Theor Appl Genet (2012) 124:1365–1373
DOI 10.1007/s00122-012-1793-y

ORIGINAL PAPER

Marker utility of miniature inverted-repeat transposable elements for wheat biodiversity and evolution

Beery Yaakov · Elif Ceylan · Katherine Domb · Khalil Kashkush

Received: 19 May 2011 / Accepted: 5 January 2012 / Published online: 28 January 2012

Abstract Transposable elements (TEs) account for up to 80% of the wheat genome and are considered one of the main drivers of wheat genome evolution. However, the contribution of TEs to the divergence and evolution of wheat genomes is not fully understood. In this study, we have developed 55 miniature inverted-repeat transposable element (MITE) markers that are based on the presence/absence of an element, with over 60% of these 55 MITE insertions associated with wheat genes. We then applied these markers to assess genetic diversity among *Triticum* and *Aegilops* species, including diploid (AA, BB and DD genomes), tetraploid (BBAA genome) and hexaploid (BBAADD genome) species. While 18.2% of the MITE markers showed similar insertions in all species indicating that those are fossil insertions, 81.8% of the markers showed polymorphic insertions among species, subspecies, and accessions. Furthermore, a phylogenetic analysis based on MITE markers revealed that species were clustered based on genus, genome composition, and ploidy level, while 47.13% genetic divergence was observed between the two main clusters, diploids versus polyploids. In addition, we provide evidence for MITE dynamics in wild emmer populations. The use of MITEs as evolutionary markers might shed more light on the origin of the B-genome of polyploid wheat.

Communicated by A. Schulman.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-012-1793-y) contains supplementary material, which is available to authorized users.

B. Yaakov · E. Ceylan · K. Domb · K. Kashkush (✉)
Department of Life Sciences, Ben-Gurion University, 84105 Beer-Sheva, Israel
e-mail: kashkush@bgu.ac.il

Present Address:
E. Ceylan
Department of Soil Science and Plant Nutrition, Faculty of Agriculture, Ege University, Izmir, Turkey

Introduction

The grass family (*Gramineae*) includes ~10,000 species classified into ~700 genera. The divergence of various groups from an ancestral progenitor occurred perhaps 80 million years ago (Kellogg 2001; Gaut 2002; Prasad et al. 2005). Polyploidy occurs with high frequency in the grasses, including autopolyploidy, allopolyploidy, and segmental allopolyploidy (e.g. maize). Wheat is the world's most widely grown cereal, with annual world production of ~600 million tons, contributing ~20% of daily human caloric intake (Dubcovsky and Dvorak 2007). Wheat (*Triticum* ssp.) is classified into three major taxonomic groups: diploid (genome AA; $2n = 14$), tetraploid (genome BBAA; $2n = 4x = 28$), and hexaploid (genome BBAADD; $2n = 6x = 42$) [see review by Feldman and Levy (2005)]. Wheat, *Triticum-Aegilops* group is used as a classical model organism to study evolution through allopolyploidization, while emmer wheat (*T. turgidum* spp. *dicoccoides*) is extensively studied as a good source for important genes in breeding programs of bread wheat. *Aegilops*, a genus generally known as goatgrass, consists of about 23 species and numerous subspecies, which have unambiguously contributed the D-genome, and most probably the B-genome to polyploid wheat, while *Triticum* has contributed the A-genome.

Bread wheat is a relatively young species, which was created ~10,000 years ago following hybridization events between the tetraploid *T. turgidum* (genome BBAA,

$2n = 4x = 28$) and the diploid *Aegilops tauschii* (genome DD, $2n = 2x = 14$) [see review Nesbitt and Samuel (1996)]. The tetraploid *T. turgidum* was created as a result of hybridization between *T. urartu* (genome AA) and an unknown genome BB diploid species as female. Ernie Sears was the first to classify the 21 chromosome pairs of bread wheat into seven homoeologous groups (Sears 1954). From intensive studies over several decades, the diploid donors of the genomes of allopolyploid wheat have been identified, and allopolyploid wheat has been re-synthesized. The leading candidate progenitor of the B-genome is *Ae. speltoides* (genome SS), as its genome is the closest to that of *T. turgidum*. In fact, the origin of the B-genome remains controversial. Nevertheless, it is believed that the B-genome is derived from the S-genome of *Aegilops* section *Sitopsis*, which includes *Ae. speltoides*, *Ae. bicornis*, *Ae. longissima*, *Ae. searsii*, and *Ae. sharonensis* (Petersen et al. 2006).

Diversity and phylogenetic relationships among *Triticum* and *Aegilops* species have been extensively studied during the last 50 years. That includes cytological (Teoh et al. 1983; Salina et al. 2006), as well as molecular biology studies, such as DNA markers in the nuclear genome (Mori et al. 1995; Sasanuma et al. 1996; Wang et al. 2000a, c; Huang et al. 2002; Kudryavtsev et al. 2004; Sallares and Brown 2004) or in organelles (Wang et al. 2000b; Haider and Nabulsi 2008). A study by (Queen et al. 2004) elucidated the phylogenetic relationships between several *Aegilops* and *Triticum* species, based on LTR-retrotransposon SSAP markers, clustering similar genomes. Polymorphism in SSAP markers relies on mutations in the recognition site of the restriction enzyme used in the assay, unlike the present study, which assesses the presence or absence of a transposon. However, phylogenetic studies that rely on transposon-based genetic markers (Kalendar et al. 2011) were not as extensively studied in wheat as they were in vertebrates. Several studies used recently active TEs with known ancestral states as genetic markers to study phylogeny and diversity among and within humans and primates (Roy-Engel et al. 2001; Salem et al. 2003; Hedges et al. 2004; Xing et al. 2007a, b). These TE-based studies have shed new light on the evolutionary history of humans and primates (Xing et al. 2007b). A recent study in wheat showed that using TE markers for wheat phylogeny might depend on the level of TE activity throughout wheat evolution (Konovalov et al. 2010).

