Mol Gen Genet (1999) 261: 883–891

ORIGINAL PAPER

B. M. Gribbon · S. R. Pearce · R. Kalendar
A. H. Schulman · L. Paulin · P. Jack
A. Kumar · A. J. Flavell

Phylogeny and transpositional activity of Ty1-*copia* group retrotransposons in cereal genomes

Received: 8 February 1999 / Accepted: 6 April 1999

Abstract The Ty1-*copia* group retrotransposon populations of barley (*Hordeum vulgare*) and bread wheat (*Triticum aestivum*) have been characterised by degenerate PCR and sequence analysis of fragments of the reverse transcriptase genes. The barley population is comprised of a highly heterogeneous set of retrotransposons, together with a collection of sequences that are closely related to the *BARE-1* element. Wheat also contains a highly diverse Ty1-*copia* retrotransposon population, together with a less prominent *BARE-1* subgroup. These data have been combined with previously published Gramineae sequences to construct a composite phylogenetic tree for this class of retrotransposons in cereal grasses. The analysis indicates that the ancestral Gramineae genome contained a heterogeneous population of Ty1-*copia* group retrotransposons, the descendants of which have proliferated to differing degrees in present-day species. Lastly, the level of recent transpositional activity of two Ty1-*copia* elements has been estimated by measuring their insertional polymorphism within species. Both transposons are highly polymorphic within all species tested. These data suggest that transposition proficiency may be a common and evolutionarily stable feature of the *Ty1-copia* group retrotransposons of cereal grasses.

Key words Retrotransposon · Transposable element · Barley · Wheat · *copia*

Communicated by H. Saedler

B. M. Gribbon · S. R. Pearce · A. J. Flavell (✉)
Department of Biochemistry, University of Dundee, MSI/WTB Complex, Dow Street, Dundee DD1 5EH, UK
e-mail: a.j.flavell@dundee.ac.uk,
Tel.: +44-1382-344244; Fax: +44-1382-322558

B. M. Gribbon · S. R. Pearce · A. Kumar
Scottish Crop Research Institute, Invergowrie, Dundee DD2 5DA, Scotland, UK

B. M. Gribbon · P. Jack
Plant Breeding International, Maris Lane, Trumpington, Cambridge CB2 2LQ, UK

R. Kalendar · A. H. Schulman · L. Paulin
Institute of Biotechnology, University of Helsinki, Viikki Biocenter, PO Box 56, Viikinkaari 9, FIN-00014 Helsinki, Finland

Introduction

Retrotransposons transpose replicatively within host genomes via RNA intermediates. There are three major retrotransposon groups, the Ty1-*copia* group, the *gypsy* group and the non-LTR retrotransposons or LINE elements (reviewed by Kunze et al. 1997). All three groups are found in plant genomes and the Ty1-*copia* group is the best characterised (Flavell et al. 1992a; Voytas et al. 1992; Kubis et al. 1998; Suoniemi et al. 1998). In at least some plant species Ty1-*copia* group retrotransposons are so numerous that they comprise major fractions of the genome (Pearce et al. 1996; SanMiguel et al. 1996).

Highly heterogeneous populations of Ty1-*copia* group retrotransposons are found in many higher plant genomes (Konieczny et al. 1991; Flavell et al. 1992a, 1992b; Voytas et al. 1992). Phylogenetic comparisons among these transposons, both within and between species, have been carried out for dicot plant species (Flavell et al. 1992b; Vanderwiel et al. 1993; reviewed by Kumar et al. 1997 and Kunze et al. 1997). These suggest that Ty1-*copia* group retrotransposons already existed early in plant evolution, and diverged into heterogeneous subgroups before modern plant orders arose; the descendants of these subgroups have been transmitted vertically to the descendant species, with horizontal transfer between species being either very rare or non-existent. In monocots similarly complex sets of interrelated Ty1-*copia* group retrotransposons are found (Hirochika and Hirochika 1993; Hirochika et al. 1996; Matsuoka and Tsunewaki 1996, 1997a, 1997b; SanMiguel et al. 1996; Pearce et al. 1997), some of which

span species boundaries (Pearce et al. 1997; Matsuoka and Tsunekawi 1999).

It is obvious that retrotransposons have been major contributors to the structures of plant genomes but the extent to which these sequences are transpositionally active in present-day species is not clear. Sequence analysis has shown that a large proportion of plant Ty1-*copia* group retrotransposons contain damaged reading frames, indicative of defective elements (Flavell et al. 1992a, 1992b; Voytas et al. 1992; Hirochika and Hirochika 1993). However, a number of plant Ty1-*copia* group retrotransposons are transcribed in their host species (Pouteau et al. 1991; Hirochika et al. 1996; Suoniemi et al. 1996b; Pearce et al. 1997); several maize elements have transposed at least once in the recent evolutionary past (Purugganan and Wessler 1994; White et al. 1994) and several tobacco and rice elements transpose in cultured cells or protoplasts (Grandbastien et al. 1989; Hirochika 1993; Hirochika et al. 1996). Lastly, high levels of insertional polymorphism within species for *BARE-1* of barley and *PDR1* of pea suggest strongly that they have been transposing at least very recently on an evolutionary time scale (Waugh et al. 1997; Ellis et al. 1998; Kalendar et al. 1999). However, these few relatively well studied elements have usually been selected on the basis of transcriptional or transpositional activity and they make up only a very small proportion of the total number of different retrotransposons in their respective genomes. The transpositional activity of the other Ty1-*copia* group retrotransposons in plant genomes remains unknown.

In this study the Ty1-*copia* group retrotransposons of barley (*H. vulgare*) and bread wheat (*T. aestivum*) are characterised and compared with the other Ty1-*copia* retrotransposons of cereal grasses to determine the diversity of this transposon group in the Gramineae and the relationships between the transposons of different species. Finally, the degree of insertional polymorphism of two Ty1-*copia* group retrotransposons within Triticeae species is measured in order to address the issue of the transpositional activity of this transposon group in the recent evolutionary past.

