Retrotransposon *BARE1* Translation, Localization, and VLP Formation in Barley

Marko Jääskeläinen

Institute of Biotechnology
Faculty of Biological and Environmental Sciences
Finnish Doctoral Program in Plant Science
University of Helsinki

MTT Agrifood Research Finland

ACADEMIC DISSERTATION

To be presented for public criticism, with the permission of the Faculty of Biological and Environmental Sciences, University of Helsinki, in the auditorium 1041, Biocenter 2, Viikinkaari 5, Helsinki, on November 16th, 2012, at 12 o'clock noon.

Helsinki 2012

Supervisor:	Professor Alan H. Schulman Director, Genomics Research, MTT Group Leader, Institute of Biotechnology, Univ. of Helsinki
Follow up group members:	Professor Jari Valkonen Department of Agricultural Sciences University of Helsinki and Professor Sarah Butcher Institute of Biotechnology University of Helsinki
Reviewers:	Professor Jari Valkonen Department of Agricultural Sciences University of Helsinki and Professor Yrjö Helariutta Institute of Biotechnology University of Helsinki
Opponent:	Doctor Josep M. Casacuberta Center for Research in Agricultural Genomics Barcelona, Spain
Custos:	Professor Tapio Palva Department of Biosciences University of Helsinki

Cover figure: Background: Barley field. Above: immunolocalization of *BARE* Gag in developing floral meristem, electron microscopy of immunocaptured *BARE* VLPs (unpublished), detail of the model of retrotransposon lifecycle (from Sabot and Schulman, 2006).

ISBN 978-952-10-8374-7 (paperback)
ISBN 978-952-10-8375-4 (PDF)

Unigrafia Oy, Helsinki University Print, Helsinki 2012

to my parents

Contents

List of original publications

This thesis is based on the following publications, which are referred in the text with their Roman numerals.

I Vicient C.M., **Jääskeläinen M.J**., Kalendar R., Schulman A.H. 2001. Active retrotransposons are a common feature of grass genomes. Plant Physiology, 125: 1283-1292.

II Jääskeläinen M., Mykkänen A.-H., Arna T., Vicient C.M., Suoniemi A., Kalendar R., Savilahti H., Schulman A.H. 1999. Retrotransposon BARE-1: expression of encoded proteins and formation of virus-like particles in barley cells. The Plant Journal, 20(4):413-422.

III Jääskeläinen, M.J., Chang W., Moisy C., Schulman A.H. Retrotransposon BARE of barley displays strong tissue-specific differences in expression. (MANUSCRIPT)

IV Chang W., **Jääskeläinen M.J**., Li S.-P., Schulman A.H. BARE retrotransposons are translated and replicated via distinct RNA pools. (SUBMITTED MANUSCRIPT)

Statement of my contribution on these articles:

Article I:

Planning: I participated in the planning of the work. Experimental part: I did 40 % of the experimental work. Writing: I took part in writing the manuscript draft and carried out revision based on instructions with my supervisor.

Article II:

Planning: I participated in the planning of the work together with my supervisor Alan Schulman. I did 60 % of the experimental work. Writing: I took part in writing the manuscript draft and carried out revisions based on interactions with my supervisor.

Article III:

Planning: I made the experimental plan together with my supervisor. Experimental part: I did 70 % of experimental work. Writing: I wrote the manuscript draft together with my supervisor and carried out revisions based on interactions with my supervisor.

Article IV:

Planning: I participated in the planning of the work. Experimental part: I participated in the VLP isolation part of the work.

Abbreviations

ABA	abscisic acid
ap/AP	aspartic proteinase gene/protein
BARE	a barley retrotransposon family
BARE1a	the fully sequenced *BARE1* clone (accession number Z17327)
BARE2	a non-autonomous *BARE* retrotransposon in barley
bp	base pair
cDNA	complementary deoxyribonucleic acid
CLEM	correlative light-electron microscopy
copia	a retrotransposon superfamily found in *Drosophila*
DIS	dimerization signal
DNA	deoxyribonucleic acid
env/ENV	envelope gene/protein
ERV	endogenous retrovirus
EST	expressed sequence tag
FRET	fluorescence resonance energy transfer
gag/GAG	GAG gene/protein
gRNA	genomic RNA
HIV	human immunodeficiency virus
int/INT	integrase gene/protein
IR	inverted repeat
IRAP	inter-retrotransposon amplification polymorphism
IRES	internal ribosome entry site
kD	kilodalton
LTR	long terminal repeat
MITE	miniature inverted repeat transposable element
MLV	murine leukemia virus
mRNA	messenger ribonucleic acid
NLS	nuclear localization signal
nt	nucleotide
ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis
PBS	primer binding site
PCR	polymerase chain reaction
pol/POL	polymerase gene/protein, encodes also several other genes, AP, INT, and RT-RH
PPT	polypurine tract
pr/PR	protease gene/protein
pre-mRNA	precursor-mRNA
PSI	packaging signal
R domain	a repeat domain present in the LTRs
REMAP	retrotransposon microsatellite amplified polymorphism
RNA	ribonucleic acid
rt-rh/RT-RH	reverse transcriptase-RNaseH gene/protein
SAM	shoot apical meristem
SDS	sodium dodecyl sulfate
snRNP	small nuclear ribonucleoprotein particle
SSR	simple sequence repeat

TATA	motif in DNA for transcription initiation
TE	transposable element
TEM	transmission electron microscopy
TIR	terminal inverted repeat
tRNA	transfer ribonucleic acid
UTL	untranslated leader, sequence upstream AUG in mRNA
VLP	virus-like particle
V(D)J	variable (diversity) joining

Abstract

Retrotransposons are major components of most eukaryotic genomes. They resemble retroviruses except that their lifecycle is limited to within the boundaries of the cell. In this thesis work, the goal was to understand the replication of the *BARE1* retrotransposon. The *BARE1*, originally found in our research group, is a member of *BARE* retrotransposon family and belongs to the Class I LTR (Long Terminal Repeat) transposable elements (TEs) of the *Copia* superfamily. Complete elements of the *BARE* family constitute about 2.9 % of the barley (*Hordeum vulgare* L.) genome. The lifecycle of an LTR retrotransposon starts with the transcription of its mRNA, followed by translation and processing of proteins needed for its lifecycle, which are the capsid protein (GAG), aspartic proteinase (AP), integrase (INT), and reverse transcriptase (RT-RH). Similarly as in retroviruses, the GAG forms a shell, the virus-like particle (VLP), which packages the nucleic acids of the element and transports them to the nucleus. The proteinase cleaves the polyprotein into the functional proteins, the reverse transcriptase copies the RNA into complementary (c)DNA, and the integrase inserts the cDNA back into the host genome.

The aims of this study were to investigate whether *BARE* and LTR retrotransposons in general in grasses are transcribed, translated, form VLPs, and are integrationally active. For transcription, we systemically searched for homologies of known retrotransposons from Expressed Sequence Tag (EST) databases. These searches revealed that the transcription of retrotransposons in grasses is widespread. Matching ESTs were found across multiple genera, whereas elements from the dicots tend to find matches only in their host species. The *BARE* transcripts were found also to be translated into proteins not only in the tribe Triticeae, but also in different species in different tribes and subfamilies of the Gramineae. These results are the first evidence for pools of retrotransposon polyproteins in plant cells, showing that *BARE*-like retrotransposons are translationally active and sufficiently conserved for immunological detection in a wide range of species in Gramineae. The *BARE1* proteins in barley were immunolocalized *in vivo* in root and shoot meristematic tissues as well as in nodes and in phloem companion cells in internodes. The latter raises the question whether the *BARE* is able to move within the plant vasculature, whereas the presence of *BARE* proteins in floral meristems suggest that newly replicated copies will be passed to the next generation, a strategy that is vital for the survival of retrotransposons in general.

We observed that the *BARE1* expresses its proteins as a polyprotein of 150 kD, which is then cleaved into the mature-sized components such as GAG and INT. We showed, for the first time in plants, that these retrotransposon proteins are abundant enough to be detected immunologically *in vivo*. Various stresses have been shown to activate retrotransposons. In the case of *BARE*, we observed that the amount of mature-sized GAG protein increases during drought, supporting earlier findings of the increase of the element copy number in arid environments. Most of the information regarding VLPs is from yeast and *Drosophila*

elements belonging to the *Copia* superfamily. Thus, we anticipated that retrotransposons in plants may also form VLPs and looked for *BARE* VLPs in barley. VLPs of three size classes were visualized from sucrose density gradient fractions positive for the *BARE1* GAG, INT, and cDNA, as well as for reverse transcriptase activity. This was the first time VLPs were demonstrated in any plant. The VLP formation is a critical step in the retrotransposon lifecycle and requires higher amounts of GAG compared to the other enzymatic components of the retrotransposon. Because the *BARE1* encodes its proteins in single open reading frame (ORF), the question rises of how the requirement of a stoichiometrically higher amount of GAG is achieved. The *BARE1* element was earlier shown to contain two TATA boxes. We show here that the first, TATA1, produces uncapped and non-polyadenylated transcripts, which are packaged into the VLP as a genomic (g)RNA dedicated to reverse transcription into complementary (c)DNA and for ultimate insertion back into the host genome. The downstream, second, TATA2 serves to initiate a shorter transcript that cannot be reverse transcribed due to the lack of repetitive region (R) at its ends. This transcript, however, is decorated with a typical cap and poly(A) tail for translation. Furthermore, we observed that some of the TATA2 products are spliced to produce a subgenomic RNA encoding only GAG, thus enabling the larger GAG production for VLP assembly. Our finding represents a first demonstration for any retrotransposon of distinct RNA pools for translation and transcription.

The final step in the lifecycle of retrotransposons is insertion back into the host genome. We detected polymorphism based on retrotransposon marker techniques in grass species and barley cultivars, suggesting that *BARE* retrotransposons were insertionally active over recent evolutionary time, and are also during and subsequent to domestication. Taken together, this thesis work shows for the first time that the plant retrotransposon *BARE1* is capable of accomplishing its lifecycle, through expressing its proteins in barley in tissues likely to pass copies on to the succeeding plant generations, and forming VLPs through specific processing and pools of RNA transcripts. Moreover, these processes likely take place in grass species other than barley as well, supporting the role of retrotransposons as contributors to growth in genome size.

1 Introduction

1.1 Transposable elements in eukaryotic genomes

Transposable elements (TEs) are ubiquitous components in eukaryotic genomes, comprising up to 85 % of the genomic DNA in plants (Liu et al., 2007; Schnable et al., 2009; Wicker et al., 2009). In contrast, the relative amount of cellular genes can be very low, less than 11% in large-genome grasses such as barley (Rostocks et al., 2002), most of the rest being repetitive DNA. Thus genome size can vary greatly, independent of organismal complexity. This phenomenon is referred as the C-value paradox (Thomas, 1971). Within grasses, the genome size can vary over 50-fold, and most of the difference can be attributed to differences in the prevalence of a group of TEs called LTR retrotransposons (Flavell et al., 1992; Suoniemi et al., 1998b; Wicker et al., 2009; Voytas et al., 1992).

The concept of transposable elements was developed in early 1950's by Barbara McClintock (McClintock, 1953). She discovered that the maize chromosomes carry 'controlling elements' that can change genetic location in the chromosome. The movement of such elements, the linked Activator (*Ac*) and Dissociation (*Ds*) element and the Suppressor-mutator (*Spm*) element, was noticed because they affected the function of nearby genes (Döring and Starlinger, 1984; McClintock, 1956). She called this movement 'transposition', and revolutionized our thinking of the genome stability and organization. In 1983 she was awarded the Nobel Prize for her pioneering work.

The movement of the TEs was experimentally verified in bacterial genomes in 1968 (Jordan et al., 1968), and McClintock's controlling elements were identified as physical segments of DNA in 1983 (Döring and Starlinger, 1984; Shure et al., 1983). The TEs were subsequently found in genomes of almost all living organisms, including yeast (*Saccharomyces cerevisiae*) and the fruit fly (*Drosophila melanogaster*), examples of the most intensively studied organisms in this respect. However, for decades the importance of TEs was not fully appreciated and they were sometimes judged as "junk DNA" (Ohno, 1972). Nowadays TEs and especially the LTR retrotransposons are considered important being responsible for changes in genome size and organization in evolution (Biémont and Vieira, 2006; Choulet et al., 2010; Hawkins et al., 2009) and also for serving as the origins of new promoters, gene domains such as for nucleic acid binding, and of mechanisms such as the V(D)J recombination in the immune system (Jones and Gellert, 2004).

In this work we concentrate on *BARE1*, the first LTR retrotransposon described in barley, discovered in 1993 by our group (Manninen and Schulman, 1993). *BARE1* is the most abundant, dispersed TE component of the barley genome and is actively transcribed in barley tissues (Suoniemi et al., 1996a; Suoniemi et al., 1996b; Wicker et al., 2009). It can

be activated by stress conditions such as drought (Kalendar et al., 2000). Furthermore, *BARE1* was shown to be the major factor in genome size dynamics in the *Hordeum* genus, where genome size variation results from loss through intra-element recombination and gain through *BARE1* integration (Vicient and Schulman, 2005). Thus *BARE1* presents an exciting object study of the lifecycle of LTR retrotransposons.

1.2 Classification of transposable elements

Transposable elements comprise two main groups (Figure 1), Class II, or DNA transposons, and Class I, the retrotransposons (Wicker et al., 2007). The retrotransposons that contain LTRs and the retroviruses resemble each other greatly (Adams et al., 1987). As a result, two classification systems have developed in parallel, where the LTR retrotransposons are classified either as TEs or viruses, depending from which side one is looking at them. The International Committee on Taxonomy of Viruses ((http://ictvonline.org) status on June 2012) classified the retrotransposons into the *Pseudoviridae* family, which at the moment includes the *Hemivirus*, *Pseudovirus*, and *Sirevirus* genera that each contain individual TEs as virus species (Havecker et al., 2004; Peterson-Burch and Voytas, 2002). However, a more detailed classification for transposable elements, which includes retroviruses as members, was also created elsewhere (Wicker et al., 2007).

The first classification system was proposed in 1989 by Finnegan (Finnegan, 1989), who divided the TEs according to their transposition intermediate: RNA (Class I or retrotransposons) or DNA (Class II or DNA transposons). Commonly, these classes were referred to "copy-and-paste" for retrotransposons and "cut-and-paste" for transposons. In the first, the original element is left in the donor site and the copy is integrated into a new site; for the "cut-and-paste" the element leaves the donor site and is reintegrated elsewhere. Later, as several new TEs were discovered that did not comfortably fit into this system, such as MITEs (miniature inverted repeats) (Jiang and Wessler, 2001) and DNA transposons exploiting the "copy-and-paste" mechanism, a new classification was proposed in 2007 for eukaryotic transposable elements (Wicker et al., 2007). Here the classification utilizes the Finnegan system, the retrotransposons in Class I and the DNA transposons in Class II, but the level of subclass is used to distinguish "copy-and-paste" and "cut-and-paste" mechanisms (Figure 1). At the moment only the Class II has both these two subclasses. Moreover, the 2007 system provided a standardized naming convention needed for annotation in the many ongoing sequencing projects.

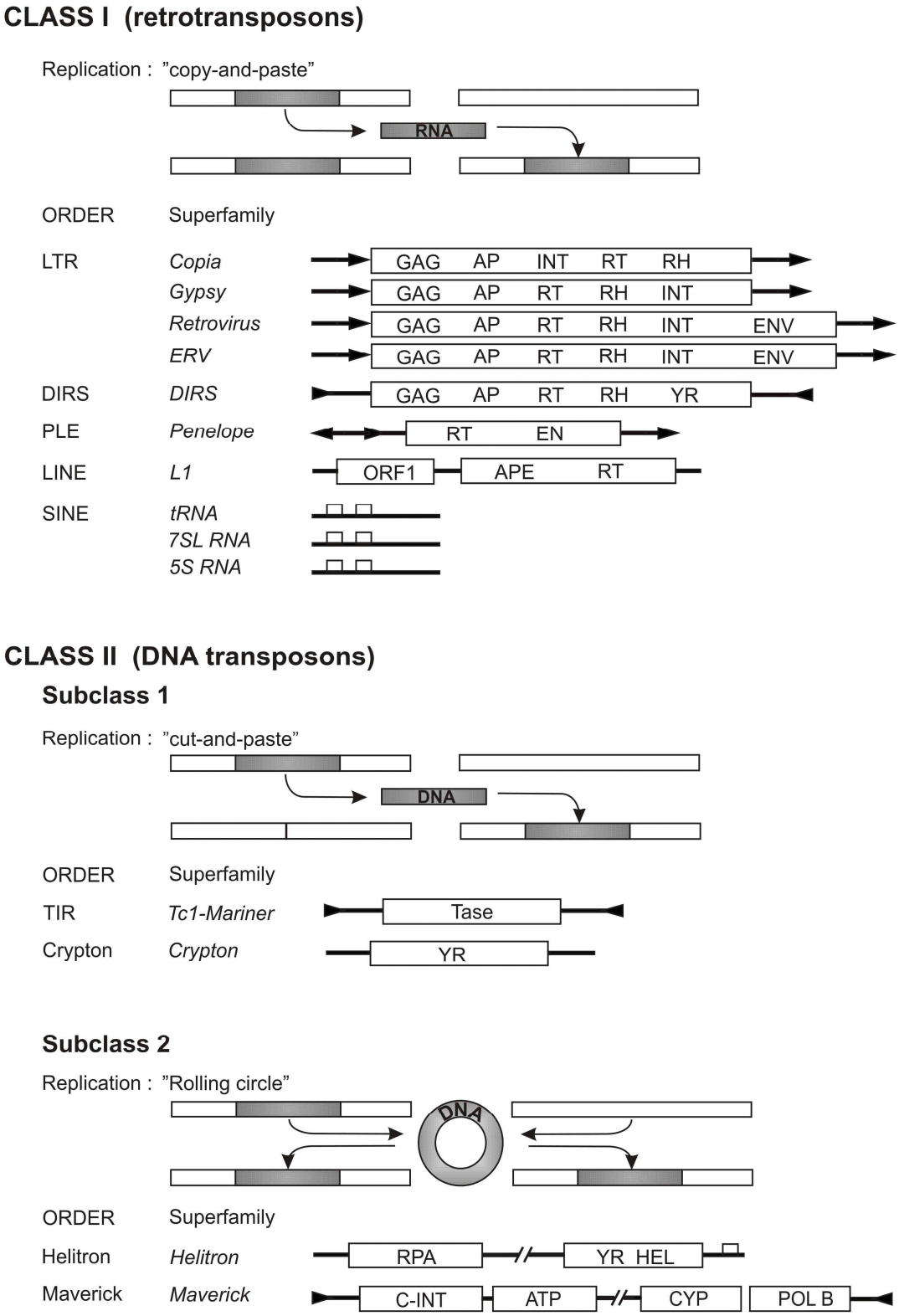


Figure 1. Classification of transposable elements into retrotransposons (Class I) and transposons (Class II). The overview of the replication mechanism of each class is shown, accompanied by a few examples of the main orders and superfamilies. Class I elements replicate by a "copy-and-paste" mechanism through an RNA intermediate, leaving the original copy in place, whereas the

Class II Subclass 1 transposons move by a "cut-and-paste" mechanism, thus leaving the original site empty. The Subclass 2 transposons replicate by a "rolling circle" mechanism that leaves the original copy in place but, like Subclass 1, there is no RNA intermediate. Abreviations: AP, aspartic proteinase; APE, apurinic endonuclease; C-INT, C-integrase; CYP, cysteine protease; EN, endonuclease; ENV, envelope protein; GAG, capsid protein; HEL, helicase; INT, integrase; ORF, open reading frame; POL B, DNA polymerase B; RH, RNase H; RPA, replication protein A; RT, reverse transcriptase; Tase, transposase; YR, tyrosine recombinase. Long terminal repeat, terminal inverted repeat, region that can contain additional ORFs, coding region, non-coding region, diagnostic feature. Adapted from Wicker et al., 2007.

