

Terminal-repeat retrotransposons in miniature (TRIM) are involved in restructuring plant genomes

Claus-Peter Witte*[†], Quang Hien Le*[‡], Thomas Bureau*[‡], and Amar Kumar*^{§5}

*Scottish Crop Research Institute, Invergowrie, DD2 5DA Dundee, Scotland; and [‡]Department of Biology, McGill University, 1205 Dr. Penfield Avenue, Montreal, QC, Canada H3A 1B1

Edited by Ronald L. Phillips, University of Minnesota, St. Paul, MN, and approved September 24, 2001 (received for review July 5, 2001)

A new group of long terminal repeats (LTR) retrotransposons, termed terminal-repeat retrotransposons in miniature (TRIM), are described that are present in both monocotyledonous and dicotyledonous plant. TRIM elements have terminal direct repeat sequences between ≈ 100 and 250 bp in length that encompass an internal domain of ≈ 100 –300 bp. The internal domain contains primer binding site and polypurine tract motifs but lacks the coding domains required for mobility. Thus TRIM elements are not capable of autonomous transposition and probably require the help of mobility-related proteins encoded by other retrotransposons. The structural organization of TRIM elements suggests an evolutionary relationship to either LTR retrotransposons or retroviruses. The past mobility of TRIM elements is indicated by the presence of flanking 5-bp direct repeats found typically at LTR retrotransposon insertion sites, the high degree of sequence conservation between elements from different genomic locations, and the identification of related to empty sites (RESites). TRIM elements seem to be involved actively in the restructuring of plant genomes, affecting the promoter, coding region and intron-exon structure of genes. In solanaceous species and maize, TRIM elements provided target sites for further retrotransposon insertions. In *Arabidopsis*, evidence is provided that the TRIM element also can be involved in the transduction of host genes.

mobile elements | transduction | nonautonomous | genome evolution | sequence analysis

Transposons, also known as mobile elements or transposable elements, are discrete genetic sequences found in nearly all eukaryotic genomes. As their name indicates, transposons have the ability to move from one location of the genome to another, generating target site duplications (TSDs) after insertion. Autonomous transposons encode the enzymes necessary for their mobilization, whereas nonautonomous elements require these proteins to be provided in trans by an autonomous element (reviewed in ref. 1).

Retrotransposons are a class of transposable element that moves via an RNA intermediate that is reverse-transcribed into extrachromosomal DNA and inserted into the genome by their encoded reverse transcriptase, RNaseH and integrase enzymes (reviewed in refs. 2 and 3). This replicative mode of transposition is similar to that used by retroviruses. However, unlike the retroviruses, which are found usually in animals, retrotransposons are found in all eukaryotes, where they are the most abundant class of mobile DNA (reviewed in refs. 4 and 5). Retrotransposons are classified into two types: those with long terminal repeats (LTRs) and those without LTRs (non-LTR retrotransposons). LTR retrotransposons are subclassified further into the *Ty1-copia* and *Ty3-gypsy* groups, which differ from each other in distinct sequence features and in the order of encoded gene products (6, 7). Retrotransposons make up a large portion of many eukaryotic genomes including humans (8, 9) and a variety of plants (4, 10, 11).

Many properties of transposable elements suggest that they are parasitic or selfish DNAs (12, 13). However, like any other component of a heritable genome, transposon DNA can serve as

a genomic resource for mutation and natural selection. Indeed, a more recent paradigm suggests that transposable elements may play a central role in the evolution of gene function and genome structure in eukaryotic organisms (for reviews see refs. 1, 4, 5, and 14–16). For example, transposons can contribute regulatory cis-factors (17, 18), alter transcript splicing (19, 20), promote exon shuffling (21), and facilitate chromosomal rearrangements or restructuring (22, 23), leading to changes in spatial and temporal expression patterns of host genes or even the generation of novel genes. Furthermore, retrotransposons are involved in altering the genome size of eukaryotic organisms either by increasing (10, 11) or decreasing (24–26) their copy numbers within the host genome.

The properties and abundance of transposable elements make it important to identify and characterize the whole spectrum of mobile elements that inhabit the eukaryotic genomes and to investigate their origin, interactions with host genomes, and contributions to the evolution of host genes and genomes. In this report a unique group of mobile retrotransposons is described. We provide evidence that these elements have some features typical of LTR retrotransposons but are unusually small and lack coding capacity for mobility-related proteins. We also demonstrate that they are present in both dicotyledonous and monocotyledonous plants and have been active in restructuring plant genomes.

Materials and Methods

Data Analysis. Terminal-repeat retrotransposons in miniature (TRIM) elements were identified initially during sequence analysis of a genomic clone containing the *Solanum tuberosum* (potato) urease gene (C.-P.W., S. Tiller, H. V. Davies, and A.K., unpublished data) and during the analysis of the genome sequence of *Arabidopsis*, where they were referred to as *Katydid* (27).

In potato, a DOTPLOT (28) revealed closely spaced direct repeats within intron 6 of the urease gene. Closer inspection indicated that the direct repeats were flanked by a 5-bp TSD and contained a primer binding site (PBS) and polypurine tract (PPT) within the region between the repeats. Initial DNA-database searches with the similarity search algorithm BLAST (29) using the complete element as the query revealed the presence of similar sequences at several genomic locations in a number of solanaceous species. Independently, TRIM elements

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: TSD, target site duplications; LTR, long terminal repeat; TRIM, terminal-repeat retrotransposons in miniature; PBS, primer binding site; PPT, polypurine tract; RESite, related to empty site; TDR, terminal direct repeat; EST, expressed sequence tag; gi, NCBI GenInfo identifier; NMD, nonsense-mediated mRNA decay.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. AJ276858, AJ276864, and AJ409162).