TEs exist in all organisms and usually make up a significant fraction of the genome. TEs are classified into two main classes based on their intermediate product (Wicker et al. 2007): RNA (class 1 elements or retrotransposons), or DNA (class 2 elements). For both classes, transposition is defined in *cis* when the element is autonomous, i.e. encodes its own proteins, or in *trans* when the element is defective (or non-autonomous), the latter being dependent on the coding abilities of the autonomous elements (Kazazian 2004; Sabot et al. 2004; Wicker et al. 2007). Surprisingly, many studies both in plants and animals, reported that in fact non-autonomous elements (both class 1 and class 2 elements) that have lost their protein-coding sequences and became miniature elements showed a high level of transpositional activity, such as the class 1 terminal-repeat retrotransposons in miniature (TRIMs) (Witte et al. 2001; Sabot et al. 2005) and the class 2 miniature inverted-repeat transposable elements (MITEs) (Yang et al. 2009).

MITEs are widespread in eukaryotic genomes; they are non-autonomous elements that are characterized by their relatively short sequence (up to a few hundred base pairs), structural similarity, conserved terminal repeats, and high copy number in some species [e.g. rice (Jiang et al. 2004), and wheat (our unpublished data)]. In plants, most MITEs are classified into two main superfamilies: *Tourist*-like with a target site preference of TAA (Bureau and Wessler 1994a), and *Stowaway*-like with a target site preference of TA (Bureau and Wessler 1994b; Jiang et al. 2004; Feschotte and Pritham 2007). Despite the lack of coding capacity, MITEs were shown to be active in rice (Jiang et al. 2003; Kikuchi et al. 2003; Nakazaki et al. 2003; Shan et al. 2005). In addition, a recent study showed that MITEs in rice might achieve their transposition activity by using transposases encoded by various autonomous transposons (Yang et al. 2009). In wheat, there are many well-characterized *Stowaway*-like MITE families that are present in high copy numbers (Wicker et al. 2001; Isidore et al. 2005; Miller et al. 2006; Cloutier et al. 2007; Choulet et al. 2010).

In this study, we have successfully developed 55 MITE markers in wheat that are based on presence/absence of an element, and are defined as a "full site" in case the element is present in the locus and as an "empty site" in case the element is absent. Then we applied these markers to study wheat diversity in *Triticum* and *Aegilops* species as well as in wild populations of emmer.

Materials and methods

Plant material and DNA isolation

Triticum and *Aegilops* species that were used in this study are: *T. urartu* (2 accessions), *T. monococcum* ssp. *aegilopoides*, *Ae. sharonensis*, *Ae. longissima*, *Ae. searsii* (4 accessions), *Ae. speltoides* (3 accessions), *Ae. tauschii* (14 accessions), *T. turgidum* ssp. *durum* (2 accessions), *T. turgidum* ssp. *dicoccoides* (wild emmer, 36 accessions), and *T. aestivum* ssp. *aestivum* (15 accessions) (see details in Supplemental Table 1). DNA was isolated from young leaves (age 4 weeks post-germination) using the DNeasy plant kit (QIAGEN).

MITE-insertion-based markers

MITE sequences were retrieved from the *Triticeae* Repeat Sequence Database (TREP; http://wheat.pw.usda.gov/ITMI/Repeats/), except for *Minos* elements (see Supplemental Table 2) which were retrieved from a BAC clone (accession number AY485644), and run in BLASTn (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Primers were designed from flanking sequence surrounding intact MITEs from the database using the Primer3 software version 0.4.0 (http://frodo.wi.mit.edu/primer3/; Supplemental Table 2). PCR reactions were prepared using 12 µl of Ultra Pure Water (Biological Industries), 2 µl of 10× Taq DNA polymerase buffer (Fisher Biotech), 2 µl of 25 mM $MgCl_2$ (Fisher Biotech), 0.8 µl of 2.5 mM dNTPs, 0.2 µl of Taq DNA polymerase (5 U/µl, Fisher Biotech), 1 µl of each site-specific primer (50 ng/µl) and 1 µl of genomic DNA (~50 ng/µl) from the mentioned wheat species. The PCR conditions for these reactions were: 94°c for 3 min, repeat (94°c for 1 min, 60°c for 1 min, 72°c for 1 min) 30 times and 72°c for 3 min. PCR products (~10 µl) were run on 1.5% agarose gels and stained with ethidium bromide (Amresco). The expected product sizes for amplicons containing or lacking an intact MITE (Supplemental Table 2) were determined against a DNA standard (100 bp ladder, Fermentas). The PCR products were purified using the Invisorb® Spin PCRapid Kit (Invitek) or extracted from the agarose gel using the MinElute® Gel Extraction Kit (QIAGEN), and sequenced with the 3730 DNA Analyzer (Applied Biosystems) at Ben-Gurion University.

Sequence analysis

For validation, the sequenced PCR products were aligned to the original BAC sequence using ClustalW2 (http://www.ebi.ac.uk/Tools/msa/clustalw2/).

Data analysis

The MITE-insertion-based band (allele) matrices were constructed by designating a PCR product with an expected size for the full site as 1 and an empty site as 0 (including lack of PCR product). Hierarchical agglomerative clustering analysis of the data with Bray-Curtis similarity and construction of the dendrogram was performed using the Primer6 software version 6.1.6 [Primer-E; (Clarke 1993)]. The similarity profile (SIMPROF) test was used on each node to assess the statistical significance of the dendrogram. SIMPROF calculates a mean profile by randomizing each variable's values and re-calculating the profile. The pi statistic is calculated as the deviation of the actual resemblance profile of the resemblance matrix with the mean profile. This is compared with the deviation of further randomly generated profiles to test for significance. Finally, significant branches ($p \leq 0.05$) in the phylogenetic tree were indicated on the tree nodes.

