Genetica **107:** 53–63, 1999.

Structure, functionality, and evolution of the *BARE*-1 retrotransposon of barley

Carlos M. Vicient[1], Ruslan Kalendar[1], Kesara Anamthawat-Jónsson[2], Annu Suoniemi[3] & Alan H. Schulman[1,*]

[1]*Institute of Biotechnology, University of Helsinki, Plant Genomics Laboratory, Viikki Biocenter, P.O. Box 56, FIN-00014, Helsinki, Finland (Phone: +358-9-191-58867; Fax: +358-9-191 58952; E-mail: alan.schulman@helsinki.fi);* [2]*Faculty of Sciences, University of Iceland, Grensásvegi 12, Reykjavík 108, Iceland;* [3]*Valio Ltd., P.O. Box 30, FIN-00039 Valio, Finland;* [*]*Author for correspondence*

Accepted 18 January 2000

Key words: BARE-1, barley, *copia*-like retrotransposon, *Hordeum*, molecular markers

Abstract

The *BARE*-1 retrotransposon is a major, active component of the genome of barley (*Hordeum vulgare* L.) and other *Hordeum* species. *Copia*-like in its organization, it consists of 1.8-kb long terminal repeats bounding an internal domain of 5275 bp which encodes a predicted polyprotein of 1301 residues. The polyprotein contains the key residues, structural motifs, and conserved regions associated with retroviral and retrotransposon GAG, aspartic proteinase, integrase, reverse transcriptase, and RNaseH polypeptides. *BARE*-1 is actively transcribed and translated. As part of our effort to understand the evolution and function of *BARE*-1, we have examined its copy number and localization. Full-length members of the *BARE*-1 family constitute 2.8% of the barley genome. Globally, they are dispersed throughout the genome, excepting the centromeric, telomeric, and NOR regions. Locally, *BARE*-1 occurs more commonly in repetitive DNA than in coding regions, forming clusters of nested insertions. Both barley and other *Hordeum* genomes contain a high proportion of *BARE*-1 solo LTRs. New techniques have been developed which exploit the insertion site polymorphism generated by *BARE*-1 integration to produce molecular markers for breeding, biodiversity, and mapping applications.

Introduction

Retrotransposons are mobile genetic elements which transpose through reverse transcription of an intermediate RNA (Grandbastien, 1992; Kumar & Bennetzen, 1999). The retrotransposons fall into two main classes, both present in plants: LTR retrotransposons, which contain long terminal repeats (Grandbastien, 1992; Bennetzen, 1996), and non-LTR retrotransposons or retroposons (Schmidt, 1999). The LTR-retrotransposons are closely related to retroviruses, resembling them in their organization, encoded proteins, and life cycle (Figure 1B). Both replicate by a cycle of transcription, reverse transcription, and integration back into the genome (Boeke & Chapman, 1991). The main difference between retrotransposons and retroviruses is that retrotransposons lack the *envelope* (*env*) domain which enables the retroviruses to spread from cell to cell (Xiong & Eickbush, 1990). However, the recent finding of a putative *env* domain in a few plant retrotransposons (Laten, Majumdar & Gaucher, 1998; Wright & Voytas, 1998) and the demonstration that the *gypsy* element in Drosophila is infectious (Kim et al., 1994; Pelisson et al., 1997), has blurred the distinction somewhat.

The LTR retrotransposons may further be divided into the *copia*-like and the *romani*- or *gypsy*-like elements, both appearing to be ubiquitous in the plants (Flavell, Smith & Kumar, 1992; Hirochika & Hirochika, 1993; Suoniemi, Tanskanen & Schulman, 1998; Voytas et al., 1992). The main difference between the two groups is the order of the encoded proteins, in particular the placement of the integrase (*in*) domain with respect to that of the reverse

transcriptase (*rt*). The ubiquity of both classes of elements in all eukaryotes and the conservation of domain order and relative placement suggests that the two classes were already in existence in the last common ancestor of the fungi, plants, and animals. The LTR retrotransposons are universal components of plant genomes and are present in often large and heterogeneous populations, forming the major class of transposable elements in plants (Flavell, Smith & Kumar, 1992; Hirochika & Hirochika, 1993; Matsuoka & Tsunewaki, 1996; Pearce et al., 1997; Suoniemi, Tanskanen & Schulman 1998; VanderWiel, Voytas & Wendel, 1993; Voytas et al., 1992).

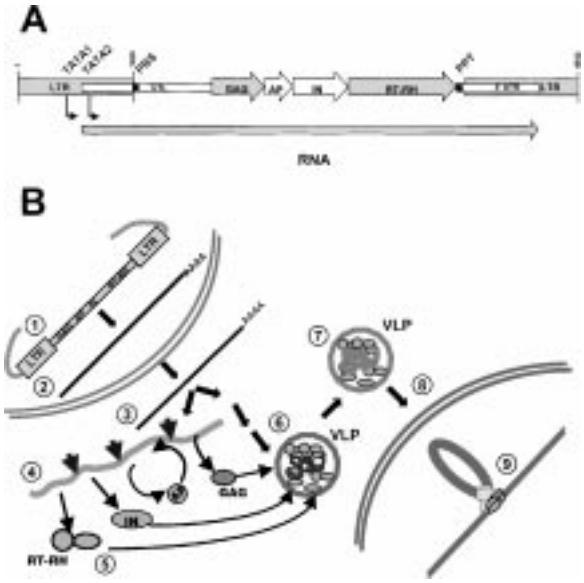

Figure 1. The *BARE*-1 retrotransposon. (A) Organization of the canonical *BARE*-1. The regions of *BARE*-1 are abbreviated as: LTR, long terminal repeat; PBS, primer binding site (the (−)-strand priming site); UTL, 5′ untranslated leader; GAG, the capsid protein; AP, aspartic proteinase; IN, integrase; RT-RH, reverse transcriptase-RNase H; PPT, polypurine tract (the (+)-strand priming site); 3′ UTR, 3′ untranslated region. All regions except the PBS and PPT are drawn to scale. The RNA is shown as would be transcribed from TATA 1. (B) Proposed life cycle of retrotransposons. The integrated element (1) is transcribed (2) and exported as mRNA (3), translated as a polyprotein (4), processed to functional units (5), assembled into a particle together with the transcript (6) which is then converted to cDNA (7) and later moved into the nucleus (8) and integrated as a new copy into the genome (9). The VLP is the virus-like particle formed of the *BARE*-1 components. Integration is evident from insertional polymorphism in close barley lines but has not been demonstrated during the course of an experiment.

