J Mol Evol (2005) 61:275–291
DOI: 10.1007/s00239-004-0168-7

JOURNAL OF MOLECULAR EVOLUTION

© Springer Science+Business Media, Inc. 2005

Variability, Recombination, and Mosaic Evolution of the Barley *BARE*-1 Retrotransposon

Carlos M. Vicient,[1]* Ruslan Kalendar,[1] Alan H. Schulman[1,2]

[1] MTT/BI Plant Genomics Laboratory, Institute of Biotechnology, University of Helsinki, FIN-00014, Helsinki, Finland
[2] Plant Breeding Biotechnology, MTT Agrifood Research Finland, Jokioinen, Finland

Received: 3 June 2004 / Accepted: 11 March 2005 [*Reviewing Editor:* Dr. Juergen Brosius]

Abstract. *BARE*-1 is a highly abundant, *copia*-like, LTR (long terminal repeat) retrotransposon in the genus *Hordeum*. The LTRs provide the promoter, terminator, and polyadenylation signals necessary for the replicational life cycle of retrotransposons. We have examined the variability and evolution of *BARE*-1-like elements, focusing on the LTRs. Three groups were found, corresponding to each of the *Hordeum* genome types analyzed, which predate the divergence of these types. The most variable LTR regions are tandem repeats near the 3′ end and the promoter. In barley (*H. vulgare* L.), two main classes of LTR promoters were defined, corresponding to *BARE*-1 and to a new class we call *BARE*-2. These can be considered as families within the group I *BARE* elements. Although less abundant in cultivated barley than is *BARE*-1, *BARE*-2 is transcriptionally active in leaves and calli. A sequenced *BARE*-2 has more than 99% similar LTRs and perfect terminal direct repeats (TDRs), indicating it is a recent insertion, but the coding region, especially *gag*, is disrupted by frameshifts and stop codons. *BARE*-2 appears to be a chimeric element resulting from retrotransposon recombination by strand switching during replication, with LTRs and 5′UTR more similar to *BARE*-1 and the rest more similar to *Wis*-2. We provide evidence as well for another form of recombination, where LTR-LTR recombination has generated tandem multimeric *BARE*-1 elements in which internal coding domains are interspersed with shared LTRs. The data indicate that recombination contributes to the complexity and plasticity of retroelement evolution in plant genomes.

Key words: *BARE*-1 retrotransposon — Barley — *Hordeum vulgare* L. — Molecular evolution — DNA recombination — LTR promoter

Introduction

Retrotransposons are mobile genetic elements, which transpose via the reverse transcription of a transcribed, intermediate RNA (McDonald 1993; Feschotte et al. 2002). Retrotransposons are abundant and widespread components of eukaryotic genomes. They are usually present in plant genomes as populations of elements in high copy number, together accounting for more than 50% of the genome (Kumar and Bennetzen 1999; Vicient et al. 1999). The sequencing of 417.5 kb of the barley (*Hordeum vulgare* L.) genome revealed at least 40% to be composed of retrotransposons (Rostoks et al. 2002); in another 103-kb region, 75% was retrotransposons (Park et al. 2004). On a local scale, retrotransposons in the grasses are frequently present as extensive nests of elements inserted into each other that surround islands of genes (SanMiguel et al. 1996; Shirasu et al. 2000), although elements specifically or preferentially

**Present address:* Laboratori CSIC-IRTA de Genetica Molecular Vegetal, Jordi Girona 18-26, 08034 Barcelona, Spain
Correspondence to: Alan H. Schulman; *email:* alan.schulman@helsinki.fi

found in centromeres (Kumekawa et al. 2000; Nomomura and Kurata 2001; Zhong et al. 2002) or telomeres (Casacuberta and Pardue 2003) or as genie insertions (Yamakazi et al. 2001) have been described in the grasses and elsewhere.

Retrotransposons are subdivided into LTR and non-LTR retrotransposons, with the former being bounded by long terminal repeats (LTRs) oriented in the same direction (Boeke and Corces 1989). The life cycle of the LTR retrotransposons resembles that of retroviruses; it comprises transcription, reverse transcription, packaging into virus-like particles, and integration of the cDNA copy back into the genome. The polyprotein encoded by the retrotransposon itself provides the reverse transcriptase, Gag structural protein, and integrase for the latter steps (Frankel and Young 1998; Kumar and Bennetzen 1999), whereas cellular RNA polymerase II is responsible for transcription. The LTRs both contain the promoter necessary for transcription and specify the terminator and polyadenylation signals needed for RNA processing. The ends of the LTRs are recognized by the integrase. In addition, LTRs also contain the R (for "repeat") region, lying between the transcription start and termination. Because the promoter functions in the 5′ LTR and the terminator in the 3′ LTR, the R region is found at both ends of the transcript. It enables the nascent (-)-strand cDNA to jump from a 5′ LTR to a 3′ LTR, a necessary step during reverse transcription.

Despite the functional importance of the LTR sequences, several reports indicate that the LTRs are one of the most rapidly evolving retrotransposon regions (Lankenau et al. 1990; Lyubomirskaya et al. 1990; Mizrokhi and Mazo 1990; Danilevskaya et al. 1997; Kalmykova et al. 2004). However, relatively few studies have focused on sequence variability in the LTR region of plant retrotransposons. A detailed analysis of regulatory regions (Casacuberta and Grandbastien 1993) and variability (Casacuberta et al. 1995) has been carried out for the tobacco (*Nicotiana tabacum* L.) retrotransposon *Tnt1* and for the maize retrotransposon *Grande1* (Garcia-Martinez and Martinez-Izquierdo 2003). The *Tnt1* element is present in only a few hundred copies in the tobacco genome (Grandbastien et al. 1989; Hirochika 1993). This retrotransposon is generally silent within the plant, although it is strongly stress-inducible (Beguiristain et al. 2001).

In contrast to *Tnt1*, retrotransposon *BARE*-1 is present in more than 1.5×10^4 copies in barley and is similarly abundant in other *Hordeum* species (Vicient et al. 1999). Elements closely related to *BARE*-1 have been found in other grasses (Matsuoka and Tsunewaki 1996; Gribbon et al. 1999; Vicient et al. 2001). The *BARE*-1 element is transcriptionally active in the plant (Suoniemi et al. 1996), and processed *BARE*-1 translation products can be detected in barley tissues and in other cereals as well (Vicient et al. 2001). The *BARE*-1 LTRs are especially long, about 1.9 kb, and contain conserved regions (Suoniemi et al. 1997; Vicient et al. 1999). Sequence examination revealed that *BARE*-1 LTRs contain two canonical TATA boxes (Manninen et al. 1993), both of them being able to direct RNA transcription but under different conditions (Suoniemi et al. 1996). The *BARE*-1 LTR can drive expression of reporter genes in transiently transformed barley protoplasts in a manner dependent on the presence of a TATA box functional *in planta* as well. Deletion analysis of the promoter allowed identification of regions important for expression in protoplasts (Suoniemi et al. 1996).

Due to the great number of *BARE*-1 elements, intrachromosomal ectopic recombination between LTRs of the same or different elements can occur, leaving behind solo LTRs (Vicient et al. 1999; Shirasu et al. 2000; Kalendar et al. 2004). Solo LTRs comprise 85% of all retroelements in the yeast genome (Kim et al. 1998). In *Hordeum*, *BARE*-1 solo LTRs are 7- to 42-fold more abundant than full-length elements (Vicient et al. 1999) and LARD solo LTRs 9-fold more abundant than the full-length elements (Kalendar et al. 2004); they are also abundant for many retrotransposons in rice (Vicient and Schulman 2002). If recombination takes place between the LTRs of different individual elements, then chromosome rearrangements could occur as has been proposed for yeast (Kim et al. 1998). In other cases, recombination between related but not identical elements could generate new variants (McClure 1991; Lerat et al. 1999; Kalmykova et al. 2004; Mugnier et al. 2005). This phenomenon has also been reported for various viruses (Lai 1995).

The genus *Hordeum*, with some 50 species, is the second largest genus in the tribe Triticae of the family Poaoeae. It includes barley and is widely distributed in both hemispheres (von Bothmer et al. 1995). The species of the genus *Hordeum* can be divided into four genomic groups, designated H, I, X, and Y, based upon analyses of chromosomal pairing during meiosis in interspecific hybrids (Jacobsen and von Bothmer 1992).

In this study, we investigated the heterogeneity found in LTR sequences of retrotransposons similar to *BARE*-1 both within and between species of *Hordeum*. Our results demonstrate the existence of three subfamilies of *BARE* LTR sequences, each one characteristic of one of the three meiotic recombination groups of the *Hordeum* genomes analyzed. The data indicate that recombination between *BARE* retroelements generates several distinct classes of products, and may be important in *BARE*-1 evolution. We also provide evidence that *BARE*-1 elements can have an influence on the evolution of the host

genome, not only by increasing genome size, but also by serving as substrates for rearrangements.

Materials and Methods

Plant Material

Accessions and provenances of all *Hordeum* accessions are as previously described (Kankaanpää et al. 1996). Seeds were germinated, and the seedlings grown for 10 days before DNA was extracted from the leaves.

DNA Extraction, PCR Amplification, Cloning, and Sequencing

DNA was extracted as previously described (Vicient et al. 1999). For amplification of genomic LTRs, primers at the ends of the LTRs were used (N referring to equal amounts of A, T, G, and C in the primer preparation at that position); "LTR-full-forward", 5′-NNTGTTGGAATTATGCCCTAGAGGCAA-3′; "LTR-full-reverse", 5′-NNTGTGGGGAACGTCGCATGGGAAAC-3′. The PCR reactions were performed using 10 ng genomic DNA, 0.2 mM each dNTP, and 1 pmol μl^{-1} each primer in a final volume of 50 μl. The mix was overlaid with paraffin oil. The reaction mixtures were heated to 95°C for 5 min, followed by 21 cycles of 94°C for 30 s, 40°C for 2 min, and 72°C for 2.5 min. Reactions were completed by incubation at 72°C for 10 min. The PCR products were purified from agarose gels (QIAEX II; Qiagen, Hilden, Germany) and cloned (pGEM-T vector system; Promega). Reactions were performed in a Minicycler (MJ Research, Waltham, MA) thermal cycler. For the amplification of the tandem arrays at C-terminal of the LTRs, the primers used were: "tan-forward", 5′-GCTGTACGTGTGCTGAACGCGGAGGTG-3′, and "tan-reverse", 5′-AACGCGGTTGATGTAGT(G/C)GAACGTC-3′. The PCR conditions were as above, but with a 1-min extension at 72°C. For the amplification of tandemly arrayed *BARE*-1 copies, primers 1 and 2 were the "full-forward" and "full-reverse," described above, primer 3 was 5′-CGGATCTGAATGTAGCAACCCGC TG-3′, and primer 4 was 5′-CTACGCATGAACCTAGCTCATG ATGCC-3′. PCR conditions were as described previously, using a 2-min extension at 72°C and 40 cycles. The primers used for specific *BARE*-1 and *BARE*-2 amplifications were, respectively, "LTR-full-forward" and "reverse," 5′-CTGGTTGGCCCACG(T/C)GAG CCATT(G/A)ATCTACAACA(C/T)A-3′ and 5′-CTG GTTGGC CCACAGTAGAGCTATAG(T/C)GCAAGCTAC-3′. PCR conditions were as above, but using a 3-min extension at 72°C and 30 cycles.