[†]C.-P.W. and Q.H.L. contributed equally to this work.

^{§5}To whom reprint requests should be addressed. E-mail: akumar@sri.sari.ac.uk.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

were identified during a survey of transposons in the *Arabidopsis* genome as part of the *Arabidopsis* Genome Initiative annotation effort (27). Iterative database analyses (with BLAST) by using TRIM sequences from previous searches as queries for new searches identified TRIM elements from several monocotyledonous and dicotyledonous plant species. In this search strategy all hits against the database were examined manually for typical features of TRIMs (e.g., a sequence of less than 540 bp containing direct repeats enclosing PBS and PPT motifs). Therefore, elements with some degree of truncation also were found for which the termini could not be mapped accurately because of deletions and/or sequence degeneracy. Sequence information of large insert clones was accessed from public sequence databases (GenBank, www.ncbi.nlm.nih.gov/GenBank/index.html, and the European Molecular Biology Laboratory, www.ebi.ac.uk/embl/index.html) between October 2000 and May 2001. Sequence analysis was performed by using programs of the UWGCG software package (version 10, University of Wisconsin Genetics Computer Group, www.gcg.com) or the EMBOSS program package (European Molecular Biology Open Software Suite, www.uk.emblnet.org/Software/EMBOSS/). Multiple alignments were generated with PILEUP (UWGCG) or CLUSTALW (30). Graphical manipulations were carried out by using GENEDOC (31). TRIM elements were mapped onto the assembled genomic sequences of *Arabidopsis* accessed at the *Arabidopsis* Genome Initiative (released in March 2001, www.Arabidopsis.org/agi.html).

Related to Empty Sites (RESites) Analysis. Sequences that were similar to the region immediately flanking an insertion but that did not harbor the insertion and had only one copy of the target site were defined as RESites (32). RESites can be used to delimit the termini of the insertion, confirm the size and sequence of the TSD, and provide supporting evidence for past mobility. RESites were identified by using the sequences immediately flanking an insertion as a query in a BLAST search as described (32).

Results

Identification, Characterization, and Distribution of TRIM Elements in Plants. Database searches and sequence analyses revealed the presence of a family of repetitive sequences from a number of plant species that share a unique set of characteristics. Specifically, these elements are short in size (mostly ≈ 350 bp), have terminal direct repeats (TDRs) of 100–250 bp (on average < 140 bp) that flank an internal domain containing a PBS (located immediately downstream of the 5' terminal repeat and complementary to the methionine tRNA), and a PPT (located immediately upstream of the 3' terminal repeat; Fig. 1A). None of the elements characterized to date encodes mobility-related proteins in the internal domain (Fig. 1). Intact elements are flanked by 5-bp direct repeats, representing the TSD that was generated after insertion. These previously uncharacterized mobile elements have been named TRIM (Fig. 1). TRIM elements are present in multiple copies and are dispersed within the host chromosomes (shown in Table 2, which is published as supporting information on the PNAS web site, www.pnas.org). TRIM elements seem ubiquitous in the plant kingdom as they have been found in both dicotyledonous and monocotyledonous species (Table 1). To date, TRIM elements from the plant species examined have similar overall lengths and structures and show some sequence similarity, especially in the TDRs (Figs. 1 and 2). For example, there is over 80 and 90% sequence identity among the TDRs of TRIM elements in *Arabidopsis* and rice, respectively. Furthermore, TDRs of TRIM elements from distantly related species show significant similarity (Fig. 2) ranging from 60 to 75% at the sequence identity level. In brief, TRIM elements are distinguished by their exceptionally small overall