Structural features of *BARE*-1

The life cycle of retrotransposons is thought to comprise the same steps as those of retroviruses (Figure 1B) except that the completed particle is not enveloped for passage out of the cell (Boeke & Corces, 1989). Save for the *env* domain required for the envelope formation, retrotransposons also contain the same basic protein products and share many of the retroviral structural characteristics. We isolated, cloned, and sequenced the first complete retrotransposon, and the first transposon of any description, to be found in barley, which we named *BARE*-1 (Barley Retroelement 1) (Manninen & Schulman, 1993). The *BARE*-1 contains all the components of a *copia*-like retrotransposon (Figure 1A). The first full-length *BARE*-1 element (Manninen & Schulman, 1993), named *BARE*-1a, is 12088 bp long, but it contains a 3135-bp insertion in its 3′ LTR. Therefore, a canonical *BARE*-1 element is predicted to be around 8.9 kb long. Southern hybridization and PCR analysis demonstrated that the majority of *BARE*-1 elements in *Hordeum* species are about of this length (Suoniemi et al., 1996; Vicient et al., 1999).

The *BARE*-1 element is bounded by LTRs which are important for the retrotransposon life cycle because they contain the promoters and terminators of transcription as well as the elements necessary for the recognition of the cDNA by integrase for integration into the genome. The length of the genomic *BARE*-1 LTRs range between 1.8 and 1.9 kb (our unpublished data), which is two to six times longer than that of the major part of the *copia*-like retrotransposons. It is unknown why or how *BARE*-1 has acquired such long LTRs. In the case of *Wis*2-1a, a wheat retrotransposon with LTRs of length similar to those of *BARE*-1 (1755 bp), it has been proposed that they increased in length through the accumulation of mutations that caused imprecise excision of the elements, and hence transduction of some flanking DNA into the LTRs (Lucas et al., 1992).

The effect that such an increase in LTR length would have on the transposition of *BARE*-1 is unclear. Large increases in the length of the LTR in the Moloney leukemia virus have only modest effects on its replication efficiency (Reddy et al., 1991; Faustinella et al., 1994), yet relatively small increases in the length of the *Ty*1 LTR strongly reduce its transposition (Lauermann, Hermankova & Boeke, 1997). The elimination of a great part of the *BARE*-1 LTR had no effect on the activity of its promoter at least in transient protoplast assays (Suoniemi, Narvanto & Schulman, 1996), suggesting that most of the LTR

sequence is not important for LTR promoter activity. The *BARE*-1 LTRs contain 6 bp imperfect inverted repeats at their ends with the canonical 5′ TG...CA 3′ terminal sequences present in most retroviruses and retrotransposons. The genome target site duplication was established to be 5 bp (Suoniemi, Schmidt & Schulman, 1997).

Reverse transcription of retroelement RNA transcripts proceed from a tRNA primer bound at the 5′ end of the internal domain, referred to as the primer binding site or PBS (Goff, 1990). Synthesis of the second strand DNA is usually initiated from a purine-rich region located just 5′ of the 3′ LTR (Heyman et al., 1995). These key functional entities are well conserved in the *BARE*-1 family. The PBS of *BARE*-1, as well as of several other plant retrotransposons, is complementary to the $tRNA_i^{Met}$ (Suoniemi, Schmidt & Schulman, 1997). The plus-strand priming site (PPT) is also highly conserved in the *BARE*-1 family.

The region between the LTRs in retrotransposons forms the internal domain. In *BARE*-1, the 5′ untranslated leader sequence (UTL, the sequence between the start of transcription and the start of translation) is about 2 kb long, unusually long compared with those of other elements and of cellular genes. Nevertheless, this sequence is conserved in length among the various copies of the element (our unpublished data). Furthermore, two other retrotransposon families, *Stonor* of maize (Marillonnet & Wessler, 1998) and *RIRE1* of rice (Noma et al., 1997) also possess a 2-kb UTL sequence. The *BARE*-1 UTL region contains at least 51 putative ATG codons. The UTL length, the large number of upstream start codons, and the greater than average sequence similarity suggests that the *BARE*-1, *RIRE1*, and *Stonor* elements may be members of a retrotransposon superfamily predating the divergence of maize, rice, and barley ancestors. Expression studies of reporter genes demonstrated that the removal of part of the UTL inside the LTR produces an increase in the promoter activity in protoplast assays (Suoniemi, Narvanto & Schulman, 1996). This data suggests that this unusually long, although conserved, region may function as a regulator of the retrotransposon activity. For example, the UTL region of the DNA (Type II) transposon *Ac* has been shown to inhibit the promoter activity of this transposable element (Scortecci et al., 1999).

The internal domain of retrotransposons encodes proteins homologous to the retroviral GAG, proteinase (AP), integrase (IN), and reverse transcriptase (RT) (Varmus & Brown, 1989; Grandbastien, 1992). Downstream from the untranslated leader, the *BARE*-1 internal domain encodes a predicted protein bearing the key residues, structural motifs, and conserved regions associated with retroviral and retrotransposon polypeptides. These residues, when aligned with their counterparts from retroviruses and other copia-like retrotransposons, are well conserved (Manninen & Schulman, 1993; Suoniemi et al., 1998; Gribbon et al., 1999).

For the *BARE*-1 IN, this conservation is quite striking (Suoniemi et al., 1998). Translationally silent substitutions are favored. The key catalytic Glu of the DD-35-E motif universal in transposases and integrases is virtually invariant in the more than 90 sequences examined. All invariant retroviral integrase residues, notably those identified to be essential by mutagenesis experiments with the HIV-1 integrase enzyme (van Gent, Oude Groeneger & Plasterk, 1992), are conserved. A model for the tertiary structure of *BARE*-1 IN shows the backbone to be as close to HIV-1 IN and avian sarcoma virus (ASV) INs, superposing with an RMSD of 1.3 Å and 1.4 Å, respectively, as either is to each other, even though the *BARE*-1 IN primary sequence is only 14% identical to these. The variable positions in *Hordeum* INs either display conservative replacements or are found on solvent-exposed surfaces where substitutions would be tolerated. This conservation was observed in *BARE*-1 sequences across the *Hordeum* genus. The sequence and structural analyses indicate that *BARE*-1 IN has been under functional constraint and purifying selection during its evolution.

BARE-1 expression

Despite the ubiquity and abundance of retrotransposons in plants, transcription (Pouteau et al., 1991; Hirochika, 1993; Royo et al., 1996; Suoniemi, Narvanto & Schulman, 1996) and reverse transcription (Hirochika et al., 1996; Lucas et al., 1995) have been demonstrated in only a few cases, one being *BARE*-1. Northern experiments (Pearce et al., 1997) and RNA protection analyses (Suoniemi, Narvanto & Schulman, 1996) demonstrated that *BARE*-1 is actively transcribed in somatic cells in barley.

The LTRs of functional retrotransposons generally contain the promoter motifs necessary for initiation by RNA polymerase II (Temin, 1981). Two putative TATA boxes were identified inside the *BARE*-1 LTRs (Figure 1A) and then shown to be functional. Primer extension analysis demonstrated that the *BARE*-1 tran-

scripts are initiated *in vivo* inside the LTR from each of the TATA boxes (Suoniemi, Narvanto & Schulman, 1996), the level of expression of these two promoters depending on the tissue analyzed. In growing, non-embryogenic calli, most of the transcripts came from TATA1, whereas embryogenic lines generated transcripts from both TATAs and leaf protoplasts from TATA2 exclusively.