Quantitative PCR for LTRs was carried out using conserved primers amplifying both *BARE*-1 and *BARE*-2 LTRs, forward 5′-TGTTGGAAATATGCCCTAGAGGCA-3′ (primer R20045, nt 1–24 at the 5′ end of the LTR) and reverse 5′-GACG GCACCTCCGCGTTCAGCACA-3′ (primer R20046, nt 1568–1591 of the LTR). The diagnostic internal domain (LTR) structure of tandem elements was quantified using a forward PPT (polypurine tract) primer 5′-CGGATCTGAATGTAGCAA CCCGCTG-3′ (primer 82574) and a reverse PBS (primer binding site) primer 5′-CTACGCATGAACCTAGCTCATGATGCC-3′ (primer 8378). The 20-μl reactions contained: 20 ng barley DNA (cv Sultan), 0.2 mM dNTPs, 0.2 μM each primer, 1U BioTools *Thermophilus thermus* polymerase (Catalog number 10.001, Biotools, B & M Labs S.A., Madrid), 1 × BioTools buffer. The reaction mixture was heated to 94°C for 2 min, then subjected to cycles of 94°C for 30 s, 55°C for 30 s, and 72°C for 2 min in an Eppendorf Master Cycler thermocycler using tubes of 0.2 ml. Reactions for the two sets of primers were compared cycle-by-cycle for up to 21 cycles; on agarose gels, the LTRs were detectable by ethidium bromide staining after 4 cycles and the tandem structure after 9 cycles. Product was quantified by scanning the amplified bands in gels stained with ethidium bromide on a Fuji imaging system (FLA-5000) using a resolution of 50 μm. Calculations were made for the logarithmic portion of the amplification reaction, assuming a doubling of the product with each cycle.

The RNA for RT-PCR was isolated with the RNAqueous kit (Ambion 9690) and then treated with RNAse-free DNAse I (Roche). The RT-PCR reactions were performed using the One-Step RT-PCR kit (Qiagen) according to the manufacturer's instructions, using 1 μg of total RNA and an amplification program comprising of 30 cycles of 45 s at 94°C, 45 s at 50°C, and 1 min at 72°C. Controls for DNA contamination consisted of reactions lacking dNTPs in the reverse transcription step but added instead at the beginning of the PCR step. The primers used for the promoter amplification were RT-forward, 5′-CCCGCTATTGG ATATTGACCGAGGAGTCCCTCGG-3′, and RT-reverse, 5′-CTGGTTCGGCCCAGGACG(G/A)CACCTCCGCGTrCAGCA CACG-3′.

Plasmid minipreps served as the templates for sequencing reactions. The reactions were catalyzed with Sequenase v2.0 (Amersham Pharmacia Biotech, Uppsala, Sweden) and resolved under standard conditions on an automated sequencing system (ALF; Amersham Pharmacia Biotech). The full-length LTR sequences and their corresponding database accession numbers are listed in Table 1. The accession numbers for the sequenced promoter regions, derived from PCR amplifications, are AJ582402-AJ582411 for genomic sequences, AJ582527–AJ582543 for sequences derived from leaf RNA, and AJ582544–AJ582561 for sequences derived from callus RNA. These accessions numbers will be available in the EMBL and GenBank databases.

Sequence Analysis

Sequence alignments were performed using ClustalW software (Higgins et al. 1994) with manual editing. Display and shading of the alignment was performed with GenDoc vers. 2.6.02 (Nicholas and Nicholas 1997). Phylogeny construction was done using the Treecon program (van de Peer and de Wachter 1997). Distances were calculated according to the Kimura 2-parameter model (Kimura 1980), and the trees were statistically evaluated using 1000 bootstrap samples. The analyses of sequence divergences were done using DnaSP program (Rozas and Rozas 1999).

Results

BARE-*1 LTR Sequences in Different* Hordeum *Species*

We have investigated sequence variability in the LTRs of the *BARE*-1 retrotransposon in the genomes of five different *Hordeum* species chosen as representatives of three meiotic recombinational types within the genus: *H. vulgare*, or cultivated barley (diploid, I genome); *H. euclaston, H. roshevitzii*, and *H. pusillum* (diploids. H genome); and *H. marinum* (diploid. X genome). These three representatives of the H genome type were chosen because *H. euclaston* and *H. pusillum* represent the smallest genomes (Kankaanpää et al. 1996; Jakob et al. 2004), the lowest number of full-length *BARE*-1 elements, and

Table 1. Origin of the full-length LTR sequences used in this study

Name	Size	Species	Genome	Origin	Type	GenBank	Reference
hvBARE-1-5	1829	*H. vulgare*	I	Lambda clone	5′LTR	Z17327	Manninen et al. (1993)
hvBARE-1-3	1851	*H. vulgare*	I	Lambda clone	3′LTR	Z17327	Manninen et al.(1993)
hvBARE-2-5	1811	*H. vulgare*	I	BAC clone	5′LTR	AJ279072	Vicient et al. (2001)
hvBARE-2-3	1810	*H. vulgare*	I	BAC clone	3′LTR	AJ279072	Vicient et al. (2001)
hvbaca-1	1830	*H. vulgare*	I	BAC clone	solo LTR	AF254799	Shirasu et al. (2000)
hvbaca-2	1817	*H. vulgare*	I	BAC clone	solo LTR	AF254799	Shirasu et al. (2000)
hvbaca-3	1822	*H. vulgare*	I	BAC clone	3′LTR	AF254799	Shirasu et al. (2000)
hvbaca-4	1825	*H. vulgare*	I	BAC clone	5-3′LTR	AF254799	Shirasu et al. (2000)
hvbaca-5	1824	*H. vulgare*	I	BAC clone	5′LTR	AF254799	Shirasu et al. (2000)
hvbacb1-5	1821	*H. vulgare*	I	BAC clone	5′LTR	AY013246	Dubcovsky et al. (2001)
hvbacb1-3	1822	*H. vulgare*	I	BAC clone	3′LTR	AY013246	Dubcovsky et al. (2001)
hvbacb2-5	1812	*H. vulgare*	I	BAC clone	5′LTR	AY013246	Dubcovsky et al. (2001)
hvbacb2-3	1819	*H. vulgare*	I	BAC clone	3′LTR	AY013246	Dubcovsky et al. (2001)
hvula	1803	*H. vulgare*	I	PCR	?	Y18767	Vicient et al. (1999)
hvulb	1732	*H. vulgare*	I	PCR	?	Y18768	Vicient et al. (1999)
hvulc	1826	*H. vulgare*	I	PCR	?	Y18769	Vicient et al. (1999)
hvuld	1719	*H. vulgare*	I	PCR	?	Y18770	Vicient et al. (1999)
hvule	1808	*H. vulgare*	I	PCR	?	Y18771	Vicient et al. (1999)
heuca	1669	*H. euclaston*	H	PCR	?	Y18780	Vicient et al. (1999)
heucb	1778	*H. euclaston*	H	PCR	?	Y18781	Vicient et al. (1999)
heuce	1131	*H. euclaston*	H	PCR	?	Y18782	Vicient et al. (1999)
heucd	1926	*H. euclaston*	H	PCR	?	AJ582609	—
heuce	1639	*H. euclaston*	H	PCR	?	AJ582610	—
hpusa	1775	*H. pusillum*	H	PCR	?	Y18783	Vicient et al. (1999)
hpusb	1788	*H. pusillum*	H	PCR	?	Y18784	Vicient et al. (1999)
hpusc	1898	*H. pusillum*	H	PCR	?	Y18785	Vicient et al. (1999)
hpusd	1564	*H pusillum*	H	PCR	?	AJ582611	—
hpuse	1887	*H. pusillum*	H	PCR	9	AJ582612	—
hrosa	1736	*H. roshevitzii*	H	PCR	?	Y18772	Vicient et al. (1999)
hrosb	1832	*H. roshevitzii*	H	PCR	?	Y18773	Vicient et al. (1999)
hrosc	1858	*H. roshevitzii*	H	PCR	?	Y18774	Vicient et al. (1999)
hrosd	1854	*H. roshevitzii*	H	PCR	?	Y18775	Vicient et al. (1999)
hmara	1877	*H. marinum*	X	PCR	?	Y18776	Vicient et al. (1999)
hmarb	1972	*H. marinum*	X	PCR	?	Y18777	Vicient et al. (1999)
hmarc	1884	*H. marinum*	X	PCR	?	Y18778	Vicient et al. (1999)
hmard	1980	*H. marinum*	X	PCR	?	Y18779	Vicient et al. (1999)
hmare	1576	*H. marinum*	X	PCR	?	AJ582608	—

the highest proportion of *BARE*-1 solo LTRs of the *Hordeum* species previously investigated (Vicient et al. 1999). The species *H. roshevitzii* has a relatively high *BARE*-1 copy number and the second highest number of solo LTRs for any *Hordeum* species examined. The X genome, which comprises only *H. marinum,* has the highest number of solo LTRs and highest ratio of LTRs to full-length elements of any *Hordeum* spp. Cultivated barley was included because it is the best-explored experimental system for *BARE*-1. Representatives of the H and Y genome groups were excluded because their *BARE*-1 population sizes, solo LTR levels, and haploid genome sizes are not distinctive.

Genomic DNAs from accessions of the five species were used for PCR amplification with two primers located at the extremes of the LTR, using low stringency conditions for primer annealing. Following PCR amplification, DNA fragments only of the expected size were detectable and corresponded to the LTRs of the same size as the original *BARE*-1a element (1.9 kb). A total of 24 fragments were cloned and sequenced, as listed in Table 1. Two of the sequences contained major deletions, *heucc* ($\approx$ 734 bp deletion) and *hmare* ($\approx$ 296 bp deletion), and the sizes of the others ranged from 1719 bp (*hvuld*) to 1971 bp (*hmarb*), with an average size of 1789 bp.

The sequences were aligned, together with the 13 full-length *BARE*-1 LTRs from *H. vulgare* that were present in the database. Neighbor-joining and parsimony methods were used to generate phylogenetic trees from the alignment and yielded similar results. The 37 sequences cluster into three major families (Fig. 1), correlating very well with the three meiotic types of *Hordeum* genomes investigated (I, H, and X). There was only one exception, *hrosd*, which will be discussed later.

