemia virus (11), and human immunodeficiency virus type 1 (HIV-1) (6, 7). A subsequent DNA strand transfer reaction, also catalyzed by IN, joins the recessed 3′ ends to the 5′ ends of the target DNA at the site of integration. Repair of the single-strand junction between the viral and target DNA, presumably mediated by host enzymes, completes the integration process by joining the 3′ end of the target to the 5′ end of the viral DNA (for reviews of retrovirus integration, see references 23 and 53). The ability of Moloney murine leukemia virus IN in vitro to use preprocessed substrates for the strand transfer reaction (11) and the accumulation of molecules with 3′ recessed ends during the course of infection (47) indicate that 3′ end processing and strand transfer are not concerted events.

In the primarily described yeast Ty1 and *Drosophila melanogaster copia* retrotransposons, the PBS and the PPT are adjacent to the LTRs (10, 13, 33). The termini of their extrachromosomal linear intermediates are therefore expected to be identical to those of the integrated elements. IN has been shown to be essential for transposition of the yeast Ty1 retrotransposon (3, 4). Blunt-ended substrates mimicking those of linear full-length Ty1 reverse transcripts are efficiently integrated when coincubated with virus-like particles as a source of IN (14) or with purified IN (41). Moreover, purified recombinant Ty1 IN does not mediate 3′ dinucleotide cleavage of Ty1 termini (42). These results led to the assumption that 3′ dinucleotide cleavage was only catalyzed by retroviral integrases. However, recent studies of the Ty3 retrotransposon of *Saccharomyces cerevisiae* provided the demonstration that 3′ dinucleotide cleavage is not restricted to retroviral IN enzymes and is indeed performed by Ty3 IN to remove two bases at each 3′ end of the linear intermediates of this element (35).

The *Nicotiana tabacum* genome contains hundreds of copies of a transposable element, Tnt1, which displays structural and functional similarities to retroviruses (26, 38). Sequence analysis of Tnt1 indicates that the putative PBS of this retrotransposon, showing complementarity to the 3′ acceptor stem of initiator methionine tRNAs from *Arabidopsis thaliana* (1), wheat germ, bean, and soybean (8, 21, 44), is separated from

* Corresponding author. Phone: (33).1.30833171. Fax: (33).1.30833099. E-mail: lucas@versailles.inra.fr.

the 5′ LTR by the dinucleotide AC and that a polypurine tract is found 1 bp upstream of Tnt1 3′ LTR (24). Thus, extrachromosomal linear DNAs originating from reverse transcription of Tnt1 are expected to have 2 bp downstream of their 3′ LTR and 1 bp upstream of their 5′ LTR. By analogy with retroviruses and Ty3, these extra base pairs would require terminal processing of Tnt1 linear intermediates of replication, a reaction anticipated to be catalyzed by Tnt1-encoded IN. RNA-mediated transposition of Tnt1 in the heterologous host *A. thaliana* has been described in a previous paper (38). In the present study, we analyze linear and circular DNA double-stranded molecules of Tnt1 in *Arabidopsis* and determine the end sequence of Tnt1 reverse transcription products by PCR-based approaches. Our results indicate that the 3′ ends of these replication intermediates possess extra nucleotides at their ends which are processed prior to integration in the host genome, as in the case of Ty3 and retroviruses.

MATERIALS AND METHODS

Plant material. A15 and A5 transgenic plants were obtained after vacuum infiltration of *A. thaliana* plantlets (ecotype Columbia) (2) with a suspension of *Agrobacterium tumefaciens* C58C1 (pGV2260) containing the Tnk23 binary vector previously described (38). Southern blot analysis of the A15-1 and A5-17 plants, derived by self-pollination from A15 and A5, indicated that they contain, respectively, one and three tandemly arranged copies of the Tnk23 binary vector T-DNA. 2S14-6-10 was derived by self-pollination from the previously described *A. thaliana* 1S14-6 transgenic plant (38) and contains one copy of the Tnk23 binary vector T-DNA and one transposed copy of Tnt1 in the homozygous state. Plants were grown in a growth chamber (Strader Honeywell) with a photoperiod of 8 h at 20°C and 225 microeinsteins/m^2/s. Leaves were harvested from the outside of 3-month-old rosettes (aging leaves). The 2S14-6-10 progeny was sprayed with 10 mM $CuCl_2$ to induce Tnt1 transcription as described by Mhiri et al. (40).

DNA extraction and Southern blot hybridization. Genomic DNA of *Arabidopsis* leaves was extracted according to the protocol described by Doyle and Doyle (12). DNAs (1 to 3 μg) were digested with restriction enzymes (Gibco BRL), fractionated on a 0.8% agarose gel by electrophoresis in 0.04 M Tris-acetate–0.001 M EDTA buffer, and transferred to a Hybond N membrane (Amersham) according to the manufacturer's instructions. The probes were labelled with [α-^{32}P]dCTP (Dupont NEN research products) as described by Feinberg and Vogelstein (15). Hybridization was carried out overnight at 42°C in a mixture of 50% formamide, 6× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate), 0.5% sodium dodecyl sulfate, 5× Denhardt's solution, and 100 μg of sonicated salmon sperm DNA per ml. All filters were washed to a stringency of 0.1× SSC–0.1% sodium dodecyl sulfate at 65°C.

Amplification of Tnt1 circle junctions. Tnt1 circle junctions were amplified by PCR with oligonucleotides located, respectively, in the U3 and U5 regions of Tnt1 LTR and directed outwards of the element. Two independent reactions were performed with the previously described Tnt77 and Tnt540 oligonucleotides (38) or with the tab2-3′ (5′-AAGCTTCACCCTCTAAAG-3′) and Tnt540 ones. PCRs were carried out in a 50-μl volume of 1.5 mM $MgCl_2$ buffer containing 0.2 mM (each) deoxynucleoside triphosphates, 20 pmol of each primer, and 0.5 U of *Taq* polymerase (United States Biochemical). Amplification was achieved with a cycle of 1 min of denaturation at 94°C, 1 min of annealing at 60°C (Tnt77) or 50°C (tab2-3′), and 1 min of polymerase extension at 72°C repeated 35 times, with a final 15-min extension at 72°C in a Perkin-Elmer Cetus DNA thermal cycler, PEC480.

Recovery of termini of Tnt1 extrachromosomal linear forms. The termini of Tnt1 extrachromosomal linear forms were amplified by ligation-mediated PCR after modification of the protocol described by Kere et al. (31).

(i) Preparation of the asymmetric adaptors. The oligonucleotide LS (5′-GC GGTGACCCGGGAGATCTGAATTC-3′) was mixed to equimolarity with oligonucleotide SS1 (5′-GAATTCAGATC-3′), SS2 (5′-ATGAATTCAGATC-3′), or SS3 (5′-ACGAATTCAGATC-3′). The mixtures were boiled for 5 min and left for 20 min at room temperature to allow renaturation of the corresponding asymmetric adaptors AD1, AD2, and AD3. The adaptor extremity corresponding to the LS 3′ end is blunt ended in AD1, while it is characterized in AD2 and AD3, respectively, by a 5′-AT-3′ dinucleotide and a 5′-AC-3′ protruding dinucleotide.

(ii) Ligation of the adaptor with genomic DNA. Twenty picomoles of the adaptors was ligated overnight at 16°C, in a total volume of 20 μl, with 500 ng of undigested genomic DNA or with 10 ng of a Tnt1-containing plasmid (pBSX1) digested by *Hin*cII or *Eco*RV. AD1 ligation was performed with 6 Weiss units of T4 DNA ligase (Pharmacia) in the presence of 0.1 mM ATP. AD2 and AD3 ligations were carried out with 10 Weiss units of *Escherichia coli* DNA ligase (United States Biochemical) in the presence of 0.1 mM β-NAD. Enzymes were heat inactivated for 10 min at 75°C, and the reaction was ethanol precipitated. The pellet was resuspended in 10 μl of distilled water, and 2 μl was used for PCR amplification.

(iii) PCR amplification. PCRs were carried out as described above with a cycle of 1 min of denaturation at 94°C, 1 min of annealing at 50°C, and 2 min of polymerase extension at 72°C repeated 35 times. The 5′ end of Tnt1 linear intermediates was selectively amplified by a two-step PCR with nested primers. After ligation with AD1, the first reaction was performed with the LS oligonucleotide in combination with the Cf5 oligonucleotide (5′-AGATCTCTCATCC TTCTTTGCC-3′) located at the beginning of the Tnt1 open reading frame. The reaction mixture was diluted 50 times, and 2 μl was used for a second amplification with the LS oligonucleotide and the Tnt77 oligonucleotide. After ligation with AD2 or AD3, the 5′ end of the Tnt1 linear intermediates was amplified in the same way, except oligonucleotides LS and tab2-3′ were used for the second amplification. Selective amplification of the 3′ end of Tnt1 extrachromosomal linear forms was achieved in the first PCR by using the LS oligonucleotide with the Avi oligonucleotide (5′-TATGCTGACCAAGGTGGTAC-3′) located at the end of the Tnt1 open reading frame. In the second PCR, oligonucleotides LS and Tnt540 or LS and Tnt79$^+$ (5′-TGACTTGGTTTGGTTGGTAGCC-3′) were used when DNA was ligated with AD1 or AD2 and AD3, respectively.

