Author Accepted Manuscript
"The final publication is available at link.springer.com"
Genetica (2020) 148:109-123
https://doi.org/10.1007/s10709-020-00094-3
Published online: 24 April 2020

Screening diversity and distribution of *Copia* retrotransposons reveals a specific amplification of *BARE1* elements in genomes of the polyploid *Hordeum murinum* complex

Malika Ourari[1], Olivier Coriton[2], Guillaume Martin[3,7], Virginie Huteau[2], Jean Keller[4], Malika-Lily Ainouche[5], Rachid Amirouche[6*], Abdelkader Ainouche[5]

[1]Laboratory of Ecology and Environment, Department of Environment Biological Sciences, Faculty of Nature and Life Sciences, Université de Bejaia, Targa Ouzemmour, 06000 Bejaia, Algeria.
[2] Institut National de Recherche en Agriculture, Alimentation et Environnement, UMR1349 INRAE-AgroCampus Ouest-Université de Rennes 1, Bât 301, INRA Centre de Bretagne-Normandie, BP 35327, 35653 Le Rheu Cedex, France.
[3] CIRAD, UMR AGAP, F-34398 Montpellier, France.
[4] LRSV, Université de Toulouse, CNRS, UPS, 24 chemin de Borde Rouge, Auzeville, 31320 Auzeville-Tolosane, France.
[5] Université de Rennes 1, UMR-CNRS 6553, EcoBio, Campus Scientifique de Beaulieu, Bât. 14A, 35042 Rennes Cedex, France.
[6*] University of Sciences and Technology Houari Boumediene, Faculty of Biological Sciences, Lab. LBPO, USTHB, BP 32 El-Alia, Bab-Ezzouar, 16111, Algiers, Algeria
[7] AGAP, Univ Montpellier, CIRAD, INRA, Montpellier SupAgro, Montpellier, France.

Corresponding author
Rachid Amirouche : ramirouche@usthb.dz

Abstract

We explored diversity, distribution and evolutionary dynamics of Ty1-*Copia* retrotransposons in the genomes of the *Hordeum murinum* polyploid complex and related taxa. Phylogenetic and fluorescent *in situ* hybridization (FISH) analyses of *reverse transcriptase* sequences identified four *Copia* families in these genomes: the predominant *BARE1* (including three groups or subfamilies, *A*, *B* and *C*), and the less represented *RIRE1*, *IKYA* and *TAR-1*. Within the *BARE1* family, *BARE1-A* elements and a subgroup of *BARE1-B* elements (named *B1*) have proliferated in the allopolyploid members of the *H. murinum* complex (*H. murinum* and *H. leporinum*), and in their extant diploid progenitor, subsp. *glaucum*. Moreover, we found a specific amplification of *BARE1-B* elements within each *Hordeum* species surveyed. The low occurrence of *RIRE1*, *IKYA* and *TAR-1* elements in the allopolyploid cytotypes suggests that they are either weakly represented or highly degenerated in their diploid progenitors. The results demonstrate that *BARE1-A* and *BARE1-B1 Copia* elements are particularly well represented in the genomes of the *H. murinum* complex and constitute its genomic hallmark. No *BARE1-A* and *-B1* homologs were detected in the reference barley genome. The similar distribution of *RT-Copia* probes across chromosomes of diploid, tetraploid and hexaploid taxa of the *murinum* complex shows no evidence of proliferation following polyploidization.

Keywords *Hordeum*, transposable elements, phylogeny, *in situ* hybridization, polyploidy

Introduction

Transposable elements represent an important and dynamic component of plant genomes (Kumar and Bennetzen 1999; Bennetzen et al. 2005). The most important group of transposable elements (TE) is composed of Long Terminal Repeat (*LTR*) retrotransposons, which includes two super families named Ty1-*Copia* and Ty3-*Gypsy*. These *LTR* retrotransposons are able to copy themselves through a replicative transposition process (*via* a RNA intermediate) and then are able to move to new chromosomal locations in the genome (Kumar and Bennetzen 1999). This mode of replication gives retrotransposons the potential to become extremely abundant in plant genomes over time, which increases the size of the host genome. In vascular plants, they can constitute more than half of the repetitive DNA and are generally widely distributed and dispersed in the genomes (Schnable et al. 2009). They are dynamic genome components with the ability to integrate new copies and facilitate intra-chromosomal recombination, and thus can cause large chromosomal rearrangements (Belyayev et al. 2010). Their insertion in close proximity to or into genes may greatly affect gene expression and function via gene disruption and/or TE-derived regulatory sequences capture (Chénais et al. 2012; and references therein). Hence, retrotransposons are an important source of genetic and phenotypic diversity in plants (Kumar and Bennetzen 1999; Mansour 2007; Lisch 2013). Transposition may be activated by various kinds of stress, including genomic shocks (e.g., hybridization, allopolyploidization) and or environmental stimuli (e.g., biotic and abiotic stress) (Granbastien 2004; Kalendar et al. 2000; Liu and Wendel 2003; Parisod et al. 2010). However, this process of transposition is controlled by different and complementary molecular and cellular mechanisms, such as: DNA methylation; miRNA and siRNA; sequence recombination and deletion. This leads to the repression of the TE proliferation and more or less rapidly leads to DNA loss and genome downsizing (Bennetzen et al. 2005; Ma et al. 2004; Ma and Bennetzen 2006; Hawkins et al. 2006, 2009; Yaakov and Kashkush 2011; Lisch and Bennetzen 2011; Chénais et al. 2012).

As a result of these various processes acting repeatedly on TEs dynamics (amplification/repression/recombination/removal) during evolution of species, plant genomes may contain various kinds of TE sequences. Therefore, TEs are represented by full length or truncated elements, likely resulting from very recent amplification, and/or by more or less degenerated but still identifiable TE fragments of various lengths, representing the witnesses of earlier amplifications that occurred in the last 5-10 million years (Devos et al. 2002; Hawkins et al. 2009; Estep et al. 2013). Otherwise, it is generally difficult even impossible to recognize highly degenerated fragments resulting from much older amplification events.

Additionally, hybridization and polyploidy have considerably complexified the impact of these processes over the evolutionary time scales (Ozkan et al. 2001; Parisod et al. 2009). Exploring the diversity and distribution of TE sequences in polyploid and their putative diploid progenitors may provide interesting insights into their evolutionary dynamics in polyploid complexes.

Hordeum L. (Poaceae) is notorious for its large genome and its high TE content. The cultivated barley, *H. vulgare* (diploid with a haploid genome size of C = 5.1 Gb) genome was found containing approximately 84% of repetitive DNA (Mayer et al. 2012), mainly composed of long terminal repeat (LTR) retrotransposons (~76%) belonging to the two LTR-superfamilies, Ty3-*Gypsy* and Ty1-*Copia* elements (Suoniemi et al. 1996; Vicient et al. 1999; Mayer et al. 2012). Among the latter, the *BARE1 Copia* family, which was the first LTR retrotransposon described in barley (Manninen and Schulman 1993), is represented by around 80,000 to 120,000 copies, accounting for 12% of the genome (Wicker et al. 2008, 2009). Examination of *BARE1* copy number, conservation and activity in various *Hordeum* species demonstrated their overall importance to genome evolution in the genus (Vicient et al. 1999). Other known transposable elements and novel repetitive sequences were detected in barley using high throughput sequencing data (Wicker et al. 2009). Polyploidy occurred independently several times in this genus, which offers opportunities to investigate the evolution of TEs throughout closely related members in polyploid groups. In this respect, the *H. murinum* L. polyploid complex, native from the Mediterranean region, appeared as an interesting group to explore (Vicient et al. 1999; Ourari et al. 2011).

The *Hordeum murinum* complex is a monophyletic group composed of one diploid subspecies (subsp. *glaucum*, 2*n* = 2*x* = 14), two tetraploid subspecies (subsp. *leporinum*, 2*n* = 4*x* = 28; subsp. *murinum*, 2*n* = 4*x* = 28) and one hexaploid subspecies (subsp. *leporinum*, 2*n* = 6*x* = 42) (Love and Love 1948; Covas 1949). Previous studies suggested an allotetraploid origin for subsp. *leporinum* and subsp. *murinum* involving the diploid subsp. *glaucum* (referred to as genome A) as one of their progenitors (the other being unknown or extinct, referred to as genome B). The subspecies *leporinum* (6*x*) would have an allohexaploid origin involving genomes A, B and another unidentified diploid parent (referred to as genome C) (Ourari et al. 2011; Brassac and Blattner 2015). The 2C genome size in the "*murinum* complex" ranged from ~8 pg in the extant diploid progenitor (*H. murinum* subsp. *glaucum*) to 17-20 pg in the tetraploid subspecies (*leporinum* and *murinum*), and reaches 26 to 30 pg in the hexaploid subsp. *leporinum*, according to Jakob and Blattner (2010) and Ourari et al. (2011).

In this study, we focus our investigations on the pattern of diversity and distribution of the *Copia*-like LTR-retrotransposons in the *H. murinum* complex, in order to shed light on the origin of the most remarkable elements detected in the tetra- and hexaploid genomes. Special attention is directed to evaluating the *BARE1* family, the best known and most abundant transposable elements in *Hordeum* (Suoniemi et al. 1996; Vicient et al. 1999; Wicker et al. 2009), which are currently classified within the large *Angela* group, one of the major evolutionary lineages of *Copia* elements in angiosperms (Wicker and Keller 2007; Neumann et al. 2019). For this purpose, we used both phylogenetic analyses and genomic fluorescent *in situ* hybridization (FISH) based on *Copia reverse transcriptase* conserved domains.

Materials and Methods

Plant material and DNA extraction

Two samples of the diploid subsp. *glaucum* originating from Algeria (M4) and Turkey (gT, USDA collection, plant ID: 402) were analysed. Tetraploid samples of subsp. *leporinum* and subsp. *murinum* were collected in Algeria (Tikjda) and France (Brittany), respectively. The hexaploid cytotype analysed was collected in Afghanistan (USDA collection, plant ID: 12910). Representatives of related diploid *Hordeum* species, *Hordeum marinum*, *H. spontaneum* and *H. bulbosum*, were sampled in Algeria (Guelt Stel, Sétif and Adekar, respectively).

Genomic DNA was extracted from 100 mg fresh young leaves following the CTAB method adapted from Doyle and Doyle (1987).

Amplification, cloning and sequencing of *reverse transcriptase* fragments

Reverse transcriptase (*RT*) sequences of Ty1-*Copia* elements were amplified using degenerate primers defined by Hirochika and Hirochika (1993) in rice:

Ty1H1 [5'-CA(A/G)ATGGA(C/T)GT(ACG/T)AA(A/G)AC-3'] and
Ty1H2 [5'- CAT(A/G)TC(A/G)TC(ACG/T)AC(A/G)TA-3'].

