ISSN 1682-296X (Print)
ISSN 1682-2978 (Online)

Bio Technology

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Biotechnology 7 (3): 544-550, 2008
ISSN 1682-296X
© 2008 Asian Network for Scientific Information

Genetic Polymorphism of Low-Molecular-Weight Glutenin Genes at *Glu-D3* Locus in a Wheat (*Triticum aestivum* L.) Variety, Chuannong 16

[1,3]Zhuo Huang, [2]Hai Long, [1]Yu-Ming Wei, [1]Ze-Hong Yan and [1,4]You-Liang Zheng
[1]Triticeae Research Institute, Sichuan Agricultural University, Yaan, Sichuan 625014, China
[2]Chengdu Institute of Biology, Science Academy of China, Chengdu, Sichuan, China
[3]Institute of Vocational Pedagogue, Sichuan Agricultural University, Dujiangyan, Sichuan 611830, China
[4]Ministry of Education Key Laboratory for Crop Genetic Resources and Improvement in Southwest China, Sichuan Agricultural University, Yaan, Sichuan 625014, China

Abstract: In this study, a pair of degenerate PCR primer was designed base on the known LMW-GS genes, with which 4 distinct sequences were obtained from a weak gluten variety of wheat, Chuannong 16, designated as *LMWCN16-1*, *LMWCN16-2*, *LMWCN16-3* and *LMWCN16-4*, respectively. The total lengths of the 4 genes were 902, 915, 909 and 927 bp, respectively. All these genes were predicted to be located at *Glu-D3* locus and amount of DNA variations were found, including 97 nucleotide substitutions and several insertions and deletions (indels). Sequence analysis indicated that the pairwise identities between the 4 genes were from 81.9 to 95.6%. *LMWCN16-2* and *LMWCN16-3* were more close related, while *LMWCN16-4* was distinguished from *LMWCN16-2* and *LMWCN16-3* by a longer repetitive domain and the different position of the seventh cysteines involved in inter-molecular disulphide bond. These differences might result in different effects on the formation of the glutenin polymer and consequently the wheat quality. *LMWCN16-1* is a putative pseudogene due to the five inframe stop codons and the frame shifting mutation in its C-terminal domain. The function loss of *LMWCN16-1* might weaken the proportion of *Glu-D3* in the total LMW-GS and bring negative influence on wheat quality.

Key words: Wheat, low-molecular-weight glutenin subunits, *Glu-D3*, quality improvement

INTRODUCTION

The endosperm is the main tissue in the wheat seed. It contains the majority of the seed- storage proteins, which are in the form of prolamins. Glutenins are among the major components and mainly consist of two types of subunits: high-molecular-weight glutenin subunit (HMW-GS) and low-molecular-weight glutenin subunit (LMW-GS), which are held together by intermolecular disulfide bonds to form the glutenin polymer. The HMW-GS represent approximately 10% of the total seed storage proteins. The important role of them has been well established (Shewry *et al*., 1992, 1995). LMW-GS represent about one-third of total seed storage proteins and 60% of glutenins (Bietz and Wall, 1973). They have been found to have pronounced effect in determining the physical properties of flour during bread-making (Gupta and MacRitchie, 1994) and pasta-making (Ruiz and Carrillo, 1995). However, the characterization of LMW-GS had been restricted due to the difficulty to resolve them by SDS-PAGE and the presence of a large number of subunits.

Coding genes of the LMW-GS are at *Glu-A3*, *Glu-B3* and *Glu-D3* loci and located on the short arms of chromosome 1A, 1B and 1D, respectively (Singh and Shepherd, 1988). The gene copy number is unknown. However, the estimates of the total gene copy number, based on southern blot analyses, varied from 10-15 (Harberd *et al*., 1985) to 35-40 (Cassidy *et al*., 1998; Sabelli and Shewry, 1991) in hexaploid wheat.