Cloning of PCR products and sequencing reaction. The number of blunt-ended PCR fragments was increased after consecutive treatments with T4 DNA polymerase and *E. coli* DNA polymerase I Klenow fragment. The amplified fragments were subsequently gel purified with a Spin-X column (Costar, Cambridge, Mass.) and cloned into the *Sma*I or *Eco*RV site of pBluescript II SK$^-$ (Stratagene). The resulting plasmids were electroporated into *E. coli* XL1 cells (Stratagene) with a Gene Pulser (Bio-Rad) according to the manufacturer's instructions. The DNA sequence of recombinant clones was determined as described before (38).

RESULTS

Linear and circular intermediates of Tnt1 are present in transgenic *Arabidopsis* leaves. Previous studies have indicated that Tnt1 is transcribed in senescing leaves of *A. thaliana* (43). We therefore looked for reverse-transcribed unintegrated copies of Tnt1 in aging leaves of *A. thaliana* transgenic plants containing Tnt1.

Total DNA was extracted from fully developed leaves sampled from the outside of the rosette of 15 3-month-old individual plants derived by self-pollination from the A15 and A5 primary transformants. Undigested DNAs from A15-1 and A5-17 plants, after electrophoresis and transfer onto a nylon membrane, were probed with an *Ssp*I fragment from Tnt1 LTR corresponding to the R and U5 regions. A fragment with a size of ca. 5.3 kbp was present in the two DNA samples (Fig. 1A), suggesting that they contained linear reverse-transcribed copies of Tnt1. Further analysis was performed with *Hin*cII-digested DNAs from individual progenies of A15 and A5 plants. DNA samples from only three plants are shown in Fig. 1B, since they were all identical. The expected two T-DNA fragments with sizes of 1.5 kbp and 600 bp hybridized with the probe in all samples. A fainter fragment with a size of ca. 500 bp was visible in the 15 A15 progenies tested as well as in the 9 A5 progenies analyzed (Fig. 1B and data not shown). This fragment corresponds to the 530-bp 3′ LTR *Hin*cII fragment expected to arise after digestion of double-stranded linear transposition intermediates of Tnt1, indicating that all plants analyzed indeed contained linear copies of Tnt1 and demonstrating that this element is reverse transcribed in aging leaves of *A. thaliana*. When genomic DNAs from the same A15 progenies were digested with *Sac*I, the R-U5 probe lit up two major fragments with sizes of ca. 2.5 and 11 kbp corresponding to the T-DNA and flanking genomic sequences (Fig. 1C). Three fainter fragments with sizes of ca. 1.3, 4, and 4.7 kbp were also detected. The two smaller fragments correspond, respectively, to the 1,340-bp 5′ LTR-containing fragment and to the 3,954-bp 3′ LTR-containing fragment generated after digestion of linear double-stranded Tnt1 elements by *Sac*I. The 4.7-kbp fragment could originate from linearization of one-LTR circles by the *Sac*I enzyme, since it was also detected after hybridization

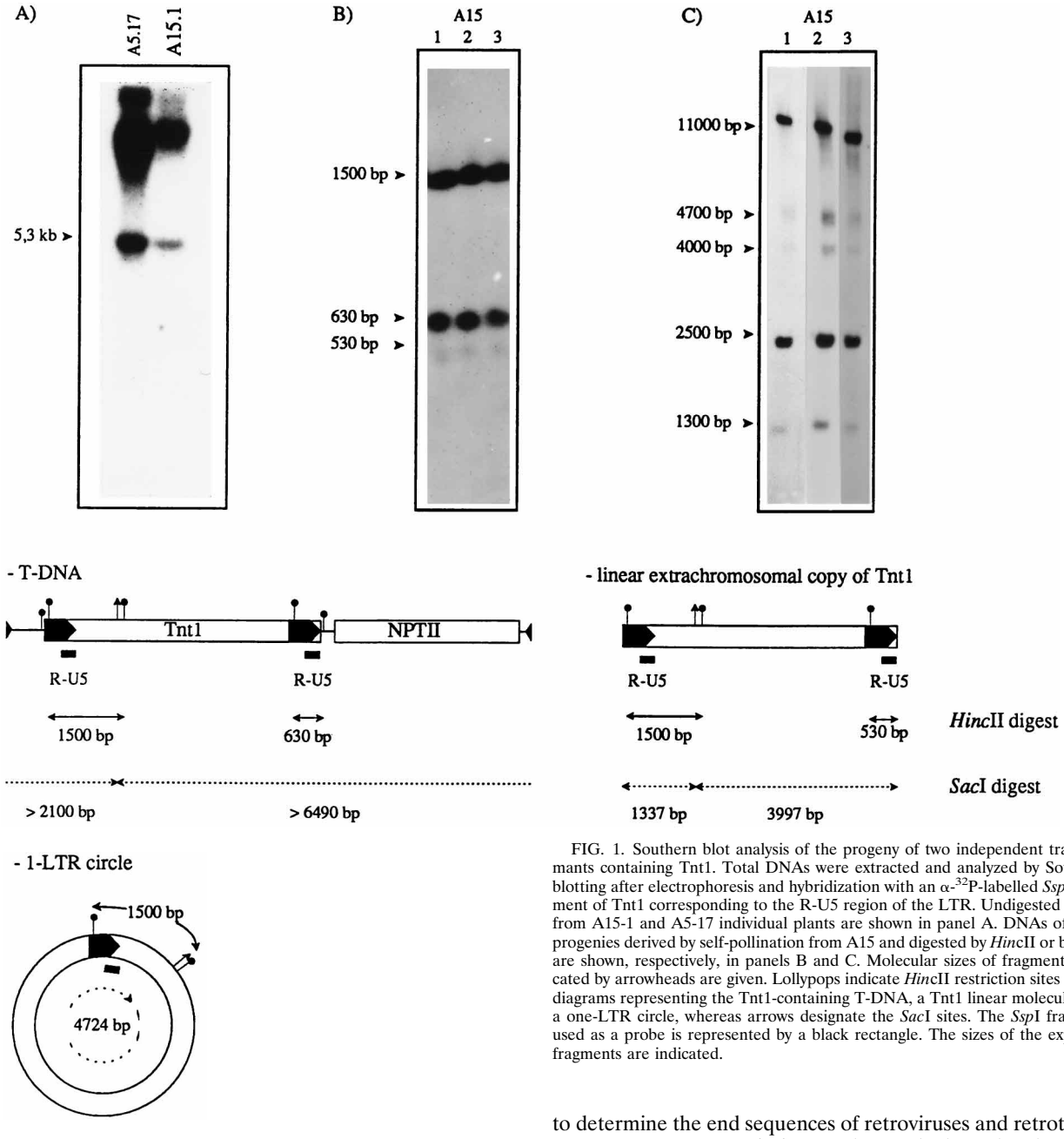


FIG. 1. Southern blot analysis of the progeny of two independent transformants containing Tnt1. Total DNAs were extracted and analyzed by Southern blotting after electrophoresis and hybridization with an α-^{32}P-labelled *Ssp*I fragment of Tnt1 corresponding to the R-U5 region of the LTR. Undigested DNAs from A15-1 and A5-17 individual plants are shown in panel A. DNAs of three progenies derived by self-pollination from A15 and digested by *Hin*cII or by *Sac*I are shown, respectively, in panels B and C. Molecular sizes of fragments indicated by arrowheads are given. Lollypops indicate *Hin*cII restriction sites on the diagrams representing the Tnt1-containing T-DNA, a Tnt1 linear molecule, and a one-LTR circle, whereas arrows designate the *Sac*I sites. The *Ssp*I fragment used as a probe is represented by a black rectangle. The sizes of the expected fragments are indicated.

of the same blot with probes corresponding to Tnt1 reverse transcriptase or *gag* domains (data not shown).

No additional bands that could correspond to two-LTR circular molecules were visible on the Southern blot shown in Fig. 1C. The presence of two-LTR head-to-tail circular molecules in total DNA extracted from aging leaves of A15-1 and A5-17 transgenic plants was specifically assayed by PCR with the Tnt540 and Tnt77 primers (Fig. 2A). PCR products of the expected size (ca. 170 bp) were detected after electrophoresis in a 1.5% agarose gel (data not shown), indicating that two-LTR head-to-tail circular molecules were present in the A15-1 and A5-17 genomic DNA samples.

Analysis of LTR circle junctions of Tnt1. Linear intermediates of transposition as well as extrachromosomal circular molecules containing two tandemly repeated LTRs have been used to determine the end sequences of retroviruses and retrotransposons' reverse transcription products. The boundary between the two LTRs in the two-LTR head-to-tail circular molecules is referred to as the circle junction (53). Two independent PCRs (reactions I and II) were performed with the A5-17 genomic DNA sample, and one PCR (reaction III) was performed with the A15-1 genomic DNA sample to amplify the Tnt1 circle junction. Additional PCR products were obtained after amplification of genomic DNA extracted from aging leaves of A15-1 pooled progenies (reaction IV). Primers Tnt77 and Tnt540 were used in reactions I and III, whereas primers tab2-3′ and Tnt540 were used in reactions II and IV. Each PCR product was cloned into a *Sma*I-digested vector, and plasmid DNAs from 43 independent clones were extracted and sequenced (Fig. 2B).