PCR conditions were as follows: 3 min (94°C), [50 sec (94°C), 50 sec (39°C), 1 min (72°C)] × 35, 7 min (72°C). The PCR products were checked on agarose gel electrophoresis and were purified with a Macherey-Nagel Nucleospin kit. The purified PCR products from all the diploid and polyploid samples were cloned using pGEM-T vectors and JM109 bacterial strains of *Escherichia coli* from Promega, according to the manufacturer's procedures. Sequencing was performed by Eurofins MWG company (Germany) using T7 primer. Only clones recognisable as *RT-Copia* sequences by comparison to reference databases (TREPbase, http://wheat.pw.usda.gov/ITMI/Repeats/; and Repbase, Jurka et al. 2005) having a size larger than 230 bp were retained for subsequent analyses; clones of lower size represent more degenerated fragments that are unusable in phylogenetic analyses. Sequence data have been deposited in GenBank (http://www.ncbi.nlm.nih.gov/) under accession numbers JN135335–JN135475.

In order to evaluate and identify the diversity of *Copia* elements in the *murinum* complex, two *RT* data matrices were constructed. One including 141 *RT* clones, generated from the *murinum* complex

samples and related taxa (*H. marinum*, *H. bulbosum*, and *H. spontaneum*) and the other containing all the above *RT* clones aligned with annotated *RT* sequences downloaded from reference databases (TREP and Repbase). DNA sequences alignment was performed using the MUSCLE program (Edgar 2004). Pairwise sequence distances (with gaped sites ignored) were estimated using uncorrected "p" with MEGA-X (Kumar et al. 2018). The two matrices were subjected to Maximum Likelihood phylogenetic analyses. The best-fitted model of sequence evolution for each sequence dataset was determined using JModeltest 62 and Maximum likelihood analyses were then performed with 1000 replicates of bootstrap (Felsenstein 1985) using MEGA X. Each group (or clade) of *RT*-domain sequences was named based on its closest phylogenetic relationship to the *RT*-references, following the *Copia* annotation used for barley and Triticeae (Suoniemi et al. 1996; Vicient et al. 1999; Wicker and Keller 2007; Wicker et al. 2009). Moreover, in order to supplement this annotation in the light of the wider and unified classification system of plant LTR-retrotransposons recently published, all *RT* sequences were also examined against a reference database of transposable element protein domains (REXdb Viridiplantae 3.0; Neumann et al. 2019), using the BLASTx+ v2.9.0 (Camacho et al. 2009) with default parameter and setting the maximum of hits per query to 10.

Retrieving the *RTs of BARE1* and *RIRE* elements in the barley genome

Additionally, we take advantage of the recent release of the barley genome (*Hordeum vulgare* genome 462R1 v1; Mascher et al. 2017; Beier et al. 2017), available in the Phytozome database (https://phytozome.jgi.doe.gov/pz/portal.html#), to verify whether some remarkable *RT* groups observed in our phylogenetic analyses have been specifically amplified in the *H. murinum* complex (see Results). Thus, all *BARE1* (*A* and *B1*) and *RIRE RT* sequences generated in our study were blasted against the barley genome of reference, using the BLASTn+ v2.6.3 algorithm. About 125,000 sequences were identified through a BLAST analysis and were clustered with CD-HIT-EST v4.6.8 (Li and Godzik 2006; Fu et al. 2012) and the following parameters: 90% of identity and a word size of 5. For each of the six obtained cluster, only 19,967 sequences having a minimum length of 270 bp that are aligned with our *RT* sequences have been retained. Finally, few representatives of these clusters of reference were combined in a same dataset with our *BARE1* (*A* and *B1*) and *RIRE RT* sequences for phylogenetic analysis. Sequences were aligned using MAFFT as implemented in Geneious v6.1.8 (Katoh et al. 2002; Kearse et al. 2012) and then subjected to Maximum Likelihood analysis using IQ-TREE v1.5 (Nguyen et al. 2015). The best-fitted evolution model was determined with ModelFinder (Kalyaanamoorthy et al. 2017). Phylogenetic analysis was performed with 10,000 SH-alrt and Ultrafast bootstraps replicates (Guindon et al. 2010; Hoang et al. 2017).

Fluorescent *in situ* hybridization

Probes derived from *RT Copia* sequences amplified by degenerate primers (Hirochika and Hirochika 1993) from the diploid subspecies *glaucum* (A genome) is used in fluorescent *in situ* hybridization technique (FISH) in order to investigate their chromosomal distribution in the different *murinum* complex cytotypes.

Paralleling the *RT*-FISH experiments, we have also carried out FISH on the same samples with 45S rDNA probe (pTa 71) to serve as control (Ourari et al. 2011). This sequence is a cloned 9-kb EcoRI fragment of a rDNA repeat unit (18S–5.8S–26S genes and spacers) isolated from *Triticum aestivum* (Gerlach and Bedbrook 1979). The probe was labeled by random priming with Alexa 594 Kit (Invitrogen Life Technology).

In situ hybridization was performed on mitotic chromosomes from roots of seedlings (~1 cm) obtained from the diploid and polyploid samples analysed in the *H. murinum* complex. The root tips were treated in α-bromonaphtalene solution at 4°C for 24 h and fixed with 3/1 (v/v) ethanol/acetic-acid. Root tips were stained in 2% aceto-carmine solution and squashed in a drop of 45% acetic acid. Labelled *RT* sequences were obtained from targeted genomes by PCR amplification with the PCR DIG Probe Synthesis Kit, according to manufacturer's instructions (Roche, Basel, Switzerland). A hybridization mixture was prepared with 25 µl of 100% formamide, 10 µl of 50% dextran sulfate, 5 µl of 20 × SSC, 0.6 µl of SDS (0.2 g/ml), 10 µg of denatured salmon sperm DNA (100 mg/ml),

approximately 100 ng of labelled *RT* sequences, 5 µl of labelled 45S rDNA (from Alexa^{594}, Kit Invitrogen Life Technology), and ultrapure water for a final volume of 50 µl. Before hybridization, slides were washed in 2 × SSC and incubated in RNase at 37°C for 1 h. Slides were then rinsed out twice for 3 min in 2 × SSC at 42°C, treated with 100 ml of pepsine at 100 µg/ml during 10 min at 37°C, rinsed out for 3 min in 2 × SSC at 42°C (step repeated twice), and plunged in a solution of paraformaldhyde (0.05 g/ml) and NaOH (0.01 M) during 10 min. Slides were rinsed out twice for 3 min in 2 × SSC at 42°C, plunged 10 min in 50% formamide and dehydrated in a graded ethanol series (70%, 90%, 100%) for 3 min each in ice-cold. Slides were air-dried at room temperature during 45 min. After a pre-denaturation step (6 min at 92°C), 50 µl of the hybridization mixture was applied to the slides. Chromosome preparations and pre-denatured probes were denatured at 85°C for 10 min and allowed to hybridize overnight at 37°C in a moist chamber. Slides were treated with 200 ml of 5% Bovine Serum Albumine in 4 × SSC Tween 20 and incubated for 5 min at room temperature. To detect the hybridization signal, 50 µl of a mixture containing 7 µl of Anti-Dig-FITC at 200 µg/ml (Roche) and 63 µl of 5% BSA were added to the slides. After 1 h at 37°C, the slides were washed 3 times in 4 × SSC for 5 min each at 37°C. Numerous cells were examined for each hybridization and two independent hybridizations were made for each taxon.

The chromosomes were mounted and counterstained in Vectashield (Vector Laboratories) containing 2.5 µg/mL 4', 6-diamidino-2-phenylindole (DAPI). Fluorescence images were captured using a CoolSnap HQ camera (Photometrics, Tucson, Arizona) on an Axioplan 2 microscope (Zeiss, Oberkochen, Germany) and analyzed using MetaVueTM (Universal Imaging Corporation, Downington, PA).

Results

RT-Copia diversity in the *H. murinum* complex

A total of 141 *RT* sequences of Ty1-*Copia* elements were obtained from all the taxa and were aligned in a data matrix. Ninety-two clones were generated from the *H. murinum* complex, namely 27 clones from subsp. *glaucum*, 23 from the tetraploid subsp. *leporinum*, 14 from subsp. *murinum* and 28 from the hexaploid subsp. *leporinum*. Moreover, 49 clones were generated to represent *Copia* elements of the closest relatives to the *H. murinum* complex, 26 from *H. marinum*, 11 from *H. spontaneum* and 12 from *H. bulbosum*. The length of this aligned matrix was 297 bp. Individual *RT Copia* fragments ranged from 232 to 279 bp in size. Uncorrected pairwise divergence (p) among the analyzed sequences varied from 0.0 to 0.466.

As our study is only focused on closely related taxa with highly similar elements, and to allow easier comparison with previously described *Copia* families, the annotation of the *RT*-domain sequences was essentially based on the classification used for barley and Triticeae (Suoniemi et al. 1996; Vicient et al. 1999; Wicker and Keller 2007; Wicker et al. 2009). A preliminary phylogenetic analysis was performed on the aligned data matrix which included the 141 *RT* sequences generated from this study and annotated *RT* sequences representing 32 *Copia* families and subfamilies obtained from TREP and Repbase, in order to identify the *Copia* diversity in the '*murinum*' complex. A simplified ML tree (using the T92+G model; Tamura 1992) is presented in Fig. 1, which includes only one *RT* reference to represent each of the *Copia* families and subfamilies detected in our sampling. All cloned sequences fell within four main clusters for which were detected homologs corresponding to the following *Copia* families: *BARE1* (including its close relatives *Angela* and *WIS* elements; not shown here), *RIRE1*, *IKYA* (incl. its close relative *IDA* elements; not shown here) and *TAR-1* (incl. its close relative *SASANDRA* elements; not shown here).

Most of the sequences obtained from the *murinum* complex (78 clones; 84.8%) clearly belong to *BARE1*, the predominant family described in barley, which is also present in wheat under *WIS* and *Angela* names (Wicker and Keller 2007; Wicker et al. 2009). Pairwise divergence (p) among the sequences in the *BARE1* cluster varied from 0.0 to 0.305. Within this family, the '*murinum'* sequences were distributed in three main groups or subfamilies (*A*, *B*, *C*) (Fig. 2). *BARE1-A*, *B* and *C* groups are identified by RLC_BARE1_A_consensus1, RLC_BARE1_B_consensus1 and RLC_BARE1_C_consensus1 obtained from reference databases TREP and Repbase, respectively.

Five *RT* clones from *H. murinum* subsp. *murinum* (4*x*) and seven from *H. murinum* subsp. *leporinum* (6*x*) showed a high homology with *RIRE1* elements, a family initially identified in *Oryza australiensis* (Noma et al. 1997). A single clone from the hexaploid subspecies *leporinum* is related to family *TAR-1* (identified in *Triticum monoccocum* and homolog to SASANDRA elements from *Triticum aestivum*), and a single one from the tetraploid subspecies *leporinum* is related to *IKYA* (identified in *H. vulgare*) which is homolog to *IDA* elements from *Triticum turgidum*. No sequence analyzed in our samples has been related to other much less abundant *Copia* families (such as *MAXIMUS*, *INGA* or *CLAUDIA*) reported in barley (Wicker et al. 2009).

