Cereal Research Communications 42(4), pp. 658–667 (2014)
DOI: 10.1556/CRC.2014.0012
First published online June 15, 2014

A New HMW-GS *1Bx23** Containing an Amino Acid Segment Similar to Collagen

L. WEI[1,2,3]**, S.G. BAI[4]**, X.J. HOU[2], J.M. LI[2], B. ZHANG[1,3], W.J. CHEN[1,3], D.C. LIU[1,3], B.L. LIU[1,3,5]* and H.G. ZHANG[1,3]*

[1]Key Laboratory of Adaptation and Evolution of Plateau Biota (AEPB), Northwest Institute of Plateau Biology, Chinese Academy of Sciences, Qinghai Xining 810001, China
[2]College of Biologic and Geographic Sciences, Qinghai Normal University, Qinghai Xining 810008, China
[3]Qinghai Province Key Laboratory of Crop Molecular Breeding, Xining 810001, China
[4]Grassland Station of Huangzhong County, Xining 811600, China
[5]State Key Laboratory Breeding Base for Innovation and Utilization of Plateau Crop Germplasm, Xining 810001, China

(Received 25 December 2013; Accepted 5 February 2014;
Communicated by F. Békés)

Among 20 awnless Tibetan wheat cultivars analyzed by SDS-PAGE, the migration rate of an HMW-GS in XM001584 and XM001593, named *1BX23**. was shown to be slightly faster than *1Bx6*. and slower than *Bx7*. Its nucleotide sequence was isolated based on homology clones. In a phylogenetic tree of *1Bx* genes, *1Bx23** was apparently clustered with *1Bx23*. Compared with *1Bx23*. eight single nucleotide replacements caused four single amino acid replacements in *1Bx23**. The deletion of "G" at base pair 1463 and insertion of "A" at 1509 bps induced a 42-nucleotide frame shift. "GQRQQAGQWQRPGQ" was replaced by "DKGNRQDNGNDRDK". The new segment cannot be found in other HMW-GSs, and it is very similar to a segment found in collagen. Moreover, an 18-nucleotide deletion made *1Bx23** six amino acids shorter than *1Bx23*. The cultivar XM001593 had 28 chromosomes, which signifies that it was tetraploid wheat, and that the new HMW-GS *1Bx23** cannot be used directly for breeding in common wheat.

Keywords: Tibetan wheat cultivar, HMW-GS, Glu-Bx, allelic variation

Introduction

Wheat is one of the most widely grown crops in the world, and is an important protein resource for humans. High molecular weight glutenin subunit (HWM-GS) is an important component of wheat grain, and accounts for approximately 10% of storage proteins (Shewry et al. 2003). Mixed with the other ingredients gliadins and low molecular weight glutenin subunits (LMW-GS), HMW-GSs confer dough elasticity and extensibility. All HMW-GSs have a short non-repetitive N-terminal domain, large central elastomeric re-

* Corresponding authors; E-mails: blliu@nwipb.cas.cn, hgzhang@nwpb.cas.cn
** These persons contributed to this research equally.

0133-3720/$20.00 © 2014 Akadémiai Kiadó, Budapest

petitive domain, and *C*-terminal domain. The *N*- and *C*-terminal domains are both likely to be rich in α-helices (Tatham et al. 1985), and the dominant structural feature of the central repetitive domain is the presence of β-reverse turns (Tatham et al. 1990).

HMW-GSs are encoded by six loci in wheat genomes. The loci situated on the long arms of homologous group-1 chromosomes, and are respectively referred to as *Glu-1A, Glu-1B* and *Glu-1D*. Each *Glu-1* locus consists of two tightly linked genes, designated as x-type and y-type, which, respectively, encode the greater and the smaller subunits (Harberd et al. 1986). The main difference between the two types of HMW-GSs is that 18 residues of the *N*-terminal remain after 33 residues are deleted in the x-type subunits, so that the domain is 81 to 89 residues long in the x-type subunits, and 104 residues long in the y-type. In contrast, the *C*-terminal domains of both types consist of 42 residues (Shewry and Halford 2003).