Six circular molecules (Fig. 2B, sequence a) contain the ACAT nucleotides between two complete LTRs of Tnt1. Since in Tnt1, the PBS is separated from the 3′ LTR by an AC

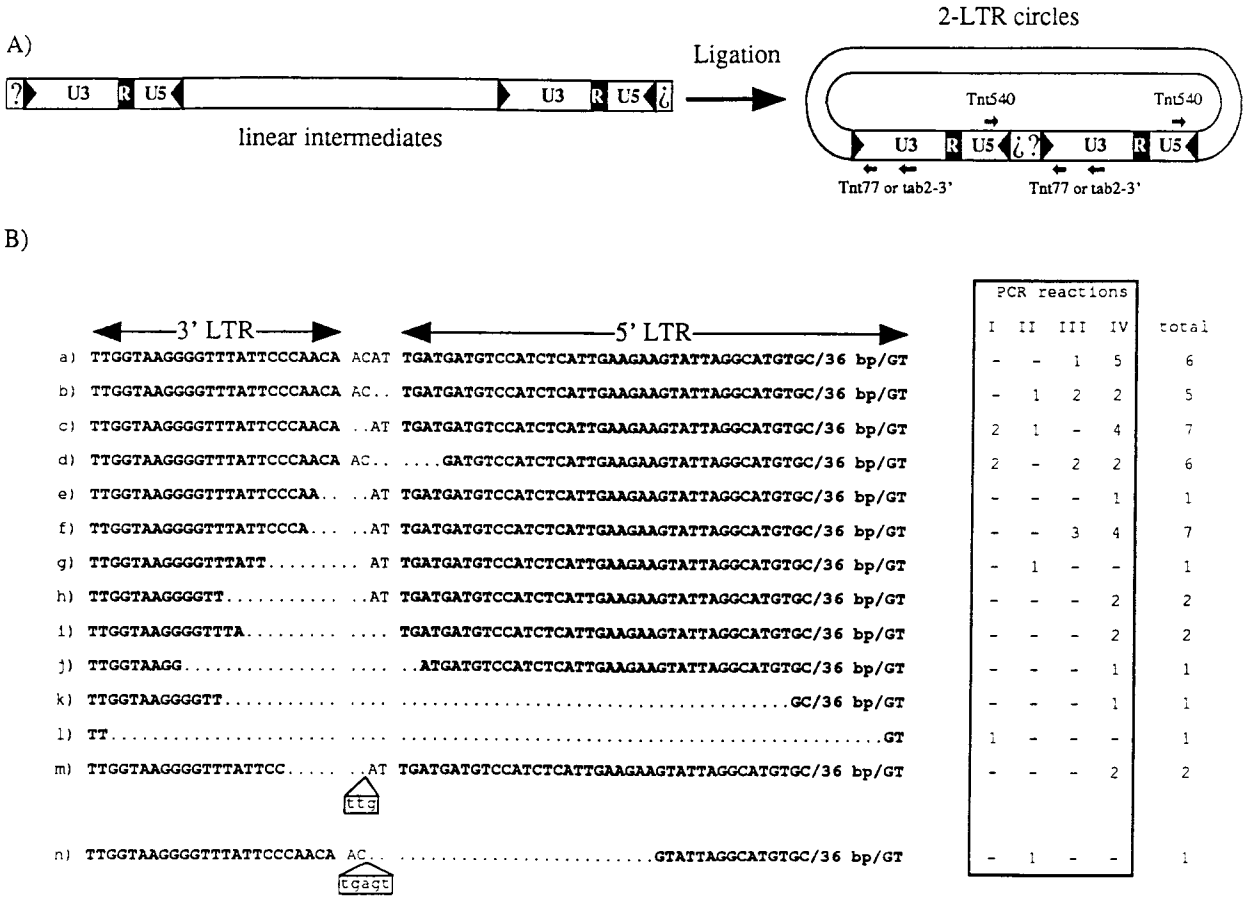


FIG. 2. Detection by PCR of Tnt1 two-LTR circles and sequence of the circle junction. (A) Schematic drawing of the linear and two-LTR circular molecules of the tobacco Tnt1 retrotransposon. The oligonucleotides (Tnt77, tab2-3′, and Tnt540) used for the PCR amplification and their positions in Tnt1 LTRs are shown by small arrows for the two-LTR circles. Question marks indicate the potential presence of extra base pairs at each end of the LTRs. (B) Sequence of 43 independent clones obtained after PCR amplification of the Tnt1 circle junction. Bold uppercase letters indicate LTR sequences; small dots indicate deletions; uppercase letters indicate Tnt1-derived extra bases; lowercase letters indicate bases of unknown origin. A5-17 genomic DNA was used as a template in reactions I and II, with, respectively, the Tnt77/Tnt540 and the tab2-3′/Tnt540 couples of oligonucleotides. Reaction III was performed with A15-1 genomic DNA with Tnt77 and Tnt540. Reaction IV was performed with DNA extracted from A15-1 pooled progenies with the tab2-3′ and the Tnt540 oligonucleotides as primers. For each PCR, the number of clones with identical sequence is indicated on the right. Because of the small number of clones analyzed per amplification reaction, the number of sequences in each class is not necessarily representative of their relative importance in vivo.

dinucleotide, one expects the formation of linear double-stranded products of reverse transcription with the extra AC/TG base pairs downstream of the 3′ LTR. The extra AT/TA base pairs upstream of the 5′ LTR are identical to the ones found immediately upstream of the Tnt1 3′ LTR and suggest that priming of the plus-strand DNA by the reverse transcriptase is initiated two nucleotides upstream of the Tnt1 3′ LTR. The a sequences shown in Fig. 2B result most probably from ligation of full-length linear reverse-transcribed copies of Tnt1, carrying two extra base pairs at each terminus compared to the integrated forms of the retrotransposon.

Twelve circular molecules (Fig. 2B, sequences b and c) contain, between the two complete LTRs of Tnt1, the AC or AT dinucleotide and could correspond to circular molecules derived from the ligation of linear intermediates of replication with a truncation of 2 bp at their 5′ or 3′ terminus. The 17 circle junctions represented in sequences d to h (Fig. 2B) suggest that the linear molecules from which they originate underwent an even more important degradation of one of their termini prior to circularization.

Five sequences (i to l in Fig. 2B) are made of truncated 3′ and 5′ LTRs, suggesting that ligation occurred after degradation of both ends of a linear intermediate of replication or after imprecise excision of a genomic Tnt1 element as proposed by Flavell and Ish-Horowicz (18) and Flavell and Brierley (17) for the *D. melanogaster copia* retrotransposon.

The last two sequences (Fig. 2B, m and n) contain, between an intact LTR end and a truncated one, an insertion of a few base pairs that might result from the action of a deoxynucleotidyltransferase-like enzyme, as proposed by Smith et al. (49), from ligation with an extrachromosomal fragment of DNA or from nontemplated nucleotide addition by the reverse transcriptase (45).

Tnt1 double-stranded DNA linear molecules are flanked by 2 bp on each side. Although the circle junctions sequenced suggested that linear reverse-transcribed copies of Tnt1 were flanked on one side by the AT/TA base pairs and on the other side by the AC/TG base pairs, only 6 of the 43 sequences analyzed contained, between the two head-to-tail LTRs, these 4 bp. A more direct analysis of Tnt1 linear double-stranded reverse-transcribed copies was therefore undertaken by a ligation-mediated PCR approach.

Undigested DNA from A15-1 transgenic plant in which linear intermediates of replication of Tnt1 had been detected by