1.2.1 Class I

Class I transposons are also called retrotransposons because they reverse the normal flow of the Central Dogma by reverse transcribing their RNA intermediate back into DNA. The genomic copy of a retrotransposon is transcribed as messenger RNA (mRNA) by RNA polymerase II and then reverse-transcribed back to DNA by a reverse transcriptase they encode. Class I transposons are divided into five orders, LTR, DIRS, PLE, LINE, and SINE, depending on the structures of their genomic DNA copies (Figure 1). The LTR retrotransposons are most abundant in plants, including crop species such as barley and wheat. The order contains two main superfamilies, *Copia* and *Gypsy*, as well as the endogenous retroviruses (ERVs) and retroviruses. It is the *Copia* superfamily, into which the *BARE1* retrotransposon, a member of the *BARE* family investigated in this study, belongs. The DIRS (*Dictyostelium* intermediate repeat sequence) are elements found in a broad range of organisms from algae to animals and fungi. They utilize a tyrosine recombinase enzyme instead of INT and have either split direct repeats or inverted repeats at their ends (Goodwin and Poulter, 2004). The PLE (*Penelope*-like elements) are mostly found in *Drosophila* but also in other animals, plants, and fungi. They contain direct or inverted LTR-like sequences, and an RT enzyme resembling telomerase more than those found in other TEs (Evgen'ev and Arkhipova, 2005). The LINEs and SINEs (the non-LTR retrotransposons) are predominantly found in mammals, including human, but more rarely in plants. The LINEs encode at least an RT enzyme and nuclease, although sometimes a *gag*-like ORF is observed. The ends are often truncated at the 5' end, probably due to incomplete reverse transcription. The 3' ends display poly(A) tails, tandem repeats, or simply A-rich regions (Wicker et al., 2007). The SINEs are a unique group, originating from accidental retrotransposition of various short (80 – 500 bp) polymerase III transcripts such as tRNA, 7SL RNA, and 5S RNA. They have their own internal promoters for transcription and rely on an RT enzyme encoded by a LINE element for their replication. The best-known SINE is the numerous *Alu* element, which is found in half a million copies in the human genome (Rowold and Herrara, 2000; Wicker et al., 2007).

1.2.2 Class II

Class II transposons are also found in almost all eukaryotes as well as in prokaryotes. They are further divided into two subclasses depending on how many DNA strands are cut during the transposition process, both strands in the "cut-and-paste" or single strands in "Rolling circle" (Kapitonov and Jurka, 2001) mechanism. The latter has also been called "copy-and-paste" (Wicker et al., 2007) or "cut-and-copy" (Schulman, 2012) mechanism. However, the Class II elements are not transcribed into an RNA intermediate as are Class I elements (Wicker et al., 2007). Subclass 1 elements comprise the orders of TIR and *Crypton*, which utilize the classical "cut-and-paste" mechanism. TIR elements have terminal inverted repeats (TIRs) which are recognizes by a transposase enzyme carried by the element. The transposase recognizes the TIRs and by cutting both DNA strands at the TIR ends cleave the element from the donor site. The *Crypton* elements lack TIRs and utilize a tyrosine recombinase enzyme in their movement (Wicker et al., 2007). Subclass 2 elements move by replication, without a double-strand cleavage. These "cut-and-copy" DNA transposons include the orders of *Helitron* and *Mavericks* (Schulman, 2012). The *Helitron* elements are found mainly in plants, mostly so far in maize, and in animals and fungi. They utilize a tyrosine recombinase enzyme and appear to replicate via a rolling circle mechanism (Kapitonov and Jurka, 2001; Poulter and Goodwin, 2005). The *Mavericks* are large TEs reaching 10 to 20 kb in size. They encode several proteins, including DNA polymerase B and an INT, with limited similarity to DNA viruses. Thus far, these elements have been found sporadically in eukaryotes other than plants (Pritham et al., 2007).

1.3 Structure of LTR retrotransposons

The LTR retrotransposons are named for the long terminal repeats (LTRs) that flank both their ends (Figure 2). The length of the LTR can vary between a hundred to several thousand base pairs, and intact copies almost always terminate into short inverted repeats 5'TG…CA'3, similar to retroviruses. The LTRs contain promoters, terminators and RNA processing signals for the transcription of the retrotransposons. Between the LTRs, most of the remaining sequence constitutes a domain typically encoding proteins in one or two open reading frames (ORFs), respectively as a single *gag-pol* –frame or as separate *gag* and *pol* –frames. Typically the gene order is 5'LTR-*gag-ap-rt-rh-int*-3'LTR for the *Gypsy* superfamily. In the *Copia* superfamily the order of *rt-rh* and *int* is inverted: 5'LTR-*gag-ap-int-rt-rh*-3'LTR. The *gag* encodes a single structural protein, the GAG, a capsid protein which assembles into the shell of the VLP. The *pol* encodes a polyprotein, which comprises three enzymes, the aspartic proteinase (AP), integrase (INT), and reverse transcriptase – RNaseH (RT-RH) (Kumar and Bennetzen, 1999; Wicker et al., 2007). The POL and/or GAG-POL polyprotein is cleaved by the AP into functional peptides.

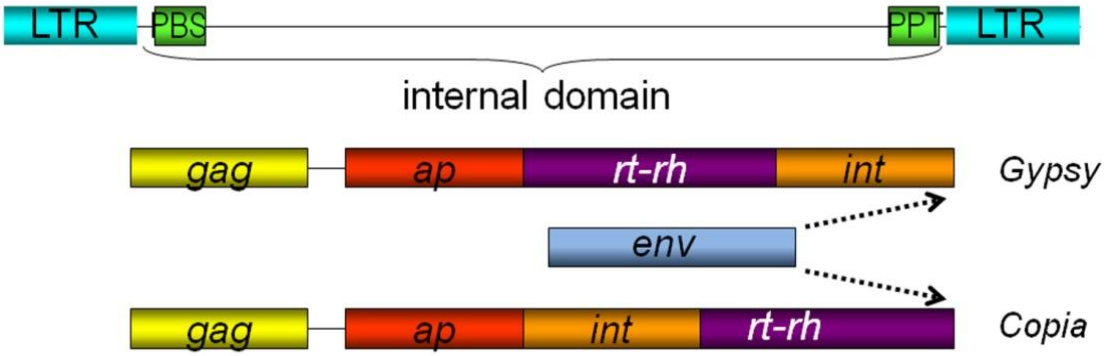


Figure 2. Structure of *Gypsy* and *Copia* superfamily LTR retrotransposons. The internal domain is located between the LTRs and carries PBS and PPT motifs for replication at its 5' and 3' ends, respectively, and protein coding domain composed of genes for *gag*, *ap*, *int*, and *rt-rh*. The *Copia* and *Gypsy* are distinguished by the order of the *int* and *rt-rh* genes. Some retrotransposons contain an *env* gene, the location of which, at the end of the protein coding domain, is depicted by arrows in the figure (Schulman and Wicker, in press).

1.3.1 LTR Retrotransposons and retroviruses

The LTR retrotransposons are considered as the ancient intracellular precursors of retroviruses (Beauregard et al., 2008; Biémont and Vieira, 2006). Retroviruses and LTR retrotransposons share a similar mechanism of replication by reverse transcription and transcription; the retrovirus gene order is similar to that of the *Gypsy* superfamily. The main difference in retroviruses is the additional *env* gene that is typically located between the *int* and 3'LTR and encodes an envelope (ENV) protein enabling their extracellular, infectious lifecycle (Eickbush and Jamburuthugoda, 2008). The retroviruses have evolved from an LTR retrotransposon, probably a *Gypsy*, after acquisition of the *env* gene (Beauregard et al., 2008; Eickbush and Jamburuthugoda, 2008; Wicker et al., 2007). Although loss of the *env* gene from a retrovirus can yield a new retrotransposon (Ribet et al., 2008), the retrotransposons are suggested to be more ancient than retroviruses because of the greater diversity (Eickbush and Jamburuthugoda, 2008). Thus the retroviruses are likely to have evolved through the gain of function of an LTR retrotransposon.

Although plant LTR retrotransposons are not considered to be infectious, a clade of *Gypsy* elements encode *env* (Vicient et al., 2001) and the *gypsy* elements in *Drosophila* were found infectious under laboratory conditions (Kim et al., 1994). The LTR retrotransposons form VLPs similarly to retroviruses by assembling their polyprotein precursors around their RNA molecule. The RNA-binding domain in the GAG is homologous to a retrovirus nucleocapsid protein, which forms the virus RNA – protein core particle, for a subset of LTR retrotransposons (Sandmeyer and Clemens, 2010) those of including *BARE1*

(Manninen and Schulman, 1993). The structure of the INT of *BARE1* is highly similar to HIV-1 and ASV viruses (Suoniemi et al., 1998a). In both LTR retrotransposons and retroviruses, the translation and encapsidation of the genomic (g)RNA is in conflict because the encapsidated gRNA is reverse transcribed and destroyed at the end. Although translation can precede reverse transcription, some viruses such as Murine Leukemia Virus (MLV) use distinct RNA pools for these activities, whereas others such as HIV-1 and -2 do not (Dorman and Lever, 2000; Messer et al., 1981).

1.3.2 Retrotransposon *BARE1*

BARE1 was the first retrotransposon described for barley (*Hordeum vulgare* L.), hence the name for BArley RetroElement 1 (Manninen and Schulman, 1993). *BARE1* was found to be a major, active retrotransposon component of the genome (Suoniemi et al., 1996a; Suoniemi et al., 1996b). The *BARE* family has three members: the *BARE1*, a fully autonomous element, *BARE2*, the non-autonomous variant, and *BARE3*, a third member, which is closest to *WIS-2* of wheat. Full-length copies constitute about 2.9 % of the barley genome, dispersed throughout except in the centromeric, telomeric, and nucleolar organizer regions (NOR) (Suoniemi et al., 1996a). Most commonly, *BARE1* is found in repetitive DNA, forming clusters and nested insertions. The lifecycle of the *BARE1*, as for TEs in general, has been thought to be entirely intracellular as it lacks the third ORF encoding the ENV.

1.3.2.1 Structure of *BARE1*

The first described *BARE1* (Manninen and Schulman, 1993; Accession Z17327) is 12088 bp long, but it contains a 3135 bp insertion in its 3' LTR. Therefore, a canonical *BARE1* element is predicted to be around 8.9 kb long (Vicient et al., 1999a) (Figure 3). The *BARE* family is placed into the *Copia* superfamily of LTR retrotransposons by the order and similarities of the proteins encoded in the polyprotein, the GAG, AP, INT, and RT-RH (Wicker et al., 2007). *BARE* is transcribed from the promoter in the 5' LTR which contains two TATA boxes, TATA1 and TATA2. The TATA2 was found to be responsible for reporter gene expression in protoplasts, and for native *BARE1* protein expression in leaf protoplasts (Suoniemi et al., 1996b). Predicted from the sequence, the polyprotein is translated from a single ORF (Tanskanen et al., 2007). The RNA is expressed in ten classes, five from each TATA box (Chang and Schulman, 2008). *BARE1* was also found expressed actively in somatic cells in barley. An abscisic acid (ABA) response element was found downstream the TATA boxes (Suoniemi et al., 1996b).

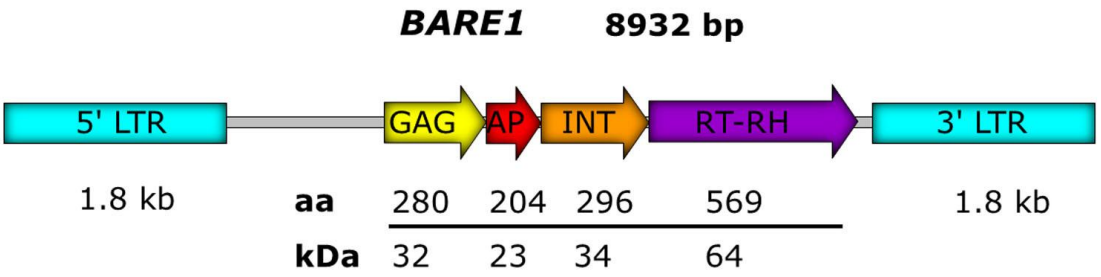


Figure 3. The structure of the *BARE1* retrotransposon. The length of the nucleotide sequence of the whole element and its LTRs, the predicted lengths of the amino acid sequences (aa) and molecular weights (kDa) of the individual proteins, GAG, AP, INT, and RT-RH are shown.

1.3.2.2 LTRs

The LTRs are 1829 bp long in *BARE* (Manninen and Schulman, 1993). The LTR is two to six times longer than that for majority of the superfamily *Copia* retrotransposons (Vicient et al., 1999a). The LTR ends contain 6 bp perfect inverted repeats flanked by 4 bp direct repeats in the host DNA. The LTR contains the promoter and two TATA boxes for transcription as well as termination, polyadenylation, and capping signals.

1.3.2.3 UTL

The region between the LTRs is the internal domain, which contains the protein coding region. Between the transcription start and translation start codon lies the 5' untranslated leader sequence (UTL). Like the LTRs, this region is also unusually long, about 2 kb, in *BARE* (Vicient et al., 1999a). The UTL is suggested to down-regulate the expression of the *BARE* because in experiments with truncated UTLs the expression is more efficient (Suoniemi et al., 1996b). The 5' UTL also contains primer binding site (PBS) at the 5' end, well conserved for $tRNA_i^{Met}$, used to prime the (-)-strand synthesis during reverse transcription (Suoniemi et al., 1997), and the dimerization signal (DIS) and packaging signal (PSI) signals, putative for *BARE*. In retroviruses such as HIV-1, the UTL contains the gRNA DIS, in the form of "kissing loop" motif, and the PSI for encapsidation of the gRNA into the VLP (Paillart et al., 2004). The second untranslated region, the UTR, is found at 3' end of the transcript. This region contains the PPT, adjacent to the 5' end of the 3' LTR, used for priming the (+)-strand cDNA synthesis.

1.3.2.4 Protein coding domain

Downstream of the UTL, the *BARE1* internal domain encodes a predicted polyprotein containing the conserved GAG, AP, INT, and RT-RH regions associated with retroviral and retrotransposon replication (Vicient et al., 1999a). The *BARE1* transcript contains at least 51 putative AUG codons within the 5' UTL (Suoniemi et al., 1998a). Multiple upstream ATGs are found in retroviruses, plant viruses, and heat shock proteins that use internal ribosome entry site (IRES) for translation initiation. This, combined with the long UTL structure that inhibits the expression of reporter genes linked to it, suggests that the translation of the *BARE1* polyprotein may use IRES (Suoniemi et al., 1996b).

1.3.2.4.5 GAG

The GAG protein is the only structural protein of LTR retrotransposons and retroviruses responsible forming the VLP or virion, and packaging the gRNA. The *gag* and *env* gene are the most rapidly changing of all retroviral genes. The *BARE1 gag* is 843 nt in length and shows little similarity to those of other LTR retrotransposons or retroviruses except for the conserved motif $CysX_2CysX_4HisX_4Cys$ at its C-terminus. This zinc-finger-like motif is identified as invariant in retroviruses, plant pararetroviruses, and retrotransposons such as *copia* in *Drosophila*, *del* 1-46 in lilly, and *Tnt1* in tobacco (Manninen and Schulman, 1993). The motif is responsible for binding nucleic acids and, in LTR retrotransposons such as yeast *Ty3* and in retroviruses, is located in a nucleocapsid (NC) domain of the GAG (Larsen et al., 2008). The other domains in retroelement GAG proteins are the matrix (MA) and capsid (CA), although *Ty3* also has a CA domain, separated from the NC by a spacer. The GAG precursor of the extensively studied HIV-1 is MA-CA-SP1-NC-SP2-p6 (the SP meaning a spacer, and p6 an additional specific protein for HIV). In HIV-1, the MA is responsible for contact with membranes through positively charged residues and a myristoyl moiety added post translationally and for trimerization of the protein. The CA is responsible for polymerization of the GAG after binding to RNA or to membranes. The NC is responsible for further GAG – GAG interactions and binding RNA through two zinc-finger-like motifs and a stretch of basic residues (Adamson and Jones, 2004). Whether *BARE1* GAG contains domains with similar functions we do not know. However, the N-terminal part of *BARE1* GAG shows similarity to the N-terminal domain of HIV-1 MA by *in silico* structure predictions. The region immediately upstream of the C-terminal zinc-finger domain shows characteristics of a spacer. This region consists of many basic residues, cannot be folded *in silico* into any secondary structure but is predicted to be solvent-accessible, and separates the predicted CA and NC of *BARE1* (unpublished results).

