DNA RESEARCH **2**, 255–261 (1995)

Molecular Analysis of a Putative Transposable Retroelement from the *Zea* Genus with Internal Clusters of Tandem Repeats

Amparo MONFORT, Carlos M. VICIENT, Regina RAZ, Pere PUIGDOMÈNECH, and José A. MARTÍNEZ-IZQUIERDO*

Departamento de Genética Molecular, CID-CSIC, Jordi Girona Salgado 18-26, 08034 Barcelona, Spain

(Received 8 December 1995)

Abstract

The molecular characterization of a recently discovered family of long repetitive sequences, termed ZLRS, is described. These elements belong to the class of moderate dispersed repetitive DNA and are specific to the *Zea* genus. An 8089-bp sequence from a *Zea diploperennis* ZLRS element have been elucidated. Sequence analysis reveals the presence of a long terminal repeat-like region, two clusters of different tandem repeats and several ORFs. On these grounds, ZLRS could be considered a new member of the superfamily of transposable retroelements. Tandems are present in the majority of ZLRS elements, they show an important stem-loop secondary structure predicted by the computer and their sequence conservation suggests a functional role.

Key words: Retrotransposon-like element; Tandem repeats; Stem-loop; (*Zea diploperennis*); (maize)

1. Introduction

Long and interspersed moderate repetitive sequences present in the genome of eukaryotes are thought to be transposable elements and include transposons and transposable retroelements.[1] The first group transposes via DNA, and good examples of them are the mobile genetic elements discovered by Barbara McClintock in maize more than 40 years ago.[2] Transposable retroelements transpose instead via an RNA intermediate that is copied to DNA by a reverse transcriptase and can be classified into three groups: retroviruses, which have not been described in plants; retrotransposons; and retroposons.[3,4] Like retroviruses, retrotransposons have internal coding *gag* and *pol* domains flanked by long terminal repeats (LTRs). Retroposons, also called non-LTR retrotransposons or LINEs (long interspersed repetitive elements) constitute a not well defined family of mobile elements that contain a variable number of open reading frames (ORFs) and lack LTRs at the ends, although a poly(A) tail in their 3′ end is frequently found instead.[5] Retroposons, usually, contain an ORF encoding protein motifs that show a weak similarity with reverse transcriptase conserved sequences, as is the case of the *Chlamydomonas reinhardtii* TOC1 transposon[6] and the human LINE-1.[7] In some transposable elements, additional ORFs are found that encode proteins, not all of them with a known function.[6,7,8]

A frequent feature observed in mammalian LINEs is the presence of tandem arrays of repeats at the 5′ end of the elements.[9] Tandem repeats are also observed in the R region of some retrovirus and retrotransposon LTRs.[10] It is not frequent to find tandem repeats in other regions of these elements. Exceptions are the *Drosophila* micropia retrotransposons, where the tandem repeats are located in the 3′ end, just upstream of the 3′-LTR;[11] and TOC1 transposable element from *Chlamydomonas reinhardtii*, where the tandem repeats are at the 5′ end of the element downstream of the putative 5′-LTR.[6] So far, no defined general functions have been attributed to the tandem repeats present in retroelements.

ZLRS elements constitute a new family of repetitive elements present in *Zea* species. They were initially detected in independent genomic clones of a *Zea diploperennis* library,[12] and have been recently identified as a family of long dispersed repetitive elements.[13] ZLRS are present in all the species of *Zea*, with a copy number of approximately 1500 per haploid genome for the modern maize (*Zea mays* ssp *mays*) and the teosinte *Z. diploperennis*. *In situ* hybridization has shown that ZLRS sequences are present in all the maize chromosomes with an interspersed pattern of distribution. In addition, their size has been estimated to be 9 kb by Southern hybridization.[13]

In this paper, we describe the structure of ZLRS elements (Fig. 1B). A combination of Southern analysis and DNA sequencing has revealed the presence of unusual

Communicated by Mituru Takanami

* To whom correspondence should be addressed. Tel. +34-3-400-6127, Fax. +34-3-204-5904

† Abbreviations: FB, foldback element(s) of *Drosophila*; HIV-1, human immunodeficiency virus type 1; HMG, high mobility group; LTR, long terminal repeat; ZLRS, *Zea* long repetitive sequence(s).