Materials and methods

Isolation and characterization of Ty1-*copia* group retrotransposons

Fragments of the genes for reverse transcriptase from Ty1-*copia* group retrotransposons were amplified by degenerate PCR from barley (cv. Blenheim) and wheat (cv. Chinese Spring) genomic DNAs, using previously described primers and conditions (Flavell et al. 1992a, 1992b; Pearce et al. 1996, 1997). The PCR products were subcloned and sequenced using Sequenase. In keeping with the nomenclature for the retrotransposons of other plants, the new barley elements have been named *Thv* (transposon, *H. vulgare*) and the wheat elements *Tta* (transposon, *T. aestivum*). We have renamed barley and rye transposons previously isolated by our groups (Flavell et al. 1992b; Pearce et al. 1997) in conformity with this system (the latter now carry the prefix *Tsc*). In all these cases, the numbers remain the same, for example clone Bar30 (Flavell et al. 1992b) has been renamed *Thv30*, and R1 (Pearce et al. 1997) is now *Tsc1*.

Sequence accession numbers

Barley (*H. vulgare*): *BARE-1*, Z17327 (Manninen and Schulman 1993); *Thv4*, AJ241322; *Thv6*, AJ241323; *Thv8*, AJ241324; *Thv12–14*, AJ241325–AJ241327; *Thv16–20*, AJ241328–AJ241331; *Thv22*, AJ241332; *Thv35*, AJ241335; *Thv37–39*, AJ241336–AJ241338; *Thv41*, AJ241339; *Thv42*, AJ241340; *Thv129*, AJ241342; *Thv130*, AJ241343 (this study); *Thv29*, AJ241333; *Thv30*, AJ241334; *Thv121*, AJ241341 (Flavell et al. 1992b).

Maize (*Zea mays mays*): *Hopscotch*, U12626 (White et al. 1994); *Opie*, U68408 (SanMiguel et al. 1996); *Mzecopia*, M94481 (Voytas et al. 1992); *Mzepol1* and *Mzepol2*, D12830 and D12831, respectively (Hirochika and Hirochika 1993).

Wheat species: *Wis-2-1A*, X63184 (Murphy et al. 1992); *WHTCOPIA*, M94498 (Voytas et al. 1992); *ABD2*, D90619; *ABD3*, D90620; *ABD12*, D90628; *ABD38*, D90649; *ABD52*, D90662; *D7*, D90675 (Matsuoka and Tsunewaki 1996); *Tta1–22*, AJ241094–AJ241115, respectively (this study).

Rice (*Oryza sativa*): *Rrt1*, Z75496; *Rrt10–13*, Z75505–Z75508, respectively; *Rrt15*, Z75510; *Rrt19*, Z75514; *Rrt20*, Z75515; *Rrt22*, Z75517 (S. Wang, direct submission to EMBL Database); *Tos1*, D12825; *Tos6*, D85865; *Tos7*, D85871; *Tos8–12*, D85866–D85870; *Tos13–19*, D85872–D85878 (Hirochika et al. 1992, 1996); *Riccopia*, M94492 (Voytas et al. 1992); *Rire*, D85597 (Noma et al. 1997).

Rye (*Secale cereale*): *Tsc1*, AJ240103; *Tsc2*, AJ240093; *Tsc3*, AJ240104; *Tsc4*, AJ240099; *Tsc5*, AJ240095; *Tsc7*, AJ240105; *Tsc8*, AJ240092; *Tsc9*, AJ240108; *Tsc15*, AJ240107, *Tsc18*, AJ240110; *Tsc19*, AJ240094; *Tsc21*, AJ240088; *Tsc23*, AJ240109; *Tsc24*, AJ240106; *Tsc25*, AJ240097; *Tsc27*, AJ240098 (Pearce et al. 1997; this study).

Oat (*Avena sativa*): *Astcopia*, M94483 (Voytas et al. 1992).

Sequence comparisons and phylogenetic trees

Computer analyses were carried out at the SEQNET facility of the Daresbury Laboratory. Nucleotide sequences were conceptually translated in all three reading frames; the resulting amino acid sequences were compared visually with corresponding peptide sequences for rye (Pearce et al. 1997) and wheat (Matsuoka and Tsunewaki 1996) to identify frameshifts. The corrected peptide sequences were aligned using CLUSTALW (Thompson et al. 1994).

The TREE/PAPA programme set (Feng and Doolittle 1990) was used for construction of phylogenetic trees. The barley tree incorporated *BARE-1* and three sequences (*Thv29*, *Thv30* and *Thv121*) generated previously (Flavell et al. 1992b). The wheat tree also incorporated *BARE-1*, three sequences from a database search and six sequences isolated by Matsuoka and Tsunewaki (1996; see Results).

The composite tree shown in Fig. 1c was generated from 49 DNA sequences which represent 202 Ty1-*copia* group retrotransposons. The 49 representative Ty1-*copia* group retrotransposon sequences were chosen as follows. First, representatives were picked for each subgroup in the trees for barley (this study), wheat (this study) and rye (Pearce et al. 1997). For instance, in Fig. 1a, sequences *Thv4* and *Thv19* were represented by *Thv19*. These representative DNA sequences were used to create a composite barley/wheat/rye tree (not shown) which identified sequences that are highly homologous between these species. In parallel, trees were constructed for the retrotransposons of rice and maize by the same method. Again, subgroup representatives were chosen from these three trees and these, together with the single oat sequence *Astcopia*, were used to construct the final composite tree (Fig. 1c).