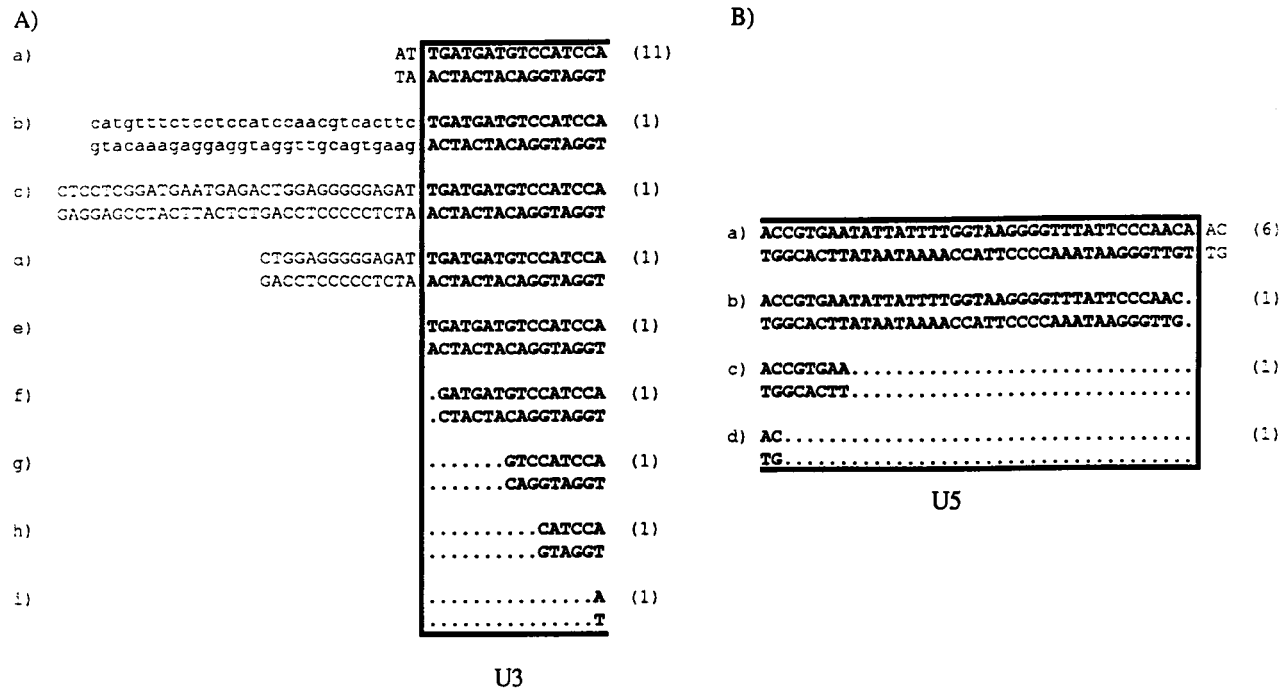


FIG. 3. End sequence of Tnt1 linear molecules. Sequences derived from one amplification reaction of the 5′ (A) or the 3′ (B) end of the blunt-ended Tnt1 linear molecules present in A15-1 genomic DNA sample are shown. Bold uppercase letters indicate LTR sequences, small dots indicate deletions, uppercase letters indicate Tnt1-derived extra base pairs, and lowercase letters indicate the sequence of the tobacco *Nia2* gene present in the Tnk23 vector used for transformation. The AT/TA base pairs correspond to the 2 bp found upstream of the Tnt1 3′ LTR, and the AC/TG base pairs correspond to the 2 bp separating the PBS from the 5′ LTR. The number of clones with identical sequences is indicated on the right.

Southern blot analysis (Fig. 1) and from pooled progenies of 2-S14-6-10 in which Tnt1 transcription had been induced by $CuCl_2$ (40) were ligated with the AD1 asymmetrical adaptor. Since AD1 is blunt ended on one side and has a 14-nt 5′ overhang on the other side, only blunt-ended double-stranded plant genomic DNA molecules were expected to be able to ligate with this adaptor. Nested PCR amplifications were performed as described in Materials and Methods to selectively amplify the 5′ or the 3′ end of Tnt1 linear molecules. After migration in a 1.5% agarose gel of the PCR products, a major fragment running at ca. 70 bp was identified for the 5′ LTR, and one running at ca. 100 bp was identified for the 3′ LTR (data not shown). No amplification was observed on control DNA that had not been subjected to ligation with AD1. The amplified fragments from A15-1 plant DNA were cloned in the *Eco*RV site of pBluescript. The sequences of 19 independent clones derived from the amplification of the 5′ LTR extremity and of 9 clones originating from that of the 3′ LTR extremity are shown in Fig. 3A and B, respectively.

Eleven of the 19 sequences obtained for the 5′ LTR of Tnt1 start with the dinucleotide AT and are followed by the U3 region (Fig. 3A, a). The fact that these nucleotides are identical to the ones found just upstream of the 3′ LTR in the Tnt1 central domain confirms that Tnt1 plus-strand synthesis is initiated 2 bp upstream of its 3′ LTR (see Fig. 5). One clone (Fig. 3A, b) contains a tobacco nitrate reductase sequence preceding Tnt1 LTR. It could be derived from a ligation event between the AD1 adaptor and a nicked blunt-ended genomic DNA molecule, since the T-DNA present in the Tnk23 vector contains a stretch of the *Nia2* gene in which Tnt1 was isolated (26). Two sequences (Fig. 3A, c and d) correspond to the Tnt1 central domain upstream of the 3′ LTR and could be attributed to the same phenomenon. If so, only the second PCR with the LS and Tnt77 primers can have allowed the synthesis of the amplified fragments, since no region homologous to the Cf5 oligonucleotide is present in the 3′ half of Tnt1. Alternatively, these two sequences could derive from aberrant reverse transcription products, the origin of which remains to be determined. The five remaining sequences in Fig. 3A (e to i) display total or partial LTR 5′ ends and correspond to deletion-containing versions of the clones analyzed above.

In the nine sequences obtained for the 3′ LTR of Tnt1 (Fig. 3B), six correspond exactly to the U5 region extended by the AC dinucleotide found between Tnt1 5′ LTR and the PBS, indicating that the majority of the linear molecules have the two expected extra base pairs at their 3′ extremity. The three others miss 1, 31, or 37 bp of the 3′ LTR.

The deletions of the termini of the LTR could be explained by the ligation of the adaptor either with nicked genomic DNA molecules or with degraded linear intermediates of Tnt1. These truncated molecules, if self-ligated, would create two-LTR head-to-tail circles lacking nucleotides on one or two of their ends, similar to the ones shown in Fig. 2B (sequences b to l).

Taken together, these results demonstrate that reverse transcription of Tnt1 generates double-stranded DNA linear intermediates of replication flanked on each side by 2 bp identical to the ones found between the 5′ LTR and the PBS for their 3′ extremity and to the ones found between the PPT and the 3′ LTR for their 5′ extremity, as illustrated in Fig. 5B (steps a to g).

Tnt1 double-stranded linear molecules are processed at their 3′-hydroxyl termini. Transposed chromosomal copies of Tnt1, either in its original host *N. tabacum* or in *Arabidopsis*, are deprived of the four extra base pairs present in the linear unintegrated intermediates of replication (16, 25, 26, 38). It is therefore logical to speculate that similar to the retroviruses

and Ty3 integrases, Tnt1 IN is capable of processing the linear blunt-ended products of reverse transcription to generate 3′-recessed DNA molecules which are then integrated into the host chromosomal DNA. This processing reaction produces, in all retroviruses and in Ty3, recessed 3′-OH ends adjacent to the dinucleotide CA at the end of the LTR. We therefore expected removal of two nucleotides on each 3′-hydroxyl end of the linear Tnt1 molecules and production of (i) a protruding 5′-AT-3′ dinucleotide upstream of the 5′ LTR and (ii) a protruding 3′-TG-5′ dinucleotide downstream of the 3′ LTR (see Fig. 5B, h). To determine whether this was the case, we designed two asymmetrical adaptors with 5′ overhangs of the short strand complementary to the expected AT (AD2) and TG (AD3) nucleotides. Ligation reactions with the *E. coli* DNA ligase were performed between these asymmetrical adaptors and A15-1 genomic DNA. Control ligation reactions between the same adaptors and blunt-ended DNA molecules obtained after treatment with the *E. coli* DNA polymerase I Klenow fragment of A15-1 genomic DNA or after digestion of a Tnt1-containing plasmid by *Hin*cII or *Eco*RV were also carried out. Primers specific for the Tnt1 5′ or 3′ end were subsequently used in combination with an adaptor-specific oligonucleotide (LS) to see whether we could detect, after PCR amplification, Tnt1 linear molecules with recessed 3′-OH ends. Two PCR amplifications were performed with the LS/Cf5 and then the LS/tab2-3′ pairs of oligonucleotides or the LS/Avi and then the LS/Tnt79^{+} pairs of oligonucleotides specific for Tnt1 5′ or 3′ end, respectively (see Materials and Methods). When the primers specific for the Tnt1 5′ end were used in the PCRs, a ca. 200-bp DNA fragment could only be amplified on the sample obtained after ligation of A15-1 DNA with the AD2 adaptor that has a 5′ overhang complementary to the AT dinucleotide expected to be present at the 5′ end of the Tnt1 linear molecules (Fig. 4B, lanes 1 to 5). Similarly, with the primers specific for the Tnt1 3′ end, a ca. 500-bp DNA fragment could only be detected after PCR amplification when ligation was performed with the AD3 adaptor characterized by a 5′ overhang complementary to the TG dinucleotide expected to be present at the 3′ end of Tnt1 linear molecules (Fig. 4B, lanes 6 to 10). These results indicate that the AD2 and AD3 adaptors are specific, respectively, for the 5′ and 3′ ends of Tnt1 linear intermediates of replication and that they cannot be ligated to blunt-ended molecules. No PCR products were obtained when the AD2 or AD3 adaptors were mixed with A15-1 genomic DNA in the absence of ligase (data not shown).

The amplified fragments were cloned into an *Eco*RV-digested vector. Sequences of the fragments corresponding to Tnt1 5′ and 3′ LTRs are given, respectively, in Fig. 4C and D. All 5′ LTR sequences carry at their extremity an upstream 5′-AT-3′ dinucleotide, and all sequences derived from the amplification of the 3′ LTR end are followed by an extra 3′-TG-5′ dinucleotide, indicating that some of the Tnt1 linear molecules present in A15-1 genomic DNA are recessed by at least two nucleotides at their 3′-hydroxyl termini. A few point mutations, introduced by Tnt1 reverse transcriptase or by the *Taq* polymerase, were found randomly distributed in some sequences (data not shown) and at the extremity of Tnt1 LTRs for three of the cloned PCR fragments (Fig. 4C and D).