Species Distribution of the Different LTR Groups

Taking the *BARE*-1 groups as defined by *Hordeum* genome type into account, we then focused on the

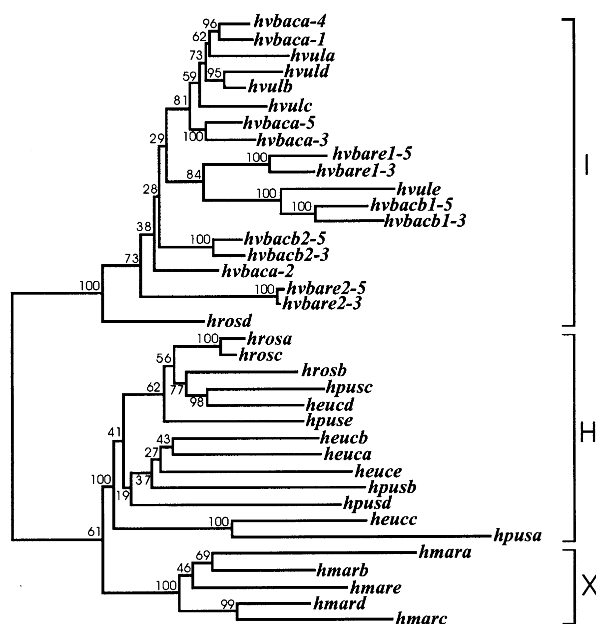

Fig. 1. Phylogenetic tree of *BARE*-1 LTR sequences. The tree was constructed by the neighbor-joining method. Sequence distances were calculated according to the Kimura two-parameter method, not taking in account insertions or deletions. The three groups of LTR sequences are indicated (I, H, X). The sequence names are as in Table 1. Bootstrap values, based on 250 replicates, of greater than 50% are shown. Horizontal lines are proportional in length to the sequence divergence; scaling of the vertical lines is only for clarity.

wealth of partial LTR sequences available for cultivated barley, *H. vulgare*. We compared these, presented in Table 2, with the consensus sequences of each of the three groups of LTRs that correspond to genome types, in order to determine to which of the groups they belong. Of the 59 barley sequences, 48 are more similar to I, which corresponds to the genome type of barley itself, 5 more similar to H (8%), 2 more similar to X (3%), and 4 more similar to H and X than to I (7%). Of the 18 wheat sequences, only one was more similar to I (6%), 7 more similar to H (39%), 2 more similar to X (11%), and 8 more similar to H and X than to H (44%). These data show that LTRs typical of the non-barley groups are also present in cultivated barley, indicating that these groups predate the divergence of the *Hordeum* genome types.

Although no LTR sequences from other *Hordeum* species are present in the database, 18 sequences from bread wheat (*Triticum aestivum* L.) representing retrotransposon *Wis*-2, which is very similar to *BARE*-1, were found. Of these, shown in Table 2, we could place only one into LTR group I and one into group X. Eight fell into group H and eight displayed equal similarity to H and X. Hence, the *Wis*-2 retrotransposons of wheat appear most closely related to the *BARE* retrotransposons of the H genome species, which include Asiatic, North American, and South American diploids.

Localization of the Nucleotide Heterogeneity

In our survey of LTR sequence variation, we quantified levels of nucleotide polymorphism within each retrotransposon family using Nei and Jin's (1989) measure of nucleotide diversity, π. Only the full-length sequences of Table 1 were analyzed. For all the sequences, the variability was 0.12193 $\pm$ 0.00669. The nucleotide variability is lower for I sequences (0.08519 $\pm$ 0.00902) and higher for the other two groups: 0.10579 $\pm$ 0.09350 for the H group, and 0.11478 $\pm$ 0.01620 for the X group. The variability is not uniformly distributed throughout the sequences. The nucleotide diversity (π) was studied using a sliding window of 50 bp and a step of 5 bp (Fig. 2). Three regions showed higher sequence diversity than the others: a short stretch next to the 5′ end; a tandem array close to the 3′ end; and the region between the two putative TATA boxes. The first hypervariable region is very short, about 23 bp, located approximately 250 bp from the 5′ end of the LTR. The variability in this segment is higher than that in the surrounding regions when considering all the sequences together, for the I and H groups, and when comparing the I-group sequences with those of the X or H groups. The data indicate that the short region has two forms, one for I and one for H and X together, which is uniform inside the families.

Tandem Repeats in the 3′ Region of the BARE-*1 LTR*

A region about 165 bp from the 3′ end of the LTR is composed of an array of tandemly repeated short sequences. In LTRs of group I, the tandem array is composed of a 12-bp unit. In groups H and X, the basic unit is 23 bp long and the first 12 bp are 100% similar to the repeat in LTRs of group I. The last 11 bp are 62% similar to the first 12 bp. The number of repeats is variable. PCR amplification with primers surrounding the tandem array, using genomic DNAs extracted from various *Hordeum* species, show a ladder of bands corresponding to different numbers of copies of the tandem repeats (Fig. 3). Two different band patterns were observed. In the species with Y or I genomes (Y, *H. murinum*; I, *H. vulgare* and *H. bulbosum*), three main bands were detected. Based on the position of the primers and a tandem unit of 12 bp, these may correspond to tandem arrays with three, five, and six repeats, respectively. This is consistent with the sequenced LTRs of *H. vulgare* in which five repeats on average were found. In the species with H or X genomes, at least nine bands were detected. Assuming a tandem unit of 23 bp, the most intense bands must correspond to tandems containing 1, 2, 3, 4, and 7 repeats.

Table 2. Database accessions of *BARE*-1-like LTR sequences

Sequence	Length (bp)	Part aligned	Similarity to consensus (%)			Species
			I	H	X	
HVGNIRE	1081	760–1930 (end)	**89**	73	85	*H. vulgare*
HVU3173-1	1059	760–1920 (end)	**80**	73	72	*H. vulgare*
HVU3 173-2	1058	760–1920 (end)	**82**	78	74	*H. vulgare*
HVE73173	1057	760–1910 (end)	**82**	75	74	*H. vulgare*
BE422053	790	1–790	**91**	87	85	*H. vulgare*
BF268128	730	890–1620	71	**73**	71	*H. vulgare*
BF620707	672	581–1253	**93**	77	76	*H. vulgare*
BE422145	634	357–991	**98**	89	87	*H. vulgare*
BE259040	630	880–1510	**67**	63	63	*H. vulgare*
HVITR1	624	1–630	90	86	86	*H. vulgare*
HV7RC	598	1–600	**91**	83	82	*H. vulgare*
HVU9760	486	380–860	**95**	90	89	*H. vulgare*
BF620457	484	746–1230	**82**	73	73	*H. vulgare*
BF618290	455	1–455	**96**	87	84	*H. vulgare*
BF620150	395	1430–1900 (END)	**92**	80	79	*H. vulgare*
BE421333	373	775–1148	**88**	79	76	*H. vulgare*
HVE1334QRC	279	1–135 1750–1900 (end)	**90**	87	88	*H. vulgare*
HVE08492R	278	1–130 1650–1820 (end)	**93**	90	90	*H. vulgare*
HVE1327QR	276	1–140 1760–1900 (end)	87	**91**	**91**	*H. vulgare*
HVE1393VR	247	1–250	**97**	89	88	*H. vulgare*
AQ248420	230	765–995	**82**	80	77	*H. vulgare*
HVE1326PR	219	1490–1900 (end)	**79**	75	74	*H. vulgare*
HVDS5	200	1640–1900 (end)	**88**	83	84	*H. vulgare*
HVDS12	198	1700–1900 (end)	**81**	76	74	*H. vulgare*
HV1334QRC	182	1–140	**90**	87	88	*H. vulgare*
HVDS97-1	181	10–190	91	**95**	**95**	*H. vulgare*
HVDS97-2	181	10–190	79	**85**	81	*H. vulgare*
HV1374YRC	165	1740–1900 (end)	**80**	77	77	*H. vulgare*
HVE1321IR	161	1670–1830 (end)	**91**	89	89	*H. vulgare*
HVE1357UR	161	1730–1900 (end)	**93**	86	86	*H. vulgare*
HVE1324NR	159	1750–1900 (end)	**99**	91	92	*H. vulgare*
HV1310KRC	151	1750–1900 (end)	85	**88**	**88**	*H. vulgare*
BE230934	150	500–650	**74**	69	70	*H. vulgare*
HV4RCX	150	1675–1830 (end)	**93**	86	87	*H. vulgare*
HV1357UR	149	1750–1900 (end)	**93**	86	86	*H. vulgare*
HVE1323LR	149	1750–1900 (end)	**97**	87	87	*H. vulgare*
HVE1329TR	149	1680–1830 (end)	**89**	84	84	*H. vulgare*
HV1R	148	1750–1900 (end)	**92**	88	87	*H. vulgare*
HV22R	148	1750–1900 (end)	**91**	84	84	*H. vulgare*
HV1371VRC	147	1–180	**96**	88	88	*H. vulgare*
HVE1395XR	147	1–150	**89**	82	82	*H. vulgare*
HV4R	146	1–150	**96**	89	89	*H. vulgare*
HVE1322KR	145	1–145	**93**	90	91	*H. vulgare*
HV1RC12DL	137	1–140	**88**	83	85	*H. vulgare*
HV19RC6DL	136	1–140	88	89	**90**	*H. vulgare*
HV13250R	135	1–120	81	81	**82**	*H. vulgare*
HV22RCX	133	1–130	**95**	89	90	*H. vulgare*
HVE1371VRC	133	1–130	**96**	88	88	*H. vulgare*
HV1328SR	132	1–130	**96**	89	89	*H. vulgare*
HV1393VR	132	1–130	**97**	89	89	*H. vulgare*
HV2RC	132	1–130	**96**	87	87	*H. vulgare*
HV1374YRCX	131	1–120	86	**91**	90	*H. vulgare*
HV1319LRC	130	1–130	**96**	89	89	*H. vulgare*
BF618289	127	1670–1797	**88**	85	84	*H. vulgare*
HV1318IRC	119	1–120	87	**94**	92	*H. vulgare*
HV19R	117	1750–1870	**93**	85	85	*H. vulgare*
HVDS4	84	1740–1820	86	**91**	90	*H. vulgare*
HV1324NR	72	1825–1900 (end)	**94**	85	84	*H. vulgare*
HV21RC	64	70–130	91	**95**	**95**	*H. vulgare*
TAWIS-21AI	1755	All	72	77	76	*T. aestivium*
BE422772	628	56–684	79	**84**	**84**	*T. aestivium*

(continued)

Table 2. Continued

Sequence	Length (bp)	Part aligned	Similarity to consensus (%)			Species
			I	H	X	
BLYACL3	606	800–1410	68	**82**	76	*T. aestivium*
BE498523	515	295–810	78	**82**	81	*T. aestivium*
TAGLUIN2	470	1310–1900	67	**72**	70	*T. aestivium*
BE398879	412	8–420	76	**81**	**81**	*T. aestivium*
AF139202	384	1–430	69	**72**	**72**	*T. aestivium*
TAE304467	364	1536–1900 (END)	70	**73**	70	*T. aestivium*
TAH2A274	325	1080–1900 (end)	80	**83**	80	*T. aestivium*
WHTHIH2A	325	1080–1250 1750–1900 (end)	68	**71**	**71**	*T. aestivium*
TAGLUIN1	300	1–300	81	88	**89**	*T. aestivium*
BE604224	270	1–270	83	**90**	**90**	*T. aestivium*
TAAWJL236	245	1530–1900 (end)	77	**80**	**80**	*T. aestivium*
AFO29897	231	1–230	79	**83**	**83**	*T. aestivium*
TAE303051	205	1–205	93	81	80	*T. aestivium*
BE497166	175	755–930	80	**86**	84	*T. aestivium*
TAREPTA2	73	1–70	84	**86**	**86**	*T. aestivium*
WHTHFH1B	52	1770–1820 (end)	74	**76**	72	*T. aestivium*

Note. For each accession, the matching region of the *BARE* LTR is displayed, together with the percent similarity to each of the three groups of LTRs that correspond to genome types. The highest degree of similarity to the consensus sequences for each accession is shown in bold.