Cysteine residues may form intermolecular disulfide bonds and alter the polymeric structure and conformation of the protein, which in turn influences the viscoelastic properties of the wheat dough. Therefore, the distribution and number of Cysteine residues are considered as important characters of HMW-GSs (Buonocore et al. 1998; Shewry and Tatham 1997). Y-type subunits have five Cysteine residues in their N-terminal domain, while x-type subunits have no more than three. 1Bx14 and 1Bx20 have one Cysteine residue, and 1Bx7, 1Bx17, all 1Dx and 1Ax subunits have three. All HMW-GSs have only one Cysteine residue in their C-terminal domain. In the central repetitive domain, Cysteine residue is also present in some special x-type subunits, but not in y-type subunits, e.g. 1Dx5 subunits. The extra Cysteine residue is taken as the major factor explaining the high quality of 1Dx5 subunits for bread making (Juhasz et al. 2001). With the exception of Cysteine residues, all other amino acids, which can cause interchain hydrogen bonds such as Glutamine or cross-links such as Tyrosine residue, are also taken into consideration in judging their effects on dough quality (Tilley et al. 2001; Anjum et al. 2007).

All six HMW-GS genes are present in most types of hexaploid cultivated wheat, but usually only 3 to 5 HMW-GSs are expressed, due to gene silencing and allelic variations (Payne 1987). Although 3 to 5 HMW-GSs exist in one wheat cultivar, there is a large amount of allele variability among the cultivars, and this variability has been proven to determine the suitability of wheat for bread making by influencing dough viscoelasticity. A large number of HMW-GS allelic variations have been identified, sequenced and documented at these loci in wheat, as well as wild relative species based on SDS-PAGE (sodium dodecyl sulfate-polyacrylamide gel electrophoresis) and homology clones (Yang et al. 2006; Yan et al. 2009). However, the information we have grasped is still not sufficient to understand their biochemical performances, quality potentials, and evolution relationships.

Here we reported a novel HMW-GS *1Bx23**, which was screened out from wheat cultivars in Tibet, China. Nucleotide and amino acid sequences of *1Bx23** were compared with other *1Bx* allelics, which should be valuable for understanding the evolution and variation mechanism of Glu-1Bx alleles, and allow us to provide new genes for improving wheat quality.

Materials and Methods

Plant materials

Twenty awnless Tibetan wheat cultivars, which were conserved at the Qinghai Duplication Seed Bank of China Crop Seed Bank, were investigated for their HMW-GS compositions.

SDS-PAGE analysis

To determine the electrophoretic mobility of each HMW-GS in SDS-PAGE, Chinese Spring (null, 7+8, 2+12) and GaoYuan182 (null, 6+8, 2+12) were used as standards. These materials were then analyzed by SDS-PAGE according to the procedure of Li (Li et al. 2007), with some modifications. One wheat grain was ground into powder, 0.02 g powder was weighed, and to it was added 200 μL HMW-GS extraction buffer (1 mg sample/100 μL extraction buffer) with 1.0 mol/L Tris-HCl (pH 6.8) containing 2% (w/v) SDS, 10% (v/v) glycerol and 3% (v/v) β-mercaptoethanol, 0.002% (w/v) bromophenol blue. These were shaken at room temperature for 1 h, then extracted at room temperature for 1 h. The suspension (extraction sample) was kept in boiling water for 10 min to denature, then centrifuged at 13,000 rpm for 10 min to obtain the supernatant. One μL supernatant solution was loaded into a sample well of 10% SDS-PAGE gel for separation of HMW-GS. The electrophoresis was performed at 20 mA by a constant current for about 2–7 h until the tracking dye (bromophenol blue) migrated off the gel. The gels were stained with 0.1% (w/v) Coomassuie Brilliant Blue R250 for 30 min and distained by 25% methanol and 10% acetic acid.

DNA extraction and polymerase chain reaction (PCR) amplification

The seeds were germinated in darkness at 20ºC for 2 weeks. Genome DNA was extracted from the young leaves by the CTAB (cetyltrimethylammonium bromide) method (Yan et al. 2009).