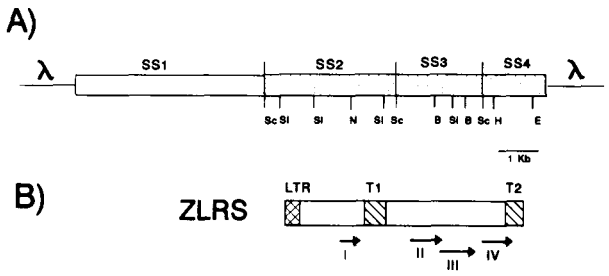

Figure 1. Restriction map and schematic representation of ZLRS elements. A) Restriction map of clone 7 ZLRS-containing element. The sequenced part of clone 7 corresponding to SS2, SS3 and SS4 segments (flanked by *Sac*I restriction sites) is indicated by a light shaded box. Vertical lines and symbols over the map represent the following restriction enzyme sites: *Bam*HI (B), *Eco*RI (E), *Hin*dIII (H), *Nae*I (N), *Sac*I (Sc) and *Sal*I (SI). B) Schematic drawing of ZLRS7 element showing the putative LTR (cross-hatched box), the two tandem arrays of repeats (hatched boxes), and the main ORFs termed by roman numbers (horizontal arrows).

clusters of tandem repeats, as well as internal ORFs and an LTR-like flanking sequence. These features suggest that ZLRS could be transposable retroelements as defined by Temin.[3]

2. Material and Methods

The source of plant material as well as lambda genomic clones 7 and 4, containing ZLRS from *Zea diploperennis*, is described elsewhere.[12,13] Southern blotting hybridization were performed essentially as described.[12] Probes were radioactively labelled by random priming.[14] DNA for sequencing was subcloned into pBluescript II KS+ vectors (stratagene, LaJolla). When needed, exonuclease III/S1 nuclease reactions were carried out[14] and the resulting plasmids with the best sizes of inserts were selected for sequencing.

DNA sequencing was performed in both strands by the dideoxy method[15] on doubled-stranded DNA templates using an Automated Laser Fluorescence System (Pharmacia), fluorescein labelled M13 universal and reverse primers and T7 polymerase. Sequence analysis was carried out with the GCG computer program package.[16] Searches for sequence similarities in updated versions of the GenEMBL and SWISSPROT Data Banks were conducted with the FASTA and TFASTA programs of the GCG package. The sequence of ZLRS element from *Z. diploperennis* lambda clone 7 will appear in the EMBL Nucleotide Sequence Database under the accession number X82087.

3. Results and Discussion

3.1. Definition of one extremity of ZLRS7 element

It was previously shown by Southern hybridization experiments with contiguous *Sac*I probes (SS1 to SS4 fragments in Fig. 1A) corresponding to the insert of *Z. diploperennis* lambda clone 7, that only SS2, SS3 and SS4 fragments formed part of the ZLRS family of repetitive elements.[13] In order to gain insight into the molecular structure of ZLRS, we have sequenced the SS2, SS3 and SS4 fragments in their entirety (8449 bp), as well as a region from clone 4 corresponding to the SS2 fragment of clone 7.[13] Sequence comparison between both clones revealed a dramatic increase in the percentage of identity (43% to 90%), beginning at the dinucleotide TG at position 361 of clone 7 (Fig. 2A). The high degree of identity between the two repetitive elements is maintained for over 500 bp (further sequence data was not available for clone 4), although only 176 bp of the sequence is shown (Fig. 2A). The initial dinucleotide of the sequence with high similarity between the two clones, TG, is the starting sequence of the great majority of LTRs from retroviruses and retrotransposons, and the end of the LTR is the inverted complementary dinucleotide, CA.[8] In order to identify the end of the ZLRS LTR, it was of interest to find the other LTR, since in functional retrotransposons the two LTRs are either identical or more than 90% similar to each other. The search of 8449 bp available sequence for the second LTR corresponding to clone 7 was unsuccessful. This was in spite of the large size (8089 bp) corresponding to the ZLRS element of clone 7. One possible explanation for the absence of the second LTR is that ZLRS7 could be a defective retrotransposable element with only one LTR. In this respect, TOC1 transposable element from *Chlamydomonas reinhardtii* has an imperfect putative 5′-LTR[6] and micropia retrotransposable elements from *Drosophila hydei* and *D. melanogaster* have imperfect, variable or absent 5′ or 3′ LTRs.[11]

Interestingly, the 5′ flanking sequence of ZLRS7, upstream of TG dinucleotide, shows a 70% similarity with the maize defective retrotransposon Cin1-like (Fig. 2B). Cin1 is a solo-LTR 691 bp long repetitive element found in maize.[17] The high similarity is lost downstream the TG of the putative LTR (Fig. 2B), suggesting that ZLRS7 was inserted into a Cin1-like element. The finding of transposable elements inserted into other transposable elements either from the same family or different types is a relatively frequent event. For example, *Physarum polycephalum* Tp1 retrotransposons have been found inserted in other Tp1 elements[18] and the inverted repeat element *Stowaway* has been found as an insertion in the *tourist* element.[19]

In order to determine if ZLRS7 has similarities with genes or structural features of other transposable elements, computer-assisted sequence analyses were per-