Therefore, four *Copia* families have been identified in genomes of the *H. murinum* complex and closely related taxa, *BARE1* (including three groups, provisionally treated here as subfamilies *A*, *B* and *C*), *RIRE1*, *IKYA* and *TAR-1,* which are indicated in Fig. 1 and reported in Table 1 (with reference to their homolog families in Triticeae). While adopting this specific classification throughout this study, comparison of our data against the recently published REXdb database, allowed to best situate these families among the main evolutionary lineages of *Copia* elements circumscribed over a wide range of taxa in the unified classification system of plant LTR-retrotransposons (Neumann et al. 2019). Accordingly, *BARE1* and *RIRE1* elements (and their homologs *WIS* and *Angela*) are supported as members of a same major lineage called "*Angela*", *IKYA* (incl. its homolog *IDA*) is related to the "*Ikeros*" lineage, whereas *TAR-1* is assigned to a lineage with the same name "*TAR*" (Table 1; Fig. 1).

A second ML analysis (using the T92+G model; Tamura 1992) was conducted on the data matrix only containing the *RT* sequences generated in this study for the *H. murinum* complex and allied diploid taxa (Fig.2). Estimates of mean pairwise sequence divergences (*i.e*., uncorrected 'p-distance') within and among groups of *RTs* depicted in the tree are given in Table 2. In the large *BARE1* clade (supported with 96% bootstrap), two divergent sequence sets were amplified from the diploid subspecies *glaucum* (2*x*): one belonging to *BARE1-A* subfamily (with 82% bootstrap support and a Mean Sequence Divergence MSD = 0.111); and the other representing part of *BARE1-B* subfamily, here after named *BARE1-B1* (with a low bootstrap of 52% and MSD = 0.098). The mean divergence between these two *RT* groups (*A vs. B1*) is p = 0.205. In addition, *BARE1-B* contained more or less well supported but distinct clusters of sequences that are species-specific to *H. spontaneum* (92% of bootstrap support and MDS = 0.047), *H. bulbosum* (52% bootstrap; MDS = 0.098), and *H. marinum* (<50% bootstrap; MDS = 0.089), with evolutionary distances among *BARE1-B* subgroups ranging from 0.107 (between the closely related *H. spontaneum* and *H. bulbosum*) to 0.144 (between the *B1* and *H. bulbosum* clusters). These two types of *RT* sequences (*A*, *B1*) are shared by allopolyploids of the *murinum* complex. The *BARE1-C* cluster (94% bootstrap) includes sequences isolated from the polyploid members of the *murinum* complex (subsp. *murinum* 4*x*, subsp. *leporinum* 4*x*, subsp. *leporinum* 6*x*) and others isolated from the related diploids (*marinum*, *spontaneum*, *bulbosum*). None of the sequences cloned from the diploid subsp. *glaucum* was detected in this cluster (*BARE1-C*) in which the mean pairwise sequences distance is p = 0.122. *RT*-sequences related to *RIRE1-I Copia*-type were detected in subspecies *murinum* (4*x*) and *leporinum* (6*x*). Three *RIRE1-I* sequences were also found in *H. marinum*. The sister relationships between the *RIRE* and *BARE1* elements (79%) support their membership as distinct subgroups in the same *Angela* lineages.

The *IKYA* group includes one sequence detected in *subsp. leporinum* 4*x* and sequences from *H. marinum,* whereas some sequences detected in subsp. *leporinum* 6*x* and *H. bulbosum* are identified as representatives of *TAR-1 Copia* elements.

Following the screening of the reference barley genome (*H. vulgare* syn. of *H. spontaneum*), we detected six clusters of *RT* sequences which showed a high sequence identity level with the *BARE1* (*A* and *B1*) and *RIRE RT* sequences generated from the *H. murinum* complex (see methods). A phylogenetic analysis was performed on a dataset including both the later *H. murinum RTs* and representatives of the barley clusters. The results (Fig. 3) show that the overwhelming majority of the barley *RTs* (representing 19,275 of the 19,967 selected) clearly clustered with the *H. spontaneum/H. bulbosum BARE1-B* subclade (in green in Fig. 3). Another group of sequences representing a substantial number of barley *RTs* (611) clustered with their *RIRE* homologs from *H. murinum* polyploids and *H. marinum.* The representatives of the few remaining barley *RT*s are related to divers other *BARE1-B RT*s (from *H. marinum, H. spontaneum, H. bulbosum and H. leporinum*). Therefore, it can be underlined that none of the barley *RT*s felt within the *BARE1-A* subfamily or the *BARE1-B1*

cluster (respectively in red and blue in Fig. 3), which then likely represent *BARE1* elements that evolved and proliferated specifically in the *H. murinum* complex. Also, it is interesting to notice that they neither felt within the specific *H. marinum BARE1-B* cluster (in purple in Fig. 3).

Distribution of *RT* elements on the chromosomes of the *H. murinum* complex

The *Copia* element probe derived from the subspecies *glaucum* (2*x*) was hybridized to metaphase chromosomes of the *murinum* complex subspecies to investigate their distribution in the genomes. The probe was hybridized to chromosomes of subspecies *glaucum* (2*x*) itself (Fig. 4a), subspecies *leporinum* (4*x*) (Fig. 4b), subspecies *murinum* (4*x*) (Fig. 4c) and subspecies *leporinum* (6*x*) (Fig. 4d).

In all surveyed taxa, this type of elements gives strong hybridisation signal that was uniformly distributed over all chromosomes, except in the NOR and centromeric regions. Moreover, the 45S labeling control was positive in each case and revealed: four 45S rDNA loci in the diploid subsp. *glaucum*, six in the two tetraploid taxa subsp. *murinum* and subsp. *leporinum*, and twelve loci in the hexaploid subsp. *leporinum*. No locus number additivity for 45S rDNA loci was shown in the tetraploids (Fig. 4e).

For this probe, similar signal intensity was found in both diploid and polyploids, and no remarkable differences of their distribution patterns was revealed among the 2*x* and 4*x*-6*x* cytotypes nor between the polyploid subgenomes.

Discussion

A fairly complete and regularly updated database, with more than two hundred TE families, is available for repeated elements in Triticeae (TREP, http://wheat.pw.usda.gov/ITMI/Repeats) (Wicker et al. 2008, 2009). The most studied species in genus *Hordeum* is *H. vulgar*e (syn. *spontaneum*; barley) (Vicient et al. 1999; Kalendar et al. 2000; Wicker et al. 2007, 2008), which genome has been recently sequenced and released (*Hv* 462R1 v.1; Mascher et al. 2017; Beier et al. 2017). Currently, no *RT*-*Copia* data are available for the *murinum* complex, which represents a widely distributed group of wild populations of *Hordeum* native from the Euro-Mediterranean region. This study presents the pattern of Ty1-*Copia* retrotransposon families detected in genomes of diploid and polyploid members of the *H. murinum* complex, using an exploratory approach based on PCR-amplified fragments and cloning of their conserved *RT* domain with universal primers (Hirochika and Hirochika 1993; Matsuoka and Tsunewaki 1996; Alix and Heslop-Harrison 2004), Also samples from their closely related diploid species (*H. marinum*, *H. bulbosum*, and *H. spontaneum*) were surveyed for comparison. Diversity and distribution of the *Copia* elements detected in these taxa are summarized in Fig. 5 and hereafter discussed. In this figure, the different *Copia* types are indicated by symbols for each taxon. To give an overview of the underlying evolutionary context of this *Copia* distribution, the taxa analyzed are placed in their phylogenetic framework, redrawn following information generated by Ourari et al. (2011) and Brassac and Blattner (2015). Also, the ploidy level (2*x*, 4*x* and 6*x*) and 2C DNA content (in picograms) are indicated for the *H. murinum* samples analyzed here, according to Ourari et al. (2011).

Diversity of *Copia* elements in *H. murinum*

The use of annotated *RT* sequences from the reference databases TREP and Repbase allowed us to identify four families belonging to the *Copia* superfamily of LTR-retrotransposon elements in the *murinum* complex, according to the specific classification used here for barley and their homologs from Triticeae: *BARE1*, *RIRE1*, *IKYA*, *TAR-1* (Table 1; Fig. 5). Comparison to the recent unified classification system of plant LTR-retrotransposons (REdb; Neumann et al. 2019) showed that these families derived from within a few ancient major *Copia* evolutionary lineages in angiosperms: "*Angela"* (incl. *BARE1*, *RIRE1*), "*Ikeros"* (incl. *IKYA*) and "*TAR"* (incl. *TAR-1*). While "*Angela*" appears as the lineage which gave rise to the most abundant *Copia* family (BARE1) in *Hordeum*, the other lineages are only poorly represented, which indicates an unequal evolutionary dynamics of the ancestral lineages in these taxa. Also, this was observed in rice (a close relative from Oryzeae) where

the most abundant *Copia* elements were generated from *"TAR"* rather than from "*Angela*" (Wicker and Keller 2007).

Previous studies clearly showed that a few TE families have reached enormous copy numbers in a way they completely dominate the repetitive genomic compartment, whereas dozens of other described TE families contribute only a few percent to the total genome (e.g., Haas et al. 2009; Wicker et al. 2009). In barley, among the 14 TE families that contributed to 50.31% of the repeated sequences, the *Copia* element *BARE1* tops the list, contributing more than 12% to the total genome (Vicient et al. 1999; Kalendar et al. 2000; Wicker et al. 2009). Accordingly, our investigation demonstrated that the overwhelming majority of the *RT*s detected in the reference barley genome (*Hv* 462R1) essentially represents *BARE1* elements, in agreement with our exploratory PCR-based results which yielded a majority of *BARE1 RT*s.

Among the much less important *Copia* families previously observed in barley (Wicker et al. 2009), only very few elements of *IKYA* and *TAR* could be identified based on their *Triticum* homologs, in spite of the reduced number of clones. The others, such as *Maximus*, *Inga*, *Claudia* were not detected in the samples surveyed here using the degenerate primers. This could be explained by either their real absence in the genomes, their very low copy number, and/or the sequencing depth. Moreover, the presumably universal primers used in this study might be not enough performing to amplify some families, due to specific sequence divergence of their *RTs* or to their high degeneracy if they are or have been present in the genomes. Further investigation of the repetitive sequences in the *H. murinum* genomes using even a low-depth highthroughput sequencing (NGS) will provide more information and help addressing these uncertainties and questions (Macas et al. 2007; Wicker et al. 2009; Piednoël et al. 2013; Staton and Burke 2015).