Based on the conserved nucleotide sequences at the 5' and 3' ends of the coding regions of the published *1Bx* HMW-GS gene, two pairs of oligonucleotide primers specific to *Glu-1Bx* gene were designed and synthesized: BxF, 5'-ATG, GCT, AAG, CGC, CTG, GTC, CT-3'; and BxR, 5'-AGC, TGC, AGA, GAG, TTC, TAT, CA-3', for amplification of the coding region. PCR was carried out using a Mastercycler pro thermocycler (Eppendorf, Germany). Each PCR reaction (50 μL) contained 2 μL (200 ng) template DNA, 4 μL of 2.5 mM dNTPs, 1 μL of each of Forward and Reverse Primers, 5 μL of 10×HiFi Taq buffer, 5 μL of GC enhancers, 0.5 μL of HiFi Taq DNA Polymerase, and 31.5 μL of ddH_2O to 50 μL volume. The PCR was programmed at 95ºC for 4 min, followed by 35 cycles of 94ºC for 1 min, 62ºC for 1 min and 72ºC for 2 min 30 s. After the amplification, the final extension was at 72ºC for 10 min.

Isolation, sequencing and comparative analysis of Glu-1Bx23*

The PCR products were separated on 1% agrose gel. The targeted DNA fragments were recovered using Easypare Quick Gel Extraction Kit (China, Beijing, TransGen Biotech

Co.), then ligated into pEASY-T1 cloning vector (China, Beijing, TransGen Biotech Co.), and used to transform competent cells of the *E. coli* DH10B strain. The full coding sequences were acquired by sequencing three positive resulting clones and a set of overlapping subclones, which were created by the nested deletion methods. The nucleotide sequences were assembled and translated into amino acid sequences using the ORF (Open Reading Frame) finder program (http://www.ncbi.nlm.nih.gov/gorf/gorf.html). The similarity comparisons of the sequence alignment among different Glu-1Bx alleles were carried out with the program BLAST on the NCBI website (http://www.ncbi.nlm.nih.gov), and alignments were made with the ClustalW program.

Phylogenetic analysis

The phylogenetic tree was constructed with the nucleotide acid sequences of ORF of 1Bx alleles using the Neighbour-Joining (NJ) method of the MEGA program (Kumar et al. 2008). During the analysis, the complete deletion option was used with respect to gaps in the aligned sequence, and the evolutionary distances were measured by calculating the p-distances for each pair of aligned sequences. The bootstrap value was estimated based on 2000 replications, using the same software.

Chromosome genotyping

Twenty seeds were, respectively, germinated at 25ºC for 48 h. The root tips, reaching up to 1.5 cM, were collected from 9:00 to 10:00, treated with a mixture of ice and water at 0–4ºC for 24 h, and fixed in Carnoy's fixative (95% ethanol: acetic acid = 3:1) at 0–4ºC for 1–8 weeks. After the sedimentation of the root tips in 45% acetic acid, the chromosome images of mitotic metaphase of the root tip cells were obtained by the root-tip squashing method with staining in the modified phenol magenta dye.

Results

SDS-PAGE analysis

Twenty Tibetan wheat cultivars were screened using SDS-PAGE for analysis of the HMW-GS component of these cultivars. After electrophoretic experiments, it was found that 15 landraces had HMW-GS 7+8 and 2+12, and three landraces had 1, 7+8, 2+12, while cultivars XM001584 and XM001593 had different HMW-GS components. Both sets had four protein bands. The electrophoretic mobility of the first band was close to *1Dx2*. The second band was slower than *1Bx7,* and faster than *1Bx6*. The third band was close to *1By8,* and the fourth band was faster than *1Dx10*. After sequencing, the fourth band was the same as *1Dx10*,* while the second was similar to *1Dx23,* and thus it was named as *1Bx23** (Fig. 1).