1.3.2.4.6 POL

The POL polyprotein is encoded by the *pol* gene, which in *BARE1* continues from the *gag* gene in same reading frame. The *pol* encodes the rest of the *BARE1* proteins: AP, INT, and RT-RH (Sabot and Schulman, 2006). The AP functions to cleave the polyprotein into its mature-sized components, the GAG, INT, and RT-RH. The *BARE1* AP has sequence similarities with other retrotransposons and retroviruses and contains three domains, the conserved aspartic proteinase active site (DTG), substrate binding, and enzyme backbone domains (Manninen and Schulman, 1993; Peterson-Burch and Voytas, 2002). In retroviruses such as HIV-1, the proteinase further processes the GAG into the MA, CA, and NC. The retroviral GAG can dimerize (Johnson et al., 2002), and as the retroviral AP is active as a dimer, it is assumed that the correct alignment of GAG molecules (in the polyprotein) would allow the activation of the AP, triggered by the microenvironment inside the actively closing particle (Adamson and Jones, 2004). The INT recognizes the ends of the LTRs and inserts the cDNA into the host genome. The *BARE* INT is well conserved in sequence and displays the catalytic Glu of the DD-35-E motif universal in transposases, in integrases of viruses such as HIV-1, and in retrotransposons. The backbone of INT in the tertiary structure model show close similarity to the HIV-1, even though the sequences are only 14 % identical (Suoniemi et al., 1998a; Vicient et al., 1999a). The function of RT-RH is to reverse-transcribe the RNA transcript, starting from the $tRNA_i^{Met}$ primer at the PBS, into cDNA and, through formation of the second primer at the PPT by the RNaseH moiety, to initiate and polymerize the plus strand cDNA and restore the LTR ends. Just as in other superfamily *Copia* retrotransposons, *BARE* contains two strongly conserved domains in the RT-RH that have served in phylogenetic studies (Flavell et al., 1992; Manninen and Schulman, 1993; Suoniemi et al., 1997).

1.4 LTR Retrotransposon lifecycle

The LTR retrotransposons lifecycle, depicted in figure 4, is based on a “copy and paste” mechanism in which the mother copy is left in place and new daughter copies are inserted back into the host genome (Frankel and Young, 1998; Sabot and Schulman, 2006). The lifecycle starts with the transcription followed by translocation of the gRNA into the cytoplasm, translation of the proteins, VLP assembly, replication of the gRNA into cDNA by reverse transcription, transport into the nucleus, and insertion back to the genome.

1.4.1 Activation and transcription

The first regulatory step in the retrotransposon lifecycle is the transcription of gRNA from the promoter in the 5'LTR. Because this RNA is transcribed by the cellular RNA polymerase II as any other cellular mRNA they, as do retroviruses, rely heavily on host factors such as transcription factors in their lifecycle (Albar et al., 2006). The retrotransposons can be induced by various cellular and environmental stresses, although also can be inhibited by them (Menees and Sandmeyer, 1996). Stress activation has been studied especially in tobacco (Beguiristain et al., 2001; Moreau-Mhiri et al., 1996), as reviewed in (Grandbastien, 1998), and for integrated endogenous pararetroviruses (Iskra-Caruana et al., 2010). The activation is controlled by the cell by several mechanisms, such as RNA interference or antisense RNA (Matsuda and Garfinkel, 2009), the cell cycle (Karst et al., 1999; Shi et al., 2007), and epigenetic silencing by methylation of the DNA. Opposite to the latter, DNA demethylation also has been shown to activate retrotransposons (La et al., 2011; Liu et al., 2004; Mirouze et al., 2009) and, depending on the levels of silencing, some elements may have permanent background activities (Vitte and Bennetzen, 2006).

1.4.2 Translation

Like transcription, also translation of LTR retrotransposons is generally accomplished as for any other cellular gene. Features needed for translation, in particular the 5´cap and 3´poly(A) tail, are added. In eukaryotes, the cap is formed by addition of a 7-methylguanosine at the 5' end; the poly(A) tail is formed by a polymerase that adds about 200 residues of adenosine nucleotides at the 3' end (Mathews and Holde, 1991; Van Der Kelen et al., 2009). This is, however, not always necessary, as many plant viruses exploit cap- and poly(A)-independent translation (Dreher and Miller, 2006; Kneller et al., 2006). Like cellular mRNAs, the transcripts are transferred into the cytoplasm and translated into polyproteins.

The AP cleaves the polyprotein into the mature-sized proteins GAG, INT, and RT-RH. The stoichiometry of the expressed proteins is critical because the GAG is required in excess amounts for the assembly of the VLPs. There are several alternative mechanisms for this: programmed frameshifting at the *gag – pol* junction, resulting in the GAG-POL – polyprotein being expressed in lower amounts than the GAG (Haoudi et al., 1997; Voytas and Boeke, 1993); deletion of the entire *pol* region from the transcript by differential splicing, which has been reported in *copia* in *Drosophila* (Brierley and Flavell, 1990); post-translational degradation of the POL proteins (Irwin and Voytas, 2001). Nevertheless, half of all examined retrotransposons encode single ORF (Gao et al., 2003).

1.4.3 Splicing of the RNA for excess GAG production

In plant genomes, about 80 % of the nuclear genes are interrupted by non-coding introns. The generation of functional mRNA from these intron-containing, precursor mRNAs (pre-mRNA) involves excising introns through a process termed pre-mRNA splicing. In plants, the introns are short, averaging 433 bp in rice, compared to humans where they average 3000 bp. The introns are rich in U or UA dinucleotides, whereas the exons are rich in G nucleotides. The exonic and intronic regions are recognized by a large RNA-protein splicing complex called spliceosome. The splicing reactions are carried out by the spliceosomal RNA, whereas the proteins assist in increasing the fidelity in proofreading the splicing signals that are recognized in the pre-mRNA. In general, higher eukaryotes have two types of spliceosomes, the major U2 type and the minor U12 type. The U2 type spliceosome consists of U1, U2, U4, U5, and U6 small nuclear ribonucleoprotein particles (snRNPs) and catalyzes the removal of introns with canonical GT-AG splice sites. The U12 type spliceosome contains U11, U12, U4atac, U5, and U6atac snRNPs and recognizes only less than 1 % of introns in *Arabidopsis* and humans (Reddy, 2007).

The splicing mechanism is also used to generate two or more mRNAs from the same pre-mRNA by using different splice sites. This mechanism termed alternative splicing can produce protein isoforms with loss or gain of function and altered cellular localization, protein stability, enzyme activity, or post-translational modifications (Reddy, 2007). Alternative splicing has been reported for the *Drosophila copia* retrotransposon, where it regulates the ratio of GAG to the POL proteins. The translation of *gag-pol* is slow compared to that of the alternatively spliced RNA for GAG (Brierley and Flavell, 1990). In *BAGY2*, an envelope-class retrotransposon, the *gag-pol* region is removed almost entirely in the alternatively spliced RNA, allowing expression of the downstream *env* gene (Vicient et al., 2001). In *Ogre*, a member of superfamily *Gypsy*, the region between the *gag-pr* ORF and *rt*/*rh-in* ORF is spliced out in preference for GAG expression (Steinbauerová et al., 2008). These examples show that the retrotransposons have evolved to exploit the cellular mechanisms of their host to optimize their replicative life cycle.

1.4.4 VLP assembly

Much of the knowledge of LTR retrotransposon VLP assembly comes from the members of *Copia* and *Gypsy* superfamilies in yeast and *Drosophila*. The assembly of retrotransposon VLPs and retroviral particles is as diverse as their sources. The common dogma for all retrotransposons and retroviruses is that the VLPs are formed of GAG (Figure 4). These VLPs, for one element, can be highly polydisperse in size as shown for the yeast *Ty* (transposon in yeast) retrotransposon (Burns et al., 1992; Palmer et al., 1997). The particle sizes are shown to depend on the length of C-terminal part of the GAG (Al-Khayat et al., 1999). Moreover, the VLP structure can be also porous as shown for *Ty*

VLPs that have large holes allowing access to the gRNA by RNases (Al-Khayat et al., 1999). The VLP assembly may, furthermore, involve various host factors, which are necessary in various phases of the retrotransposon (Downs and Jackson, 1999) and retrovirus lifecycles (Albar et al., 2006).

The assembly of the VLPs may or may not require the presence of the gRNA. Typically two gRNAs are packaged. For this, the gRNA contains the DIS motif as well as PSI motif, which is used for selectively packaging the gRNA into the VLP by GAG. Little is known about how GAG selects the PSI, but the binding happens through hydrogen bonding (Gao et al., 2002) similarly as in retroviruses. In retroviruses, the gRNA is required as a structural "scaffolding" element (Muriaux et al., 2001). Nevertheless, GAG can form retrovirus VLPs on its own and, in the absence of the correct gRNA, the VLPs are filled with cellular RNAs (Rulli et al., 2007), as has been shown also for phages (Legendre and Fastrez, 2005; Pickett and Peabody, 1993).

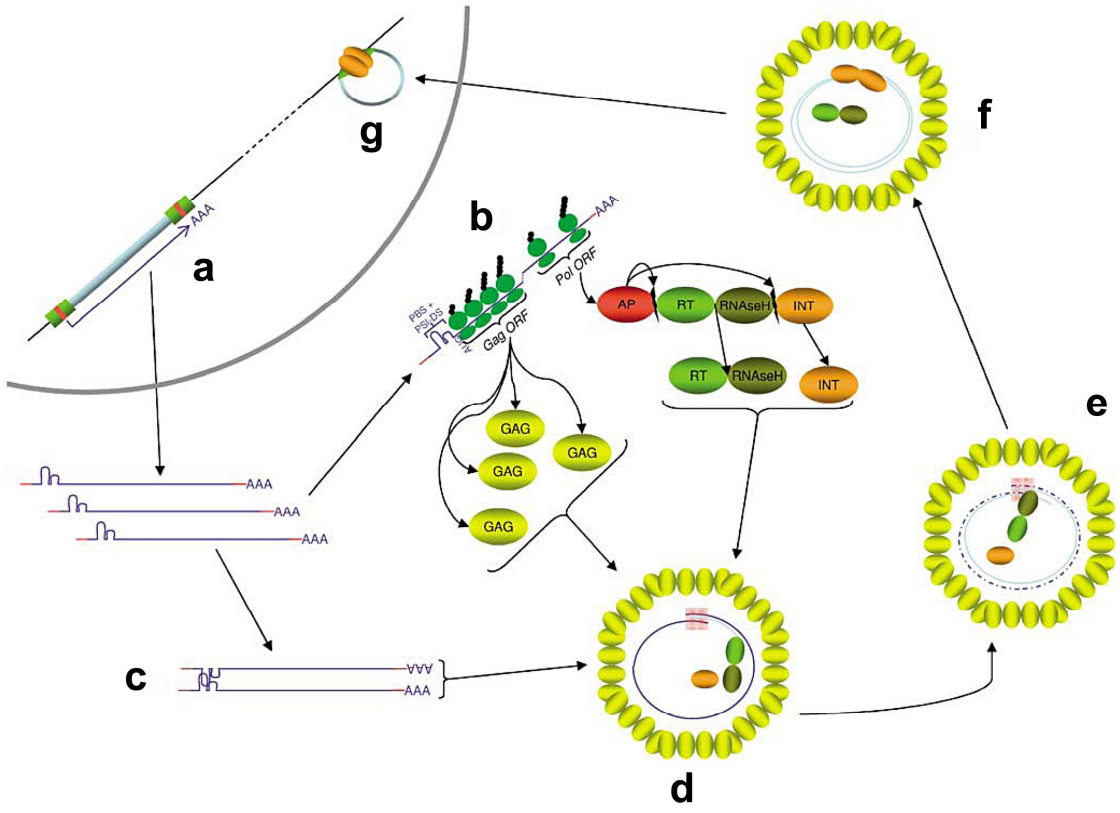


Figure 4. Theoretical lifecycle of the LTR retrotransposon. (a) Transcription of the mRNA starting from the 5' LTR R region and ending at the 3' LTR R region, and transport into cytoplasm. (b) Translation of the proteins GAG and POL. The POL is further cleaved by the proteinase to yield the RT-RH and INT. (c) Dimerization of the gRNA for packaging. (d) Formation of the VLP by the GAG and packaging of gRNA and the enzymes RT-RH and INT. (e) Reverse transcription into cDNA and degradation of the gRNA. (f) Completion of the transcription into double-stranded cDNA. (g) Transport into nucleus and enzymatic reactions to insert the cDNA copy of the retrotransposon into a new genomic location (Sabot and Schulman, 2006).

1.4.5 Replication

In general, the GAG assembles into VLPs together with the gRNA, which in turn is reverse-transcribed by the RT-RH into duplex cDNA (Figure 5). The gRNA has the overall composition: 5'-R-U5-PBS-coding region-PPT-U3-R-3' (R= repeat, U5= unique 5' sequence, PBS= primer binding site, PPT=polypurine tract, U3= unique 3' sequence). This gRNA serves as a template for the RT, and the reaction is primed usually with the 3' end of a cellular tRNA, which matches to the PBS (Figure 5, A). In rare cases, other mechanisms such as self priming (Atwood-Moore et al., 2006) or *trans* priming (Haag et al., 2000) has been reported. Similar to VLP assembly, assisting host factors may also be needed as shown for yeast *Ty1* and *Ty3* (Karst et al., 2000). The reverse transcription may also require the presence of the INT as shown for *Ty1* (Wilhelm and Wilhelm, 2006).
The cDNA synthesis proceeds, from the tRNA 3' end at the PBS through the U5 and R regions until it reaches the end of the gRNA molecule generating the "strong stop" (-) strand cDNA (Figure 5, B). The RNaseH then digests the template gRNA from 5' end, thus revealing a single-stranded cDNA 3' end complementary to the R region of the gRNA 3' end (Figure 5, C). The first template switch occurs when the R regions of the cDNA and the 3' end of the gRNA hybridize (Figure 5, D). Reverse transcription to generate the (-) strand then continues to proceed along the gRNA (Figure 5, E). The RNaseH digests the gRNA template as the cDNA is synthesized. Simultaneously, the (+) strand synthesis is primed by RNA fragments at the PPT, and the DNA is synthesized until the end of U5 in the (-) strand cDNA template to generate the (+) strand strong-stop cDNA (Figure 5, F). The second template switch then occurs (Figure 5, G), and the double-strand cDNA is completed (Figure 5, H) (Sabot and Schulman, 2006; Schulman and Wicker, in press; Telesnitsky and Goff, 1997). The lifecycle is completed by the INT that inserts the cDNA back into the genome following the steps described below.

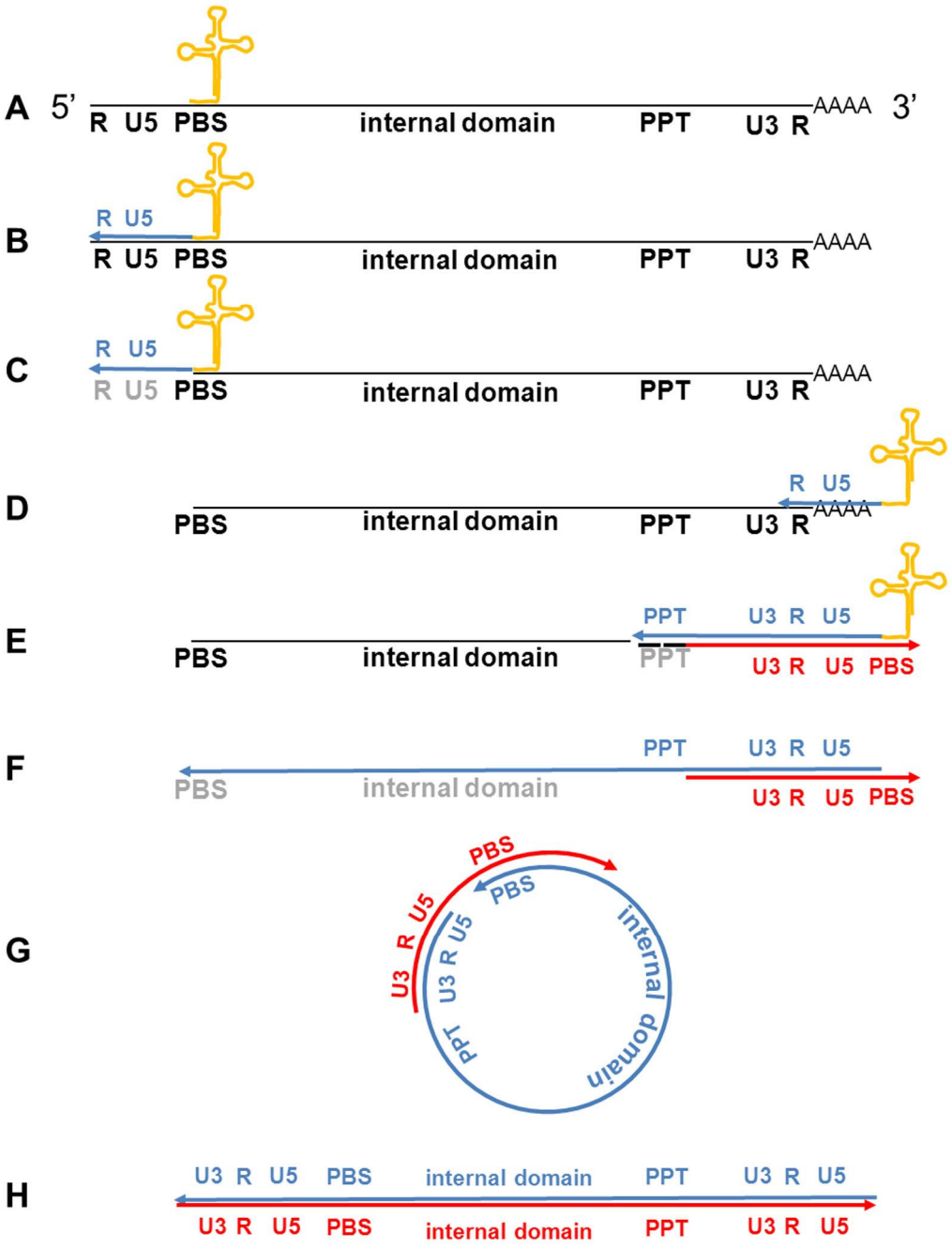


Figure 5. LTR retrotransposon replication. Binding of the tRNA onto the PBS site in gRNA (A), first strand synthesis creating a "strong stop" (-) cDNA (B), digestion of the template gRNA at the DNA-RNA hybrid (C), first template switch through pairing of the R regions of the cDNA and the 3'end of the gRNA (D), continuation of reverse transcription of the (-) strand cDNA and priming of the (+) strand cDNA at the PPT after digestion of the gRNA (E), the gRNA is completely digested and "strong stop" (+) strand cDNA is formed (F), second template switch (G), and completion of the double strand cDNA synthesis (H) (Schulman and Wicker, in press).

