Theor Appl Genet (2004) 109: 1093–1104
DOI 10.1007/s00122-004-1726-5

ORIGINAL PAPER

W. Li · Y. Wan · Z. Liu · K. Liu · X. Liu · B. Li · Z. Li · X. Zhang · Y. Dong · D. Wang

Molecular characterization of HMW glutenin subunit allele *1Bx14*: further insights into the evolution of *Glu-B1-1* alleles in wheat and related species

Received: 22 December 2003 / Accepted: 6 May 2004 / Published online: 29 July 2004

Abstract 1Bx14 is a member of the high molecular weight (HMW) glutenin subunits specified by wheat *Glu-B1-1* alleles. In this work, we found that the full-length amino acid sequence of 1Bx14 derived from cloned coding region was similar, but not identical, to that of 1Bx20. In the N-terminal domains of 1Bx14 and 1Bx20, the last two of the three cysteine residues, which are conserved in 1Bx7, 1Bx17 and homoeologous 1Ax and 1Dx subunits, were replaced by tyrosine residues. In the 5′ flanking regions (−900 to −1,200 bp relative to the start codon), a novel miniature inverted-repeat transposable element insertion was present in *1Bx14* and *1Bx20* but not *1Bx7* and *1Bx17*. *1Bx14* and *1Bx20* like alleles were readily found in tetraploid wheat subspecies but not several S genome containing *Aegilops* species. Phylogenetic analysis showed that the four molecularly characterized *Glu-B1-1* alleles (*1Bx7*, *1Bx14*, *1Bx17*, *1Bx20*) could be divided into two allelic lineages. The lineage represented by *1Bx7* and *1Bx17* was more ancient than the one represented by *1Bx14* and *1Bx20*. Combined, our data establish that *1Bx14* and *1Bx20* represent a novel subclass of *Glu-B1-1* alleles. Based on current knowledge, potential mechanism involved in the differentiation of two *Glu-B1-1* lineages is discussed.

Communicated by F. Salamini

Electronic Supplementary Material Supplementary material is available for this article at http://dx.doi.org/10.1007/s00122-004-1726-5

W. Li · Y. Wan · Z. Liu · K. Liu · X. Liu · B. Li · Z. Li · D. Wang (✉)
Institute of Genetics and Developmental Biology, Chinese Academy of Sciences,
Beijing, 100101, China
e-mail: dwwang@genetics.ac.cn
Tel.: +86-10-64889380
Fax: +86-10-64854467

X. Zhang · Y. Dong
Institute of Crop Germplasm Research, Chinese Academy of Agricultural Sciences,
Beijing, 100081, China

Introduction

In bread wheat, the processing quality is largely determined by the composition of a group of seed storage proteins (named glutenin subunits) that are deposited in the wheat grains (Lawrence and Shepherd 1981; Shewry et al. 1995). The high molecular weight (HMW) glutenin subunits are encoded by the *Glu-1-1* (*x* type) and *Glu-1-2* (*y* type) genes contained in the *Glu-1* locus (Lawrence and Shepherd 1981; Payne 1987). Since 1970 s, a large body of evidence from genetic, biochemical and transgenic studies has shown conclusively that HMW glutenin subunits are the major determinants of the processing quality of wheat grains (Shewry and Halford 2002). In addition to wheat, orthologous HMW glutenin subunits have also been found in *Aegilops* and rye species (William et al. 1993; Wan et al. 2000, 2002; De Bustos et al. 2001; Liu et al. 2003). In barley, the D-hordeins are structurally related to HMW glutenin subunits (Halford et al. 1992).

Extensive comparisons of HMW glutenin subunit genes have shown that the genomic space occupied by the *x* or *y* type genes may be below 10 kb (Anderson et al. 2002). Within the 2.5-kb region upstream of the coding sequence of well characterized HMW glutenin subunit genes (e.g., *1Ax2**, *1Ay*, *1Bx7*, *1Dx5* and *1Dy10*), there are usually a conserved matrix attachment region (MAR) and the regulatory sequences required for gene transcription (Anderson et al. 2002; Rampitsch et al. 2000). Several types of miniature inverted-repeat transposable elements (MITEs) have been found within the MAR elements although the relative positions of the MITE insertions differ among different subunit genes. *Waffle* was inserted in the MAR element in *1Ax2** and *1Dx5*, whereas *Stowaway-Ta3* was found present in the MAR element of *1Ay* (Anderson et al. 2002). In another investigation, an uncharacterized MITE sequence was found in the 5′ flanking region of *1Bx20* from a durum wheat variety (Anderson et al. 1998). Compared to the 5′ flanking region, the non-coding sequence immediately after the 3′ end of the open reading frame (ORF) of a HMW glutenin subunit gene is usually short (400–500 bp) (Anderson et

al. 2002). At the protein level, the primary structures of *x* and *y* types of HMW glutenin subunits are similar and composed of a signal peptide (removed from the protein in mature seeds), an N-terminal domain, a central repetitive domain and a C-terminal domain (Shewry et al. 1995). In either *x* or *y* type of HMW glutenin subunits, the cysteine residues are usually conserved in both numbers and positions (Shewry et al. 1995). However, in the recently characterized 1Bx20 subunit, the last two of the three cysteine residues in the N-terminal domain, are replaced by tyrosine residues (Shewry et al. 2003). Both the conserved cysteine residues and the size of the repetitive domains contribute to the high order structure of HMW glutenin subunits. The former is involved in the formation of inter- or intramolecular disulphide bonds; the latter may promote intermolecular interactions through hydrogen bonding (Shewry et al. 2002). In this respect, it would be very interesting to identify more 1Bx20 like subunits and to study the effect of the reduction in conserved cysteine residues on their high order structures and their function in the end use qualities of wheat grains.

Compared to above studies, fewer investigations have been conducted on molecular evolution of HMW glutenin subunit genes. The primary structure of a HMW glutenin subunit may originally be formed by the triplication of an ancestral domain and the subsequent acquisition of a repetitive domain (Kreis et al. 1985). The later events leading to the formation of various *Glu-1* loci (such as *Glu-A1*, *Glu-B1* and *Glu-D1*) and the multiple alleles of a given locus have not been specifically addressed in past literatures. There is clear evidence for the evolution of novel allelic subunit through change in the number of the conserved cysteine residues. For example, the 1Dx5 is a novel subunit conferring good processing properties in bread wheat varieties because of the presence of a cysteine residue in its repetitive domain (Shewry et al. 1995; Anderson et al. 1989). Based on phylogenetic analysis using a conserved region (241–243 bp) located immediately upstream of the start codon of HMW glutenin subunit genes, researchers have estimated that the *Glu* gene duplication event (i.e., the differentiation of *x* and *y* types of HMW glutenin subunit genes) occurred 7.2–10.0 MYA (Allaby et al. 1999). The origin of A, B and D genomes (and hence the differentiation of the *Glu-A1*, *Glu-B1* and *Glu-D1* loci) may be dated 5.0–6.9 MYA (Allaby et al. 1999). The *Glu-B1-1* alleles from cultivated wheats might be divided into two subgroups that diverged 1.4–2.0 MYA (Allaby et al. 1999). *1Bx7*, *1Bx17* and several alleles amplified from an archaeological wheat sample were contained in one subgroup. However, the identities of the alleles contained in the other subgroup were not clear owing to the lack of sufficient sequence information.

The objectives of the studies reported in this paper are to characterize novel HMW glutenin subunit gene alleles and to investigate the evolutionary biology of this important group of genes. In the sections below, we describe molecular characterization of 1Bx14 and its gene and further insights into the evolutionary biology of *Glu-B1-1* alleles based on our results and those published previously.

Materials and methods

Plant materials

The plant materials used in this study included hexaploid wheat varieties (Xiaoyan 54, Chinese Spring, Bobwhite, L88-6 and L86-69), tetraploid wheats (*Triticum turgidum* ssp. *dicoccoides*, *T. turgidum* ssp. *dicoccon*, *T. turgidum* ssp. *turgidum*, two accessions for each subspecies), and S genome containing *Aegilops* species (*Ae. speltoides*, six accessions; *Ae. searsii*, seven accessions; *Ae. longissima*, three accessions; *Ae. bicornis*, two accessions). L88-6 and L86-69 were both derived from the crosses between the mutant lines of Olympic and Gabo (Lawrence et al. 1988; Reddy and Appels 1993; Barro et al. 1997). The composition of HMW glutenin subunits in the two lines was 1Ax1, 1Bx17, 1By18, 1Dx5 and 1Dy10 (Reddy and Appels 1993; Barro et al. 1997).

SDS-PAGE and N-terminal protein sequencing

High molecular weight glutenin subunits were preferentially extracted from seed materials and were separated using SDS-PAGE as described elsewhere (Wan et al. 2000). The four HMW glutenin subunits in Chinese Spring (1Bx7, 1By8, 1Dx2, 1Dy12) were used as electrophoretic mobility standards in SDS-PAGE analysis. Using SDS-PAGE, the composition of HMW glutenin subunits in Xiaoyan 54 had previously been found to be 1Ax1, 1Bx14, 1By15, 1Dx2 and 1Dx12 (Zhang et al. 2002). To verify this finding we analyzed HMW glutenin subunits of Xiaoyan 54 by N-terminal protein sequencing as described previously (Wan et al. 2002).