Phylogenetic analysis of the *RT* sequences generated from this study showed that *BARE1* C*opia* elements diversified into three main phylogenetic groups more or less well supported, provisionally treated here as subfamilies *A*, *B*, and *C*. Further detailed analyses, including more *Copia* sequences, and using alignments of separate and concatenated protein domain sequences should be helpful to test and adjust this classification at the subfamily level, if needed. Subfamily '*A*' is particularly well represented in the *murinum* complex, which suggests that the polyploids (4*x* and 6*x*) most likely inherited their *BARE1-A* elements at least from their extant diploid progenitor, subsp. *glaucum* (Fig. 5). The subfamily '*B*' shows a greater diversity and is amplified in all the taxa analyzed (including the extra-*murinum* diploids examined here). It is noteworthy that some *BARE1-B* elements specifically cluster according to their taxonomic origin in three subclades: "*murinum* complex", "*marinum*", "*bulbosum*" and "*spontaneum*" (Figs. 2 and 5). This result suggests that following their inheritance from a common ancestor, *BARE1-B* elements separately experienced specific waves of amplification after speciation. There is an increasing number of studies illustrating such waves of amplification subsequent to speciation in other taxonomic groups, as for instance in *Nicotiana*, *Gossypium* or Orobanchaceae (Grandbastien 2004; Hawkins et al. 2006; Piednoël et al., 2013). Indeed, distribution of transposable element families in their host genomes is not necessarily homogeneous. Some elements are restricted to one or two host species, while others may be distributed in several species and genera of the same plant family; which is obvious from this study and indicates that different evolutionary patterns (*i.e*., differential lineage-specific amplification of TEs) and mechanisms (accumulation/loss) have shaped the *Copia* profile during the evolutionary history of each species. Many plant retrotransposons are activated by stress or environmental changes (Granbastien 2004; Parisod et al. 2010). For instance, it has been demonstrated significant *BARE1* copy number differences among individuals in a wild population of barley (*Hordeum spontaneum*), in response to a microgeographical gradient of xericity. High elevation, high exposure to sun-light and dryness were correlated with a highest *BARE1* copy number (Kalendar et al. 2000). The environmental history of the host will therefore play a crucial role in the amplification dynamics.

Besides, as these elements are dispersed and inserted in various points throughout the genomes they represent a useful potential source of insertion polymorphism for genetic diversity and phylogenetic estimates (Kalendar et al. 1999, 2011). A recent study using inter-retroelement amplified polymorphism (IRAP) markers in the *H. murinum* complex from Iran revealed patterns of genetic diversity correlated with taxonomic groups, ploidy levels and geographic origin (Sharifi-Rigi et al. 2014).

The sequences obtained from the *glaucum* diploid genome were distributed in *BARE1-A* and *BARE1-B1;* Figs. 2 and 5) also containing sequences from the tetraploid and hexaploid *murinum* taxa. Therefore, this study allowed identification of two groups of *BARE1* elements (*A* and *B1*) that are specific to the *H. murinum* complex. Interestingly, this is supported by our screening of the reference barley genome (*Hv* 462R1) in which no homologous *BARE1-A* and -*B1 RTs* could be detected (Fig. 3). Also, this result lends additional support to subsp. *glaucum* as one of the parents of the polyploids (Jakob and Blattner 2010; Tanno et al. 2010; Ourari et al. 2011; Cuadrado et al. 2013; Brassac and Blattner 2015). In allopolyploid cotton, most TEs inherited from their parental diploid genomes remain closely related to their respective parental orthologs (Hu et al., 2010). Consequently, the authors were able to determine the parental genome origin of many TE sequences from the allopolyploid genome. The other weakly represented *Copia* elements only detected in the polyploids of the *H. murinum* complex *via* the PCR-based approach (such as for instance, *BARE1-C*, *RIRE1, IKYA* and *TAR-1*) suggest that they may have originated from their unknown (most likely extinct) parent (other than *glaucum*). However, it cannot be ruled out that these elements might have been present in the *glaucum* genome before being degenerated and/or eliminated, or yet present at a very low copy number. For example, the PCR-based approach did not reveal traces of *RIRE1 RTs* in our *H. spontaneum* (syn. *vulgare*), whereas our investigations allowed detection of the presence of these weakly represented elements in the reference barley genome, which is likely due to a low depth cloning and sequencing. As mentioned above, generating NGS data on *Hordeum* complex should allow detecting the potential low copy transposable elements, if they are present. Although *BARE1-C*, *RIRE1, IKYA* and *TAR*-1 elements were found in both *H. murinum* subspecies and other diploid *Hordeum* species analyzed in this study (*H. marinum, H. bulbosum* or *H. spontaneum*), it is unlikely that one of them represents a possible progenitor of the *H. murinum* allopolyploids, according to previous congruent phylogenies based on nuclear genes (see Fig. 5), which provided evidence that these taxa are clearly distinct from the *H. murinum* complex (Jakob and Blattner 2010; Tanno et al. 2010; Ourari et al. 2011). The latter pattern of relationships found additional support and was more accurately defined, using phylogenomic analysis from second-generation sequencing data (Brassac and Blattner 2015), and molecular cytogenetic evidence (Cuadrado et al. 2013). Besides, the potential activation of TEs in genomes following a genomic stress, such as hybridization, has been demonstrated in different systems, as shown for instance in the *Zizania* × *Oryza* hybrid (Liu and Wendel 2000) and in the cotton and wheat allopolypoids (Kraitshtein et al. 2010). It has been shown that different TE families originating from their parental genomes might be differentially amplified following hybridization/allopolyploidization (Parisod et al. 2010; Kraitshtein et al. 2010). Senerchia et al. (2014) also identified different evolutionary trajectories from TE sequences in wild wheat, as demonstrated by proliferation of Sabine-*RTs* in particular polyploids, but massive elimination in others.

In this study, no remarkable change in the FISH signal intensity using *RT* probes (Fig. 4) was detected according to ploidy levels in the *murinum* complex. Even the FISH method only provides a rough estimate of the *Copia* abundance and scattering in the genomes, no striking change in *Copia* element intensity was observed in the *murinum* polyploids (4*x*, 6*x*), as compared at least to their extant diploid parental genome (*H. murinum* subsp. *glaucum*). This suggests that no significant *Copia* amplification followed allopolyploidization events in this complex, or alternatively that amplification could have been followed by subsequent mechanisms leading to sequence DNA loss. Interestingly, previous data (indicated in Fig. 5) showed that the tetraploid cytotypes (subsp. *murinum* and subsp. *leporinum*), which display a nearly similar genome size, have a larger genome (+ 0.75-1 pg of DNA) than the approximate value corresponding to the double of the genome size of their extant diploid progenitor (subsp. *glaucum*) (Ourari et al. 2011). However, regarding that the genome sizes of the other diploid progenitors of the "*murinum*" polyploids are unknown, and that only *Copia* elements (among the other TEs) have been surveyed, no conclusion can be made here on the causes of such eventual genome size increase in the tetraploid cytotypes following allopolyploidization. Further analyses based on NGS data and appropriate bioinfiormatic tools (Novak et al. 2010; Staton and Burke 2015) should allow to get a more complete qualitative and quantitative overview on the role of the different types of repetitive sequences (not only *Copia* elements) in the eventual genome increase in tetraploids.

Different situations have been reported regarding the dynamics of TEs in synthetic and natural allopolyploids. For instance, only few changes were observed in experimentally resynthetisized polyploids in *Gossypium* (Liu et al. 2001), whereas the retroelement content increased three-fold in soybean following polyploidy (Innes et al. 2008). Nevertheless, it has been shown that rearrangements resulting in TE loss might occur early after the allopolyploidization event and might hide the amplification effects (Kraitshtein et al. 2010; Yaakov and Kashkush 2011).

We also observed that the *Copia*-signal was similarly distributed across all the chromosomes and does not distinguish the three *murinum* subgenomes, suggesting that the unknown parental subgenomes of the 4*x* and 6*x* polyploids already possessed these *Copia* elements. This also, suggests that these subgenomes are closely related, which agree with previous phylogenetic and cytogenetic evidence (Ourari et al. 2011; Cuadrado et al. 2013; Brassac and Blattner 2015). Cuadrado et al. (2013) were able to go further and to distinguish these subgenomes using appropriate SSR-FISH markers.

Distribution of the *Copia* elements in the genomes of the *H. murinum* complex

There is evidence from molecular cytogenetic results that retrotransposons show preferential insertion into some genomic regions. For example, many Ty1-*Copia* retrotransposons are preferentially found within euchromatic regions. Their insertion in close proximity to or directly into genes greatly affects gene expression and function (Kumar and Bennetzen 1999; Grandbastien et al. 2005; Mansour 2007).

In the *H. murinum* complex, the diploid *RT-Copia* probes uniformly label the chromosome arms, except at the sub-telomeric and peri-centromeric regions, where the signal is lacking. This *RT* distribution appears quite similar to that observed in *H. vulgare* using *Copia* / *Gypsy* annotations and *RT* domains (Mascher et al. 2017), suggesting few changes in the overall organization of *Copia* in the genus. This pattern was earlier seen for *BARE1* using LTR probes in barley (Suoniemi et al. 1996; Vicient et al. 1999), and is consistent with the evidence generated from whole genome sequencing of barley (Mayer et al. 2012).

Also, this is in accordance with data reported in number of plant species such as rice, maize, wheat, barley and sugarcane, in which the centromeric region rather contain Ty3-*Gypsy* elements than *Copia* elements (Hirsch and Jiang 2012). Brandes and co-workers (1997) demonstrated that the Ty1-*Copia* group is dispersed throughout the euchromatic regions, but absent from regions where specialized tandem repeats are expected to lie, such as centromeres, telomeres, heterochromatin, and the nucleolus organizing region (NOR). Additionally, the retroelements are largely excluded, from centromeric satellite arrays. The data suggest that centromeric satellite arrays are under selection for their function in chromosome movement, much like genic regions (SanMiguel et al. 1996).

In conclusion:

In this study, the diversity of the *Copia*-like LTR-retrotransposons occurring in the genomes of the *H. murinum* polyploid complex has been investigated, using a PCR-based approach of their conserved *RT* domain. In spite of some expected sampling biases and limits of the method (particularly, performance of universal primers, and sequencing depth), the PCR-based exploration provided a first overview on the diversity, distribution and relative importance of *Copia* elements in the "*H. murinum*" complex and related taxa.

Therefore, it was shown that Ty1-*Copia* retroelements are well represented in the genomes of the *H. murinum* L. complex. These elements belong to four *Copia* families referred to *BARE1*, *RIRE1*, *IKYA*, and *TAR-1*. In accordance with studies using NGS-based data on barley, *BARE1* also appeared as the most important *Copia* family detected in these genomes. Among the three major subfamilies (*A, B*, *C*) found in *BARE1*, *BARE1-B* elements showed a great diversity which is represented in all taxa analyzed, including *H. bulbosum*, *H. spontaneum*, *H. marinum* and *H. murinum*. *BARE1-A* and *BARE1-B* elements appear to have been specifically amplified in the *H. murinum* complex, thus providing a genomic hallmark to this taxonomic group. In contrast, the other *BARE1-C*, *IKYA* and *TAR-1* elements are poorly represented in the *murinum* polyploid cytotypes and seem absent in the diploid subspecies *glaucum*. These elements could come from the other two unknown diploid parents of the polyploids or might result from amplification following allopolyploidization events.