Reporter gene constructs driven by the LTR sequences likewise established that the LTR contains at least one active promoter (Suoniemi, Narvanto & Schulman, 1996). In barley protoplasts, only TATA2 is essential for the reporter gene activity. The region upstream of TATA1 contains positive and negative regulatory elements. The negative regulatory elements upstream of TATA1 also affect the activity of the TATA2 promoter. The region between TATA1 and TATA2 showed only positive effects on reporter activity. Curiously, in an analysis of several LTR sequences from different *Hordeum* species, this region showed the highest sequence variability (our unpublished data). Similarities between promoter regulatory motifs, including ABREs (ABA responsive elements), were found within the LTR promoter region. Experiments are in progress to examine their functionality.

Following transcription, the next step in the retrotransposon's cell cycle is translation. We have recently demonstrated that *BARE*-1 mRNA is translated and that the GAG and IN components of the polyprotein are processed into polypeptides of the expected size. Some of the GAG protein forms virus-like particles which contain, together with integrase and *BARE*-1 cDNA, reverse transcriptase activity (Jääskeläinen et al., 1999). The final step of the retrotransposon life cycle is integration into the host genome.

BARE-1 is very prevalent in the barley genome

Unlike classical DNA transposons, integrated copies are not excised and their transposition is replicative. Hence, each transcript of a retrotransposon has the formal potential to be integrated as cDNA back into the genome, giving rise to additional transcripts following integration. A short consideration shows how disruptive this would be to the genome if the process were unregulated: In our hands, total RNA and total DNA may be extracted from tissue in about equal quantities, mRNA comprising some 10% of the total RNA. The average diploid genome size of barley is approximately 9.0 pg (Kankaanpää, Mannonen & Schulman, 1996). Hence, one might expect about 0.9 pg mRNA per cell. Taking the average transcript to encode a 50-kD protein, and therefore, the average mRNA length to be 1.35 kb, 0.9 pg of RNA would contain 2.02×10^{-18} mol of transcripts, comprising 1.2×10^{6} transcripts in total. Rare transcripts are present at abundances of 0.05% of the total or below. Even a transcript of 0.01% abundance would nevertheless be present at about 122 copies per cell given the calculations above. The *BARE*-1 transcripts can be detected by RNase protection assays (Suoniemi, Narvanto & Schulman, 1996) and its translation products detected immunologically (Jääskeläinen et al., 1999), so the transcript is likely not to be extremely rare. Even if it were – on the order of 100 copies per cell – efficient integration of even 10% of these transcripts as cDNA per cell cycle, followed by their subsequent expression, would quickly explode the genome. These new copies would be heritable if the integrations occurred in clones of cells ultimately giving rise to gametes. That this hasn't happened indicates that much regulation of *BARE*-1 replication or integration is taking place.

As might be expected from this, plant retrotransposons present a great variation in copy numbers from a few copies to many thousands (Grandbastien, 1992; Bennetzen, 1996), sometimes presenting great variations between species in the same genus (Joseph, Sentry & Smith, 1990). Plant species with small genomes, such as *Arabidopsis thaliana*, do not appear to tolerate high copy numbers of such elements, whereas some plants with large genomes have upwards of 50% of their genomes composed of retrotransposons (Pearce et al., 1996; SanMiguel et al., 1996).

The *BARE*-1 family is a major, dispersed component of the genome (Suoniemi et al., 1996). Among the species in the *Hordeum* genus, *BARE*-1 is present on average in 14,000 copies, and in barley in 16,000 copies (Vicient et al., 1999), the number varying across the genus (Figure 2A). Based on these copy numbers and on the genome sizes (Kankaanpää, Mannonen & Schulman, 1996), full-length *BARE*-1 comprises from 0.8% to 5.7% of the genome in the genus (Vicient et al., 1999), this measure also varying across the genus (Figure 2B). The *BARE*-1 copy number and genome size are positively correlated (Vicient et al., 1999), indicating that *BARE*-1 is an important contributor, although not the sole, to the differences in genome size among the species of the genus *Hordeum*. The abundance of *BARE*-1 is apparent from its repeated and inadvertent cloning by various researchers seeking barley gene sequences and by its presence on most barley

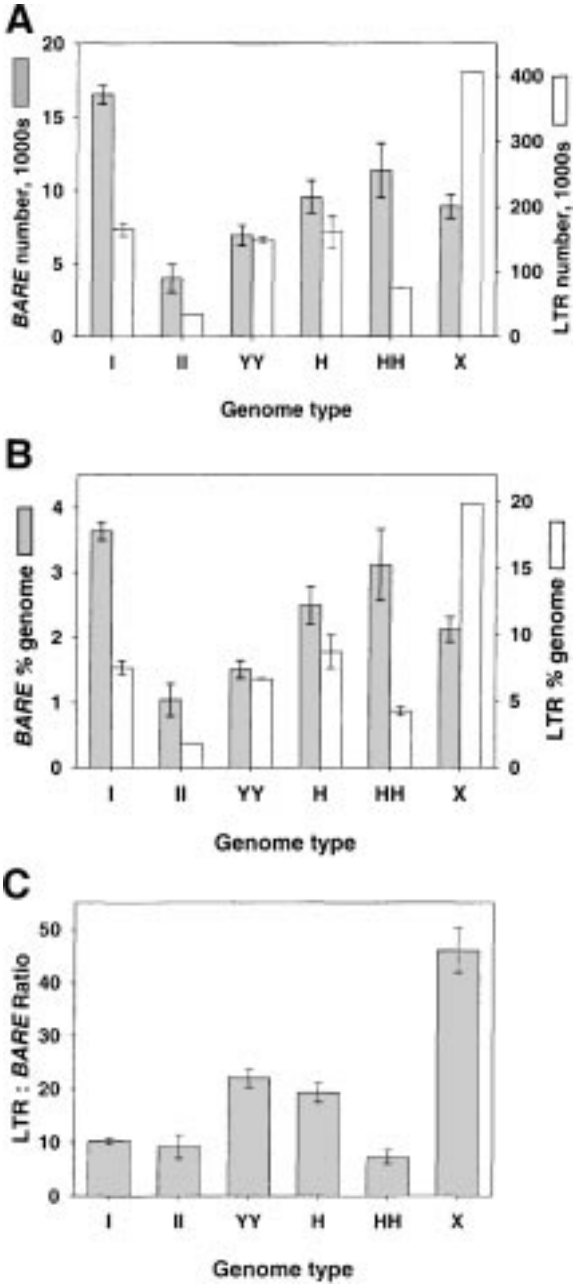

Figure 2. *BARE*-1 copy number in the genus *Hordeum*. (A) Copies per basic (*x*) genome. The shaded bars display the average number of *in* and *rt* copies (left axis) which estimates the number of *BARE*-1 elements; open bars represent the average number of LTR copies (right axis). Error bars are SEM. The data are pooled from Vicient et al. (1999) according to species groups based on meiotic pairing efficiency: I-genome, 12 accessions of *H. vulgare* and *H. spontaneum*; II-genome, *H. bulbosum*; YY-genome, two *H. murinum* tetraploid subspecies; *H*-genome, 10 American diploids; *HH*-genome, *H. depressum* and *H. jubatum*; XX-genome, *H. marinum* subsp. *gussoneanum*. (B) *BARE*-1 genome share. Axes are labeled as in (A). (C) Ratio of LTRs to full-length *BARE*-1 copies. Axes are labeled as in (A).