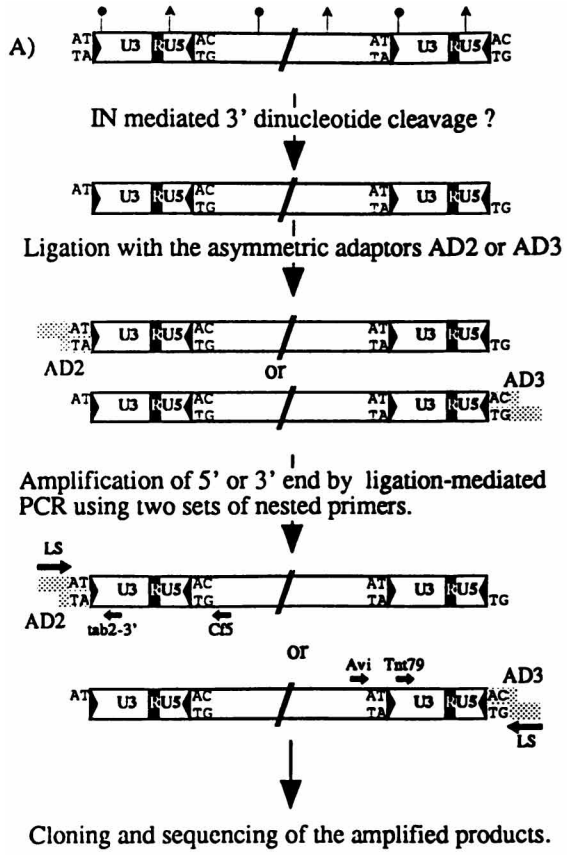
DISCUSSION

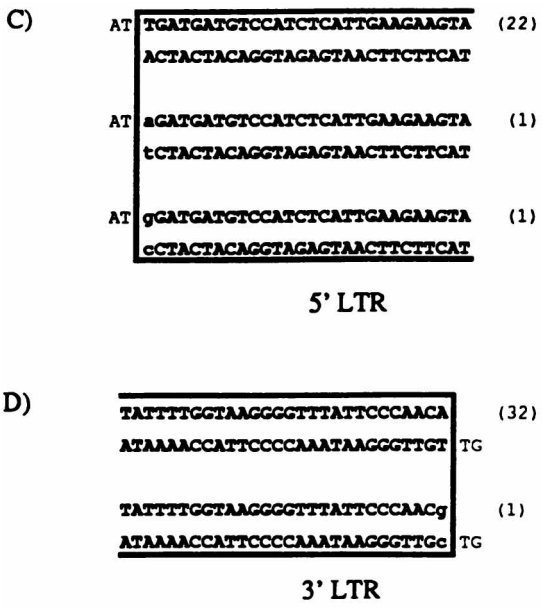
The two AC/TG base pairs found downstream of the 3′ LTR in the reverse-transcribed unintegrated copies of Tnt1 correspond to the ones found between the 5′ LTR and the region complementary to the tRNA$_i^{Met}$ of *A. thaliana* in the Tnt1 sequence. This complementarity extends from the end of the tRNA to the 12th ribonucleotide at position 65, defining for this retrotransposon a 12-bp PBS starting at the TGG triplet (Fig. 5A). It has been shown that during the synthesis of the plus-strand strong-stop DNA, the reverse transcriptase is able to reverse transcribe a portion of the tRNA attached to the 5′ end of the minus-strand strong-stop DNA and that modified ribonucleotides of the tRNA act as stop signals for this enzyme (22, 36, 51). The tRNA$_i^{Met}$ of *A. thaliana* presumably used by Tnt1 as a primer for the initiation of the minus-strand DNA is identical in sequence to that of wheat, in which a 2′-O-ribosylated G has been identified at position 64 (32). The full-length strong-stop DNA should therefore extend 14 nucleotides downstream of the LTR, up to the 12th ribonucleotide of the primer tRNA attached to the 5′ end of the minus-strand template. Since the complementarity between the Tnt1 PBS sequence and the last 12 ribonucleotides of the tRNA$_i^{Met}$ is perfect, it is likely that contrary to the case of the yeast Ty1 and Ty3 retrotransposons (35, 36), full-length strong-stop DNA species are used as intermediates of replication during Tnt1 reverse transcription.

The presence of the AT/TA base pairs upstream of the 5′ terminus of linear extrachromosomal forms of Tnt1, which correspond exactly to the 2 bp found immediately upstream of the Tnt1 3′ LTR, allows us to precisely map the 3′ end of Tnt1 PPT at position 4722 of this element. The 5′ end of Tnt1 PPT can only be inferred from the sequence and may be positioned at nucleotide 4713. Tnt1 PPT as defined here would have the sequence GGAGGGGGA and would be composed of 10 straight purines (Fig. 5A).

Full-length linear molecules of Tnt1 possess the AT/TA base pairs upstream of their 5′ end and the AC/TG base pairs downstream of their 3′ end. However, integrated copies of the tobacco Tnt1 retrotransposon end with the conserved TG. . .CA at their termini. The extra base pairs shown to be present at each terminus of the extrachromosomal linear forms are therefore removed prior to or during integration into the host genome. We have used ligation-mediated PCR to demonstrate that at least in some of the linear intermediates of the tobacco retrotransposon Tnt1, the AT and the AC dinucleotides 3′ to the conserved CAs at the termini of their 5′ and 3′ LTRs, respectively, are indeed removed prior to integration. One cannot exclude the possibility that some of the PCR products shown in Fig. 4C and D arose after ligation between the adaptors and replication intermediates recessed by more than two nucleotides at their 3′ ends. However, the data presented here, together with the fact that all integrated Tnt1 elements have intact LTR ends, suggest that the two nucleotides adjacent to the dinucleotide CA are processed at each 3′-hydroxyl terminus of Tnt1 linear intermediates prior to integration into the host genome, in a process similar to that described for retroviruses and for the yeast Ty3 retrotransposon (Fig. 5B, step h).

Three sequences out of 57 shown in Fig. 4C and D possess the expected AT or GT dinucleotide adjacent to an AG, GG, or CG instead of the conserved TG dinucleotide at the end of the Tnt1 LTR. These molecules could have been generated after IN 3′ dinucleotide cleavage of substrates with mutated ends, suggesting that Tnt1-encoded IN is able to cleave the phosphodiester bond 3′ to a CT, a CC, or a CG dinucleotide instead of the conserved CA. However, we find this explanation unlikely, since it has been shown that for retroviruses, mutations in the conserved CA introducing an A or a T residue sustain cleavage and/or integration (47, 52), whereas mutations introducing a G residue do not (6). This suggests that among the three cases we describe here, only the one with a CA-to-CT mutation (Fig. 4C) may derive from IN-mediated cleavage of a





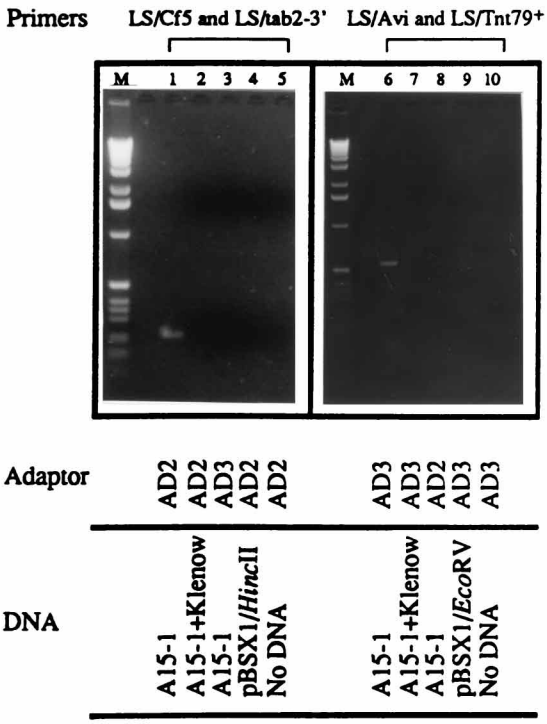


FIG. 4. The 3′ dinucleotide cleavage of Tnt1 linear transposition intermediates. (A) Schematic representation of the strategy used to search for Tnt1 linear molecules with 3′ recessed ends. IN-mediated 3′ dinucleotide cleavage is expected to remove two nucleotides on each 3′-hydroxyl end of the linear Tnt1 molecules to produce a protruding 5′-AT-3′ dinucleotide upstream of the 5′ LTR and a protruding 3′-TG-5′ dinucleotide downstream of the 3′ LTR. Two asymmetrical adaptors with 5′ overhangs of the short strand complementary to the expected AT (AD2) and TG (AD3) dinucleotides were ligated with genomic DNA. The occurrence of Tnt1 linear molecules with 3′ recessed ends was assayed by PCR with the LS oligonucleotide in combination with primers specific for the Tnt1 5′ or 3′ ends. Dotted boxes with a protruding AT indicate asymmetric adaptor AD2, dotted boxes with a protruding GT indicate asymmetric adaptor AD3, and small arrows indicate oligonucleotides used for PCR amplification. *Hin*cII and *Eco*RV sites are represented, respectively, by lollypops and arrows on the schematized linear molecule of Tnt1. (B) PCR amplification performed after ligation of undigested DNA from *Arabidopsis* aging leaves with the AD2 and AD3 adaptors. PCR amplifications were performed with the LS/Cf5 and then the LS/tab2-3′ pairs of oligonucleotides (1–5) or with the LS/Avi and then the LS/Tnt79^{+} pairs of oligonucleotides (6–10) specific for the 5′ and the 3′ ends of Tnt1, respectively. Prior to amplification, undigested genomic DNA from A15-1 aging leaves (1, 3, 6, 8), Klenow-treated A15-1 DNA (2, 7), and Tnt1-containing plasmid DNA (pBSX1) digested either by *Hin*cII (4) or *Eco*RV (9) were ligated either with the AD2 adaptor or with the AD3 adaptor. Negative controls with no added template DNA are also shown (5, 10). One-tenth of each reaction mixture was loaded on the gel. M, molecular size marker (1-kb ladder; BRL). (C and D) Sequences of the fragments derived from the 5′ or the 3′ termini of Tnt1 linear molecules with recessed 3′ ends. Bold uppercase letters indicate LTR sequences, uppercase letters indicate the extra dinucleotide left after 3′ cleavage, and lowercase letters indicate point mutations in the LTR. The number of clones with identical sequences is indicated on the right.

substrate bearing a terminal mutation introduced by Tnt1 reverse transcriptase. Consistent with this is the observation of such a CA-to-CT mutation in one out of 12 sequences derived from transposed copies of the tobacco retrotransposon Tnt1 in *A. thaliana* (16). The mutations observed could have been introduced by Tnt1 reverse transcriptase or by the *Taq* DNA polymerase, which are both error-prone polymerases. Twenty-two point mutations (over 22.3 kb sequenced) were found in the U3 or U5 regions of the amplified fragments (data not shown). Only one of them corresponds to a mutation already observed in a variable position of the Tnt1 U3 region (9) and may be attributable to Tnt1 reverse transcriptase. This, together with the fact that we performed two cycles of PCR amplification, suggests that the observed intermediates with noncanonical ends are due to PCR artifacts.