In situ hybridization method (FISH) confirmed the abundance and similar distribution of *RT* elements throughout the diploid and polyploid genomes (excepted in the centromeric and NOR

regions). No remarkable change in retroelement intensity was detectable in the *murinum* polyploids (4*x* and 6*x*), as compared to their diploid parental genome (*H. murinum* subsp. g*laucum*), showing no evidence of *Copia* amplification following allopolyploidization in the *H. murinum* complex.

Overall, the results are consistent with the main findings observed elsewhere in barley, and provided new information and insights on the *Copia* elements occurring in this complex. Therefore, this study provides a consistent framework (including questions to be addressed) that paves the way for further deepest NGS-based evaluation of the repetitive content of the *H murinum* genomes to better understand their evolutionary dynamics.

Acknowledgements
Both the Algerian and French Ministries of Education and Research and the French-Algerian PHC joint research program are thanked for funding support to our collaborative Project Tassili 08 MDU 724.

Conflict of interest
The authors declare that they have no conflict of interest.

Data Archiving
Sequence data have been submitted to GenBank under accession numbers JN135335-JN135475.

References

Alix K, Heslop-Harrison JS (2004) The diversity of retroelements in diploid and allotetraploid *Brassica* species. Plant Mol Biol 54(6):895–909. doi:10.1007/s11103-004-0391-z

Beier S, Himmelbach A, Colmsee C, Zhang X-Q, Barrero RA, Zhang Q et al. (2017) Construction of a map-based reference genome sequence for barley, *Hordeum vulgare* L. Scientific Data 4(1): 170044-170044. https://doi.org/10.1038/sdata.2017.44

Belyayev A, Kalendar R, Brodsky L, Nevo E, Schulman AH, Raskina O (2010) Transposable elements in a marginal plant population: temporal fluctuations provide new insights into genome evolution of wild diploid wheat. Mobile DNA 1(1): 6. doi:10.1186/1759-8753-1-6

Bennetzen JL, Ma J, Devos KM (2005) Mechanisms of recent genome size variation in flowering plants. Ann Bot 95 (1): 127–132. doi:10.1093/aob/mci008

Brandes A, Heslop-Harrison JS, Kamm A, Doudrick RL, Schmidt T (1997) Comparative analysis of the chromosomal and genomic organization of Ty1-*Copia*-like retrotransposons in Pteridophytes, Gymnosperms and Angiosperms. Plant Mol Biol 33(1):11–21. doi:10.1023/A:1005797222148

Brassac J, Blattner FR (2015) Species-Level Phylogeny and Polyploid Relationships in *Hordeum* (Poaceae) Inferred by Next-Generation Sequencing and In Silico Cloning of Multiple Nuclear Loci. Syst Biol 64(5): 792–808. doi:10.1093/sysbio/syv035. PMID:26048340

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL (2009) BLAST+: architecture and applica-tions. BMC Bioinformatics 10: 421. doi :10.1186/1471-2105-10-421

Chénais B, Caruso A, Hiard S, Casse N (2012) The impact of transposable elements on eukaryotic genomes: From genome size increase to genetic adaptation to stressful environments. Gene 509(1): 7–15. doi.org/10.1016/j.gene.2012.07.042. PMID: 22921893

Covas G (1949) Taxonomic observations on the North American species of *Hordeum*. Madrono 10: 1–21

Cuadrado A, Carmona A, Jouve N (2013) Chromosomal Characterization of the Three Subgenomes in the Polyploids of *Hordeum murinum* L.: New Insight into the Evolution of This Complex. PloS ONE 8(12). e81385. doi.org/10.1371/journal.pone.0081385

Devos KM, Brown JK, Bennetzen JL (2002) Genome size reduction through illegitimate recombination counteracts genome expansion in *Arabidopsis.* Genome Res 12(7):1075–1079. doi:10.1101/gr.132102

Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem Bull 19(1): 11–15

Edgar RC (2004) MUSCLE: A multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics 5(1):113. doi:10.1186/1471-2105-5-113

Estep MC, DeBarry JD, Bennetzen JL (2013) The dynamics of LTR retrotransposon accumulation across 25 million years of panicoid grass evolution. Heredity 110(2):194–204. doi:10.1038/hdy.2012.99

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39(4): 783–791. doi:10.2307/2408678

Fu L, Niu B, Zhu Z, Wu S, Li W (2012) CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics 28:3150–3152. Doi:10.1093/bioinformatics/bts565.

Gerlach WL, Bedbrook JR (1979) Cloning and characterization of ribosomal RNA genes from wheat and barley. Nucleic Acids Res 7(7):1869–1885. doi:10.1093/nar/7.7.1869

Grandbastien MA (2004) Stress activation and genomic impact of plant retrotransposons. J Soc Biol 198(4): 425–432. PMID:15969350

Grandbastien MA, Audeon C, Bonnivard E, Casacuberta JM, Chalhoub B, Costa AP et al. (2005) Stress activation and genomic impact of Tnt1 retrotransposons in Solanaceae. Cytogenet Genome Res 110(1–4):229–41. doi:10.1159/000084957

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel, O (2010) New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. Systematic Biology 59(3):307–321. doi:10.1093/sysbio/syq010

Haas BJ, Kamoun S, Zody MC, Jiang RHY, Handsaker RE, Cano LM et al. (2009) Genome sequence and analysis of the Irish potato famine pathogen *Phytophthora infestans*. Nature 461(7262):393–8. doi:10.1038/nature08358

Hawkins JS, Kim H, Nason JD, Wing RA, Wendel JF (2006) Differential lineage-specific amplification of transposable elements is responsible for genome size variation in *Gossypium.* Genome Res 16(10):1252–1261. doi/10.1101/gr.5282906

Hawkins JS, Proulx SR, Rapp RA, Wendel JF (2009) Rapid DNA loss as a counterbalance to genome expansion through retrotransposon proliferation in plants. Proc Natl Acad Sci USA 106 (42):17811–17816. doi:10.1073/pnas.0904339106

Hirochika H, Hirochika R (1993) Ty1-*copia* group retrotransposons as ubiquitous components of plant genomes. Jpn J Genet 68(1):35–46. doi:10.1266/ggs.68.35

Hirsch CD, Jiang J (2012) Centromeres: sequences, structure, and biology. In Wendel JF, Greilhuber J, Dolezel J, Leitch IJ et al. (Ed). Plant Genome Diversity, Volume 1: Plant Genomes, Their Residents, and Their Evolutionary Dynamics. Springer-Verlag, Wien, pp 59–70. doi:10.1007/978-3-7091-1130-7_4.

Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Le SV (2017) UFBoot2: Improving the ultrafast bootstrap approximation. Mol Biol Evol 35(2):518–522doi:10.1093/molbev/msx281

Hu G, Hawkins JS, Grover CE, Wendel JF (2010) The history and disposition of transposable elements in polyploid *Gossypium.* Genome 53(8):599–607. doi:10.1139/g10-038

Innes RW, Ameline-Torregrosa C, Ashfield T (2008) Differential accumulation of retroelements and diversification of NB-LRR disease resistance genes in duplicated regions following polyploidy in the ancestor of soybean. Plant Physiol 148(4):1740–1759. doi:10.1104/pp.108.127902

Jakob SS, Blattner FR (2010) Two extinct diploid progenitors were involved in allopolyploid formation in the *Hordeum murinum* (Poaceae: Triticeae) taxon complex. Mol Phylogenet Evol 55(2):650–659. doi.org/10.1016/j.ympev.2009.10.021

Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J et al. (2005) Repbase Update, a database of eukaryotic repetitive elements. Cytogenet Genome Res 110(1–4):462–467. doi:10.1159/000084979

Kalendar R, Flavell AJ, Ellis THN, Sjakste T, Moisy C, Schulman AH (2011) Analysis of plant diversity with retrotransposon-based molecular markers. Heredity 106(4):520–530. doi:10.1038/hdy.2010.93

Kalendar R, Grab T, Regina M, Suoniemi A, Schulman AH (1999) IRAP and REMAP: two new retrotransposon-based DNA fingerprinting techniques. Theor Appl Genet 98(5):704–711. doi:10.1007/s001220051124

Kalendar R, Tanskanen J, Immonen S, Nevo E, Schulman AH (2000) Genome evolution of wild barley (*Hordeum spontaneum*) by *BARE1* retrotransposon dynamics in response to sharp microclimatic divergence. Proc Natl Acad Sci USA 97(12):6603–6607. doi:0.1073/pnas.110587497

Kalyaanamoorthy S, Minh B, Wong T et al. (2017) ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods 14:587–589. https://doi.org/10.1038/nmeth.4285

Katoh K, Misawa K, Kuma K, Miyata T (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res 15-30:3059-66 Nucleic Acids Res. 15;30(14):3059-66

Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S et al. (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 15-28(12):1647–9. doi:10.1093/bioinformatics/bts199

Kraitshtein Z, Yaakov B, Khasdan V (2010) The genetic and epigenetic dynamics of a retrotransposon after allopolyploidization of wheat. Genetics 186(3):801–812. doi:10.1534/genetics.110.120790

Kumar A, Bennetzen JL (1999) Plant retrotransposons. Ann Rev Genet 30:479–532. doi:10.1146/annurev.genet.33.1.479

Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Mol Biol Evol 35(6):1547–1549 doi:10.1093/molbev/msy096

Li W, Godzik A (2006) Cd-Hit: a Fast Program for Clustering and Comparing Large Sets of Protein or Nucleotide Sequences. Bioinformatics 22(13):1658-9. DOI:10.1093/bioinformatics/btl158

Lisch D (2013) How important are transposons for plant evolution? Nature Reviews Genetics 14 (1):49–61. doi:10.1038/nrg3374

Lisch D, Bennetzen JL (2011) Transposable element origins of epigenetic gene regulation. Curr Opin Plant Biol 14(2):156–161. doi.org/10.1016/j.pbi.2011.01.003

Liu B, Brubaker CL, Mergeai G, Cronn RC, Wendel JF (2001) Polyploid formation in cotton is not accompanied by rapid genomic changes. Genome 44(3):321–330. doi:10.1139/gen-44-3-321

Liu B, Wendel JF (2000) Retroelement activation followed by rapid repression in interspecific hybrid plants. Genome 43(5):874–880. doi :10.1139/g00-058

Liu B, Wendel JF (2003) Epigenetic phenomena and the evolution of plant allopolyploids. Mol Phylogenet Evol 29(3):365–379. doi:10.1016/S1055-7903(03)00213-6

Löve A, Löve D (1948) Chromosome numbers of northern plant species. Dept, of Agr. Report Ser. B, No 3, Iceland.

