BARE retrotransposons produce multiple groups of rarely polyadenylated transcripts from two differentially regulated promoters

Wei Chang and Alan H. Schulman

1 MTT/Bl Plant Genomics Laboratory, Institute of Biotechnology, Viikki Biocenter, University of Helsinki, P.O. Box 56, Helsinki, Finland, and
2 Plant Genomics, Biotechnology and Food Research, MTT Agrifood Research Finland, Myllytie 10, FIN-31600 Jokioinen, Finland

Received 11 March 2008; revised 29 April 2008; accepted 19 May 2008; published online 15 July 2008.

Summary

The BARE retrotransponson family comprises more than 10^4 copies in the barley (Hordeum vulgare) genome. The element is bounded by long terminal repeats (LTRs, 1829 bp) containing promoters and RNA-processing motifs required for retrotransposon replication. Members of the BARE1 subfamily are transcribed, translated, and form virus-like particles. Very similar retrotransposons are expressed as RNA and protein in other cereals and grasses. The BARE2 subfamily is, however, non-autonomous because it cannot produce the GAG capsid protein. The pattern of plant development implies that inheritance of integrated copies should critically depend, in the first instance, on cell-specific and tissue-specific expression patterns. We examined transcription of BARE within different barley tissues and analyzed the promoter function of the BARE LTR. The two promoters of the LTR vary independently in activity by tissue. In embryos TATA1 was almost inactive, whereas transcription in callus appears to be less tightly regulated than in other tissues. Deletion analyses of the LTR uncovered strong positive and negative regulatory elements. The promoters produce multiple groups of transcripts that are distinct by their start and stop points, by their sequences, and by whether they are polyadenylated. Some of these groups do not share the common end structures needed for template switching during replication. Only about 15% of BARE transcripts are polyadenylated. The data suggest that distinct subfamilies of transcripts may play independent roles in providing the proteins and replication templates for the BARE retrotransposon life cycle.

Keywords: retrotransposon life cycle, promoter, transcriptional regulation, polyadenylation, mRNA, barley.

Introduction

The long terminal repeat (LTR) retrotransposons are ubiquitous in eukaryotes and comprise large proportions of many plant genomes (Feschotte et al., 2002; Kumar and Bennetzen, 1999; Schulman and Kalander, 2005). The life cycle of active LTR retrotransposons comprises the same basic steps, facilitated by the same major proteins, as that of retroviruses except that the assembled particle is not enveloped for passage out of the cell (Vicient et al., 2001b; Wicker et al., 2007; Wright and Voytas, 2002). These steps comprise transcription, translation of the proteins encoded by the retrotransposon, packaging into a virus-like particle, reverse transcription, nuclear localization, and finally integration of the cDNA copy back into the genome.

Transcription of a retrotransposon therefore represents the first regulatory opportunity because it controls production of the RNA that serves as the template for both reverse transcription and translation. Long terminal repeat retrotransposon transcriptional promoters and their regulatory regions are contained within the LTRs. Although the process of reverse transcription renders the two LTRs identical, the 5’ LTR drives transcription whereas the 3’ LTR serves as the transcription terminator.

The final replication step, integration, must occur in cells clonally ancestral to gametes to be heritable in plants that reproduce only sexually. Hence, tissue-specific control mechanisms can affect the evolutionary success and geno-
mic prevalence of retrotransposon families. Some, generally low-copy, retrotransposons are normally transcriptionally silent but are stress-inducible (Grandbastien et al., 2005; Liu et al., 2004; Takeda et al., 1998), whereas abundant elements such as BARE appear more generally active (Jääskeläinen et al., 1999). Transcriptional analysis of the latter is needed to clarify control of their propagation and production of heritable copies.

BARE, a well-characterized and active Copia LTR retrotransposon family, is bounded by 1.8-kb LTRs that surround a central coding domain. The BARE elements are highly abundant in Hordeum genomes, which they have helped to shape (Kalander et al., 2000; Shirasu et al., 2000; Suoniemi et al., 1996a; Vicient et al., 1999). Furthermore, closely related copies are found in other cereals, which are both transcribed (Vicient and Schulman, 2002) and translated (Vicient et al., 2001a). The BARE family is divided between two main subfamilies, BARE1 and BARE2 (Tanskanen et al., 2007). The recently identified BARE2 subfamily is highly similar to BARE1 but lacks a functional coding region for the capsid protein GAG, due to a specific, conserved deletion of the start codon (Tanskanen et al., 2007; Vicient and Schulman, 2005). However, it contains the signals and structures needed for packaging into BARE1 virus-like particles (VLPs) and replication by the BARE1 enzymes (Sabot and Schulman, 2006). In barley, BARE2 is actually more abundant than the BARE1 host. This suggests that BARE2 is a non-autonomous subfamily of retrotransposons that is parasitic on BARE1 (Sabot and Schulman, 2006; Tanskanen et al., 2007). Such parasitism appears to require co-expression of the BARE1 and BARE2 transcripts and the BARE1 proteins, but this has not been investigated until now.