Results

Developing MITE-insertion-based markers

Based on publicly available wheat sequences, we have successfully developed 55 markers that are based on the presence/absence of a MITE insertion. Nearly half (28/55) of the sequences were from the B-genome, while 18 of the 28 (64.2%) B-genome sequences where from chromosome 3B. The rest of the 27 sequences were from the D and A genomes, or cDNA sequences. The 55 markers were developed from DNA sequences containing seven *Stowaway*-like MITE families (Supplemental Table 2): (1) thirteen markers of a ~162 bp-long *Thalos*; (2) four markers of a ~340 bp-long *Fortuna*; (3) nine markers of a ~85 bp-long *Athos*; (4) ten markers of a ~155 bp-long *Oleus*; (5) five markers of a 237 bp-long *Minos*; (6) four markers of a ~350 bp-long *Eos*; and (7) ten markers of a 127 bp-long *Pan*. Of the 55 studied MITE insertions, 33 (60%) were found to be associated with genes or coding sequences, while 17 of the 33 (51.5%) were inserted in introns of known genes such as *KN1 homeobox protein*, *acetyl-coenzyme A carboxylase*, *transcription factor S-II domain*, *peptide transporter* and *oxidoreductase NAD-binding domain* (see Supplemental Table 2). The remaining 22 insertions were found in repetitive regions (12 insertions), or in intergenic regions (10 insertions within 55 kb upstream or downstream of known genes).

In all cases, the primers for PCR analysis were designed to amplify the MITE insertion and flanking host sequences (~100 bp from each side of the intact element). Thus, the expected size of a PCR product will be the size of the MITE insertion plus the flanking sequences, this we termed "full site". In the case of an "empty site", a lack of a MITE insertion, the size of the PCR product will be shorter, containing only the flanking sequences. An example of a site-specific PCR for *Athos* (Atho-AF029897 in Supplemental Table 2), which was inserted in intron 2 of an *acetyl-coenzyme A carboxylase* gene, is shown in Fig. 1. In this example, the expected full site is 432 bp, while the empty site is 347 bp. While most of the diploid species have an empty site, except for *T. urartu* (both TMU06 and TMU38 accessions) that had a full site, both the tetraploid and hexaploid *Triticum* species showed two bands (both an empty and a full-site). Sequence analysis of the extracted PCR bands from *T. monococcum* (empty site) and from *T. urartu* (full site) confirmed that indeed the difference between the higher and the lower bands in the gel is a

presence or absence of an *Athos* element (Fig. 2). For all the 55 MITE markers, the primer sequences, expected PCR product sizes, chromosome or genome location, and annotation of flanking sequences are described in detail in Supplemental Table 2.

Genotyping *Triticum* and *Aegilops* species using MITE markers

Overall, 65 accessions (see Supplemental Table 1) of five diploid *Aegilops* species (diploid SS and DD genomes) and five *Triticum* species (diploid AA, tetraploid BBAA and hexaploid BBAADD genomes) were genotyped. Of the 55 markers, 10 (18.2%) showed monomorphic bands (expected full-site) in all species, indicating that they are ancient insertions. The other 45 markers were polymorphic among species and subspecies, while 4 main banding patterns were observed: (1) expected full site band (e.g. lane 2, Fig. 1); (2) expected empty site band (e.g. lane 1, Fig. 1); (3) two bands [corresponding to full and empty site (e.g. lane 10, Fig. 1)]; or (4) no amplification at all. The latter could occur as a result of lack of primer annealing because of a mutation or chromosomal aberration at the specific site. It is very important to note that the negative PCR amplifications were not the result of PCR artifacts because the PCRs were repeated at least twice and a positive control was used for DNA quality, together with the internal control, namely using the same DNA templates for all markers. In summary, each one of the diploid *Triticum* and *Aegilops* species showed positive results (full/empty site) for at least 31 of the 55 markers, while the tetraploid and hexaploid species showed positive results for at least 49 of the 55 markers. Specifically, in ~70% of the cases, the source of polymorphism among individuals was a presence or absence of a MITE insertion, meaning positive PCR products (full, empty or both sites) were seen. Examples of genotyping MITE markers in *Triticum* and *Aegilops* species are shown in Fig. 3. In some cases (e.g. patterns D and E in Fig. 3) there is a clear indication for possible MITE activation following allopolyploidization process, i.e. pattern D notes an insertion that occurred in the allotetraploid, and pattern E notes an insertion that occurred in the allohexaploid.

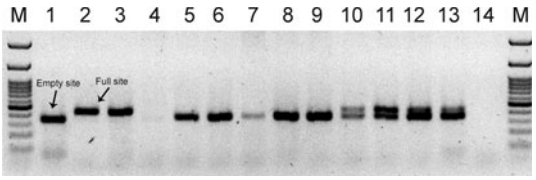

Fig. 1 Site-specific PCR analysis using primers that flanked an *Athos* insertion (Atho-AF029897 in Supplemental Table 2) in *Triticum* and *Aegilops* species: *1 T. monococcum*, *2 T. urartu* (TMU06), *3 T. urartu* (TMU38), *4 Ae. sharonensis*, *5 Ae. longissima*, *6 Ae. searsii*, *7 Ae. speltoides*, *8 Ae. tauschii* (TQ27), *9 Ae. tauschii* (TA1682), *10 T. dicoccoides*, *11 T. durum*, *12 T. aestivum* (6256), *13 T. aestivum* (TAA01), and *14* negative control (H_2O was used as a template in PCR). *M* denotes the 100 bp DNA ladder (Fermentas) that was used. *Bands* corresponding to either full site (432 bp) or empty site (347 bp) are indicated. Note that bands were isolated from the gel and sequenced (see Fig. 2)

Genetic diversity and phylogenetic trees based on MITE insertions in *Triticum* and *Aegilops* species

In order to gain insights into wheat phylogeny and evolution, we assessed the genetic diversity among and within various *Triticum* and *Aegilops* species using the MITE markers. Because MITEs are considered one of the most recently active elements in plants (Yang et al. 2009), we expected to find high genetic diversity in their insertion sites throughout wheat evolution. As stated above, except for the 10 monomorphic MITE insertions in all tested species, 45 markers were polymorphic in the *Triticum-Aegilops* group; while 23 of the 45 markers showed monomorphic bands in all *Aegilops* species but were polymorphic in *Triticum* species, 7 of the 45 markers showed monomorphic bands in all *Triticum* species but were polymorphic in *Aegilops* species. In addition, one MITE insertion was unique to *Aegilops*, while 13 insertions were unique to *Triticum* [e.g. no MITE insertion was detected in *Aegilops* species, while it was detected in tetraploid or hexaploid *Triticum* species (see Fig. 3, patterns d and e)].