Ma J, Bennetzen JL (2006) Recombination, rearrangement, reshuffling, and divergence in a centromeric region of rice. Proc Natl Acad Sci USA 103(2):383–8. doi:10.1073/pnas.0509810102

Ma XF, Fang P, Gustafson JP (2004) Polyploidization-induced genome variation in triticale. Genome 47(5):839-848. doi:10.1139/g04-051

Macas J, Neumann P, Navrátilová A (2007) Repetitive DNA in the pea (*Pisum sativum* L.) genome: comprehensive characterization using 454 sequencing and comparison to soybean and *Medicago truncatula*. BMC. BMC Genomics 8:427. https://doi.org/10.1186/1471-2164-8-427

Manninen I, Schulman AH (1993) *BARE1*, a *copia*-like retroelement in barley (*Hordeum vulgare* L.). Plant Mol Biol 22(5):829–846. doi:10.1007/BF00027369

Mansour A (2007) Epigenetic activation of Genomic Retrotransposon. J Cell Mol Biol 6(2):99–107

Mascher M, Gundlach H, Himmelbach A, Beier S, Twardziok SO, Wicker T et al (2017) A chromosome conformation capture ordered sequence of the barley genome. Nature 544(427). doi:10.1038/nature22043

Matsuoka Y, Tsunewaki K (1996) Wheat Retrotransposon Families Identified by Reverse Transcriptase Domain Analysis. Mol. Biol. Evol. 13(10):1384-1392

Mayer KF, Waugh R, Brown JW, Schulman A, Langridge P, Platzer M et al. (2012) A physical genetic and functional sequence assembly of the barley genome. Nature 491(7426):711–716. doi:10.1038/nature11543

Neumann P, Novák P, Hoštáková N, Macas J (2019) Systematic survey of plant LTR retrotransposons elucidates phylogenetic relationships of their polyprotein domains and provides a reference for element classification. Mobile DNA 10:1. https://doi.org/10.1186/s13100-018-0144-1

Lam-Tung Nguyen, Heiko A., Arndt von Haeseler, Bui Quang Author Notes

Nguyen LT, Schmidt HA, Haeseler A, Minh BQ (2015) IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol 32(1):268–274. https://doi.org/10.1093/molbev/msu300

Noma K, Nakajima R, Ohtsubo H, Ohtsubo E (1997) *RIRE1*, a retrotransposon from wild rice *Oryza australiensis*. Genes Genet Syst 72(3):131–140. doi:10.1266/ggs.72.131

Novak P, Neumann P, Macas J (2010) Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data. BMC Bioinformatics 11: 378

Ourari M, Ainouche A, Coriton O, Huteau V, Brown S, Misset MT et al. (2011) Diversity and evolution of the *Hordeum murinum* polyploid complex in Algeria. Genome 54(8):639–654. doi:10.1139/g65-040

Ozkan H, Levy AA, Feldman M (2001) Allopolyploidy-induced rapid genome evolution in the wheat (*Aegilops-Triticum*) group. Plant Cell 13(8):1735–1747. doi:10.1105/TPC.010082

Parisod C, Salmon A, Zerjal T, Tenaillon M, Grandbastien MA, Ainouche M (2009) Rapid structural and epigenetic reorganization near transposable elements in hybrid and allopolyploid genomes in *Spartina*. New Phytol 184(4):1003–1015. doi:10.1111/j.1469-8137.2009.03029.x

Parisod C, Alix J, Just M, Petit V, Mhiri C, Ainouche M et al. (2010). Impact of transposable elements on the organisation and function of allopolyploid genomes. New Phytol 186(1):37–47. doi:10.1111/j.1469-8137.2009.03096.x

Piednoël M, Carrete-Vega G, Renner SS (2013) Characterization of the LTR retrotransposon repertoire of a plant clade of six diploid and one tetraploid species. The Plant Journal 75: 699–709. doi: 10.1111/tpj.1223

SanMiguel P, Tikhonov A, Jin YK, Motchoulskaia N, Zakharov D, Melake-Berhan A et al. (1996) Nested retrotransposons in the intergenic regions of the maize genome. Science 274(5288):765–768. doi:10.1126/science.274.5288.765

Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S et al. (2009) The B73 maize genome: complexity, diversity, and dynamics. Science 326(5956):1112–1115. doi:10.1126/science.1178534

Senerchia N, Felber F, Parisod C (2014) Contrasting evolutionary trajectories of multiple retrotransposons following independent allopolyploidy in wild wheats. New Phytol 202(3):975–985. doi:10.1111/nph.12731

Sharifi-Rigi P, Saeidi H, Rahiminejad MR (2014) Genetic diversity and geographic distribution of variation of *Hordeum murinum* as revealed by retroelement insertional polymorphisms in Iran. Biologia 69(4):469–477. doi:10.2478/s11756-014-0340-5

Staton SE, Burke JM (2015) Transposome: a toolkit for annotation of transposable element families from unassembled sequence reads Bioinformatics 31:1827–1829. doi: 10.1093/bioinformatics/btv059

Suoniemi A, Anamthawat-Jonsson K, Arna T, Schulman AH (1996) Retrotransposon *BARE1* is a major dispersed component of the barley *(Hordeum vulgare* L.) genome. Plant Mol Biol 30(6):1321-1329. doi:10.1007/BF00019563

Tamura K (1992) Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G+C content biases. Mol Biol Evol 9:678–68. DOI: 10.1093/oxfordjournals.molbev.a040752

Tanno K, von Bothmer R, Yamane K, Takeda K, Komatsuda T (2010) Analysis of DNA sequence polymorphism at the cMWG699 locus reveals phylogenetic relationships and allopolyploidy within *Hordeum murinum* subspecies. Hereditas 147(1):34–42. doi:10.1111/j.1601-5223.2009.02142.x

Vicient CM, Suoniemi A, Anamthawat-Jónsson K, Tanskanen J, Beharav A, Nevo E et al. (1999) Retrotransposon *BARE1* and its role in genome evolution in the genus *Hordeum*. Plant Cell 11(9):1769–1784. doi:10.1105/tpc.11.9.1769

Wicker T, Keller B (2007) Genome-wide comparative analysis of *copia* retrotransposon in triticeae, rice, and *Arabidopsis* reveals conserved ancient evolutionary lineages and distinct dynamics of individual *copia* families. Genome Res 17(7):1072–1081. doi:10.1101/gr.6214107

Wicker T, Narechania A, Sabot F, Stein J, Giang VTH, Graner A et al. (2008) Low-pass shotgun sequencing of the barley genome facilitates rapid identification of genes, conserved non-coding sequences and novel repeats. BMC Genomics 9: 518. doi:10.1186/1471-2164-9-518

Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B et al. (2007) A unified classification system for eukaryotic transposable elements. Nature Reviews Genetics 8(12):973–982. doi:10.1038/nrg2165

Wicker T, Taudien S, Houben A, Keller B, Graner A, Platzer M et al. (2009) A whole-genome snapshot of 454 sequences exposes the composition of the barley genome and provides evidence for parallel evolution of genome size in wheat and barley. Plant J 59(5):712–722. doi:10.1111/j.1365-313X.2009.03911.x

Yaakov B, Kashkush K (2011) Methylation, Transcription, and Rearrangements of Transposable Elements in Synthetic Allopolyploids. International Journal of Plant Genomics, vol. 2011, Article ID 569826, http://dx.doi.org/10.1155/2011/569826

Table 1 Annotation of the *Copia*-like elements detected in the *Hordeum murinum* L. complex and closely related taxa.

Annotation used in this study[a]	**Homolog families** [a,b]	**REXdb Copia lineage**[c]
Family/subfamily		
BARE1-A	*WIS; Angela*	*Angela*
BARE1-B	*WIS; Angela*	*Angela*
BARE1-C	*WIS; Angela*	*Angela*
RIRE1		*Angela*
IKYA	*IDA*	*Ikeros*
TAR-1	*SASANDRA*	*TAR*
[a] Folowing the specific classification for barley and Triticeae (Suoniemi et al. 1996; Vicient et al. 1999; Wicker et al. 2009)		
[b]From TREP and Repbase		
[c]From the unified classification system of plant LTR-retrotransposons (Neumann et al. 2019)		

Table 2 Estimates of average pairwise sequence divergence (p-distance) within (in bold) and between groups of *Copia*-like elements in the *Hordeum murinum* L. complex using MEGA X software (Kumar et al. 2018)

Groups of *Copia* elements	***BARE1***						***RIRE1***	***IKYA***	***TAR-1***
	A	***B***				***C***			
		B1	***marinum***	***bulbosum***	***spontaneum***				
BARE1-A	**0.111**								
BARE1-B1	0.205	**0.098**							
BARE1-B marinum	0.196	0.121	**0.089**						
BARE1-B bulbosum	0.209	0.144	0.136	**0.090**					
BARE1-B spontaneum	0.201	0.126	0.133	0.107	**0.047**				
BARE1-C	0.231	0.167	0.171	0.171	0.151	**0.122**			
RIRE1	0.358	0.311	0.326	0.292	0.288	0.296	**0.098**		
IKYA-IDA	0.366	0.358	0.360	0.352	0.333	0.355	0.380	**0.044**	
TAR-1	0.430	0.410	0.411	0.404	0.408	0.408	0.362	0.387	**0.133**

TITLES AND LEGENDS TO FIGURES

Fig. 1 Maximum likelihood tree (using the T92+G evolutionary model) of 141 *reverse transcriptase* sequences representing *Copia-like* retrotransposons cloned in the *H. murinum* L. complex and its closest diploid relatives*, H. marinum, H. bulbosum* and *H. spontaneum*. Bootstrap values higher than 50% are given for the main nodes. This unrooted tree includes annotated *RT* sequences from referee databases, TREP and REPbase, which are indicated by arrowed branches. Four families are circumscribed in t he tree, following the specific *Copia* classification used for barley and Triticeae.: *BARE1, RIRE1-I, IKYA, TAR*-1. Also is superimposed in the tree, the lineage assignment of these *Copia* families according to the recently published unified classification system of plant LTR-retrotransposons (Neumann et al. 2019)

Fig. 2 Maximum likelihood tree (using the T92+G evolutionary model) of *reverse transcriptase* sequences representing Ty1-*Copia* retrotransposons families identified in the *H. murinum* complex and their c losely related diploid taxa, *H. marinum*, *H. spontaneum* and *H. bulbosum*. Annotation assignments of the main clades and subclades of interest are indicated in the tree, following the specific *Copia* classification used for barley and Triticeae.: *BARE1* (*A*, *B*, *C*)*, RIRE1-I, IKYA*, *TAR-1*. Taxonomic origins of the *RT*-clones are given at the right of the tree for each clade. Bootstrap values (as %) are indicated on the nodes of the main clades

Fig. 3 Maximum likelihood tree generated from the analysis of a dataset including (i) all *BARE1* (*A* and *B1*) and *RIRE RT* sequences cloned and sequenced in this study from the *H. murinum complex*, *H. marinum*, *H. bulbosum* and *H. spontaneum*, and (ii) 24 *RT* sequences from the reference barley genome (highlighted in yellow in the tree) representing clusters of retrotransposon which show the highest sequence similarity with the *BARE1* (*A* and *B1*) and *RIRE RT* of the *H. murinum* complex. Annotation and taxonomic origins of the *RT*-clones are indicated in the tree. None of the barley *RTs* fell within the *BARE1 -A* or *BARE1-B1* subclades of *H. murinum*

