Genetica (2006) 128:439–447
DOI 10.1007/s10709-006-7725-5

ORIGINAL PAPER

Morgane, a new LTR retrotransposon group, and its subfamilies in wheats

François Sabot · Pierre Sourdille · Nathalie Chantret · Michel Bernard

Received: 6 October 2005 / Accepted: 1 March 2006

Abstract Transposable elements are the main components of grass genomes, especially in *Triticeae* species. In a previous analysis, we identified a very short element, *Morgane*_CR626934-1; here we describe more precisely this unusual element. *Morgane*_CR626934-1 shows high sequence identity (until 98%) with *ESTs* belonging to other possible small elements, expressed under abiotic and biotic stress conditions. No putative functional polyprotein could be identified in all of these different *Morgane*-like sequences. Moreover, elements from the *Morgane*_CR626934-1 subfamily are found only in wheats and *Agropyrum* genomes and among these species, only *Ae. tauschii* and *T. aestivum* present a high copy number of these elements. They are highly conserved in wheat genomes (95.5%). Based on the uncommon characteristics of the described *Morgane*-like elements, we proposed to classify them in a new group within the Class I LTR retrotransposon, the *Morgane* group.

Keywords Evolution · Genome · Grass · Transposable elements · Wheat

Abbreviations

EST	Expressed Sequence Tag
LTR	Long Terminal Repeat
TE	Transposable Element
TRIM	Terminal Repeats in Miniature
LARD	LArge Retrotransposon Derivative

Introduction

Wheat and its relatives include diploid and polyploid species, among which durum and bread wheats (*Triticum turgidum* ssp. *durum* and *Triticum aestivum*, respectively) are the main representatives. The latter is an allohexaploid species carrying three different subgenomes named A, B and D ($2n = 6\times = 42$, AA BB DD), probably created after two independent and spontaneous hybridization events bringing together the three different subgenomes. The first hybridization step was at the origin of the *T. turgidum* species (AA BB genome), bringing together the two diploid genomes from *T. monococcum* ssp. *urartu* (A^uA^u genome) and one or more species of the *sitopsis* section of the *Aegilops* sp. (SS genomes, supposed progenitors of the BB genome). This tetraploid durum wheat and the wild diploid species *Ae. tauschii* (DD genome) hybridized to give the modern bread wheat, 8,000–10,000 years ago (reviewed in Levy and Feldman 2002). The present cultivated hexaploid bread wheat thus carries a large and complex allohexaploid genome of about 17,000 Mb, representing 120 and 40 times the *Arabidopsis* and rice genomes, respectively (Bennett and Leitch 1995), and made

Electronic Supplementary Material Supplementary material is available for this article at http://www.dx.doi.org/10.1007/s10709-006-7725-5

F. Sabot · P. Sourdille · M. Bernard (✉)
UMR INRA/UBP 1095 Amélioration & Santé des Plantes, 234 Avenue du Brézet, F-63039 Clermont-Ferrand Cedex, France
e-mail: michel.bernard@clermont.inra.fr

N. Chantret
INRA, CIRAD-AMIS-Biotrop Avenue Agropolis, F-34398 Cedex 5 Montpellier, France

Present Address:
F. Sabot
Plant Genomics Laboratory, Institute of Biotechnology, FIN-00710 Helsinki, Finland

of near 80% of repeated sequences, mostly transposable elements TEs (Bendich and McCarthy 1970; Flavell et al. 1977; Keller and Feuillet 2000). These TEs are endogenous mobile genomic sequences, which may play an important role in chromatin structure and genome plasticity (reviewed in Kumar and Bennetzen 1999 and Sabot et al. 2004). They are generally classified following firstly their transposition intermediates and secondly their structure and sequence homologies. Class I elements transpose via RNA intermediate, and can be distinguished as LTR (*Long Terminal Repeats*) retrotransposons and non-LTR retrotransposons. Within the LTR retrotransposons subclass, elements are further grouped in *copia*, *gypsy*, *athila*, TRIMs (*Terminal Repeats In Miniature*, Witte et al. 2001), and LARDs (*LArge Retrotransposon Derivatives*, Kalendar et al. 2004) elements, according to their structure and their sequence (Sabot et al. 2004).

In the comparative analysis of the ''*Hardness*'' locus in homoeologous sequences from related wheats (Chantret et al. 2004, 2005), we identified a new TE on the D genome from *T. aestivum*, using the *LTR_STRUC* program (which detects LTR retrotransposons based on their specific structures, McCarthy and McDonald 2003). Thus, on the CR626934 sequence, we annotated this new Class I element within the third intron of the CR626934.3 hypothetical gene. The homoeologous gene was present in the corresponding homoeologous sequence of *Ae. tauschii* (CR626926 BAC, D diploid genome), but without this insertion in its third intron. This TE was short (1.8 kb), and was called *Morgane*_CR626934-1 (Fig. 1), according to the *TREP* nomenclature (*Triticeae* REPeat sequence database, http://wheat.pw.usda.gov/ITMI/Repeats/, Wicker et al. 2002).

In the present work, we studied the structure, the possible origin, the spreading and the unusual repartition of one subfamily of this type of element within the grass genomes, especially in *Triticum* genus.

Materials and methods

Materials

The various species used in the present analysis are listed in Table 1 in Supplementary Data. The accessions of durum and bread wheat were provided by the INRA Genetic Resource Centre from Clermont–Ferrand (France). The DNA of *Spartina townsendii* was kindly provided by Dr M. Ainouche (Rennes University, France), rice DNA by Dr E. Guiderdoni (CIRAD, Montpellier, France), maize DNA by P. Barret (INRA, Clermont–Ferrand, France) and other non-wheat grass DNA by L. Zhang (INRA, Clermont–Ferrand, France). The *Ha* locus sequences are deposited in GenBank under the accession number CR626926 for the *Ae. tauschii* sequence and CR626934 for the *T. aestivum* D sequence (Chantret et al. 2005). The different complete internal cloned sequences from the *Morgane*_CR626934-1 subfamily are deposited in GenBank under the accession numbers AY675322–AY675344.