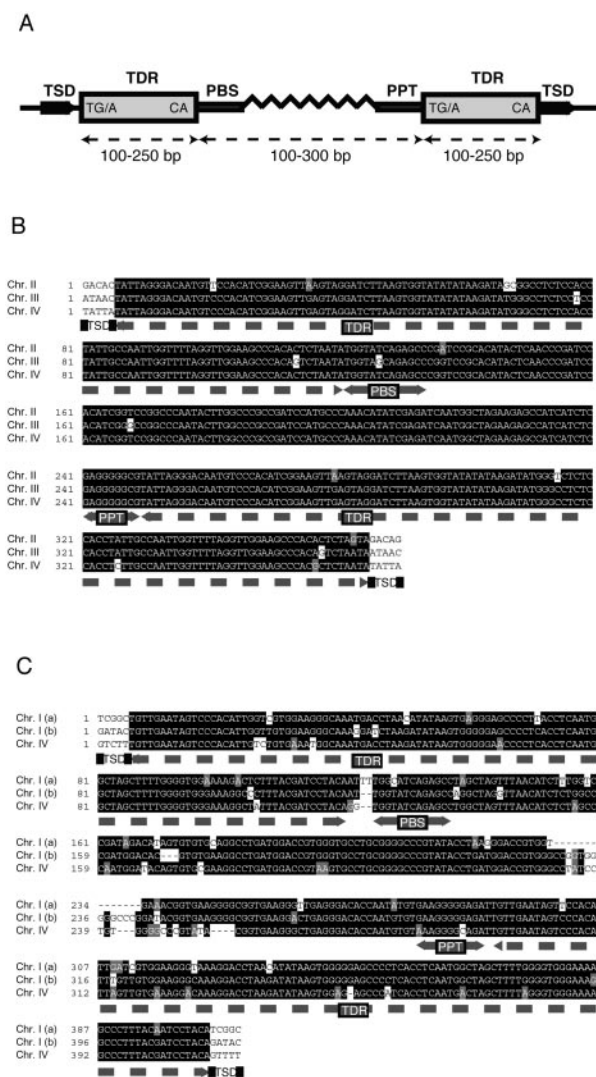


Fig. 1. Schematic diagram of the general structure of TRIM elements. (A) The archetypal TRIM contains the following sequence features: TSD, TDRs (flanking shaded boxes), PBS, and PPT. Start and end bases of TDRs are given in the flanking boxes. The possible length of different elements is given underneath (maximum length < 540 bp). (B) Multiple alignment of selected TRIM elements (*Katydid-At1*) elements from *A. thaliana*. The alignment was created with CLUSTALW by using default parameters. The BOXSHADE program was used for shading. Structural features (abbreviations as described for A) are marked underneath. The elements in chromosomes II (Chr. II, gi 6598569, position 33,850–33,481), III (Chr. III, gi 7629988, position 45,624–45,993), and IV (Chr. IV, gi 3309259, position 95,888–95,519) are shown. (C) Multiple alignment of selected TRIM elements from rice (*Oryza sativa*). Alignment and abbreviations are as described above. The elements from chromosomes I (Chr. I(a), gi 10800055, position 29,385–29,802), I (Chr. I(b), gi 9711848, position 74,966–74,558), and IV (Chr. IV, gi 5777612, position 77,595–77,183) are shown.

size (< 540 bp) and presence in multiple copies in different locations within the host genome with highly conserved TDRs and short internal sequences (≈ 100 – 300 bp). Moreover, despite the complete lack of coding capacity for mobility-related proteins (which makes it impossible to classify TRIM elements conventionally into Ty1-*copia* or Ty3-*gypsy* groups), they are able to transpose in trans, making them the smallest active LTR retrotransposons known to date.

In *Arabidopsis thaliana*, a total of 43 TRIM elements were identified (Table 1) and subdivided into three groups on the basis of overall sequence similarity: *Katydid-At1*, -*At2*, and -*At3*.

Table 1. TRIM elements found in plant sequences

Organism	Intact*	Solo LTR	Truncated†	Total	EST‡
Brassicaceae					
<i>Arabidopsis thaliana</i>	21	5	17	43	1
Solanaceae					
<i>Solanum tuberosum</i>	3	2	3	8	4
<i>Solanum ochranthum</i>	1	0	0	1	0
<i>Lycopersicon pennellii</i>	0	0	1	1	1
<i>Lycopersicon esculentum</i>	3	0	7	10	3
<i>Nicotiana tabacum</i>	1	0	0	1	0
Fabaceae					
<i>Lotus japonicus</i>	0	0	2	2	1
<i>Medicago sativa</i>	1	0	0	1	0
<i>Medicago trunculata</i>	0	7	5	11	9
<i>Phaseolus vulgaris</i>	1	1	0	2	0
<i>Glycine max</i>	0	0	1	1	1
Poaceae					
<i>Oryza sativa</i>	8	4	4	16	1
<i>Zea mays</i>	1	0	1	2	0

*Elements with small internal deletions were counted as intact.

†Elements for which one or both termini could not be resolved.

‡Each EST represents a TRIM insertion in a different gene (duplicate ESTs for the same gene were not counted).

Thirty-one copies of *Katydid-At1* are present including 12 truncated elements (i.e., termini could not be mapped accurately). *Katydid-At2* is present in seven copies including three truncated elements, whereas there are five copies of *Katydid-At3*, of which two are truncated. *Katydid-At1* elements have TDRs of 116 bp (Fig. 1B). The left and right TDRs are nearly identical in many elements. Because LTRs are identical initially after element insertion, the high sequence identity between the TDRs suggests that *Katydid-At1* elements were recently, and are perhaps still, active (33). The fact that three RESites were identified (Fig. 3) and the high nucleotide similarity between some *Katydid-At1* elements (up to 98%, Fig. 1B) also attest to recent mobility. In contrast, similarity between members of the TDRs of *Katydid-At2* or *-At3* was lower, indicating that they may be ancient insertions (Table 2). *Katydid-At2* elements have TDRs similar to *Katydid-At1* in size (≈ 115 bp) but have only limited similarity in sequence (Fig. 2). Although one *Katydid-At3* element shared structural characteristics with other TRIM elements, two members had unusual features (Table 2). For example, one element (gi 4263586) has terminal inverted repeats rather than TDRs.

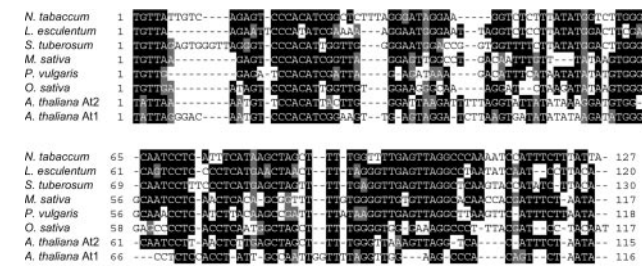


Fig. 2. Multiple alignment of selected 5' TDR sequences of TRIM elements from different species. Sequences from *Nicotiana tabacum* (gi 9392606, position 3,821–3,947), *Lycopersicon esculentum* (gi 4220970, position 38–157), *S. tuberosum* (gi14599414, position 1,522–1,651), *M. sativa* (gi 19642, position 8,088–8,204), *P. vulgaris* (gi 2576326, position 1,878–1,761), *O. sativa* (gi 10800055, position 29,390–29506), *A. thaliana Katydid-At2* (gi 7649355, position 68,877–68,994), and *A. thaliana Katydid-At1* (gi 6598490, position 7722–7607) are shown. The alignment was created with CLUSTALW (gap opening parameter, 10; gap extension parameter, 1). The program BOXSHADE was used for the shading.