1.4.6 Transport into the nucleus and integration

The cDNA is transferred to the nucleus within the VLP and inserted back into the host genome by the INT that recognizes the LTR ends and the target site in genomic DNA (Sabot and Schulman, 2006; Sandmeyer and Clemens, 2010; Wicker et al., 2007). Transport to the nucleus requires either a nuclear localization signal (NLS), which is generally located in the GAG or INT, or direct contact with the nuclear pore factor by the GAG (Balasundaram et al., 1999; Dang and Levin, 2000; Kenna et al., 1998). The importance of the NLS for transport has been shown in yeast. Because the nuclear membrane in yeast is permanently intact and does not break down during mitosis, yeast retrotransposons need to have a functional NLS. In *Ty1*, the NLS is in the INT and the element utilizes the classical nuclear protein import pathway (McLane et al., 2008). In *Ty3*, not only does the VLP interacts directly with the nuclear pore complexes through GAG, but also the INT has an NLS (Beliakova-Bethell et al., 2009). The insertion sites are generally within euchromatin and vary depending on the retrotransposon in question. The *Gypsy* family shows strict target DNA selection, while other LTR retrotransposons show no specificity.

1.4.7 Autonomous and non-autonomous elements

All transposable elements can lose some of their activities by mutations, and thus become non-autonomous, unable to move within the genome by themselves (Figure 6). The high error rates of transcription and reverse transcription are the source for the observed high variability between individual copies, resulting in point mutations, premature stop codons, and frame shifts (Boutabout et al., 2001). However, the lost activities, such as transposase in Class II or reverse transcriptase in Class I, can be provided *in trans* by the autonomous members of the same family that are still functional or by members of other families (Sabot and Schulman, 2006). The autonomous *BARE1* has such a non-autonomous partner, *BARE2* in barley that has lost the GAG function. Nevertheless, the *BARE2* has retained the critical LTR end structures and signals for its replication: PBS, DIS, and PSI as well as the POL region. Thus, *BARE2* needs only the GAG of *BARE1* (Tanskanen et al., 2007).

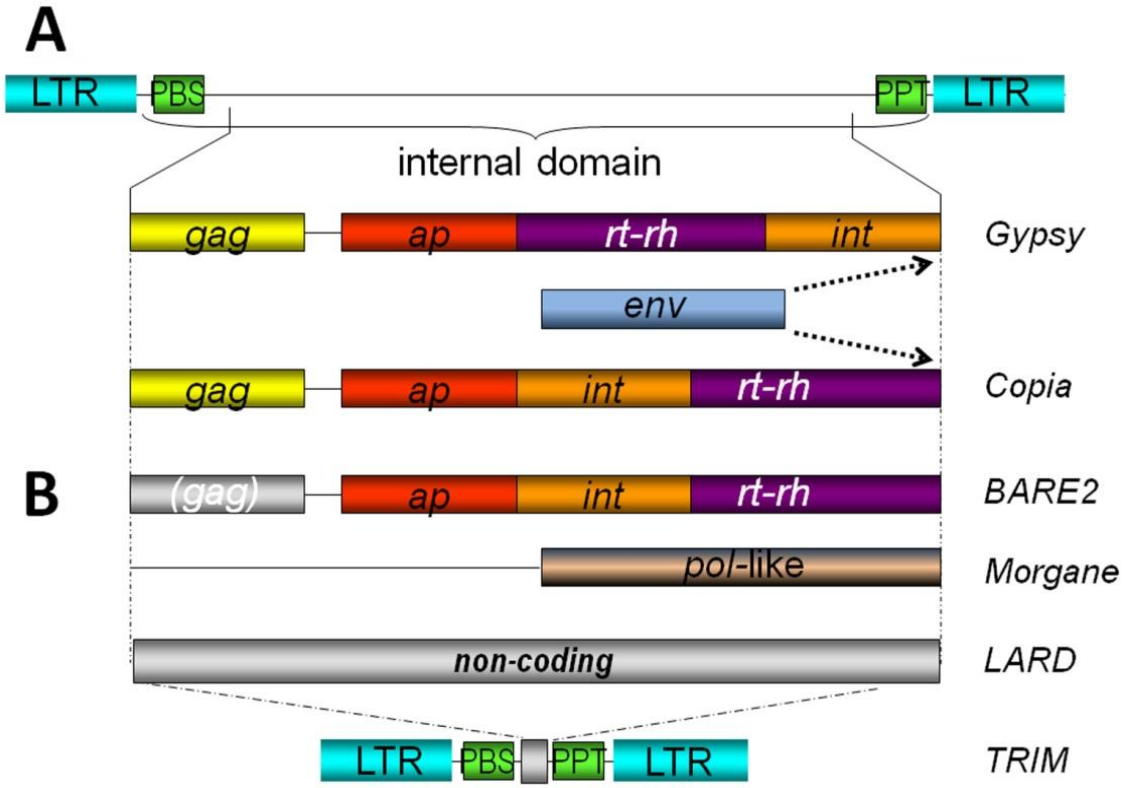


Figure 6. Autonomous (A) and non-autonomous (B) LTR retrotransposons in plants. The *BARE2* lacks the functional *gag* gene. Other examples of non-autonomous elements are *Morgane*, which have a degenerate or truncated, but recognizable open reading frame; *LARD* (Large Retrotransposon Derivative) elements which have a long internal domain with conserved structure but no coding capacity; and *TRIM* (Terminal Repeat retrotransposon In Miniature) elements which have no internal domain except the PBS and PPT signals (Schulman and Wicker, in press).

1.5 Impact of retrotransposons on genomes

The completion of the LTR retrotransposon lifecycle results in multiple new copies inserted into new locations in the host genome. The insertions can be anywhere, although nested insertions into places already occupied by TEs may be preferred because insertions into or next to cellular genes may lead to inactivation or otherwise abnormal activities. The retrotransposons can also be "domesticated" in such a way to gain benefit for the cell. This and other examples are reviewed in (Gogvadze and Buzdin, 2009). Combining the replicative nature of their mobilization and their large size (5 to 10 kb), the LTR retrotransposons have great potential to influence genome size. As a result, cereal genes appear as small "gene islands" within massive "repeat seas" (Ananiev et al., 1998; Panstruga et al., 1998; SanMiguel et al., 1996) with synteny and collinearity being maintained between species (Bennetzen, 2000; Keller and Feuillet, 2000; Moore et al.,

1995). As the genome size expands through retroelement proliferation, the LTRs of the elements provide targets for intrachromosomal ectopic recombination between LTRs of the same or different elements, thus allowing the genome to shrink. This process leaves solo LTRs behind (Vicient et al., 1999b) that, in the yeast genome, comprise up to 85% of all retroelements (Kim et al., 1998 138). In barley, the solo LTRs of *BARE1* are 7- to 42-fold more abundant than full-length elements (Vicient et al., 1999b).

1.6 Retrotransposons as molecular markers

The integration of LTR retrotransposons creates new joints between the element and genomic DNA that can be detected as molecular fingerprints. Because the LTR retrotransposons use a copy-and-paste replication method, variation in the fingerprints generally manifests newly integrated copies. However, sequence variations at the priming sites due to mutation and insertions, or deletions, or recombination between the priming sites can create fingerprint polymorphisms that are not due to retrotransposon activity. The molecular fingerprints themselves are revealed as products generated by PCR using primers matching the LTRs facing away from the internal region of the LTR, combined with primers matching features in the genomic DNA. These features can be repetitive sequences such as microsatellites (in the retrotransposon-microsatellite amplified polymorphism [REMAP] method; (Kalendar et al., 1999)), another retroelement (in the inter-retrotransposon amplified polymorphism [IRAP] method; (Kalendar et al., 1999)), or a restriction site adapter (in the sequence-specific amplified polymorphism [SSAP] method; (Waugh et al., 1997)). We used the IRAP PCR technique that amplifies regions between two retrotransposons to reveal the polymorphism and possible candidates for new insertions.

2 Aims of the study

The *BARE* has been earlier shown to be transcribed in barley tissues such as leaves and also in cultured cells (Suoniemi, 1996b). The aims of this study were to investigate whether *BARE* and LTR retrotransposons in general in grasses are transcribed, translated, form VLPs, and are integrationally active. These aims and major techniques are listed here below.

1) Study the transcriptional, translational, and integrational activities of retrotransposons and *BARE* among grasses. Retrotransposon EST database searches, antibodies to *BARE1* GAG for protein detection, and IRAP PCR techniques were used to detect these activities.

2) Study the locations the *BARE* is active in barley *in* vivo. Because plants do not set aside a germ line, the lifecycle must be carried out in tissues giving rise to gametes. Immunoblots, immunolocalization, and *in situ* hybridizations were used to detect *BARE* GAG and transcripts.

3) Study the activity of *BARE* transposition machinery in barley, the formation of the polyprotein, and the components it carries, including the GAG and the INT. For this, antibodies to GAG and INT were produced and used for detection on immunoblots.

4) Study whether *BARE1* forms VLPs in barley *in vivo*. For this we used retroviral RNA encapsidation and VLP formation as a model in our attempt to isolate VLPs from barley. Immunoblots, PCR techniques, and electron microscopy were used to detect and visualize the VLPs.

5) Study how the excess levels of GAG are formed for VLP assembly and detect the RNA pools for translation and encapsidation by using reverse transcription PCR techniques on various extracts and VLP preparations from barley.

3 Materials and methods

Table 1. Methods used in this work are listed here. The details are described in the original publications.

Method	Described and used in			
Antibody production	I	II		
Bacterial expression	I	II		
Culture of barley cells *in vitro*		II		IV
Cloning	I	II		IV
DAPI staining			III	
Electron microscopy		II		
Electrophoresis of nucleic acids	I	II		IV
EST database searches	I			
Fixation and paraffin embedding of plant tissues			III	
Immunoblotting	I	II	III	
Immunohistochemistry			III	
In situ hybridization			III	
IRAP	I			
Plant DNA isolation				IV
Plant RNA isolation				IV
PCR	I	II		IV
Polyribosome isolation				IV
Protein purification	I	II		
REMAP		II		
Reverse transcriptase assay		II		
RLM-PCR				IV
RT-PCR				IV
SDS-polyacrylamide gel electrophoresis	I	II	III	
Sectioning of plant tissues			III	
Sequencing				IV
Sucrose gradient ultrasentrifugation		II		IV
Virus like particle isolation		II		IV

4 Results and discussion

We studied *BARE* retrotransposon activity in several grass species, mainly barley, on the levels of transcription, translation, and VLP formation in order to better understand the *BARE* lifecycle. Based on the earlier studies, *BARE* is actively transcribed in barley (Suoniemi et al., 1996b) from conserved LTRs containing two TATA boxes. The *BARE* has also been found to have contributed to the differential increase of the genome size in *Hordeum* species (Vicient et al., 1999b). Since the first reports of active retrotransposons in plants (Hirochika, 1993; Pouteau et al., 1991; Royo et al., 1996; Suoniemi et al., 1996a) many more have been found (Todorovska, 2007).

4. 1 Active retrotransposons are a common feature of grass genomes

In previous work, we showed the *BARE* retrotransposon is actively transcribed in barley (Suoniemi et al., 1996b; Suoniemi et al., 1997), thus having the potential to influence the genome size. We wanted to study how widespread this phenomenon is among other grasses with large genomes. For this, we examined evidence for *BARE* and other retrotransposon activity by systemically searching EST databases for transcripts, and by applying immunoblots to detect translational products to confirm the viability of the *BARE* transcripts.

4.1.1 Retrotransposon transcription is a common phenomenon especially in grasses

Wide-scale analysis of retrotransposon transcription became possible when large-scale expressed sequence tag (EST) databases were developed and made publicly available. They offered a new source to look for transcriptionally active genes in organisms. Each EST represents a short stretch of an mRNA transcript generally extending from the 3' end. We searched these databases for homologies to the LTRs or internal domains of known retrotransposons and found 934 matches (1.2 ‰) of the total ESTs (I, Table 1). However, taxonomically representative sets of ESTs for monocots and dicots were not available at the time of our study, and were (perhaps still) skewed towards crop species. The monocots, mostly grasses, showed the highest average fraction of retrotransposon-containing accessions (1.75 ‰) compared with the two conifer species (1.40 ‰) and dicots (0.92 ‰). This expression of retrotransposons in grasses can be described as moderate (on a scale of high, moderate, rare) according to *in silico* transcriptional profiling of EST data sets (Bortoluzzi et al., 2000).

We used sequences of known retrotransposons, 10 from dicots and 27 from monocots, including 14 from superfamily *Copia* and 18 from *Gypsy*, as query sequences against the

EST database summarized in Table II (I, Table II). The activity of retrotransposons inferred as EST matches was observed to be low and no inter-generic matches were found for any dicots. In contrast, the monocots showed 22 times more EST matches, the *BARE1* being among one of the most numerous (I, Table II). Furthermore, members of the *Copia* and *Gypsy* superfamilies, when used as queries, find related and expressed elements in grass species other than their origin.

4.1.2 Retrotransposon EST abundance correlates with genome size

It has been shown earlier that the retrotransposon number and genome size is positively correlated in barley (Kalendar et al., 2000; Vicient et al., 1999b). Because transcription is the first key step in the lifecycle ultimately leading to integration, we assumed that expression of the retrotransposons as witnessed by ESTs may roughly correlate with the genome size. Indeed, we observed a strong and significant correlation between genome size and plant group, the grasses having larger genomes, and weak but not significant association between genome size and ESTs matching retrotransposons. In dicots alone, the correlation between genome size and EST fraction is both strong and significant (r_p = 0.895, P = 0.006). There are, however, two caveats: first, the EST databases were constructed from many tissue types but the new retrotransposon insertions are heritable only if they occur in those tissues giving rise to gametes. Second, the ESTs are derived from a mixture of cDNA constructions and sequencing methods, producing reads of varying length, and are in general partial in nature rather than full-length cDNA sequences. As a result some retrotransposon transcripts may not have been detected by our searches. It is also impossible to differentiate retrotransposon transcripts originating from LTRs and read-through transcripts from cellular promoters into solo LTRs or into adjacent full-length retrotransposons. However, the tendency of grass genes to cluster into gene islands would decrease the likelihood of read-through transcripts arising from gene promoters, at least for most retrotransposons in this group of plants.

4.1.3 *BARE1* transcripts are translated in various grasses

Transcription does not necessarily mean that the gene is also expressed as a protein. Because of this, we investigated further whether retrotransposons giving rise to the ESTs that were similar to *BARE* in other grass species are also translated. The *BARE1* retrotransposon contains a single ORF (I, Figure 1), which encodes the classical retroelement polyprotein components of the *Copia* superfamily (Wicker et al., 2007): GAG, integrase (INT), aspartic proteinase (AP), and reverse transcriptase-RNaseH (RT-RH) (Manninen and Schulman, 1993). To study the expression of these proteins, we raised two antisera to the subcloned and expressed *BARE1* GAG proteins. The first antiserum was directed to the N-terminal half of the GAG protein (II, Figure 1) and the second antiserum to the full-length GAG protein (I, Figure 2).

We recognized *BARE* proteins from all extracts from barley, wheat, rye, oat, Lyme grass, and rice. The immunoresponse were specific, with both antibodies compared with the preimmune sera (I, Figure 2 right). The full-length anti-GAG detected proteins of 150 kDa, 90 kDa, and 31.5 kDa, corresponding directly to the predicted sizes of 146.9 kDa for the *BARE* polyprotein and 90 kDa for the processed form after endoproteolytic cleavage of the RT-RH domain. The predicted size for GAG is 32 kDa. The GAG detected by the antiserum raised to the partial-length GAG was 31 kDa (I, Figure 2), slightly smaller than the one detected with the full-length anti-GAG. This differing but specific immunoresponse is evidently due to the variation in the *BARE1* elements and is detectable because the peptides used as antigens in antiserum production had only 85 % identity in the overlapping GAG region. Furthermore, the GAG is shown to be the most variable protein among all retrotransposons (Eickbush and Jamburuthugoda, 2008). Thus, the protein recognition patterns may reflect variations in the processing kinetics of distinct *BARE* subfamilies. Proteins of variable molecular mass of 34 kDa from rice as well as 53 and 54 kDa were also detected and appear to be equivalents of the GAG or its processing products. These results are the first evidence for pools of retrotransposon polyproteins in plant cells as well as the first that the *BARE* translation products are sufficiently well conserved for immunological cross reaction among the *Gramineae*.