Methods

Plant DNA was extracted by a standard CTAB method (Murigneux et al. 1993), and BAC DNAs were isolated by a classical alkaline lysis protocol (Sambrook and Russell 2001). PCR DNA bulks were constituted with an equal amount of each single DNA, as indicated in Table 1.

Primers for standard PCR were designed with *FastPCR* v3.2.130 (Kalendar 2004, available at http://www.biocenter.helsinki.fi/bi/bare-1_html/oligos.htm) on the entire CR626934 BAC sequences. The primer sequences are 5′-tcagcaggtatcccaggagc-3′ for *Morgane*-f and 5′-aggccctgtg gatggtgctg-3′ for *Morgane*-r, and located in position 610–630 and 1570–1590 of the *Morgane*_CR626934-1 sequence, respectively (Fig. 1). These primers were designed and used in PCR as in Sabot et al. (2005b).

The Southern blots were performed with a standard alkaline technique for membrane transfer (Sambrook and Russell 2001) as in Sabot et al. (2005b), after digestion by *Hin*dIII (which does not cut in the *Morgane*_CR626934-1 sequence). The probes used were the PCR products obtained from the amplification of *Morgane* on the CR626934 BAC DNA with *Morgane*-f and *Morgane*-r primers. The hybridization stringency allowed classical sequence recognition with 80% of homology (subfamily level).

One microlitre of each PCR product was cloned in the pGEM-T vector in JM109 strain of *E. coli*, as recommended by suppliers (*Promega*) and sequenced by *Genome Express*

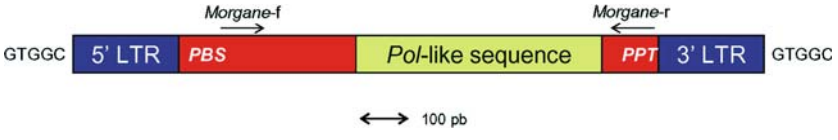

Fig. 1 Schematic representation of *Morgane* element. The LTR and the *Pol*-like homologous sequences are shown, and the primers used for internal sequence amplification are symbolized with arrows. PBS = Primer Binding Site, PPT = PolyPurine Tract, LTR = Long Terminal Repeat, *Pol* = Polyprotein ORF

(Meylan, France). Sequences were aligned and compared after manual editing using *ClustalX* (Higgins et al. 1994, http://www.ebi.ac.uk/clustalw), and *SIM & LALNVIEW* tools available at the *ExPASy* website (http://au.expasy.org/tools/sim-nucl.html, Huang and Miller 1991).

Results

Classification of the Morgane_CR626934-1 element

The *Morgane*_CR626934-1 element was detected with *LTR_STRUC* within the CR626934 BAC sequence: it has short but perfect LTRs (286 bp) and is 1,871 bp long. This element harbours conserved 5 bp Target-Site Duplication (TSD) GTGGC/GTGGC. Its LTRs classically start with TG, end with CA and are 99.7% identical. Its PBS sequence (Primer Binding Site, needed for the negative strand reverse transcription of LTR retrotransposon) was clearly identified as 5′-attggtatctagagccacaaattttt-3′. *LTR_STRUC* proposed a potential PPT (*PolyPurine Tract*, required for the cDNA (+)-strand synthesis) as 5′-accatccacagggcctgcggtgcctc-3′ (on the (+)-strand): this sequence did not answer the canonical purine-tract of the normally identified PPTs and was unusually long, but some other LTR retrotransposons showed the same unusual features. *Morgane*_CR626934-1 has an unusual intronic location in the CR626934 BAC, as retroelements are generally found nested in large intergenic spacers in grasses. The intronic location is here confirmed by the comparison with the homoeologous CR626926 BAC sequence. This BAC has the same predicted gene in the same homoeologous location, with the same exonic sequences, but without the insertion of any element such as *Morgane*_CR626934-1 (Chantret et al. 2005).

No clear homology was found at the nucleic acid level (*BLASTn*) with any already identified transposable elements within the genomic databases. The best homologies on *RepBase* (Jurka 1998) were observed with *OSR* (O*ryza* s*ativa* r*etroelement*) element from rice and *REINA* element from maize, which are both Class I LTR retrotransposons belonging to the *gypsy* group. These homologies hold on the second part of the sequence, and only with *tBLASTx* analyses (from position 994 to 1506 for *REINA*, expect = $5e^{-32}$, 32% identical, 50% positives). The first homology found in the *TREP* databases is with the putative protein (*PTREP740*) from another *gypsy* element, *Sukkula*_AF427791-1 from barley, between position 766 and 1431; this homology was obtained by *BLASTx* search, with a low score of 51.2 and an expect of $2e^{-08}$ only (24% identical, 36% positive). Recently, the *Sukkula* elements were reclassified as *LARDs*, and not as *gypsy* element (Kalendar et al. 2004). These *LARDs* elements do not encode any protein and are dependent of yet unknown active partners to retrotranspose. However, *LARDs* are Class I LTR retrotransposons too. Using the *EMBL non-redundant protein* database (with *BLASTx*), the observed hits are against multiple putative *gypsy*-like polyproteins from *Oryza sativa*, but with various stop-codons scattered all over this homology. With all the different *TIGR* databases (pseudo-molecules, repeats, genomic and translated sequences databases) from rice and other cereals, no hits better than those previously obtained with the other databases are obtained. These bioinformatic analyzes enlighten only a possible, but far, relationship between *Morgane*_CR626934-1 and the *gypsy* Class I LTR retrotransposons (50% maximum of positive amino acids at the proteic level). Thus, *Morgane*_CR626934-1 has an internal sequence potentially related to reverse transcriptase and *Pol* ORF (*Polyprotein Open Reading Frame*) of Class I LTR retrotransposon, but with a lot of mutations leading to numerous stop codons, even if *LTR_STRUC* could predict a short ORF of 372 bp.