More generally, retrotransposon transcripts are either translated, reverse-transcribed, or both. It has remained an open question for both retrotransposons and retroviruses as to whether the self-same transcripts serve both functions. Here, in order to shed light on the dynamics of the BARE life cycle and the interplay of BARE1 and BARE2 subfamilies, we have examined the transcripts of this retrotransposon family. Promoter choice in different intact tissues, control of transcription, transcription initiation, and termination and polyadenylation were investigated.

**Results**

**A comparison of BARE promoter choice in different barley tissues**

The BARE retrotransposon is unusual in containing two promoters, TATA1 and TATA2 (TATA-box like motifs; Figure 1). The question of how both promoters may be under selection for maintenance prompted us to investigate the relative activity of TATA1 and TATA2 in leaves,
Probes P1 and P2 were designed to match the consensus sequence of the available BARE LTRs. The products of the probes protected by cellular RNA and by control sense-strand RNA, together with the other reaction controls (Figure 2), show that the probes are specific. Each probe generates protected fragments corresponding to the sizes expected from hybridization to the main transcript classes (described below and in Figure 5). However, because TATA2 is downstream of TATA1, the P2 signal is derived from transcription of both TATA1 and TATA2, whereas P1 reports TATA1 activity only. Relative expression levels from TATA1 and TATA2 were calculated so as to take this into account.

We compared the strength of TATA1 and TATA2 by using the 18S RNA probe P18 as the universal internal control. The curves of Figure 2(b) are normalized to the highest peak and compressed along the Y-axis; hence, the P1 and P2 peaks appear similar in height. Embryos showed the highest total signal as measured by the P2 probe (4.57 \texttimes \text{the signal from 18S}), followed by roots (3.75), callus (3.25), shoots (3.0), and leaves (1.9). The TATA1 promoter, as measured by the P1 probe, showed fairly similar activities between the different tissues of about 1 \times 18S. Therefore, differences in total transcript levels between the tissues can be attributed to differences in TATA2 activity. When the P1 and P2 signals were normalized to a constant 18S signal, TATA2 accounted for an activity of 41.1 normalized units in embryos, 33.4 in roots, 24.5 in callus and shoots, and 12.5 in leaves.

The probes could also bind to read-through transcripts not driven by the LTR. To check this, we searched the dbEST database, which contains about 479 000 barley expressed sequence tags (ESTs), for matches to BARE by BLASTN. Only one possible read-through transcript from barley and two from wheat were seen (data not shown). Hence, the RNase protection results are unlikely to have been affected by read-through transcripts.

Control of BARE transcription in barley tissues

Previous assays of the promoter activity of the BARE1 LTR were carried out only in leaf protoplasts (Suoniemi et al., 1996b), which may not accurately reflect expression in organized tissues. Transient expression assays (Figure 3) were carried out by particle bombardment using full-length and deleted LTR constructs (Table 1 and Figure 4) in order to reveal regulatory elements. We investigated leaves, embryos, and callus cultures, but concentrated on leaves (Figure 4a) as they are better bombardment targets.

The removal of 654 bp from the 5’ end of the LTR to form construct C reduces promoter activity by 30% in leaves (Figure 4a), indicating the presence of positive elements upsteam of this position. However, construct GH, containing only the TATA2 box, yields as much activity as the full-length LTR, suggesting that negative regulatory elements are present between nucleotide position (nt) 963 and nt
1444. Construct BP contains only the final 568 bp of the LTR and displays 52% of the activity seen with the full-length LTR, pointing to a strong positive regulatory element for TATA2 in the 166-bp piece between nt 1444 and nt 1611 that differentiates constructs BP and GH.

We investigated the role of regions downstream of TATA2 on reporter enzyme expression in leaves. When 426 bp at the 3’ end of construct C is deleted to form construct CΔ2 (Figure 4a), LUC activity returns to that of a full-length LTR. Because enzyme activity rather than mRNA per se is measured in the assay, this increase in activity may be due to deletion of the many previously reported (Suoniemi et al., 1996b) ATG start codons upstream of the reporter gene, rather than to an increase in promoter strength. Construct T2 contains most (except for the 5’ 38 nt) of the positive regulatory regions of construct GH and of the full-length LTR between nt 309 and 962, but it also retains one-quarter of the negative regulatory region defined by the comparison between constructs C and GH. It displays only 40% of full LTR activity, thereby narrowing the extent of the negative and positive regulatory regions to, respectively, 127 bp between nt 1444 and 1611 and 38 bp between nt 1444 and 1481.