Fig. 3. Identification of RESites for TRIM elements (*Katydid-At1*). Sequences harboring the *Katydid-At1* insertion (depicted by the black boxes with arrowheads) are shown above the corresponding RESites. The positions on clone and gi numbers are indicated. Target sequences and TSDs are underlined. Insertion in gi 3702730 is a solo LTR.

Because a perfect 5-bp flanking TSD can be identified, it is possible that this unique element was mobilized with this structure. The other element is truncated at one end and yet still appears to have a 5-bp TSD. Structural rearrangements were observed also within other *Arabidopsis Katydid* elements. Although most *Katydid-At1* elements have TDRs, “solo” and “triple” direct repeat structures were identified (Fig. 6, which is published as supporting information on the PNAS web site). Solo and triple direct repeat elements are also immediately flanked by a 5-bp TSD. Solo direct repeats were generated most likely by illegitimate recombination between TDRs of an intact *Katydid-At1* element (10, 24, 25). The middle repeat of the triple direct repeats-containing element is unlikely to be a solo direct repeat, because flanking TSDs are not present. Within the triple repeats the TDRs share 93% with each other but showed lower (89 and 90%) similarity with the middle direct repeat (gi 3150006). It currently is unclear as to the precise mechanisms that generated these unusual TRIM elements.

Unlike other *Arabidopsis* retrotransposons, which cluster in the pericentromeric and centromeric regions where gene density is low, *Katydid-At1* is dispersed throughout the genome (Table 2). In addition to the elements mined from nuclear sequences, we have identified a *Katydid-At1* insertion in the mitochondrial genome of *Arabidopsis* (gi 1785729). Similar to most elements found in the mitochondrial genome of *Arabidopsis* (34), this *Katydid-At1* element is truncated and degenerate. Transposable elements make up $\approx 4\%$ of the total mitochondrial DNA and are predominately retrotransposon fragments that originated from reverse transcripts of nuclear genomic elements (34). Similarly, we consider that the mitochondrial *Katydid-At1* element is likely to be of nuclear origin, because it shares sequence similarity with nuclear insertions and no other such elements could be identified in mitochondrial genomes.

In rice, TRIM elements from different genomic locations are highly similar, having nearly identical TDRs and highly conserved internal domains (Fig. 1C). These features of TRIM elements attest to the recent activity of these elements in the rice genome. At the date of this analysis (March 2001), 16 rice TRIM elements could be found in the sequence database. There were eight complete and three truncated elements, and five were present only as solo LTRs (Tables 1 and 2). Similar to *Arabidopsis Katydid* elements, the rice elements also are located on a variety of chromosomes, and some of the insertions are near or within putative genes and overlap with predicted ORFs (e.g., gi 9711848 and 5777612 in Table 2). However, only one expression sequence tag (EST, coding for a protein with similarity to leucine-rich repeat-like proteins, gi 4715547) with partial similarity to a TDR of a rice TRIM was identified. In maize, only one complete TRIM element has been found to date. This element possesses unusually long TDRs of ≈ 230 bp and a short internal domain of 72 bp. However, similar to TRIM elements from other plants, this element contains a PBS and a PPT motif and 5-bp TSDs.

In contrast to *A. thaliana* and rice, only limited genomic

sequence data are available for members of the family of solanaceous and leguminous species (Table 1). In both groups ancient TRIM insertions were found (Table 2, gi 9392606) including truncated elements (Table 2, gi 19200 and 170430), solo LTRs (Table 2, gi 21590 and 8747823), and more complex structures, probably generated by several insertion-deletion events involving more than one element (Table 2, gi 9251206 and 10764221). However, TRIM elements with well conserved sequences were identified also (Table 2, gi 9562370 and 19642), indicating that TRIM elements have a long evolutionary history in these plants and that some still may be active.

Retrotransposons Nested Within TRIM Elements. The TRIM in intron 6 of one of the alleles of the urease gene from potato (*S. tuberosum*, cv. Désirée (gi 14599411) contains an insertion of a truncated LTR retrotransposon of the Ty1-*copia* group (Table 2). In the wild species *Solanum ochranthum*, the TRIM in intron 6 of urease sustained a long nuclear interspersed element insertion, a non-LTR retrotransposon 30 bp downstream of the Ty1-*copia* insertion found in potato (gi 14599445, Table 2). In both cases the insertion was located within the internal sequence of the TRIM. Another case for TRIM elements being targeted by Ty1-*copia* retrotransposons was found in the maize genome. In this case, a complete element of ≈ 9 kb inserted within the 3' TDR of a TRIM element (Table 2).