Cloning and bacterial expression of the complete ORF of *1Bx14*

Genomic DNA was extracted from the etiolated seedlings of Xiaoyan 54 as described previously (Wan et al. 2002). For amplifying the complete coding sequence of *1Bx14* from Xiaoyan 54 using genomic PCR, a pair of degenerate primers (P1 and P2, Supplementary Fig. S1 and Table S1) was designed according to the nucleotide sequences conserved in the 5′ or 3′ ends of the ORFs of published HMW glutenin subunit genes (Liu et al. 2003). The cycling parameters for the genomic PCR were the same as those reported previously (Liu et al. 2003). The amplified products included four fragments ranging from approximately 1.8–2.5 kb. Three fragments (whose size was about 1.9, 2.4 and 2.5 kb, respectively) were separately purified and cloned into pGEM-T Easy vector (Promega). By restriction enzyme digestion mapping and partial DNA sequencing, the 1.9 and 2.5 kb fragments were found to represent the ORFs of *1Dy12* and *1Dx2*, respectively. The 5′ and 3′ sequences of the 2.4 kb fragment (in four independent plasmid clones) were highly similar to those of *1Bx20*. This fragment was deduced to represent the ORF of *1Bx14* in Xiaoyan 54 and was completely sequenced. Potential mistakes brought about by genomic PCR using degenerate primers (P1 and P2) were corrected by additional PCR experiments amplifying the sequences flanking the 5′ or 3′ ends of the 2.4 kb fragment (see below).

To confirm the cloned 2.4-kb fragment as the coding sequence of *1Bx14*, a set of bacterial expression experiments was conducted. The 2.4-kb fragment was reamplified using primers P3 and P4 (Supplementary Fig. S1 and Table S1) in order to remove the coding sequence for the signal peptide and to introduce restriction enzyme sites for subsequent cloning work. The reamplified fragment was cloned into the bacterial expression vector pET-30a (Invitro-

gen). The resultant expression construct pET-1Bx14 was induced to express the mature protein of 1Bx14 in bacterial cells as detailed in a previous publication (Wan et al. 2002). The electrophoretic mobility of the 1Bx14 protein produced in the bacterial cells was compared to that of 1Bx14 extracted from the seeds of Xiaoyan 54 using SDS-PAGE.

Cloning and sequencing the 5′ and 3′ flanking sequences of *1Bx14* ORF

The 5′ flanking sequence of *1Bx14* ORF was amplified using genomic PCR (as described above) with primers P5 and P1Bx14/20 (Supplementary Fig. S1 and Table S1). P5 was designed based on a sequence element that was strictly conserved in the 5′ flanking regions of *1Ax2**, *1Bx7*, *1Bx20* and *1Dx5*, whereas P1Bx14/20 was derived from the sequence coding for the six amino acid residues (ITVSPG) in the N-terminal domains of 1Bx14 and 1Bx20. The 3′ flanking sequence of *1Bx14* ORF was amplified by genomic PCR using primers P1Bx14 and P6 (Supplementary Fig. S1 and Table S1). P1Bx14 was derived from the sequence encoding the seven amino acid residues (AMCRLEG) in the C-terminal domain of 1Bx14. P6 was designed based on a sequence element strictly conserved in the 3′ flanking regions of *1Ax2**, *1Bx7* and *1Dx5*. Taken together, the nucleotide sequence of *1Bx14* determined in this study was 4,021 bp (GenBank accession AY367771). During above experiments, the desired PCR fragments were purified, cloned in the pGEM-T Easy vector, and were sequenced from both strands. The final nucleotide sequences for the 5′ or 3′ regions of *1Bx14* were each constructed based on the sequencing results of three independent clones.

Molecular analysis of the MITE insertion in the 5′ flanking region of *1Bx14*

To determine the copy numbers of the *Tripper* element in bread wheat, Southern hybridization experiments were conducted using genomic DNA samples of Xiaoyan 54 and Chinese Spring. Genomic DNA samples were digested with either *Hind*III or *Nsi*I, separated in agarose gels, transferred onto nylon membrane, and hybridized using a *Tripper* specific probe (Sambrook et al. 1989). The ^{32}P-labeled probe was prepared using the DNA fragment of *Tripper* (238 bp, amplified by PCR using primers P7 and P8, Supplementary Fig. S1 and Table S1) and the RadPrime DNA Labeling System (Invitrogen). To assess the influence of *Tripper* insertion on the transcription directed by the 5′ flanking region of *1Bx14*, two expression constructs (pM1Bx14PR-GUS, p1Bx14PR-GUS) were prepared. A DNA fragment containing *Tripper* (−1,140~+3, relative to the start codon) was amplified by PCR with primers P9 and P10 (Supplementary Fig. S1 and Table S1). The amplified fragment, after digestion with *Kpn*I and *Nco*I, was cloned into the pJIT166 vector (http://www.pgreen.ac.uk) that had previously been digested with the same enzymes. The resulted construct pM1Bx14PR-GUS was used to investigate the presence of *Tripper* insertion on the expression of the *GUS* marker gene. A DNA fragment lacking *Tripper* (−868~+3, relative to the start codon) was amplified using primers P10 and P11 (Supplementary Fig. S1 and Table S1). The resulted fragment was cloned into pJIT166 (as described above), giving rise to p1Bx14PR-GUS that was used to assess the absence of *Tripper* on the expression of the *GUS* marker gene. The two expression constructs were tested in a transient expression assay as described previously (Oñate et al. 1999). Briefly, gold particles coated with the DNA of the expression constructs were delivered into the endospermic tissues (20 tissues per bombardment) extruded out from the developing seeds of the bread wheat variety Bobwhite at 12–14 days after flowering using the PDS-1000/HE system (BIO-RAD). The bombardment was repeated three times for each construct. The results of the bombardment experiments were calculated as the mean numbers of GUS spots per endospermic tissue segment plus standard deviations.

Detection of potential *1Bx14* and *1Bx20* like alleles in *Aegilops* species and tetraploid wheat subspecies

The existence of potential *1Bx14* and *1Bx20* like alleles in four diploid, S genome containing *Aegilops* species and three tetraploid wheat subspecies was investigated using genomic PCR with primers P5 and P1Bx14/20 (Supplementary Fig. S1 and Table S1). For successful amplifications, the desired fragments were cloned and sequenced (as described above). The nucleotide sequences of the cloned fragments were constructed using sequence information derived from at least three independent clones.

Reinvestigation of nucleotide sequence of *1Bx17*

While carrying out the studies in this paper, we found that the nucleotide sequence for a part of the 5′ flanking region (965 bp upstream of the start codon) of the *1Bx17* allele (from L86-69, designated here as *1Bx17-86*) reported previously (Reddy and Appels 1993) differed, unexpectedly, in several locations from those of the homologous regions in *1Ax2**, *1Bx7*, *1Bx20*, and *1Dx5* (Supplementary Fig. S2). This prompted us to reinvestigate the nucleotide sequence of *1Bx17*. Using the primers P5 and P1Bx17 (Supplementary Table S1), we amplified a DNA fragment of about 1.25 kb from L88-6, which would cover a part of the 5′ flanking region of *1Bx17* and the entire sequence encoding the N-terminal domain of 1Bx17 protein. This fragment was subsequently cloned and sequenced. Compared to *1Bx17-86*, the sequence amplified from L88-6 (constructed from three independent clones, designated as *1Bx17-88*) was more similar to its orthologous sequences in *1Ax2**, *1Bx7*, *1Bx20*, and *1Dx5* (Supplementary Fig. S2). So for the DNA or protein alignments that involved the 5′ flanking region of *1Bx17* or the protein sequence of the N-terminal domain of 1Bx17 in this paper, we employed the sequences derived from our own investigations using L88-6 (Supplementary Figs. S2 and S3).

DNA and protein sequence analyses and evolutionary investigations

For multiple alignments of DNA or protein sequences, the ClustalW program (Thompson et al. 1994) was generally used. For maximizing the similarities among the repetitive domains of 1Bx7, 1Bx14, 1Bx17 and 1Bx20, some manual adjustment to the multiple alignment was required. For predicting potential secondary structure of the *Tripper* element, the MFOLD program (http://bioweb.pasteur.fr/seqanal/interfaces/mfold-simple.html) was employed with default options. To investigate the phylogenetic relationship of *1Bx14* and *1Bx20* with previously characterized *Glu-1-1* alleles (represented by *1Ax2**, *1Bx7, 1Bx17* and *1Dx5*), a multiple alignment was created with homologous nucleotide sequences (the 5′ flanking sequences plus the ones encoding the N-terminal domains) using the ClustalW program. This alignment file was converted to mega format at the MEGA website (Version 2, http://www.oup-usa.org/sc/0195135857) for building phylogenetic trees using neighbor joining, minimal evolution or parsimony programs (Nei and Kumar 2000).