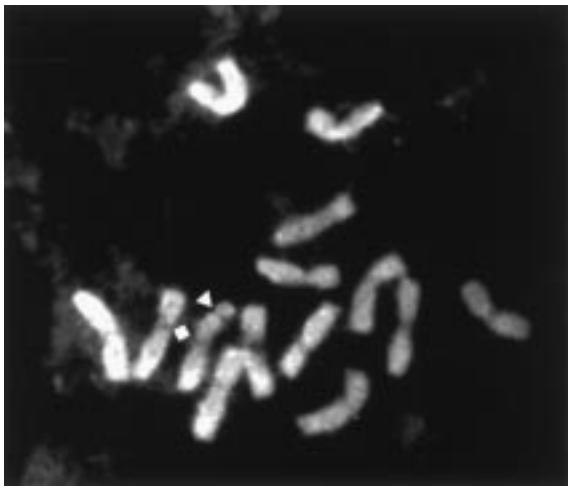

Figure 3. *In situ* hybridization of a *BARE*-1a GAG probe to barley chromosomes. The diamond points to two adjacent centromeres (left) which, like the nucleolar organizing region (NOR, triangle), remain unlabeled. The probe was labeled with biotin, hybridized to the metaphase chromosomes using the same stringency and conditions as in Vicient et al. (1999), detected using ExtrAvidin-FITC (Sigma, USA), and visualized in a Leitz epifluorescence microscope at 1000 H magnification with a FITC filter.

BAC clones (our unpublished data). Moreover, degenerate PCR followed by sequence analysis of fragments of the reverse transcriptase revealed that, although the barley genome contains a highly heterogeneous population of *copia*-like retrotransposons, *BARE*-1-like elements are very common (Gribbon et al., 1999).

The chromosomal distribution of *BARE*-1 was studied by *in situ* hybridization (Suoniemi et al., 1996; Waugh et al., 1997; Vicient et al., 1999) and confirms the copy number observations. The *BARE*-1 elements appear uniformly distributed over the whole of all chromosomes, excepting paracentromeric regions, telomeres, and nucleolar organizer regions (NORs), when the LTR or protein-coding domains (Figure 3) serve as the probe. When the chromosomes of *Hordeum pusillum* Nuttal and *Hordeum euclaston* Steudel were analyzed, a similar pattern was observed (Vicient et al., 1999).

The role of *BARE*-1 in *Hordeum* genome evolution

The available data, taken together, indicates that *BARE*-1 is both active and very abundant. This combination could be strongly mutagenic but apparently barley cells manage to survive the presence and activity of *BARE*-1. Currently, no barley mutation has been demonstrated to be produced by the insertion of a

BARE-1 element. This may be related to a preference for integration into repetitive DNA. Sequencing of *BARE*-1 flanking regions demonstrated that 69% of the flanking sequences were retrotransposons, 62% of these being *BARE*-1 elements (Suoniemi, Schmidt & Schulman, 1997). While this may indicate only a differential survival of such insertions and not the enzymatic preference of the IN, the success of inter-*BARE*-1 amplification reactions used in the IRAP marker system (Kalender et al., 1999) indicates that *BARE*-1 elements are generally clustered near each other in the genome. Nevertheless, high levels of insertional polymorphism has been detected within various *Hordeum* species (Gribbon et al., 1999; Kalendar et al., 1999), between closely-related barley cultivars (Kalendar et al., 1999), and even among individuals of a population (our unpublished results), strongly suggesting that *BARE*-1 have been transposing very recently.

Clustering of retroelements has been reported elsewhere, for *Drosophila* and maize (Pimpinelli et al., 1995; SanMiguel et al., 1996). A nested retrotransposon insertion pattern has also been observed for maize (SanMiguel et al., 1996) and may be common where such elements are abundant. Retrotransposon-rich areas would represent a safe target for new insertions, minimizing the disruption of cellular genes. Again, this may represent only differential survival of such insertions. In plants, the germ line is not sequestered early in development. If a deleterious retrotransposon insertion occurs in a meristematic cell which would give rise to the floral meristem, this cell may be replaced by a non-mutated neighbor. In this way, it is possible that lethal or deleterious retrotransposon insertions in the apical meristem would not contribute to the formation of germ cells upon the differentiation of the floral meristem. As a consequence, only tolerated, relatively non-deleterious insertions – therefore likely to be in non-essential DNA including other retrotransposons – would be transmitted to the next generation.

Determination of the copy number for both the *BARE*-1 LTR and the internal regions, RT and IN, as well as analyses of BAC clones demonstrated that *Hordeum* genomes contain a large excess of *BARE*-1 LTR sequences relative to the internal domain. In various species there are 7- to 42-fold more LTRs than internal sequences (Figure 2C), departing from the expected 2:1 ratio of intact retrotransposons (Vicient et al., 1999). In cultivated barley, the LTRs in excess of those at *BARE*-1 ends account for a further 3% of the barley genome (Figure 2B). The excess of LTRs may have their origin in homologous recombination between the LTRs of a single element, which would remove the internal regions and leave behind a single recombinant LTR (Figure 4A). Solo LTRs have been observed occasionally in other plant retrotransposons (Shepherd et al., 1984; Day et al., 1988; Deragon et al., 1994; SanMiguel et al., 1996; Bevan et al., 1998; Chavanne et al., 1998; Chen, SanMiguel & Bennetzen, 1998), though not in the abundance seen in *Hordeum.*

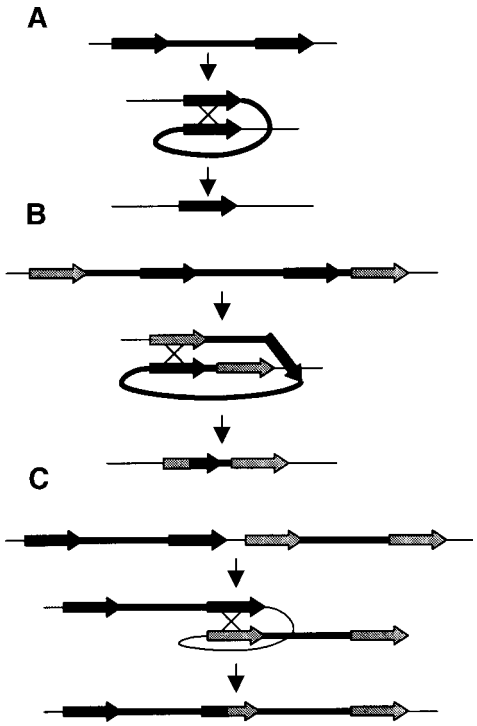

Figure 4. Recombinational processes between LTRs. (A) Intra-element recombination and generation of a solo LTR. (B) Inter-element recombination between nested elements and generation of partially deleted elements. (C) Inter-element recombination between closely inserted elements and generation of tandem elements sharing one LTR.