Finally, the best hits were obtained using *EST* (*Expressed Sequences Tag*) databases. The best homology was found with the clone *AZO3104L08* from leaves of *T. aestivum* (cv. Renan), obtained from abiotic nitrogen stress assays on bread wheat (*Genoplante* Program Library). This *EST* is homologous to *Morgane*_CR626934-1 on its largest part (98.4% of the *EST*, Fig. 2). This homology recovers first, the end of the 5′ LTR and the starting of the internal region, and second, the end of the internal region, including the potential PPT. The whole part of this internal region is homologous to already identified *Pol*-like sequences on the two sequences. It appears that the *AZO3104L08 EST* might be an expressed version of a shorter *Morgane*-like element. The second best hit was the clone *wlm1.pk0008.a8* from *T. aestivum*, which matched at 98.5% to the LTR sequences from *Morgane*_CR626934-1, and was issued from an infection assay of bread wheat seedlings by *Erysiphe graminis f. sp. tritici* (powdery mildew). The ending reverse complement sequence of this *EST* (upward the poly-A sequence) is homologous to the ending part of *Morgane*-CR626934-1. This homology is limited to the first part of the LTRs, probably the U3 and R regions (see below). The upstream sequence in the *EST* is also homologous to a potential *Pol*-like sequence, which let us suppose that it is another version of an expressed *Morgane*-like element. These two *ESTs* possess thus homology with *Pol*-like sequences, but as for the *Morgane*-CR626934-1 element, these *Pol*-like sequences are full of stop-codons. All the other hits observed with *ESTs* came from biotic and abiotic stress conditions libraries and showed the same kind of homologies with *Morgane*_CR626934-1. None of these *EST* homologies came from the *CR626934.3 hypothetical gene* in which *Morgane*_CR626934-1 was inserted.

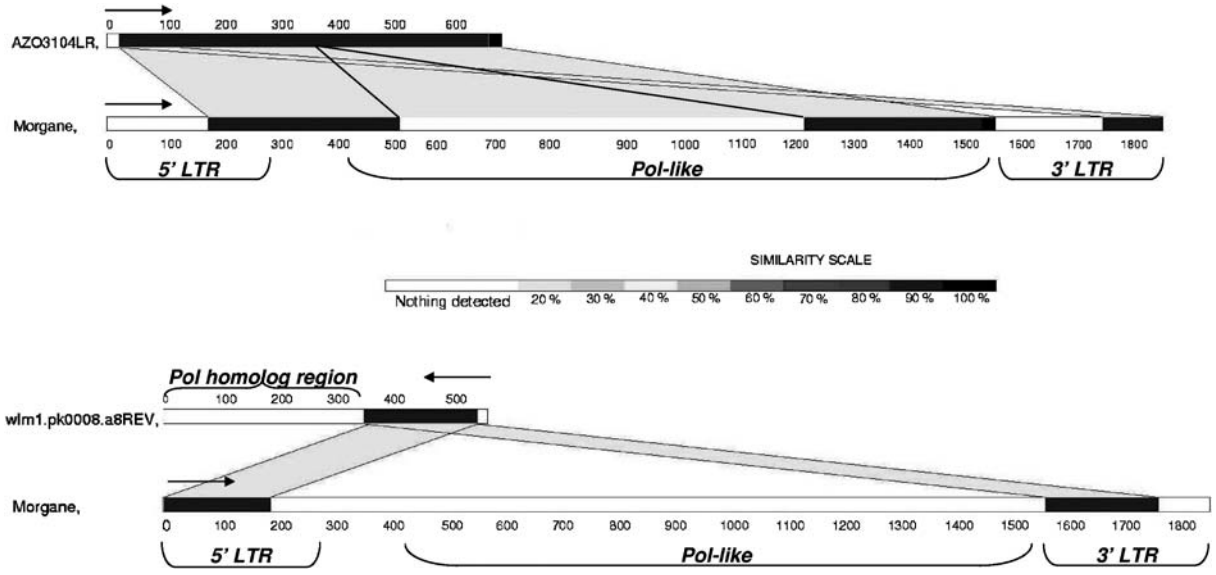

Fig. 2 Alignments between *Morgane*-CR626934-1 and *ESTs AZO3104L08* and *wlm1.pk0008.a8*. *LALNVIEW* and *SIM* alignment tools from the *ExPASy* website services were used here. The plain arrows indicate the forward direction of each sequence. LTR = Long Terminal Repeat, *Pol* = Polyprotein