For inferring the ancestral amino acid sequences of the N-terminal domains of Glu-1-1 subunits, the computer program ANCESTOR (Zhang and Nei 1997) was employed. For this purpose, a phylogenetic tree was constructed with the amino acid sequences of the N-terminal domains of 1Ax1, 1Ax2*, 1Bx7, 1Bx14, 1Bx17, 1Bx20, 1Dx2 and 1Dx5 (and the corresponding region of barley D-hordein as an outgroup) using the neighbor joining method (in the MEGA website). The topology of the phylogenetic tree and the aligned amino acid sequences were then used as input information

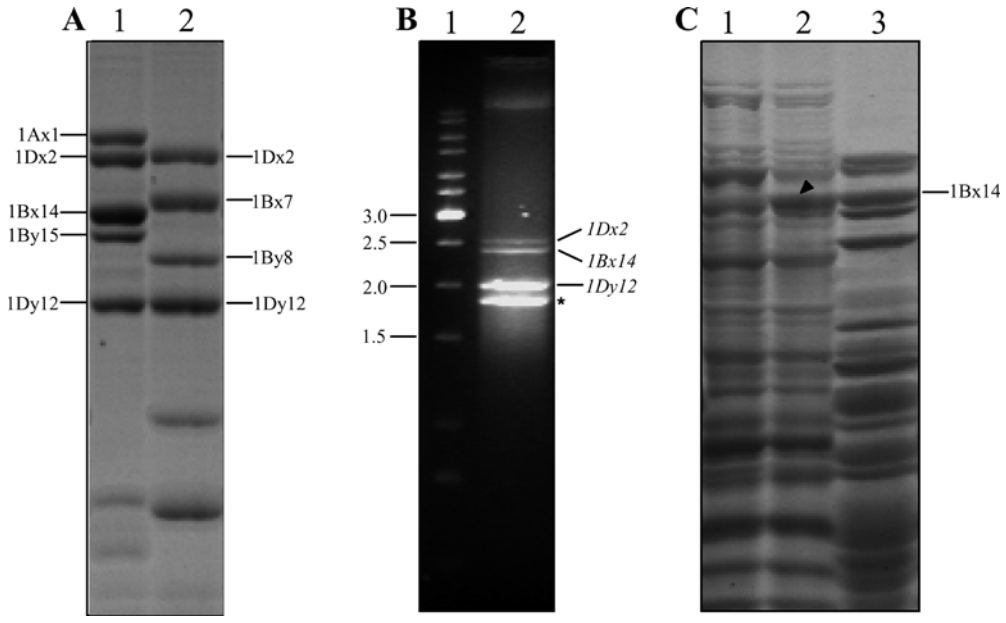

Fig. 1a–c Characterization of HMW glutenin subunits expressed in the hexaploid wheat variety Xiaoyan 54 by SDS-PAGE and molecular cloning of the complete coding region of the *Glu-B1-1* allele *1Bx14*. **a** SDS-PAGE analysis showed that five HMW glutenin subunits (1Ax1, 1Dx2, 1Bx14, 1By15 and 1Dy12) were expressed in Xiaoyan 54 (*lane 1*) as compared to the four subunits (1Dx2, 1Bx7, 1By8 and 1Dy12) expressed in the hexaploid wheat variety Chinese Spring (*lane 2*). **b** Amplification of the complete coding regions of HMW glutenin subunit genes in Xiaoyan 54 via genomic PCR using degenerate primers (P1 and P2). By cloning and sequencing analysis, the three fragments whose size was above 1.9 kb were found to represent the ORFs of *1Dx2*, *1Bx14* and *1Dy12*, respectively (*lane 2*). The fragment marked by an *asterisk* was not characterized because its size may be below that of a functional HMW glutenin subunit gene ORF. The DNA markers (kb) were contained in *lane 1*. **c** Bacterial expression of *1Bx14* coding sequence. The expression of 1Bx14 mature protein (*lane 2*, indicated by an *arrowhead*) was detected in the IPTG induced bacterial culture (*lane 2*). In contrast, overexpression of 1Bx14 mature protein was not observed in the control bacterial culture that was not induced by IPTG (*lane 1*). The bacterially expressed 1Bx14 mature protein showed an electrophoretic mobility identical to that of 1Bx14 subunit extracted from the seeds of Xiaoyan 54 (*lane 3*)

for computing the ancestral amino acid sequences at the different nodes of the phylogenetic tree.

To estimate the divergence time between *1Bx7* and *1Bx14* allelic lineages, the genomic sequence (5′ flanking sequence plus the one encoding the signal peptide and the N-terminal domain) of *1Bx14* (1,468 bp) was aligned to its homologous sequences in *1Bx7* (1,269 bp) or *1Bx17* (1,268 bp). The divergence time was calculated as described previously (Sanderson 1998). We used the average nucleotide substitution rate of 6.5×10^{-9} per site per year calculated for barley *ADH* genes (Gaut et al. 1996). This substitution rate has recently been used successfully for estimating the divergence time between two low molecular glutenin subunit genes in *T. monococcum* (Wicker et al. 2003).

For generating the alignments described in above analyses, appropriate DNA (or protein) sequences of *1Ax1*, *1Ax2**, *1Bx7*, *1Bx17*, *1Bx20*, *1Dx2*, *1Dx5* and the barley D-hordein gene were retrieved from the GenBank. The EMBL accession numbers for *1Ax1*, *1Ax2**, *1Bx7*, *1Bx20*, *1Dx2*, *1Dx5* and the barley D-hordein gene are X61009, M22208, X13927, AJ437000, X03346, X12928, and AY268139, respectively. The accession number of the 1Bx17 protein sequence (Reddy and Appels 1993) is JC2099. The nucleotide sequence for the 5′ flanking region of *1Bx20* was derived from a previous publication (Anderson et al. 1998).

Results

N-terminal protein sequencing of 1Bx14 and homoeologous subunits in Xiaoyan 54

In the bread wheat variety Xiaoyan 54, the complement of expressed HMW glutenin subunits has previously been identified to be 1Ax1, 1Bx14, 1By15, 1Dx2 and 1Dy12 based on electrophoretic mobility comparisons (Zhang et al. 2002, Fig. 1a). We tried to confirm the expression of 1Ax1, 1Bx14 and 1By15 in Xiaoyan 54 by direct protein sequencing. The 18 residues obtained for 1Ax1 (EGEASGQLQCERELQEHS) were indeed identical to those in the published 1Ax1 subunit. The 20 residues obtained for 1Bx14 (EGEASGQLQCERELRKRELE) were identical to those in the previously reported 1Bx20 subunit. The 15 residues found for 1By15 (EGEASRQLQ-CERELQ) were identical to those in 1By9, 1Dy10 and 1Dy12.

Characterization of *1Bx14* coding sequence and the primary structure of deduced 1Bx14 protein

Among the several molecularly characterized *x* type HMW glutenin subunits, 1Bx 20 is unusual in that its N-terminal domain contains only one conserved cysteine residue