Two of the species have a slightly different pattern. In *H. roshevitzii*, the band corresponding to two repeats is more intense than it is in the other species. This species is the one with the lower average number of repeats in the sequences. In contrast, in *H. marinum* the average size of the bands seems to be higher and bands corresponding to 10 repeats or more are detectable. Accordingly, the sequences in this species have a higher average number of repeats. These data confirm the overall sequence-based placement of LTRs into the I, H, and X groups and suggests that Y genomes contain LTRs more similar to those of the I group.

Variability in the Promoter Region and Differential Expression of BARE *Subfamilies*

Upstream from the tandem array, the region containing the two *BARE* promoters was the most variable of the LTR. For the LTRs examined, especially for the within-group or pairwise comparisons, more sequence divergence was found in the region surrounding the second TATA box than surrounding the first (Fig. 2). The amount of variability surrounding the second TATA box is lower within the LTR groups than between them, but nevertheless remains higher than that for the LTR as a whole. The region upstream of the first TATA box shows little variability in the I-group LTRs; this box was earlier shown to contribute little to *BARE*-1 expression in barley (Suoniemi et al. 1996).

To better understand the variability in this region, we PCR-amplified, cloned, and sequenced more genomic sequences from it. We designed primers in the conserved flanks surrounding this region and cloned, from *H. vulgare*, ten genomic sequences, five from each of two independent amplification reactions. A total of about 470 bp were compared, located between the two TATA boxes. When a phylogenetic tree was generated from alignments of these sequences together with *Wis*-2 (ALIGN_000706, available via the SRS tool from the EMBLALIGN database at http://srs.ebi.ac.uk/srs6bin/cgi-bin/wgetz?-page + query + -libList + EMBLALIGN + -newId), the results were similar but not identical to those obtained previously (representatives are presented below in Fig. 5). All the sequences not from *H. vulgare* clustered together except *hrosd* and *Wis*-2, as we show above. The others form two clusters. One contains *Wis*-2 and *H. vulgare* hvBARE-2-5 and hvBARE-2-3 (the two LTRs of the same element, in accession AJ279072).

A low level of transcription in unstressed plant tissues has been demonstrated for several retrotransposons (Vicient et al. 2001) including *BARE*-1 (Suoniemi et al. 1996). Because the variable region includes a fragment previously shown to be necessary for the promoter activity of the *BARE*-1a LTR (Suoniemi et al. 1996), we decided to check the sequence conservation of the region in transcripts. We generated cDNAs from *H. vulgare* leaves and calli. PCR-amplified the variable region, and cloned and sequenced the products. A total of 17 sequences from leaves and 18 from calli were produced and aligned with the genomic DNA sequences described above (ALIGN_000706). The variable region of the alignment (Fig. 4) divides the

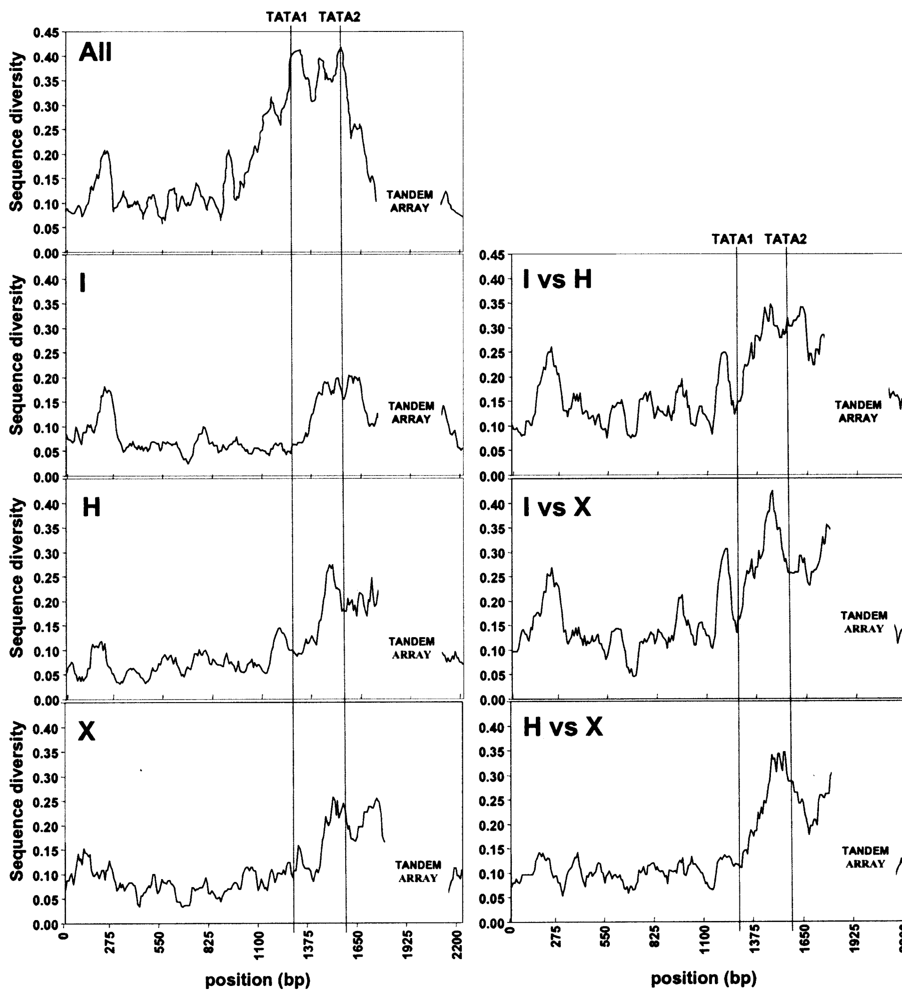

Fig. 2. LTR sequence divergence. The values on the x-axis correspond to the nucleotide position in the LTR alignment. The values on the y-axis are the nucleotide diversity (π) measure, the percentage of divergent nucleotides relative to the number of informative bases, calculated using a sliding window of 50 bp and a step of 5 bp, by the DNASp program (Rozas and Rozas 1999). Insertions and deletions were considered as single mutational events. The position of the two TATA boxes and the tandem array are indicated. Data are shown for all sequences described in Table 1 together, for each LTR group, and for pairwise comparisons between groups.

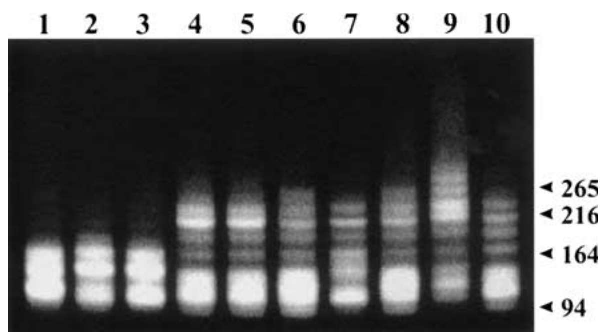

Fig. 3. Variation in the copy number of the tandem arrays in the LTR. PCR amplification of the *BARE*-1 LTR region containing the tandem array was carried out and the products separated by agarose gel electrophoresis. An ethidium bromide–stained gel is shown. Multiple bands indicate the presence of multiple classes of LTRs differing in their number of repeats. The templates were from: 1, *H. vulgare*; 2, *H. bulbosum*; 3, *H. murinum*; 4, *H. euclaston*; 5, *H. pusillum*; 6, *H. brachyanterum*; 7, *H. roshevitzii*; 8, *H. depresum*; 9, *H. marinum*; 10, *H. patagonicum*. Size markers in bp are indicated on the right.

sequences into three groups. The upper two in the alignment show good conservation across the alignment, whereas the third, which includes *Wis*-2, is considerably more variable. These three groups differ much more in the promoter than in other parts of the sequence.

A phylogenetic tree generated from the genomic and cDNA sequences (Fig. 5) defines three major groups. The non-barley *Hordeum* sequences, except hrosd, are separated from a cluster containing *Wis*-2, the hvBARE-2-5 and hvBARE-2-3 sequences, and 15 cDNAs, 10 from leaves and 5 from calli, with a bootstrap strength of 70%. The other clade, distinct by a bootstrap value of 99%, contains all the other barley genomic sequences, hrosd, 13 calli cDNAs and 7 from leaves. The tree topology was unaffected by gap score. The RNA sequences were interspersed among the *H. vulgare* DNA sequences, although their relative abundance in the sequence groups was different from the DNA clones. This may reflect differences between the groups in their present transcriptional activity and their historic integrational success.

The two strongly supported I-group clades or families were named *BARE*-1 and *BARE*-2. The RNA clones predominated in the *BARE*-2 clade, which contains only two genomic sequences but 15 of

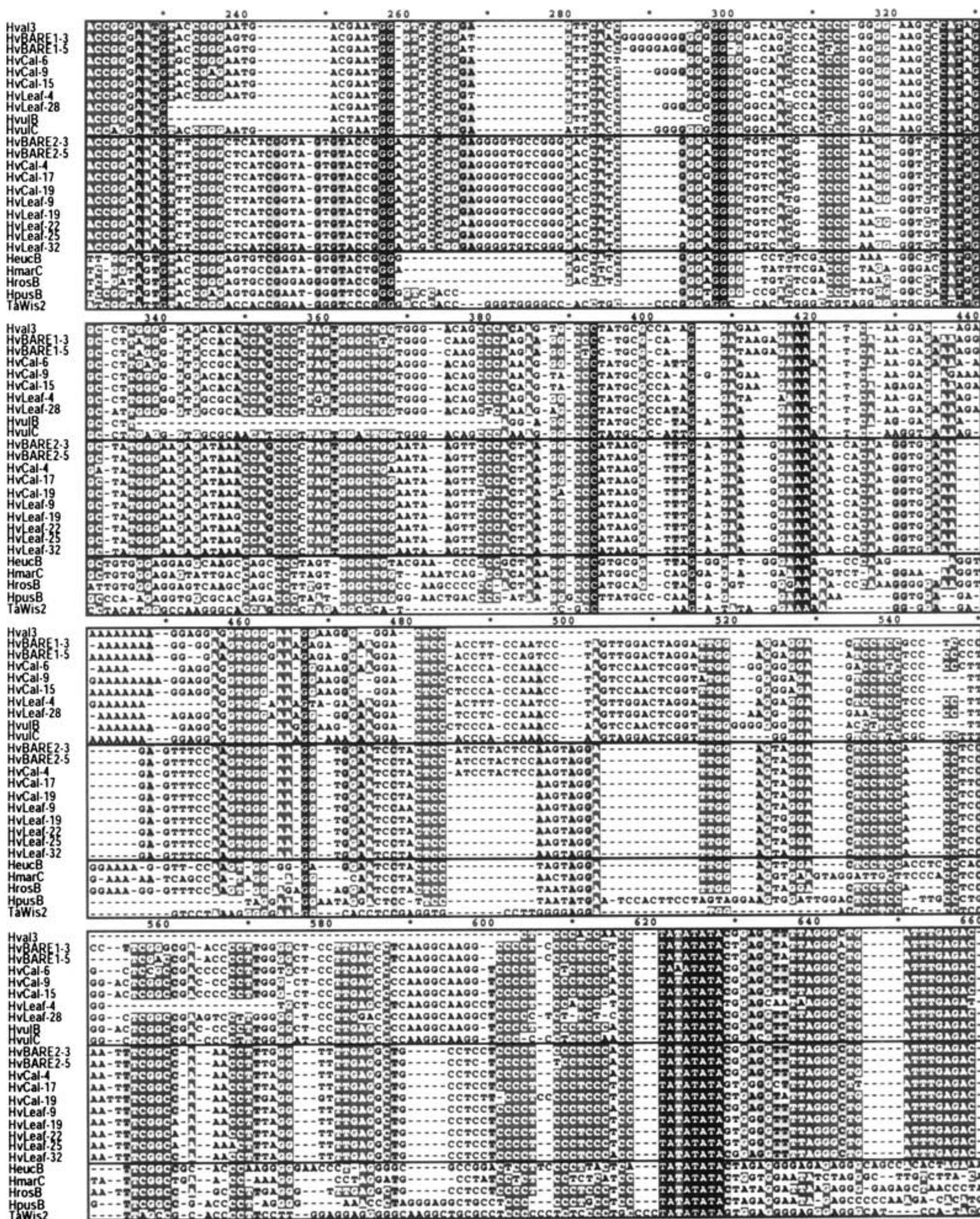

Fig. 4. Alignment of *BARE* LTR promoter sequences. Ten sequences from the complete alignment (ALIGN_000706) were selected from each of the two major clusters and five from the diverse group that is from the H and X genomes. The variable region between the TATA boxes is presented. Three groups of LTR sequences are indicated by boxes over the alignment, with *BARE*-1 and *BARE*-2 together comprising the group I LTRs. The sequence accession names are as in Table 1; HvCal sequences are derived from callus cDNA, HvLeaf sequences from leaf cDNA. Accessions for these are listed in ALIGN_00706 and in Materials and Methods.