Overall, a matrix of 550 alleles (full vs. empty site) in the ten *Triticum* and *Aegilops* species was constructed and used as an input in the Primer6 software (see Materials and methods). The output phylogenetic tree clearly showed two main clusters: diploid and polyploid species (Fig. 4), with 47.13% divergence between them. Note that in some cases, two bands (corresponding to both empty and full-site) were seen in tetraploids and/or hexaploids (see examples in Figs. 1, 3). In this case, the analysis was repeated twice: first by using the empty site band in the matrix, and second by using the full-site band. The two analyses resulted in similar phylogenetic trees.

In the diploid cluster (Fig. 4), *Triticum* species were clustered together, while the other cluster contained *Aegilops* species, with 41.73% divergence between the two clusters. Moreover, there is another significant clustering between the tetraploids and the hexaploids, with 25.74% divergence between them.

In the *Aegilops* cluster (Fig. 4), *searsii* and *speltoides* were clustered together, with 25% divergence between them, while *longissima* and *sharonensis* (with only 8.3% divergence between them) were clustered together with *tauschii*.

Furthermore, ~10% divergence based on MITE markers was observed among different accessions of the same

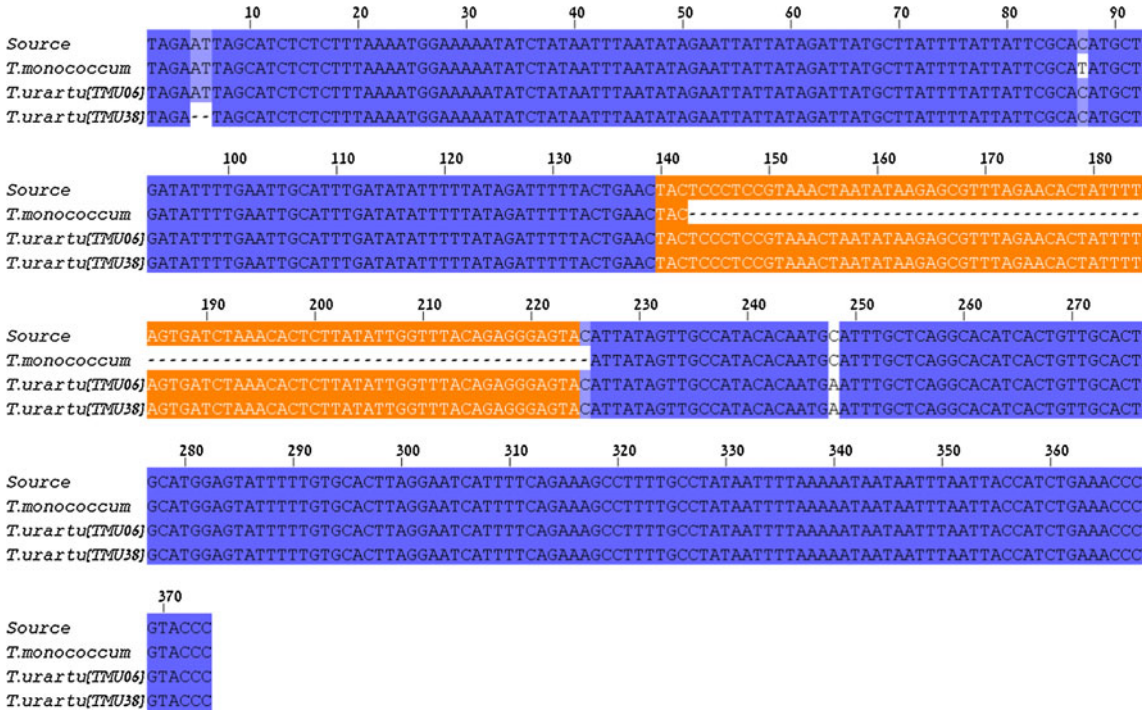

Fig. 2 Multiple sequence alignment of sequenced PCR bands (see Fig. 1) corresponding to *T. monococcum*, *T. urartu* (TMU06), *T. urartu* (TMU38), and the source sequence from the NCBI database (AF029897), from which primers were designed. The *Athos* sequence (85 bp) is indicated in *orange*, while flanking sequences are indicated in *blue*. Both TMU06 and TMU38 contain the element, while *T. monococcum* lacks the element (color figure online)

species (such as *aestivum* and *taushii*, see Supplemental Fig. 1). However, the clustering among the different accessions of the same species was statistically insignificant. These results indicate that MITEs showed a high activity level during the evolution of wheat, and most probably are still active today (see more details in discussion).

It is important to clarify that in the above analysis (full vs. empty site, Fig. 4), the "no score" allele was considered as an "empty site" allele. Of the 550 alleles, 184 (33.45%) were in fact "no score" alleles, which could result from mutations at the primer sequence site, or from deletion of the entire locus. In this case, the consideration of "no score" as an empty site would be misleading. In order to ensure that the lack of amplification in some of the samples did not skew the results, experiments were performed using additional primers flanking each one of the MITE elements (see Supplemental Table 3). In all cases, no clear PCR products corresponding to either full or empty site were seen. In any case, an additional analysis was performed using three allelic states (full, empty and no score) and the resulting phylogenetic tree (Supplemental Fig. 2) was similar to that achieved by two allelic states (see Fig. 4). Finally, it is important to note that the fact that about half of the markers (28 of 55) were originated from the B-genome did not significantly reduce the power of the analysis. This was tested by generating phylogenetic trees using similar number of markers representing the A, B, and D-genomes by removing B-derived markers randomly in each analysis (see Supplemental Fig. 4).

Evidence for MITE dynamics in wild emmer populations

Extensive studies were performed to assess the genetic diversity between and within populations of wild emmer wheat using allozymes or DNA markers (Nevo et al. 1982; Nevo and Beiles 1989; Fahima et al. 1999, 2002). To assess the dynamics of MITEs in different populations of emmer wheat and to test whether MITEs contribute to the genetic diversity among wild emmer populations, 4 Israeli populations (9 accessions for each population) of emmer wheat from Mount Aamasa, Amiad, Jaba, and Mount Hermon) were genotyped using 27 MITE markers.