The inactivity of construct T0, lacking TATA1, TATA2, and all LTR sequences downstream of TATA2, establishes that the first 910 bp of the LTR do not contain a cryptic promoter. Likewise, the inactivity of construct H demonstrates that the last 338 bp of the LTR lack an independent promoter. The results reported here with constructs GH and T2 correspond

Table 1 Constructs used in particle bombardment for investigation of BARE transcription regulation

<table>
<thead>
<tr>
<th>Construct</th>
<th>LTR fragment</th>
<th>Nucleotide range</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>LTR</td>
<td>Xhol</td>
<td>309–2179</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>PvuI–Xhol</td>
<td>963–2179</td>
<td></td>
</tr>
<tr>
<td>GH</td>
<td>Ddel–Xhol</td>
<td>1444–2179</td>
<td></td>
</tr>
<tr>
<td>BP</td>
<td>StyI–Xhol</td>
<td>1611–2179</td>
<td></td>
</tr>
<tr>
<td>CΔ2</td>
<td>PvuI–SacI</td>
<td>963–1753</td>
<td></td>
</tr>
<tr>
<td>H</td>
<td>SacI–Xhol</td>
<td>1757–2179</td>
<td>No TATA box</td>
</tr>
<tr>
<td>T0</td>
<td>BseEI</td>
<td>309–1217</td>
<td>No TATA box</td>
</tr>
<tr>
<td>Δ4</td>
<td>Stul</td>
<td>309–1420</td>
<td>Only TATA1</td>
</tr>
<tr>
<td>T1</td>
<td>StyI + SacI–Xhol</td>
<td>309–1614 + 1757–2179</td>
<td>Only TATA1</td>
</tr>
<tr>
<td>SA</td>
<td>Stul–SacI</td>
<td>1421–1753</td>
<td>Only TATA2</td>
</tr>
<tr>
<td>SN</td>
<td>Stul–SnaBI</td>
<td>1421–2043</td>
<td>Only TATA2</td>
</tr>
<tr>
<td>T2</td>
<td>BseEI + AviI–Xhol</td>
<td>309–1217 + 1481–2179</td>
<td></td>
</tr>
</tbody>
</table>

*The full-length LTR ranges over nt 309–2137, with the untranslated leader extending to the putative translational start at nt 3392. An intact 5’ end begins at nt 309.

*Numbering as in accession number Z17327.

Figure 3. Histochemical analysis of the BARE long terminal repeat (LTR) promoter by GUS activity staining in different tissues following bombardment.
(a) Seven-day-old leaf.
(b) Twenty-four-hour germinated embryo.
(c) Forty-eight-hour-old shoot.
(d) Seven-day callus culture.
(e) Seven-day root.

Figure 4. Contribution of BARE long terminal repeat (LTR) regions to promoter activity in bombarded tissues.
(a) Comparison of various constructs in leaves. A diagram of the constructs (full details in Table 1) is shown on the left, luciferase activity relative to the full-length LTR on right. Error bars are shown.
(b) Comparison of responses in embryo (black) and callus (brown), relative to the full-length LTR in each tissue.

© 2008 The Authors
well to those from the earlier protoplast assays (Suoniemi et al., 1996b). However, construct BP has more than 60% of the LTR activity in leaves and less than 50% in protoplasts, and construct C gives 90% of the LTR activity in leaves but 125% in protoplasts. These differences may relate to variation in the presence of regulatory factors in protoplasts compared to intact tissues, to the stressed state of protoplasts, or to both factors.

Callus tissue is a useful experimental system for studying the BARE life cycle due to the abundance of BARE translation products in callus (as there is also in embryos) compared with their dearth in leaves (Jääskeläinen et al., 1999; Vicent et al., 2001a). For transcriptional comparisons between embryos and callus, we chose constructs to test the relative strengths of TATA1 and TATA2: the full-length LTR, BP, and SA, having only TATA2, and T1 and ?4 (Figure 4b). Construct T1 contains TATA1 and the rest of the LTR, but lacks TATA2 and its downstream region. Construct ?4 contains TATA1 and the initial part of LTR but lacks TATA2 and both its upstream and downstream regions. The results were normalized to the function of the full-length LTR, and showed that both TATA1 and TATA2 were active in callus. TATA2 had more absolute activity in embryos than in callus. Construct SA showed that the upstream regulatory region of TATA2 is sufficient to give strong expression in the embryo, with expression in callus approaching the full-length LTR. Compared with the full-length LTR, construct BP gave about 40% activity in embryos and about 75% in callus, differing in turn from leaves and protoplasts as described above.

**Determination of the start site for BARE transcripts**

Given the tissue-specific differences in promoter choice, we have examined the 5' ends of BARE transcripts from various tissues in more detail by 5' rapid amplification of cDNA ends (RACE). For TATA1, we were also thereby able to separate the results for BARE1 and BARE2 (Figure 5). Shoots yielded clones for transcripts from both BARE subfamilies and TATA boxes. Callus cells contained both TATA1-driven BARE2 transcripts and TATA2-driven BARE transcripts. Multiple, closely spaced 5' ends were obtained, with four different 5' ends for TATA1 products in 13 sequenced RACE clones and five different 5' ends for TATA2 products among 24 RACE clones (Figures 5 and 6). The nucleotide polymorphisms visible within the sequences indicate that the transcripts are derived from different groups of BARE elements, rather than representing incomplete extensions. Only one sequence was found that was prematurely truncated at the 5' end.