In recent years, more efforts have been focused on the isolation and characterization of LMW-GS genes. A number of LMW-GS genes were reported from wheat and its relatives (Cassidy *et al*., 1998; Ikeda *et al*., 2002; Wang *et al*., 2005; Huang *et al*., 2005; Shang *et al*., 2005; Yue *et al*., 2005; Ma *et al*., 2006). The general structure of a typical LMW-GS was thus predicted including a signal peptide of 20 amino acids, a short N-terminal domain (13 amino acids) that usually contains the first cysteine residue, a highly variable repetitive domain rich in glutamine codons and a conserved C-terminal domain, in which the other 7 cysteines presented.

The full-length of LMW-GS genes vary from about 900 to 1150 bp and are not interrupted by introns. Thus,

Corresponding Author: You-Liang Zheng, Triticeae Research Institute, Sichuan Agricultural University, Yaan, Sichuan 625014, China Tel: +86-835-2882620 Fax: +86-835-2883153

it is possible to isolate these genes directly from genomic DNA amplification (Lee *et al*., 1999). In this study, we attempted to isolate the LMW-GS genes, by PCR, from a week gluten variety of wheat, Chuannong 16, which was featured by the ideal plant architecture and high yield, but possessed the considerable potential of quality improvement. The objectives of this study were to investigate the coding regions of LMW-GSs in Chuannong 16 and to discuss their relationships to wheat quality.

MATERIALS AND METHODS

Plant material and preparation of genomic DNA: A Chinese wheat variety with weak gluten, Chuannong 16 was used for the isolation of LMW-GS genes. Chuannong 16 was developed and maintained by Triticeae Research Institute, Sichuan Agricultural University, which was initiated in 1990 and passed the national variety certification in 2003. The young leaves were harvested and immediately submerged in liquid nitrogen. Following grinding with mortar and pestle, pre-warmed extraction buffer (65°C), modified to include a final concentration of 2% (w/v) CTAB, 1.4 M NaCl, 100 mM Tris-HCl (pH = 8.0), 50 mM EDTA was added and then incubated for 2 h at 65°C. Following phenol/chloroform extraction, precipitation and ethanol wash, the DNA was dissolved in TE (1 mM Tris.HCl pH 8.0, 0.1 mM EDTA).

PCR amplification: PCR amplification of LMW-GS gene ORFs was conducted in 50 μL volume, containing 100-300 ng of genomic DNA, 200 μmol L^{-1} of each dNTPs, 50 μmol L^{-1} of each oligonucleotide primers, 3U *Ex taq* DNA polymerase (TaKaRa Biotechnology, Dalian, China) and 1xPCR buffer (supplied with the DNA polymerase). The amplification of the complete ORFs of LMW-GS gene was conducted with a degenerate primer set P1: forward, 5'-ACCATGAAGACCTTCCTC*G/A*TCTTT-3' and reverse, 5'-GAGTTGGTG*G/C*CTACTGATAA-3', which was designed on the basis of the known LMW-GS genes deposited in the GenBank and EMBL. The cycling parameters were: 94°C for 5 min followed by 30 cycles of 94°C for 45 sec, 63°C for 45 sec and 72°C for 1 min and a final extension at 72°C for 5 min.

Cloning of PCR products and DNA sequencing: PCR products were separated in 0.8% agarose gels. The desired DNA fragments were recovered from the gel, followed by ligation with the pMD18-T vector (TaKaRa Biotechnology, Dalian, China). The DNA and vector were combined in a 3:1 molar ratio and were ligated overnight at constant 16°C using 1xligase buffer and 150 U of T4 DNA ligase. Transformation of the recombinant plasmid into *E. coli* DH5α competent cells followed standard procedures (Sambrook *et al*., 1989). Transformed cells were plated on LB-agar containing 0.04 mg mL^{-1} Isopropyl-1-thio-β-D-galactoside (IPTG), 0.05 mg mL^{-1} of 5-bromo-4-chloro-3-indolyl-β-D-galactoside (X-Gal) and 0.05 mg mL^{-1} ampicillin and were incubated over night at 37°C. White colonies were amplified using M13 primers to identify the clones with an insert. Several positive colonies were sequenced in both directions by commercial company (TaKaRa Biotechnology, Dalian, China).