Fig. 4 FISH karyotype analysis in the *H. murinum* complex using *RT-Copia* probe from *H. murinum* subsp. *glaucum* (in green) and 45S rDNA (in red). (a): *H. murinum* subsp. *glaucum* (2*x*); (b): *H. murinum* subsp. *leporinum* (4*x*); (c): *H. murinum* subsp. *murinum* (4*x*); (d): *H. murinum* subsp. *leporinum* (6*x*). Chromosomes carrying rDNA are identified in (e). Chromosomes were counterstained with DAPI (blue). Scale bars = 5 µm

Fig. 5 Distribution of retrotransposon Ty1-*Copia* subfamilies detected throughout the diploid and polyploid members of the *H. murinum* complex and its close diploid relatives, using amplification with universal primers and cloning of their *reverse transcriptase* domain. The phylogenetic tree of the *Hordeum* taxa is redrawn from Ourari et al. (2011) and Brassac and Blattner (2015). *2C* DNA content (in picogrammes) is indicated in brackets for the *H. murinum* samples analyzed here, according to Ourari et al. (2011)

Fig.1

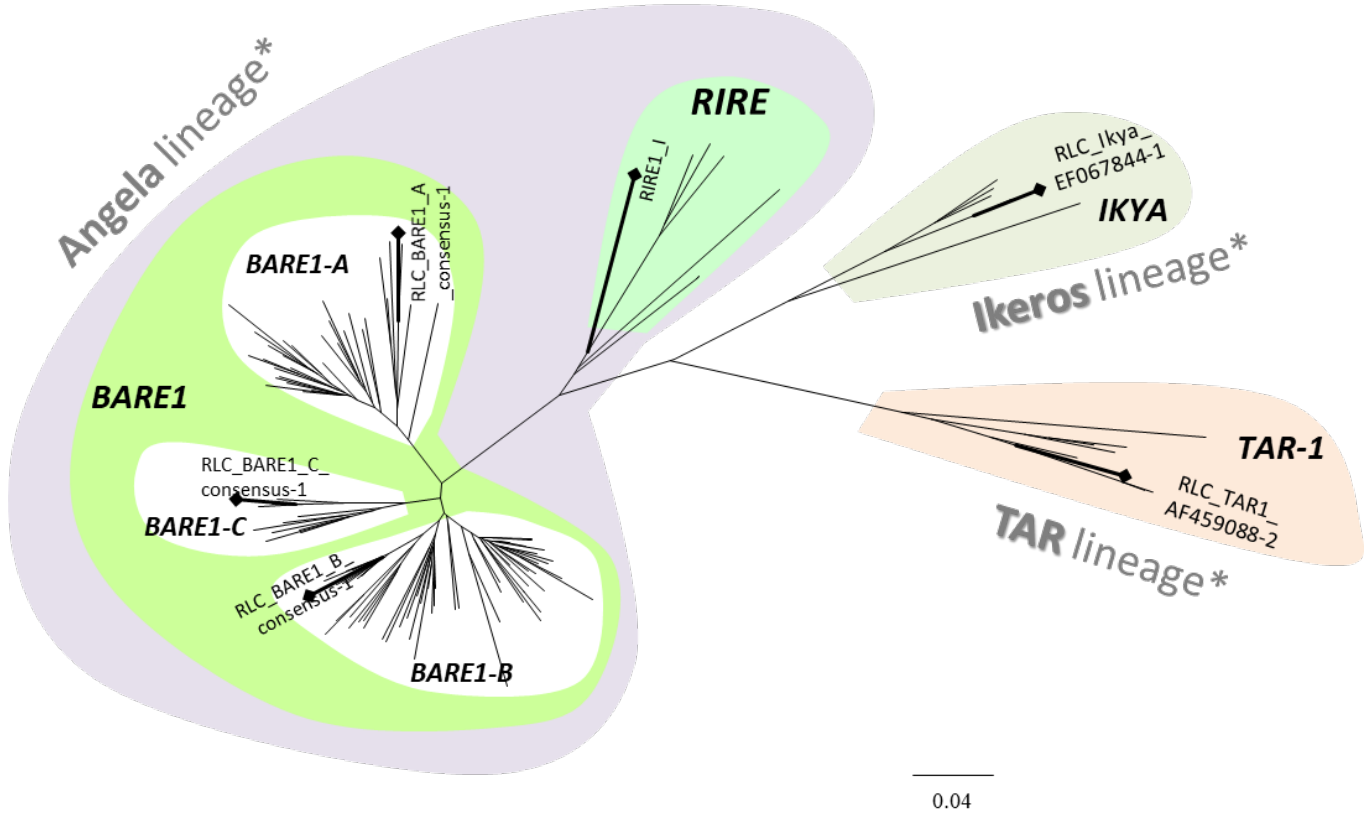

Angela lineage*
RIRE
RIRE1_I
RLC_Ikya_
EF067844-1
IKYA
Ikeros lineage*
BARE1-A
RLC_BARE1_A
_consensus-1
BARE1
RLC_BARE1_C_
consensus-1
BARE1-C
RLC_BARE1_B_
consensus-1
BARE1-B
TAR-1
RLC_TAR1_
AF459088-2
TAR lineage*
0.04

Fig.2

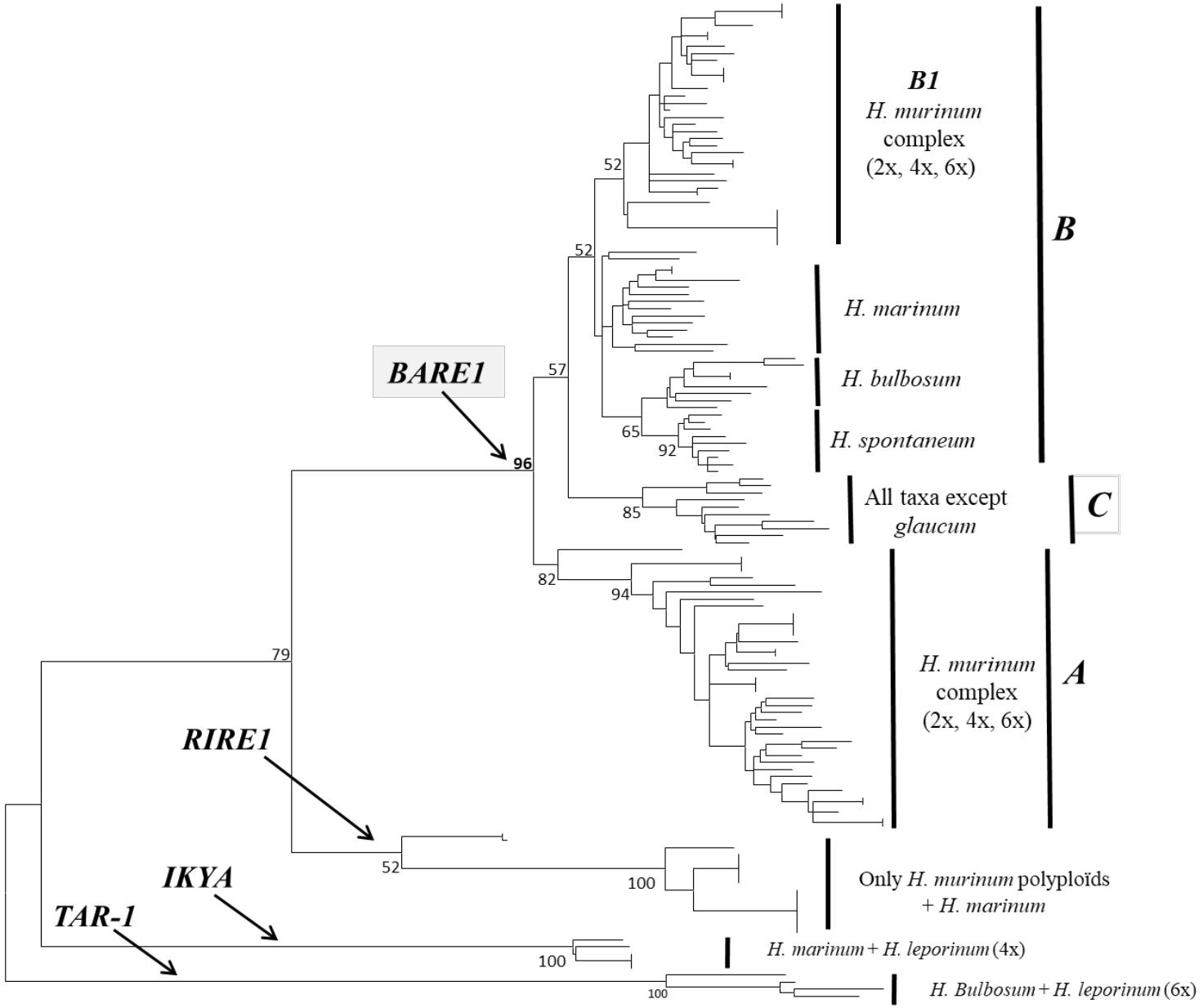

B1
H. murinum
complex
(2x, 4x, 6x)
B
H. marinum
H. bulbosum
H. spontaneum
BARE1
All taxa except
glaucum
C
H. murinum
complex
(2x, 4x, 6x)
A
RIRE1
Only H. murinum polyploïds
+ H. marinum
IKYA
H. marinum + H. leporinum (4x)
TAR-1
H. Bulbosum + H. leporinum (6x)
52
52
57
65
92
96
85
82
94
79
52
100
100
100