Our results clearly indicate that linear intermediates with 3′ recessed ends accumulate in tissues where transcription has

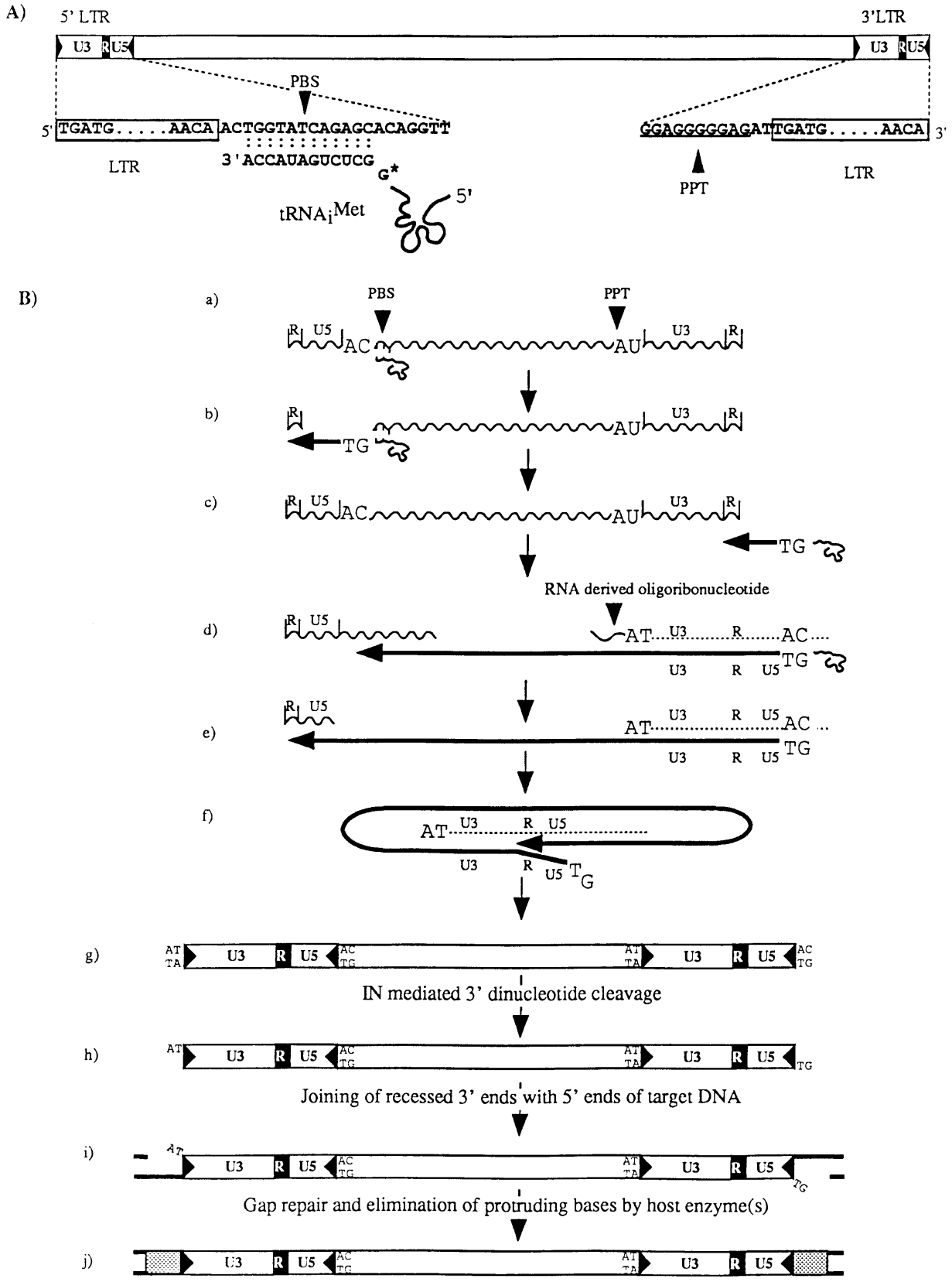


FIG. 5. Structure of the tobacco Tnt1 retrotransposon and proposed model for its reverse transcription and integration. (A) Structure of the tobacco Tnt1 retrotransposon. The PBS is shown annealed with *A. thaliana* $tRNA_i^{Met}$. G* indicates the 2′-O-ribosylated purine found at position 64 in all cytoplasmic initiator tRNAs from plants and fungi. The Tnt1 PBS is 12 bp long and is separated from the 5′ LTR by an AC dinucleotide. The Tnt1 PPT, presumably used as a primer for plus-strand DNA synthesis, is underlined. It is 10 bp long and is separated from the 3′ LTR by an AT dinucleotide. (B) Model for Tnt1 reverse transcription and integration based

TABLE 1. Distance between the PBS and the 5′ LTR of different retrotransposons

Retrotransposon	Host	Primer tRNA	No. of bp between PBS and 5′ LTR	GenBank accession no. (reference)
blood	*Drosophila melanogaster*	$tRNA^{Arg}$	0	x81127
copia	*Drosophila melanogaster*	$tRNA_i^{Met}$ [a]	0	x04456 (33)
micropia	*Drosophila melanogaster*	$tRNA^{Leu}$	0	x14037
1731	*Drosophila melanogaster*	$tRNA_i^{Met}$ [a]	0	x07656
412	*Drosophila melanogaster*	$tRNA^{Arg}$	0	x04132
mdg1	*Drosophila melanogaster*	$tRNA^{Arg}$	0	x59545
Ty1	*Saccharomyces cerevisiae*	$tRNA_i^{Met}$	0	m11351 (10)
Ty2	*Saccharomyces cerevisiae*	$tRNA_i^{Met}$	0	x03840
Ty4	*Saccharomyces cerevisiae*	$tRNA^{Asn}$	0	x67284
Ty5	*Saccharomyces cerevisiae*	$tRNA_i^{Met}$ [a]	0	u19263
Tf1	*Schizosaccharomyces pombe*	Self-priming	0	m38526 (37)
Tp2	*Physarium polycephalum*	$tRNA_i^{Met}$ [a]	0	x52770
Osser	*Volvox carteri*	$tRNA_i^{Met}$ [a]	0	x69552
Tst1	*Solanum tuberosum*	$tRNA_i^{Met}$	0	x52287
Bare-1	*Hordeum vulgare*	$tRNA_i^{Met}$	1	z17327
Stonor	*Zea mays*	$tRNA_i^{Met}$	1	s46937
Zeon-1	*Zea mays*	$tRNA^{Lys}$	1	u11059
Wis 2-1A	*Triticum aestivum*	$tRNA_i^{Met}$	1	x63184
Mag	*Bombyx mori*	?[b]	1	x17219
Ty3	*Saccharomyces cerevisiae*	$tRNA_i^{Met}$	2	m34549 (35)
Tnt1	*Nicotiana tabacum*	$tRNA_i^{Met}$	2	x13777 (this study)
Tto1	*Nicotiana tabacum*	$tRNA_i^{Met}$	2	d83003
Bs 1	*Zea mays*	$tRNA_i^{Met}$	2	m25397
Ta 1-3	*Arabidopsis thaliana*	$tRNA_i^{Met}$	2	x13291
Tos 1-1	*Oryza sativa*	$tRNA_i^{Met}$	2	x66521
SIRE-1	*Glycine max*	$tRNA_i^{Met}$	2	u22103
B 104	*Drosophila melanogaster*	?[b]	2	z48503
IAP	*Mesocricetus auratus*	$tRNA^{Phe}$	2	m10134
VL 30	*Mus musculus*	$tRNA^{Gly}$?	2	x17124
ART	*Gallus gallus*	$tRNA^{Trp}$	2	125262
Del 1	*Lilium henryi*	$tRNA_i^{Met}$	3	x13886
Athila	*Arabidopsis thaliana*	$tRNA^{Glu}$	3	x81801
Ulysses	*Drosophila virilis*	$tRNA^{Lys}$	3	x56645

[a] Reverse transcription is primed by an internal fragment of initiator methionine tRNA.
[b] ?, primer tRNA is not known or has been inferred by us from the element sequence.