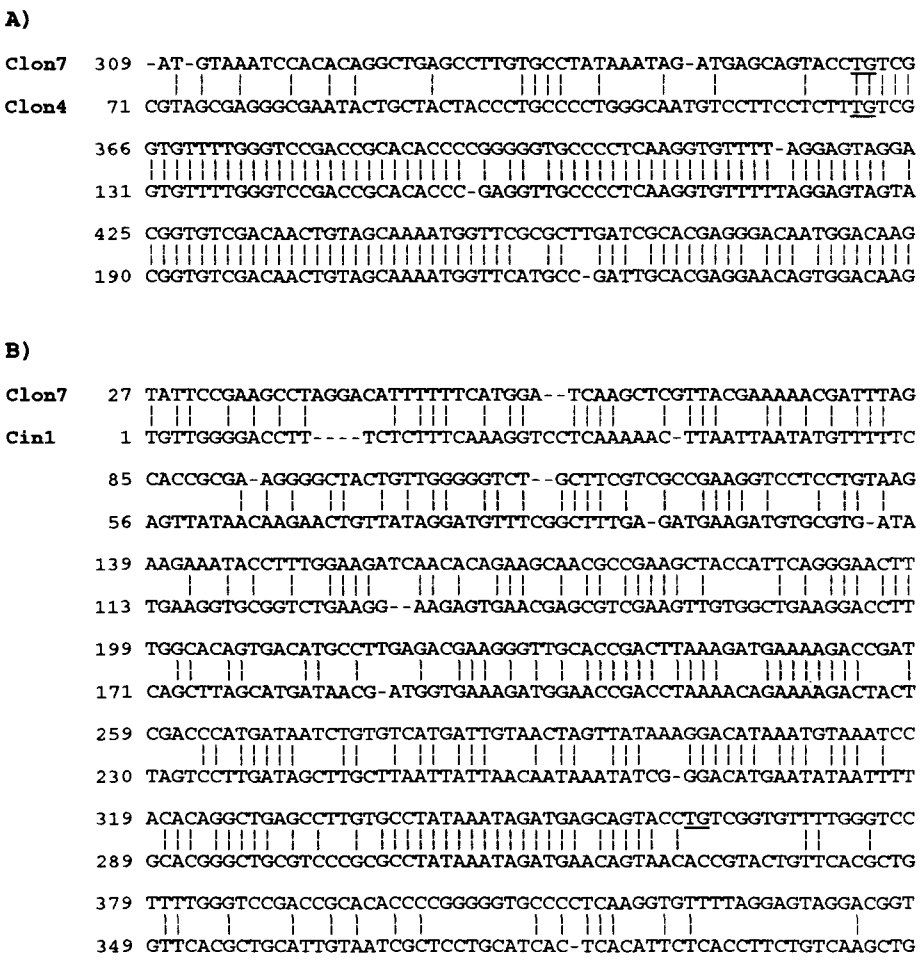

Figure 2. The 5′-end region of ZLRS7. A) Sequence alignment of clone 7 and 4 ZLRS-containing elements. The TG dinucleotide, representing the beginning of the putative LTR where the two sequences start to show high similarity, is underlined. B) Sequence alignment of ZLRS clone 7 and Cin1 defective retroelement.[17] The TG dinucleotide of ZLRS7 (the same as in A) where both sequences start to diverge, is underlined.

formed, that revealed the presence of several ORFs and arrays of tandem repeats (Fig. 1B).

3.2. Open reading frames

Unless otherwise stated, the cut-off value of the ORFs selected for further analysis was 500 bp. ORFs II to IV met this criterion and ORFI, 471 bp long, was selected because it contains the sequence of the previously characterized ZEAR repetitive sequence.[12] The analysis ORFs I to IV included searches in DNA and protein data banks and codon usage determination.

ORFs II to IV are clustered between two tandem arrays of repeats (Fig. 1B). This compact distribution of ORFs has also been described for the maize Bs1 and *C. reinhardtii* TOC1 defective retrotransposons.[6,20] Some of the ORFs from these two elements show weak similarities with reverse transcriptase and other typical proteins of retrotransposons. However, no significant similarities were found among ZLRS ORFs II to IV and those proteins or other known genes or proteins.