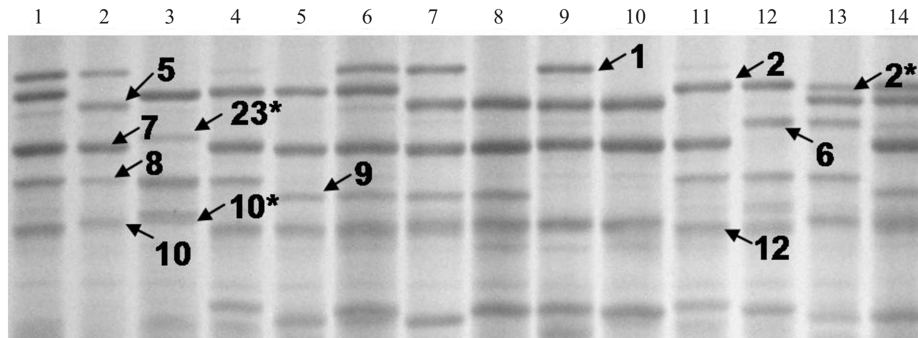

Figure 1. SDS-PAGE analysis of the novel HMW-GS Bx23*

Notes: Lanes 4, 11: Chinese Spring; Lane 1: XM001351; Lane 2: MY010533; Lane 3: XM01593; Lane 5: MY013262; Lane 6: ZM009790; Lane 7: MY013263; Lane 8: MY013255; Lane 9: MY012807; Lane 10: MY013701; Lane 12: XM001331; Lane 13: MY013257; Lane 14: MY014014

Nucleotide sequence isolation

Primers BxF and BxR specific for DNA fragments of *1Bx* gene were used to carry out the total genomic PCR of Tibet wheat cultivar XM001593. The PCR amplification produced approximately 2.4 kb DNA, which corresponds to the complete coding region of *Glu-1Bx* (Fig. 2). Then the DNA fragments were recovered and cloned into a T-easy vector. By sequencing the three different resulting clones and a series of overlapping subclones, which were created by nested deletion methods, the complete gene sequence of HMW-GS *1Bx23** was obtained and deposited in the GenBank database under the accession number

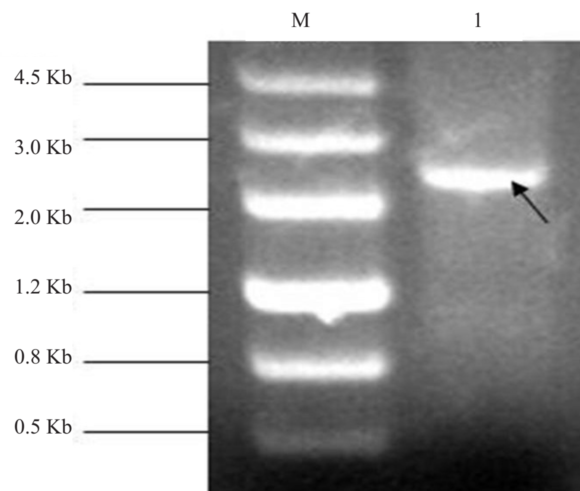

Figure 2. PCR amplification of the HMW-GS gene *Glu-Bx23**

Notes: Lane M: DNA marker; Lane 1: XM001593

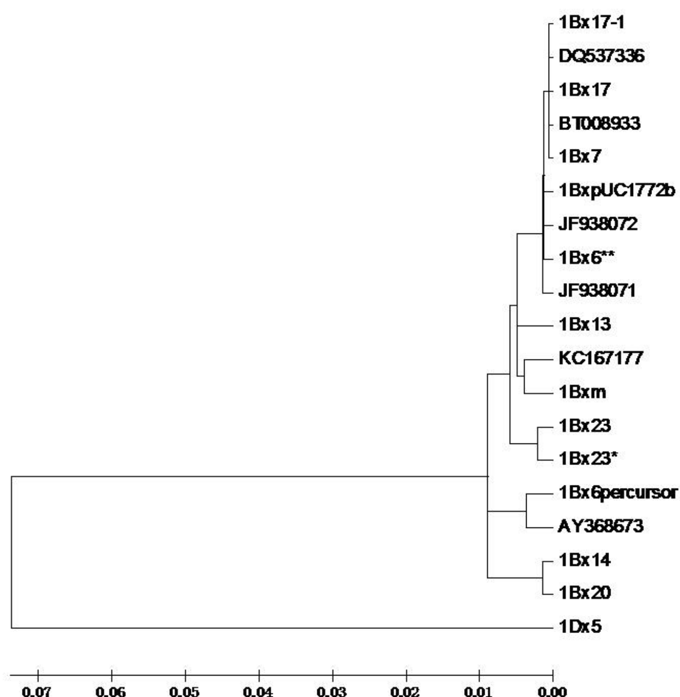

Figure 3. Phylogenetic relationship of nucleotide acid sequence of HMW-GS gene *Bx23** with previously characterized genes