Of the 27 MITE markers, 16 (59.26%) were monomorphic in the four populations, meaning that all individuals in all populations have similar MITE insertions. Another 4 (14.81%) were polymorphic in one population, 4 (14.81%) were polymorphic in two populations, 1 (3.7%) was polymorphic in three populations, and 2 (7.41%) were polymorphic in all populations. This data indicates that MITEs have been active in wild emmer wheat populations.

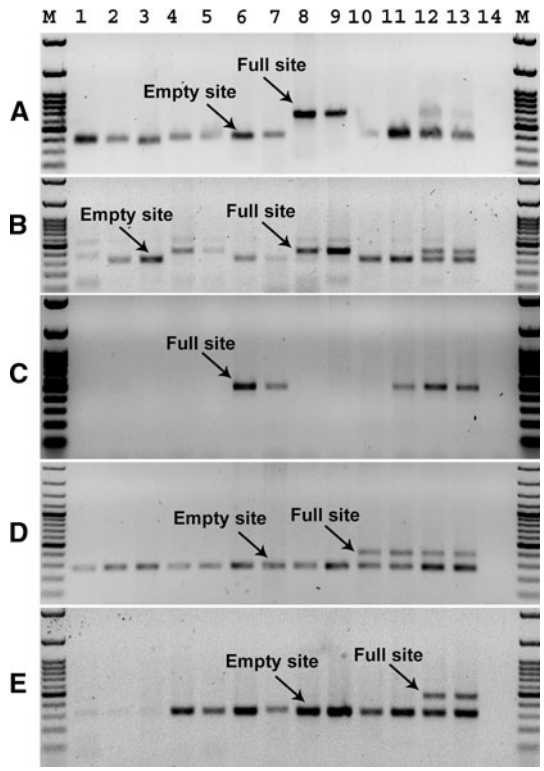

Fig. 3 Site-specific PCR analysis using primers that flanked (see Supplemental Table 2): **a** *Oleus* (Oleu-FN564429), **b** *Athos* (Atho-AB201447), **c** *Oleus* (Oleu-FN564430), **d** *Pan* (Pan-FN564428), and **e** *Thalos* (Thal-AF536819), in *Triticum* and *Aegilops* species: *1 T. monococcum*, *2 T. urartu* (TMU06), *3 T. urartu* (TMU38), *4 Ae. sharonensis*, *5 Ae. longissima*, *6 Ae. searsii*, *7 Ae. speltoides*, *8 Ae. tauschii* (TQ27), *9 Ae. tauschii* (TA1682), *10 T. dicoccoides*, *11 T. durum*, *12 T. aestivum* (6256), *13 T. aestivum* (TAA01), and *14* negative control (H_2O was used as a template in PCR). *M* denotes the 100 bp DNA ladder (Fermentas) that was used. *Bands* corresponding to either full site or empty site are indicated

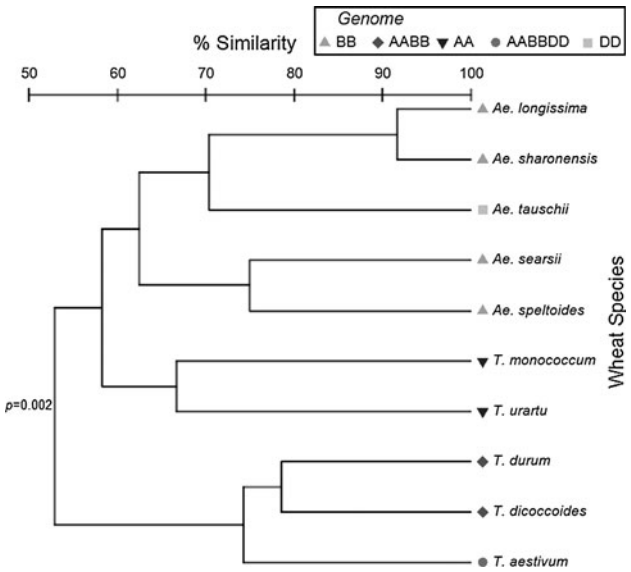

Fig. 4 Bootstrapped phylogenetic tree of 10 *Triticum* and *Aegilops* species based on MITE markers. The level of genetic similarity is indicated on *top*. The index (*top right*) indicates the genome composition of each species. The statistical significance (*p* value) is indicated in the tree

However, when we tested whether MITEs have a unique proliferation in wild emmer populations, no statistically significant inter-population or intra-population variation in the presence of MITEs was found by the Primer6 software, and no significant phylogenetic structure could be derived from the data (Supplemental Fig. 3). Furthermore, ~10% divergence based on MITE markers was observed between the emmer populations.

Discussion

In this study, we have retrieved MITE-containing sequences from the publicly available wheat database, and developed markers that are based on presence/absence of an element, similar to the previously reported assay for retrotransposable elements, named RBIP (Flavell et al. 1998). Next, using these markers, we have genotyped 79 accessions from 10 *Triticum* and *Aegilops* species and have constructed a phylogenetic tree. We found that, based on MITE insertional polymorphism, we can learn about the evolutionary history of wheat with relatively good resolution. In addition, we tested the pattern of 27 MITE insertions in four Israeli populations of wild emmer and found evidence for MITE activity. The data of this study clearly indicates the utility of MITEs as good markers for the study of wheat diversity and evolution, similarly to what was shown in barley (Lyons et al. 2008).

MITE markers and genetic diversity: evidence of activity

The main advantage of the MITE markers that we have developed in this study is that they are based on the presence or absence of an element. Such an assay can easily be applied to small-sized elements such as MITEs. In addition, by sequencing the empty-site-related PCR bands we can search for footprints and validate the transposition of the element from those "donor" sites. Comparing the sequences from full sites of different samples and source sequences also validates their common origin. Thus, such markers can give an indication not only of the genetic divergence that was caused by MITE transposition, but also of the history of MITE activity in each species. Moreover, 60% of the 55 studied MITEs were associated with wheat genes, whereas ~51% of those were inserted into introns. In wheat, an in silico study showed that ~43% of the MITE insertions are associated with genes (Sabot et al. 2005). In addition, MITEs were shown to be associated

with plant genes (Bureau and Wessler 1994a, b; Jiang et al. 2004; Naito et al. 2009), and were shown to interfere with the expression of rice genes (Naito et al. 2009).