Insertion Within Genes. TRIM elements contribute to coding and untranslated regions of plant genes. However, in many cases this notion is based on the observation that TRIM elements overlap with sequences predicted to contain a gene by computer-assisted intron/exon assignment. Although experimental evidence confirming the presence of a gene at predicted locations is often still lacking, the frequent presence of TRIM elements in ESTs from species of the Solanaceae and Fabaceae (Tables 1 and 2) demonstrates that insertion into transcribed coding sequences is common at least in some species. In some cases, ESTs provided evidence for the insertion of TRIM elements into coding sequences with similarity to known genes, for example glycolate oxidase in tomato (gi 10902684), a nucleic acid binding domain/leucine-rich repeat-like protein in potato (gi 9562370), or a P450-like protein in *Medicago trunculata* (gi 10520041). In solanaceous and leguminous species TRIM elements were found also in promoter regions and introns of genes (Table 2). For example, in potato and *S. ochranthum* elements were found in intron 6 of the urease gene (gi 14599411, gi 14599414, and gi 14599445) in *Medicago sativa* in intron 9 of the nodulin-25 gene (gi 19642), in potato in the promoter region of the proteinase inhibitor II gene (gi 21553), and in *Phaseolus vulgaris* a TRIM solo terminal repeat was found in the promoter region of the chalcone isomerase gene (gi 20981, Table 2). In contrast, TRIM elements are found rarely in ESTs from *Arabidopsis* and rice, although they often overlap with computer-predicted genes (Table 2).

Two examples of *Katydid-At1* elements altering host genes in *Arabidopsis* are shown in Fig. 4. In the first case (Fig. 4A and C) a *Katydid-At1* element (gi 3510340, position 22,675–22,946) overlaps with the transcription and possible translation start site and the first exon-intron boundary of a gene encoding a cytochrome P450-like protein (gi 3510340, position 19,519–22,694). That the gene is transcribed and correctly spliced is indicated by the existence of two ESTs corresponding to this gene (gi 5841198 and 1053954). The insertion of *Katydid-At1*, probably at the 5' terminus of a preexisting gene, introduced a transcription start site (a TATA motif within the terminal repeat is present ≈ 40 bp upstream of transcription start). The element is truncated in the 3'-terminal repeat ≈ 20 bp downstream of the first exon-intron boundary. Thus, the first intron was either created *de novo*, or the 5' boundary of an already existing intron was replaced because

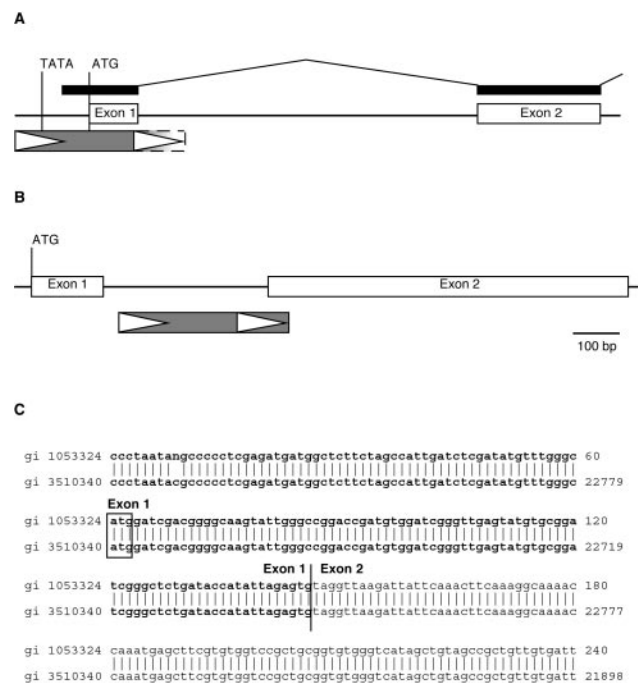


Fig. 4. Examples of TRIM (*Katydid-At1*) contributions to gene structures. The gray boxes with white arrowheads depict *Katydid-At1* elements. Open boxes represent exons. The predicted start of translation (ATG) and TATA box are indicated. (A) The *Katydid-At1* element in gi 3510340 (position 22,675–22,946) contributes sequences for the promoter, first exon, and first exon-intron boundary of a gene encoding a cytochrome P450-like gene. This insertion is truncated. Corresponding identical EST sequences (gi 5841198 and gi 1053954) are represented by thick black bars. Thin lines connecting thick black bars represent a putative intron that is not found in the EST. (B) The *Katydid-At1* found on gi 7209738 contributes a splice site to a predicted cytochrome P450 gene. The scale is as described for A. (C) Sequence alignment of *Katydid-At1* from A showing the putative translation start codon (boxed) and exon boundaries.

of the insertion of the element. Additionally, the potential use of a translation start site within the region between the two TDRs introduces 28 amino acids encoded by the element to the N terminus of the protein. Fig. 4B shows a case in which a *Katydid-At1* element appears to have created a new intron within a gene, although no EST data to support this hypothesis are available.

Transduction of Host Genes. A *Katydid-At1* element located in the *Arabidopsis* genome contains an ORF (gi 3241926) nearly identical to a putative cellular gene annotated as a nonsense-mediated mRNA decay (NMD) trans-acting factor but that actually is more similar to the yeast gene *sen1*. NMD is responsible for the degradation of prematurely terminated mRNAs (35), whereas SEN1 is involved in the endonucleolytic cleavage of pre-tRNAs (36). Both NMD and SEN1 contain a large domain involved in degrading ribonucleotides (35). The ORF linked to the *Katydid-At1* elements lacks part of the first and second exons and does not contain intron sequences (Fig. 5). This structure most likely was generated by the recombination of a *Katydid* element with an mRNA of NMD during the transposition process. The presence of this ORF within *Katydid-At1* is very reminiscent of the product of cellular gene transduction by retroviruses leading to the formation of oncoviruses (37). Previously, cellular gene transduction was documented for another plant retroelement, the maize *BsI* element (37–41). Although the cellular NMD (c-NMD) and transduced NMD (r-NMD, where r- refers to retrotransposon) both are located on chro-