the RNA sequences. The *BARE*-1 family, in contrast, contains the other 20 RNA sequences together with 23 genomic sequences. Of the RNA sequences, a higher proportion from callus is found in the *BARE*-1 clade, whereas the *BARE*-2 cluster has relatively more leaf RNA clones. The third clade comprises a diverse group of sequences from the H and X LTR groups.

BARE-*2 Is a Chimeric and Defective Retrotransposon*

In order to clarify the nature of the *BARE*-2 elements, a full-length *BARE*-2 retrotransposon was sequenced from a barley BAC clone (AJ279072). This particular *BARE*-2 element is inserted into a *Bagy2* retrotransposon. The sequences *hvBARE*-2-5 and *hvBARE*-2-3 correspond to the LTRs of this *BARE*-2

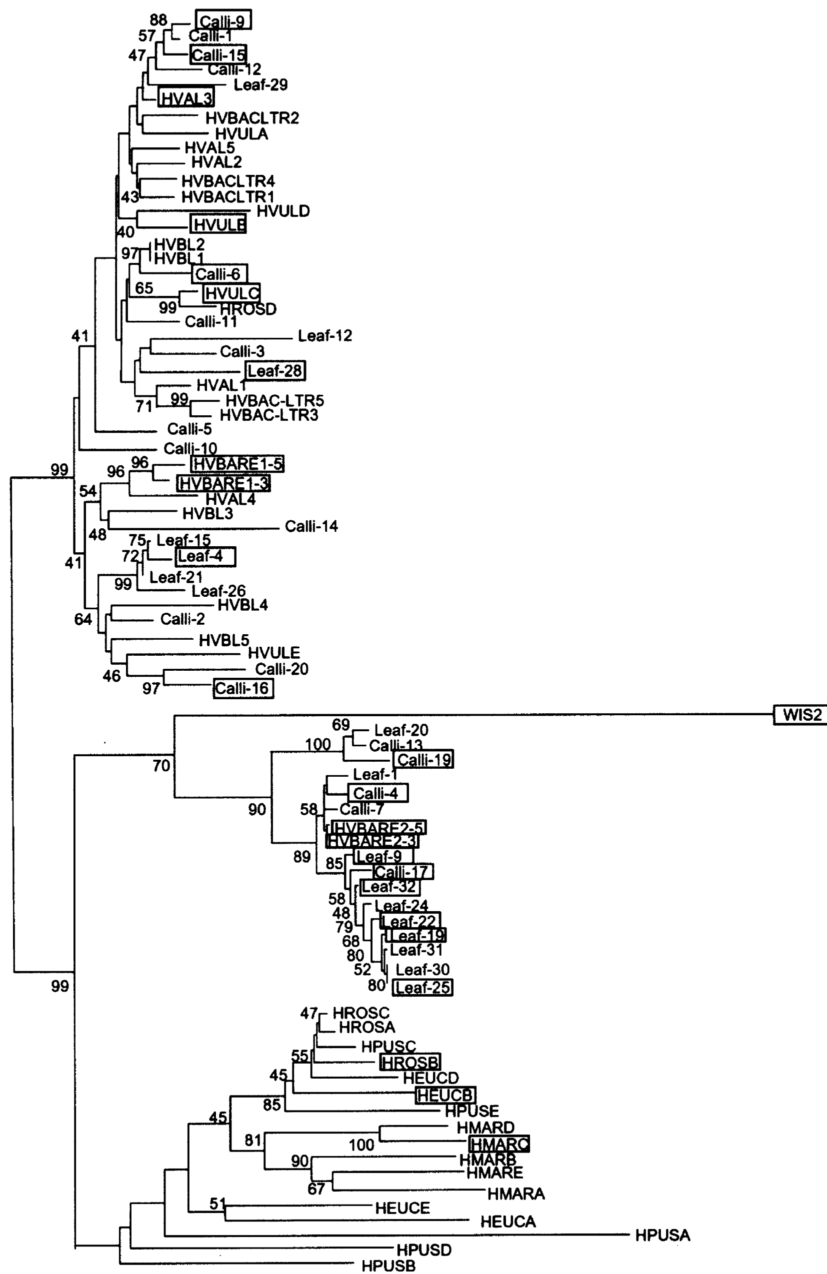

Fig. 5. Phylogenetic tree of the *BARE* LTR promoter sequences. The tree is constructed with the neighbor-joining method. Sequence distances were calculated according to the Kimura two-parameter method, not taking in account insertions or deletions. Accessions in the alignment are listed in ALIGN_00706 and in Materials and Methods. Sequences labeled Calli are derived from callus cDNA and those labeled Leaf are from leaf cDNA. Sequences included in Figure 4 are boxed. Bootstrap values greater than 35%, produced from 500 replicates, are shown. Horizontal line lengths are proportional to evolutionary divergence; those of the vertical lines are only for clarity.

element. The full element is 8615 bp long and contains LTRs of 1811 and 1810 bp. A perfect target site duplication, GGTAC, was found at the insertion site, indicating that the integration event is relatively recent. When translated, however, the deduced *BARE*-2 polyprotein is interrupted by stop codons and frameshifts, as are many *BARE*-1 copies (Suoniemi et al. 1998), especially within the *gag* region. Furthermore, the putative ATG start codon of *BARE*-1a is deleted in *BARE*-2.

When the sequence of *BARE*-2 was compared with the related *BARE*-1 and *Wis*-2a elements (Fig. 6), a sharp dichotomy in sequence relatedness was observed. The untranslated leader region and all of the LTR of *BARE*-2, except the first 240 bases, were more similar to *BARE*-1 than to *Wis*-2a. However, the other parts of the element including the *gag* and *pol* (proteinase, integrase, and reverse transcriptase) coding regions were more similar to *Wis*-2a. The abrupt shift in similarity suggests that *BARE*-2 is chimeric, the product of a recombinational event.

The deletion in *BARE*-2 of the segment containing the start codon of the polyprotein provided a means of designing primers specific for *BARE*-1 and *BARE*-2 elements. These were used with genomic DNAs of other Triticeae members, including *Elymus repens* and various *Hordeum* and *Triticum* species (Fig. 7). By this standard, both *BARE*-1 and *BARE*-2 appear to be present in all species tested except in *Elymus repens*, which failed to amplify a *BARE*-2 product.

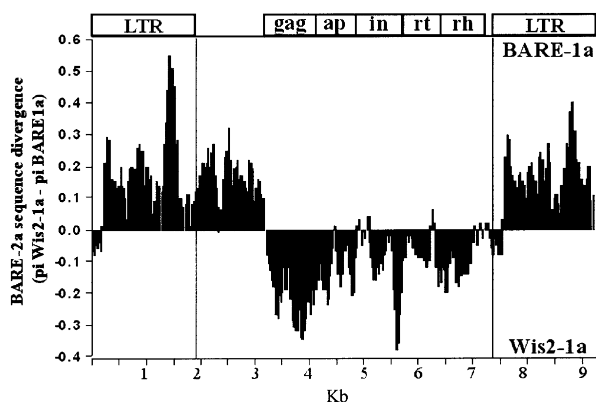

Fig. 6. The chimeric nature of *BARE*-2. The x-axis corresponds to the position in the *BARE*-2 retrotransposon. The value on the y-axis was calculated as the percentage of divergence from *Wis*-2 minus percentage of divergence from *BARE*-1. Insertion and deletions were considered as single mutational events. Divergence was calculated in a window of 60 nucleotides, advanced in steps of 6 nucleotides.

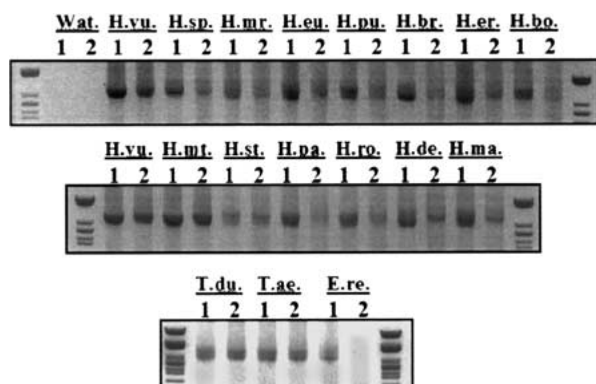

Fig. 7. Presence of *BARE*-2 elements in species of the genera *Hordeum*, *Triticum*, and *Elymus*. PCR amplifications were carried out using primers that specifically amplify sequences of the *gag* region of *BARE*-1 (1) and *BARE*-2 (2) elements. The figure shows a negative image of an ethidium bromide–stained agarose gel following electrophoresis. Wat., water control; H.vu., *Hordeum vulgare*; H.sp. *Hordeum spontaneum*; H.mr., *H. murinum*; H.eu., *H. euclaston*; H.pu., *H. pusillum*; H.br., *H. brachyanterum*. H.er., *H. erectifolium*; H.bo., *H. bogdoni*; H.mt., *H. muticum*; H.st., *H. stenostachys*; H.pa., *H. patagonicum*; H.ro., *H. roshevitzii*; H.de., *H. depresum*; H.ma., *H. marinum*; T.du., *T. durum*; T.ae., *T. aestivum*; E.re., *Elymus repens*.