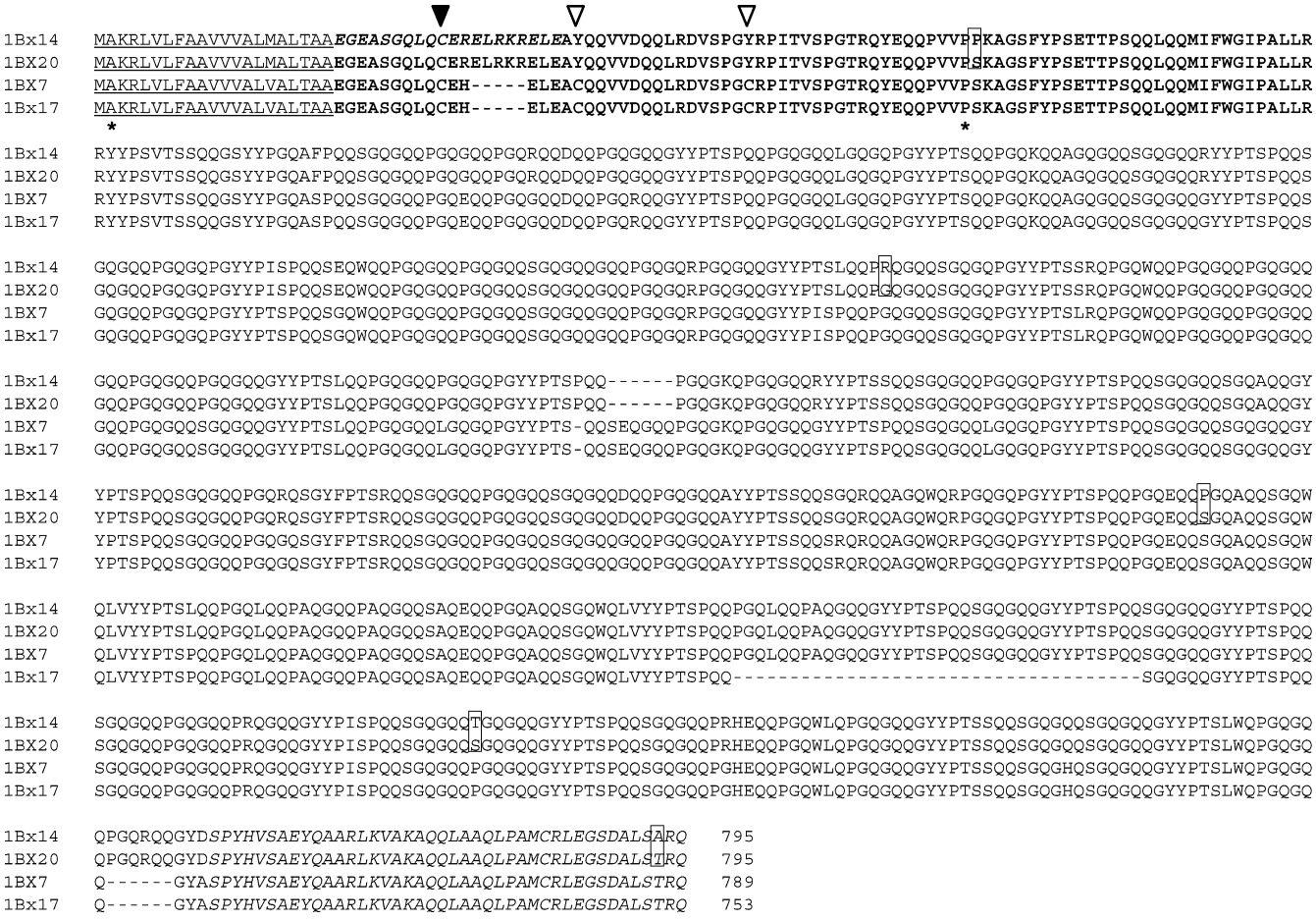

Fig. 2 A comparison of the amino acid sequences of four Glu-B1-1 subunits (1Bx14, 1Bx20, 1Bx7 and 1Bx17). This comparison shows that the four subunits possess an identical primary structure composed of signal peptide (*underlined region*), N-terminal domain (represented by *bold letters*), C-terminal domain (represented by *italicized letters*), and the central repetitive domain situated in between the N and C-terminal domains. The *filled arrowhead* indicates the cysteine residue that is strictly conserved among the four subunits. The *empty arrowheads* mark the two locations where the cysteine residues that are conserved in 1Bx7 and 1Bx17 are replaced by tyrosine residues in 1Bx14 and 1Bx20. The *boxed regions* indicate the five differences (caused by substitutions of single residues) between the amino acid sequences of 1Bx14 and 1Bx20. In the 1Bx17 sequence determined previously (Reddy and Appels 1993), the two locations marked by *asterisks* were occupied by "T" and "A", respectively. According to the results of our own investigation on 1Bx17 in this work (Supplementary, Fig. S3B), the residues in the two marked locations have now been changed into "A" and "P", respectively. The first 20 residues of the deduced 1Bx14 mature protein (represented by *italicized*, *bold letters*) are identical to those determined by direct protein sequencing

(Shewry et al. 2003). The finding on the N-terminal sequence of 1Bx14 in the protein sequencing experiment prompted us to further investigate if this subunit would be similar to 1Bx20 in possessing only one conserved cysteine residue in its N-terminal domain. We amplified the complete ORF of the *1Bx14* gene using primers P1 and

Table 1 Some properties of the mature protein of 1Bx14 compared to those of previously reported 1Ax, 1Bx and 1Dx subunits

	Number of amino acid residues				Number of cysteine residues			
	N-terminal domain	Repetitive domain	C-terminal domain	Total	N-terminal domain	Repetitive domain	C-terminal domain	Total
1Ax1	86	681	42	809	3	0	1	4
1Ax2*	86	666	42	794	3	0	1	4
1Bx7	81	645	42	768	3	0	1	4
1Bx17	81	609	42	732	3	0	1	4
1Bx14	86	646	42	774	1	0	1	2
1Bx20	86	646	42	774	1	0	1	2
1Dx2	88	687	42	817	3	0	1	4
1Dx5	89	687	42	818	3	1	1	5

The unprocessed *x* type HMW glutenin subunits contain signal peptides, which are removed from the mature proteins after targeting to the protein bodies

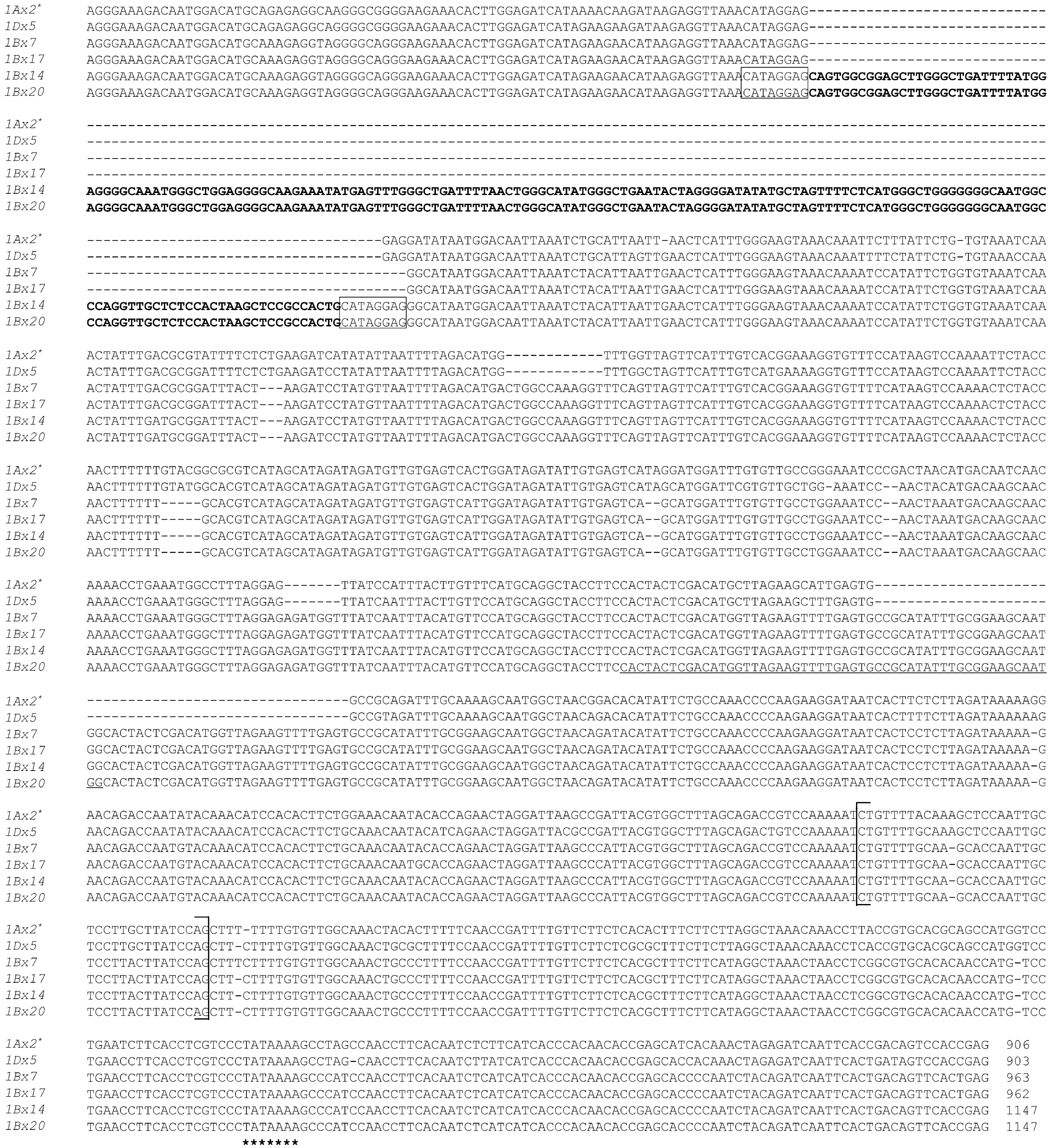

Fig. 3 A comparison of the 5′ flanking sequences of four *Glu-B1-1* alleles (*1Bx7*, *1Bx17*, *1Bx14*, *1Bx20*) with homologous sequences from representative *Glu-A1-1* (*1Ax2**) and *Glu-D1-1* (1Dx5) alleles. The MITE elements (*Tripper*) present in the 5′ flanking regions of *1Bx14* and *1Bx20* are represented by **bold letters**. The target site duplications caused by the *Tripper* element are *boxed*. The *underlined* sequence element (54 bp) is tandemly duplicated in all four *Glu-B1-1* alleles. However, only a part of this element is present in the 5′ flanking regions of *1Ax2** and *1Dx5* and it is not tandemly duplicated. The region in *brackets* is the enhancer element conferring seed specific expression of HMW glutenin subunit genes. The TATA box is marked by *asterisks*

P2 (Fig. 1b). The size of *1Bx14* ORF was 2,391 bp (including the six nucleotides coding for the tandem stop codons at the end of the ORF). When expressed in bacterial cells, the cloned *1Bx14* coding sequence yielded a polypeptide showing an electrophoretic mobility identical to that of 1Bx14 subunit extracted from the seeds (Fig. 1c), indicating that the cloned sequence was an accurate representation of the *1Bx14* ORF. The amino acid