Sequence alignments (Figure 6) demonstrate that TATA1 transcripts starting at position 1348 (numbered according to accession Z17327) belong to the BARE1 subfamily (Figure 6a, group A) and are highly similar to the BARE1 type element (Z17327). This position is 42-bp downstream of TATA1. Transcripts starting at positions 1351, 1363, and 1383 (Figures 5 and 6b) belong to the BARE2 subfamily. The TATA1 promoter region and transcriptional starts of BARE2 LTRs contain many small indels and point mutations compared with BARE1, hence the nucleotide numbering according to BARE1 is imprecise. The TATA2 transcripts initiate at positions 1675, 1678, and 1685 (Figures 5 and 6). The majority of TATA2 shoot transcripts begin 19-bp downstream of the TATA box, less than the predicted 33 bp. The TATA2 box, visible in the transcripts originating from TATA1, is completely conserved at position 322 in the alignment (Figure 6b). Due to the similarity of the sequence downstream of TATA2 between BARE1 and BARE2 (Figure 6b), it is difficult to assign TATA2 transcripts to a particular subfamily.

Callus transcripts differ from shoot transcripts not only by start site but also by sequence (Figures 5 and 6b, groups I, J). Given the active BARE translation in callus, it is striking that of 37 sequenced transcripts none of the TATA1 products from callus were derived from BARE1, which is translationally competent. This indicates that GAG translation products seen in callus are likely to be from BARE1/TATA2 transcripts.

---

**Figure 5.** Transcriptional start sites for BARE.

(a) Diagram of the promoter region of the long terminal repeat (LTR). White box, untranscribed region; gray box, untranslated leader according to the longest transcript.

(b) Diagram, abundance, and types of 5' rapid amplification of cDNA ends (RACE) products cloned from shoot transcripts. Diagonally hatched boxes indicate transcripts of TATA1 and stippled boxes transcripts of TATA2. 'Group' refers to sequences sharing the same start site. 'Copies' is the number cloned and sequenced of each group out of a total of 35. 'Variants' are distinct sequences within each group. 'Element' is the retrotransposon subfamily from which the transcripts are derived, with ? meaning that subfamily cannot be resolved. 'Start' refers to the transcriptional start position for each group.
The most common TATA2 transcripts from callus begin 10 bp further downstream than those found in shoots (nt 352 in the alignment; position 1686, Z17327). The callus transcripts, furthermore, begin only 1–4 bp downstream of the putative RNA cap site (Manninen and Schulman, 1993a). These data, taken together, suggest that a different set of BARE elements is transcribed in callus compared with those in differentiated plant tissues, and that BARE expression in callus may be less tightly regulated as well.

BARE transcript polyadenylation

Regulation of the polyadenylation of the retrotransposon transcript is virtually unexplored in plants. Retroviruses may produce both polyadenylated and non-polyadenylated transcripts (Ratnasabapathy et al., 1990); many non-polyadenylated virus transcripts are efficiently translated (Gallie and Walbot, 1990; Zeenko and Gallie, 2005). We investigated the frequency of polyadenylation BARE transcript by cloning BARE RNA independent of the presence of poly(A) tails, using RNA ligase to ligate linkers to total RNA prior to PCR (RLM-PCR). Following 3' linker ligation, RT-PCR was carried out with primer RLM-2 matching BARE in combination either with E1820, which amplifies only poly(A) RNA, or with primer E2146, which matches the linker and amplifies independent of the presence of a poly(A) tail (Figure 7). To investigate the proportion of polyadenylated transcripts among all BARE transcripts, more than 100 of these products were cloned, then screened by PCR with the same primers.

![Figure 6. Alignment of BARE rapid amplification of cDNA ends (RACE) products with the BARE1 and BARE2 DNA sequences. Letters refer to transcript groups, numbers to the variant type within each group (for details see Figure 5).](image)

(a) Alignment of group A transcripts to BARE1. Transcripts of group A are aligned separately from the others due to the large number of indels in the start region between BARE1 and BARE2.

(b) Alignment of group B–J transcripts (for details see Figure 5) to BARE1 and BARE2 genomic sequences.

![Figure 7. Amplification of BARE 3' ends by RNA ligase-mediated (RLM) rapid amplification of cDNA ends (RACE).](image)

Left, amplification only of transcripts containing a poly(A) tail using primer E1820. Right, amplification of all BARE 3' ends using primer E2146. RT(−) control reactions lacked reverse transcriptase. A 100-bp ladder is indicated. The poly(A) products, although not visible in the ‘all tails’ lane, are faintly present and can be isolated as clones from the reaction.
A total of 15.5% contained poly(A) tails. The tails varied from 32 to 36 nt in length.