Chimeric LTRs in Hordeum

Comparisons between *BARE*-2 and the related *BARE*-1 and *Wis*-2a sequences suggested that *BARE*-2 elements might generally be chimeras. This was tested further. We examined the *hrosd* sequence because, although it is from *H. roshevitzii* of the H genome set of species, it belongs to group I, which is mainly composed of *H. vulgare* sequences. The sequence was compared with the consensus sequences for the I, H, and X groups, using a sliding window. The *hrosd* is more similar to H LTRs in the first 520 bases and in the last 200, but is more similar to those

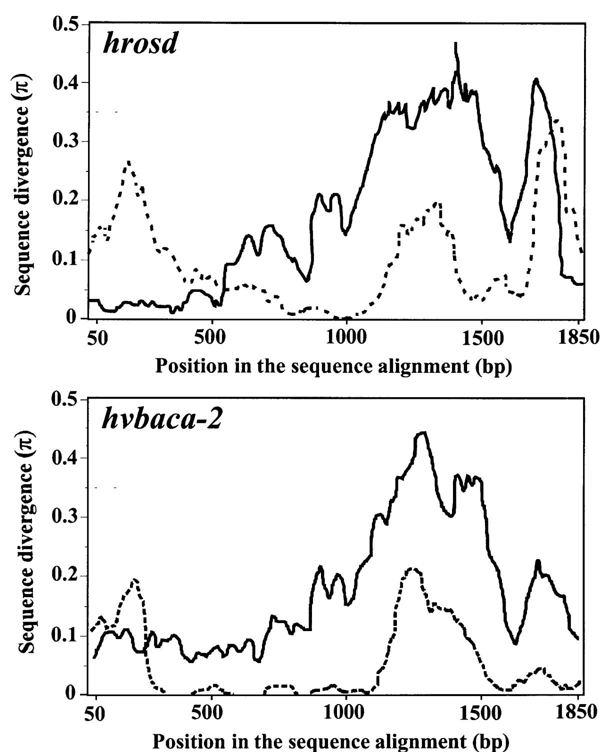

Fig. 8. Evidence for recombination between two LTR sequences belonging to different LTR groups. The values on the x-axis correspond to the position in the LTR alignment. The values on the y-axis were calculated as the percentage of divergence from the consensus sequences of groups I (dotted line) and H (bold line). Insertion and deletions were considered as single mutational events. Divergence was calculated for a window of 60 nucleotides, advanced in steps of 6 nucleotides. The LTR sequence is indicated in each graph.

of group I in the more variable middle region (Fig. 8A). This suggests that the *hrosd* LTR is a chimera between the I and the H LTR types. In order to confirm this, we performed a phylogenetic analysis using different parts of the alignment between the *hrosd* and consensus sequences (not shown). The trees generated are consistent with the previous results; in trees based on the central part of the alignment, *hrosd* fits into the I clade, whereas in trees using the LTR extremes, it clusters with the H clade.

Similar analyses were performed with all other LTR sequences in the study, looking for new cases of chimeric sequences. Five more were detected, with one shown in Fig. 8B. In *hvbaca2*, approximately the first 250 bp are more similar to group H members, and the rest to group I. The same situation was found in *hvbacb2-5*, *hvbacb2-3*, hv*BARE*-2-5, and hv*BARE*-2-3. These results suggest the existence of recombination between LTRs of the different groups to create chimeric LTRs.

Tandem Multimeric BARE-*1 Insertions*

Chimeric retroelements and LTRs of the kind described above may arise through strand switching

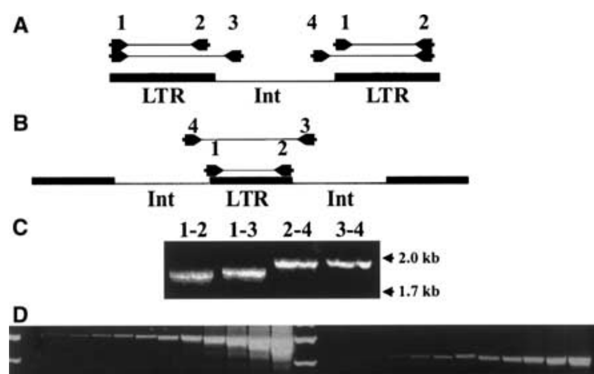

Fig. 9. Tandem *BARE*-1 elements. The positions of the primers, assigned numbers, are shown, respectively, on an intact retrotransposon (**A**), labeled for the LTR and the internal ("ORF") domains, and on a tandem recombinant structure (**B**). The presence of *BARE*-1 internal domains flanking single LTRs in the genome *Hordeum vulgare* was confirmed (**C**) by PCR using primers 3 and 4, matching the internal region of *BARE*-1. The PCR amplification produced a single band of expected size (lane 3-4), which is less abundant than those corresponding to full-length LTRs independent of whether they are from solo LTRs, full-length elements, or tandem elements (lanes 2–4). Lanes 1–2 and 1–3 are, respectively, from all LTRs and the left LTRs of full-length elements. Lane 3–4 was loaded with fivefold more PCR product than the others. The cloning and sequencing of products from lane 3–4 confirmed their origin from tandem structures (data not shown). Products of quantitative PCR are shown in (**D**). The left side of the gel displays lanes with 10–21 cycles of amplification for the tandem recombinant structure. The right side displays 4–15 cycles of amplification for a *BARE* LTR region. The markers visible on the left correspond to 1200 and 1300 bp.

during reverse transcription, by crossovers or gene conversion, or via intrachromosomal recombination. In the latter case, recombination between LTRs of a single element generates solo LTRs. However, if recombination were to take place between the right LTR of one element and the left LTR of another element downstream from the first, a tandem multimeric element could be generated. This would consist of two internal regions and three LTRs, one of which abuts both coding regions. We tested and confirmed the presence of tandem copies of *BARE*-1, sharing one LTR, in the barley genome by PCR.

Outward-facing primers matching the internal regions (respectively, PBS and RNAse H) of *BARE*-1 were designed (Fig. 9). From these primers, two adjacent or nearby full-length elements would yield a product containing two LTRs and possibly intervening genomic DNA of varying size. A nested insertion of one full-length *BARE*-1 into the LTR of another would generate products distinguishable by their sequence, organization, and size. However, the PCR amplification did not reveal such events but instead produced a single band of the size expected for a tandem multimeric structure produced by LTR-LTR recombination. Sequencing of one of these fragments confirmed that they correspond to a shared LTR located between two internal domains.

The prevalence of the tandem structures was estimated by quantitative PCR. The presence of internal domains flanking an LTR was detected with PBS and PPT primers. The abundance of this structure was compared to that of *BARE* LTRs using conserved LTR primers. Using the two reaction products from genomic DNA as templates, we established that the two primer pairs were matched in their efficiency of amplification. The primers were also tested for amplification efficiency over a range of annealing temperatures (55°C to 65°C) and were found to be specific, comparable in efficiency, and robust over that range. The reactions were analyzed only over their logarithmic amplification range (Fig. 9D; cycles 6 to 14 for the LTR, 11 to 19 for the tandem repeat). The *BARE* LTRs were on average 28.4 times more prevalent in the target DNA (barley cv. Sultan) than were LTR-internal domain tandem structures. The *BARE* LTRs were previously estimated at about 1.3×10^5 copies per haploid genome equivalent in barley. Based on this, about 4.6×10^3 such tandem structures are present in the genome.

Discussion

Sequence Heterogeneity in BARE-*1 LTR Sequences*

Replication of retrotransposons is very error-prone, due to the lack of proofreading repair activity by RNA polymerase and reverse transcriptase. As a consequence, the replication of a single retrotransposon can generate a population of closely related, but not identical, sequences resembling the "quasispecies" populations described for RNA viruses (Domingo et al. 1985; Casacuberta et al. 1995). Furthermore, individual copies of retrotransposons are not expected to be under strong selection to maintain function and would accumulate mutations at the neutral rate following their integration. However, those retrotransposon copies with appropriate expression patterns, efficient mechanisms of replication and integration, and non deleterious integration preferences will tend to predominate in the population over time, leading to a feedback loop of purifying selection for functionality (Suoniemi et al. 1998).

If replication errors and mutational drift were the only factors responsible for the sequence variation found in the *BARE*-1 LTR, it might be expected that the variation would be randomly distributed throughout. Our results indicate clearly that this is not the case. The localization of the nucleotide variation within the *BARE*-1 LTR indicates that selective pressure is directed specifically. This can be understood in terms of the LTR function and suggests that there is sufficient transcriptional and integration

activity to provide a selective feedback loop for these retrotransposons.

A retrotransposon under selection will display conservation in those regions that are essential for replication. Although LTRs are the most rapidly evolving region of the LTR retrotransposons (Kulguskin et al. 1981; Lankenau et al. 1990; Lyubomirskaya et al. 1990; Mizrokhi and Mazo 1990), they also contain some functionally important regions: the terminal segments, recognized by integrase, the promoter and enhancer elements, and the RNA processing signals. The data here show that the termini of *BARE* LTRs are conserved. Two functional TATA boxes have been detected in *BARE*-1 LTRs and the regions important for TATA2-promoter activity have been determined (Suoniemi et al. 1996). The region upstream of the TATA1 promoter, which is relatively inactive in transient assays (Suoniemi et al. 1996), is well conserved between the *BARE*-1 and *BARE*-2 families. However, the region of TATA2 necessary for its activity is the least conserved region of the LTR, not only between the families, but also within them. Similar results were reported for *Tnt1* (Casacuberta et al. 1995; Vernhettes et al. 1998), *copia* (Matyunina et al. 1996), and retroviruses (Carpenter et al. 1991; Maury et al. 1997; Montano et al. 1997), although these are single-promoter systems.

High variability in the promoter regions opens the possibility of different transcriptional profiles, a form of niche differentiation, for retrotransposon subfamilies. The sequences reported here for the promoter region show that although *BARE*-2 comprises 92% of the genomic *BARE* copies, it contributes only 72% to the total number of cDNAs derived from callus tissue and just 41% of the total cDNAs sequenced from leaves. Taken together with the distinct promoter sequences of *BARE*-1 and *BARE*-2, these data indicate that these two retrotransposon families are differentially regulated, and that the pools of factors regulating *BARE* transcription vary from tissue to tissue.

The predominance of particular LTR groups in each *Hordeum* genome type suggests that certain forms were favored during the amplification of the *BARE* families, which happened, at least to a great degree, after speciation. In cultivated barley, which we have analyzed more extensively than the other species, all three groups (I, H, X) were found. Therefore, the groups appear to predate the divergence of the *Hordeum* genome types, but in each genome type one group has come to predominate. The higher proportion of full-length *BARE*-1 elements (Vicient et al. 1999) and a lower LTR sequence diversity among the I-group elements in cultivated barley suggest that growth in *BARE* copy number in barley may have been more recent and more pervasive than in the other species investigated. A conceptually similar phenomenon has been reported for the *1731* retrotransposon family of *Drosophila* (Kalmykova et al. 2004). Variants with both altered transcriptional profiles due to changes in the LTR sequence and altered translational strategy due to loss of frameshifting have supplanted the more ancient forms.

Chimeric Retrotransposons Generated Through Template Switching

In addition to sequence variation derived from replicational errors, retrotransposons are subject to recombinational mutagenesis. Four forms of recombination can be distinguished: template-switching during reverse transcription to generate a chimeric cDNA; integrase-catalyzed integration of one element into another; LTR-LTR recombination, generating either solo LTRs or tandem arrays of LTRs and internal domains; and allelic recombination and gene conversion between homologous chromosomes. Ectopic, interchromosomal recombination between retrotransposons is likely suppressed, due to the disruption of chromosomal integrity and the consequent lethality it would cause. The analyses of *BARE* retrotransposons reported here provide evidence for template switching and for the generation of tandem arrays through LTR-LTR recombination.