Fig.3

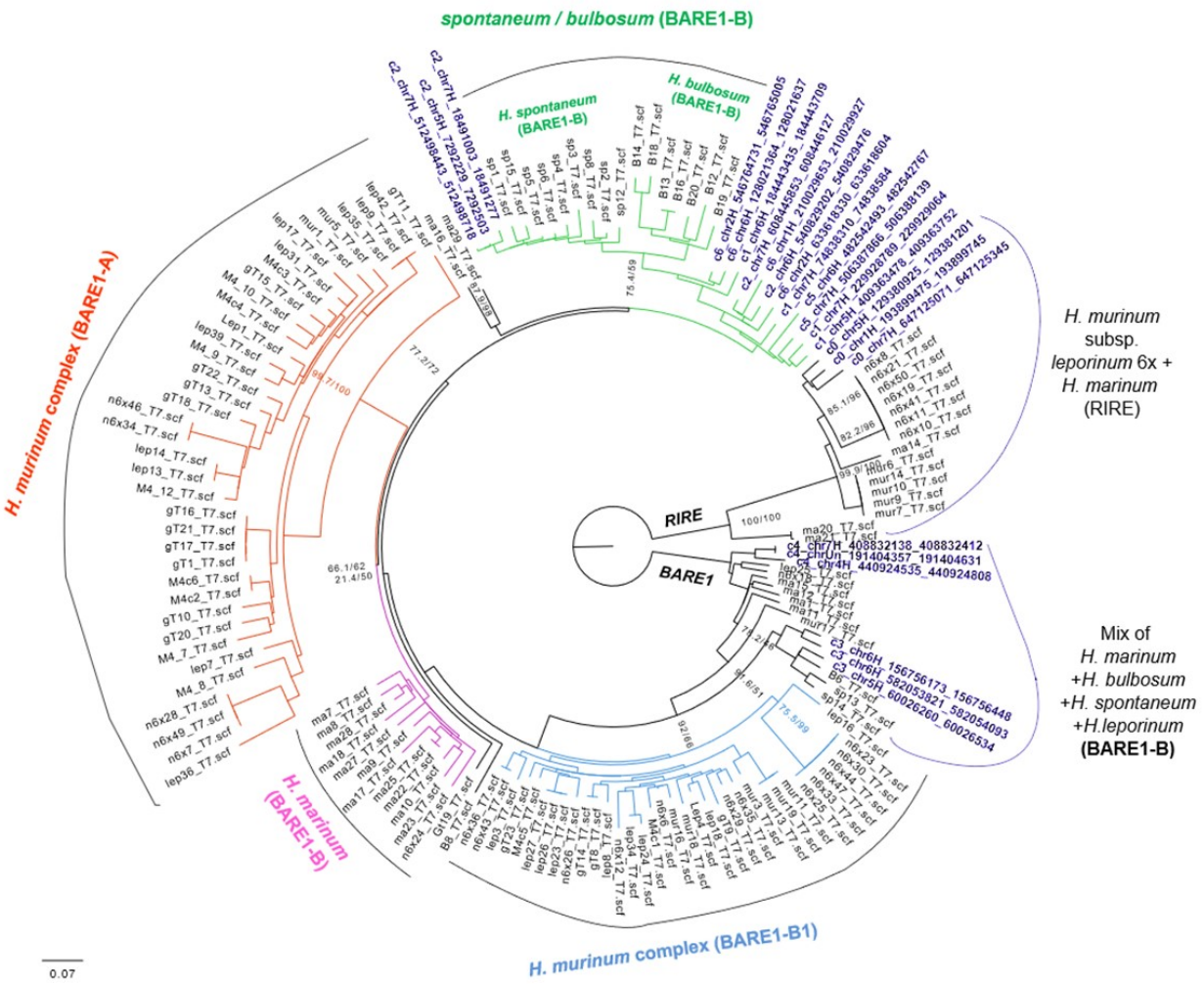

spontaneum / bulbosum (BARE1-B)
H. spontaneum (BARE1-B)
H. bulbosum (BARE1-B)
H. murinum complex (BARE1-A)
H. marinum (BARE1-B)
H. murinum complex (BARE1-B1)
RIRE
BARE1
H. murinum subsp. leporinum 6x + H. marinum (RIRE)
Mix of H. marinum +H. bulbosum +H. spontaneum +H.leporinum (BARE1-B)
0.07

Fig.4

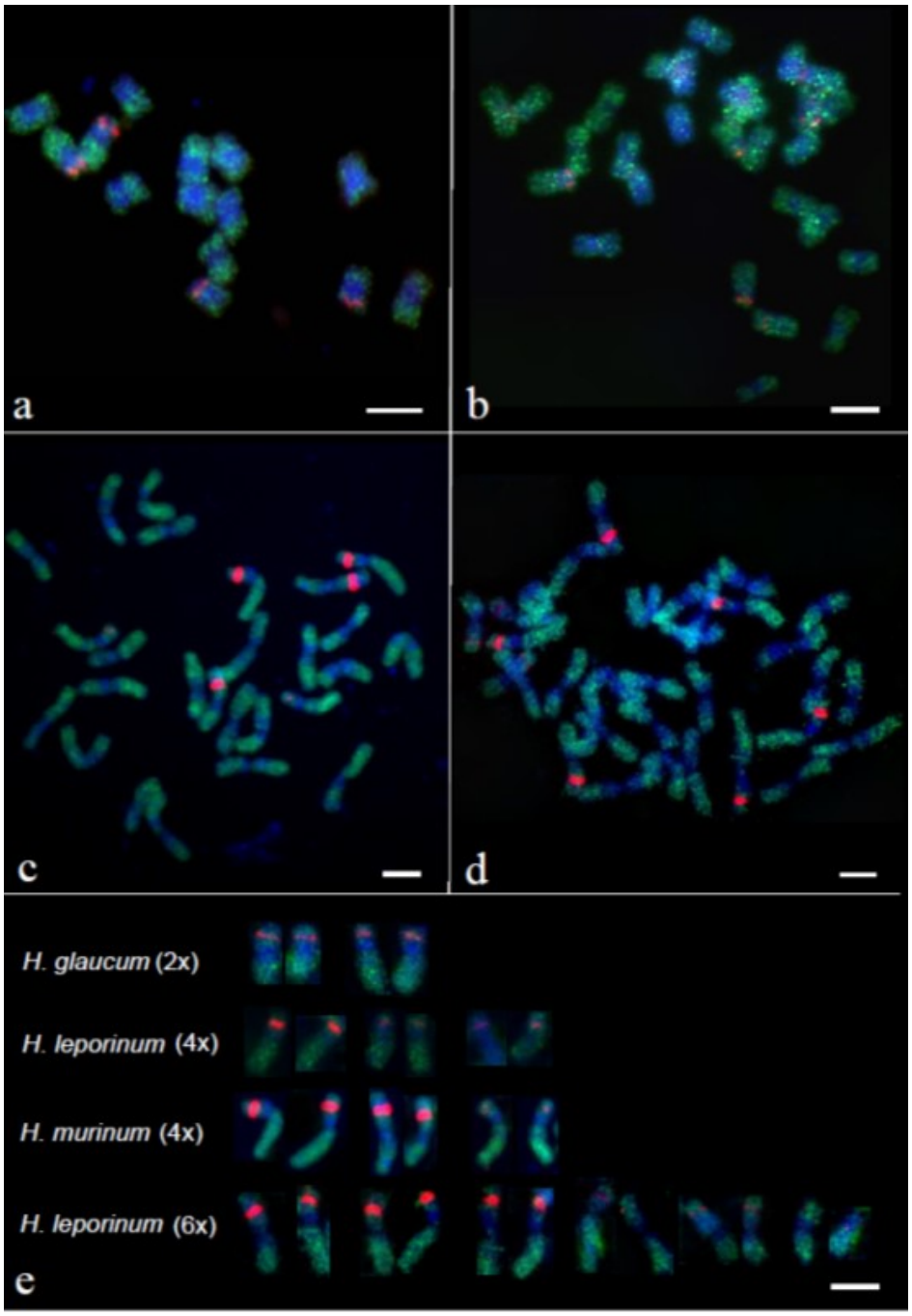
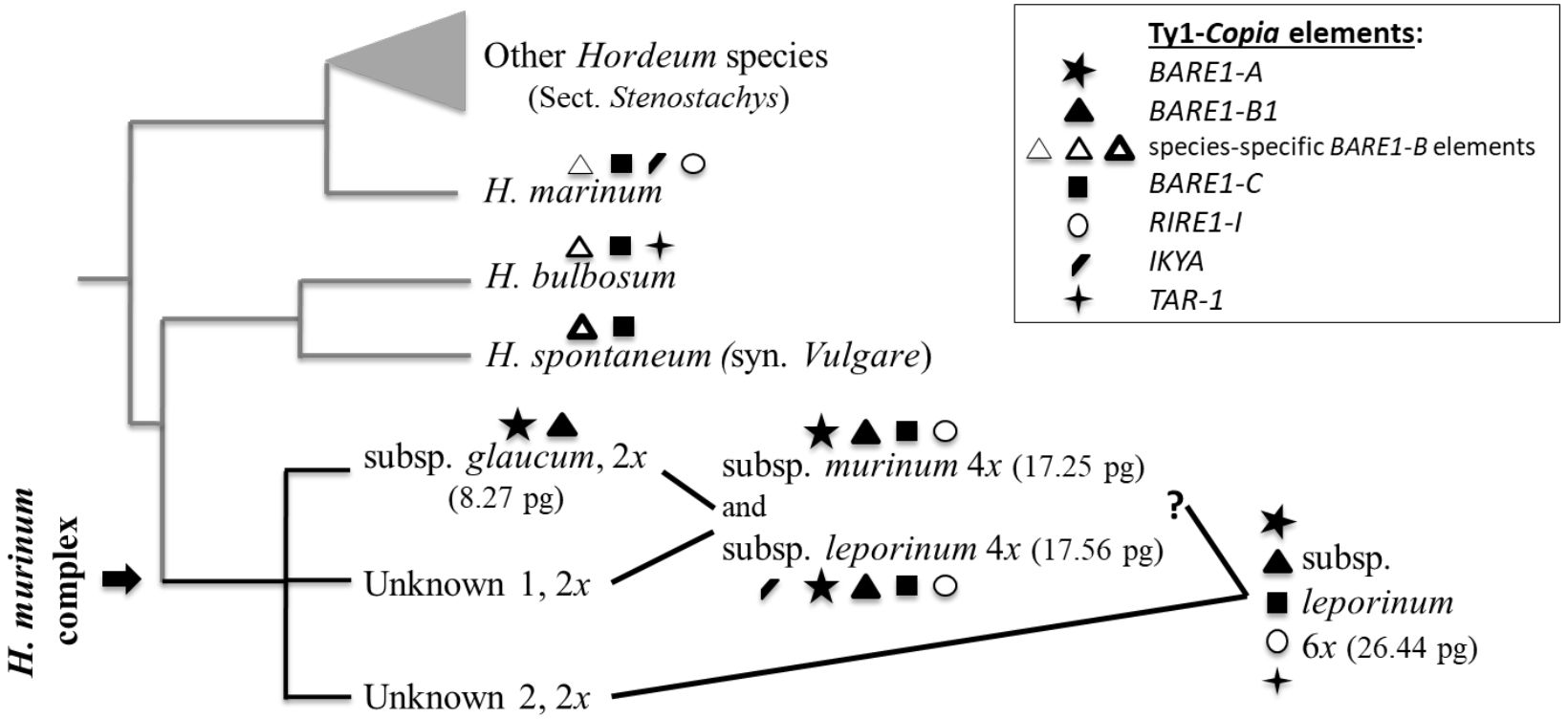


Fig.5


Ty1-*Copia* elements:
BARE1-A
BARE1-B1
species-specific *BARE1-B* elements
BARE1-C
RIRE1-I
IKYA
TAR-1
Other *Hordeum* species
(Sect. *Stenostachys*)
H. marinum
H. bulbosum
H. spontaneum (syn. *Vulgare*)
H. murinum complex
subsp. *glaucum*, 2*x*
(8.27 pg)
Unknown 1, 2*x*
Unknown 2, 2*x*
subsp. *murinum* 4*x* (17.25 pg)
and
subsp. *leporinum* 4*x* (17.56 pg)
?
subsp.
leporinum
6*x* (26.44 pg)