been induced and suggest that Tnt1-encoded integrase, as do the retroviral IN (for a review, see references 29 and 54) and the yeast Ty3 enzyme (35), catalyzes 3′ dinucleotide cleavage to generate linear intermediates with an AT overhang on one side and a GT overhang on the other side. Further study is needed to confirm that the 3′ dinucleotide cleavage evidenced here is indeed performed by the tobacco Tnt1 retrotransposon IN.

on the retroviral model. The LTRs are divided into three parts: R, corresponding to the small repeats found at each end of the RNA, and U3 and U5, derived from the unique 3′ and 5′ regions of the RNA, respectively. After reverse transcription of retroviruses and LTR retrotransposons, a full-length double-stranded DNA copy with two complete LTRs is generated from a message that initiates and ends within the LTRs of the element. Wavy lines, RNA; heavy line with arrowhead, minus-strand DNA; dotted line, plus-strand DNA; heavy line, host DNA; dotted boxes, direct repeats resulting from Tnt1 integration in the host genome. The 2 bp separating the PBS from the 5′ LTR as well as the ones separating the PPT from the 3′ LTR are shown. (Step a) Only one genomic RNA molecule is shown for convenience, but two copies of this RNA are probably packaged in a virus-like particle, as in the case of retroviruses. The primer tRNA is annealed with a genomic RNA molecule. (Step b) Reverse transcription, primed by using the 3′-OH end of the primer tRNA, gives rise to the minus-strand strong-stop DNA when the reverse transcriptase reaches the 5′ end of the RNA. During this process, the RNA moiety of the RNA-DNA hybrid is removed by the reverse transcriptase-associated RNase H activity. (Step c) The minus-strand strong-stop DNA is transferred to the 3′ end of a second RNA molecule. (Step d) The minus-strand strong-stop DNA is elongated for extensive minus-strand DNA synthesis. During this process, the RNA moiety of the RNA-DNA hybrid is degraded by the RNase H activity. An RNase H-resistant oligoribonucleotide corresponding to the PPT is used by the reverse transcriptase as a primer for the synthesis of plus-strand strong-stop DNA. During the synthesis of this short DNA molecule, at least part of the primer tRNA is copied. (Step e) The minus-strand DNA is extended to the 5′ end of the RNA template, and the primer tRNA is removed by the RNase H activity. (Step f) The minus strand is extended by displacement synthesis along the plus-strand template, and the plus-strand is completed by using the minus-strand as a template. (Step g) Reverse transcription of Tnt1 genomic RNA generates a linear intermediate with two extra base pairs at each end compared with integrated forms. (Step h) The extra dinucleotides 3′ to the conserved CA in the minus strand are cleaved off by the IN protein. (Step i) Tnt1 3′ recessed ends are joined with the 5′ ends of a 5-bp staggered cut in the host DNA catalyzed by the integrase. (Step j) Host enzymes remove the protruding bases at the 5′ ends of Tnt1 and repair the gap, generating a 5-bp duplication at the target site.

Analysis of retrotransposon sequences available from the databases strongly suggests that the tobacco Tnt1 and the yeast Ty3 may not be the only retrotransposons for which integrase is able to perform 3′ dinucleotide cleavage, since many different elements possess a PBS separated from the 5′ LTR by a few base pairs (Table 1). Based on the distance between the PBS and the 5′ LTR, four classes of retrotransposons can be distinguished.

The first class of elements comprising the yeast Ty1 and the *D. melanogaster copia* retrotransposons has a PBS directly adjacent to the LTR. Consequently, extrachromosomal forms of these elements are not expected to possess extra base pairs at their ends and should be integrated without further modifications. Studies with purified Ty1 IN have shown, albeit indirectly, that extrachromosomal linear and integrated forms of this retrotransposon are identical and that blunt linear molecules can be integrated (41, 42).

The second class of elements, similar to the case of HIV-1 (46, 56), has a PBS separated from the 5′ LTR by 1 bp. Extrachromosomal linear forms of such elements should possess one extra base pair downstream of their 3′ LTR compared with their integrated forms. However, sequencing of the circle junction of HIV-1 (28, 49, 58) has shown that extrachromosomal forms of this retrovirus possess 2 bp downstream of the 3′ LTR. It is thus possible that the cleavage mediated by the RNase H encoded by these elements to remove the primer tRNA attached to the 5′ end of the minus-strand template occurs in the CCA triplet between the terminal A and the adjacent C, giving rise to extrachromosomal forms with 2 bp instead of 1 bp downstream of their 3′ LTR, as has been proposed for HIV-1 (58).

In the third class of elements, including the yeast Ty3 and the tobacco Tnt1 retrotransposons, the PBS and the 5′ LTR are separated by 2 bp, and extrachromosomal linear forms of these elements possess two extra base pairs downstream of their 3′ LTR that are removed prior to integration (see reference 35 and this study). This class of retrotransposons, together with the first one, represents about 75% of the retrotransposons (Table 1). It contains many plant retroelements, including the tobacco retrotransposons Tnt1 and Tto1. The sequence analysis performed here, together with the analysis of the tobacco Tto1 retrotransposon circular forms (27), allows us to propose that Tto1 linear extrachromosomal intermediates would have two extra nucleotides downstream of the 3′ LTR, as is the case reported here for Tnt1. However, unlike Tnt1 linear DNA, Tto1 linear forms would have three extra nucleotides upstream of the 5′ LTR, since the circle junction consensus for this element (AGAAT) is 5 bp long (27). This strongly suggests that Tto1-encoded integrase cleaves the end of linear DNA asymmetrically, as has been proposed for HIV-2 (57).

A fourth class of retrotransposons includes a few elements whose PBS is separated from the 5′ LTR by 3 bp, suggesting that their linear forms would have three extra base pairs downstream of the 3′ LTR and that the element's encoded IN should be able to cleave three nucleotides instead of two nucleotides 3′ to the conserved CA.

The *gypsy*-like elements (e.g., *gypsy*, tom, 17.6, and TED) (19, 39, 48, 50), with a PBS overlapping the 5′ LTR, may constitute a fifth class of elements. The presence in these elements of a third open reading frame homologous to the retroviral *env* gene, together with the fact that some of them are able to form infectious particles (34, 50), suggests, however, that they are more closely related to retroviruses than to retrotransposons. They have therefore not been included in Table 1.

In conclusion, we have shown here that the PBS and the PPT of the tobacco retrotransposon Tnt1 used as primers for the initiation of minus- and plus-strand DNAs are separated by two nucleotides from the 5′ and 3′ LTRs, respectively, giving rise to extrachromosomal linear intermediates with two extra base pairs at each end compared with the integrated copies as observed for retroviruses (23, 53) and the yeast Ty3 retrotransposon (35). We have also shown that the 3′ ends of such linear intermediates are processed at each LTR extremity. This cleavage is thought to be mediated by Tnt1-encoded integrase by analogy with retroviral systems. Our data, together with the report by Kirchner and Sandmeyer (35) for the yeast Ty3 retrotransposon, reinforce the idea that 3′ dinucleotide cleavage is not restricted to retroviral enzymes.

ACKNOWLEDGMENTS

We thank S. Boussion, J. Goujaud, and J. Talbotec for technical assistance. We are also grateful to M.-A. Grandbastien and H. Vaucheret for critical reading of the manuscript.

Frank Feuerbach is supported by a fellowship from the MRES.

REFERENCES

1. **Akama, K., and S. Tanifuji.** 1989. Nucleotide sequence of a methionine initiator tRNA gene of *Arabidopsis thaliana*. Plant Mol. Biol. **13:**599–600.
2. **Bechtold, N., J. Ellis, and G. Pelletier.** 1993. *In planta Agrobacterium* mediated gene transfer by infiltration of adult *Arabidopsis thaliana* plants. C. R. Acad. Sci. Paris **316:**1194–1199.
3. **Boeke, J. D., D. 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. **Braiterman, L. T., and J. D. Boeke.** 1994. Ty1 in vitro integration: effects of mutation in *cis* and in *trans*. Mol. Cell. Biol. **14:**5731–5740.
5. **Brown, P. O., B. Bowerman, H. E. Varmus, and J. M. Bishop.** 1989. Retroviral integration: structure of the initial covalent product and its precursor, and a role for the viral IN protein. Proc. Natl. Acad. Sci. USA **86:**2525–2529.
6. **Bushman, F. D., and R. Craigie.** 1991. Activities of human immunodeficiency virus (HIV) integration protein in vitro: specific cleavage and integration of HIV DNA. Proc. Natl. Acad. Sci. USA **88:**1339–1343.
7. **Bushman, F. D., T. Fujiwara, and R. Craigie.** 1990. Retroviral DNA integration directed by HIV integration protein in vitro. Science **249:**1555–1558.
8. **Canaday, J., P. Guillemaut, and J. Weil.** 1980. The nucleotide sequence of the initiator transfer RNAs from bean cytoplasm and chloroplasts. Nucleic Acids Res. **8:**999–1008.
9. **Casacuberta, J. M., S. Vernhettes, and M.-A. Grandbastien.** 1995. Sequence variability within the tobacco retrotransposon Tnt1 population. EMBO J. **14:**2670–2678.
10. **Chapman, K. B. A., 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. **Craigie, R., T. Fujiwara, and F. D. Bushman.** 1990. The IN protein of Moloney murine leukemia virus processes the viral DNA ends and accomplishes their integration in vitro. Cell **62:**829–837.
12. **Doyle, J. J., and J. L. Doyle.** 1990. Isolation of plant DNA from fresh tissue. Focus **12:**13–15.
13. **Eibel, H., J. Gafner, A. Stotz, and P. Philippsen.** 1981. Characterization of the yeast mobile genetic element Ty1. Cold Spring Harbor Symp. Quant. Biol. **45:**609–617.
14. **Eichinger, D. J., and J. D. Boeke.** 1990. A specific terminal structure is required for Ty1 transposition. Genes Dev. **4:**324–330.
15. **Feinberg, A. P., and B. Volgestein.** 1983. A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. **132:**6–13.
16. **Feuerbach, F.** Unpublished observations.
17. **Flavell, A. J., and C. Brierley.** 1986. The termini of extrachromosomal linear *copia* elements. Nucleic Acids Res. **14:**3659–3669.
18. **Flavell, A. J., and D. Ish-Horowicz.** 1983. The origin of extrachromosomal circular *copia* elements. Cell **34:**415–419.
19. **Friesen, P. D., and M. S. Nissen.** 1990. Gene organization and transcription of TED, a lepidopteran retrotransposon integrated within the baculovirus genome. Mol. Cell. Biol. **10:**3067–3077.
20. **Fujiwara, T., and K. Mizuuchi.** 1988. Retroviral DNA integration: structure of an integration intermediate. Cell **54:**497–504.
21. **Ghosh, H. P., K. Ghosh, M. Simsek, and U. L. RajBhandary.** 1982. Nucleotide sequence of wheat germ cytoplasmic initiator methionine transfer ribonucleic acid. Nucleic Acids Res. **10:**3241–3247.
22. **Gilboa, E., S. W. Mitra, S. P. Goff, and D. Baltimore.** 1979. A detailed model of reverse transcription and a test of crucial aspects. Cell **18:**93–100.