Sequence analyses: Amino acids prediction and sequence alignments were conducted by computer programs DNAMAN and BioEdit (Hall, 1999). Sequences similarity was analyzed using BLASTN program on the NCBI website (www.ncbi.nlm.nih.gov).

RESULTS

Isolation of LMW-GS genes from Chuannong 16: The primers were designed based on the conserved nucleotides of 5' and 3' ends of coding region aiming for the amplification of complete ORF of LMW-GS gene. A single expected band, approximate 900 bp was detected by agarose gel electrophoresis. The purified DNA fragment was transformed into *E. coli* DH5α competent cells after being ligated to pMD18-T vector. Directly sequencing of the positive colonies resulted in 4 clones, designated as *LMWCN16-1*, *LMWCN16-2*, *LMWCN16-3*, *LMWCN16-4*, respectively, which were confirmed as LMW-GS genes by performing BLASTN program on NCBI website. And higher sequence homology between them and known LMW-GS genes were also found (data not shown).

Sequence analyses: The total lengths of the four genes were: 902 bp for *LMWCN16-1*, 915 bp for *LMWCN16-2*, 909 bp for *LMWCN16-3* and 927 for *LMWCN16-4*. *LMWCN16-2*, *LMWCN16-3* and *LMWCN16-4* contain uninterrupted ORF, consisting of 303, 301 and 307 amino acid residues, respectively. *LMWCN16-1* is a putative pseudogene due to the presence of five inframe stop codons in the C-terminal domain. Alignment of nucleotide sequences show high degree of similarity between the 4 genes (Fig. 1) and the sequence identities between them are more than 81% (Table 1). *LMWCN16-2* and *LMWCN16-3* seemed to be more closely related, while *LMWCN16-4* was most distinct from the others (Table 1).

Chromosomal location prediction: Long *et al*. (2005) reported that the LMW-GS genes with same N-terminal domain sequences could be cluster to a group and located

Fig. 1: Alignment of nucleotide sequences of LMW-GS genes in this study

Table 1: Sequence identities between the 4 LMW-GS genes in this study

LMW-GIS-genes	*LMWCN16-1*	*LMWCN16-2*	*LMWCN16-3*	*LMWCN16-4*
LMWCN16-1	1	0.918	0.926	0.819
LMWCN16-2		1	0.956	0.832
LMWCN16-3			1	0.855
LMWCN16-4				1

Table 2: Motif comparison of the repetitive domains

	Repeat motifs			
LMW-GS	*LMWCN16-1*	*LMWCN16-2*	*LMWCN16-3*	*LMWCN16-4*
	QQQ	QQQ	QQQ	QQQ
	PLPPQQ	PLPPQQ	PLPPQQ	PLPPQQ
	SFSQQ	SFSQQ	TFPQQ	TFPQQ
	PPFSQQQQQ	PPFSQQQQQ	PPFSQQQQQ	PPFSQQQQQQ
	PLPQQ	PLPQQ	PLPQQ	PLPQQ
	PSFSQQQ	PSFSQQQ	PSFSQQQ	PSFSQQQ
				PPFWQQQ
	PPFSQQQ	PPFSQQQ	PPFSQQQ	PPFSQQQ
	PILPQG	PILSQQ	PILSQQ	PILPQQ
	PPFSQQTQ	PPFSQQQQ	PPFSQQQQ	PPFSQQQQ
	PVLPQQ	PVLPQQ	PVLPQQ	LVLPQQ
	SPFSQQQQ	SPFSQQQQ	SPFSQQQQ	PPFSQQQQ
	LILPPQQQQQ	LVLPPQQQQQQ	LVLPPQQQQQQ	PVLPPQQ
				SPFPQQQQQHQQ
	LPQQQ	LVQQQ	LVQQQ	LVQQQ
No. of amino acid	85	86	86	102
Gln content	38	40	41	48
Pro content	20	18	18	23

on the same chromosome arm, which provide a convenient method to predict the chromosomal locations of a cloned sequence. Therefore, *LMWCN16-1*, *LMWCN16-3* and *LMWCN16-4* belong to group 6 and *LMWCN16-2* belong to group 7.2, which both were D-genome specific gene groups.