Northern analysis was performed with RNAs from different teosinte and maize tissues, using them as probes in hybridizations, DNA fragments containing ORFs II to IV. No transcripts were detected for those ORFs (data not shown). However, as previously published, small transcripts were found using a probe fragment of 270 bp corresponding to ORFI, termed ZEAR.[12] Transcripts of around 820 nucleotides are found in *Zea diploperennis* and several maize tissues including the poly(A)$^+$ RNA fraction.[12] Retroviruses and retrotransposons produce a full-size transcript from the promoter located in the 5′ LTR, termed genomic RNA. In addition, some retroelements produce a variety of RNAs from the genomic RNA by splicing of the genomic RNA or by transcription from

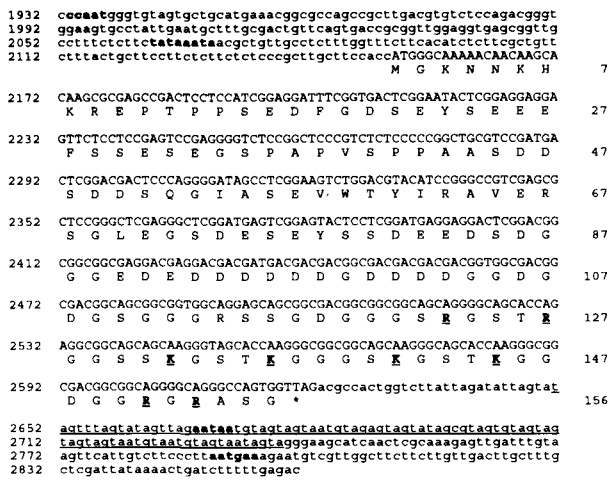

Figure 3. Nucleotide sequences of ZLRS7 ORFI and flanking regions. Lower-case letters correspond to ORFI flanking sequences. TATA and CCAAT boxes and putative polyadenylation signals are shown as bold letters. The sequence rich in stop codons is underlined. The deduced amino acid sequence of the ORFI is shown as single letter code below the nucleotide sequence. The basic amino acids, taking part of the sequence similar to the motif found in the DNA-binding domain of HIV-1 integrase, are underlined and in boldface. Numbering of the nucleotide and amino acid sequences are shown on the left and right margins respectively.

internal promoters located in both strands,[8,21] as is the case with primate and rodent LINEs.[21,22] As we did not detect the full-size ZLRS RNA, we favoured as the source for ORFI RNA the transcription from an internal promoter *versus* the spliced product of a genomic RNA.

Flanking sequences of ORFI show *cis*-elements typical of eukaryotic genes such as CCAAT and TATA consensus boxes and putative polyadenylation signals at the 3′ end (Fig. 3). A peculiar feature observed in the 3′ region is the presence of an array of stop codons (Fig. 3, underlined) with unknown function. Moreover, the ORF1 translation initiation context, 5′-CCACCATGG-3′, perfectly matches the consensus for eucaryotic genes, 5′-CCA/GCCATGG-3′.[23] All these features suggest that ORFI may be transcribed from its own promoter within the ZLRS element. Retroelement internal promoters have been described for *Drosophila* poly(A)-type retrotransposons,[21] LINE-1 elements[7] and *D. hydei* micropia retrotransposon.[11]

The estimated codon usage frequency for ORFI is close to that compiled from 129 genes of maize.[24] On the other hand, the average codon frequency of ORF1 determined by the CodonPreference program of the GCG software package (1,0), based on the codon usage of maize genes, was significantly higher than the value for a random sequence (0,82). On the contrary, the average codon frequencies for ORFII, ORFIII and ORFIV are very close to the values obtained for the random sequence, indicating that is less likely that those ORFs were translated.

ORFI (the initial ATG codon of which is located at position 2152 of the genomic clone 7) shows a high G+C content (69%), typical of maize genes (Fig. 3). The ORFI deduced protein (157 residues) is rich in glycine (22%), serine (19%) and aspartic and glutamic acids (16 and 11%). In spite of the absence of significant similarity between ORFI-encoded protein and known proteins, there is a stretch of basic amino acids at the C-terminal part of the protein (Fig. 3, bold underlined) resembling the motif $KX_3KX_3KX_4RX_3RX_2RX_4KX_3KX_3K$, found in the C-terminal part of the retroviral HIV-1 integrase, which is the region containing the DNA binding activity of the endonuclease.[25] In addition, the very high acid amino acid content of the predicted protein, as well as the asymmetrical distribution of basic and acidic amino acids, are features found in the plant HMG1/2 family of proteins, which are known to bind DNA.[26] In summary, we have found indirect evidence suggesting that the ORFI transcript may be translated. First, the codon usage frequency is very close to the average of maize genes; second, ORFI transcripts are polyadenylated[12] as are the majority of mRNAs coding for protein in eukaryotes.