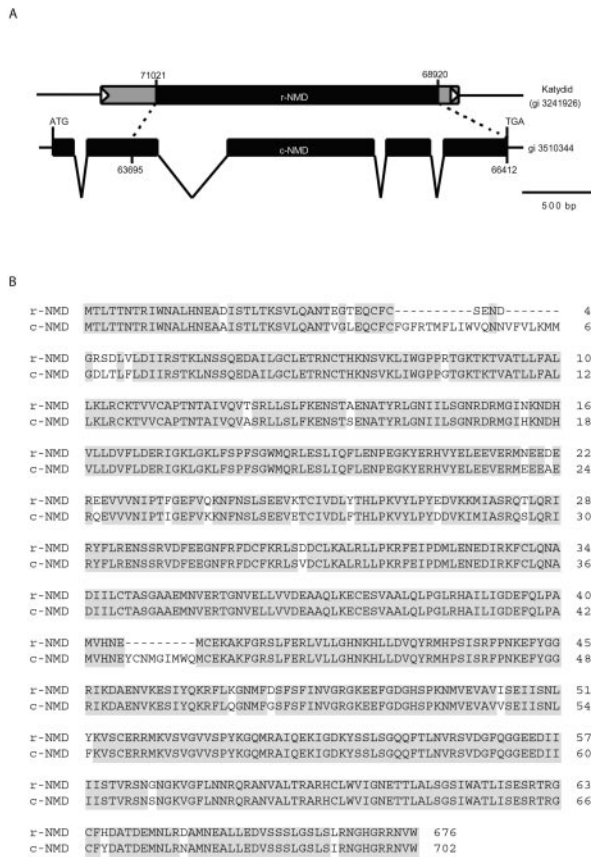


Fig. 5. TRIM (*Katydid-At1*) involvement in the transduction of a cellular gene. (A) Diagram illustrating the region of similarity (93%) between the cellular NMD gene (c-NMD) and the transduced version (r-NMD). The intron sequences on c-NMD (bent thin lines) are absent in r-NMD. Exons are represented by black boxes, and *Katydid-At1* sequences and LTRs are represented by gray boxes and arrowheads, respectively. The gi numbers and positions on clones are indicated. (B) Amino acid sequence alignment of r-NMD and c-NMD. Identical residues are shaded in gray.

mosome 5, they are not linked tightly, nor are there other *Katydid-At1* elements near c-NMD in the *A. thaliana* (Columbia ecotype) genomic sequence. This precludes the possibility that transduction resulted from a read-through transcript.

Discussion

Computer-assisted data mining of transposable elements has proven to be the most efficient and informative route to the understanding of not only mobile element evolution but also their role in eukaryotic gene and genome evolution. Although many mined elements reinforce traditional paradigms concerning sequence- and structure-based transposon classification, unusual elements occasionally are mined that force a re-evaluation of element type definition. TRIM elements are one such type of element. Their size (<540 bp) is considerably less than that of typical LTR retrotransposons (5–15 kb), which makes them the smallest TDR-containing mobile elements discovered to date. Despite their small size and lack of coding capacity, TRIM elements are obviously mobile as indicated by several observations: (i) the high sequence similarity between elements from different genomic locations (Fig. 1 B and C), (ii) the wide distribution within the individual genomes (Table 2), and (iii) the identification of RESites (Fig. 3).

Despite their small size, TRIM elements have some typical features of LTR retroelements: possessing TDRs, a 5-bp duplication of the target site sequence, and PBS and PPT motifs,

which are essential for the mobilization of LTR retrotransposons. Furthermore, the solo TDRs of TRIM elements are reminiscent of solo LTRs and presumably arose by a similar mechanism (10, 24, 25). Unlike LTR retrotransposons, however, TRIM elements lack mobility-related coding sequences such as *gag* and *pol* domains (Fig. 1; refs. 3 and 4). A full-length TRIM with coding capacity for mobility-related proteins has not yet been discovered, although the sequencing of an *Arabidopsis* genome is nearly complete. The lack of an autonomous *Arabidopsis* TRIM element may merely reflect ecotype distribution or stochastic loss. In any case, TRIM elements probably are mobilized in trans by other retroelements. The mobility of elements lacking coding capacity is well documented. For example, defective or nonautonomous DNA transposons (e.g., class II elements) can be mobilized by functional autonomous element counterparts (1). The short nuclear interspersed element *Alu* is the most abundant element in the human genome, yet similar to other short nuclear interspersed elements, it has no coding capacity (42). The *Bs1* elements of the Ty1-*copia* group retrotransposons in maize have been active recently but lack a *pol* domain (43). Recently human long nuclear interspersed elements were shown to mobilize transcripts not derived from long nuclear interspersed element sequences (44). Furthermore, several retroviral vectors have been designed that lack coding capacity and only transpose in the presence of mobility-related proteins in trans (45). The abundance of streamlined nonautonomous elements such as short nuclear interspersed elements and TRIMs in eukaryotic genomes may reflect a unique evolutionary relationship with their autonomous counterparts and the host genome.

A TRIM-like structure found in barley (gi 9623334) may have been generated by a combination of transposition and unequal recombination events among two identical LTR retrotransposons (25). Although this model would not necessarily account for the presence of PBS and PPT, it suggests a mechanism by which the TRIM progenitor might have evolved originally from LTR retroelements. In any case, because retroelement classification relies on internal domains such as *gag* and *pol* and no TRIM-related retroelement containing these domains could be identified, we cannot be certain whether TRIM elements are related to Ty1-*copia* or Ty3-*gypsy*-like retrotransposons or retroviruses or perhaps originated independently.