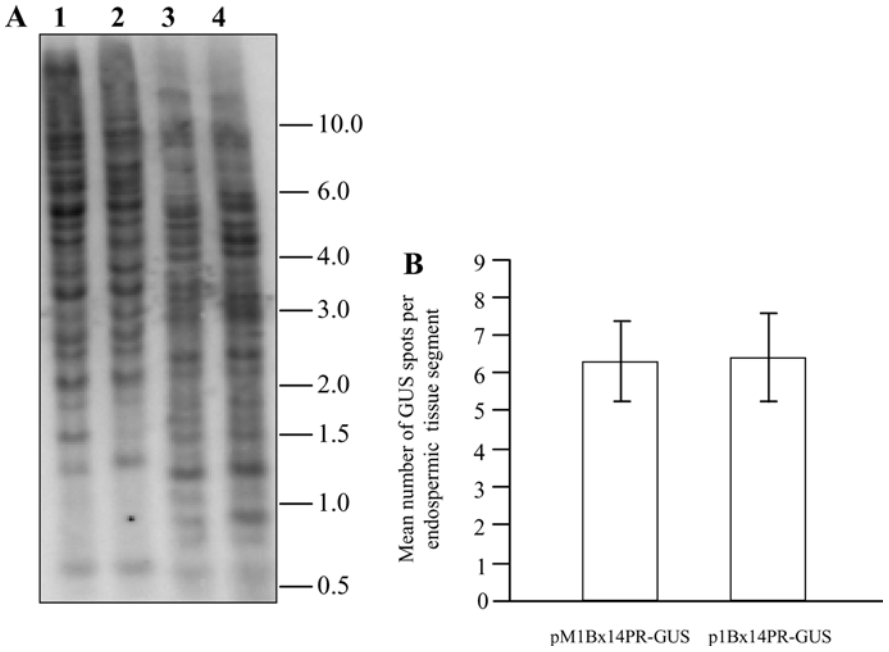

Fig. 4a, b Molecular analysis of the *Tripper* element present in the 5′ flanking region of *1Bx14*. **a** Southern hybridization analysis, using *Nsi*I (*lanes 1* and *2*) or *Hind*III (*lanes 3* and *4*) digested genomic DNA samples of Xiaoyan 54 (*lanes 1* and *3*) and Chinese Spring (*lanes 2* and *4*), showed that *Tripper* existed as multiple copies in the hexaploid genome of bread wheat. The size of the DNA markers (kb) is shown on the right side of the graph. **b** Influence of *Tripper* on the transcription (expression) of the *GUS* gene directed by the 5′ flanking region of *1Bx14* assessed using particle bombardment mediated transient assay. The 5′ flanking region of *1Bx14* in the expression construct pM1Bx14PR-GUS contained the *Tripper* element, whereas the *Tripper* element was deleted from 5′ flanking region of *1Bx14* in the construct p1Bx14PR-GUS via PCR mutagenesis. Based on the average numbers of GUS spots per endospermic tissue segment, the two expression constructs did not differ significantly in terms of *GUS* gene expression, indicating that the presence or absence of *Tripper* may not substantially affect the transcription directed by the 5′ flanking region of *1Bx14*

sequence deduced from *1Bx14* ORF contained 795 amino acid residues (Fig. 2). The primary structure of 1Bx14 was identical to that of the HMW glutenin subunits characterized previously and composed of a signal peptide, an N-terminal domain, a central repetitive domain and a C-terminal domain (Fig. 2, Table 1). The first 20 residues of the deduced 1Bx14 mature protein were identical to those determined by direct protein sequencing (Fig. 2). Amino acid sequence comparison showed that 1Bx14 was most closely related to that of 1Bx20. The two sequences differed in five positions (one in the N-terminal domain, three in the repetitive domain, and one in the C-terminal domain) involving the substitutions of single amino acid residues (Fig. 2). Interestingly, in the N-terminal domains of 1Bx14 and 1Bx20, the last two of the three cysteine residues, which were conserved in 1Bx7 and 1Bx17 (and all other Glu-1-1 subunits characterized so far), were replaced by tyrosine residues (Fig. 2).

Structural features of the 5′ flanking region of *1Bx14*

The 5′ flanking region (−1,147 bp relative to the start codon) of *1Bx14* was obtained by PCR amplification using the primers P5 and P1Bx14/20 and was aligned to homologous regions in *1Ax2**, *1Dx5*, *1Bx7*, *1Bx17* and *1Bx20* (Fig. 3). This alignment showed that the 5′ flanking region of *1Bx14* was similar to those of *1Bx7*, *1Bx17* and *1Bx20* (higher than 90% identities). Among the compared regions, a typical MITE insertion (Fig. 3, sequence represented by bold letters) was found in *1Bx14* sequence. This MITE insertion was shared by *1Bx20* but not *1Ax2**, *1Dx5*, *1Bx7* and *1Bx17* (Fig. 3). The *waffle* insertions previously identified in the 5′ flanking regions of *1Ax2** and *1Dx5* (Anderson et al. 2002) were located more upstream than the MITE insertions seen in the 5′ flanking region of *1Bx14* and *1Bx20*, and were not shown in Fig. 3. A tandem duplication of a sequence element consisted of 54 nucleotides (Fig. 3, underlined) was found in all four *Glu-B1-1* alleles. A part of this duplicated sequence element was also present in the 5′ flanking regions of *1Ax2** and *1Dx5*, but it was not duplicated (Fig. 3). The enhancer element (Fig. 3, sequence in brackets, Thomas and Flavell 1990) and the TATA box (Fig. 3, marked by asterisks) were conserved in the 5′ flanking regions of all compared alleles.

The nucleotide sequences of the MITE insertions in the 5′ flanking regions of *1Bx14* and *1Bx20* were identical. The MITE had 14 bp terminal inverted repeat (TIR, 5′-CAGTGGCGGAGCTT-3′, Fig. 3). The insertion of the MITE produced 8 bp target site duplication (TSD, 5′-CATAGGAG-3′, Fig. 3, boxed regions). Nucleotide sequence comparisons suggested that the MITE associated with the 5′ flanking regions of *1Bx14* and *1Bx20* was not related to any other MITEs identified previously and was therefore given a new name *Tripper*. Southern hybridiza-

tion analysis showed that *Tripper* existed as multiple copies in the genome of hexaploid bread wheat varieties (Fig. 4a). Computer modeling indicated that the nucleotide sequence of *Tripper* could potentially form a stable secondary structure composed of stems and loops (data not shown). The location of the *Tripper* element and its potential to form secondary structure led us to test if *Tripper* could affect the transcription directed by the 5′ flanking region of *1Bx14*. Interestingly, we found that the presence or absence of *Tripper* in the 5′ flanking region of *1Bx14* did not significantly affect the expression of the *GUS* marker gene (Fig. 4b) in transient expression assays mediated by particle bombardment.

Absence of *1Bx14* and *1Bx20* like alleles in diploid *Aegilops* species containing various types of S genomes

Owing to the unusual properties in the promoter regions of *1Bx14* and *1Bx20* and in the amino acid sequences of the two subunits, it was interesting to investigate if *1Bx14* and *1Bx20* like alleles would be present in the ancestral species donating the B genome. Because the precise identity of the ancestral species donating the B genome is still not known, we tried to amplify *1Bx14* and *1Bx20* like alleles from four *Aegilops* species containing various S genomes, to which the B genome of tetraploid and hexaploid wheats may be related. Using the PCR primers P5 and P1Bx14/20, the anticipated fragment from the S genome *Aegilops* species would be about 1.35 kb (including 5′ flanking region, the coding sequence for the signal peptide, and the sequence encoding the first 45 amino acid residues of the N-terminal domain). In repeated PCR experiments, the expected 1.35 kb fragment was not detected in any of the six accessions of *Ae. speltoides*, seven accessions of *Ae. searsii*, three accessions of *Ae. longissima*, and two accessions of *Ae. bicornis*. In stead, a 1.1-kb fragment was found in the majority of the *Aegilops* accessions. The nucleotide sequence of this 1.1 kb fragment was highly similar to the 5′ flanking region plus the coding sequences for the signal peptide and the first 45 amino acid residues of the N-terminal domain in *1Bx7* or *1Bx17* rather than to the corresponding regions in *1Bx14* or *1Bx20*. These results indicated that *1Bx14* and *1Bx20* like alleles might not be present in the S genome containing *Aegilops* species. In contrast, using the same PCR conditions *1Bx14* and *1Bx20* like alleles were readily detected in two (*T. turgidum* ssp. *dicoccon* and *T. turgidum* ssp. *turgidum*) of the three tetraploid subspecies (Supplementary Fig. S4).

Evolutionary analyses of *Glu-B1-1* alleles

The results described above on comparative analyses of the amino acid sequences and the 5′ flanking regions of *1Bx14* and its alleles (*1Bx7, 1Bx17*, *1Bx20*) suggested that *1Bx14* and *1Bx20* might represent a novel group of HMW glutenin subunit gene alleles. To study the evolutionary biology of *1Bx14* and *1Bx20*, we asked three questions. (1) What was the phylogenetic relationship of *1Bx14* and *1Bx20* with previously characterized *Glu-1-1* alleles? (2) Was the N-terminal domain possessing three conserved cysteine residues (exemplified by 1Bx7, 1Bx17, and homoeologous 1Ax and 1Dx subunits) more ancestral than those containing only one conserved cysteine residue (represented by 1Bx14 and 1Bx20), or vice versa? (3) What was the divergence time between *1Bx14* and *1Bx20* type of alleles and the *1Bx7* and *1Bx17* type of alleles?