Notes: The genes are Glu-B1-1 alleles *(1Bx23*, 1Bx7, 1Bx6**, 1Bx23, 1Bx20, 1Bx17, 1Bx14, 1Bx17-1, BT008933, DQ537336, 1BxpUC1772b, JF938072, JF938071, 1Bx13, KC167177, 1Bxm, 1Bx6percursor, AY368673)* and *1Dx5*. The bootstrap values were calculated based on 2000 replications

KF995273. It was indicated that *Glu-Bx23** was a *1Bx* gene, and that it was very close to *Glu-Bx23* after the retrieval of the *Glu-Bx23** gene sequence in the NCBI net service (http://www.ncbi.nlm.nih.gov/blast).

Phylogenetic analysis

To analyze the relationship between *1Bx23** and the other *1Bx* glutenin, 55 nucleotide sequences of *Glu-1Bx* genes were downloaded from NCBI. However, only 17 genes were taken into consideration after excluding approximation. These were *1Bx17-1, BT008933, 1Bx17, DQ537336, 1Bx7, 1BxpUC1772b, JF938072, 1Bx6**, JF938071, 1Bx23, 1Bx13, KC167177, 1Bxm, 1Bx6percursor, AY368673, 1Bx14* and *1Bx20*. Through the construction of phylogenetic trees, sequence comparisons were performed among the *Glu-1Bx* to determine the relatedness and the divergent time. As shown in Fig. 3, the phylogenetic tree exhibits three branches among *Glu-1Bx,* two alleles (1Bx14, 1Bx20) in branch 1, 2 alleles (1Bx6percursor, AY368673) in branch 2, and the remaining 14 alleles were divided into branch 3. *1Bx23** and *1Bx23* were very close, and both belonged to 3. It was evident that

Figure 4. Comparison of nucleotide (a) and deduced amino acid sequence (b) between *1Bx23* and *1Bx23**

Notes: The Genbank accession numbers for *1Bx23* and *1Bx23** were AY553933 and KF995273, respectively. The short bars indicated that the nucleotide or amino acids were the same for two genes, while the broken boxes show the deletions or the insertions of nucleotides or amino acids

the present of *1Bx23** and *1Bx23* may have appeared later than the split of *1Bx14* and *1Bx20, 1BX6percursor* and *AY368673,* but before the separation of *1Bx7* and *1Bx17,* and other *1Bx* genes.

Compare 1Bx23* *to* 1Bx23

The complete gene sequence of *Glu-Bx23** was 2370 bps in the nucleotide sequence and 789 amino acid residues in the deduced protein sequence, which was close to the size of the previously published *Glu-Bx23* gene with 2388 bps nucleotides and 795 amino acid residues (Yan et al. 2009). The comparison data showed that the sequence of the complete coding region of *1Bx23* gene presented a homology of 98.83% with the *1Bx23** gene. a single-base "G" at base pair 1463, and an 18-base "GCACAAGGGCAACAACCA" from 1642 to 1659 bps were deleted in the sequence of *1Bx23**. A single-base insertion of "A" was found at 1509 bps downstream from the start point of the sequence. Two nucleotide substitutions of "AA" to "GG" appeared at the positions of 1899 and 1900 bps. There were also seven single-base nucleotide substitutions in this sequence. These were "A" to "G" at 768 bps, "T" to "C" at 1602 bps, "G" to "A" at 1689 bps, "A" to "G" at 1731 bps, "G" to "A" at 1736 bps, "G" to "A" at 2359 bps, and "G" to "A" at 2367 bps (Fig. 4a).

The deduced amino acid sequence of *1Bx23** had a similar primary structure to that of the published 1Bx23, containing 21 amino acid residues for the signal peptide, 81 residues of a conservative N terminal domain, 645 residues of repetitive central domains, and the last 42 residues of the *C*-terminal domains. Among the deduced amino acid sequences, the differentiation between *1Bx23** and *1Bx23* was that there were a deletion of a hexapeptide unit (PAQGQQ) at the position of 553–558 residues from the start point of the amino acid sequence in its repetitive domain, a 14-amino acid replacement unit (GQRQQAGQWQRPGQ/DKGNRQDNGNDRDK) from residues 488 to 501, and three single-amino acid replacements (Q/K, Q/R, G/R) at 503, 579 and 634 residues in its repetitive domain, and one single-amino acid replacement (T/A) at 787 residues in its C terminal domain (Fig. 4b).