3.3. Clusters of tandem repeats

Downstream of ORFI (position 2930) there is an array of tandem repeats, termed tandem 1, composed of 9 repeats 81 bp long with more than 90% identity on average among them (Fig. 4A). Interestingly, every repeat is composed of two subunits 40–41 bp long with more than 60% similarity between them. At position 7387 there is another tandem, termed tandem 2, composed of 5 repeats 89 bp long, with more than 95% identity on average (Fig. 4A). Tandem 1 and tandem 2 do not show sequence similarity, although they share the presence of 7 stop codons per repeat (Fig. 4A, boldface letters). This number of stops codons is higher than expected for a random sequence (double) and their role, if any, is unknown.

Southern hybridization analysis of genomic DNA from maize with probes encompassing the tandems of Teosinte DNA show the presence of both tandems in the majority ZLRS elements. This is a conclusion based on the intensity of bands (48 hours exposure time) and the absence of any band other than those found at the expected position for tandems (Fig. 5). Interestingly, a ladder pattern is observed on tandem 2, indicating that different ZLRS elements have different numbers of repeats in tandem 2. and a large band is observed in tandem 1. The majority of ZLRS have between 8 and 13 (648–1052 bp) copies for tandem 1, and 4 to 8 (356 to 712 bp) copies for tandem 2 (Fig. 5). The absence of prominent bands hybridizing with tandem-containing probes (apart from the expected bands) even at longer exposures (not shown), indicates

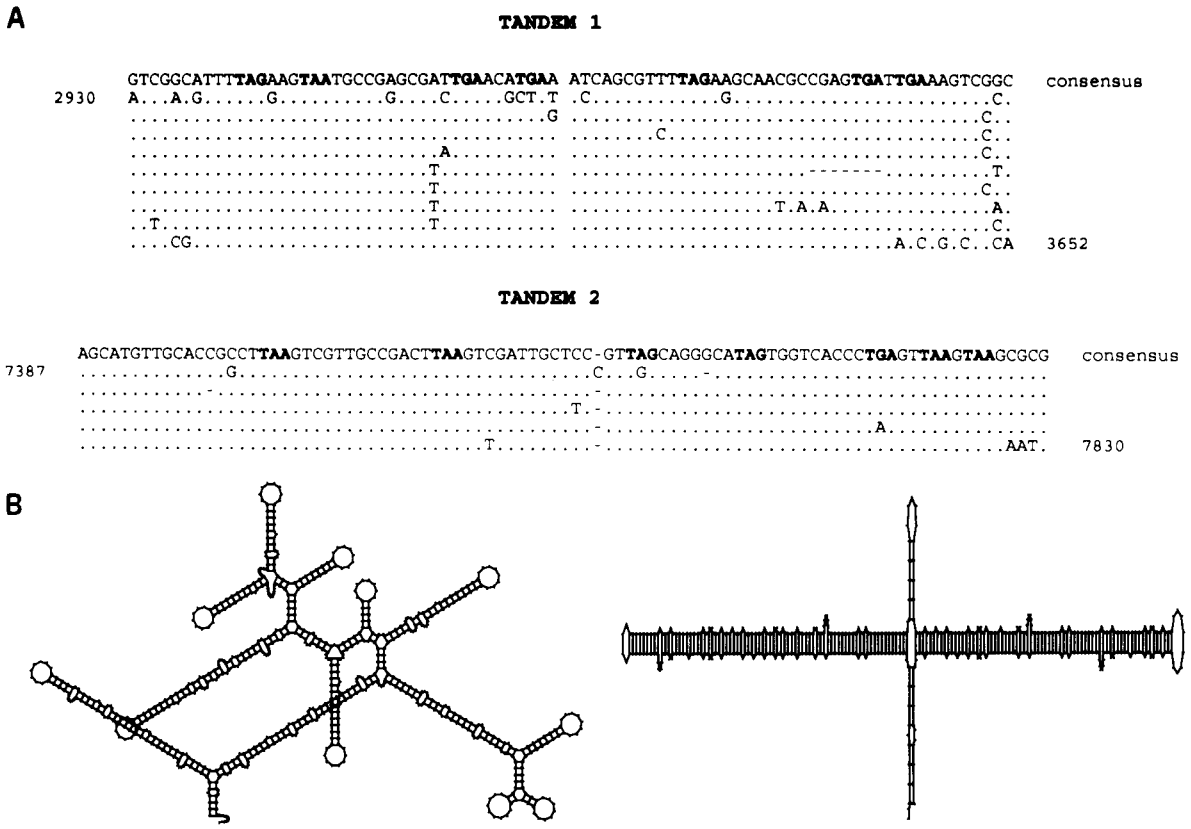

Figure 4. Tandem repeats alignment and secondary structures of tandems. A) Sequence alignment of the repeats of both tandem 1 and tandem 2 with the consensus sequences on the top of each alignment. The consensus sequences have been generated using in each position the nucleotide present in at least 50% of the repeats. Dots represent consensus nucleotides and only nucleotides which differ are shown. Dashes represent the absence of a nucleotide at that position. Trinucleotides corresponding to stop codons are shown as bold letters. B) Secondary structures generated by the FOLD program of the GCG package,[16] corresponding to the strand of DNA shown in A for tandem 1 (left) and tandem 2 (right) complete sequences.

the absence of comparable tandem arrays in the genome, outside ZLRS elements. Note that not all the bands correspond exactly to the size predicted from the restriction map shown, which is for *Zea diploperennis* ZLRS 7. This may be due to *Mva* I restriction fragment length polymorphism among ZLRS elements, especially considering the fact that DNA is from BMS maize and not from *Z. diploperennis*.