4.2 Localization of the *BARE1* retrotransposon shows tissue-specific expression in barley *in vivo*

Because retrotransposons are under purifying selection, and to get inherited, they need to carry out their lifecycle in tissues giving rise to gametes. Hence, we wanted to study in more detail in which tissues *BARE* is active. The organs we studied represent multiple tissue types mixed together, creating similar problems as in the EST work, where we may have missed some active TEs because weak signals from single cells could have been diluted by strong signals from their neighbors (Paper I). This time, we used a combination of immunolocalization and *in situ* hybridization to examine *BARE* expression precisely on a tissue-specific level to reveal this part of the *BARE* lifecycle.

We started by examining the presence of the GAG in total protein extracts from barley by immunoblotting with the full-length anti-GAG antibody. We detected the 150 kDa polyprotein, 90 kDa intermediate, and the 31 kDa mature size GAG in virtually all tissues: callus, developing embryo, germinating embryo, scutellum, root tip, root stele, leaf, internode, node, shoot apical meristem, and flower (III, Figure 1B). However, the relative proportions of each protein form varied among these organs. The polyprotein was nearly missing in internodes and weakly detected in root steles. The mature GAG was most abundant in both developing and germinating embryos, scutellum, and nodes. The 90 kDa intermediate showed the most stable relative amounts across the samples. In addition, there

were weakly detected 54 kDa, 55 kDa proteins, identical in size to the predicted GAG-AP intermediate, and a 64 kDa protein of unknown identity.

The variation of the levels of *BARE* proteins between barley organs suggested that the *BARE* may be differentially expressed in a tissue-specific manner. This supported the need for a detailed study of the tissues. Next, we carried out immunolocalizations on the same barley organs with the GAG antibodies in order to pinpoint in which cell type the *BARE* proteins are expressed. The GAG showed intense localization in apical meristematic regions in root tips and shoot apical meristems (SAMs). This was consistent with our expectation that *BARE* would be expressed in cells that can clonally give rise to gametes. Developing floral primordia as well as prefertilization ovules, scutellum, and aleurone of the germinating embryos also displayed intense localization.

4.2.1 Roots

In roots, particularly in the root cap, root initial cells, and the cell files of the cortex that are expected to develop into vasculature, a strong GAG localization was seen (III, Figure 2A). Preimmune serum controls showed no localization (III, Supplemental Figure 1C). The localization spanned the whole stele of the root tip at the meristem (III, Figure 2B) and resembled the locations of auxin transport in *Arabidopsis* (Marchant et al., 1999). In general, the localization spanned the whole cytoplasm. However, in the root cap, the GAG was also found in cytoplasmic granules (III, Figure 2E), which are most likely amyloplasts containing starch granules serving as statoliths, as they were not found in other root tissues. While the GAG seemed to be restricted to cytoplasm, it was still also present in the perinuclear region during stages of nuclear division (III, Figure 2F-G) and thus it may have access to the chromatin at certain stages (III, Figure 2H). The nuclei seemed to be devoid of GAG as shown by the combination of immunolocalization and DAPI staining (III, Figure 2E-H) and, furthermore, the amounts of mature GAG detected in immunoblots were low compared with that of the polyprotein. Taken together, the localized protein in root meristem cells is most likely the *BARE* polyprotein. The missing of the mature GAG protein may be explained by hindrance in the polyprotein processing, faster turnover of the GAG, or *BARE* RNA splicing.

4.2.2 Shoot apical meristems and embryo

The GAG also displayed strong expression in the axillary SAM at 18 days after germination (III, Figure 3A). The GAG was localized in the meristematic tissues that give rise to the leaf primordia, as well as in inflorescences which had developed into rudimentary spikes. The preimmune serum control showed no localization (III, Figure 3B).

The GAG was also abundant in the parenchyma adjacent to the provascular tissue as well as in the mature vascular tissues spanning across the parenchymatous ground tissue, which was devoid of GAG (III, Figure 3A). The GAG seems to accumulate during the development into the SAM and floral meristems because it was not found in an earlier stage, 4 days after germination in the embryonic shoot (III, Supplemental Figure 2A,B). However, at this stage, the scutellum, the aleurone, and the vascular tissues together with adjacent parenchymatous tissues showed intense staining with the GAG antibody (III, Supplemental Figure 2A,D). In addition, the localization region in the scutellum was separated from the embryo by a sharp layer of approximately two cells devoid of staining (III, Supplemental Figure 2C). Because of this sharp polarization in GAG localization, we examined the correlation of the protein and RNA expression in this tissue by *in situ* hybridization (III, Figure 4). In general, the amounts of the GAG and RNA corresponded, the scutellum showing more intense staining over the embryonic shoot. Interestingly, the scutellar cell layer devoid of GAG seemed to express the *BARE* RNA most intensively (III, Figure 4B,C,D) suggesting this RNA is not used for *BARE* protein synthesis.

4.2.3 Ovules

In *Arabidopsis*, the genome was shown to be demethylated in ovule cells and to express TEs (Slotkin et al., 2009). Because of this phenomenon, we examined barley pre-fertilization ovules for *BARE* GAG and RT protein expression. The ovule tissues including chalaza, inner- and outer integument, antipodal cells, nucellus, and provascular tissues displayed strong localization with the full-length GAG antibody (III, Figure 5B). The RT was localized in the same tissues, but more weakly than for GAG as expected for an enzyme.

4.2.4 Vascular tissues

Because the GAG was localized to the provascular tissues in developing flowers and embryos, we examined vascular tissues also from adult plants. Here, the GAG was found strongly localized in the vascular bundles of internodal tissue as seen in cross sections (III, Figure 6A,B,D,E) and longitudinal sections (III, Figure 6G,H). The preimmune serum control showed no localization (III, Figure 6C,F,I). The axillary stems showed the most intense localization, whereas in the main stem the localization is specific to the companion cells of the phloem sieve tubes (III, Figure 6D,E). The companion cells showed dense staining also in their nuclei (III, Supplemental Figure 5), suggesting that GAG is being targeted into the nucleus in these cells.

4.3 Expression of *BARE* protein components in barley *in vivo*

While the *BARE* was found to be transcribed and translated in various grasses in a constitutive manner, we wanted to understand whether the *BARE* is also able to produce functional proteins in order to accomplish its lifecycle. Also, we wanted to find out whether the expression is increased by environmental factors. The treatment of cultured barley cells with various chemicals and environmental conditions that promote cellular stress did not increase the levels of the extracted GAG protein (unpublished results). Only drought treatments provided evidence for *BARE* retrotransposon induction in barley.

4.3.1 The effect of drought on *BARE* expression

Earlier findings suggested that the *BARE* is stress-induced not only because it has an abscisic acid (ABA) response element in its LTR (Suoniemi et al., 1996b), but also because the numbers of copies in the genome vary between moist and dry habitats (Kalendar et al., 2000). We examined the effect of drought on *BARE* protein expression in two barley varieties, cv. Bomi and cv. Golden Promise. We observed increase in the amount of GAG protein in the internode and node tissues of the drought treated cv. Golden Promise plants compared to the normally watered plants (III, Figure 7). In general, the cv. Golden Promise was also phenotypically severely affected by the drought, the whole plant wilted and flaccid. The drought had less severe effects on cv. Bomi, in which the leaves died in an orderly fashion, one by one, starting from the oldest ones. In these plants the *BARE* protein levels were not increased compared to the controls (data not shown). Only a little uncleaved polyprotein was detected in drought-treated plants whereas, in the normally watered plants, the polyprotein was readily detected. Interestingly, the levels of the 90 kDa intermediate remained unchanged throughout.

4.3.2 The *BARE1* GAG is already present in early stages of germination

In germinating grains, the GAG was localized particularly in scutellum, which serves the germinating embryo by secreting hydrolytic enzymes, absorbing glucolytic and proteolytic products, and translocating them to the embryo (Potokina et al., 2002). Among the many differentially expressed genes and transcription factors in the scutellum, the GAG accumulation pattern is similar to that reported for the SAD transcription factor, which is a DOF-class protein that binds the pyrimidine box and activates transcription of a GA-induced promoter (Isabel-LaMoneda et al., 2003). This is not surprising, as the *BARE* LTR contains 15 putative promoter elements (data not shown), including motifs for AMYBOX2, DOF, and GAMYB associated with the gibberelling response (Gaur et al.,

2011; Gubler et al., 1999; Haseneyer et al., 2008; Yanagisawa and Schmidt, 1999). Taken together, these observations suggest that *BARE* transcription could hence be gibberellin-regulated.

4.3.3 The expressed *BARE1* polyprotein is processed into functional GAG, INT, and RT proteins *in vivo*

The *BARE1* is expressed as a polyprotein that is cleaved into mature products including INT in barley. We raised antiserum also to the INT to detect whether the *BARE1* polyprotein processing is completed. The partial-length anti-GAG (we had not produced the full-length anti-GAG yet) and INT antisera recognized mature 32.5 kDa GAG and 34.3 kDa INT, virtually identical in size to the ones predicted from the *BARE1* sequence, in extracts from resting and germinating barley embryos as well as from leaves and cultured cells (II, Figure 1b,c). The INT antiserum was specific, showing no cross reactivity with the anti-GAG and *vice versa* (II, Figure 1a). This represented, for the first time, a demonstration that the translation products of a plant retrotransposon are sufficiently abundant to be detected immunologically *in vivo*.

4.3.4 The *BARE1* GAG protein is produced in excess compared to POL

The titer response of the partial-length anti-GAG and anti-INT antibodies were almost equal, showing similar immunoresponses against the *E.coli* expressed GAG and INT (II, Figure 1a). However, on immunoblots from barley extracts, the GAG showed a much stronger response compared to the INT (II, Figure 1b,c). This indicates the GAG is present at many times higher levels than INT *in vivo*, even though the GAG and INT appear to be expressed as part of the same polyprotein. This excess of GAG can be explained by its requirement in greater amounts than the INT and RT-RH for the VLP assembly, as has been shown for several non-plant retrotransposons (Haoudi et al., 1997; Voytas and Boeke, 1993).

4.3.5 Excess GAG is produced by alternatively splicing of transcripts

Retroviruses and retrotransposons often encode their GAG protein from a separate ORF for the production of the excess GAG needed for assembly of the VLPs. This is generally achieved by translational frameshifting of -1 or +1 nucleotides by the ribosome. We earlier pondered whether the *BARE* may use a putative frameshift in AP region to produce the excess GAG detected in immunoblots in relation to the INT. This is, however, inconsistent

with the other members of the *Copia* superfamily which have been shown to contain only a single ORF (Gao et al., 2003; Wicker et al., 2007) and also with the detailed examination of *BARE1*, which showed only one ORF (Tanskanen et al., 2007); hence, the question remained.

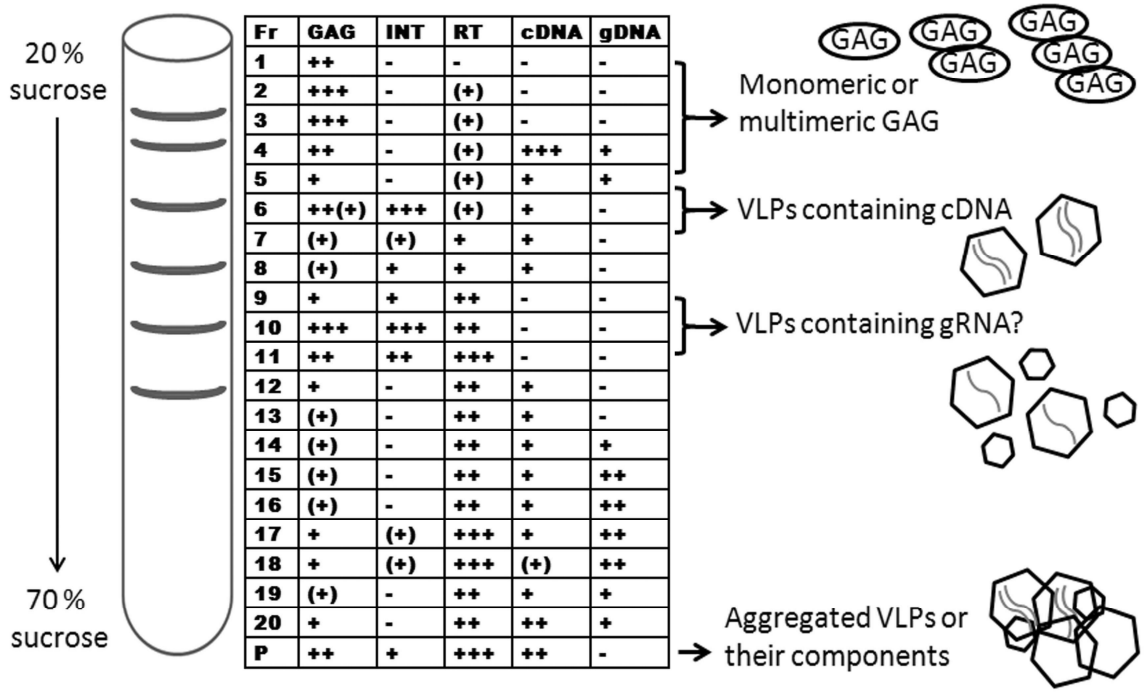
The amplification of *BARE* from total RNA, isolated from cells of cultured barley callus, with the primers LS1 and AP4 (IV, Table S1, Figure S1B) showed two products, a main one and a shorter one (IV Figure 1A). The shorter product was not detected among the products amplified from genomic DNA (IV Figure 1B). The shorter product was also amplified from RNA of barley embryos, leaves, and roots. Sequencing confirmed that the main product was derived equally from either *BARE1* or *BARE2* transcripts, whereas the shorter one was solely from *BARE1,* and represented 12.5 % of the total *BARE1* products. The shorter product contained a deletion of 104 nt at the beginning of AP domain, flanked by GT and AG at the left and right borders (IV, Figure S1A). There was no equivalent to the short form in the genome, suggesting the RNA is alternatively spliced. Furthermore, the GT…AG junctions match to the splice site consensus (Schuler, 2008) and were identified as well by *in silico* prediction. Moreover, the high content (33.6-37.5 %) of U nucleotides in the deleted part is typical for plant introns (Ko et al., 1998). The splice junction was found to be two nucleotides after the end of the predicted *gag* coding region, creating a stop codon three amino acids beyond the predicted end of the GAG. Thus, the spliced RNA can express only GAG, of which the predicted size (32.1 kDa) is virtually the same as we have detected earlier.

If the spliced and unspliced *BARE* RNAs are competent for protein expression, they are generally expected to display the features needed for translation, in particular the 5´cap and 3´poly(A) tail. We observed the transcripts of TATA2 to be capped (IV, Figure 2BC) and polyadenylated whereas the TATA1 transcripts were not. Furthermore, the TATA2 transcripts were 303-338 nt shorter due to the distance between TATA boxes, and thus lacked the R domain critical for reverse transcription, suggesting this RNA was dedicated for protein production only. To determine if the capping and splicing occur on the same RNA molecules, cap assays were performed on total RNA from callus followed by 5' RLM-PCR and a second round of PCR to detect the spliced and unspliced forms. We found that the fraction containing capped RNAs displayed the spliced and unspliced forms in similar proportions as the spliced and unspliced RNAs detected in total RNA (IV, Figure S2A), suggesting that only the TATA2 products are spliced. This was further confirmed by similar assays, in which the fraction of poly(A) RNAs showed identical relative amounts of splicing (IV, Figure S2B) but no uncapped transcripts (IV, Figure 2E). Taken together, these results indicate that both spliced and unspliced forms of TATA2 products are capped and polyadenylated. If the spliced form is used for excess GAG expression, it should be associated with polyribosomes. This was confirmed, because the RT-PCR using polyadenylated RNA isolated from the polyribosomes showed the same spliced band (IV, Figure 4A and B).

4.4 The translation products of *BARE* migrate as VLPs in sucrose gradients

Translation of a retrotransposon is generally followed by the assembly of VLPs. In non-plant retrotransposons such as the *Ty1* (Eichinger and Boeke, 1988; Hajek and Friesen, 1998) and in retroviruses such as the HIV-1 (Frankel and Young, 1998), VLPs have been shown to contain RNA or cDNA, RT, and INT within a capsid consisting of GAG. We used the *Ty1* and HIV-1 systems as conceptual models to study whether the *BARE* retrotransposons also assemble VLPs. The barley cell culture extracts showing the strongest GAG expression were fractionated using isopycnic sucrose density gradient ultracentrifugation to reveal whether the immunoreactive components would migrate together, which would indicate the presence of VLPs. We examined the total protein content, presence of GAG and INT immunologically, the activity of RT, and the DNA content in the gradient fractions. Light-scattering bands were observed, and used to trace the *BARE* components. We were unable to link the light-scattering material directly to *BARE* VLPs although the GAG and INT reproducibly was detected in these fractions. Because the RNA of the retrotransposon is reverse transcribed by RT, we examined the presence of cDNA in the fractions by PCR, using primers for the *gag* region. As a control, we assayed the presence of contaminating genomic (g)DNA using a REMAP PCR technique that amplifies regions between microsatellites or simple sequence repeats (SSRs) and the LTR ends in the genome (Kalendar et al., 1999). Because the retrotransposon cDNA terminates directly at the ends of the LTRs, this DNA will not be amplified by the REMAP PCR.