Recombination between LTRs would be expected to reduce the complement of functional retrotransposons in the genome, limiting but not eliminating the contribution of *BARE*-1 to the genome size. Consistent with these observations, the LTR excess is inversely correlated with the proportion of genome occupied by *BARE*-1. As seen in Figure 2, wild and cultivated barley (I-genome species) have simultaneously the highest average *BARE*-1 copy number per genome (2A), fraction of the genome (2B), and lowest ratio of LTRs to full-length elements. The basic genome of the tetraploid, II-genome species, *H. bulbosum*, has com-

paratively few *BARE*-1 elements. This suggests that, for this species, pressure for a larger genome was met by polyploidization rather than by retrotransposon activation. Selective pressures for a small genome would lead to an accumulation of solo LTRs, as has been observed in the small genome of yeast (Kim et al., 1998; Jordan and McDonald, 1999). The X-genome species, *H. marinum* subsp. *gussoneanum*, has an average number of *BARE*-1 elements relative to other species in the genus, but appears at the same time to have a very active *BARE*-1 population, held in check by simultaneously high rates of recombination, resulting in the highest number, genome fraction, and relative abundance of LTRs observed (Figure 2). Other aspects of the role of *BARE*-1 propagation and recombination are presented elsewhere (Vicient et al., 1999).

Considering the *BARE*-1 abundance and insertion pattern (Suoniemi, Schmidt & Schulman, 1997; Vicient et al., 1999), inter-LTR recombination may lead to general reduction in the repetitive DNA content with few deleterious effects. Recombination events could occur not only between the two LTRs of the same *BARE*-1 element but also between different nested *BARE*-1 elements or between LTRs inserted near to one another in a matrix of repetitive DNA. In the first case, that of nested insertions, the LTR of the newest insertion could recombine with the LTRs of the oldest one, eliminating in certain cases something more than one internal region and one LTR (Figure 4B). In the second case, that of nearby insertions, recombination between LTRs could eliminate non-*BARE*-1 DNA located between LTRs (Figure 4C), generating tandem insertions sharing one of the LTRs. Preliminary results suggest the existence of such structures in the barley genome (our unpublished data). Because *BARE*-1 tends to insert into repetitive, non-essential DNA, the removal of the flanking DNA would probably therefore not be deleterious.

Evolution of *BARE*-1 beyond the *Hordeum* genus

BARE-1 and *BARE*-1-like sequences are widely distributed within the *Triticeae* tribe. Sequences homologous to *BARE*-1 have been detected in wheat, rye, and oat (Gribbon et al., 1999). Moreover, *BARE*-1-like sequences have been detected in rice (Noma et al., 1997), maize (Hirochika & Hirochika, 1993) and *Agropyron cristatum* (Wang & Wein, 1995). High levels of insertional polymorphisms have been detected in barley, wheat, rye, and oat, indicating that *BARE*-1 is actively transposing, or has been active recently, in all *Triticeae* species and not just in cultivated barley (Kalendar et al., 1999; Gribbon et al., 1999).

BARE-1 as genetic tool

The high copy number and the widespread and uniform chromosomal distribution of the *BARE*-1 element in the barley genome provides an excellent platform for the development of DNA-based marker systems. Such marker systems seek to produce a 'fingerprint', or pattern of DNA bands which can be separated electrophoretically to produce a pattern characteristic of a particular genotype. To generate a multiplicity of bands and maximize the potential to detect genetic polymorphisms, all molecular marker systems except RFLP (restriction fragment length polymorphism) are based on prevalent, widespread elements in the genome. Most of these systems rely on PCR amplification, either from random primers (RAPD), from primers to adapters matching particular restriction sites (AFLP), from primers to the simple sequence repeats (SSRs) to amplify between microsatellites (ISSR), or from primers flanking specific microsatellite loci. Retrotransposons can be used for markers because integration of a daughter copy creates new joints between genomic DNA and the conserved LTRs.

To detect polymorphisms for retrotransposon insertion, the marker systems that have been recently developed generally rely on PCR amplification between an LTR and some component of the flanking genomic DNA. The S-SAP (Sequence-Specific Amplified Polymorphism) method employs an outward-facing LTR primer and an adapter primer such as is used in AFLP (Gribbon et al., 1999; Waugh et al., 1997). The RBIP (Retrotransposon-Based Insertional Polymorphism) method is more akin to the microsatellite marker approach, using primers flanking specific retrotransposon insertion sites (Ellis et al., 1998; Flavell et al., 1998). We have developed two methods, REMAP (Retotransposon-Microsatellite Amplified Polymorphism) and IRAP (Inter-Retrotransposon Amplified Polymorphism) which require neither restriction enzyme digestion nor ligation to generate the marker bands (Kalendar et al., 1999). The IRAP bands are generated from two nearby LTRs using outward-facing primers annealing to LTR target sequences. In REMAP, amplification between LTRs proximal to simple sequence repeats such as microsatellites produces the marker bands.