Tandem arrays of repeats have also been described in some transposable retroelements like TOC1-defective retrotransposon,[6] *Trypanosoma brucei gambiense* SLACs and mammalian LINEs[9,27] and *Drosophila* micropia retrotransposon.[11] Micropia-DhMi2 from *D. hydei* is the only case besides ZLRS with two tandem arrays of repeats. The two tandems are very close to each other and proximal to the 3′-LTR. The position of the tandems in the other elements is either at the 5′ end of the element (mammalian LINEs) or very close to it (TOC1 and SLACS elements). No defined functions have been attributed to transposable retroelement tandems, although their conserved location among elements from the same family and their high sequence conservation suggest a function for them. The sequence conservation among repeats, once the tandem was established, is probably due to an unequal crossing-over mechanism,[28] besides the functional constriction that may have operated.

No significant sequence similarity has been found between ZLRS tandems and any other described to date. An interesting structural peculiarity is the presence of short direct and inverted repeats inside the individual repeats of the two tandems. The inverted repeats could form potential stable hairpin structures for some of them. The most stable structures have 8-8 bases for the stem-loop and −11.8 kcal of free energy (6 in total for tandem 1), and 9-4 bases for the stem-loop and −11.2 kcal of free energy (5 in total for tandem 2). Surprisingly, the Fold program of GCG package using the whole tandem as a query sequence predicted important secondary structures for tandem 1 and especially for tandem 2 (Fig. 4B). The corresponding free energies were −202.1 kcal (T1, 723

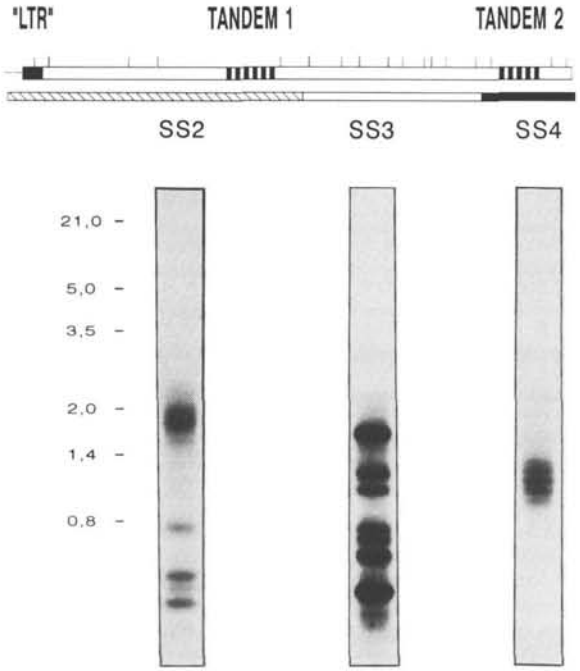

Figure 5. ZLRS Southern blot analysis. Vertical lines over the ZLRS schematic drawing represent *Mva*I restriction sites. Genomic DNA from Black Mexican Sweet maize leaves was isolated as described.[32] The DNA (4 μg per lane) was digested to completion with a restriction enzyme which does not cut internally to the two tandem arrays of repeats (*Mva*I), fractionated by electrophoresis in 0.8% agarose gel and transferred to Nytran-N membranes (Schleicher and Schuell) according to the manufacturer's instructions. Hybridizations with SS2, SS3 and SS4 radioactively labelled probes were performed as described in Material and Methods. After hybridization and washing, the filters were exposed to films (Agfa Curix RP2) for 48h at $-70°C$ with intensifying screens. The size of molecular weight markers in the left margin are indicated in kilobase pairs. Bands smaller than 270 bp are not shown in the figure.

bp) and -145 kcal (T2, 446 bp). The stem-loop structure of tandem 2 (Fig. 4B, right) resembles the structures of *tourist* and *stowaway* inverted repeat elements.[19,29] These two families of elements are associated with genes of maize (*tourist*) and monocotyledonous and dicotyledonous genes (*stowaway*). Other elements showing potential hairpin structures are the foldback (FB) elements of *Drosophila*.[30]

The five best stem-loop structures of tandem 2 generated by the program Multifold of the GCG package are different (not shown) and the increment of free energy between the most and less probable ones is very low (0.7 kcal). The corresponding increment of free energy between the most and less probable stem-loop structures of the five best ones for tandem 1 is higher (3.0 kcal). Therefore, the probability that stem-loop structures corresponding to ZLRS tandem repeats can actually occur is relatively high, in particular for tandem 2.