**BARE transcript termination**

A repeat, the R domain, is present at both ends of a retrotransposon transcript. It is formed during transcription by the LTR segment lying between the start of transcription and its termination. It allows pairing between the RNA transcript and the first cDNA strand, essential for replication. Given the variations in transcript polyadenylation and initiation position, we examined whether transcript termination sites also vary, and how this would affect the R domain. The termination sites were established by sequencing both polyadenylated and non-polyadenylated clones derived from the T4 RNA ligase-mediated (RLM) cloning. Eight cloned products were checked for insertion length, and five were sequenced. Of these, four terminate at positions 11539, 11559, 11561, and 11561 (by alignment with Z17327). This places the termination points 21 and 23 bp immediately after TATA1. One terminated 189 bp downstream from the TATA1 initiation point. The more upstream termination sites do not provide any R domain to transcripts initiated from TATA2 (Figure 1d). However, the non-polyadenylated clones displayed termination sites 60, 105, 500, and 697nt (at positions 11598, 11642, 12037, and 12147 by alignment with Z17327) beyond the TATA1 initiation site. The data indicate that multiple transcript classes are produced, displaying distinct R-domain structures.

**Discussion**

Success for retrotransposons may be defined as the ability to propagate in the genome to high copy number without seriously compromising the success of the host itself. In the absence of strong negative selection, a positive feedback loop leads to the increase in proportion of the retrotransposon population to represent ones that are expressed under conditions where daughter copies can be effectively integrated and inherited. In this regard, the BARE family can be regarded as highly successful, particularly the BARE2 subfamily parasitic on BARE1. BARE is abundant (Vicient et al., 2001a), transcribed (Suoniemi et al., 1996b), translated (Jääskeläinen et al., 1999), and associated with insertional polymorphisms even at single sites (Kalender et al., 2000).

The first step in retrotransposon replication is transcription of the integrated copies. Transcription and transcript processing offer avenues of cellular control over retrotransposon activity. Here, we have examined the transcription and transcripts of BARE retrotransposons in barley tissues. The BARE LTR initiates transcription and is unusual in possessing two promoters. We found that both promoters are active in callus. TATA2 is much more active in embryos than in leaves, with intermediate activity in roots and callus, whereas TATA1 is virtually inactive in embryos. The results from the RNase protection assays, bombardment assays, and 5' RACE are all consistent regarding the activity of TATA1 versus TATA2, although the assays report somewhat different relative values. The results suggest that, because new daughter copies in embryos but not in leaves are inheritable, control of TATA2 plays a much more important role than control of TATA1 in the evolution of the BARE family. The analyses of LTR promoter constructs indicate that positive regulatory regions are located between nt 309 and 963 and nt 1444 and 1611, while a negative regulatory region can be found between nt 963 and nt 1444. The nt 309-963 region, when searched against a plant cis-acting element database (http://www.dna.affrc.go.jp/PLACE/), reveals many potential regulatory motifs, the roles of which remain to be explored.

We found four start sites for TATA1 and five for TATA2. For TATA1, cDNA sequence and start site correlations show that different BARE groups favor particular start sites. In shoots, both BARE1 and BARE2 use TATA1, but only BARE2 does so in callus. Of the BARE2 transcripts, the one cloned from shoots has a different start site, further downstream, from any of the 11 isolated from callus, which belong to several distinct sequence groups. Different start sites were also seen for TATA2-driven transcripts in both shoots and callus, though for these it is not possible to distinguish those originating from BARE1 from those of BARE2.

Our earlier results (Suoniemi et al., 1996b) indicated a start for TATA1 at nt 1323 and for TATA2 at nt 1670, both for callus. Our current results (Figure 5) differ from these by 25–60 nt for TATA1 and from 5 to 19 nt for TATA2, which is unsurprising. The earlier positions were estimated from the relative mobilities of RNA fragments in sequencing ladders, whereas the current positions have been determined directly by sequencing. Furthermore, the earlier estimates were made without reference to BARE2, which was unknown at the time.

Whole-genome transcriptional analyses indicate that putative alternative transcriptional start sites are not uncommon in Arabidopsis (Alexandrov et al., 2006). However, BARE represents a multi-gene family on a scale not seen for cellular genes in Arabidopsis or elsewhere. Furthermore, the presence not only of multiple start sites but also of multiple stop sites and non-polyadenylated transcripts shows that retrotransposon transcription differs radically from general cellular transcription in plants. Variation in the start and termination positions correlates with sequence variation within BARE, implying that BARE subgroups may differ in their replication rather than each individual BARE being transcribed from multiple sites.