The presence of TRIM elements in dicotyledonous and monocotyledonous species indicates that these elements probably are ubiquitous in the plant kingdom. A comparison of the TDR sequences of TRIM elements from seven different plant species indicated significant sequence similarities (Fig. 2; also see *Results*). This significant sequence similarity among 5' TDRs is surprising given that LTRs typically share little sequence similarity between members of Ty1-*copia* or Ty3-*gypsy* elements in different plant species (see refs. 1 and 4 and references therein). The relative conservation of TDRs of TRIM elements probably reflects selective constraints resulting in the TDRs only containing the minimal set of cis-factors necessary for the retrotransposition process. The high level of conservation in structure and size of TRIM elements from a wide range of plant species together with clear indications of mobility strongly suggest that TRIM elements are a form of streamlined LTR retrotransposon.

TRIM Elements Contribute to the Restructuring of the Host Genomes.

Retrotransposons have been reported to contribute to the coding regions of genes in plants (see refs. 4 and 14 and references therein). Recent data from the Human Genome Sequence has also revealed many cases of *bona fide* and predicted proteins that contain L1 and *Alu* sequences (9, 46). Similar to other LTR retrotransposons, there are indications that TRIM elements have been involved intimately in reshaping genomes and in gene evolution. TRIM elements are

found dispersed throughout the genome in *Arabidopsis* and rice (Table 2). Because TRIM elements are comparatively small in size, it can be argued that they are less likely to cause serious disruptions of promoter and intron integrity compared with insertions of the much larger LTR retrotransposons. Alternatively, some LTR retrotransposons have been suggested to sustain deletions immediately after insertion near genes and thereby lessening the impact of transposition events (14, 19, 47). TRIM elements also might be able to contribute more effectively to ORFs of genes (Figs. 4 and 6), especially if the element itself already contains long ORFs (e.g., the *Katydid-At1* element on gi 3241926). Apart from the potential direct influence of TRIM elements on genes, they seem to act indirectly as target sites for both LTR and non-LTR retrotransposons in *Solanum* plants and maize (Table 2). It has been reported previously that LTR retrotransposons act as target sites for the insertions of other retrotransposons in maize (10, 33) and barley (24, 25). Nested insertions may reflect the propagation of these elements into genomic regions where retrotransposons already have been established and thus may limit the deleterious effects of retrotransposon activity on host fitness (4, 10).

The ability of TRIM elements to transduce a host gene was documented for the *Katydid-At1* element in *A. thaliana* (Fig. 5). Together with the maize *Bs1* element, these are the only plant examples of retrotransposon-mediated transduction, a phenomenon thought to be limited to the acquisition of proto-oncogenes in viruses (37–40). Recent data from the human genome-sequencing project revealed many examples of long nuclear

interspersed elements, specifically L1 elements, with transduced genomic sequences located in their 3'-flanking region (21, 46, 48, 49). Approximately 15–21% of these L1 elements harbor downstream sequences between 30 and 970 bp in length, meaning that $\approx 1\%$ of the human genome is the result of transduced sequences. Thus, transduction may account in part for the expansion of host genomes and may facilitate exon shuffling, and amplified gene sequences may serve as the raw material for the evolution of new genes.

In conclusion, TRIM elements are the smallest known LTR retrotransposons identified to date and are important players in plant genome evolution. Moreover, like other LTR retrotransposons, they have been involved actively in the restructuring of plant genomes by acting as target sites for retrotransposon insertions by altering host gene structure and by also being capable of transducing host genes. Although only plant TRIM elements have been reported here, it is anticipated that TRIM-like elements would be found in other eukaryotic organisms.

We thank Nabil Elrouby and John Jones for critical reading of an earlier version of this manuscript. A.K. and C-P.W. acknowledge the funding of the Scottish Executive & Environmental Rural Affairs Department (SEERAD), and C-P.W. also acknowledges the European Commission Training and Mobility of Researchers program. This work was funded in part by grants from the National Science and Engineering Research Council (NSERC) of Canada (to T.B.) and a David Stewart McGill Majors Fellowship (to Q.H.L.).