Spreading of the Morgane_CR626934-1 subfamily in DD genomes of wheat

In order to assess the distribution of the short *Morgane*_CR626934-1 element subfamily in the different *Triticeae* and *Poaceae* genomes, PCR assays were performed with the *Morgane*-r and *Morgane*-f primer pairs (Fig. 1) on DNA from different wheat relatives (Fig. 3a–c) and non-wheat grass species (Fig. 3d and data not shown). The hexaploid wheats (AA BB DD genome, Fig. 3a, c), the synthetic hexaploid wheat (W7984, AA BB × DD genome, Fig. 3a) and the *Ae. tauschii* (DD genome, Fig. 3a) lanes showed a large amount of amplification product of ~1 kb long (expected 980 bp), and a minor band (ca. 500 bp, which was coherent with the homology observed with the *AZO3104L08 EST*). In the diploid species as well as in the ''Russian'' polyploid wheats (*T. timopheevi* A^mA^m GG genome and *T. zhukovskyii* A^uA^u A^mA^m GG genome) and in *Ae. variabilis* (UU S^vS^v genome), the amplifications were much weaker for the 1 kb band (as for the 500 bp one). This result indicates either a high level of copy of the *Morgane*_CR626934-1 subfamily, which is detected only when the DD genome is present, or mispriming due to sequence variations. Interestingly, *Ae. ventricosa* did not show a strong PCR amplification signal (just a very faint amplification similar to ''*non-D*'' genomes), while its classical genomic formula is DD M^vM^v (Fig. 3a). In the same way, *Morgane*_CR626934-1 elements are widely present in almost all bread wheats (Fig. 3c) but not in durum wheat, even in the more ancestral types of durum wheat (Fig. 3b). *Morgane*_CR626934-1 elements from this subfamily were not amplified at all in rice, rye and barley (Fig. 3d), maize, *Brachypodium sylvaticum* and *Spartina townsendii* (data not shown), and only poorly amplified in the *Agropyrum elongatum* species (Fig. 3d). *Agropyrum elongatum* amplification level was similar to that obtained in ''non-*Ae. tauschii* D genome'' wheat species (Fig. 3). Southern experiments with *Morgane*_CR626934-1 probes on wheat species and varieties confirmed that no highly longer version of *Morgane* can be detected (but shorter are detected, as expected following the *EST* analysis), and second that *Morgane*_CR626934-1 subfamily seems to be overrepresented in species with ''*Ae. tauschii* D genome'', i.e. *Ae. tauschii*, synthetic wheat W7984 and all hexaploid bread wheats tested yet (data not shown). The tetraploid wheats as well as the ''non-*Ae. tauschii* D'' diploid wheats only showed a limited and weak hybridization signal, which tended to confirm a low copy number of this specific subfamily already supposed after the PCR experiments. The non-wheat species did not show hybridization with *Morgane*_CR626934-1 probe in standard stringency (data not shown).

We cloned *Morgane* PCR products from different wheat species DNA bulks (Table 1, Supplementary data) and sequenced 192 individual clones. These clones were normally homologous to the internal part of *Morgane*_CR626934-1, as they were amplified with *Morgane*-r and *Morgane*-f primers (position 610–1590; Fig. 1). The sequence alignments of some of these *Morgane*_CR626934-1 related fragments are shown in Fig. 4. These sequences showed a high degree of identity between each other, with an overall identity of 95.5% (from 90% to 100%). Few complete