We found, furthermore, that the termination site depends on whether a BARE transcript is polyadenylated. Only about 16% of BARE transcripts were polyadenylated. Although de-adenylation is a key step in mRNA degradation (Beelman
and Parker, 1995; Feldbrügge et al., 2001), two features of the cloned cDNAs suggest that those lacking poly(A) tails are not de-adenylated transcripts. First, the length of the poly(A) tails was homogeneous even when the cDNAs were cloned by T4 RNA ligase and without an oligo-dT primer, showing no evidence of the poly(A) shortening typical of de-adenylation (Feldbrügge et al., 2001). Furthermore, the non-adenylated transcripts terminate at distinct points (60, 105, 500 nt) further downstream from TATA1 in the 3’ LTR than the termination sites of the polyadenylated transcripts (most, 20 nt downstream from TATA1). Hence, the non-adenylated transcripts are not shortened versions of those with poly(A) tails.

Retrotransposons are transcribed in the nucleus, like cellular genes, but are replicated and translated in the cytoplasm. Distinct classes of RNA utilize different export pathways (Cullen, 2003). Polyadenylation, lacking in most BARE transcripts, is an important identity element in defining and controlling export of mRNA. Our results therefore beg the question of nuclear export of non-adenylated BARE transcripts. In preliminary experiments (W.C. and A.S., unpublished results), isolated BARE VLPs appear to contain exclusively or mostly non-polyadenylated BARE RNA, providing evidence not only for the export but also for the packaging of these transcripts. Cellular histone and some retroviral transcripts, which like BARE are both unspliced and not polyadenylated, rely on a 3’ stem–loop structure and recognition by the mRNA exporter receptor TAP (Erkmann et al., 2005; LeBlanc et al., 2007). It remains to be seen if a similar mechanism acts on BARE transcripts.

Retrotransposon transcripts have two distinct roles, as mRNA for translation and as templates for cDNA synthesis, and whether the same RNA serves both purposes is unknown (Sabot and Schulman, 2006). As a prelude to replication, retrotransposon transcripts are generally packaged two to a VLP. Examination of the BARE packaging signals indicates that BARE2 can be packaged together with BARE1 in trans-complementation of its lack of its own GAG (Sabot and Schulman, 2006; Tanskanen et al., 2007). However, it is not the BARE1–BARE2 distinction but that of transcript structure more generally that is critical. The data here indicate that multiple transcript classes are produced from BARE, displaying distinct R-domain structures, with the polyadenylated RNAs containing short ones (Figure 1d). End structures have been shown to be critical in replication of synthetic plant retrotransposon constructs (Böhmdorfer et al., 2005).

Here, we have demonstrated that major classes of native BARE transcripts would not be able to engage in strand transfer during cDNA synthesis if co-packaged, due to the lack of a common R domain stemming from variation in the initiation and termination sites. Only a minority of the transcripts with long predicted R domains were found, for example, among shoot RNA. This provides an indication that the trans-preference concept for retroviral RNAs, where two pools of mRNAs co-exist – one to direct synthesis of the proteins required for the life cycle and the other to serve as the template for the replication of the virus genome (Nikolaitchik et al., 2006) – applies to the LTR retrotransposons as well.

The structure of the RNA transcript probably also plays a critical role, not only in LTR retrotransposon replication but also in translation. Translation of cellular mRNA involves the formation of a loop, where the 5’ cap and 3’ poly(A) tail are brought together through the interaction of a variety of initiation factors and poly(A)-binding proteins (Kawaguchi and Bailey-Serres, 2002). The long R domains, present in the BARE transcripts that are not polyadenylated, may provide for a strategy similar to that of many plant RNA viruses, where stem–loop-mediated translation is used (Dreher and Miller, 2006). BARE2 transcripts must be packaged into BARE1 GAG, yet which population of BARE transcripts is translated remains to be established. At least in callus, only TATA2 produces translationally competent BARE transcripts. The translational competence of non-polyadenylated BARE transcripts needs to be established.

In conclusion, several features of our data suggest that a dynamic system regulates the transcription and replication of distinct BARE subfamilies. First, the existence of two TATA boxes allows independent selective forces to act on their corresponding regulatory regions. Although TATA1 is virtually inactive in embryos, which can give rise to clones of cells ultimately leading to gametes, it is nonetheless maintained. At present we do not know, due to technical limitations, whether TATA1 is active in floral tissues directly giving rise to gametes. Second, subfamilies of BARE1 and BARE2 give rise to mRNAs with distinct start sites, and these may replicate as separate populations due to the presence of different R domains. Lastly, the presence of both polyadenylated and non-polyadenylated transcripts of varying structures suggests the existence of separately replicating pools. The compartmentalization of BARE copies within the genome by expression, RNA processing, and replication creates avenues for segmental evolution of the family and interacts with an opposing force, that of recombination between distinct groups of these retrotransposons (Sabot and Schulman, 2007; Tanskanen et al., 2007; Vicient and Schulman, 2005).