The result of a typical co-migration experiment for *BARE* components in a density gradient fractionation is summarized in Figure 7 (II, Figure 2 and 3). The majority of the total proteins and most of the GAG was detected in the four top fractions of the gradient. This GAG is interpreted to be monomers or small multimers. Deeper in the gradient, fraction 6 contained both GAG and INT as well as cDNA and some light-scattering material, but no RT activity. Further below these, fraction 10 contained the GAG, INT, RT, and the light-scattering material, but the cDNA was missing. The fractions 17 and 18 contained GAG, RT activity, and light-scattering material. The presence of key *BARE* components together in at least two distinct fractions suggests they are assembled into VLPs of at least two subtypes. One, in fraction 10, contains RNA because RT activity was detected but no cDNA. In the other, fraction 6, reverse transcription appears to have been completed as we detected the cDNA, and no RT activity. This suggests that the RT is released from the VLP. The INT is also shown to be required for reverse transcription in *Ty1* retrotransposons of Copia superfamily in yeast, where INT acts in *trans* to activate the RT (Wilhelm and Wilhelm, 2006).



Fr	GAG	INT	RT	cDNA	gDNA
1	++	-	-	-	-
2	+++	-	(+)	-	-
3	+++	-	(+)	-	-
4	++	-	(+)	+++	+
5	+	-	(+)	+	+
6	++(+)	+++	(+)	+	-
7	(+)	(+)	+	+	-
8	(+)	+	+	+	-
9	+	+	++	-	-
10	+++	+++	++	-	-
11	++	++	+++	-	-
12	+	-	++	+	-
13	(+)	-	++	+	-
14	(+)	-	++	+	+
15	(+)	-	++	+	++
16	(+)	-	++	+	++
17	+	(+)	+++	+	++
18	+	(+)	+++	(+)	++
19	(+)	-	++	+	+
20	+	-	++	++	+
P	++	+	+++	++	-

Figure 7. Analysis of *BARE1* components in isopycnic sucrose density gradients. The presence of GAG, INT, RT activity, cDNA, and gDNA in the gradient fractions (Fr) 1-20 and pellet (P) are indicated by "-" not detected, (+) barely detectable, and +, ++, +++ indicating the level of positive detection. The linear sucrose concentrations are shown on the left and the positions of the light scattering material observed in the sucrose after centrifugation is drawn within the drawing representing the ultracentrifugation tube. The VLPs were observed from fractions 6 and 10. Schematic representation of these VLPs are shown on the right.

In addition, there might be VLP assembly intermediates or preintegration complexes formed by individual components, because weak positive signals were detected in the neighbouring fractions. Moreover, the presence of cDNA in the absence of contaminating gDNA confirmed that the *BARE* transcripts are reverse-transcribed and at least some of the RT activity detected in the fractions comes from *BARE*. The RT activity assay we use to detect RT, being enzymatic, is not specific for *BARE* RT. Consistent with this, we detected RT activity all over the gradient, starting from fraction 7. The highest peaks of RT activity were seen in fraction 11, adjacent to the second major GAG and INT peak in fraction 10, and in fraction 17. The positions of the peaks of GAG, INT, cDNA, and RT activity in the fractions are reproducible over many replicates. The pellet typically contained more total protein as well as GAG, INT, and RT activity than did the fractions, which we interpret as resulting from aggregation, which is common for virus and VLP preparations.

4.4.1 VLP structures detected in barley

When we began our work, no VLPs had been reported in plants. In *Drosophila* and yeast, the formation of VLPs was documented, the yeast *Ty1* being perhaps the best characterized (Burns et al., 1992; Palmer et al., 1997; Roth, 2000). Because VLP formation is important for the lifecycle of retrotransposons (Kirchner and Sandmeyer, 1993; Merkulov et al., 1996), we studied whether *BARE* also forms VLPs. The fractions containing GAG were visualized in transmission electron microscopy by negative staining. We recognized VLP-like structures in fractions 6 and 10 (II, Figure 4). The particular structures detected in fraction 6 were spherical and approximately 40 nm in diameter and similar to those of reported earlier for *gypsy* MDG4 from *Drosophila* (Syomin et al., 1993). Fraction 10 contained two classes of particles, smaller (10.0 ±0.3 nm) and larger (~35 nm) almost identical to the negatively stained yeast *Ty1* VLPs (Burns et al., 1992; Palmer et al., 1997; Roth, 2000). Furthermore, *Ty1* also shows smaller (11-16 nm) particles composed only of GAG and larger (15-39 nm) particles containing also INT and RT. While we were not able to identify directly the visualized structures as *BARE* VLPs, because they did not withstand the immunoprobing procedures, their co-migration with the *BARE* components GAG and INT, together with RT activity, indicates that at least some of them are indeed *BARE* VLPs (Figure 7). There are also other transposable elements residing in the barley genome, such as *romani* (Suoniemi et al., 1998b), *BAGY-2* (Vicient et al., 2001), and other elements of the superfamilies *Gypsy* and *Copia* (Wicker et al., 2009), which may contribute to the presence of the visualized structures and the RT activity detected.

4.4.2 VLPs are packaged with TATA 1 transcripts

Finally, the question remains which one of the *BARE* RNAs are packaged into the VLPs. We showed that the VLPs in fraction 6 contained cDNA and no reverse transcriptase activity whereas in fractions 10 and 11 the situation was reversed. We isolated the VLPs and analyzed the ends of the RNA molecules associated with them. The RNA was isolated from pooled VLP fractions 9-11 from the sucrose density gradient (II, Figure 2). The 5' RLM-RACE was performed (IV, Figure 4C) and the products sequenced, revealing that only non-polyadenylated RNA having end points that correspond to the TATA1 products respectively of *BARE1* and *BARE2* are packaged into the VLPs. This was consistent with the TATA1 transcripts being the only RNA species that contain the R domain necessary for reverse transcription into cDNA, even though the conserved DIS and PSI sequences are also present in the TATA2 transcripts (unpublished, data not shown). Thus, we can conclude that the *BARE1* forms two types of VLPs. The first is particles containing GAG, INT, RT-RH, and gRNA from TATA1. These particles migrate deeper into the gradient and are thus more dense than the other type of VLPs, which contain GAG, INT, and cDNA. Evidently, the RT-RH is released from this second type of VLPs once it is no longer needed.

4.5 *BARE* appears integrationally active in many grasses

We studied the *BARE* insertional polymorphisms using IRAP method. In this method, primers matching the *BARE* LTR were used to generate PCR products between two elements near enough to each other to allow amplification. We detected several polymorphic PCR bands in each Tribe *Triticeae* accession tested, as well as in timothy and oat, which are members of the *Avenae* tribe, all in the *Pooideae* subfamily, and in *Spartina maritima* and *Spartina alterniflora* in the *Chloridoideae* subfamily (I, Figure 3). Rice, while showing translated protein products of *BARE*-like elements, produced only a single band in this study, suggesting that the *BARE* elements are either clustered or that the LTRs are dissimilar. The latter seems more likely because BLAST searches with the *BARE1* LTR-UTL sequence against the published rice genome showed no matches (Jääskeläinen, unpublished). Polymorphism may not necessarily manifest the new insertions, as described in the introduction (chapter 1.6, page 28). Nevertheless, in many cases where individual bands could be cloned and the empty sites found for the flanking DNA in accessions lacking the bands, we were able to establish the presence of new insertions (Belyayev et al., 2010).

Our data indicate that the *BARE* is capable of accomplishing its lifecycle, expressing proteins, assembling VLPs, replicating, and inserting back into the genome even today, not only in barley, but also in many species among grasses. The insertional activity expands the genome, a process requiring control by the cell because integrations are not entirely neutral in their effect. In case of *BARE*, this control apparently lies either on the transcriptional or post-transcriptional level for RNA or somewhere between VLP formation and integration because we saw no apparent limitation in translation. The *BARE*-like elements are estimated to have been insertionally active throughout the last 3 million years (Wicker and Keller, 2007) after the divergence of the above mentioned species from their last common ancestor. The similar results we obtained from other retroelements as well suggest that a broad activity of retrotransposons across the grasses may be a general phenomenon.

5 Conclusion and future prospects

LTR Retrotransposons occupy a large fraction of the genome in most eukaryotes. They have become the major constituents of the genome because of the replicative nature of their lifecycle, forcing the cellular genes into the minority. Their long co-evolution with the cell led into exploitation of cellular enzymes and mechanisms, while many cellular mechanisms evolved to suppress their potentially harmful proliferation. Moreover, because the LTR retrotransposons share strong similarities to retroviruses, suggesting a common ancestor, it is important to understand details of their lifecycle and their consequences. One of the most active LTR retrotransposons in barley, *BARE*, provided thus an interesting subject for the study of the LTR retrotransposon lifecycle.

In this study, the aspects of *BARE* transcription, translation, VLP formation, and the locations in which these activities take place in barley were investigated. We searched the EST databases for LTR retrotransposons and observed that transcription of retrotransposons, including *BARE*, is a common phenomenon in grasses. Protein analyses of barley, wheat, rye, oat, rice, and lyme grass showed for the first time that the *BARE* transcript pools included also fully functional ones without premature stop codons and encode sufficiently enough conserved proteins for immunological detection among wide range of species in Gramineae. Furthermore, polymorphisms were detected between the investigated species, which indicate of new insertions that occurred at different time points after the divergence from a common ancestor. However, a more thorough sequence-based bioinformatics approach is needed to analyze the gain and loss of new *BARE* retrotransposons, due to the various inferring factors in scoring PCR-based polymorphisms.

In barley cells, *BARE1* was observed to be translated into a polyprotein of predicted size, and processed into mature-sized GAG and INT. This was the first demonstration that a plant retrotransposon proteins are abundant enough to be detected immunologically *in vivo*. The assembly of VLP requires an excess amount of GAG compared to the other components of the POL. The *BARE1*, having only one ORF for the polyprotein, solved this challenge by having a proportion of its transcripts alternatively spliced. Three distinct classes of RNA transcripts were observed, one that is capped, polyadenylated and translated into the polyprotein, a second one lacking the cap and poly(A) tail destined for packaging into the VLP and for reverse transcription, and a third one, which is capped, polyadenylated, and spliced in such a way to translate only GAG. Taking the data together, a model for the RNA pools for translation and RNA packaging into the VLPs can be derived (IV, Figure S4). The *BARE* system, which we have described, is reminiscent of the MLV retrovirus (Messer et al., 1981) in having distinct pools for translation and reverse transcription, in contrast to the single pool used by HIV (Dorman and Lever, 2000).

We detected three types of VLPs formed *in vivo* in barley, which are correlated to the positions of *BARE1* GAG, INT, and cDNA in sucrose density gradients. Two of the VLP forms were linked to RT activity and contained no cDNA, whereas one form lacked the RT activity but contained the cDNA. Thus, *BARE* first forms a particle containing the gRNA, and then reverse transcribes this into cDNA within the VLP. The size of the detected VLPs vary greatly, possible due to various assembly forms or mutations in the GAG. The latter has been reported to have drastic effects on the VLP morphology, even increasing the VLP size eightfold (Brookman et al., 1995; Martin-Rendon et al., 1996; Merkulov et al., 1996). However, clear proof that the VLPs observed were actually *BARE* VLPs was not obtained, because the particles did not withstand immunolabeling.

For retrotransposon survival, it is important that the newly inserted copies are passed into successive plant generations. In this context, tissue in which the lifecycle is carried out is critical, as the retrotransposons are thought to be unable to leave the cell. Immunolocalizations for *BARE1* proteins in barley tissues were pursued already at the beginning of this thesis work, but only after finding an effective method to unmask the antigens we were able to show localizations. We found that both the root and shoot meristems expressed GAG, consistent with a strategy to follow the clonal cell line into gametes. In roots, the vascular initials and root cap had expression patterns resembling the auxin transport pathways in *Arabidopsis* (Marchant et al., 1999). In these cells there is, however, little GAG compared to the polyprotein, suggesting a hindrance in the AP processing step, which could be a replication control point. The AP could be inhibited in the roots as there are inhibitors there reported for plants (Headey et al., 2010). Similarly, also callus cells contain relatively little mature GAG compared to the polyprotein. Shoot apical meristems were sites for the most active GAG expression. Also the scutellum of germinating embryo and phloem companion cells of the vascular tissues had GAG.

In contrast to the roots, the relative amount of mature 31 kDa GAG in the embryo and scutellum is several times higher compared to other tissues such as the node and internode. Simultaneously, the 150 kDa polyprotein is almost absent, supporting either complete processing of the polyprotein and preservation of the GAG or movement of the GAG from elsewhere. In the vascular tissues, the GAG is either expressed or moved in to the phloem companion cells, as the nuclei of these cells displayed positive staining for the GAG. The movement is supported by the strong localization of the GAG along the vasculature in the axillary stems and at the ends of the main stems near the shoot apical meristems. If the movement is real, this raises then the question of what method *BARE1* uses to leave the cell into the phloem, because the canonical element lacks the movement protein, the ENV. To answer this, a chimeric element with a detectable tag could be constructed, transformed into barley, and then it whereabouts followed over time.

Drought was observed to increase the levels of GAG in the internodes and nodes in the barley cv. Golden Promise. This can be explained by the presence of an ABA response element (Suoniemi et al., 1996b) in the *BARE* LTR. Interestingly ABA is also translocated

in the phloem (Hartung et al., 2002) and is synthesized in phloem companion cells and xylem parenchyma cells as well as in root tips and hypocotyls (Koiwai et al., 2004) in response to water stress (Davies and Zhang, 1991). Taken together, the data suggest that *BARE* is activated and expressed in these cells. Furthermore, *BARE* can evidently complete its lifecycle more readily in plants experiencing drought stress because over the long run, the copy number of *BARE* correlates with water scarcity (Kalendar et al., 2000).

In germinating grains, the GAG accumulation pattern matched that for the SAD transcription factor. The presence of gibberellin response motifs in the *BARE* LTR suggests that *BARE* is gibberellin regulated. Whether this is in line with the strategy of *BARE* propagation or simply coincidental remains to be solved. The observation that embryos during early stages of germination do not express the *BARE* proteins in the SAM, but only following the transition into floral meristems, is in line with a general strategy for getting newly inserted copies inherited. It is generally held that uncontrolled replication is selected against due to mutagenic effects of integration. In meristems however, such insertions would end up killing cells, which could simply be replaced by the neighboring cells. The *BARE* family as well as other TEs are often found in "seas" of repetitious DNA between "gene islands" (Wicker et al., 2005). These repeat regions may be kept silent epigenetically (Saze and Kakutani, 2011) as well as post-transriptionally silenced by small RNAs (Slotkin, 2010; Slotkin and Martienssen, 2007). According to our results, at least some *BARE* elements in the genome are active. Thus, the control of the *BARE* lifecycle in barley may lie in the level of VLP assembly, reverse transcription or integration into the genome.

We plan to study *BARE* activity in the localized tissues. The formation of specific gRNA destined for packaging raises the question how this gRNA separated from the other RNAs destined for translation, and where in the cell the VLP formation happens. The packaging of the gRNA is not completely specific, as shown for retroviruses (Rulli et al., 2007). Thus, limitation of the assembly into certain locations in the cell would increase the likelihood that a correct RNA becomes packaged. Specific P-bodies are shown to be the assembly sites for yeast *Ty1* (Checkley et al., 2010) and *Ty3* elements (Beliakova-Bethell et al., 2006). To study this for *BARE*, methods such as correlative light-electron microscopy (CLEM) or fluorescence resonance energy transfer (FRET) could be used to reveal the locations where the gRNA and GAG are complexing. Expression of the element and individual new insertions may also happen in the meristems only in individual cells, thus would not be possible to detect with conventional PCR-based methods. For this, we plan to use a laser-capture microdissection method to detect the tissue and cell specific RNA levels as well as the chromatin states for epigenetics.

6 Acknowledgements

This work was carried out at the MTT/BI Plant Genomics Lab, a joint laboratory of MTT Agrifood Research Finland and the Institute of Biotechnology, University of Helsinki. Financial funding was provided by the Academy of Finland and the Finnish Doctoral Program in Plant Science (FDPPS) for research, together with the Chancellor's Travel Grant, which supported travel to scientific meetings.

The Institute of Biotechnology has been an astonishing and stimulating environment to work over the years, thanks to the Director, Professor Tomi Mäkelä, former Director Professor Mart Saarma, and all the people working here. Also the people in the Department of Biosciences, especially Professor Tapio Palva, who serves as my custos, Docent Pekka Heino, and Jussi Alho are warmly thanked for their help and guidance of my studies during this project. I would like to thank the FDPPS and Dr Karen Sims-Huopaniemi for their help in my studies and with many practical issues. Anita Tienhaara is thanked for the help with the layout of the cover.

My greatest and warmest thanks go to my supervisor Professor Alan Schulman, who gave me the opportunity carry out this study and supported, guided and inspired me during my years in the field of transposable elements. I especially thank Alan for the good spirit, environment, and freedom of our lab, which made me feel privileged to work there.

I warmly thank Professor Jari Valkonen and Professor Yrjö Helariutta for pre-reading the manuscript and giving valuable comments before the final publication of my thesis. I want to thank the members of my follow-up group, Professor Jari Valkonen and Professor Sarah Butcher, for all their guidance and feedback; these encouraged me greatly in my work. I would also like to thank Alan Schulman for proofreading the manuscript and help with the English grammar.