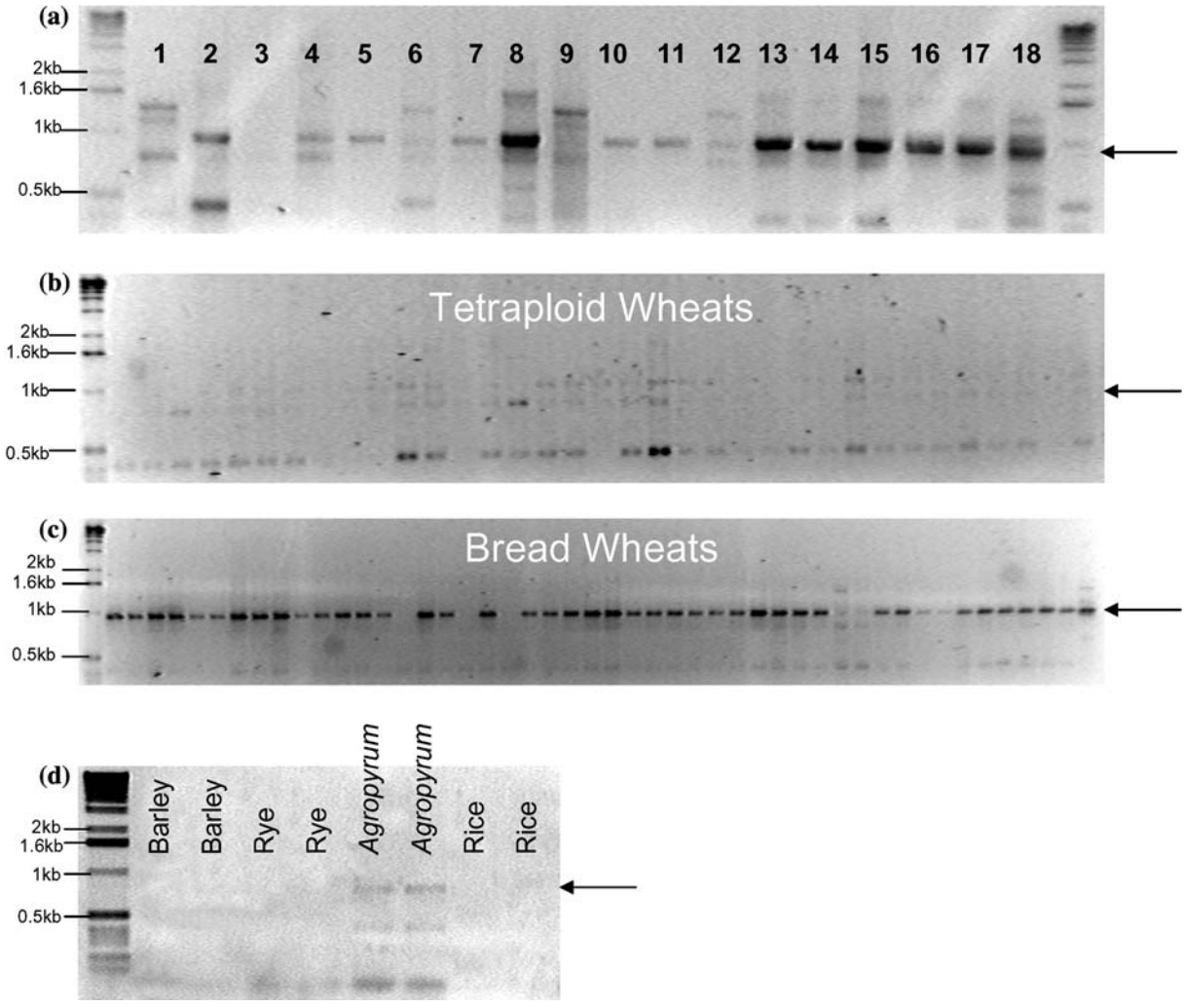

Fig. 3 PCR amplifications of *Morgane*_CR626934-1 subfamily in wheat species and grass genomes. Arrows indicate the expected length of the PCR product (~1 kb). PCR products are flanked by standard 1 kb DNA Ladder. (**a**) PCR amplifications with *Morgane*-f and -r primers. The following DNA species were used [lane number]: *T. monococcum ssp. urartu* [1]; *T. monococcum ssp. monococcum* [2]; *T. monococcum ssp. boeoticum* [3]; *Ae. searsii* [4]; *Ae. longissima* [5]; *Ae. speltoides* [6]; *Ae. bicornis* [7]; *Ae. tauschii* [8]; *T. timopheevi* [9]; *T. zhukovskyii* [10]; *Ae. ventricosa* [11]; *Ae. variabilis* [12]; *Synthetic* wheat W7984 [13]; *T. compactum* var. Hérisson Barbu [14]; *T. spelta* var. Epeautre Blond [15]; *T. sphaerococcum* [16]; *T. macha* [17]; BAC CR626934 DNA [18]. (**b**) PCR assays with *Morgane*-f + *Morgane*-r on 48 representative species of tetraploid wheats (AA BB), (**c**) on 47 representative accessions of hexaploid bread wheats (AA BB DD), and (**d**) on four non-wheat grass species (two reproducible PCR each)

sequences homologous to *Morgane*_CR626934-1 could be cloned from ''non-*Ae. tauschii* D'' wheat species, the others showing stronger homologies with either the *AZO3104L08* or *wlm1.pk0008.a8 ESTs* (in sequences as in size). The ''complete'' clones are not highly different between D and non-D wheat species (90% of identity), and without large differences within a group of sequences: the A sequences are 94.8% homologous, and the D sequences are 96.5% homologous. However, as no ''complete'' sequences could be cloned in the polyploid AB, we can suppose that the copies from this subfamily from AB genomes have been deleted or highly mutated, as we can clone A & B copies separately.

Discussion

Classification of the Morgane-like elements

All the elements belonging to the LTR retrotransposons subclass possess minimal specific structures: *Long terminal Repeats*, *Primer Binding Site* and *PolyPurine Tract*, and *Target-Site Duplication* (Kumar and Bennetzen 1999; Sabot et al. 2004). As the *Morgane*_CR626934-1 element harboured two LTR-like sequences, an identified PBS, a potential PPT, specific TSD, as well as homologies with *Pol*-like sequences, we can classify this element as a Class I LTR retrotransposon. Moreover, *Morgane*-like elements

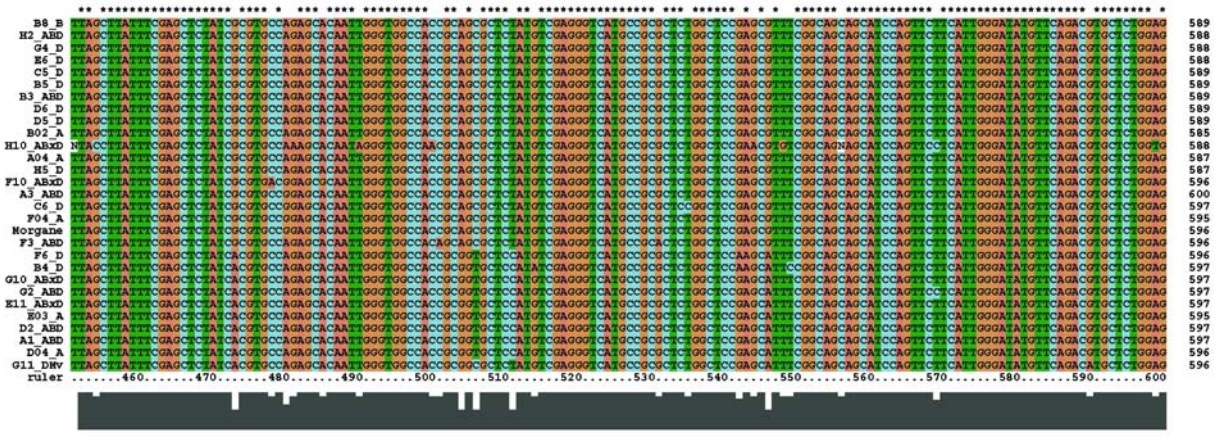

Fig. 4 *ClustalX* alignments of the cloned ''complete'' sequences with the original *Morgane*_CR626934-1 sequence. The clones are labelled with a letter and a number and the genome from which they originate (i.e. A3_ABD is the clone A3, and was cloned on the ABD genome)