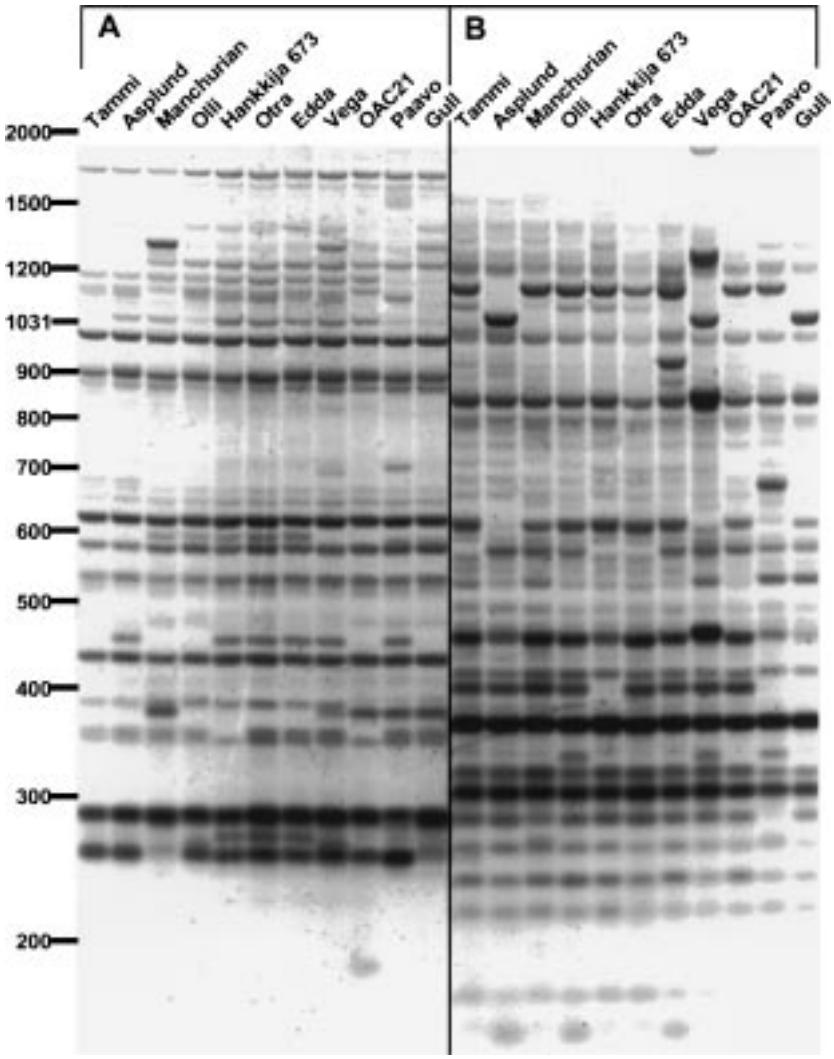

Figure 5. REMAP banding pattern of barley varieties. The image is a negative of an agarose gel stained with ethidium bromide. (A) Pattern from PCR amplification with LTR-A primer for *BARE*-1 and microsatellite primer $(GCT)_6C$. (B) Amplification with LTR-A primer and microsatellite primer $(AGC)_6C$. Primer LTR-A corresponds to the LTR reverse primer described in Kalendar et al. (1999); reaction and electrophoresis conditions are as detailed there. For the varieties depicted: Olli and Asplund are parents of Tammi; Asplund and Vega are parents of Edda; Edda and Tammi are parents of Otra; Otra and Paavo are parents of Hankkija 673. Size markers in bp are shown on the left axis.

Retrotransposons have several advantages as molecular markers over those of other systems such as RAPD, AFLP, or ISSR, described above. Their abundance and dispersion can yield many marker bands, the pattern possessing a high degree of polymorphism due to transpositional activity. The LTR termini are highly conserved even between families (Suoniemi, Schmidt & Schulman, 1997), yet longer primers can be tailored to specific families. Unlike DNA transposons, the new copies are inserted but not removed. Even intra-element recombination resulting in the conversion of a full-length element to a solo LTR either does not affect the marker band and furthermore can be exploited to produce a new marker band in RBIP. Retrotransposon families may vary in their insertional activity, as is seen clearly by the pattern of nested insertions in maize (SanMiguel et al., 1996), allowing the matching of the family used for marker generation to the phylogenetic depth required. Furthermore, the length and conservation of primers to the LTRs facilitate cloning of interesting marker bands and the development of new retrotransposons for markers (Pearce et al., 1999).

The single greatest advantage of the retrotransposon-based marker systems may be their ability to

track an insertion event and its subsequent vertical radiation through a pedigree or phylogeny (Shimamura et al., 1997). Since retrotransposon methods are PCR-based, they may be multiplexed through the use of fluorescent dyes and subjected to analysis on automated sequencers. Retrotransposon marker bands are likely to be more stable than those of other methods. This is because sequence conservation of the LTRs is driven by selection for activity, unlike the essentially neutral evolution of the general repetitive DNA on which AFLP markers, employing the common *Eco*RI and *Mse*I restriction site adapters (Reamon-Buttner, Schmidt & Jung, 1999), and RAPD markers appear to be based. The REMAP and IRAP methods, not requiring restriction enzyme digestion and subsequent linker ligation, should be unaffected by genomic DNA methylation and also be less sensitive to template cleanliness.

The multitude of bands generated by the REMAP method is consistent with *BARE*-1 being inserted into regions of repetitive DNA, and the banding pattern observed in the IRAP method also supports the idea that *BARE*-1 tends to form local clusters. We have successfully exploited IRAP and REMAP to map a disease resistance gene for net blotch of barley (O. Manninen, R. Kalendar, J. Robinson, and A.H. Schulman, submitted), to characterize barley varieties (Kalendar et al., 1999), and even to distinguish individuals within a population of wild barley, *Hordeum spontaneum* C. Koch (our unpublished results). An example of the resolution of the REMAP technique is shown in Figure 5, where polymorphisms in a set of Finnish and Swedish barley varieties related by pedigree (Reino Aikasalo, Boreal OY, Finland, personal communication) are visualized. All bands present can be traced to one or other parent.

Conclusions

Plant retrotransposons as research subjects lagged behind the type II or DNA transposons. However, in recent years it has become clear that not only do they vastly outnumber DNA transposons in both absolute and relative terms in most plants but they are also an appreciable part of the total complement of DNA in the genome. In the cereals and elsewhere, retrotransposons have furthermore been shown to have an active role in genome evolution. The *BARE*-1 retrotransposon comprises upwards of 3% of the barley genome as full-length elements and another 3% as solo LTRs. The *BARE*-1 family is ubiquitous throughout the genus *Hordeum*, widespread in the tribe *Triticeae*, and is recognizably related to elements (*RIRE1*, *Stonor*) in phylogenetically distant species of the family *Gramineae*. It is transcriptionally, translationally, and insertionally active, and contains well-conserved retroviral-like protein-coding domains. We are currently analyzing the *BARE*-1 life cycle and its control in an effort to understand how the replicative potential of this active retrotransposon is regulated. We are furthermore exploiting *BARE*-1 as a breeding and biodiversity marker for *Hordeum* and related genera.

Acknowledgements

The work presented here received support under the European Union Directorate for Biotechnology research program on Molecular Tools for Biodiversity as well as under the Genome Research Programme of the Academy of Finland.

References

Bennetzen, J.L., 1996. The contributions of retroelements to plant genome organization, function and evolution. Trends Microb. 4: 347–353.

Bevan, M., I. Brancroft, E. Bent, K. Love, H. Goodman et al., 1998. Analysis of 1.9 Mb of contiguous sequence from chromosome 4 of *Arabidopsis thaliana*. Nature 391: 485–488.

Boeke, J.D. & V.G. Corces, 1989. Transcription and reverse transcription of retrotransposons. Ann. Rev. Microbiol. 43: 403–434.

Boeke, J.D. & K.B. Chapman, 1991. Retrotransposition mechanisms. Curr. Opin. Cell Biol. 3: 502–507.

Chavanne, F., D.X. Zhang, M.F. Liaud & R. Cerff, 1998. Structure and evolution of Cyclops: a novel giant retrotransposon of the Ty3/Gypsy family highly amplified in pea and other legume species. Plant Mol. Biol. 37: 363–375.