- Kunze, R., Saedler, H. & Lonnig, W. E. (1997) *Adv. Bot. Res.* **27**, 331–470.
- Boeke, J. D. & Corces, V. G. (1989) *Ann. Rev. Microbiol.* **43**, 403–434.
- Boeke, J. D. & Stoye, J. P. (1997) in *Retroviruses*, eds Varmus, H., Hughes, S. & Coffin, J. (Cold Spring Harbor Lab. Press, Plainview, NY), pp. 343–345.
- Kumar, A. & Bennetzen, J. L. (1999) *Annu. Rev. Genet.* **33**, 479–532.
- Kidwell, M. G. & Lisch, D. (2000) *Trends Ecol. Evol.* **15**, 95–99.
- Doolittle, W. F., Feng, D. F., Johnson, M. S. & McClure, M. A. (1989) *Q. Rev. Biol.* **64**, 1–29.
- Xiong, Y. & Eickbush, T. H. (1990) *EMBO J.* **9**, 3353–3362.
- Smit, A. F. (1999) *Curr. Opin. Genet. Dev.* **9**, 657–663.
- Li, W.-H., Gu, Z., Wang, H. & Nekrutenko, A. (2001) *Nature (London)* **409**, 847–849.
- SanMiguel, P., Tikhonov, A., Jin, Y. K., Motchoulskaia, N., Zakharov, D., Melake-Berhan, A., Springer, P. S., Edwards, K. J., Lee, M., Avramova, Z. & Bennetzen, J. L. (1996) *Science* **274**, 765–768.
- Pearce, S. R., Harrison, G., Li, D., Heslop-Harrison, J. S., Kumar, A. & Flavell, A. J. (1996) *Mol. Gen. Genet.* **250**, 305–315.
- Doolittle, W. F. & Sapienza, C. (1980) *Nature (London)* **284**, 601–603.
- Orgel, L. E. & Crick, F. H. C. (1980) *Nature (London)* **284**, 604–607.
- Wessler, S. R., Bureau, T. E. & White, S. E. (1995) *Curr. Opin. Genet. Dev.* **5**, 814–821.
- McDonald, J. F., Matyunina, L. V., Wilson, S., Jordan, I. K., Brown, N. J. & Miller, W. J. (1997) *Genetica (The Hague)* **100**, 3–13.
- Kazazin, H. H. & Moran, J. V. (1998) *Nat. Genet.* **19**, 19–24.
- Lister, C., Jackson, D. & Martin, C. (1993) *Plant Cell* **5**, 1541–1553.
- Yang, Z., Boffelli, D., Boonmark, N., Schwartz, K. & Lawn, R. (1998) *J. Biol. Chem.* **273**, 891–897.
- Marillonnet, S. & Wessler, S. R. (1997) *Plant Cell* **9**, 967–978.
- Brodbeck, D., Amherd, R., Callearts, P., Hintermann, E., Mayer, U. A. & Affolter, M. (1998) *Cell Biol.* **17**, 621–633.
- Moran, J. V., DeBerardinis, R. J. & Kazazin, H. H. (1999) *Science* **283**, 1530–1534.
- Agrawal, A., Eastman, Q. M. & Schatz, D. G. (1998) *Nature (London)* **394**, 744–751.
- Végh, Z., Vincze, E., Kadirov, R., Toth, G. & Kiss, G. B. (1990) *Plant Mol. Biol.* **15**, 295–306.
- Vicient, C. M., Suoniemi, A., Anamthawat-Johnson, K., Tanskanen, J., Beharav, A., Nevo, E. & Schulman, A. H. (1999) *Plant Cell* **11**, 1769–1784.
- Shirasu, K., Schulman, A. H., Lahaye, T. & Schulze-Lefert, P. (2000) *Genome Res.* **10**, 908–915.
- Petrov, D. A. (2001) *Trends Genet.* **17**, 23–28.
- Arabidopsis* Genome Initiative (2000) *Nature (London)* **408**, 796–815.
- Maizel, J. V. & Lenk, R. P. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 7665–7669.
- Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. (1997) *Nucleic Acids Res.* **25**, 3389–3402.
- Thomson, J. D., Higgins, D. G. & Gibson, T. J. (1994) *Nucleic Acids Res.* **22**, 4673–4680.
- Nicholas, K. B., Nicholas, H. B., Jr. & Deerfield, D. W. I. (1997) *EMBNEW NEWS* **4**, 14.
- Le, Q. H., Wright, S., Yu, Z. & Bureau, T. (2000) *Proc. Natl. Acad. Sci. USA* **97**, 7376–7381.
- SanMiguel, P., Gaut, B. S., Tikhonov, A., Nakajima, Y. & Bennetzen, J. L. (1998) *Nat. Genet.* **20**, 43–45.
- Unsel, M., Marienfeld, J. R., Brandt, P. & Brennecke, A. (1997) *Nat. Genet.* **15**, 57–61.
- Jacobson, A. & Peltz, S. W. (1996) *Annu. Rev. Biochem.* **65**, 693–739.
- DeMarini, D. J., Winey, M., Ursic, D., Webb, F. & Culbertson, M. R. (1992) *Mol. Cell Biol.* **12**, 2154–2164.
- Varmus, H. E. (1984) *Annu. Rev. Genet.* **18**, 553–612.
- Bureau, T. E., White, S. E. & Wessler, S. R. (1994) *Cell* **77**, 479–480.
- Jin, Y.-K. & Bennetzen, J. L. (1994) *Plant Cell* **6**, 1177–1186.
- Palmgren, M. G. (1994) *Plant Mol. Biol.* **25**, 137–140.
- Elrouby, N. & Bureau, T. (2001) *J. Biol. Chem.*, in press.
- Deiningner, P. L. (1989) in *Mobile DNA*, eds Berg, D. E. & Howe, M. M. (Am. Soc. Microbiol., Washington, DC) pp. 619–636.
- Jin, Y.-K. & Bennetzen, J. L. (1989) *Proc. Natl. Acad. Sci. USA* **86**, 6235–6239.
- Esnault, C., Maestre, J. & Heidmann, T. (2000) *Nat. Genet.* **24**, 363–367.
- Kim, H. S., Kim, S. & Robbins, P. D. (2000) *Adv. Virus Res.* **55**, 545–663.
- International Human Genome Sequencing Consortium (2001) *Nature (London)* **409**, 860–921.
- White, S. E., Habera, L. E. & Wessler, S. R. (1994) *Proc. Natl. Acad. Sci. USA* **91**, 11792–11796.
- Pickeral, O. K., Makalowski, W., Boguski, M. S. & Boeke, J. D. (2000) *Genome Res.* **10**, 411–415.
- Goodier, J. L., Ostertag, E. M. & Kazazin, H. H. (2000) *Hum. Mol. Genet.* **9**, 653–657.