A possible role of the ZLRS tandem repeat stem-loop structures could be to pause and eventually to terminate transcription, as has been suggested for stem-loop structures in eukaryotic gene transcription termination among other arrangements.[31]

Concerning the origin of ZLRS tandem repeats, one can envisage that repeats constituting the tandem must have been generated by saltatory replication of the master copy within a ZLRS followed by amplification. Alternatively, ZLRS tandem repeats could have been inverted repeated or FB-like elements inserted in a preexisting transposable retroelement that became integrated in it and later amplified as a new element.

In summary, we have presented evidence suggesting that ZLRS elements can be a new type of transposable retroelement with a composite structure. Indeed, ZLRS elements share properties with retrotransposons (LTR-like sequences), with inverted repeat elements and FB elements (stem-loop structures), and with mammalian LINEs (tandem repeats at one end of the element). The absence of clear target duplications flanking ZLRS7, makes it difficult to define the second end of the element. The possibility that ZLRS7 is a truncated element, as has been proposed for the majority of mammalian LINEs[7] can not be discounted. A common characteristic of ZLRS and the above mentioned transposable elements is the presence of ORFs. Among the ZLRS ORFs, only ORFI is predicted to be translated. If this were the case, the ORFI protein might have DNA-binding activity, as can be inferred from the resemblance of the ORFI deduced amino acid sequence to proteins with such activity. More experiments are needed to demonstrate this point. Finally, more experiments would be needed to demonstrate that ZLRS elements are active transposable retroelements.

Acknowledgements: We thank Mrs. A. Pons (CID-Automatic Sequencing Service) for technical support in DNA sequencing. We are indebted to Dr. J. M. Casacuberta for critical reading of the manuscript and to Dr. S. Jackson for carefully revising it. This work has been supported by grants from DGPC (PB90-0129 and PB93-0043) to J.A.M.I. A.M. was the recipient of a postdoctoral reincorporation fellowship (CIRIT). C.M.V. was the recipient of an FPI fellowship (CIRIT). R.R. was the recipient of a PFPI fellowship (MEC).

References

1. Schweizer, D., Strehl, S., and Hagemann, S. 1990, Plant repetitive DNA elements and chromosome structure, In Chromosomes today (Fredga, K., Kihlman, B. A., Bennett, M. D. eds) Vol. 10, pp. 33–44, Unwin Hyman, London.
2. McClintock, B. 1951, Chromosome organization and genic expression, Cold Spring Harbor Symposium of