Isolation and sequence characterisation of the *Thv19* retrotransposon

A barley genomic lambda library (a gift of Robbie Waugh) was screened using the 0.27-kb PCR fragment of *Thv19* as a probe. DNA from a positive clone was screened by a modified SSAP (sequence-specific amplification polymorphism) approach (S. R. Pearce, C. M. Stuart-Rogers, A. Kumar and A. J. Flavell, manuscript submitted) to isolate the RNaseH-3′ LTR junction region of the *Thv19* transposon.

Analysis of *Thv19* and *BARE-1* insertional polymorphism

SSAP analysis (Waugh et al. 1997) was used to determine polymorphism levels associated with the elements *BARE-1* and *Thv19*. SSAP analysis of *BARE-1* in the Triticeae (Fig. 2) used a conventional ^{33}P-based manual sequence analysis system. The *BARE-1*-specific primer was 5′-CTAGGGCATAATTCCAACAA-3′, corresponding to the first 19 bases of the *BARE-1* LTR, facing outwards from the 5′ LTR, plus one selective base (A) to inhibit internal priming within the *BARE-1* element from the 3′ LTR. The adaptor primer was *Pst*I-specific with selective bases CGA. SSAP analysis of *Thv19* and *BARE-1* in *H. vulgare* cultivars (Fig. 3) used 5′-GCCCAACCGACCAGGTTGTTACAG-3′, corresponding to bases 48–25 of the *Thv19* LTR (Accession No. AJ241330) or the same *BARE-1* primer as above. In both cases a Cy5 fluorescent tag was attached to the 5′ end of the transposon-specific primer. The adaptor primer was *Pst*I-specific, with selective bases CG, in both cases. The products were sequenced on an ALFExpress (Pharmacia) sequencing apparatus, and the output trace file scored and displayed with Fragment Manager software (Pharmacia). This generated both a table of band positions and a simulated fluorogram, in which the bands represent the peak positions of Cy5 fluorescence.

Results

Characterisation of Ty1-*copia* group retrotransposons of *H. vulgare* and *T. aestivum*

Ty1-*copia* group retrotransposon sequences were amplified from *H. vulgare* and *T. aestivum* by degenerate PCR of a part of the reverse transcriptase gene, using primers which have been successfully used in a variety of higher plant species (Flavell et al. 1992a, 1992b; Pearce et al. 1996, 1997). The PCR products were subcloned, random inserts were sequenced and inferred amino acid sequences were aligned (data not shown).

As expected, a heterogeneous collection of Ty1-*copia* group retrotransposons is found in both species. To visualize the diversity within these transposon populations, phylogenetic trees were constructed for the two sets of sequences (Fig. 1a, b). Previously described Ty1-*copia* group retrotransposons from wheat and barley were included to give an optimal picture for these species. Six of the previously described wheat sequences represent six of the Ty1-*copia* group retrotransposon 'families' defined by Matsuoka and Tsunewaki (1996) and based on a collection of 97 sequences isolated by a similar approach to that used here. *ABD3*, *ABD2*, *ABD12*, *ABD38*, *D7* and *ABD52* represent W families 7, 4, 5, 3, 2 and 1, respectively, of Matsuoka and Tsunekawi (1996); *Wis-2-1A* (Murphy et al. 1992) belongs to W-family 6 (see Table 1).

The two trees (Fig. 1a, b) show the high level of sequence heterogeneity within both species of cereals. Nearly half of the barley sequences are close homologues of the *BARE-1* element. We collectively call these, and other equally closely related sequences from other species, the *BARE-1* subgroup. [Subgroups are defined as sets of related sequences separated by branch lengths of less than 30 distance units (Feng and Doolittle 1987) in the corresponding phylogenetic tree. This corresponds to greater than 80% nucleotide sequence identity.] The predominance of the *BARE-1* subgroup in the barley retrotransposon population is consistent with the known high copy number of *BARE-1* in barley (approximately 7% of the genome; Suoniemi et al. 1996a) and suggests that *BARE-1* is a significant component of the entire Ty1-*copia* group retrotransposon population in this species.

The barley tree shows the strong contribution of the *BARE-1* subgroup, together with a heterogeneous population of sequences. The wheat tree resembles the barley tree, with a *BARE-1* subgroup accompanied by several related subgroups (notably the *Tta12* subgroup), separated from a highly diverse collection of sequences.

The spectrum of wheat sequences obtained in this study is surprisingly different from that obtained in a previous study by a similar degenerate PCR approach (Fig. 1b; Matsuoka and Tsunewaki 1996; 1999). All seven W-families described by these authors cluster close to or within three subgroups in the upper half of the wheat tree (*BARE-1*, *Tta12* and *Tta15* subgroups; Fig. 1b). Therefore, the spectrum of Ty1-*copia* group retrotransposons revealed here is far wider than that indicated in the earlier study. This probably results from differences in PCR amplification, as the two studies used different primer pairs (Flavell et al. 1992b; Matsuoka and Tsunewaki 1996).

Phylogenetic relationships among Ty1-*copia* group retrotransposons from cereal grasses

The wheat and barley sequences reported here were combined with corresponding sequences for other Ty1-*copia* group retrotransposons from Gramineae to gain an overall picture of the diversity of this transposon group in the cereal grasses (Fig. 1c; Table 1). To simplify the tree structure, which is based upon 202 sequences, closely related transposons are represented by individual sequences (Table 1; see Materials and methods). For instance, the *BARE-1* subgroup of barley and wheat (Fig. 1a, b) is represented by *Thv20* and *BARE-1*, respectively.