are transcribed in bread wheat, at least during biotic as well as abiotic stresses, as previously described for other Class I LTR retrotransposons (Mhiri et al. 1997; Takeda et al. 1998; Melayah et al. 2001): the main detected homologous sequences to *Morgane*_CR626934-1 are in *EST* databases. The two examples detailed here (*AZO3104L08* and *wlm1.pk0008.a8* sequences) are surely expressed from the promoter within the LTR, and do not come from a hybrid transcript derived from an external *Pol*II promoter. Actually, even if the *AZO3104L08* sequence is partial, it starts in the homologous potential R region of the 5′ LTR of *Morgane*_CR626934-1, and continues from the LTR until the internal region, homologous to the *Pol*-like sequence from *Morgane*_CR626934-1 (but lacks most of it, Fig. 2), as a standard LTR-driven expression. Thus, either this *EST* is a shorter evolution of the *Morgane*_CR626934-1, or *Morgane*_CR626934-1 is a longer evolution of this *EST*. As the *Pol*-like region of *Morgane*_CR626934-1 is larger than the one from the *EST*, we could reasonably suppose that the *EST* is the shorter form, probably created through the same mechanism, which could lead to *Morgane*_CR626934-1 from longer active elements. The reverse complement *wlm1.pk0008.a8 EST* is also homologous to the LTRs from *Morgane*_CR626934-1, directly after its poly-A tail, on the U3-R part of the LTR (standard location in a LTR-driven expression). It should be noticed that the non-homologous remaining part to *Morgane*_CR626934-1 in this *EST* is whatever homologous to numerous *gypsy Pol*-like sequences from various grasses. Thus, this sequence is probably the end of a complete transcript from an element which is close to *Morgane*_CR626934-1, and which is probably as small as *Morgane*_CR626934-1 is.

The *Morgane*_CR626934-1 itself is unusually located in the third intron of the *CR626934.3* hypothetical gene. This location is quite remarkable for a LTR retrotransposon, as in *Triticeae* they are generally nested in long stretch of TEs outside of gene islands (Sabot et al. 2005a), either because of a specific insertion preference outside of these islands or a specific elimination from the gene islands. Anyhow, no *EST* related to a chimerical transcript between the element and the *CR626964.3* hypothetical gene has been reported yet in the databases. The small size of this element, most presumably, allows the correct splicing of this intron and should not disrupt the activity of this gene. So, here we probably observed an exceptional case of intronic insertion of a *Morgane* element.

Sequence homologies in genomic databases (*BLASTx* and *tBLASTx*) of all *Morgane*-like elements would categorize them as *gypsy* elements (50% positive amino acid), but *gypsy* are generally rather long elements (10–12 kb), with a coding internal sequence (mutated or not) related to *Gag* and *Pol* proteins (Kumar and Bennetzen 1999; Sabot et al. 2004). *Morgane* elements are short (1.8 kb), with numerous stop-codons in their internal homologous *Pol*-like sequence. Bioinformatic analyses did not allow us to identify any putative functional protein, which could *cis*-mobilize the elements from the *Morgane* family, and moreover no *Gag* ORF homology was detected. As it was previously suggested for ''classical'' non-autonomous LTR retrotransposons (*LARDs* & *TRIMs*, for example), this element should be *trans*-mobilized by another LTR retrotransposon protein complex (Kalendar et al. 2004). *LARDs* elements are also supposed to have derived from *gypsy* elements, but they are long (12 kb at least), with very long LTR (~4.5 kb), have a non-coding internal sequence with potential short ORFs and numerous stop codons and a conserved secondary RNA structure (Kalendar et al. 2004). *TRIMs* are short LTR retrotransposons (about 500 bp maximum), but with internal sequences non-related to *Pol*-like sequences (Witte et al. 2001). These two groups of LTR retrotransposons are non-autonomous (i.e. they do not

encode active *Gag/Pol* sequences), and it is possible that *TRIMs* are derived from *LARDs*, which themselves might originate from *gypsy* elements. The *LARDs* to *TRIMs* transition could have been performed by internal reductive deletion, in the internal sequence as in the LTR sequences. The *Morgane*-like elements described here are longer than the already described *TRIMs*, and possess a potential short ORF related to *Pol*-like sequences, but which is smaller than the ORFs found in *LARDs*. Thus, the *Morgane*-like elements may be an intermediate form between *LARDs* and *TRIMs*, or directly between *gypsy* and *TRIMs*.

Morgane elements seem to be the first family of such small *Morgane*-like elements, which are different from *TRIMs* because of their homologies with polyprotein sequences (even with stop-codon within), also different from active LTR elements (such as *gypsy*, *copia* and *athila*) because of their small size and obvious lack of autonomy and of *Gag* ORF, and finally different from *LARDs* because of their really much smaller size. Moreover, we did not yet identify any *trans*-activator of *Morgane* elements, i.e. the active element able to mobilize *Morgane*. Such an element would probably belong to *gypsy* LTR retrotransposon group.

We thus suggest that there are *Morgane* elements (from the same family of *Morgane*_CR626934-1, with the example of the *AZO3104L08 EST*) and *Morgane*-like elements (with the same structure and specificity, as for the *wlm1.pk0008.a8 EST*). Besides, we think that such *Morgane*-like elements will be found in other plant genomes in in future analyse.

So, we propose to temporarily classify all these *Morgane* and *Morgane*-like elements in a new LTR retrotransposons group, the *Morgane* group, until we can identify either their *trans*-activator or their origin.