Chromosome counting

1Bx23 originated from *Triticum turgidum,* and *1Bx23** from Tibet cultivar XM01593. Whether this cultivar belonged to *Triticum turgidum* or *Triticum aestivum* was determined by chromosome counting. Mitotic chromosomes were prepared for chromosome counting using the root tip squashing method. It was proven, as shown in Fig. 5, that the somatic chromosome number of XM01593 was 2n = 2X = 28, which signifies that it is tetraploid.

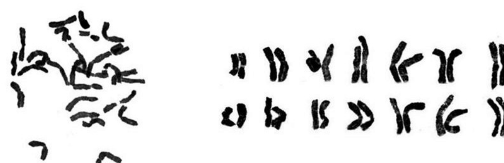

Figure 5. Chromosome morphology and numbers of XM01593

Discussion

Neither mutation nor silencing of HMW-GS genes is not lethal for wheat, thus wheat can endure any changes in the coding region of HMW-GS, and record these changes in future generations (Liu et al. 2008). Moreover, every HMW-GS has a large central repetitive domain. There are some repeats based on combinations of motifs of tripeptides, such as "GQQ"; hexapeptides, such as "PGQGQQ" and "PGQLQQ"; and nonapeptides, such as "LRQGQQGQQ". The tripeptides also appear in tandem with hexapeptides, forming nine-residue motifs, such as "PGQGQQGQQ". Some nonapeptides are interspersed with hexapeptides, forming 15-residue motifs, such as "PGQGQQLRQGQQGQQ". These types of repetitive sequences are very easy to delete or insert during the process of genome replication. In this study, an 18-nucleotide acid based deletion with the sequence of "GCACAAGGGCAACAACCA" was deleted in *1Bx23**. The repetitive sequences was six times of three nucleotide acids, encoding integer amino acid "PAQGQQ", and so its lost do not change amino acid sequence behind.

*1Bx23** has a different electrophoresis rate than *1Bx23*. *1Bx23** is between *1Bx6* and *1Bx7*, while *1Bx23* is between *1Bx7* and *1Bx8*. It may be inferred that *1Bx23** is slower than *1Bx23*, but *1Bx23** have six fewer amino acids, and a lower molecular weight than 1*Bx23*. Both *1Bx23* and *1Bx7* have 795 amino acids, but *1Bx7* is slower than *1Bx23,* which may be due to several single amino acid changes between *1Bx7* and *1Bx23*. Therefore, single-amino acid replacements, 6-amino acid deletion, and replacement of the 14 amino acids between *1Bx23** and *1Bx23* are likely to be the reasons for the different electrophoresis rates. Of course, more research must be performed to fully explain this phenomenon.

Compared with 1Bx23, one nucleotide was inserted at 1509 bps and one nucleotide was deleted at base pair 1463 in *1Bx23*,* which induced a frame shift. "GQRQQAG QWQRPGQ" was replaced by "DKGNRQDNGNDRDK". "DKGNRQDNGNDRDK" had eight charged amino acids, four acidic amino acids, four basic amino acids, and four polar amino acids, while "GQRQQAGQWQRPGQ" had two charged amino acids, two basic amino acids, six polar amino acids, and two hydrophobic amino acids. This difference is likely to play an important role in the changing electrophoresis rate. Moreover, the sequence "DKGNRQDNGNDRDK" cannot be found in the amino acid sequence of any HMW-GS. Using the BLAST software program on the NCBI website, it was shown to be similar to the repetitive amino acid sequence "DKGNSQDKGNSRDK" of collagen alpha-1(IV) chain in *Pseudopodoces humilis*. Collagen is a major protein found in animals, and is responsible for elasticity and tenacity. Further research must be performed to identify the effects of this sequence and *1Bx23** on the rheological characteristics of dough.

Acknowledgements

This work is financially supported by the National Natural Science Foundation of China (nos 31071417 and 31260322), State Key Laboratory Breeding Base for Innovation and Utilization of Plateau Crop Germplasm, Pilot Projects of Designer Breeding by Molecular Module, and West Light Foundation of The Chinese Academy of Sciences.