The 202 sequences in this analysis are indicated by boxes which are colour-coded by species (Fig. 1c). The transposon names corresponding to these boxes are shown on the tree in Fig. 1c if they are representative sequences or in Table 1 if they are not. This allows easy identification of transposon subgroups which span species boundaries (subgroups containing boxes of different

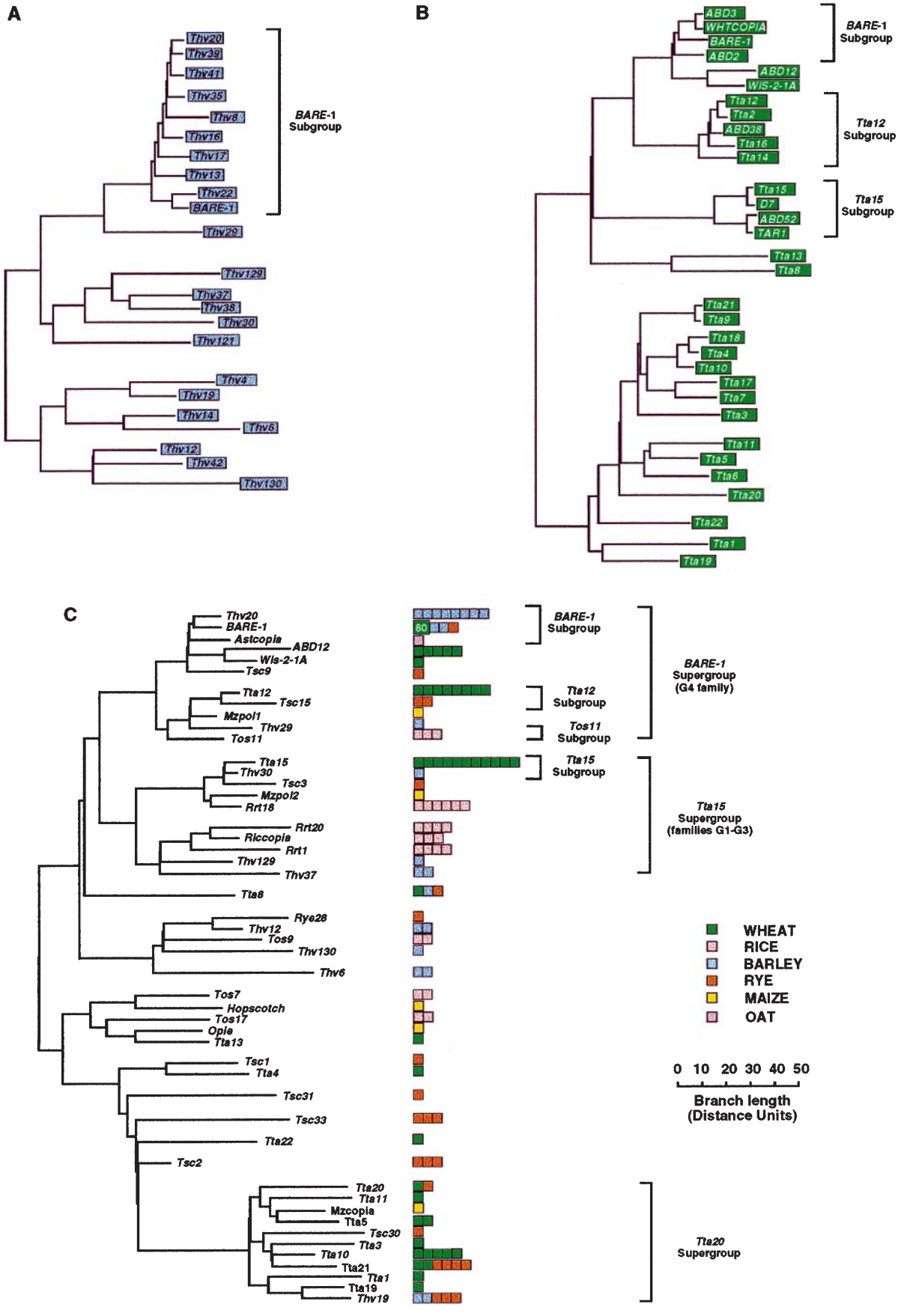
A
Thv20
Thv39
Thv41
Thv35
Thv8
Thv16
Thv17
Thv13
Thv22
BARE-1
Thv29
Thv129
Thv37
Thv38
Thv30
Thv121
Thv4
Thv19
Thv14
Thv6
Thv12
Thv42
Thv130
BARE-1 Subgroup
B
ABD3
WHTCOPIA
BARE-1
ABD2
ABD12
WIS-2-1A
Tta12
Tta2
ABD38
Tta16
Tta14
Tta15
D7
ABD52
TAR1
Tta13
Tta8
Tta21
Tta9
Tta18
Tta4
Tta10
Tta17
Tta7
Tta3
Tta11
Tta5
Tta6
Tta20
Tta22
Tta1
Tta19
BARE-1 Subgroup
Tta12 Subgroup
Tta15 Subgroup
C
Thv20
BARE-1
Astcopia
ABD12
Wis-2-1A
Tsc9
Tta12
Tsc15
Mzpol1
Thv29
Tos11
Tta15
Thv30
Tsc3
Mzpol2
Rrt18
Rrt20
Riccopia
Rrt1
Thv129
Thv37
Tta8
Rye28
Thv12
Tos9
Thv130
Thv6
Tos7
Hopscotch
Tos17
Opie
Tta13
Tsc1
Tta4
Tsc31
Tsc33
Tta22
Tsc2
Tta20
Tta11
Mzcopia
Tta5
Tsc30
Tta3
Tta10
Tta21
Tta1
Tta19
Thv19
60
BARE-1 Subgroup
Tta12 Subgroup
Tos11 Subgroup
Tta15 Subgroup
BARE-1 Supergroup (G4 family)
Tta15 Supergroup (families G1-G3)
Tta20 Supergroup
WHEAT
RICE
BARLEY
RYE
MAIZE
OAT
0 10 20 30 40 50
Branch length (Distance Units)