Comparison of primary structure of LMW-GSs: The leader sequences of the deduced LMW-GSs encode a typical signal peptide consisting of 20 amino acid residues and were conserved among the four deduced polypeptides (Fig. 2a). Following the signal peptide was a short and highly conserved N-terminal domain consisting of 13 amino acids. A cysteine at position 5 was a common feature in this domain. The repetitive domain of LMW-GS features with tandem variations of short peptide motifs rich in proline and glutamine (Cassidy *et al*., 1998). The number of repeats present in this domain is mainly responsible for length variation of LMW-GS (D'Ovidio and Masci, 2004). *LMWCN16-1*, *LMWCN16-2* and *LMWCN16-3* consisted of 13 repeat motifs with conserved positions (Table 2), while the repeat domain of *LMWCN16-4*, which was composed of 15 repeat motifs, was longer than the others. The PPFSQQQQ was the most common repeat motif with variations in the number of glutamines. A unique motif SPFPQQQQQHQ could be only found in *LMWCN16-4*.

C-terminal domain could be subdivided to three domains. Domain I (Fig. 2c) was highly conserved in which five cysteines were found, as that in all known LMW-GS (Cassidy *et al*., 1998). All four LMW-GSs isolated in this study consist of 75 amino acids in this domain and the difference between were the amino acids substitutions at 12 positions. Domain II (Fig. 2d) contains a cysteine residue and was identified by its high glutamine composition. The position of the cysteine in *LMWCN16-4* was distinguished from other 3 LMW-GSs and more close to Domain I. Twenty-nine glutamines, 34, 33 and 27 glutamines were observed in *LMWCN16-1*, *LMWCN16-2*, *LMWCN16-3* and *LMWCN16-4*, respectively, which covered 50% of amino acids in this domain. Further more, 2-residue insertions were found in *LMWCN16-2* and *LMWCN16-4*, respectively and *LMWCN16-4* lost a peptide of 12 amino acids. Two stop codons replacing 2 glutamines were found in *LMWCN16-1* in this domain. Domain III (Fig. 2e) was composed of 51 residues among the *LMWCN16-2*, *LMWCN16-3* and *LMWCN16-4* and contained the last conserved cysteine. However, the sequence composition of *LMWCN16-1* in this domain, predicted from nucleotides, did not show significant similarity to other 3 LMW-GSs and contained 3 stop codons.

DISCUSSION

DNA diversity at *Glu-D3* locus and stop condons: Using a pair of degenerate oligonucleotide primers, 4 LMW-GS genes were isolated from a common wheat variety