To approach the first question, we conducted phylogenetic analysis of *1Ax2**, *1Bx7, 1Bx14*, *1Bx17*, *1Bx20* and *1Dx5* using the 5′ flanking sequences plus the sequences encoding the signal peptides and N-terminal domains. The selection of these sequences for phylogenetic analysis was based on the following reasoning. First, with the exception of the silenced alleles, HMW glutenin subunit genes are specifically and highly expressed in the endospermic tissue of developing seeds, indicating that the *cis*-elements that control tissue specificity and expression level of different HMW glutenin subunit genes are well conserved in the 5′ flanking regions. Second, the coding sequences for the signal peptides and N-terminal domains are also relatively conserved among different HMW glutenin subunit genes, probably owing to the important roles of the signal peptides (in targeting the newly synthesized subunits into protein bodies) and N-terminal domains (in maintaining the high order structure of the subunits). Third, the conservations in the 5′ flanking sequences and the sequences encoding the signal peptides and N-terminal domains suggest that these regions are subject to progressive changes during the evolution of HMW glutenin subunit genes. They are therefore phylogenetically informative. Fourth, the coding sequences for the repetitive domains are not suitable for phylogenetic investigations because they contain repetitive motifs that interfere with the correct alignment of the sequences to be compared. The phylogenetic tree thus constructed had two clades, one composed of *Glu-B1-1* alleles and the other of *Glu-A1-1* and *Glu-D1-1* alleles (represented by *1Ax2** and *1Dx5*, respectively) (Fig. 5). In the former clade, there were clearly two branches: one composed of *1Bx7* and *1Bx17* and the other of *1Bx14* and *1Bx20* (Fig. 5). Furthermore, the division of the two *1Bx* branches was supported by high bootstrap values (Fig. 5), indicative of strong statistic support for the existence of two *Glu-B1-1* allelic lineages.

For investigating the second question, we attempted to infer the ancestral amino acid sequence for the N-terminal domain of the *x* type HMW glutenin subunits (immediately before the differentiation of the two 1Bx allelic lineages) using the computer program ANCESTOR (Zhang and Nei 1997). To this end, a phylogenetic tree was constructed using the amino acid sequences of the signal peptides and N-terminal domains of 1Ax1, 1Ax2*, 1Bx7, 1Bx14, 1Bx17, 1Bx20, 1Dx2 and 1Dx5 and the corresponding regions of barley D-hordein as an outgroup (Fig. 6a). The results showed clearly that, immediately

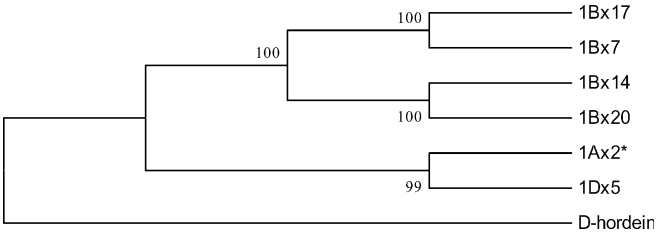

Fig. 5 Phylogenetic relationship of *1Bx14* and *1Bx20* with previously characterized *Glu-B1-1* (*1Bx7*, *1Bx17*), *Glu-A1-1* (represented by *1Ax2**) and *Glu-D1-1* (represented by *1Dx5*) alleles. The rootless phylogenetic tree was constructed based on a multiple alignment of the 5′ flanking sequences plus the sequences encoding the signal peptides and N-terminal domains of the six *Glu-1-1* alleles and the barley D-hordein gene (used here as an outgroup). The bootstrap values were calculated based on 500 replications

before the separation of the two allelic lineages of Glu-B1-1 subunits (in node 14, Fig. 6a and b), the configuration of the N-terminal domain was the one possessing three conserved cysteine residues, thus indicating that the N-terminal domain that contained three conserved cysteine residues was likely to be more ancestral.

The differentiation of the two *Glu-B1-1* allelic lineages may occur either during the evolution history of tetraploid wheat or in the ancestral species denoting the B genome. If the former scenario was true, then the divergence time for the two *Glu-B1-1* allelic lineages would be about 0.5 MYA because tetraploid wheat was formed approximately 0.5 MYA. By aligning the genomic sequences of *1Bx7* and *1Bx14* (or those of *1Bx17* and *1Bx14*) and calculating the numbers of total nucleotide substitutions, the divergence time for *1Bx7* and *1Bx14* types of alleles was estimated to be 0.46±0.15 MYA (Table 2). This indicated that the divergence time of the two *Glu-B1-1* allelic lineages might coincide with the timing of the tetraploidization event that led to the formation of tetraploid wheat. The relatively short divergence time between the two *Glu-B1-1* allelic lineages was against the alternative scenario that the origin of the two lineages could be traced back to the ancestral species denoting the B genome. Moreover, the results from our PCR experiments on the four S genome containing *Aegilops* species indicated that the B genome ancestral species, although expressing *1Bx7* and *1Bx17* like alleles, might not encode *1Bx14* and *1Bx20* like alleles.

Discussion

In the work described in this paper, we reported, for the first time, molecular information on the nucleotide sequence of the HMW glutenin subunit allele *1Bx14* and

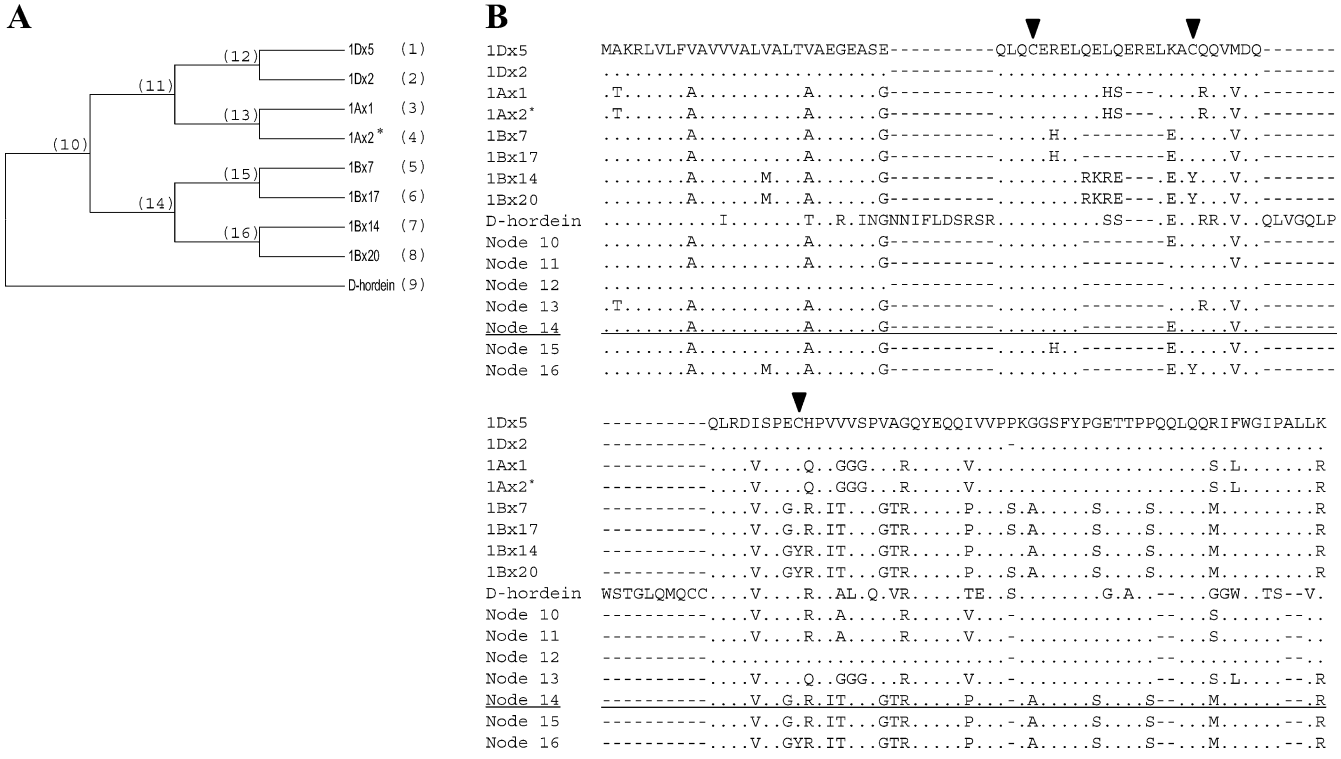

Fig. 6a, b Inference of ancestral amino acid sequences for the N-terminal domains of *x* type HMW glutenin subunits using the ANCESTOR program (Zhang and Nei 1997). **a** A phylogenetic tree generated using the amino acid sequences of the signal peptides plus N-terminal domains of eight *x* type subunits and the barley D-hordein (used here as an outgroup). The *numbers in brackets* indicate nodal positions. **b** Prediction of ancestral amino acid sequences at nodes 10–16. *Periods* indicate amino acid residues identical to the ones in the 1Dx5 sequence. *Hyphens* indicate deletions of single or multiple amino acid residues. The ancestral amino acid sequence inferred for Node 14 (*underlined*) contains the three cysteine residues (indicated by *arrowheads*) conserved in the N-terminal domains of the majority of the HMW glutenin subunits characterized so far, indicating that the N-terminal domain containing three conserved cysteine residues is likely to be more ancestral to the one possessing one conserved cysteine residue

Table 2 Estimation of divergence time between *1Bx7* and *1Bx14* allelic lineages