23. **Goff, S. P.** 1992. Genetics of retroviral integration. Annu. Rev. Genet. **26:** 527–544.
24. **Grandbastien, M., C. Audeon, J. M. Casacuberta, P. Grappin, H. Lucas, C. Moreau, and S. Pouteau.** 1994. Functional analysis of the tobacco Tnt1 retrotransposon. Genetica **93:**181–189.
25. **Grandbastien, M.-A.** Unpublished observations.
26. **Grandbastien, M.-A., A. Spielmann, and M. Caboche.** 1989. Tnt1, a mobile retroviral-like transposable element of tobacco isolated by plant cell genetics. Nature **337:**376–380.
27. **Hirochika, H., and H. Otsuki.** 1995. Extrachromosomal circular forms of the tobacco retrotransposon Tto1. Gene **165:**229–232.
28. **Hong, T., K. Drlica, A. Pinter, and E. Murphy.** 1991. Circular DNA of human immunodeficiency virus: analysis of circle junction nucleotide sequences. J. Virol. **65:**551–555.
29. **Katz, R. A., and A. M. Skalka.** 1994. The retroviral enzymes. Annu. Rev. Biochem. **63:**133–173.
30. **Katzman, M., R. A. Katz, A. M. Skalka, and J. Leis.** 1989. The avian retroviral integration protein cleaves the terminal sequences of linear viral DNA at the in vivo sites of integration. J. Virol. **63:**5319–5327.
31. **Kere, J., R. Nagaraja, S. Mumm, A. Ciccodicola, M. d'Urso, and D. Schlessinger.** 1992. Mapping human chromosomes by walking with sequence-tagged sites from end fragments of yeast artificial chromosome inserts. Genomics **14:**241–248.
32. **Kiesewetter, S., G. Ott, and M. Sprinzl.** 1990. The role of modified purine 64 in initiator/elongator discrimination of $tRNA_i^{Met}$ from yeast and wheat germ. Nucleic Acids Res. **18:**4677–4682.
33. **Kikuchi, Y., Y. Ando, and T. Shiba.** 1986. Unusual priming mechanism of RNA-directed DNA synthesis in *copia* retrovirus-like particles of Drosophila. Nature **323:**824–826.
34. **Kim, A., C. Terzian, P. Santamaria, A. Pelisson, N. Prud'homme, and A. Bucheton.** 1994. Retrovirus in invertebrates: the gypsy retrotransposon is apparently an infectious retrovirus of *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA **91:**1285–1289.
35. **Kirchner, J., and S. B. Sandmeyer.** 1996. Ty3 integrase mutants defective in reverse transcription or 3′-end processing of extrachromosomal Ty3 DNA. J. Virol. **70:**4737–4747.
36. **Lauermann, V., K. Nam, J. Trambley, and J. D. Boeke.** 1995. Plus-strand strong-stop DNA synthesis in retrotransposon Ty1. J. Virol. **69:**7845–7850.
37. **Levin, H. L.** 1995. A novel mechanism of self-primed reverse transcription defines a new family of retroelements. Mol. Cell. Biol. **15:**3310–3317.
38. **Lucas, H., F. Feuerbach, K. Kunert, M.-A. Grandbastien, and M. Caboche.** 1995. RNA-mediated transposition of the tobacco retrotransposon Tnt1 in Arabidopsis thaliana. EMBO J. **14:**2364–2373.
39. **Marlor, R. L., S. M. Parkhurst, and V. G. Corces.** 1986. The *Drosophila melanogaster gypsy* transposable element encodes putative gene products homologous to retroviral proteins. Mol. Cell. Biol. **6:**1129–1134.
40. **Mhiri, C., J.-B. Morel, S. Vernhettes, J. M. Casacuberta, H. Lucas, and M.-A. Grandbastien.** 1997. The promoter of the tobacco Tnt1 retrotransposon is induced by wounding and by abiotic stress. Plant Mol. Biol. **33:**257–266.
41. **Moore, S. P., and D. J. Garfinkel.** 1994. Expression and partial purification of enzymatically active recombinant Ty1 integrase in *Saccharomyces cerevisiae*. Proc. Natl. Acad. Sci. USA **91:**1843–1847.
42. **Moore, S. P., M. Powers, and D. J. Garfinkel.** 1995. Substrate specificity of Ty1 integrase. J. Virol. **69:**4683–4692.
43. **Moreau-Mhiri, C., J.-B. Morel, C. Audeon, M. Ferault, M.-A. Grandbastien, and H. Lucas.** 1996. Regulation of expression of the tobacco Tnt1 retrotransposon in heterologous species following pathogen-related stresses. Plant J. **9:**409–419.
44. **Palmer, J., and W. R. Folk.** 1987. Isolation and sequence analysis of a nuclear $tRNA^{met\text{-}i}$ gene from soybean. Plant Mol. Biol. **8:**47–51.
45. **Preston, B. D.** 1996. Error-prone retrotransposition: rime of the ancient mutators. Proc. Natl. Acad. Sci. USA **93:**7427–7431.
46. **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, I. 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 **313:**277–284.
47. **Roth, M. J., P. L. Schwartzberg, and S. P. Goff.** 1989. Structure of the termini of DNA intermediates in the integration of retroviral DNA: dependence on IN function and terminal DNA sequence. Cell **58:**47–54.
48. **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 *Drosophila melanogaster*. Nature **312:**659–661.
49. **Smith, J. S., S. Kim, and M. J. Roth.** 1990. Analysis of long terminal repeat circle junctions of human immunodeficiency virus type 1. J. Virol. **64:**6286–6290.
50. **Tanda, S., J. L. Mullor, and V. G. Corces.** 1994. The *Drosophila* tom retrotransposon encodes an envelope protein. Mol. Cell. Biol. **14:**5392–5401.
51. **Taylor, J., T. W. Hsu, C. Yeate, and W. S. Mason.** 1979. Synthesis and integration of avian sarcoma virus DNA. Cold Spring Harbor Symp. Quant. Biol. **44:**1091–1096.
52. **Van Den Ent, F. M. I., C. Vink, and R. H. A. Plasterk.** 1994. DNA substrate requirements for different activities of the human immunodeficiency virus type 1 integrase protein. J. Virol. **68:**7825–7832.
53. **Varmus, H., and P. Brown.** 1989. Retroviruses, p. 53–108. *In* D. E. Berg and M. M. Howe (ed.), Mobile DNA. American Society for Microbiology, Washington, D.C.
54. **Vink, C., and R. H. A. Plasterk.** 1993. The human immunodeficiency virus integrase protein. Trends Genet. **9:**433–437.
55. **Vora, A. C., M. L. Fitzgerald, and D. P. Grandgenett.** 1990. Removal of 3′-OH terminal nucleotides from blunt-ended long terminal repeat termini by the avian retrovirus integration protein. J. Virol. **64:**5656–5659.
56. **Wain-Hobson, S., P. Sonigo, O. Danos, S. Cole, and M. Alizon.** 1985. Nucleotide sequence of the AIDS virus, LAV. Cell **40:**9–17.
57. **Whitcomb, J. M., and S. H. Hughes.** 1991. The sequence of human immunodeficiency virus type 2 circle junction suggests that integration protein cleaves the ends of linear DNA asymetrically. J. Virol. **65:**3906–3910.
58. **Whitcomb, J. M., R. Kumar, and S. H. Hughes.** 1990. Sequence of the circle junction of human immunodeficiency virus type 1: implications for reverse transcription and integration. J. Virol. **64:**4903–4906.