Quantitative Biology, **16**, 13–47.
3. Temin, H. M. 1989, Reverse transcriptases. Retrons in bacteria, *Nature*, **339**, 254–255.
4. Grandbastien, M. A. 1992, Retroelements in higher plants, *Trends Genet*, **8**, 103–108.
5. Weiner, A. M., Deininger, P. L., and y Efstratiadis, A. 1986, Nonviral retroposons: genes, pseudogenes, and transposable elements generated by the reverse flow of genetic information, *Annu. Rev. Biochem.*, **55**, 631–661.
6. Day, A. and Rochaix, J. D. 1991, A transposon with an unusual LTR arrangement from Chlamydomonas reinhardtii contains an internal tandem array of 76bp repeats, *Nucleic Acids Res.*, **19**, 1259–1266.
7. Thayer, R. E., Singer, M. F., and Fanning, T. G. 1993, Undermethylation of specific LINE-1 sequences in human cells producing a LINE-1-encoded protein, *Gene*, **133**, 273–277.
8. Varmus, H. and Brown, P. 1989, Retroviruses, In Mobile DNA (Berg, D. E. and Howe, M. M. eds), Washington, DC. Am. Soc. Microbiol., pp. 53–108.
9. Lloyd, J. A. and Potter, S. S. 1988, Distinct subfamilies of primate L1Gg retroposons, with some elements carrying tandem repeats in the 5′ region, *Nucleic Acids Res.*, **16**, 6147–6156.
10. Xiong, Y., Burke, W. D., and Eickbush, T. H. 1993, Pao, a highly divergent retrotransposable element from Bombyx mori containing long terminal repeats with tandem copies of the putative R region, *Nucleic Acids Res.*, **21**, 2117–2123.
11. Lankenau, D. H. 1993, The retrotransposon family micropia in *Drosophila* species, In Transposable elements and Evolution (McDonald, J. F. ed.), Kluwer Academic Publishers, Netherlands, pp. 232–241.
12. Raz, R., Puigdomènech, P., and Martinez-Izquierdo, J. A. 1991, A new family of repetitive nucleotide sequences is restricted to the genus *Zea*, *Gene*, **105**, 151–158.
13. Aledo, R., Raz, R., Monfort, A., Vicient, C. M., Puigdomènech, P., and Martinez-Izquierdo, J. A. 1995, Chromosome localization and characterization of a family of long interspersed repetitive DNA elements from the genus Zea, *Theor. Appl. Genet.*, **90**, 1094–1100.
14. Sambrook, J., Fristch, E. F., and Maniatis, T. 1989, Molecular Cloning: a Laboratory Manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
15. Sanger, F., Nicklen, S., and Coulson, A. R. 1977, DNA sequencing with chain-terminating inhibitors, *Proc. Natl. Acad. Sci. USA.*, **74**, 5463–5467.
16. Devereux, J., Haeberli, P., and Smithies, O. 1984, A comprehensive set of sequence analysis programs for the VAX, *Nucleic Acids Res.*, **12**, 387–395.
17. Shepherd, N. S., Schwarz-Sommer, Z., Blumnerg vel Spalve, J., Gupta, M., Wienand, U., and Saedler, H. 1984, Similarity of the Cin1 repetitive family of *Zea* mays to eukaryotic transposable elements, *Nature*, **307**, 185–187.
18. Rothnie, H. M., McCurrach, K. J., Glover, L. A., and Hardman, N. 1990, Retrotransposon-like nature of Tp1 elements: implications for the organisation of highly repetitive, hypermethylated DNA in the genome of Physarum polycephalum, *Nucleic Acids Res.*, **19**, 279–286.
19. Bureau, T. E. and Wessler, S. R. 1994, Stowaway: a new family of inverted repeats with the genes of both monocotyledonnous and dicotyledonous plants, *Plant Cell*, **6**, 907–916.
20. Johns, M. A., Babcock, M. S., Fuerstenberg, S. M., Fuerstenberg, S. I., Freeling, M., and Simpson, R. B. 1989, An unusually compact retrotransposon in maize, *Plant. Mol. Biol.*, **12**, 633–642.
21. Boeke, J. D. and Corces, V. G. 1989, Transcription and reverse transcription of retrotransposons, *Annu. Rev. Microbiol.*, **43**, 403–434.
22. Loeb, D. D., Padgett, R. W., Hardies, S. C., Shehee, W. R., Comer, M. B., Edgell, M. H., and Hutchinson, C. A. 1986, The sequence of a large L1Md element reveals a tandemly repeated 5′ end and several features found in retrotransposons, *Mol. Cell. Biol.*, **6**, 168–182.
23. Kozak, M. 1984, Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs, *Nucleic Acids Res.*, **12**, 857–872.
24. Wada, K., Wada, Y., Ishibashi, F., Gojobori, T., and Ikemura, T. 1992, Codon usage tabulated from the GenBank genetic sequence data, *Nucleic Acids Res.*, **20**, 2111–2118.
25. Woerner, A. M. and Marcus-Sekura, C. J. 1993, Characterization of a DNA binding domain in the C-terminus of HIV-1 integrase by deletion mutagenesis, *Nucleic Acids Res.*, **21**, 3507–3511.
26. Grasser, K. D. 1995, Plant chromosomal high mobility group (HMG) proteins, *Plant. J.*, **7**, 185–192.
27. Aksoy, S., Williams, S., Chang, S., and Richards, F. F. 1990, SLACS retrotransposon from Trypanosoma brucei gambiense is similar to mammalian LINEs, *Nucleic Acids Res.*, **18**, 785–792.
28. Weiner, A. M. and Denison, R. A. 1982, Either gene amplification or gene conversion may maintain the homogeneity of the multigene family encoding human U1 small nuclear RNA, *Cold Spring Harbor Symp. Quant. Biol.* **47**, 1141–1149.
29. Bureau, T. E. and Wessler, S. R. 1992, Tourist: a large family of small inverted repeat elements frequently associated with maize genes, *Plant Cell*, **4**, 1283–1294.
30. Smith, P. A. and Corces, V. G. 1991, Drosophila transposable elements: mechanisms of mutagenesis and interactions with the host genome, *Adv. Genet.*, **29**, 229–300.
31. Proudfoot, N. J. 1989, How RNA polymerase II terminates transcription in higher eukaryotes, *Trends Biochem. Sci.*, **14**, 105–110.
32. Dellaporta, S. L., Wood, J., and Hicks, J. B. 1983, A plant DNA minipreparation: version II, *Plant. Mol. Biol. Rep.*, **1** (4), 19–21.