Fig. 1A–C Phylogenetic trees of Ty1-*copia* group retrotransposons from cereal grasses. Divergences (in distance units; Feng and Doolittle 1987) between retrotransposon sequences are indicated by *horizontal branch lengths*, vertical lengths have no significance. All trees are unrooted, so that the relative lengths of the two horizontal sections of the deepest branches have no significance. However, the sum of the two lengths is equal to the distance separating these from the other sequences. **A** Barley (*H. vulgare*). The *BARE-1* subgroup (see text) is indicated. **B** Wheat (*T. aestivum*). Three subgroups (*BARE-1*, *Tta12* and *Tta15*) are shown. **C** Composite tree for 49 Gramineae Ty1-*copia* group retrotransposons, representing the diversity of 202 transposons. The representative sequences are shown in the tree and the sequences they represent are indicated by *boxes* which are colour-coded by species and named in Table 1. Each box represents a single transposon sequence, except for one in the *BARE-1* subgroup which represents 80, comprising W-families 4 and 7 of Matsuoka and Tsunewaki (1996). Particular transposon subgroups and supergroups mentioned in the text are shown, together with the corresponding G-families described by Matsuoka and Tsunekawi (1999)

colours) and it reveals species-specific bias in transposon distribution within the tree (shown by an asymmetric distribution of coloured boxes across the tree).

Inspection of the Gramineae tree reveals several interesting results. First, the tree structure for the *BARE-1*-related sequences is expanded by the addition of retrotransposons from rye (*Tsc23*) and oat (*Astcopia*). Close homologues of *BARE-1* are therefore broadly distributed among the species of the tribe Triticeae. A further set of sequences also clusters close to *BARE-1* in the composite tree. This includes *ABD12* (the representative for W-family 5; Table 1; Matsuoka and Tsunewaki 1996), *Wis-2-1A* and *Tsc9*. This set of 99 transposon sequences is clearly separated from the nearest sequence relatives, which include eight wheat sequences, two from rye, the *Mzpol1* element of maize, *Thv29* of barley and the *Tos11* subgroup of rice (Fig. 1c). These 114 sequences are very well resolved from all the others in the tree and we have named them collectively the *BARE-1* supergroup to reflect this higher level of organisation. [Supergroups contain members separated by less than 100 distance units, corresponding to approximately 60% nucleotide or 40% amino acid sequence identity. The correspondence between supergroups and the four G-families described by Matsuoka and Tsunekawi (1999) is shown in Fig. 1c.]

The Ty1-*copia* group retrotransposons of individual species show different distributions within the Gramineae tree (Fig. 1c). Wheat and rye sequences form the most diverse species sets, with at least one representative from each species being found in many of the major clades. They are particularly common in the *BARE-1* supergroup, the *Tta15* subgroup and the *Tta20* supergroup (Fig. 1c). In contrast, the barley sequences are more tightly localised within the tree, although most of the supergroups contain barley retrotransposons. The rice Ty1-*copia* group retrotransposon sequences are more restricted within the tree. They are completely absent from the bottom third of the tree and the majority of them (17/26) are members of the *Tta15* supergroup. This supergroup also contains 11 wheat Ty1-*copia* group retrotransposons and a few maize, rye and barley sequences. These data suggest that the *Tta15* supergroup has proliferated in rice and wheat and has persisted, but not thrived, in the other Gramineae genomes. Lastly, the few maize elements used are very widely distributed within the composite tree.

Table 1 Closely related Ty1-*copia* group retrotransposons

Representative transposon	Closely related transposons
Thv20	*Thv39*, *Thv41*, *Thv35*, *Thv8*, *Thv16*, *Thv17*, *Thv13*
BARE-1	*Thv22*, *ABD3* (W-family 7; contains 60 sequences)[a], *WHTCOPIA*, *ABD2* (W-family 4; contains 19 sequences)[a], *Tsc23*
ABD12	W-family 5 (contains five sequences)[a]
Tta12	*Tta2*, *Tta16*, *ABD38* (W-family 3; contains four sequences)[a], *Tta14*
Tsc15	*Tsc24*
Tos11	*Ret12*, *RIRE*
Tta15	*D7* (W-family 2; contains two sequences)[a], *ABD52* (W-family 1; contains seven sequences)[a], *TAR1*
Ret18	*Rrt13*, *Rrt10*, *Tos14*, *Tos15*, *Tos16*
Ret20	*Tos12*, *Tos18*, *Rrt19*
RICCOPIA	*Tos13*, *Rrt11*
Ret1	*Rrt15*, *Tos6*, *Rrt22*
Thv37	*Thv38*
Tta8	*Thv121*, *Tsc7*
Thv12	*Thv42*
Tos9	*Tos10*
Thv6	*Thv14*
Tos7	*Tos8*
Tos17	*Tos19*
Tsc33	*Tsc40*, *Tsc21*
Tsc2	*Tsc37*, *Tsc8*
Tta20	*Tsc35*
Tta5	*Tta6*
Tta10	*Tta18*, *Tta4*, *Tta7*, *Tta17*
Tta21	*Tta9*, *Tsc38*, *Tsc25*, *Tsc27*, *Tsc4*
Thv19	*Thv4*, *Tsc18*, *Tsc19*, *Tsc5*

[a] W families as defined by Matsuoka and Tsunewaki (1996)

Polymorphism of *BARE-1* in Triticeae species

BARE-1 is highly polymorphic among *H. vulgare* cultivars, indicating that it has been transposing within this species in the recent past (Waugh et al. 1996; Kalendar et al. 1999). The data discussed above show that *BARE-1* is widely distributed within the Triticeae tribe. To determine whether it is transpositionally active in other Triticeae species, its insertional polymorphism was analysed by SSAP (Fig. 2). The results show significant levels of insertional polymorphism within *T. aestivum*, *S. cereale* (rye) and *A. sativa* (oat). The polymorphism level is lowest among the *T. aestivum* cultivars (lanes 1–7). This is not surprising, as this species has a low level of polymorphism (Gale et al. 1990). Polymorphism levels are highest among the three rye cultivars (lanes 12–14), with few bands shared between any pair of accessions. The oat cultivars (lanes 15–17) are less polymorphic than rye but more so than *T. aestivum*. We conclude that *BARE-1* has been transposing in all three Triticeae species.