The high genetic divergence (~41%) in MITE insertion sites between *Triticum* and *Aegilops* that were observed in this study (Fig. 4) and the divergence within different accessions of the same species such as *T. aestivum* (Supplemental Fig. 1) indicates that most probably MITEs were recently active in wheat. The phylogenetic tree that was constructed based on MITE insertional polymorphism clearly classified the *Triticum* and *Aegilops* species, and even classified species with different genome compositions (AA, BB, DD, BBAA, and BBAADD) into specific groups. Such a high-resolution phylogenetic tree could only be observed if MITEs were indeed recently active and show species-unique proliferations. The power of the phylogenetic tree was retained even when a third allelic state (no score) was used (see Supplemental Fig. 3), and when the number of B-genome derived MITE markers was reduced (see Supplemental Fig. 4). Note that the majority rule consensus phylogenetic tree generated after randomly reducing the number of B-genome markers distanced the *Ae. searsii* and *Ae. speltoides* from the other two *Sitopsis* species (see more details below). Additional support for MITEs activity came from the analysis of MITE insertions in wild emmer populations, as ~40% of the tested insertions showed polymorphism in at least one population. This data, together with the finding that MITEs are active in synthetic wheat allopolyploids (unpublished data), provide evidence for the transpositional activity of MITEs throughout the evolution of the *Triticum-Aegilops* group as well as in recent lineages of *durum* and modern wheat. Note that because of the limited number of MITE insertions that were tested in wild emmer populations, which resulted in insignificant differences among populations (see Supplemental Fig. 3), we cannot draw a general conclusion regarding the associations of MITE proliferation with environment.

What can MITEs tell us about the origin of the B-genome of bread wheat?

Although it is generally accepted that *Ae. speltoides* was the B-genome donor (Riley et al. 1958), there is much evidence to the contrary (Kimber 1966, 1974; Sears 1969; Johnson 1972), and thus the exact identity of the B-genome donor is still an enigma. Studies that are more recent have reported *Ae. searsii* (termed *T. searsii*) as the B-genome donor (Nath et al. 1983, 1984). Feldman and Levy (2005) provided some explanations on the ambiguous nature of the B-genome, which include: (1) the diploid donor of the B-genome being extant but not yet discovered; (2) the diploid donor of the B-genome undergoing massive genomic changes after the formation of the allotetraploid; and (3) the B-genome of wheat evolving through introgression of chromosomal segments from other allopolyploid or diploid species. Our data clearly show that among the four *Sitopsis* species that were used in this study, *Ae. searsii* was always clustered together with *Ae. speltoides*, while *Ae. longissima* was clustered together with *Ae. sharonensis* (Fig. 4). The clusters that were observed based on MITE markers showed ~90% genetic similarity between *Ae. longissima* and *Ae. sharonensis*, while ~75% genetic similarity was seen between *Ae. searsii* and *Ae. speltoides*. These data together with our unpublished data, which show that some B-genome sequences in hexaploid wheat were unique to *speltoides*, while others were unique to *searsii*, suggest that both *searsii* and *speltoides* are the best candidates to contribute the B-genome to wheat, and that it is hard to favorite one of the two species to be the donor of the B-genome. Further support of this claim came from the majority rule consensus phylogenetic tree (Supplemental Fig. 4), where it can clearly be seen that because the majority (27 out of the 28) of the B-derived markers where generated from *T. aestivum* or *T. turgidum* genomic libraries, only the *Ae. searsii* and *Ae. speltoides* cluster was out-grouped from the other clusters in the tree. This strongly indicates the relationship between *Ae. searsii*, *Ae. speltoides* and polyploid wheat. In support of this finding, Salse et al. (2008) have recently reported that *Ae. speltoides* itself cannot be the only contributor of the B-genome of wheat. As there is no evidence for the existence of a diploid species that might have donated the B-genome, we hypothesize, based on this study, that the B-genome donor might have been ancestral to *Ae. searsii* and *Ae. speltoides*, and that this divergence occurred after the formation of the allotetraploid.

In summary, this study shows that MITEs are potential evolutionary markers in wheat and might provide new insights into the origin of the B-genome. The enormous TE content and activity throughout wheat evolution might provide new insights into the study of wheat biodiversity. Future studies on TE composition in *Triticum* and *Aegilops* species, and more specifically in *Ae. searsii* and *Ae. speltoides*, will probably shed more light on the evolutionary history of wheat.

Acknowledgments We would like to thank Moshe Feldman and Sergei Volis for providing some of the plant material, and Avi Levy for helpful discussions. This work was supported by a grant from the Israeli Science Foundation (grant # 142/08) to K. K.

References

Bureau TE, Wessler SR (1994a) Mobile inverted-repeat elements of the tourist family are associated with the genes of many cereal grasses. Proc Natl Acad Sci USA 91(4):1411–1415

Bureau TE, Wessler SR (1994b) Stowaway—a new family of inverted repeat elements associated with the genes of both monocotyledonous and dicotyledonous plants. Plant Cell 6(6):907–916

Choulet F, Wicker T, Rustenholz C, Paux E, Salse J, Leroy P, Schlub S, Le Paslier MC, Magdelenat G, Gonthier C, Couloux A, Budak H, Breen J, Pumphrey M, Liu SX, Kong XY, Jia JZ, Gut M, Brunel D, Anderson JA, Gill BS, Appels R, Keller B, Feuillet C (2010) Megabase level sequencing reveals contrasted organization and evolution patterns of the wheat gene and transposable element spaces. Plant Cell 22(6):1686–1701

Clarke KR (1993) Nonparametric multivariate analyses of changes in community structure. Aust J Ecol 18(1):117–143