The full-length *BARE*-2 retrotransposon that is described here displays abrupt switches in sequence similarity between two related families of elements, *BARE*-1 and *Wis*-2. Hence, it appears that this *BARE*-2 is a mosaic or chimeric element generated by stand switching during replication. Retrotransposons (Gabus et al. 1998; Feng et al. 2000) and retroviruses (Hu and Temin 1990) are known to package two RNA templates. In yeast, 14 of the 32 elements previously identified as Ty1 are actually Ty1/Ty2 hybrid elements (Jordan and McDonald 1998, 1999); template switching between packaged Ty1 RNAs occurs with a high frequency (Wilhelm et al. 1999). Recombination within the LTR that affects the regulatory region of the *Drosophila* retrotransposon 412 has recently been reported (Mugnier et al. 2005). In plants, the sole examples of retrotransposon chimeras are the presence of segments of the nonautonomous Dasheng element in retrotransposon RIRE2 (Jiang et al. 2002a).

Template switching is better explored in retroviral than in retrotransposon replication and is a well-known phenomenon (reviewed by Mikkelsen and Pedersen 2000). Although template switching takes place as a normal part of reverse transcription, during the transfer of (−)-strand and (+)-strand strong-stop DNAs (Hu and Temin 1990; Marr and Telesnitsky 2003), it can occur by the jumping of the growing DNA strand to the co-packaged alternative template

at many places across the retroviral genome (Jetzt et al. 2000; Moumen et al. 2001). The process can take place despite mismatched nucleotides (Yu et al. 1998; Marr and Telesnitsky 2003), but template secondary structure is important in determining its likelihood (Mikkelsen and Pedersen 2000; Moumen et al. 2001). The process is thought to increase retroviral fitness through the creation of diversity and altered tropisms and by repair of nonfunctional regions (Mikkelsen and Pedersen 2000). The *BARE*-2 and other chimeric elements reported here support the observation that template switching during replication also may generate diversity for plant retrotransposons, conferring advantages similar to those gained by retroviruses.

In both *hrosd* and *hvbaca2*, sharp increases in sequence divergence relative to the consensus occur near the start of transcription mapped for TATA1 (Suoniemi et al. 1996). The position of the shift away from the consensus sequence suggests it could have arisen through strand switching to a heterologous RNA during reverse transcription. An alternative is gene conversion. Although there is no evidence that gene conversion favors LTRs, at least in yeast it appears to play an important role in retrotransposon evolution (Jordan and McDonald 1999). The size of two conversion tracks in maize was recently estimated in 0.9–1.5 kb (Dooner and Martínez-Ferez 1997), which is similar to the central, divergent region of the *hrosd* sequence. However, although gene conversion may be important in the highly recombinogenic yeast genome, it appears to play a role in only a minority of gene families in the genomes of barley and other cereals (Zhang et al. 2001).

Two lines of evidence suggest that *BARE*-2 is an active retrotransposon despite its nonfunctional protein-coding domain. First, the terminal direct repeats (TDRs) of the *BARE*-2 insertion into *Bagy*-2 are identical and its LTRs are >99% similar, implying a recent integration event. Second, it is more transcriptionally active, especially in leaves, than *BARE*-1, which is an active element. The 5-bp TDRs, furthermore, indicate that the *BARE*-2 copy that was analyzed is not a consequence of post-insertional recombination. The prevalence and conservation in plants of nonautonomous retrotransposons such as TRIM (Witte et al. 2001) and *Dasheng* elements (Jiang et al. 2002a,b), the latter of which are members of the insertionally polymorphic LARD class (Kalendar et al. 2004), show that possession of an open reading frame is not required for the evolutionary success of a retrotransposon.

Recombination Between BARE-1 *Elements*

The means by which retrotransposons are reverse transcribed produces identical LTRs in the cDNA copy that is integrated. Direct repeats were earlier shown experimentally to recombine with an additive frequency, dependent on their length (Puchta and Hohn 1991). Short TDRs normally flank full-length retroelements and are produced as a consequence of integration by the staggered cuts made at the target site by integrase. The TDRs found on the flanks of solo LTRs in the barley genome have been interpreted as resulting from recombination between LTRs of a single retrotransposon, deleting everything in between (Shirasu et al. 2000). The very high ratio of solo LTRs to full-length *BARE*-1 elements in barley implies that LTR-LTR recombination is frequent for this retrotransposon (Vicient et al. 1999).

Recombination between LTRs of two different individual elements can generate a range of products, depending on which two LTRs, distal or proximal in relationship to each other, recombine. Recombination between the right LTR of a nested element with the left LTR of the surrounding element was reported for a single instance in barley (Shirasu et al. 2000). Here, we have demonstrated that LTR-LTR recombination in barley generates repeat units consisting of two internal domains flanking a single, recombinant LTR. This kind of structure has been found in yeast for Ty1 and Ty5 (Ke and Voytas 1997; Kim et al. 1998). Our estimates by quantitative PCR show that these structures are not rare, but are present in about 4.6×10^3 copies per haploid genome. Furthermore, the tandem structure represents only one of four possible outcomes of the recombination of an LTR from one element with that from another (the others being a solo LTR and in two cases a single intact element). *In silico* analyses showed that, of the solo LTRs of 11 low- and middle-copy-number families in the rice genome, 11% appear to be the product of interelement recombination (Ma et al. 2004), though these are less frequent in *Arabidopsis* (Bennetzen et al. 2005). In the orthologous regions of the rice subspecies that have been examined, Ma et al. (2004) estimate that half of the LTR retrotransposons inserted over the last 5 million years have undergone LTR-LTR recombination.

Two alternatives for the production of tandem retrotransposon structures have been proposed for yeast. One involves the demonstrated recombination between cDNAs and integrated copies (Ke and Voytas 1999) by a single-strand annealing mechanism in silent DNA. The other possibility is recombination between a one-LTR circle, such as generated by LTR-LTR recombination within a single element, and an integrated copy (Kim et al. 1998). However, such structures appear rare in the yeast genome (Kim et al. 1998); experimental frequencies are only high when mutations in the integrase or LTR termini block normal integration (Sharon et al. 1994). Hence, the significance for plants of these alternative mech-

anisms for the generation of tandem elements remains to be clarified.

The recombination between LTRs of a single element to generate solo LTRs has been suggested as a mechanism to counteract constant genome expansion due to retrotransposon propagation (Vicient et al. 1999; Shirasu et al. 2000; Bennetzen 2000; Kalendar et al. 2004). However, this process is not fully efficient because it removes only a single internal domain and LTR, together ~7.2 kb for *BARE*-1, and leaves behind a potentially long LTR (1.8 kb for *BARE*-1). Recombination between LTRs of different elements can remove large segments of intervening DNA in one step to generate either solo LTRs with dissimilar flanking TDRs (terminal direct repeats) or tandem elements of the sort demonstrated here.

Recombination between LTRs of different elements to generate a tandem structure, however, can be very deleterious if it removes intervening genes as well. The "gene island," "repeat sea" organization typical of large cereal genomes (Panstruga et al. 1998; Feuillet and Keller 1999; Shirausu et al. 2000; Rostocks et al. 2002; Wei et al. 2002; Park et al. 2004) limits the number of genes interspersed between retrotransposon pairs that could be subject to such a mechanism. Although tandem structures are almost 30-fold less prevalent than solo LTRs, of which there are about 6×10^5 in the barley genome, their number is about 5% that of the estimated 3×10^4 genes in the genome. Furthermore, we would not have detected tandem structures that have been disrupted by further recombination events or by nested insertions. Dispersion of retrotransposons and genes would have subjected at least 1600 genes to loss though recombination. Hence, genome organization where LTR-LTR recombination is frequent may be driven by simultaneous pressures to limit genome expansion and to retain cellular genes.

Acknowledgments. This work was supported by projects 44404 and 47581 from the Academy of Finland. Anne-Mari Narvanto is thanked for her always excellent technical assistance. David Kudrna and Andris Kleinhofs (Washington State University) are thanked for their gift of the barley BAC analyzed here.

References

Beguiristain T, Grandbastien MA, Puigdomènech P, Casacuberta JM (2001) Three Tnt1 subfamilies show different stress-associated patterns of expression in tobacco. Consequences for retrotransposon control and evolution in plants. Plant Physiol 127:212–221

Bennetzen J (2000) Transposable element contributions to plant gene and genome evolution. Plant Mol Biol 42:251–269

Bennetzen JL, Ma J, Devos KM (2005) Mechanisms of recent genome size variation in flowering plants. Ann Appl Biol 95:127–132

Boeke JD, Corces VG (1989) Transcription and reverse transcription of retrotransposons. Annu Rev Microbiol 43:403–434

Carpenter S, Alexandersen S, Long MJ, Perryman S, Chesebro B (1991) Identification of a hypervariable region in the long terminal repeat of equine infectious anemia virus. J Virol 65:1605–1610

Casacuberta E, Pardue ML (2003) Transposon telomeres are widely distributed in the *Drosophila* genus: TART elements in the virilus group. Proc Natl Acad Sci USA 100:3363–3368

Casacuberta JM, Grandbastien M-A (1993) Characterization of LTR sequences involved in the protoplast specific expression of the tobacco Tnt1 retrotransposon. Nucleic Acids Res 21:2087–2093

Casacuberta JM, Vernhettes S, Grandbastien M-A (1995) Sequence variability within the tobacco retrotransposon *Tnt1* population. EMBO J 14:2670–2678

Danilevskaya ON, Arkhipova IR, Traverse KL, Pardue ML (1997) Promoting in tandem: the promoter for telomere transposon HeT-A and implications for the evolution of retroviral LTRs. Cell 88:647–655

Domingo E, Martinez-Salas E, Sobrino F, de la Torre JC , Portela A, Ortín J, López-Galindez C, Pérez-Breña P, Villanueva N, Nájera R, VandePol S, Steinhauer D, DePolo N, Holland J (1985) The quasispecies (extremely heterogeneous) nature of viral RNA genome populations: biological relevance, a review. Gene 40:1–8

Dooner HK, Martínez-Ferez IM (1997) Recombination occurs uniformly within the bronze gene, a meiotic recombination hotspot in the maize genome. Plant Cell 9:1633–1646

Feng YX, Moore SP, Garfinkel DJ, Rein A (2000) The genomic RNA in Ty1 virus-like particles is dimeric. J Virol 74:10819–10821

Feschotte C, Jiang N, Wessler S (2002) Plant transposable elements: Where genetics meets genomics. Nat Rev Genet 3:329–341

Feuillet C, Keller B (1999) High gene density is conserved at syntenic loci of small and large grass genomes. Proc Natl Acad Sci USA 96:8265–8270

Frankel AD, Young JA (1998) HIV-1: Fifteen proteins and an RNA. Annu Rev Biochem 67:1–25

Gabus C, Ficheux D, Rau M, Keith G, Sandmeyer S, Darlix JL (1998) The yeast Ty3 retrotransposon contains a 5′-3′ bipartite primer-binding site and encodes nucleocapsid protein NCp9 functionally homologous to HIV-1 NCp7. EMBO J 17:4873–4880

García-Martínez J, Martínez-Izquierdo JA (2003) Study on the evolution of the *Grande* retrotransposon in the *Zea* genus. Mol Biol Evol 20:831–841