**Experimental procedures**

**Plant materials**

Barley (Hordeum vulgare L.) cv. Himalaya was germinated aseptically at room temperature for 24 h before harvesting embryos. For roots and shoots, Himalaya seeds were sown in 15-cm pots filled with vermiculite and grown in a controlled environment room illuminated with 18 h light at 5 × 104 lux 103μEm−2 sec−1 photosynthetically active radiation (PAR). Fresh shoots and roots were
excised after 3.5 days and leaves were collected after 7 days. Callus lines derived from barley cv. Kymppi were harvested 10–14 days after subculture. Tissues of the same age as prepared for bombardments were frozen in liquid nitrogen and stored at −70°C for total RNA isolation.

In vitro synthesis and gel purification of probes

Probes for ribonuclease protection assays were synthesized by in vitro transcription using T7 RNA polymerase on templates of synthetic DNA containing the T7 promoter. Several partially single-stranded templates were prepared by annealing a synthetic oligonucleotide, complementary only to the promoter region of the longer bottom strand of the template. The resulting oligonucleotide probes are shown relative to the LTR in Figure 1. The templates for probe P1 and its complementary strand were generated by in vitro transcription from annealed oligos A and B, and oligos A and C, respectively. The template for probe P2 and its complementary strand were generated from oligos A and D, and oligos A and E, respectively. The oligos consisted of: oligo A: 5'-GGATCCTATGTTAGGGATCC-3'; oligo B, 5'-CCGGAGTGCAATGAGTGCTTGGAGATCC-3'; oligo C, 5'-GGGAGATCCGGTACGCGCTATTAGGATCC-3'; oligo D, 5'-TTTAGGGGTTAGGGGAGTACCTTATAGTGA-3'; oligo E, 5'-GGGAAATTCTGTCTCAATTACGGGACCTAAACCCTAGTGATCGAGCCATC-3'. The same templates were also used for making probes to detect antisense RNA.

In vitro transcription reactions were made with the T7 MEGA-shortscript kit (Ambion, http://www.ambion.com/) according to the manufacturer’s protocol. Fluorescent probes were labeled with fluorescein-12-UTP at a 1:1 ratio with labeled UTP. The resulting probes (Figure 1) were: P1, 5'-TTTAGGGGTTAGGGGAGTACCTTATAGTGA-3'; P2, 5'-TTTAGGGGTTAGGGGAGTACCTTATAGTGA-3'. The fluorescent probes were present at a 12-fold (P1) to 28-fold (P2) molar excess with respect to the template. The size of the P1 and P2 probes, the position of the LTR–LUC constructs, and the size of the P1 and P2 probes, the probes were present at a 12-fold (P1) to 28-fold (P2) molar excess over the target transcripts.

Transient transformation

Transient transformation was carried out by bombardment (PDS 1000/He particle device, Bio-Rad, http://www.bio-rad.com/), using a vacuum of 28 in Hg, a He pressure of 800 lb in−2 (p.s.i.), and a target distance of 8 cm. Three pieces of leaf, each 1.5 cm in length, or 25 embryos were placed in the center of a bombardment plate containing 0.8% solidified agar. Concentrated callus culture was placed on the center of a plate with a 1.5-cm diameter flat surface.

The different LTR–LUC constructs were mixed with a GUS control plasmid (pBI221) at a ratio of 1:1, and precipitated onto gold particles (~1 μm diameter) prior to bombardment. The precipitation mixture contained 1.25 mg gold particles, 8.25 μg plasmid DNA, 25 μl CaCl₂ (2.5 M), and 10 μl spermidine-base free (Sigma S0266, http://www.sigmaaldrich.com/) at 0.1 M. The mixture was centrifuged briefly and the supernatant removed. After washing with 100% ethanol, the particle–DNA pellet was resuspended in 30 μl ethanol. For each bombardment, 7.5 μl of particle–DNA suspension was spread onto the surface of the macrocarrier. After bombardment, tissues were incubated on the sealed agar plate for 24 h at room temperature in the dark. Preliminary experiments showed that bombardment at room temperature rather than at 4°C, using a full-length LTR promoter, gave four-fold higher expression. Room temperature was used in all subsequent assays. Signal strength was reduced if the particles were coated with CaCl₂ but no spermidine-free base. Microprojectile aggregates were dispersed by bombardment.

LUC and GUS assays

Expressed LUC and GUS proteins were extracted for enzymatic measurement by submerging the bombarded tissues in extraction buffer containing 50 mM Na phosphate pH 7.4, 1% polyvinylpyrrolidone (PVP) 360 000, 2 mM EDTA, and 20 mM DTT and then homogenizing. The extracts were pelleted in a microcentrifuge for 10 min at 4°C. Cleared lysates of 10 μl were assayed both with the Promega LUC kit (Promega E1901, http://www.promega.com/) and Tropix GUS-Light kit (Applied Biosystems, http://www.appliedbiosystems.com/) according to the manufacturers’ directions. In each experiment, two sets of bombardments were carried out, one containing an LTR–uidA deletion construct mixed with the full-length LTR–uidA control and the other containing the full-length LTR–uidA fusion mixed with the full-length LTR–uidA control. Each pair was measured in four replicates. The final results were calculated as the ratio EXO-LUC/LTR-LUC:LTR-LTR/LTR-GUS, where EXO-LUC is the LTR deletion fused to luc, LTR–LUC is the full-length LTR fused to luc, and LTR–GUS is the full-length LTR fused to uidA.