Cloutier S, McCallum BD, Loutre C, Banks TW, Wicker T, Feuillet C, Keller B, Jordan MC (2007) Leaf rust resistance gene Lr1, isolated from bread wheat (*Triticum aestivum* L.) is a member of the large psr567 gene family. Plant Mol Biol 65(1–2):93–106

Dubcovsky J, Dvorak J (2007) Genome plasticity a key factor in the success of polyploid wheat under domestication. Science 316(5833):1862–1866

Fahima T, Sun GL, Beharav A, Krugman T, Beiles A, Nevo E (1999) RAPD polymorphism of wild emmer wheat populations, *Triticum dicoccoides*, in Israel. Theor Appl Genet 98(3–4):434–447

Fahima T, Roder MS, Wendehake K, Kirzhner VM, Nevo E (2002) Microsatellite polymorphism in natural populations of wild emmer wheat, *Triticum dicoccoides*, in Israel. Theor Appl Genet 104(1):17–29

Feldman M, Levy AA (2005) Allopolyploidy—a shaping force in the evolution of wheat genomes. Cytogenet Genome Res 109(1–3): 250–258

Feschotte C, Pritham EJ (2007) DNA transposons and the evolution of eukaryotic genomes. Annu Rev Genet 41:331–368

Flavell AJ, Knox MR, Pearce SR, Ellis THN (1998) Retrotransposon-based insertion polymorphisms (RBIP) for high throughput marker analysis. Plant J 16(5):643–650

Gaut BS (2002) Evolutionary dynamics of grass genomes. New Phytol 154(1):15–28

Haider N, Nabulsi I (2008) Identification of *Aegilops* L. species and *Triticum aestivum* L. based on chloroplast DNA. Genet Resour Crop Evol 55(4):537–549

Hedges DJ, Callinan PA, Cordaux R, Xing J, Barnes E, Batzer MA (2004) Differential Alu mobilization and polymorphism among the human and chimpanzee lineages. Genome Res 14(6):1068–1075

Huang SX, Sirikhachornkit A, Faris JD, Su XJ, Gill BS, Haselkorn R, Gornicki P (2002) Phylogenetic analysis of the acetyl-CoA carboxylase and 3-phosphoglycerate kinase loci in wheat and other grasses. Plant Mol Biol 48(5):805–820

Isidore E, Scherrer B, Chalhoub B, Feuillet C, Keller B (2005) Ancient haplotypes resulting from extensive molecular rearrangements in the wheat A genome have been maintained in species of three different ploidy levels. Genome Res 15(4):526–536

Jiang N, Bao ZR, Zhang XY, Hirochika H, Eddy SR, McCouch SR, Wessler SR (2003) An active DNA transposon family in rice. Nature 421(6919):163–167

Jiang N, Feschotte C, Zhang XY, Wessler SR (2004) Using rice to understand the origin and amplification of miniature inverted repeat transposable elements (MITEs). Curr Opin Plant Biol 7(2):115–119

Johnson BL (1972) Protein electrophoretic profiles and the origin of the B genome of wheat. Proc Natl Acad Sci USA 69(6): 1398–1402

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

Kazazian HH (2004) Mobile elements: drivers of genome evolution. Science 303(5664):1626–1632

Kellogg EA (2001) Evolutionary history of the grasses. Plant Physiol 125(3):1198–1205

Kikuchi K, Terauchi K, Wada M, Hirano HY (2003) The plant MITE mPing is mobilized in anther culture. Nature 421(6919):167–170

Kimber G (1966) Estimate of the number of genes involved in the genetic suppression of the cytological diploidisation of wheat. Nature 212:317–318

Kimber G (1974) A reassessment of the origin of the polyploid wheats. Genetics 78(1):487–492

Konovalov FA, Goncharov NP, Goryunova S, Shaturova A, Proshlyakova T, Kudryavtsev A (2010) Molecular markers based on LTR retrotransposons BARE-1 and Jeli uncover different strata of evolutionary relationships in diploid wheats. Mol Genet Genomics 283(6):551–563

Kudryavtsev AM, Martynov SP, Broggio M, Buiatti M (2004) Evaluation of polymorphism at microsatellite loci of spring durum wheat (*Triticum durum* Desf.) varieties and the use of SSR-based analysis in phylogenetic studies. Russ J Genet 40(10):1102–1110

Lyons M, Cardle L, Rostoks N, Waugh R, Flavell AJ (2008) Isolation, analysis and marker utility of novel miniature inverted repeat transposable elements from the barley genome. Mol Genet Genomics 280(4):275–285

Miller AK, Galiba G, Dubcovsky J (2006) A cluster of 11 CBF transcription factors is located at the frost tolerance locus Fr-A(m)2 in *Triticum monococcum*. Mol Gen Genomics 275(2):193–203

Mori N, Liu YG, Tsunewaki K (1995) Wheat phylogeny determined by Rflp analysis of nuclear-DNA.2. Wild tetraploid wheats. Theor Appl Genet 90(1):129–134

Naito K, Zhang F, Tsukiyama T, Saito H, Hancock CN, Richardson AO, Okumoto Y, Tanisaka T, Wessler SR (2009) Unexpected consequences of a sudden and massive transposon amplification on rice gene expression. Nature 461(7267):U1130–U1232

Nakazaki T, Okumoto Y, Horibata A, Yamahira S, Teraishi M, Nishida H, Inoue H, Tanisaka T (2003) Mobilization of a transposon in the rice genome. Nature 421(6919):170–172

Nath J, McNay JW, Paroda CM, Gulati SC (1983) Implication of *Triticum searsii* as the B-genome donor to wheat using DNA hybridizations. Biochem Genet 21(7–8):745–760

Nath J, Hanzel JJ, Thompson JP, McNay JW (1984) Additional evidence implicating *Triticum searsii* as the B-genome donor to wheat. Biochem Genet 22(1–2):37–50

Nesbitt M, Samuel D (1996) From staple crop to extinction? The archaeology and history of the hulled wheats. In: Hammer K, Heller J (eds) Hulled wheats proceedings of the first international workshop on hulled wheats promoting the conservation and use of underutilized and neglected crops 4, pp 41–100