Grandbastien M-A, Spielmann A, Caboche M (1989) *Tnt1*, a mobile retroviral-like transposable element of tobacco isolated by plant cell genetics. Nature 337:376–380

Gribbon BM, Pearce SR, Kalendar R, Schulman AH, Paulin L, Jack P, Kumar A, Flavell AJ (1999) Phylogeny and transpositional activity of *Ty1-copia* group retrotransposons in cereal genomes. Mol Gen Genet 261:883–891

Higgins D, Thompson J, Gibson T, Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

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

Hu WS, Temin HM (1990) Genetic consequences of packaging two RNA genomes retroviral particle: pseudodiploidy and high rate of genetic recombination. Proc Natl Acad Sci USA 87:1556–1560

Jacobsen N, von Bothmer R (1992) Supraspecific groups in the genus *Hordeum*. Hereditas 116:21–24

Jakob SS, Meister A, Blattner FR (2004) The considerable genome size variation of *Hordeum* species (Poaceae) is linked to phylogeny, life form, ecology, and speciation rates. Mol Biol Evol 21:860–869

Jetzt AE, Yu H, Klarmann GJ, Ron Y, Preston BD, Dougherty JP (2000) High rate of recombination throughout the Human Immunodifficiency Virus Type 1 genome. J Virol 74:1234–1240

Jiang N, Jordan IK, Wessler SR (2002a) Dasheng and RIRE2. A nonautonomous long terminal repeat element and its putative autonomous partner in the rice genome. Plant Physiol 130:1697–1705

Jiang N, Bao Z, Temnykh S, Cheng Z, Jiang J, Wing RA, McCouch SR, Wessler S (2002b) *Dasheng*: a recently amplified non-autonomous LTR element that is a major component of pericentromeric regions in rice. Genetics 161:1293–1305

Jordan IK, McDonald JF (1998) Evidence for the role of recombination in the regulatory evolution of *Saccharomyces cerevisiae* Ty elements. J Mol Evol 47:14–20

Jordan IK, McDonald JF (1999) Tempo and mode of Ty element evolution in *Saccharomyces cerevisiae*. Genetics 151:1341–1351

Kalendar R, Vicient CM, Peleg O, Anamthawat-Jonsson K, Bolshoy A, Schulman AH (2004) LARD retroelements: Conserved, non-autonomous components of barley and related genomes. Genetics 166:1437–1450

Kalmykova AI, Kwon DA, Rozovsky YM, Hueber N, Capy P, Maisonhaute C, Gvozdev VA (2004) Selective expansion of the newly evolved genomic variants of retrotransposon *1731* in the *Drosophila* genomes. Mol Biol Evol 21:2281–2289

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

Ke N, Voytas DF (1997) High frequency cDNA recombination of the *Saccharomyces* retrotransposon Ty5: The LTR mediates formation of tandem elements. Genetics 147:545–556

Ke N, Voytas DF (1999) cDNA of the yeast retrotransposon Ty5 preferentially recombines with substrates in silent chromatin. Mol Cell Biol 19:484–494

Kim JM, Vanguri S, Boeke JD, Gabriel A, Voytas DF (1998) Transposable elements and genome organization: A comprehensive survey of retrotransposons revealed by the complete *Saccharomyces cerevisiae* genome sequence. Genome Res 8:464–478

Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16:111–120

Kulguskin V, Ilyin YV, Georgiev GP (1981) Mobile dispersed genetic element MDG1 of *Drosophila melanogaster*: nucleotide sequence of long terminal repeats. Nucleic Acids Res 9:3451–3464

Kumar A, Bennetzen J (1999) Plant retrotransposons. Annu Rev Genet 33:479–532

Kumekawa N, Hosouchi T, Tsuruoka H, Kotani H (2000) The size and sequence organization of the centromeric region of *Arabidopsis thaliana* chromosome 5. DNA Res 7:315–321

Lai MM (1995) The molecular biology of hepatitis delta virus. Annu Rev Biochem 64:259–286

Lankenau DH, Huijser P, Jansen E, Miedema K, Hennig W (1990) DNA sequence comparison of *micropia* transposable elements from *Drosophila hydei* and *Drosophila melanogaster*. Chromosoma 99:111–117

Lerat E, Capy P (1999) Retrotransposons and retroviruses: Analysis of the envelope gene. Mol Biol Evol 16:1198–1207

Lyubomirskaya NV, Arkhipova IR, Ilyin YV, Kim AI (1990) Molecular analysis of the *gypsy* (mdg4) retrotransposon in two *Drosophila melanogaster* strains differing by genetic instability. Mol Gen Genet 223:305–309

Ma J, Devos K, Bennetzen JL (2004) Analyses of LTR-retrotransposon structures reveal recent and rapid genomic DNA loss in rice. Genome Res 14:860–869

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

Marr SF, Telesnitsky A (2003) Mismatch extension during strong stop strand transfer and minimal homology requirements for replicative template switching during Moloney murine leukemia virus replication. J Mol Biol 330:657–674

Matsuoka Y, Tsunewaki K (1996) Wheat retrotransposon families identified by reverse transcriptase domain analysis. Mol Biol Evol 13:1384–1392

Matyunina LV, Jordan IK, McDonald JF (1996) Naturally occurring variation in copia expression is due to both element (cis) and host (trans) regulatory variation. Proc Natl Acad Sci USA 93:7097–7102

Maury W, Perryman S, Oaks JL, Seid BK, Crawford T, McGuire T, Carpenter S (1997) Localized sequence heterogeneity in the long terminal repeats of *in vivo* isolates of equine infectious anemia virus. J Virol 71:4929–4937

McClure MA (1991) Evolution of retroposons by acquisition or deletion of retrovirus-like genes. Mol Biol Evol 8:835–856

McDonald JF (1993) Evolution and consequences of transposable elements. Curr Opin Genet Dev 3:855–864

Mikkelsen JG, Pedersen FS (2000) Genetic reassortment and patch repair by recombination in retroviruses. J Biomed Sci 7:77–99

Mizrokhi LJ, Mazo AM (1990) Evidence for horizontal transmission of the mobile element jockey between distant *Drosophila* species. Proc Natl Acad Sci USA 87:9216–9220

Montano MA, Novitsky VA, Blackard JT, Cho NL, Katzenstein DA, Essex M (1997) Divergent transcriptional regulation among expanding human immunodeficiency virus type 1 subtypes. J Virol 71:8657–8665

Moumen A, Polomack L, Roques B, Buc H, Negroni M (2001) The HIV-1 repeated sequence R as a robust hot-spot for copy-choice recombination. Nucleic Acids Res 29:3814–3821

Mugnier N, Biémont C, Vieira C (2005) New regulatory regions of Drosophila *412* retrotransposable element generated by recombination. Mol Biol Evol:doi:10.1093/molbev/msi060

Nei M, Jin L (1989) Variances of the average numbers of nucleotide substitutions within and between populations. Mol Biol Evol 6:290–300

Nicholas KB, Nicholas HB Jr (1997) GeneDoc: A tool for editing and annotating multiple sequence alignments. Distributed by the author (www.psc.edu/biomed/genedoc)

Nomomura K, Kurata N (2001) The centromere composition of multiple repetitive sequences on rice chromosome 5. Chromosoma 110:284–291

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

Park Y-J, Dixit A, Yoo J-W, Bennetzen J (2004) Further evidence of microcolinearity between barley and rice genomes at orthologous regions. Mol Cells 17:492–502

Puchta H, Hohn B (1991) A transient assay in plant cells reveals a positive correlation between extrachromosomal recombination rates and the length of homologous overlap. Nucleic Acids Res 19:2693–2700

Rostocks N, Park YJ, Ramakrishna W, Ma J, Druka A, Shiloff BA, SanMiguel PJ, Jiang Z, Brueggeman R, Sandhu D, Gill K, Bennetzen JL, Kleinhofs A (2002) Genomic sequencing reveals gene content, genomic organization, and recombination relationships in barley. Funct Integr Genomics 2:51–59

Rozas J, Rozas R (1999) DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. Bioinformatics 15:174–175

SanMiguel P, Tikhonov A, Jin YK, Motchoulskaia N, Zakharov D, Melake-Berhan A, Springer PS, Edwards KJ, Lee M, Avramova Z, Bennetzen JL (1996) Nested retrotransposons in the intergenic regions of the maize genome. Science 274:765–768

Sharon G, Burkett TJ, Garfinkel DJ (1994) Efficient homologous recombination of Ty1 element cDNA when integration is blocked. Mol Cell Biol 14:6540–6551

Shirasu K, Schulman AH, Lahaye T, Schulze-Lefert P (2000) A contiguous 66 kb barley DNA sequence provides evidence for reversible genome expansion. Genome Res 10:908–915

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

Suoniemi A, Schmidt D, Schulman AH (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 MS, Schulman AH (1998) The core domain of retrotransposon integrase in *Hordeum*: Predicted structure and evolution. Mol Biol Evol 15:1135–1144

Van de Peer Y, de Wachter R (1997) Construction of evolutionary distance trees with TREECON for Windows: accounting for variation in nucleotide substitution rate among sites. Comput Appl Biosci 13:227–230

Vernhettes S, Grandbastien M-A, Casacuberta JM (1998) The evolutionary analysis of the Tnt1 retrotransposon in *Nicotiana* species reveals the high variability of its regulatory sequences. Mol Biol Evol 15:827–836

Vicient CM, Schulman AH (2002) *Copia*-like retrotransposons in the rice genome: Few and assorted. Genome Lett 1:35–47

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

Vicient CM, Jaskelainen M, Kalendar R, Schulman AH (2001) Active retrotransposons are a common feature of grass genomes. Plant Physiol 125:1283–1292

von Bothmer R, Jacobsen N, Baden C, Jørgensen RB, Linde-Laursen I (1995) An ecogeographical study of the genus *Hordeum*. International Plant Genetic Resources, Rome

Wei F, Wing RA, Wise RP (2002) Genome dynamics and evolution of the *Mla* (powdery mildew) resistance locus in barley. Plant Cell 14:1903–1917

Wilhelm M, Boutabout M, Heyman T, Wilhelm FX (1999) Reverse transcription of the yeast Ty1 retrotransposon: The mode of first strand transfer is either intermolecular or intramolecular. J Mol Biol 288:505–510

Witte CP, Le QH, Bureau T, Kumar A (2001) Terminal-repeat retrotransposons in miniature (TRIM) are involved in restructuring plant genomes. Proc Natl Acad Sci USA 98:13778–13783

Yamazaki M, Tsugawa H, Miyao A, Yano M, Wu J, Yamamoto S, Matsumoto T, Sasaki T, Hirochika H (2001) The rice retrotransposon *Tois17* prefers low-copy-number sequences as integration targets. Mol Genet Genomics 265:336–344

Yu H, Jetzt AE, Ron Y, Preston BD, Dougherty JP (1998) The nature of Human Immunodeficiency Virus Type 1 strand transfers. J Biol Chem 273:28384–28391

Zhang L, Pond SK, Gaut BS (2001) A survey of the molecular evolutionary dynamics of twenty-five multigene families from four grass taxa. J Mol Evol 52:144–156

Zhong CX, Marshall JB, Topp C, Mroczek R, Kato A, Nagaki K, Birchler JA, Jiang J, Dawe RK (2002) Centromeric retroelements and satellites interact with maize kinetochore protein CENH3. Plant Cell 14:2825–2836