© 2008 The Authors
This approach minimizes variation in promoter activity due to differences between experiments in the efficiencies of sample bombardment and extraction.

**Histochemical staining**

Visualization of GUS activity by staining was performed by submerging tissues, which had been bombarded withLTR–GUS constructs, for 48 h at room temperature in a stain buffer containing 2 mM 5-bromo-4-chloro-3-indolyl-beta-D-galactopyranoside (X-Gal), 0.1% Triton X-100, 0.1 M NaHPO₄. The blue stain was then visualized by light microscopy.

5' RACE-PCR

5' RACE-PCR was used to determine 5' RNA ends. Total RNA was first treated with RNase-free DNase and then purified on a column (Qiagen 28204, http://www.qiagen.com/). The DNase-treated RNA was tested for DNA contamination by PCR, using two primer pairs for integrase (Suoniemi et al., 1998) that amplify well from the highly repeated BARE integrase domain present in barley genomic DNA. Untreated total RNA and genomic DNA served as controls for the DNA contamination test. The reverse transcription reaction for production of cDNA was carried out using the SMART RACE cDNA Amplification Kit (Clontech K1811-1, http://www.clontech.com/) on 1 μg callus or short RNA according to the manufacturer’s instructions. The positions of the specific primers, P-BBS (5'-CTAGGGATCCACGCACACCTGG-3') and ATGA (5'-CTGGATTGCCACCGCCTCAGGACG-3'), are shown in Figure 1. They respectively matched the BARE PBS, which is found adjacent to the 5'LTR, and the beginning of GAG coding region, which is internal to the PBS. For the second step, an additional gene-specific primer, PW1 (5'-AAGCTAGGACGACGCTCGGAGGTTGGT-3'), was used for the PCR amplification. The RACE experiments were carried out with a touchdown PCR (Don et al., 1991) protocol to increase the specificity of the products. The reaction consisted of: five cycles of 94°C for 5 sec, 72°C for 2 min; five cycles of 94°C for 5 sec, 70°C for 10 sec, 72°C for 2 min; 35 cycles of 94°C for 5 sec, 68°C for 10 sec, 72°C for 2 min; a final extension at 72°C for 5 min; storage at 4°C. The PCR product was purified (Minielute PCR purification kit, Qiagen), then cloned into the pGEM-T vector (Promega). Colonies containing inserts were screened by PCR, using the gene-specific primer PW1 together with the Universal primer mix from the kit. Control reactions lacking reverse transcriptase produced no product.

**Analysis of 3’ transcript ends**

To determine the sequence and presence of polyadenylation at the 3’ ends of BARE transcripts, RNA ligase-mediated (RLM) RT-PCR was carried out. Total RNA was isolated (RNAasy kit, Qiagen) then ligated to the phosphorylated linker E2147 (5'-GaagaAGGTGAACAACGCGAGGTTGGT-3'). The RNA was first denatured at 95°C for 2 min then cooled on ice before adding to the reaction. The ligation reaction was incubated at 37°C for 3.5 h and then purified on a MiniElute RNA clean-up column (Qiagen). The cDNA was synthesized using a first-strand synthesis kit (Clontech) with oligo E2146, which is complementary to the E2147 linker, as the primer. The cDNA was amplified using the conditions described above for 5' RACE PCR.

For polyadenylated transcripts, RT-PCR was carried out as follows. Total RNA isolated as above was treated three times with a DNA-free kit (Ambion AM-1906). To test for DNA contamination, PCR was first carried out directly on 5 μg RNA using a pair of well-conserved primers to the LTR. To make cDNA, 1 μg RNA was primed with oligo E1820 (AAGCAGTGGTAAACACGCGAGGTTGGT) and nested PCR was then carried out in two rounds, the first round with primers RLM-1 (CGCAGGGTCTGCACACTTAAGGTTC) and E2146 (AAGCAGTGGTAAACACGCGAGGTTGGT), which matches part of E1820 and the second round with primers RLM-2 (GTGTGGTACC-GAATGTTGGTCCGAGGAGTTGC) and E2146, respectively. The RT-minus controls produced no amplification products. The RT-PCR products were cloned and sequenced.

**Acknowledgements**

We thank Päivi Laamanen for her excellent technical assistance with sequencing gels for the NPA work, Annu Suonimi and Anne-Mari Narvanto for their contributions to the LTR constructs, and Lars Paulin for helpful technical suggestions. We also thank Carlos Vicent and François Sabot for discussion and comments. This work was supported by a grant to WC from the Academy of Finland (project 106949).

**References**