References

Anjum, F.M., Khan, M.R., Din, A., Saeed, M., Pasha, I., Arshad, M.U. 2007. Wheat gluten: High molecular weight glutenin subunits – Structure, genetics, and relation to dough elasticity. J. Food Sci. **72**:R56–R63.

Buonocore, F., Bertin, L., Ronchi, C., Bekes, F., Caporale, C., Lafiandra, D., Gras, P., Tatham, A.S., Greenfield, J.A., Halford, N.G. 1998. Expression and functional analysis of M-r 58 000 peptides derived from the repetitive domain of high molecular weight glutenin subunit 1 Dx5. J. Cereal Sci. **27**:209–215.

Harberd, N.P., Bartels, D., Thompson, R.D. 1986. DNA restriction-fragment variation in the gene family encoding high molecular weight (HMW) glutenin subunits of wheat. Biochem. Genet. **24**:579–596.

Juhasz, A., Tamas, L., Karsai, I., Vida, G., Lang, L., Bedo, Z. 2001. Identification, cloning and characterisation of a HMW-glutenin gene from an old Hungarian wheat variety, Bankuti 1201. Euphytica **119**:75–79.

Kumar, S., Nei, M., Dudley, J., Tamura, K. 2008. MEGA: A biologist-centric software for evolutionary analysis of DNA and protein sequences. Brief Bioinform. **9**:299–306.

Li, X., Zhang, Y., Gao, L., Wang, A., Ji, K., He, Z., Appels, R., Ma, W., Yan, Y. 2007. Molecular cloning, heterologous expression, and phylogenetic analysis of a novel y-type HMW glutenin subunit gene from the G genome of *Triticum timopheevi*. Genome **50**:1130–1140.

Liu, S.W., Gao, X., Xia, G.M. 2008. Characterizing HMW-GS alleles of decaploid *Agropyron elongatum* in relation to evolution and wheat breeding. Theor. Appl. Genet. **116**:325–334.

Payne, PI. 1987. Genetics of wheat storage proteins and the effect of allelic variation on bread making quality. Annu. Rev. Plant Physiol. **38**:141–153.

Shewry, P.R., Halford, N.G. 2003. Genetics of wheat gluten proteins. Advances in Genetics **49**:111–184.

Shewry, P.R., Tatham, A.S. 1997. Disulphide bonds in wheat gluten proteins. J. Cereal Sci. **25**:207–227.

Shewry, P., Halford, N., Tatham Yves, A. 2003. The high molecular weight subunits of wheat glutenin and their role in determining wheat processing properties. Adv. Food Nutr. Res. **45**:219–302.

Tatham, A.S., Drake, A.F., Shewry, P.R. 1990. Conformational studies of synthetic peptides corresponding to the repetitive regions of the high-molecular-weight (Hmw) glutenin subunits of wheat. J. Cereal Sci. **11**:189–200.

Tatham, A.S., Miflin, B.J., Shewry, P.R. 1985. The beta-turn conformation in wheat gluten proteins – Relationship to gluten elasticity. Cereal Chem. **62**:405–412.

Tilley, K.A., Benjamin, R.E., Bagorogoza, K.E., Okot-Kotber, B.M., Prakash, O., Kwen, H. 2001. Tyrosine cross-links: Molecular basis of gluten structure and function. J. of Agricultural and Food Chem. **49**:2627–2632.

Yan, Y.M., Jiang, Y., An, X.L., Pei, Y.H., Li, X.H., Zhang, Y.Z., Wang, A.L., He, Z., Xia, X., Bekes, F. 2009. Cloning, expression and functional analysis of HMW glutenin subunit 1By8 gene from Italy pasta wheat (*Triticum turgidum* L. ssp. *durum*). J. Cereal Sci. **50**:398–406.

Yang, Z.J., Li, G.R., Shu, H.L., Liu, C., Feng, J., Chang, Z.J., Ren, Z.L. 2006. Molecular characterization of high molecular weight glutenin subunit allele 1Bx23 in common wheat introduced from hexaploid triticale. Hereditas **143**:159–166.