Morgane_CR626934-1 subfamily spreading

Elements from the *Morgane*_CR626934-1 subfamily were only detected in wheats and very closely related species such as *A. elongatum*. Therefore, this specific subfamily might have ‘‘appeared’’ after the divergence between barley, rye and wheat, but before the radiation of *Agro-*

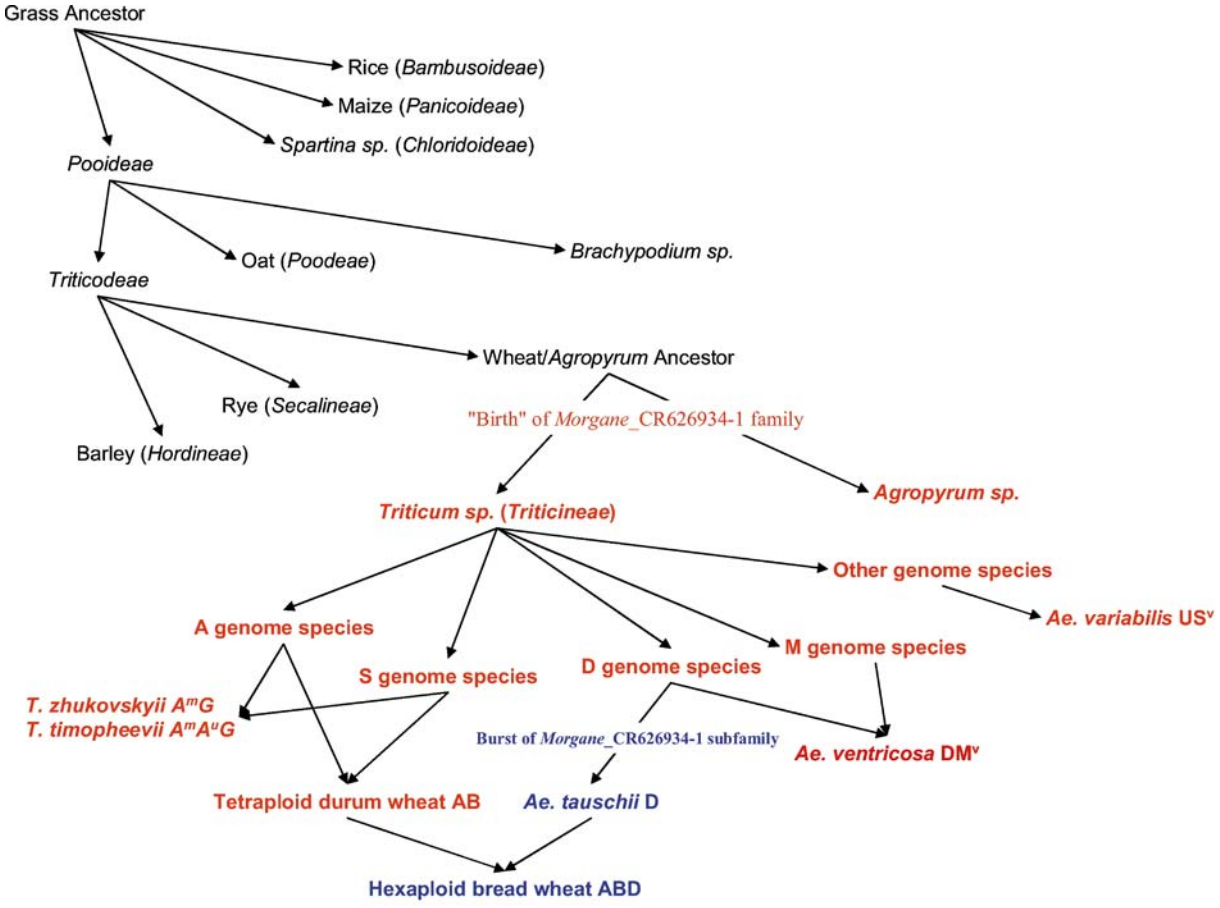

Fig. 5 Model for the spreading of the *Morgane*_CR626934-1 subfamily during the *Triticineae* and grass genomes evolution. Species with a low copy number are in red, and species with a high copy number are in blue

pyrum sp. and wheat, around 6 MYA (Fig. 5), Wicker et al., 2001. Moreover, this subfamily was subjected to differential amplification between closely related genomes of wheats (Fig. 3). In the ''non D'' diploid species, as well as *Ae. ventricosa*, *Ae. variabilis* and ''Russian'' wheat species, its copy number is lower than in the AA BB × DD, AA BB DD and *Ae. tauschii* DD genomes. The DD genome from *Ae. tauschii* may have been subjected to specific amplification of this specific subfamily after radiation of the different diploid wheat species 4–5 MYA (Fig. 5). As a matter of fact, the sequence variations between different *Morgane*_CR626934-1-like elements from different species are minor, which support the hypothesis of a recent amplification. Moreover, *Morgane* as *Morgane*-like elements are transcribed in bread wheat, at least under biotic and abiotic stress conditions, which is a characteristic of active LTR promoters (reviewed in Kumar and Bennetzen 1999 and Sabot et al. 2004). Then, *Morgane*_CR626934-1 subfamily, and probably all elements from this group, are recent and probably still active elements, with the *Morgane*_CR626934-1 subfamily mainly restricted to *Triticineae* genomes.

The clones from internal part belonging to this subfamily are highly conserved at the nucleic level (95.5%), without notable differences between genomes even 4–5 millions years after radiation (Levy and Feldman 2002). Such conservation level was also observed with *TRIMs*, which show a high level (60–75%) of nucleic acid conservation between non-related species such as rice and *Arabidopsis*, even after 60 millions years of divergence (Witte et al. 2001). This high level of conservation could reflect a biological role for *TRIMs*. In the same way, the end of the internal sequences of *LARDs* is well-conserved, and also possesses a conserved secondary RNA structure, suggesting that it is therefore under selective pressure for nucleotide conservation for a putative biological role of the sequence itself (chromatin regulation or else) (Kalendar et al. 2004). Thus, the sequence conservation observed here may indicate that these elements also play an important role in the evolution of the *Triticeae*.