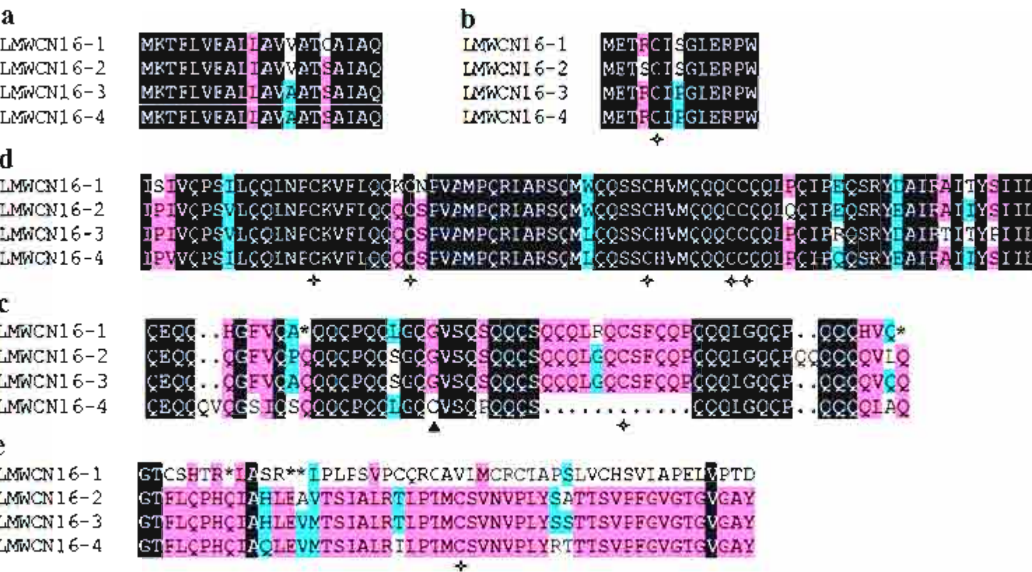

Fig. 2: Alignment of the deduced amino acid sequences, (a) signal peptide, (b) N-terminal domain, (c) C-terminal domain I, (d) C-terminal domain II and (e) C-terminal domain III. A star indicates a stop codon, a dot indicates a inserted gap, a crossing star indicates a conserved cysteine and the unique cysteine is indicated by a triangle

Chuannong 16. According to the results described by Long *et al*. (2005), it could be predicted that the 4 LMW-GS genes were located at *Glu-D3*. A total of 97 nucleotide substitutions were found in the 4 genes, including 65 transversions (A-G and T-C) and 32 transitions (A-C, A-T, G-C and G-T) (Fig. 1, 3). Forty-six, about 47% of total substitutions are nonsynonymous. Insertion or deletion is also response for the difference between the four sequences. Four insertions (total of 54 nucleotides) and a 36 bp-deletion were present in *LMWCN16-4*. In *LMWCN16-2*, a 6 bp-deletion was detected. A 3 bp- and a 4 bp-deletion occurred in *LMWCN16-1*. The latter led to the frameshift mutation resulting in the C-domain III.

LMWCN16-1 was a putative pseudogene due to the 5 inframe stop codons present in the C-terminal domain III. The origins of the first two stop codons at positions 667 and 808 might be explained by the same transversion (C→T) occurred in the first base of the codon encoding the glutamine (CGA→TGA) (Fig. 1). A four-nucleotide deletion occurred at position 816 caused the frame-shift mutation which changed the reading frame of the following sequence (from position 820 to the end) and produced the consequent three stop codons.

Primary structure and association with wheat quality: Except the *LMWCN16-1*, the other 3 LMW-GS genes in this study were likely functional cause the transcripts and protein products of same types the LMW-GS genes were detected in wheat (Ikeda *et al*., 2002, 2006) as well as its diploid D genome progenitor (Johal *et al*., 2004). The deduced amino acid sequences of signal peptide and N-terminal domain were conserved among 3 genes, which all belonged to LMW-m type and contained the first cysteine residue at position 5, which was involved in inter-molecular disulphide bonds.

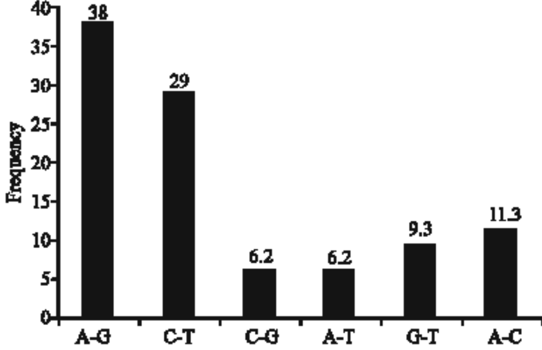

Fig. 3: The frequency distribution of nucleotide substitutions

Patacchini (2003) reported that a truncated repetitive domain decreased the mixing-time, a parameter that is positively correlated with glutenin polymer size. Application of flexibility modeling to a 42K LMW-GS indicated that the repetitive domain, particularly where stretches of glutamines were present was highly flexible.

These high flexibility regions were predicted to surround the cysteine residues involved in inter-molecular disulphide bonds to facilitate polymerization (D'Ovidio and Masci, 2004). These results suggested the length of repetitive domain and glutamines content might have significant affect on wheat quality. The *LMWCN16-4* had a longer repetitive domain and more glutamines than *LMWCN16-2* and *LMWCN16-3* and thus might possess different functional properties.