Chen, M., P. SanMiguel & J.L. Bennetzen, 1998. Sequence organization and conservation in sh2/a1-homologous regions of sorghum and rice. Genetics 148: 435–443.

Day, A.M., M. Schimer-Rahire, M.R. Kuchka, S.P. Mayfield & J.D. Rochaix, 1988. A transposon with an unusual arrangement of long terminal repeats in the green alga *Chlamydomonas reinhardtii*. EMBO J. 7: 1917–1927.

Deragon, J.M., B.S. Landry, T. Pelissier, S. Tutois, S. Tourmente & G. Picard, 1994. An analysis of retroposition in plants of SINEs from *Brassica napus*. J. Mol. Evol. 39: 378–386.

Ellis, T.H.N., S.J. Poyser, M.R. Knox, A.V. Vershinin & M.J. Ambrose, 1998. Polymorphism of insertion sites of Ty1-*copia* class retrotransposons and its use for linkage and diversity analysis in pea. Mol. Gen. Genet. 260: 9–19.

Faustinella, F., H. Kwon, F. Serrano, J.W. Belmont, C.T. Caskey et al., 1994. A new family of murine retroviral vectors with extended multiple cloning sites for gene insertion. Hum. Gene Ther. 5: 307–312.

Flavell, A.J., D.B. Smith & A. Kumar, 1992. Extreme heterogeneity of Ty1-copia group retrotransposons in plants. Mol. Gen. Genet. 231: 233–242.

Flavell, A.J., M.R. Knox, S.R. Pearce & T.H.N. Ellis, 1998. Retrotransposon-based insertion polymorphisms (RBIP) for high throughput marker analysis. Plant J. 16: 643–650.

Goff, S.P., 1990. Retroviral reverse transcriptase: synthesis, structure and function. J. Acquire. Immune Defic. Syn. Dr. 3: 817–831.

Grandbastien, M.A., 1992. Retroelements in higher plants. Trends Genet. 8: 103–108.

Gribbon, B.M., S.R. Pearce, R. Kalendar, A.H. Schulman, L. Paulin, P. Jack, A. Kumar & A.J. Flavell, 1999. Phylogeny and transpositional activity of Ty-*copia* group retrotransposons in cereal genomes. Mol. Gen. Genet. 261: 883–891.

Heyman, T., B. Agouti, S. Riant, D.X. Wilhelm & M.L. Wilhelm, 1995. Plus-strand DNA synthesis of yeast retrotransposon Ty1 is initiated at two sites: PPT1 next to the 3′ and PPT2 within the pol gene. PPT1 is sufficient for Ty1 transposition. J. Mol. Biol. 253: 291–303.

Hirochika, H., 1993. Activation of tobacco retrotransposons during tissue culture. EMBO J. 12: 2521–2528.

Hirochika, H. & R. Hirochika, 1993. Ty1-*copia* group retrotransposons as ubiquitous components of plant genomes. Jpn. J. Genet. 68: 35–46.

Hirochika, H., K. Sugimoto, Y. Otsuki, H. Tsugawa & M. Kanda, 1996. Autonomous transposition of the tobacco retrotransposon *Tto1* in rice. Plant Cell 8: 725–734.

Jääskeläinen, M., A.H. Mykkänen, T. Arna, C. Vicient, A. Suoniemi, R. Kalendar, H. Savilahti & A.H. Schulman, 1999. Retrotransposon *BARE*-1: expression of encoded proteins and formation of virus-like particles in barley cells. Plant J. 20: 413–422.

Jordan, I.K. & J.F. McDonald, 1999. Tempo and mode of *Ty* element evolution in *Saccharomyces cerevisiae*. Genetics 151: 1341–1351.

Joseph, J.L., J.W. Sentry & D.R. Smyth, 1990. Interspecies distribution of abundant DNA sequences in *Lilium*. J. Mol. Evol. 30: 146–154.

Kalendar, R., T. Grob, M. Regina, A. Suoniemi & A.H. Schulman, 1999. IRAP and REMAP: Two new retrotransposon-based DNA fingerprinting techniques. Theor. Appl. Genet. 98: 704–711.

Kankaanpää, J., L. Mannonen & A.H. Schulman, 1996. The genome sizes of *Hordeum* species show considerable variation. Genome 39: 730–735.

Kim, A., C. Terzian, P. Santamaria, A. Pelisson, N. Prud'homme & A. Bucheton, 1994. Retroviruses in invertebrates: the *gypsy* retrotransposon is apparently an infectious retrovirus of *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA 91: 1285–1289.

Kim, J.M., S. Vanguri, J.D. Boeke, A. Gabriel & D.F. Voytas, 1998. Transposable elements and genome organization: A comprehensive survey of retrotransposons revealed by the complete *Saccharomyces cerevisiae* genome sequence. Genome Res. 8: 464–478.

Kumar, A. & J. Bennetzen, 1999. Plant Retrotransposons. Annu. Rev. Genet. 33: 479–532.

Laten, H., A. Majumdar & E.A. Gaucher, 1998. SIRE-1, a copia/Ty1-like retroelement from soybean, encodes a retroviral envelope-like protein. Proc. Natl. Acad. Sci. USA 95: 6897–6902.

Lauermann, V., M. Hermankova & J.D. Boeke, 1997. Increased length of long terminal repeats inhibits *Ty*1 transposition and leads to the formation of tandem multimers. Genetics 145: 911–922.

Lucas, H., F. Feuerbach, K. Kunert, M.-A. Grandbastien & M. Caboche, 1995. RNA-mediated transposition of the tobacco retrotransposon *Tnt*1 in *Arabidopsis thaliana*. EMBO J. 14: 2364–2373.

Lucas, H., G. Moore, G. Murphy & R.B. Flavell, 1992. Inverted repeats in the long-terminal repeats of the wheat retrotransposon Wis2-1A. Mol. Biol. Evol. 9: 716–728.

Manninen, O., R. Kalendar, J. Robinson & A.H. Schulman. Application of *BARE*-1 retrotransposon markers to map a major resistance gene for net blotch in barley. Submitted.

Manninen, I. & A.H. Schulman, 1993. *BARE*-1, a *copia*-like retroelement in barley (*Hordeum vulgare* L.). Plant Mol. Biol. 22: 829–846.

Marillonnet, S. & S.R. Wessler, 1998. Extreme structural heterogeneity among the members of a maize retrotransposon family. Genetics 150: 1245–1256.

Matsuoka, Y. & K. Tsunewaki, 1999. Evolutionary dynamics of Ty1-*copia* group retrotransposons in grass shown by reverse transcriptase domain analysis. Mol. Biol. Evol. 16: 208–217.

Noma, K., R. Nakajima, H. Ohtsubo & E. Ohtsubo, 1997. *RIRE1*, a retrotransposon from wild rice *Oryza australiensis*. Genes Genet. Syst. 72: 131–140.

Pearce, S.R., G. Harrison, J.S. Heslop-Harrison, A.J. Flavell & A. Kumar, 1997. Characterization and genomic organisation of Ty1-*copia* group retrotransposons in rye (*Secale cereale*). Genome 40: 617–625.