Nevo E, Beiles A (1989) Genetic diversity of wild emmer wheat in Israel and Turkey—structure, evolution, and application in breeding. Theor Appl Genet 77(3):421–455

Nevo E, Golenberg E, Beiles A, Brown AHD, Zohary D (1982) Genetic diversity and environmental associations of wild wheat, *Triticum-dicoccoides*, in Israel. Theor Appl Genet 62(3):241–254

Petersen G, Seberg O, Yde M, Berthelsen K (2006) Phylogenetic relationships of *Triticum and Aegilops* and evidence for the origin of the A, B, and D genomes of common wheat (*Triticum aestivum*). Mol Phylogenet Evol 39(1):70–82

Prasad V, Stromberg CAE, Alimohammadian H, Sahni A (2005) Dinosaur coprolites and the early evolution of grasses and grazers. Science 310(5751):1177–1180

Queen RA, Gribbon BM, James C, Jack P, Flavell AJ (2004) Retrotransposon-based molecular markers for linkage and

genetic diversity analysis in wheat. Mol Gen Genomics 271(1):91–97

Riley R, Unrau J, Chapman V (1958) Evidence on the origin of the B genome of wheat. J Hered 49:90–98

Roy-Engel AM, Carroll ML, Vogel E, Garber RK, Nguyen SV, Salem AH, Batzer MA, Deininger PL (2001) Alu insertion polymorphisms for the study of human genomic diversity. Genetics 159(1):279–290

Sabot F, Simon D, Bernard M (2004) Plant transposable elements, with an emphasis on grass species. Euphytica 139(3):227–247

Sabot F, Guyot R, Wicker T, Chantret N, Laubin B, Chalhoub B, Leroy P, Sourdille P, Bernard M (2005) Updating of transposable element annotations from large wheat genomic sequences reveals diverse activities and gene associations. Mol Genet Gen 274(2):119–130

Salem AH, Ray DA, Xing J, Callinan PA, Myers JS, Hedges DJ, Garber RK, Witherspoon DJ, Jorde LB, Batzer MA (2003) Alu elements and hominid phylogenetics. Proc Natl Acad Sci USA 100(22):12787–12791

Salina EA, Lim KY, Badaeva ED, Shcherban AB, Adonina IG, Amosova AV, Samatadze TE, Vatolina TY, Zoshchuk SA, Leitch AR (2006) Phylogenetic reconstruction of *Aegilops* section Sitopsis and the evolution of tandem repeats in the diploids and derived wheat polyploids. Genome 49(8):1023–1035

Sallares R, Brown TA (2004) Phylogenetic analysis of complete 5' external transcribed spacers of the 18S ribosomal RNA genes of diploid *Aegilops* and related species (Triticeae, Poaceae). Genet Resour Crop Evol 51(7):701–712

Salse J, Chague V, Bolot S, Magdelenat G, Huneau C, Pont C, Belcram H, Couloux A, Gardais S, Evrard A, Segurens B, Charles M, Ravel C, Samain S, Charmet G, Boudet N, Chalhoub B (2008) New insights into the origin of the B genome of hexaploid wheat: evolutionary relationships at the SPA genomic region with the S genome of the diploid relative *Aegilops speltoides*. BMC Genomics 9:555

Sasanuma T, Miyashita NT, Tsunewaki K (1996) Wheat phylogeny determined by RFLP analysis of nuclear DNA.3. Intra- and interspecific variations of five *Aegilops sitopsis* species. Theor Appl Genet 92(8):928–934

Sears ER (1954) The aneuploids of common wheat. Res Bull Univ Missouri Agric Exp Stn 572:1–59

Sears ER (1969) Wheat cytogenetics. Annu Rev Genet 3:451–468

Shan XH, Liu ZL, Dong ZY, Wang YM, Chen Y, Lin XY, Long LK, Han FP, Dong YS, Liu B (2005) Mobilization of the active MITE transposons mPing and Pong in rice by introgression from wild rice (*Zizania latifolia* Griseb.). Mol Biol Evol 22(4):976–990

Teoh SB, Miller TE, Reader SM (1983) Intraspecific variation in C-banded chromosomes of *Aegilops-Comosa* and *A. Speltoides*. Theor Appl Genet 65(4):343–348

Wang C, Shi SH, Wang JB, Zhong Y (2000a) Phylogenetic relationships of diploid species in *Aegilops* inferred from the ITS sequences of nuclear ribosomal DNA. Acta Botanica Sinica 42(5):507–511

Wang GZ, Matsuoka Y, Tsunewaki K (2000b) Evolutionary features of chondriome divergence in *Triticum* (wheat) and *Aegilops* shown by RFLP analysis of mitochondrial DNAs. Theor Appl Genet 100(2):221–231

Wang JB, Wang C, Shi SH, Zhong Y (2000c) ITS regions in diploids of *Aegilops* (Poaceae) and their phylogenetic implications. Hereditas 132(3):209–213

Wicker T, Stein N, Albar L, Feuillet C, Schlagenhauf E, Keller B (2001) Analysis of a contiguous 211 kb sequence in diploid wheat (*Triticum monococcum* L.) reveals multiple mechanisms of genome evolution. Plant J 26(3):307–316

Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O, Paux E, SanMiguel P, Schulman AH (2007) A unified classification system for eukaryotic transposable elements. Nat Rev Genet 8(12):973–982

Witte CP, Le QH, Bureau T, Kumar A (2001) Terminal-repeat retrotransposons in miniature (TRIM) are involved in restructuring plant genomes. Proc Natl Acad Sci USA 98(24):13778–13783

Xing J, Wang H, Zhang Y, Ray DA, Tosi AJ, Disotell TR, Batzer MA (2007a) A mobile element-based evolutionary history of guenons (tribe Cercopithecini). BMC Biol 5:5

Xing J, Witherspoon DJ, Ray DA, Batzer MA, Jorde LB (2007b) Mobile DNA elements in primate and human evolution. Yearb Phys Anthropol 50(50):2–19

Yang GJ, Nagel DH, Feschotte C, Hancock CN, Wessler SR (2009) Tuned for transposition: molecular determinants underlying the hyperactivity of a Stowaway MITE. Science 325(5946):1391–1394