	Number of aligned sites[a]	Distance[b]	MYA
1*Bx*7/1*Bx*14	1,268	0.006±0.002	0.46±0.15
1*Bx*17/1*Bx*14	1,268	0.006±0.002	0.46±0.15

For calculating the divergence time, an average nucleotide substitution rate of 6.5×10^{-9} per site per year was used

[a]The DNA sequences used for the alignment were those covering the 5′ flanking regions plus the regions encoding the signal peptides and N-terminal domains of the compared alleles

[b]Distances (and SD) were calculated using the complete deletion option and a variety of nucleotide substitution models (*p*-distance, Tajima-Nei, Kimura 2-parameter, Jukes-Cantor, Tamura 3-parameter, Tamura-Nei) as implemented in the MEGA website. But identical values were obtained in all cases

the amino acid sequence of its protein product. Using high fidelity genomics PCR, the nucleotide sequences covering the coding region as well as the 5′ and 3′ flanking regions of *1Bx14* were obtained. Each nucleotide sequence was constructed using sequencing information derived from multiple independent clones. Potential errors brought about by the use of degenerate primers (e.g., P1 and P2) were corrected by amplifying and sequencing additional DNA fragments containing the regions from which the degenerate primers were originally derived. Using bacterial expression experiment, the correctness of the amplified coding region sequence of *1Bx14* was confirmed. Finally, the first 20 residues of the mature protein of 1Bx14 deduced from the cloned coding region were found to match exactly with those determined by direct protein sequencing. Taken together, it can be concluded that the molecular information we generated for *1Bx14* and its protein product is reliable and is hence suitable for investigating structural differentiations and evolution of *Glu-B1-1* alleles in bread wheat and related tetraploid wheat species.

It has long been shown that, in both tetraploid and hexaploid wheats, the *Glu-B1* locus encodes more *x* (and *y*) alleles than the *Glu-A1* and *Glu-D1* loci (Payne et al. 1984). Prior to our study, molecular information on the gene structure of *Glu-B1-1* alleles was only available for *1Bx7*, *1Bx17* and *1Bx20*. The results reported here suggest that *1Bx14* was similar, but not identical, to *1Bx20*. The insertion of the *Tripper* element in the 5′ flanking region and the possession of a novel ORF that encodes HMW glutenin subunit with fewer conserved cysteine residues are unique features of *1Bx14* and *1Bx20*. However, the amino acid sequences of 1Bx14 and 1Bx20 subunits differed in five locations (involving the substitutions of single amino acid residues). This may have been the cause for the difference in the electrophoretic mobilities of the two subunits in SDS-PAGE analysis (Payne et al. 1984). It is also likely that the two subunits may differ functionally because past studies have demonstrated that allelic HMW glutenin subunits that are more than 95% identical in amino acid sequences (e.g., 1Dx2 and 1Dx5) possess drastic differences in their effects on the processing qualities of bread wheat varieties (Shewry et al. 1995; Shewry and Halford 2002). In view of above discussions, it is possible that there may be both similarities and differences in the function of 1Bx14 and 1Bx20. Because both subunits contain fewer conserved cysteine residues, their high order structures may be more similar to each other than to those of 1Bx7 or 1Bx17 that contain the whole complement of conserved cysteine residues. The 1Bx20 subunit has been postulated to confer poor processing properties in bread wheat based on in vitro incorporation assays (Shewry et al. 2003). It will be important to examine whether or not the expression of 1Bx14 is also associated with poor processing properties in bread wheat in future investigations. We noticed that, in past literatures, 1Bx14 and 1Bx20 were deduced to contain different numbers of cysteine residues in their proteins based on the separation of reduced or reduced and alkylated protein samples through reverse phase high performance liquid chromatography (RP-HPLC, 37). However, estimation of cysteine residues in protein samples by RP-HPLC may not be precise. For example, amino acid analysis showed that 1Bx20 and 1Bx13 contained similar numbers of cysteine residues, yet the two subunits differed substantially in their behavior during RP-HPLC separations (Margiotta et al. 1993).

Central to the evolution of HMW glutenin subunit genes are the duplication event that gives rise to *x* and *y* types of genes, the divergence of various *Glu-1* loci, and the differentiation of multiple alleles for a given *Glu-1* locus. In past investigations (Allaby et al. 1999), a time frame for the above evolutionary events was given, and some evidence on the differentiation of *Glu-B1-1* alleles was uncovered. The availability of the knowledge on *1Bx14* and its deduced protein permitted us to conduct a better analysis of the evolutionary relationships among *Glu-B1-1* alleles characterized so far. Based on our work, several new insights into the evolution of *Glu-B1-1* alleles were produced. First, our analysis demonstrated that four molecularly characterized *Glu-B1-1* alleles (*1Bx7*, *1Bx14*, *1Bx17*, *1Bx20*) possessed clear differences in their 5′ flanking regions (in terms of *Tripper* insertion) and their amino acid sequences (with respect to the numbers of the conserved cysteine residues in the N-terminal domain). Second, our phylogenetic analysis showed that the four *Glu-B1-1* alleles could be classified into two allelic lineages with strong statistical support. The results on the inference of the more ancestral amino acid sequence of the N-terminal domain suggest that the lineage represented by *1Bx7* and *1Bx17* is more ancient than the one represented by *1Bx14* and *1Bx20*. Third, by using longer stretches of nucleotide sequences that were judged to be phylogenetically informative, we estimated that the *1Bx7* and *1Bx14* allelic lineages probably diverged 0.46 ±0.15 MYA. Because the divergence of the two allelic lineages coincided with the timing of the formation of the tetraploid wheat, we hypothesize that the tetraploidization event might have some connection with the differentiation of the two *Glu-B1-1* lineages (see below). The two *Glu-B1-1* lineages uncovered previously (Allaby et al. 1999) diverged 1.4–2.0 MYA. Because this divergence time is

substantially earlier than the time at which the tetraploidization event occurred, the two *Glu-B1-1* lineages reported by previous investigators (Allaby et al. 1999) might have existed in the ancestral species denoting the B genome. Considering the fact that the *Glu-B1* locus possesses a significantly greater diversity of alleles than the *Glu-A1* and *Glu-D1* loci in tetraploid and hexaploid wheats, the differentiation of *Glu-B1-1* alleles at different evolutionary stages (i.e., in the ancestral species, in the tetraploid wheat, or after the formation of the hexaploid wheat) is possible. However, further studies, involving the characterization of more *Glu-B1-1* alleles with diverse molecular structures, are needed to verify this possibility.

Tripper, a novel MITE inserted in the 5′ flanking regions of *1Bx14* and *1Bx20*, was found to exist as multiple copies in the genome of hexaploid wheat. It is interesting to find that the nucleotide sequence of the *Tripper* element in *1Bx20* (from tetraploid wheat) was identical to that in *1Bx14* (from hexaploid wheat). Considering that the positions of the *Tripper* insertion in *1Bx14* and *1Bx20* were also identical, it is likely that the *Tripper* insertion in *1Bx14* and *1Bx20* took place before the split of the two genes (i.e., the two insertions did not occurred independently). In past literatures, MITE insertions have been found in numerous plant genes, either in 5′ flanking regions, introns, or 3′ untranslated regions (Wessler 1998). More recently, evidence supporting active MITE transposition has been found in rice, a distant relative of wheat (Jiang et al. 2003; Kikuchi et al. 2003; Nakazaki et al. 2003). MITE insertion can have three potential consequences on genes. First, the insertion may be lethal, which mutates the normal function of the inserted gene. For example, the insertion of the *mPing* MITE in rice *Rurm1* gene caused a slender mutation of the glume (Kikuchi et al. 2003). Second, the inserted element may form a part of the gene structure. For example, the MITE *Ditto-Os2* may provide the TATA box for the transcription of a rice gene homologous to maize *knotted-1* (Wessler 1998). Third, MITE insertion modifies gene expression pattern or the biochemical function of the product of the inserted gene. In this case, MITE insertion would contribute directly to the natural evolution of a functional allele with new property. An interesting example in the literature is that the evolution of a new phosphate transporter gene allele might be linked to a MITE insertion in the 5′ flanking region (Rausch et al. 2001). It has been proposed that "genomic shocks" (caused by biotic and abiotic stresses) may enhance the activities of transposons (McClintock 1984). In this respect, it is important to find that tissue culture and γ-radiation stresses stimulated MITE transposition and that polyploidization activated the transcription of a retrotransposon (Kikuchi et al. 2003; Kashkush et al. 2003). Based on above discussions (and the results described in this paper), it is tempting for us to speculate that the evolution of the *1Bx14* lineage from *1Bx7* like alleles might be linked to *Tripper* insertions in the 5′ flanking regions of *1Bx14* like alleles, and that this insertional event might be triggered by tetraploidization during the formation of tetraploid wheat. In our transient expression assays, *Tripper* insertion did not affect the transcription directed by the promoter region of *1Bx14*, indicating that the evolution of the 1Bx14 like subunits may not be linked to an alteration in the expression level of their coding genes.