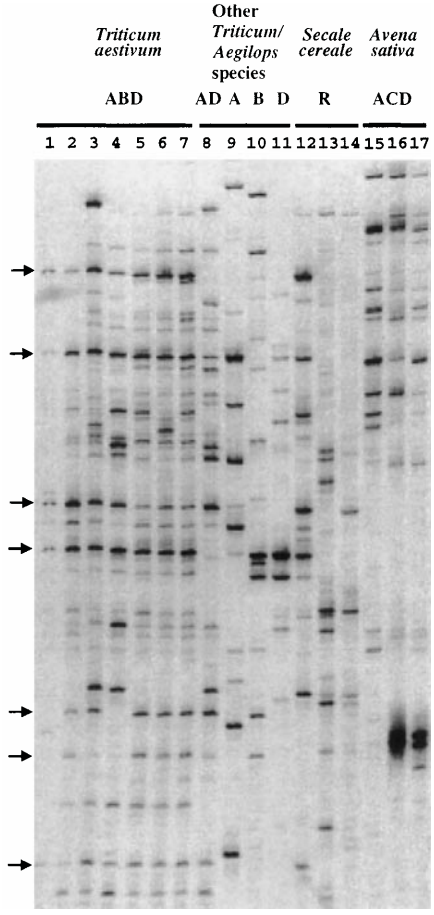

Fig. 2 Insertional polymorphism of *BARE-1* retrotransposons in Triticeae species revealed by SSAP analysis. Bands shared between species are indicated by *arrows* on the left of the Figure. The genome constitutions of *Triticum* and *Aegilops* species are indicated at the top. Lanes 1–7, *T. aestivum* cultivars Chinese Spring (lane 1), Sicco (lane 2), Tjahle (lane 3), Mahti (lane 4), Herzog (lane 5), Petrus (lane 6) and Astron (lane 7); lanes 8–11, other wheat species: *T. durum* (lane 8), *T. urartu* (lane 9), *Ae. speltoides* (lane 10) and *Ae. squarrosa* (lane 11); lanes 12–14; *S. cereale* cultivars KingII (lane 12), Dominant (lane 13) and Cannay (lane 14); lanes 15–17; *A. sativa* cultivars Gerald (lane 15), Kelt (lane 16) and Image (lane 17)

There is a surprising amount of apparent band sharing between some of the species used in Fig. 2. The most obvious of these are arrowed on the left of Fig. 2. Bands shared between *T. aestivum* and the other *Triticum* and *Aegilops* species, which contain different combinations of the three ancestral genomes A, B and D (lanes 8–11), may indicate which genomes have donated particular markers to the *T. aestivum* marker set. There also appears to be band sharing between wheat, rye and oat, which could correspond to ancient shared insertion sites.

Polymorphism of a novel Ty1-*copia* group retrotransposon in *Hordeum*

Most individual copies of plant Ty1-*copia* group retrotransposons carry mutations which disrupt their coding potential (Konieczny et al. 1991; Flavell et al. 1992a, 1992b; Voytas et al. 1992; Hirochika and Hirochika 1993; Manninen and Schulman 1993), implying that the transpositional activity of the great majority of plant Ty1-*copia* group retrotransposons is low or zero. We have tested this proposition by selecting one barley Ty1-*copia* element at random (*Thv19*) and asking whether it is polymorphic within the genus *Hordeum*. To achieve this, LTR sequence data for *Thv19* was needed. Therefore, a genomic clone of *Thv19* was isolated, LTR sequence data obtained (see Materials and methods) and *Thv19*-based SSAP carried out on DNAs from ten *H. vulgare* cultivars. The results clearly show a high level of polymorphism (Fig. 3). The majority of the *Thv19*-associated SSAP bands are polymorphic among nine *H. vulgare* accessions. The corresponding polymorphism level for *BARE-1* for these accessions is somewhat higher (Fig. 3). This is not surprising, because

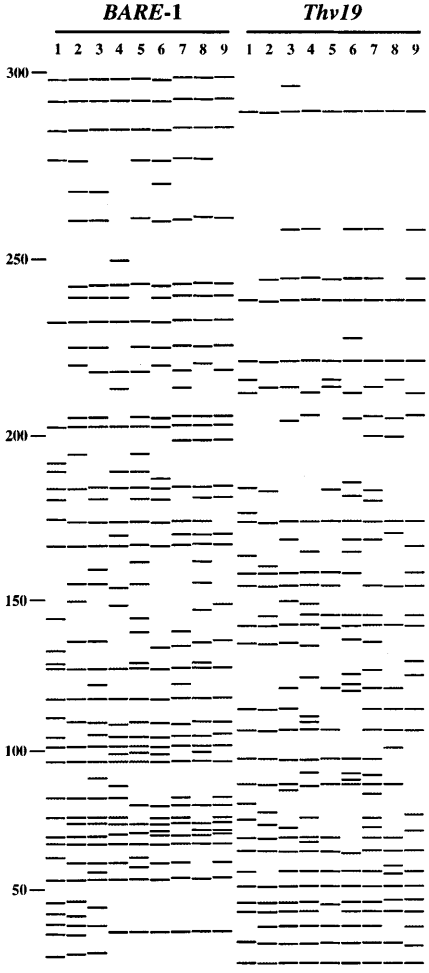

Fig. 3 Insertional polymorphism of *BARE-1* and *Thv19* retrotransposons in *H. vulgare* cultivars, revealed by SSAP analysis. Lanes 1–9, the *H. vulgare* cultivars Djau Kabutak (lane 1), Pallidum 107 (lane 2), Pallidum (lane 3), Bomi (lane 4), Pomo (lane 5), Schooner (lane 6), Kustaa (lane 7), Tampar (lane 8) and Forrest (lane 9). Size standards in bp are shown on the left

BARE-1 is extremely polymorphic within *H. vulgare* (Waugh et al. 1997; Kalendar et al. 1999).