Pearce, S.R., G. Harrison, D. Li, J.S. Heslop-Harrison, A. Kumar & A.J. Flavell, 1996. The Ty1-*copia* group retrotransposons in *Vicia* species: copy number, sequence heterogeneity and chromosomal localisation. Mol. Gen. Genet. 250: 305–315.

Pearce, SR, C. Stuart-Rogers, M.R. Knox, A. Kumar, T.H.N. Ellis & A.J. Flavell, 1999. Rapid isolation of plant Ty1-*copia* group retrotransposon LTR sequences for molecular marker studies. Plant J. 19: 711–717.

Pelisson, A., L. Teysset, F. Chalvet, A. Kim, N. Prud'homme, C. Terzian & A. Bucheton, 1997. About the origin of retroviruses and the co-evolution of the *gypsy* retrovirus with the *Drosophila flamenco* host gene. Genetica 100: 29–37.

Pimpinelli, S., M. Berloco, L. Fanti, P. Demitri, E. Bonaccorsi et al., 1995. Transposable elements are stable components of *Drosophila melanogaster* heterochromatin. Proc. Natl. Acad. Sci. USA 92: 3804–3808.

Pouteau, S., E. Huttner, M.A. Grandbastien & M. Caboche, 1991. Specific expression of the tobacco Tnt1 retrotransposon in protoplasts. EMBO J. 10: 1911–1918.

Reamon-Buttner, S.M., T. Schmidt & C. Jung, 1999. AFLPs represent highly repetitive sequences in *Asparagus officinalis* L. Chrom. Res. 7: 297–304.

Reddy, S.J., J.V. Degregori, H. von Melchinger & E. Ruley, 1991. Retrovirus promoter trap vector to induce *lacZ* gene fusion in mammalian cells. J. Virol. 65: 1507–1515.

Royo, J., N. Nass, D.P. Matton, S. Okamoto, A.E. Clarke & E. Newbigin, 1996. A retrotransposon-like sequence linked to the S-locus of *Nicotiana alata* is expressed in styles in response to touch. Mol. Gen. Genet. 250: 180–188.

SanMiguel, P., A. Tikhonov, Y.K. Jin, N. Motchoulskaia, D. Zakharov, A. Melake-Berhan, P.S. Springer, K.J. Edwards, M. Lee, Z. Avramova & J.L. Bennetzen, 1996. Nested retrotransposons in the intergenic regions of the maize genome. Science 274: 765–768.

Schmidt, T., 1999. LINEs, SINEs and repetitive DNA: non-LTR retrotransposons in plant genomes. Plant Mol. Biol. 40: 903–910.

Scortecci, K.C., R. Raina, N.V. Fedoroff & M.A. van Sluys, 1999. Negative effect of the 5′-untranslated leader sequence on *Ac* transposon promoter expression. Plant Mol. Biol. 40: 935–944.

Shepherd, N.S., Z. Schwarz-Sommer, J. Blumberg vel Spalve, M. Gupta, U. Wienand & H. Saedler, 1984. Similarity of the Cin1 repetitive family of *Zea mays* to eukaryotic transposable elements. Nature 307: 185–187.

Shimamura, M., H. Yasue, K. Ohshima, H. Abe, H. Kato, T. Kishiro, M. Goto, I. Munechika & N. Okada, 1997. Molecular evidence from retrotransposons that whales form a clade within even-toed ungulates. Nature 388: 666–670.

Suoniemi, A., K. Anamthawat-Jonsson, T. Arna & A.H. Schulman, 1996. Retrotransposon *BARE*-1 is a major, dispersed component of the barley (*Hordeum vulgare* L.) genome. Plant Mol. Biol. 30: 1321–1329.

Suoniemi, A., A. Narvanto & A.H. Schulman, 1996. The *BARE*-1 retrotransposon is transcribed in barley from an LTR promoter active in transient assays. Plant Mol. Biol. 31: 295–306.

Suoniemi, A., D. Schmidt & A.H. Schulman, 1997. *BARE*-1 insertion site preferences and evolutionary conservation of RNA and cDNA processing sites. Genetica 100(1–3): 219–230.

Suoniemi, A., J. Tanskanen, O. Pentikäinen, M.S. Johnson & A.H. Schulman, 1998. The core domain of retrotransposon integrase in *Hordeum*: predicted structure and evolution. Mol. Biol. Evol. 15: 1135–1144.

Suoniemi, A., J. Tanskanen & A.H. Schulman, 1998. *Gypsy*-like retrotransposons are widespread in the plant kingdom. Plant J. 13: 699–705.

Temin, H.W., 1981. Structure, variation and synthesis of retrovirus long terminal repeat. Cell 27: 1–3.

VanderWiel, P.L., D.F. Voytas & J.F. Wendel, 1993. *Copia*-like retrotransposable element evolution in diploid and polyploid cotton (*Gossypium* L.). J. Mol. Evol. 36: 429–447.

van Gent, D.C., A.M.M. Oude Groeneger & R.H.A. Plasterk, 1992. Mutational analysis of the integrase protein of human immunodeficiency virus type 2. Proc. Natl. Acad. Sci. USA 89: 9598–9602.

Varmus, H. & P. Brown, 1989. Retroviruses, pp. 53–108 in Mobile DNA, edited by D.E. Berg and M.M. Howe. Am. Soc. Microbiol. Washington, DC.

Vicient, C.M., A. Suoniemi, K. Anamthawat-Jonsson, J. Tanskanen, A. Beharav, E. Nevo & A.H. Schulman, 1999. Retrotransposon *BARE*-1 and its role in genome evolution in the genus *Hordeum*. Plant Cell 11: 1769–1784.

Voytas, D.F., M.P. Cummings, A. Koniczny, F.M. Ausubel & S.R. Rodermel, 1992. *Copia*-like retrotransposons are ubiquitous among plants. Proc. Natl. Acad. Sci. USA 89: 7124–7128.

Wang, R.R. & J.Z. Wein, 1995. Variations of two repetitive DNA sequences in several *Triticeae* genomes revealed by polymerase chain reaction and sequencing. Genome 38: 1221–1229.

Waugh, R., K. McLean, A.J. Flavell, S.R. Pearce, A. Kumar, B.B. Thomas & W. Powell, 1997. Genetic distribution of *BARE*-1-like retrotransposable elements in the barley genome revealed by sequence-specific amplification polymorphisms (S-SAP). Mol. Gen. Genet. 253: 687–694.

Wright, D.A. & D.F. Voytas, 1998. Potential retroviruses in plants: *Tat1* is related to a group of *Arabidopsis thaliana Ty3*/*gypsy* retrotransposons that encode envelope-like proteins. Genetics 149: 703–715.

Xiong, Y. & T.H. Eickbush, 1992. Origin and evolution of retroelements based upon their novel reverse transcriptase sequences. EMBO J. 9: 3353–3362.