I would like to thank the people in the histology lab, especially Eija Koivunen who helped me many times by giving me good hints, for sample processing or sectioning of the paraffin blocks. The microscopy would not have been so successful without this help. I also spent considerable amounts of time working with the transmission electron microscopes over the years, although many of the localization results there are still awaiting publications. Thus, I also want to specially thank all the people at the Institute Electron Microscopy Unit. Director Docent Eija Jokitalo and the former Director, Professor Michael Hess, are warmly thanked for helping me in the planning the sample processing and microscopy, as are Mervi Lindman, Arja Strandell, Pirkko Leikas-Lazanyi, Dr Helena Vihinen, and Antti Salminen for help in sample processing and with all kinds of practical issues. I would like to thank Professor Dennis Bamford and his lab for teaching me about the ultracentrifugation and gradient fractionation, which stood as a cornerstone in my work. Professor Harri Savilahti is also thanked for the collaboration, discussions, and

guidance during the VLP work. I want to thank Dr Anna-Helena Saariaho (née Mykkänen), who was also co-author in my first paper, for joining the project and sharing the load by working with the enzymatic assays.

The Schulman research group working with the transposable elements has been like a transposon itself. We have moved from building to building and thus refreshed our lab shelves as well as our minds from time to time. The spirit of our group has been both great and international thanks to many visitors I have been privileged to work with. I want to thank all the present and previous members of our research group, especially Docent Ruslan Kalendar, Dr Carlos Vicient, Dr François Sabot, Dr Cédric Moisy, Dr Wei Chang, Jaakko Tanskanen, Elitsur Yaniv, Anne-Mari Narvanto, and Ursula Lönnqvist. I sincerely want to thank the former members of our group, Dr Annu Suoniemi and Tiina Arna, who gave me the first insights into the retrotransposons way back at the very beginning of my work. I would like to thank Dr Carlos Vicient, also a co-author on the first papers, for his help with the constructs for antibody production. We still have the Spanish sun in our lab. Special thanks to Docent Ruslan Kalendar, with whom I shared the office for a long time, for spontaneous discussions on all imaginable subjects both in and out of science. Humor was an unavoidable spice in all of our talks.

Many experiments necessarily do not flourish into publication, due to the unexpected behavior of the object of investigation, but I feel the people still deserve the warmest thanks: Dr François Sabot with whom I did yeast-3-hybrids, and Docent Moshe Finel with whom we expressed the proteins in Baculovirus. Professor Adrian Goldman is thanked warmly for the opportunity to try to crystallize the GAG protein. May there be a day when we understand why our proteins behaved as badly as they did. I am thankful to Dr Wei Chang with whom we did one of the important findings of the *BARE* lifecycle that led into manuscript. Warm thanks also to Dr Cédric Moisy, who came to our laboratory and joined the project. I will remember the good times in lab and the various leisure time activities you organized. It was great to sweat at the poker table or to “shoot” fellow lab mates. Thanks to Jaakko Tanskanen for being there for bioinformatics. Special thanks to Anne-Mari Narvanto, and Ursula Lönnqvist who helped me so many times in the lab. Thanks to Elitsur Yaniv, you were a good company and spirit maker!

Life outside the laboratory is an important thing to have, especially during the short summer. In this sense, I would like to thank my friends, especially Taneli and Laura, Janne, and Tapio with whom I have spent many holidays, travels to the ends of the world by bicycle, plane, or “hiking”, thus helping me annually to reload my batteries for the year ahead. Finally, I wish to express my deepest gratitude to my parents, Lea and Ari, who I am so lucky to have, who let me “experiment” with “real” plants in the garden that also helped me to switch off from work, and most importantly for the constant support and understanding - for everything.

Helsinki, October 2012

7 References

Adams, S.E., Mellor, J., Gull, K., Sim, R.B., Tuite, M.F., Kingsman, S.M., and Kingsman, A.J. (1987). The functions and relationships of Ty-VLP Proteins in yeast reflect those of mammalian retroviral proteins. Cell *49*, 111-119.

Adamson, C.S., and Jones, I.M. (2004). The molecular basis of HIV capsid assembly - five years of progress. Reviews in Medical Virology *14*, 107-121.

Al-Khayat, H.A., Bhella, D., Kenney, J.M., Roth, J.F., Kingsman, A.J., Martin-Rendon, E., and Saibil, H.R. (1999). Yeast Ty retrotransposons assemble into virus-like particles whose T-numbers depend on the C-terminal length of the capsid protein. Journal of Molecular Biology *292*, 65-73.

Albar, L., Bangratz-Reyser, M., Hébrard, E., Ndjiondjop, M.-N., Jones, M., and Ghesquière, A. (2006). Mutations in the eIF(iso)4G translation initiation factor confer high resistance of rice to Rice yellow mottle virus. The Plant Journal *47*, 417–426.

Ananiev, E.V., Phillips, R.L., and Rines, H.W. (1998). Complex structure of knob DNA on maize chromosome 9. Retrotransposon invasion into heterochromatin. Genetics *149*, 2025-2037.

Atwood-Moore, A., Yan, K., Judson, R.L., and Levin, H.L. (2006). The self primer of the long terminal repeat retrotransposon Tf1 is not removed during reverse transcription. Journal of Virology *80*, 8267-8270.

Balasundaram, D., Benedik, M.J., Morphew, M., Dang, V.-D., and Levin, H.L. (1999). Nup124p is a nuclear pore factor of Schizosaccharomyces pombe that is important for nuclear import and activity of retrotransposon Tf1. Molecular and Cellular Biology *19*, 5768-5784.

Beauregard, A., M., and Belfort, M. (2008). The take and give between retrotransposable elements and their hosts. Annual Review of Genetics *42*, 587-617.

Beguiristain, T., Grandbastien, M.-A., Puigdome`nech, P., and Casacuberta, J.M. (2001). Three Tnt1 subfamilies show different stress-associated patterns of expression in tobacco. Consequences for retrotransposon control and evolution in plants. Plant Physiology *127*, 212-221.

Beliakova-Bethell, N., Beckham, C., Giddings, T.H. Jr., Winey, M., Parker, R., and Sandmeyer, S. (2006). Virus-like particles of the Ty3 retrotransposon assemble in association with P-body components. RNA *12*, 94-101.

Beliakova-Bethell, N., Terry, L.J., Bilanchone, V., DaSilva, R., Nagashima, K., Wente, S.R., and Sandmeyer, S. (2009). Ty3 nuclear entry is initiated by viruslike particle docking on GLFG nucleoporins. Journal of Virology *83*, 11914-11925.

Belyayev, A., Kalendar, R., Brodsky, L., Nevo, E., Schulman, A.H., and Raskina, O. (2010). Transposable elements in a marginal plant population: temporal fluctuations provide new insights into genome evolution of wild diploid wheat. Mobile DNA *1*, 6-6.

Bennetzen, J. (2000). Comparative sequence analysis of plant nuclear genomes. Microcolinearity and its many exceptions. Plant Cell *12*, 1021-1029.

Biémont, C., and Vieira, C. (2006). Junk DNA as an evolutionary force. Nature *443*, 521-524.

Bortoluzzi, S., D'Alessi, F., Romualdi, C., and Danieli, G.A. (2000). The human adult skeletal muscle transcriptional profile reconstructed by a novel computational approach. Genome Research *10*, 344-349.

Boutabout, M., Wilhelm, M., and Wilhelm, F.X. (2001). DNA synthesis fidelity by the reverse transcriptase of the yeast retrotransposon Ty1. Nucleic Acids Research *29*, 2217-2222.

Brierley, C., and Flavell, A.J. (1990). The retrotransposon copia controls the relative levels of its gene products post-transcriptionally by differential expression from its two major mRNAs. Nucleic Acids Research *18*, 2947-2951.

Brookman, J.L., Stott, A.J., Cheeseman, P.J., Adamson, C.S., Holmes, D., Cole, J., and Burns, N.R. (1995). Analysis of TYA protein regions necessary for formation of the Ty1 virus-like particle structure. Virology *212*, 69-76.

Burns, N.R., Saibil, H.R., White, N.S., Pardon, J.F., Timmins, P.A., Richardson, S.M., Richards, B.M., Adams, S.E., Kingsman, S.M., and Kingsman, A.J. (1992). Symmetry, flexibility and permeability in the structure of yeast retrotransposon virus-like particles. The EMBO Journal *11*, 1155-1164.

Chang, W., and Schulman, A.H. (2008). BARE retrotransposons produce multiple groups of rarely polyadenylated transcripts from two differentially regulated promoters. The Plant Journal *56*, 40-50.

Checkley, M.A., Nagashima, K., Lockett, S.J., Nyswaner, K.M., and Garfinkel, D.J. (2010). P-body components are required for Ty1 retrotransposition during assembly of retrotransposition-competent virus-like particles. Molecular and Cellular Biology *30*, 382-398.

Choulet, F., Wicker, T., Rustenholz, C., Paux, E., Salse, J., Leroy, P., Schlub, S., Le Paslier, M.-C., Magdelenat, G., Gonthier, C.*, et al.* (2010). Megabase level sequencing reveals contrasted organization and evolution patterns of the wheat gene and transposable element spaces. The Plant Cell *22*, 1686-1701.

Dang, V.-D., and Levin, H.L. (2000). Nuclear import of the retrotransposon Tf1 is governed by a nuclear localization signal that possesses a unique requirement for the FXFG nuclear pore factor Nup124p. Molecular and Cellular Biology *20*, 7798-7812.

Davies, W.J., and Zhang, J. (1991). Root signals and the regulation of growth and development of plants in drying soil. Annual Review of Plant Physiology and Plant Molecular Biology *42*, 55-76.

Dorman, N., and Lever, A. (2000). Comparison of viral genomic RNA sorting mechanisms in human immunodeficiency virus type 1 (HIV-1), HIV-2, and Moloney murine leukemia virus. Journal of Virology *74*, 11413-11417.

Downs, J.A., and Jackson, S.P. (1999). Involvement of DNA end-binding protein Ku in Ty element retrotransposition. Molecular and Cellular Biology *19*, 6260-6268.

Dreher, T.W., and Miller, W.A. (2006). Translational control in positive strand RNA plant viruses. Virology *344*, 185-197.

Döring, H.P., and Starlinger, P. (1984). Barbara McClintock's controlling elements: Now at the DNA level. Cell *39*, 253-259.

Eichinger, D.J., and Boeke, J.D. (1988). The DNA intermediate in yeast Ty1 element transposition copurifies with virus-like particles: Cell-free Ty1 transposition. Cell *54*, 955-966.

Eickbush, T.H., and Jamburuthugoda, V.K. (2008). The diversity of retrotransposons and the properties of their reverse transcriptases. Virus Research *134*, 221-234.

Evgen'ev, M.B., and Arkhipova, I.R. (2005). Penelope-like elements--a new class of retroelements: distribution, function and possible evolutionary significance. Cytogenetic and Genome Research *110*, 510-521.

Finnegan, D.J. (1989). Eukaryotic transposable elements and genome evolution. Trends in Genetics *5*, 103-107.

Flavell, A.J., Dunbar, E., Anderson, R., Pearce, S.R., Hartley, R., and Kumar, A. (1992). Ty1-copia group retrotransposons are ubiquitous and heterogeneous in higher plants. Nucleic Acids Research *20*, 3639-3644.

Frankel, A.D., and Young, J.A. (1998). HIV-1: Fifteen proteins and an RNA. Annual Review of Biochemistry *67*, 1-25.

Gao, X., Havecker, E.R., Baranov, P.V., Atkins, J.F., and Voytas, D.F. (2003). Translational recoding signals between gag and pol in diverse LTR retrotransposons. RNA *9*, 1422-1430.

Gao, X., Rowley, D.J., Gai, X., and Voytas, D.F. (2002). Ty5 gag mutations increase retrotransposition and suggest a role for hydrogen bonding in the function of the nucleocapsid zinc finger. Journal of Virology *76*, 3240–3247.

Gaur, V.S., Singh, U.S., and Kumar, A. (2011). Transcriptional profiling and in silico analysis of Dof transcription factor gene family for understanding their regulation during seed development of rice Oryza sativa L. Molecular Biology Reports *38*, 2827-2848.

Gogvadze, E., and Buzdin, A. (2009). Retroelements and their impact on genome evolution and functioning. Cellular and Molecular Life Sciences *66*, 3727-3742.

Goodwin, T.J., and Poulter, R.T. (2004). A new group of tyrosine recombinase-encoding retrotransposons. Molecular Biology and Evolution *21*, 746-759.

Grandbastien, M.A. (1998). Activation of plant retrotransposons under stress conditions. Trends in Plant Science *3*, 181-187.

Gubler, F., Raventos, D., Keys, M., Watts, R., Mundy, J., and Jacobsen, J.V. (1999). Target genes and regulatory domains of the GAMYB transcriptional activator in cereal aleurone. Plant Journal *17*, 1-9.

Haag, A.L., Lin, J.-H., and Levin, H.L. (2000). Evidence for the packaging of multiple copies of Tf1 mRNA into particles and the trans priming of reverse transcription. Journal of Virology *74*, 7164-7170.

Hajek, K.L., and Friesen, P.D. (1998). Proteolytic processing and assembly of gag and gag-pol proteins of TED, a baculovirus-associated retrotransposon of the gypsy family. Journal of Virology *72*, 8718-8724.

Haoudi, A., Rachidi, M., Kim, M.H., Champion, S., Best-Belpomme, M., and Maisonhaute, C. (1997). Developmental expression analysis of the 1731 retrotransposon reveals an enhancement of Gag-Pol frameshifting in males of Drosophila melanogaster. Gene *196*, 82-93.

Hartung, W., Sauter, A., and Hose, E. (2002). Abscisic acid in the xylem: where does it come from, where does it go to? Journal of Experimental Botany *53*, 27-32.

Haseneyer, G., Ravel, C., Dardevet, M., Balfourier, F., Sourdille, P., Charmet, G., Brunel, D., Sauer, S., Geiger, H.H., Graner, A.*, et al.* (2008). High level of conservation between genes coding for the GAMYB transcription factor in barley (Hordeum vulgare L.) and bread wheat (Triticum aestivum L.) collections. Theoretical and Applied Genetics *117*, 321-331.

Havecker, E.R., Gao, X., and Voytas, D.F. (2004). The diversity of LTR retrotransposons. Genome Biology *5* (6), article 225.

Hawkins, J.S., Proulx, S.R., Rapp, R.A., and Wendel, J.F. (2009). Rapid DNA loss as a counterbalance to genome expansion through retrotransposon proliferation in plants. Proceedings of the National Academy of Sciences USA *106*, 17811-17816.

Headey, S.J., MacAskill, U.K., Wright, M.A., Claridge, J.K., Edwards, P.J., Farley, P.C., Christeller, J.T., Laing, W.A., and Pascal, S.M. (2010). Solution structure of the Squash Aspartic Acid Proteinase Inhibitor (SQAPI) and mutational analysis of pepsin inhibition. The Journal of Biological Chemistry *285*, 27019-27025.

Hirochika, H. (1993). Activation of tobacco retrotransposons during tissue culture. The EMBO Journal *12*, 2521-2528.

Irwin, P.A., and Voytas, D.F. (2001). Expression and processing of proteins encoded by the Saccharomyces retrotransposon Ty5. Journal of Virology *75*, 1790-1797.

Isabel-LaMoneda, I., Diaz, I., Martinez, M., Mena, M., and Carbonero, P. (2003). SAD: a new DOF protein from barley that activates transcription of a cathepsin B-like thiol protease gene in the aleurone of germinating seeds. Plant Journal *33*, 329-340.

Iskra-Caruana, M.-L., Baurens, F.-C., Gayral, P., and Chabannes, M. (2010). A four-partner plant–virus interaction: Enemies can also come from within. Molecular Plant-Microbe Interactions *23*, 1394-1402.

Jiang, N., and Wessler, S.R. (2001). Insertion preference of maize and rice miniature inverted repeat transposable elements as revealed by the analysis of nested elements. Plant Cell *13*, 2553-2564.

Johnson, M.C., Scobie, H.M., Ma, Y.M., and Vogt, V.M. (2002). Nucleic acid-independent retrovirus assembly can be driven by dimerization. Journal of Virology *76*, 11177-11185.

Jones, J.M., and Gellert, M. (2004). The taming of a transposon: V(D)J recombination and the immune system. Immunological Reviews *200*, 233-248.

Jordan, E., Saedler, H., and Starlinger, P. (1968). 0^0 and strong polar mutations in the *gal* operons are insertions. Molecular and General Genetics *102*, 353-360.

Kalendar, R., Grob, T., Regina, M., Suoniemi, A., and Schulman, A.H. (1999). IRAP and REMAP: Two new retrotransposon-based DNA fingerprinting techniques. Theoretical and Applied Genetics *98*, 704-711.

Kalendar, R., Tanskanen, J., Immonen, S., Nevo, E., and Schulman, A.H. (2000). Genome evolution of wild barley (*Hordeum spontaneum*) by BARE-1 retrotransposon dynamics in response to sharp microclimatic divergence. Proceedings of the National Academy of Sciences USA *97*, 6603-6607.

Kapitonov, V.V., and Jurka, J. (2001). Rolling-circle transposons in eukaryotes. Proceedings of the National Academy of Sciences USA *98*, 8714-8719.

Karst, S.M., Rütz, M.-L., and Menees, T.M. (2000). The yeast retrotransposons Ty1 and Ty3 require the RNA lariat debranching enzyme, Dbr1p, for efficient accumulation of reverse transcripts. Biochemical and Biophysical Research Communications *268*, 112-117.

Karst, S.M., Sadeghi, N., and Menees, T.M. (1999). Cell cycle control of reverse transcriptase activity for the yeast retrotransposon Ty3. Biochemical and Biophysical Research Communications *254*, 679-684.

Keller, B., and Feuillet, C. (2000). Colinearity and gene density in grass genomes. Trends in Plant Science *5*, 246-251.

Kenna, M.A., Brachmann, C.B., Devine, S.E., and Boeke, J.D. (1998). Invading the yeast nucleus: a nuclear localization signal at the C terminus of Ty1 integrase is required for transposition *in vivo*. Molecular and Cellular Biology *18*, 1115-1124.

Kim, A., Terzian, C., Santamaria, P., Pelisson, A., Prud'homme, N., and Bucheton, A. (1994). Retroviruses in invertebrates: the gypsy retrotransposon is apparently an infectious retrovirus of Drosophila melanogaster. Proceedings of the National Academy of Sciences USA *91*, 1285-1289.