Discussion

Retrotransposons make up a large proportion of the total DNA of higher plants, especially those with large genomes. They also represent one of the most fluid of genomic components, varying greatly in copy number over relatively short evolutionary timescales. They are thus arguably one of the most important factors affecting the structural evolution of higher plant genomes. Therefore, it is important to determine how the basic parameters for these mobile elements vary during the evolution of higher plants. Two important parameters are population structure (how many copies of which kinds of transposons are present) and transpositional activity (what proportion of which transposons are active and what are their transposition rates). This study is intended as a step towards measuring these parameters for the agronomically most important family of plants, the Gramineae.

The phylogeny of Ty1-*copia* group retrotransposons in Gramineae

This study confirms and extends the study of Matsuoka and Tsunekawi (1999) to reveal the extreme complexity of the population structure of Ty1-copia group retrotransposons in the Gramineae. A detailed description of the Ty1-*copia* group retrotransposons of barley is now available for the first time, comparable data for wheat have effectively been doubled, and a similar number of sequences from rye have been incorporated. These data, when combined with existing data for rye, rice, maize and oat, reveal several levels of organisation. The lowest organisational level is the subgroup, which effectively represents a distinct transposon, such as *BARE-1*. Other, deeper levels of organisation are also seen, with clustering of subgroups to form supergroups, such as the *BARE-1* and *Tta15* supergroups.

These rather arbitrary groupings probably reflect the archaeology of the transposons. Subgroups are often shared between closely related species but are not shared among more distant species pairs such as rice and the Triticeae, which diverged roughly 50 Myr ago (Wolfe et al. 1989). This suggests that the subgroups which are discernible now are descended from ancestral transposons which were already distinct from each other in the species lineages that would lead to rice and the Triticeae. In contrast, supergroups are usually spread across the entire extent of the Gramineae, suggesting that their last common ancestral transposon existed before this family arose approximately 60 Myr ago (Wolfe et al. 1989). Taken together, these data suggest that the skeleton of the present-day complex spectrum of Ty1-*copia* group retrotransposons was established before the cereal grasses arose, and its complexity has continued to grow since then, with successful subgroup members amplifying and diverging to form new subgroups, and ancestral subgroups diverging into higher-level groupings.

The distributions of the sequences of individual species within the Gramineae transposon tree are intriguing. The even distribution and high levels of heterogeneity for wheat and rye elements is not surprising, because these genomes probably contain very large amounts of Ty1-*copia* sequences and previous studies in *Vicia* have shown that retrotransposon copy numbers correlate with sequence heterogeneity (Pearce et al. 1996). The barley data are more surprising. These sequences are quite localised within the tree, even though the largest branches contain these elements. Barley also contains very large amounts of Ty1-*copia* sequences, but a large fraction of this is *BARE-1* (Suoniemi et al. 1996a). Perhaps *BARE-1* has restricted the amplification of the other Ty1-*copia* sequences by competing for available genomic space. Lastly, the rice Ty1-*copia* group retrotransposons are the most homogeneous of the four species sets studied in depth here (Fig. 1c). This is underlined by the analysis of 23 new rice sequences (Matsuoka and Tsunekawi 1999), of which only three fall outside existing subgroups. This homogeneity may correlate with the lower copy numbers of these sequences in rice (Hirochika et al. 1992, 1996).

Transposition of *Ty1*-copia group retrotransposons within Triticeae species

A large amount of descriptive data now exists for the wide variety retrotransposons in higher plants. Much less is known about which elements transpose and how frequently this happens during plant species evolution (SanMiguel et al. 1998). This obviously has profound effects on the evolution of the structures of the transposon populations and the overall genome structure of each species, because retrotransposition is a replicative process which can eventually produce hundreds of thousands of copies, comprising major fractions of the plant genome.

Previous studies have shown that *BARE-1* is transcribed in barley (Suoniemi et al. 1996b) and other Triticeae species (Pearce et al. 1997). *BARE-1* insertions are highly polymorphic in barley, demonstrating that it has been transposing in the recent evolutionary past (Waugh et al. 1997; Kalendar et al. 1999). This study shows that *BARE-1* has also been transposing in other Triticeae species, strongly suggesting that the transpositional activity of this mobile element has persisted at a high level for millions of years in multiple species.

The relative rates of transposition in different species lineages for *BARE-1* and the other cereal grass retrotransposons remain to be determined. The levels of *BARE-1* SSAP polymorphism within species revealed in Fig. 2 are probably misleading, because the overall levels of nucleotide polymorphism in these species are very

different, as a result of selective breeding (Gale et al. 1990). To determine relative transposition rates for *BARE-1* within different Triticeae species, *BARE-1* polymorphism levels will need to be compared with other kinds of sequence polymorphism whose rates should be invariant between species.

The presence of *BARE-1* SSAP bands that are shared between different Triticeae species is interesting (Fig. 2). This could be coincidental, but that seems unlikely, considering their high frequency. The possibility that they reflect *BARE-1* internal sequence polymorphism revealed by SSAP priming from the 3′ LTR into the interior of the transposon is also very unlikely, because four specific single base substitutions in *BARE-1* would be required to produce such a band from the primer pair used in this experiment (the 3′ terminal base of the SSAP primer and the three bases in the transposon interior; Manninen and Schulman 1993). A more likely possibility is that they represent *BARE-1* insertions which occurred in the common ancestor of these species and have been preserved in each lineage. This could be tested by isolating and sequencing the SSAP bands.