As an additional ''benefit'' to the *Morgane*_CR626934-1 subfamily analyses, we found that the DD genome from *Ae. tauschii* may be more different from the *Ae. ventricosa* DD subgenome than expected. The amplification burst of the *Morgane*_CR626934-1 subfamily elements seems to have occurred only in the DD genome of *Ae. tauschii* clad, after the divergence between *Ae. tauschii* and *Ae. ventricosa* diploid ancestor (Fig. 5). This transposition burst could have arised following either an activating stress or the formation of a ''Master Copy'' of *Morgane*_CR626934-1 subfamily in the genome of *Ae. tauschii*. Such ''Master Copies'' are under the control of a strong constitutive promoter (*Pol*II), and are able to transcribe their RNA faster than the other copies (reviewed in Sabot et al. 2004). Such a difference in TE content and amplification may lead to another genomic structure, and thus possibly to another wheat species, as reviewed in Bennetzen et al. (2005).

Here we described the structure of a potentially new type of Class I element and the specific distribution of one of its subfamilies. The structure of these elements is quite different from those belonging to the already identified LTR retrotransposons groups. Moreover, at least one subfamily of this group (The *Morgane*_CR626934-1 one) has been subjected to specific amplifications within closely related wheat genomes and/or between tribes of grass genomes. Further studies and extended analyses will allow us to clarify the origin of the *Morgane* group and the complete mechanism of its transposition.

Acknowledgements Authors wish to thank Vanessa Ferraria for her technical assistance, Catherine Feuillet for her constructive corrections, Bastien Laubin for his help in the bioinformatics verifications, and the editor and reviewers for their useful recommendations. This work was supported by a grant of the French *Institut National de la Recherche Agronomique* and the *Région Auvergne*.

References

Bendich AJ, McCarthy BJ (1970) DNA comparisons among barley, oats, rye, and wheat. Genetics 65:545–565

Bennett MD, Leitch IJ (1995) Nuclear DNA amounts in angiosperms. Ann Bot 76:113–176

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

Chantret N, Cenci A, Sabot F, Anderson O, Dubcovsky J (2004) Sequencing of the *Triticum monococcum Hardness* locus reveals good microcolinearity with rice. Mol Genet and Genomics 271:377–386

Chantret N, Salse J, Sabot F, Rahman S, Bellec A, Laubin B, Dubois I, Sourdille P, Joudrier P, Gautier M-F, Cattolico L, Beckert M, Aubourg S, Weissenbach J, Caboche M, Bernard M, Leroy P, Chalhoub B (2005) Molecular basis of evolutionary events that shaped the *Hardness*, *Ha*. locus in diploid and polyploidy wheat species, *Triticum* and *Aegilops*. Plant Cell 17:1033–1045

Flavell RB, Rimpau J, Smith DB (1977) Repeated sequence DNA relationship in four cereals genomes. Chromosoma 63:205–222

Higgins D, Thompson J, Gibson TJ (1994) CLUSTALW: 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

Huang X, Miller W (1991) A Time-efficient, linear-space local similarity algorithm. Adv Appl Math 12:337–357

Jurka J (1998) Repeats in genomic DNA: mining and meaning. Curr Opin Struct Biol 8:333–337

Kalendar R, Vicient CM, Peleg O, Anamthawat-Jonsson K, Bolshoy A, Schulman AH (2004) *LArge Retrotransposon Derivatives*: Abundant, conserved but non-autonomous retroelements of Barley and related genomes. Genetics 166:1437–1450

Keller B, Feuillet C (2000) Colinearity and gene density in grass genomes. Trends Plant Sci 5:246–251

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

Levy AA, Feldman M (2002) The impact of polyploidy on grass genome evolution. Plant Physiol 130:1587–1593

McCarthy EM, McDonald JF (2003) LTR_STRUC: a novel search and identification program for LTR retrotransposons. Bioinformatics 19:362–367

Melayah D, Bonnivard E, Chalhoub B, Audeon C, Grandbastien M-A (2001) The mobility of the tobacco *Tnt*1 retrotransposon correlates with its transcriptional activation by fungal factors. Plant J 28:159–168

Mhiri C, Morel JB, Vernhettes S, Casacuberta JM, Lucas H, Grandbastien M-A (1997) The promoter of the tobacco *Tnt1* retrotransposon is induced by wounding and by abiotic stress. Plant Mol Biol 33:257–266

Murigneux A, Barloy D, Leroy P, Beckert M (1993) Molecular and morphological evaluation of doubled haploid lines in maize. 1 – Homogeneity within DH lines. Theor Appl Genet 86:837–842

Sabot F., Guyot R., Wicker T., Chantret N., Salse J., Laubin B., Leroy P., Sourdille P, Chalhoub B, Bernard M (2005a) Updating transposable elements annotations from large wheat genomic sequences reveals diverse activities and gene association of elements. Mol Genet Genomics 274:119–132

Sabot F, Sourdille P, Bernard M (2005b) Advent of a new retrotransposon structure: the long form of the *Veju* elements. Genetica 125:325–335

Sabot F, Simon D, Bernard M (2004) Plant transposable elements, with an emphasis on grass species. Euphytica 137:227–247

Sambrook J, Russel A (2001) Molecular cloning a laboratory manual. Cold Spring Harbor: Laboratory Press, Cold Spring Harbor, New York

Takeda S, Sugimoto K, Otsuki H, Hirochika H (1998) Transcriptional activation of the tobacco retrotransposon *Tto1* by wounding and methyljasmonate. Plant Mol Biol 36: 365–376

Wicker T, Stein N, Albar L, Feuillet C, Schlagenhauf E, Keller B (2001) Analysis of a 211 kb sequence in diploid wheat, *Triticum monococcum* L. reveals multiple mechanisms of genome evolution. Plant J 26: 307–316

Wicker T, Matthews DE, Keller B (2002) TREP: a database for *Triticeae* repetitive elements. Trends Plant Sci 7:561–562

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