In conclusion, we have characterized 1Bx14 and its coding and promoter sequences. Comparative analysis of *1Bx14* and other HMW glutenin subunit genes has provided new insights into structural differentiation and evolution of *Glu-B1-1* alleles. 1Bx14, together with 1Bx20, constitute a novel subclass of HMW glutenin subunits with fewer conserved cysteine residues in their proteins. Their genes represent an allelic lineage distinct from the one containing *1Bx7* and *1Bx17*. The precise mechanisms causing the divergence between *1Bx7* and *1Bx14* allelic lineages are currently unknown. But they may be linked to the polyploidization event and the dynamics of MITE insertions, both of which have profoundly affected the constitutions and activities of the genomes of grass species (Jiang and Wessler 2001; Ozkan et al. 2001; Feschotte et al. 2002; Kashkush et al. 2002).

Acknowledgements We thank Professors Peter Shewry, Rudi Appels and Domenico Lafiandra and Drs Olin Anderson and Zhongyi Li for discussion and help in this work. We are grateful to Dr. Jianzhi Zhang (University of Michigan, USA) for guidance in using the computer program "ANCESTER". This work was supported by grants from the Ministry of Science and Technology of China (2002CB111301, 2001AA222091) and a biotechnology grant from the Chinese Academy of Sciences.

References

Allaby RG, Banerjee M, Brown TA (1999) Evolution of the high molecular weight glutenin loci of the A, B, D, and R genomes of wheat. Genome 42:296–307

Anderson OD, Greene FC, Yip RE, Halford NG, Shewry PR, Malpica-Romero JM (1989) Nucleotide sequences of the two high molecular weight glutenin genes from the D-genome of a hexaploid bread wheat, *Triticum aestivum* L. cv. Cheyenne. Nucleic Acids Res 17:461–462

Anderson OD, Abraham-pierce FA, Tam A (1998) Conservation in wheat high molecular weight glutenin gene promoter sequences: comparisons among loci and among alleles of the *Glu-B1-1* locus. Theor Appl Genet 96:568–576

Anderson OD, Larka L, Christoffers MJ, Mccue KF, Gustafson JP (2002) Comparison of orthologous and paralogous DNA flanking the wheat high molecular weight glutenin genes: sequence conservation and divergence, transposon distribution, and matrix-attachment regions. Genome 45:367–380

Barro F, Rooke L, Békés F, Gras P, Tatham AS, Fido R, Lazzeri PA, Shewry PR, Barceó P (1997) Transformation of wheat with high electrophoretic mobility subunit genes results in improved functional properties. Nat Biotechnol 15:1295–1299

De Bustos A, Rubio P, Jouve N (2001) Characterization of two gene subunits on the 1R chromosome of rye as orthologs of each of the *Glu-1* genes of hexaploid wheat. Theor Appl Genet 103:733–742

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

Gaut BS, Morton BR, Mccaig BC, Clegg MT (1996) Substitution rate comparisons between grasses and palms: synonymous rate differences at the nuclear gene *Adh* parallel rate differences at the plastid gene *rbcL*. Proc Natl Acad Sci U S A 93:10274–10279

Halford NG, Tatham AS, Sui E, Daroda L, Dreyer T, Shewry PR (1992) Identification of a novel beta-turn-rich repeat motif in the D hordeins of barley. Biochem Biophys Acta 1122:118–122

Jiang N, Wessler SR (2001) Insertion preference of maize and rice miniature inverted repeat transposable elements as revealed by the analysis of nested elements. Plant Cell 13:2553–2564

Jiang N, Bao Z, Zhang X, Hirochika H, Eddy SR, Mccouch SR, Wessler SR (2003) An active DNA transposon family in rice. Nature 421:163–167

Kashkush K, Feldman M, Levy AA (2002) Gene loss, silencing and activation in a newly synthesized wheat allotetraploid. Genetics 160:1651–1659

Kashkush K, Feldman M, Levy AA (2003) Transcriptional activation of retrotransposons alters the expression of adjacent genes in wheat. Nat Genet 33:102–106

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

Kreis M, Forde BG, Rahman S, Miflin BJ, Shewry PR (1985) Molecular evolution of the seed storage proteins of barley, rye and wheat. J Mol Biol 183:499–502

Lawrence GJ, Shepherd KW (1981) Chromosomal location of genes controlling seed protein in species related to wheat. Theor Appl Genet 59:25–31

Lawrence GJ, Macritchie F, Wrigley CW (1988) Dough and baking quality of wheat lines deficient in glutenin subunits controlled by the *Glu-A1*, *Glu-B1* and *Glu-D1* loci. J Cereal Sci 7:109–112

Liu Z, Yan Z, Wan Y, Liu K, Zheng Y, Wang D (2003) Analysis of HMW glutenin subunits and their coding sequences in two diploid *Aegilops* species. Theor Appl Genet 106:1368–1378

Margiotta B, Colaprico G, D'Ovidio R, Lafiandra D (1993) Characterization of high M_r subunits of glutenin by combined chromatographic (RP-HPLC) and electrophoretic separations and restriction fragment length polymorphism (RFLP) analyses of their coding genes. J Cereal Sci 17:221–236

McClintock B (1984) The significances of responses of the genome to challenge. Science 226:792–801

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

Nei N, Kumar S (2000) Molecular evolution and phylogenetics. Oxford University Press, UK

Oñate L, Vicente-carbajosa J, Lara P, Díaz I, Carbonero P (1999) Barley BLZ2, a seed-specific bZIP protein that interacts with BLZ1 *in vivo* and activates transcription from the GCN4-like motif of B-hordein promoters in barley endosperm. J Biol Chem 274:9175–9182

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

Payne PI (1987) Genetics of wheat storage protein and the effect of allelic variation on breed making quality. Ann Rev Plant Physiol 38:141–153

Payne PI, Holt LM, Jackson EA, Law CN (1984) Wheat storage proteins: their genetics and their potential for manipulation by plant breeding. Philos Trans R Soc Lond B 304:359–371

Rampitsch C, Jordan MC, Cloutier S (2000) A matrix attachment region is located upstream from the high molecular glutenin gene *bx7* in wheat (*Triticum aestivum* L.). Genome 43:483–486

Rausch C, Daram P, Brunner S, Jansa J, Laloi M, Leggewie G, Amrhein N, Bucher M (2001) A phosphate transporter expressed in arbuscule-containing cells in potato. Nature 414:462–466

Reddy P, Appels R (1993) Analysis of a genomic DNA segment carrying the wheat high molecular weight (HMW) glutenin Bx17 subunit and its use as an RFLP marker. Theor Appl Genet 85:616–624

Sambrook J, Fritsch EF, Maniatis T (1989) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor

Sanderson MJ (1998) Estimating rate and time in molecular phylogenies: beyond the molecular clock? In: Soltis DE, Soltis PS, Doyle JJ (eds) Molecular systematics of plants II: DNA sequencing. Kluwer, Boston/Dordrecht/London, pp 242–264

Shewry PR, Halford NG (2002) Cereal seed storage proteins: structures, properties and role in grain utilization. J Exp Bot 53:947–958

Shewry PR, Tatham AS, Barro P, Lazzeri P (1995) Biotechnology of breadmaking: unraveling and manipulating the multi-protein gluten complex. Bio/Technology 13:1185–1190

Shewry PR, Halford NG, Belton PS, Tatham AS (2002) The structures and properties of gluten: an elastic protein from wheat grain. Philos Trans R Soc Lond B 357:133–142

Shewry PR, Gilbert SMA, Savage WJ, Tatham AS, Wan YF, Belton PS, Wellner N, D'ovidio R, Békés F, Halford NG (2003) Sequence and properties of HMW subunit 1Bx20 from pasta wheat (*Triticum durum*) which is associated with poor end use properties. Theor Appl Genet 106:744–750

Thomas MS, Flavell RB (1990) Identification of an enhancer element for the endosperm-specific expression of high molecular weight glutenin. Plant Cell 2:1171–1180

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

Wan Y, Liu K, Wang D, Shewry PR (2000) High-molecular-weight glutenin subunits in the *Cylindropyrum* and *Vertebrata* section of the *Aegilops* genus and identification of subunits related to those encoded by the Dx alleles of common wheat. Theor Appl Genet 101:879–884

Wan Y, Wang D, Shewry PR, Halford NG (2002) Isolation and characterization of five novel high molecular weight subunit genes from *Triticum timopheevi* and *Aegilops cylindrica*. Theor Appl Genet 104:828–839

Wessler SR (1998) Transposable elements associated with normal plant genes. Physiol Plant 103:581–586

Wicker T, Yahiaoui N, Guyot R, Schlagenhauf E, Liu ZD, Doubcovsky J, Keller B (2003) Rapid genome divergence at orthologous low molecular weight glutenin loci of the A and A^m genomes of wheat. Plant Cell 15:1186–1197

William MDHM, Peña RJ, Mujeeb-Kazi A (1993) Seed protein and isozyme variations in *Triticum tauschii* (*Aegilops squarrosa*). Theor Appl Genet 87:257–263

Zhang J, Nei M (1997) Accuracies of ancestral amino acid sequences inferred by the parsimony, likelihood, and distance methods. J Mol Evol 44:S139-S146

Zhang XY, Pang BS, You GX, Wang LF, Jia JZ, Dong YC (2002) Allelic variation and genetic diversity at *Glu-1* loci in Chinese wheat (*Triticum aestivum* L.) germplasms. Scientia Agricultura Sinica 35:1302–1310