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

Kirchner, J., and Sandmeyer, S. (1993). Proteolytic processing of Ty3 proteins is required for transposition. Journal of Virology *67*, 19-28.

Kneller, E.L., Rakotondrafara, A.M., and Miller, W.A. (2006). Cap-independent translation of plant viral RNAs. Virus Research *119*, 63-75.

Ko, C.H., Brendel, V., Taylor, R.D., and Walbot, V. (1998). U-richness is a defining feature of plant introns and may function as an intron recognition signal in maize. Plant Molecular Biology *36*, 573-583.

Koiwai, H., Nakaminami, K., Seo, M., Mitsuhashi, W., Toyomasu, T., and Koshiba, T. (2004). Tissue-specific localization of an abscisic acid biosynthetic enzyme, AAO3, in Arabidopsis. Plant Physiology *134*, 1697-1707.

Kumar, A., and Bennetzen, J.L. (1999). Plant retrotransposons. Annual Review of Genetics *33*, 479-532.

La, H., Ding, B., Mishra, G.P., Zhou, B., Yang, H., del Rosario Bellizzi, M., Chen, S., Meyers, B.C., Peng, Z., Zhu, J.-K.*, et al.* (2011). A 5-methylcytosine DNA glycosylase/lyase demethylates the retrotransposon Tos17 and promotes its transposition in rice. Proceedings of the National Academy of Sciences USA *108*, 15498-15503.

Larsen, L.S.Z., Beliakova-Bethell, N., Bilanchone, V., Zhang, M., Lamsa, A., DaSilva, R., Hatfield, W., Nagashima, K., and Sandmeyer, S. (2008). Ty3 nucleocapsid controls localization of particle assembly. Journal of Virology *82*, 2501-2514.

Legendre, D., and Fastrez, J. (2005). Production in Saccharomyces cerevisiae of MS2 virus-like particles packaging functional heterologous mRNAs. Journal of Biotechnology *117*, 183-194.

Liu, R., Vitte, C., Ma, J., Mahama, A.A., Dhliwayo, T., Lee, M., and Bennetzen, J.L. (2007). A GeneTrek analysis of the maize genome. Proceedings of the National Academy of Sciences USA *104*, 11844-11849.

Liu, Z.L., Han, F.P., Tan, M., Shan, X.H., Dong, Y.Z., Wang, X.Z., Fedak, G., Hao, S., and Liu, B. (2004). Activation of a rice endogenous retrotransposon Tos17 in tissue culture is accompanied by cytosine demethylation and causes heritable alteration in methylation pattern of flanking genomic regions. Theoretical and Applied Genetics *109*, 200-209.

Manninen, I., and Schulman, A.H. (1993). BARE-1, a copia-like retroelement in barley (Hordeum vulgare L.). Plant Molecular Biology *22*, 829-846.

Marchant, A., Kargul, J., May, S.T., Muller, P., Delbarre, A., Perrot-Rechenmann, C., and Bennett, M.J. (1999). AUX1 regulates root gravitropism in Arabidopsis by facilitating auxin uptake within root apical tissues. The EMBO Journal *18*, 2066-2073.

Martin-Rendon, E., Marfany, G., Wilson, S., Ferguson, D.J.P., Kingsman, S.M., and Kingsman, A.J. (1996). Structural determinants within the subunit protein of Ty1 virus-like particles. Molecular Microbiology *22*, 667-679.

Mathews, C.K., and Holde, K.E. (1991). Structural gene transcription and mRNA processing. In Biochemistry (Redwood City: The Benjamin/Cummings Publishing Company), pp. 1019-1020.

Matsuda, E., and Garfinkel, D.J. (2009). Posttranslational interference of Ty1 retrotransposition by antisense RNAs. Proceedings of the National Academy of Sciences USA *106*, 15657-15662.

McClintock, B. (1953). Induction of instability at selected loci in maize. Genetics *38*, 579-599.

McClintock, B. (1956). Controlling elements and the gene. Cold Spring Harbor Symposium in Quantitative Biology *21*, 197-216.

McLane, L.M., Pulliam, K.F., Devine, S.E., and Corbett, A.H. (2008). The Ty1 integrase protein can exploit the classical nuclear protein import machinery for entry into the nucleus. Nucleic Acids Research *36*, 4317-4326.

Menees, T.M., and Sandmeyer, S.B. (1996). Cellular stress inhibits transposition of the yeast retrovirus-like element Ty3 by a ubiquitin-dependent block of virus-like particle formation. Proceedings of the National Academy of Sciences USA *93*, 5629-5634.

Merkulov, G.V., Swiderek, K.M., Brachmann, C.B., and Boeke, J.D. (1996). A critical proteolytic cleavage site near the C terminus of the yeast retrotransposon Ty1 Gag protein. Journal of Virology *70*, 5548-5556.

Messer, L.I., Levin, J.G., and Chattopadhyay, S.K. (1981). Metabolism of viral RNA in murine leukemia virus-infected cells; evidence for differential stability of viral message and virion precursor RNA. Journal of Virology *40*, 683-690.

Mirouze, M., Reinders, J., Bucher, E., Nishimura, T., Schneeberger, K., Ossowski, S., Cao, J., Weigel, D., Paszkowski, J., and Mathieu, O. (2009). Selective epigenetic control of retrotransposition in Arabidopsis. Nature *461*, 427-431.

Moore, G., Devos, K.M., Wang, Z., and Gale, M.D. (1995). Cereal genome evolution. Grasses, line up and form a circle. Current Biology *5*, 737-739.

Moreau-Mhiri, C., Morel, J.B., Audeon, C., Ferault, M., Grandbastien, M.A., and Lucas, H. (1996). Regulation of expression of the tobacco Tnt1 retrotransposon in heterologous species following pathogen-related stresses. The Plant Journal *9*, 409-419.

Muriaux, D., Mirro, J., Harvin, D., and Rein, A. (2001). RNA is a structural element in retrovirus particles. Proceedings of the National Academy of Sciences USA *98*, 5246-5251.

Ohno, S. (1972). So much 'junk' in our genome. Brookhaven Symposia in Biology *23*, 366-370.

Paillart, J.-C., Shehu-Xhilaga, M., Marquet, R., and Mak, J. (2004). Dimerization of retroviral genomes: an inseparable pair. Nature Reviews Microbiology *2*, 461-472.

Palmer, K.J., Tichelaar, W., Myers, N., Burns, N.R., Butcher, S.J., Kingsman, A.J., Fuller, S.D., and Saibil, H.R. (1997). Cryo-electron microscopy structure of yeast Ty retrotransposon virus-like particles. Journal of Virology *71*, 6863-6868.

Panstruga, R., Büschges, R., Piffanelli, P., and Schulze-Lefert, P. (1998). A contiguous 60 kb genomic stretch from barley reveals molecular evidence for gene islands in a monocot genome. Nucleic Acids Research *26*, 1056-1062.

Peterson-Burch, B.D., and Voytas, D.F. (2002). Genes of the Pseudoviridae (Ty1/copia Retrotransposons). Molecular Biology and Evolution *19*, 1832–1845.

Pickett, G.G., and Peabody, D.S. (1993). Encapsidation of heterologous RNAs by bacteriophage MS2 coat protein. Nucleic Acids Research *21*, 4621-4626.

Potokina, E., Sreenivasulu, N., Altschmied, L., Michalek, W., and Graner, A. (2002). Differential gene expression during seed germination in barley (Hordeum vulgare L.). Functional & Integrative Genomics *2*, 28-39.

Poulter, R.T., and Goodwin, T.J. (2005). DIRS-1 and the other tyrosine recombinase retrotransposons. Cytogenetic and Genome Research *110*, 575-588.

Pouteau, S., Huttner, E., Grandbastien, M.A., and Caboche, M. (1991). Specific expression of the tobacco Tnt1 retrotransposon in protoplasts. The EMBO Journal *10*, 1911-1918.

Pritham, E.J., Putliwala, T., and Feschotte, C. (2007). Mavericks, a novel class of giant transposable elements widespread in eukaryotes and related to DNA viruses. Gene *390*, 3-17.

Reddy, A.S.N. (2007). Alternative splicing of pre-messenger RNAs in plants in the genomic era. Annual Review of Plant Biology *58*, 267-294.

Ribet, D., Harper, F., Dupressoir, A., Dewannieux, M., Pierron, G., and Heidmann, T. (2008). An infectious progenitor for the murine IAP retrotransposon: Emergence of an intracellular genetic parasite from an ancient retrovirus. Genome Research *18*, 597-609.

Rostocks, N., Park, Y.J., Ramakrishna, W., Ma, J., Druka, A., Shiloff, B.A., SanMiguel, P.J., Jiang, Z., Brueggeman, R., Sandhu, D.*, et al.* (2002). Genomic sequencing reveals gene content, genomic organization, and recombination relationships in barley. Functional and Integative Genomics *2*, 51-59.

Roth, J.F. (2000). The yeast Ty virus-like particles. Yeast *16*, 785-795.

Rowold, D.J., and Herrera, R.J. (2000). Alu elements and the human genome. Genetica *108*, 57-72.

Royo, J., Nass, N., Matton, D.P., Okamoto, S., Clarke, A.E., and Newbigin, E. (1996). A retrotransposon-like sequence linked to the S-locus of Nicotiana alata is expressed in styles in response to touch. Molecular and General Genetics *250*, 180-188.

Rulli, S.J. Jr., Hibbert, C.S., Mirro, J., Pederson, T., Biswal, S., and Rein, A. (2007). Selective and nonselective packaging of cellular RNAs in retrovirus particles. Journal of Virology *81*, 6623-6631.

Sabot, F., and Schulman, A.H. (2006). Parasitism and the retrotransposon life cycle in plants: A hitchhiker's guide to the genome. Heredity *97*, 381-388.

Sandmeyer, S.B., and Clemens, K.A. (2010). Function of a retrotransposon nucleocapsid protein. RNA Biology *7*, 642-654.

SanMiguel, P., Tikhonov, A., Jin, Y.K., Motchoulskaia, N., Zakharov, D., Melake-Berhan, A., Springer, P.S., Edwards, K.J., Lee, M., Avramova, Z*., et al.* (1996). Nested retrotransposons in the intergenic regions of the maize genome. Science *274*, 765-768.

Saze, H., and Kakutani, T. (2011). Differentiation of epigenetic modifications between transposons and genes. Current Opinion in Plant Biology *14*, 81-87.

Schnable, P.S., Ware, D., Fulton, R.S., Stein, J.C., Wei, F., Pasternak, S., Liang, C., Zhang, J., Fulton, L., Graves, T.A*., et al.* (2009). The B73 maize genome: complexity, diversity, and dynamics. Science *326*, 1112-1115.

Schuler, M.A. (2008). Splice site requirements and switches in plants. Current Topics in Microbiology and Immunology *326*, 39-59.

Schulman, A.H. (2012). Hitching a ride: Non-autonomous retrotransposons and parasitism as a lifestyle. In: Plant Transposable Elements - Impact on Genome Structure & Function, Grandbastien M-A & Casacuberta JM, eds. Series: Topics in Current Genetics. Springer Verlag.(In press)

Schulman, A.H., and Wicker, T. (in press). A Field Guide to Transposable Elements: their diversity, origins, propagation mechanisms, and evolutionary significance. In Molecular Genetics and Epigenetics of Plant Transposons, N. Fedoroff, ed. (Hoboken, NJ USA: John Wiley & Sons).

Shi, X., Seluanov, A., and Gorbunova, V. (2007). Cell divisions are required for L1 retrotransposition. Molecular and Cellular Biology *27*, 1264-1270.

Shure, M., Wessler, S., and Fedoroff, N. (1983). Molecular identification and isolation of the Waxy locus in maize. Cell *35*, 225-233.

Slotkin, R.K. (2010). The epigenetic control of the Athila family of retrotransposons in Arabidopsis. Epigenetics *5*, 483-490.

Slotkin, R.K., and Martienssen, R. (2007). Transposable elements and the epigenetic regulation of the genome. Nature Reviews Genetics *8*, 272-285.

Slotkin, R.K., Vaughn, M., Borges, F., Tanurdzić, M., Becker, J.D., Feijó, J.A., and Martienssen, R.A. (2009). Epigenetic reprogramming and small RNA silencing of transposable elements in pollen. Cell *136*, 461-472.

Steinbauerová, V., Neumann, P., and Macas, J. (2008). Experimental evidence for splicing of intron-containing transcripts of plant LTR retrotransposon Ogre. Molecular Genetics and Genomics *280*, 427-436.

Suoniemi, A., Anamthawat-Jónsson, K., Arna, T., and Schulman, A.H. (1996a). Retrotransposon BARE-1 is a major, dispersed component of the barley (Hordeum vulgare L.) genome. Plant Molecular Biology *30*, 1321-1329.

Suoniemi, A., Narvanto, A., and Schulman, A.H. (1996b). The BARE-1 retrotransposon is transcribed in barley from an LTR promoter active in transient assays. Plant Molecular Biology *31*, 295-306.

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

Suoniemi, A., Tanskanen, J., Pentikäinen, O., Johnson, M.S., and Schulman, A.H. (1998a). The core domain of retrotransposon integrase in Hordeum: Predicted structure and evolution. Molecular Biology and Evolution *15*, 1135-1144.

Suoniemi, A., Tanskanen, J., and Schulman, A.H. (1998b). Gypsy-like retrotransposons are widespread in the plant kingdom. The Plant Journal *13*, 699-705.

Syomin, B.V., Kandror, K.V., Semakin, A.B., Tsuprun, V.L., and Stepanov, A.S. (1993). Presence of the gypsy (MDG4) retrotransposon in extracellular virus-like particles. FEBS Letters *323*, 285-288.

Tanskanen, J.A., Sabot, F., Vicient, C., and Schulman, A.H. (2007). Life without GAG: The BARE-2 retrotransposon as a parasite´s parasite. Gene *390*, 166-174.

Telesnitsky, A., and Goff, S.P. (1997). Reverse transcriptase and the generation of retroviral DNA. In Retroviruses, J.M. Coffin, S.H. Hughes, and H.E. Varmus, eds. (Cold Spring Harbor Laboratory Press), pp. 121-160.

Thomas, C.A. (1971). The genetic organization of chromosomes. Annual Review of Genetics *5*, 237-256.

Todorovska, E. (2007). Retrotransposons and their role in plant - genome evolution. Biotechnology & Biotechnological Equipment *21*, 294-305.

Van Der Kelen, K., Beyaert, R., Inzé, D., and De Veylder, L. (2009). Translational control of eukaryotic gene expression. Critical Reviews in Biochemistry and Molecular Biology *44*, 143-168.

Vicient, C.M., Kalendar, R., Anamthawat-Jonsson, K., Suoniemi, A., and Schulman, A.H. (1999a). Structure, functionality, and evolution of the BARE-1 retrotransposon of barley. Genetica *107*, 53-63.

Vicient, C.M., Kalendar, R., and Schulman, A.H. (2001). Envelope-class retrovirus-like elements are widespread, transcribed and spliced, and insertionally polymorphic in plants. Genome Research *11*, 2041-2049.

Vicient, C.M., Kalendar, R., and Schulman, A.H. (2005). Variability, recombination, and mosaic evolution of the barley BARE-1 retrotransposon. Journal of Molecular Evolution *61*, 275-291.

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

Vitte, C., and Bennetzen, J.L. (2006). Analysis of retrotransposon structural diversity uncovers properties and propensities in angiosperm genome evolution. Proceedings of the National Academy of Sciences USA *103*, 17638-17643.

Voytas, D.F., and Boeke, J.D. (1993). Yeast retrotransposons and tRNAs. Trends in Genetics *9*, 421-426.

Voytas, D.F., Cummings, M.P., Konieczny, A.K., Ausubel, F.M., and Rodermel, S.R. (1992). Copia-like retrotransposons are ubiquitous among plants. Proceedings of the National Academy of Sciences USA *89*, 7124-7128.

Waugh, R., McLean, K., Flavell, A.J., Pearce, S.R., Kumar, A., Thomas, B.B.T., and Powell, W. (1997). Genetic distribution of BARE-1-like retrotransposable elements in the barley genome revealed by sequence-specific amplification polymorphisms (S-SAP). Molecular and General Genetics *253*, 687-694.

Wicker, T., and Keller, B. (2007). Genome-wide comparative analysis of copia retrotransposons in Triticeae, rice, and Arabidopsis reveals conserved ancient evolutionary lineages and distinct dynamics of individual copia families. Genome Research *17*, 1072-1081.

Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J.L., Capy, P., Chalhoub, B., Flavell, A.J., Leroy, P., Morgante, M., Panaud, O.*, et al.* (2007). A unified classification system for eukaryotic transposable elements. Nature Reviews Genetics *8*, 973-982.

Wicker, T., Taudien, S., Houben, A., Keller, B., Graner, A., Platzer, M., and Stein, N. (2009). A whole-genome snapshot of 454 sequences exposes the composition of the barley genome and provides evidence for parallel evolution of genome size in wheat and barley. The Plant Journal *59*, 712-722.

Wicker, T., Zimmermann, W., Perovic, D., Paterson, A.H., Ganal, M., Graner, A., and Stein, N. (2005). A detailed look at 7 million years of genome evolution in a 439 kb contiguous sequence at the barley Hv-eIF4E locus: recombination, rearrangements and repeats. The Plant Journal *41*, 184-194.

Wilhelm, M., and Wilhelm, F.X. (2006). Cooperation between reverse transcriptase and integrase during reverse transcription and formation of the preintegrative complex of Ty1. Eukaryotic Cell *5*, 1760-1769.

Yanagisawa, S., and Schmidt, R.J. (1999). Diversity and similarity among recognition sequences of Dof transcription factors. Plant Journal *17*, 209-214.