The insertion sites of *Thv19* are highly polymorphic within barley. The only other barley element studied in this way, *BARE-1*, shows even higher levels of polymorphism. Both transposons were selected at random, without reference to their transpositional activity (this work; Manninen and Schulman 1993). This suggests that transposition proficiency of Ty1-*copia* subgroups may be quite common among the Ty1-*copia* group retrotransposons of cereal grasses. This is in apparent contradiction to the well known preponderance of defective retrotransposons in plants (Konieczny et al. 1991; Flavell et al. 1992a, 1992b;Voytas et al. 1992; Hirochika and Hirochika 1993). These two findings could be reconciled if plant retrotransposon subgroups are mostly made up of untranscribed defective copies, together with a relatively small number of transcribed, transposition-proficient elements which are solely responsible for propagating the subgroup. This is known to be the case for the non-LTR retrotransposons and *Alu* elements of mammals (reviewed by Deininger and Batzer 1995; Boeke 1997). Experiments that focus on the actively transposing Ty1-*copia* group retrotransposons of cereal grasses are needed to test this model.

Acknowledgements We thank Robbie Waugh for the gift of a barley genomic lambda library and DNA samples. We thank Yoshihiro Matsuoka for interesting discussions and for communicating unpublished data. This work was supported by Grant 960508 from the European Commission under the Framework IV Biotechnology Program.

References

Boeke JD (1997) LINEs and *Alu*s – the poly A connection. Nature Genet 16:6–7

Deininger PL, Batzer MA (1995) SINE master genes and population biology. In Maraia R (ed) The impact of short interspersed elements (SINEs) on the host genome. RG Landes-Springer, Austin, pp 43–60

Ellis THN, Poyser SJ, Knox MR, Vershinin AV, Ambrose MJ (1998) Ty1-*copia* class retrotransposon insertion site polymorphism for linkage and diversity analysis in pea. Mol Gen Genet 260:9–19

Feng DF, Doolittle RF (1987) Progressive sequence alignment as a prerequisite to correct phylogenetic trees. J Mol Evol 25:351–360

Feng DF, Doolittle RF (1990) Progressive alignment and phylogenetic tree construction of protein sequences. Methods Enzymol 183:375–387

Flavell AJ, Dunbar E, Anderson R, Pearce SR, Hartley R, Kumar A (1992a) Ty1-*copia* group retrotransposons are ubiquitous and heterogeneous in higher plants. Nucleic Acids Res 20:3639–3644

Flavell AJ, Smith DB, Kumar A (1992b) Extreme heterogeneity of Ty-*copia* group retrotransposons in plants. Mol Gen Genet 231:233–242

Gale MD, Chao S, Sharp P (1990) RFLP mapping in wheat: progress and problems. In: Gustafson JL (ed) Gene manipulation in plant improvement II. Plenum Press, New York, pp 353–363

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

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

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

Hirochika H, Fukuchi A, Kikuchi F (1992) Retrotransposon families in rice. Mol Gen Genet 233:209–216

Hirochika H, Kazuhiko S, Otsuki Y, Tsugawa H, Kanda M (1996) Retrotransposons of rice involved in mutations induced in tissue culture. Proc Natl Acad Sci USA 93:7783–7788

Kalendar R, Grob T, Regina M, Suoniemi A, Schulman A (1999) IRAP and REMAP: two new retrotransposon-based fingerprinting techniques. Theor Appl Genet, in press

Konieczny A, Voytas DF, Cummings MP, Ausubel FM (1991) A superfamily of *Arabidopsis thaliana* retrotransposons. Genetics 127:801–809

Kubis SE, Heslop-Harrison JS, Desel C, Schmidt T (1998) The genomic organization of non-LTR retrotransposons (LINEs) from three *Beta* species and five other angiosperms. Plant Mol Biol 36:821–831

Kumar A, Pearce SR, McLean K, Harrison G, Heslop-Harrison JS, Waugh R, Flavell AJ (1997) The Ty1-*copia* group retrotransposons in plants: genomic organisation, evolution and use as molecular markers. Genetica 100:205–217

Kunze R, Saedler H, Lönnig W-E (1997) Plant transposable elements. Adv Bot Res 27:331–470

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

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

Matsuoka Y, Tsunewaki K (1997a) Presence of wheat retrotransposons in Gramineae species and the origin of wheat retrotransposon families. Genes Genet Syst 72:335–343

Matsuoka Y, Tsunewaki K (1997b) Origin and transmission of some types of family 1 wheat retrotransposons in the two related genera *Triticum* and *Aegilops*. Genes Genet Syst 72:345–351

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

Murphy GJ, Lucas H, Moore G, Flavell RB (1992) Sequence analysis of *WIS-2-1A*, a retrotransposon-like element from wheat. Plant Mol Biol 20:991–995

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

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

Pearce SR, Harrison G, Heslop-Harrison JS, Flavell AJ, Kumar A (1997) Characterisation and genomic organisation of Ty1-*copia* group retrotransposons in rye (*Secale cereale*). Genome 40:617–625

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

Purugganan MD, Wessler SR (1994) Molecular evolution of *magellan*, a maize *Ty3/gypsy*-like retrotransposon. Proc Natl Acad Sci USA 91:11674–11678

SanMiguel P, Tikhonov A, Jin Y-K, 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

SanMiguel P, Gaut BS, Tikhonov A, Nakijama Y and Bennetzen JL (1998) The paleontology of intergene retrotransposons of maize. Nature Genet 20:43–45

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

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

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

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

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

Voytas DF, Cummings MP, Konieczny A, Ausubel FM, Rodermel SR (1992) *copia*-like retrotransposons are ubiquitous among plants. Proc Natl Acad Sci USA 89:7124–7128

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

White SE, Habera LF, Wessler SR (1994) Retrotransposons in the flanking regions of normal plant genes: a role for *copia*-like elements in the evolution of gene structure and expression. Proc Natl Acad Sci USA 91:11792–11796

Wolfe KH, Gouy M, Yang Y-W, Sharp PM, Li W-H (1989) Date of the monocot-dicot divergence estimated from the chloroplast DNA sequence data. Proc Natl Acad Sci USA 86:6201–6205