The C-terminal domain I and domain III among the 3 LMW-GSs were also conserved, containing 6 conserved cysteines to form the intra-molecular disulphide bonds (Fig. 2c, d). However, interestingly, a 12-peptide (including several glutamines) lost in the glutamine-rich domain II of *LMWCN16-4* and the position of a cysteine involved in inter-molecular disulphide bond was distinct to *LMWCN16-2* and *LMWCN16-3*. These features might also have some influence on the formation of the glutenin polymers.

It was proposed that *Glu-D3* locus was much larger and possessed more LMW-GS genes than *Glu-A3* and *Glu-B3* loci (Long *et al*., 2005; Huang and Cloutier, 2008). It probably was one of the reasons why the genes obtained by degenerate primers in this study were all from *Glu-D3* locus and possess abundant genetic polymorphism. It also suggested that the LMW-GS from D genome of wheat probably has important role in the formation of glutenin polymer. Therefore, the function loss of *LMWCN16-1* might weaken the influence of the *Glu-D3* on wheat quality. Further research on the function analysis of *LMWCN16-2*, *LMWCN16-3* and *LMWCN16-4* is necessary for the improvement of end-using quality of Chuannong 16 by either traditional breeding or using biotechnology strategies.

ACKNOWLEDGMENTS

This research was supported by the Science and Technology Foundation (N200701) of Dujiangyan Campus of Sichuan Agricultural University, the National High Technology Research and Development Program of China (863 program) and the Foundation for the Author of National Excellent Doctoral Dissertation of PR China (No. 200357 and 200458). Y.-M. Wei was supported by the Program for New Century Excellent Talents in University of China.

REFERENCES

Bietz, J.A. and J.S. Wall, 1973. Isolation and characterization of gliadin-like subunits from glutenins. Cereal Chem., 50: 537-547.

Cassidy, B.G., J. Dvorak and O.D. Anderson, 1998. The wheat low-molecular-weight glutenin genes: Characterization of six genes and progress in understanding gene family structure. Theor. Applied Genet., 96: 743-750.

D'Ovidio, R. and S. Masci, 2004. The low-molecular-weight glutenin subunits of wheat gluten. J. Cereal Sci., 39: 321-339.

Gupta, R.B. and F. MacRitchie, 1994. Allelic variation at gluten subunit and gliadin loci, *Glu-1*, *Glu-3* and *Gli-1* of common wheats. II. Biochemical basis of the allelic effects on dough properties. Cereal Chem., 19: 19-29.

Hall, T.A., 1999. BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl. Acid. Symp. Ser., 41: 95-98.

Harberd, N.P., D. Bartels and R.D. Thompson, 1985. Analysis of the gliadin multigene loci in bread wheat using nullisomic-tetrasomic lines. Mol. Gener. Genet., 198: 234-242.

Huang, Z., H. Long, Z.H. Yan, Y.M. Wei and Y.L. Zheng, 2005. Sequence analysis of LMW-GS genes from *Aegilops tauschii* containing HMW glutenin subunits 5+12 with good quality. High Technol. Lett., 15: 67-72.

Huang, X.Q. and S. Cloutier, 2008. Molecular characterization and genomic organization of low molecular weight glutenin subunit genes at the Glu-3 loci in hexaploid wheat (*Triticum aestivum* L.). Notice: Undefined index: institute in D:\scialert.net \how2cite\reflist.php on line 7710.1007/s00122-008-0727-1.

Ikeda, T.M., T. Nagamine, H. Fukuoka and H. Yano, 2002. Identification of new low-molecular-weight glutenin subunit genes in wheat. Theor. Applied Genet., 104: 680-687.

Ikeda, T.M., E. Araki, Y. Fujita and H. Yano, 2006. Characterization of low-molecular-weight glutenin subunit genes and their protein products in common wheats. Theor. Applied Genet., 112: 327-334.

Johal, J.M., C. Gianibelli, S. Rahman, M.K. Morell and K.R. Gale, 2004. Characterization of low-molecular-weight glutenin genes in *Aegilops tauschii*. Theor. Applied Genet., 109: 1028-1040.

Lee, Y.K., M. Ciaffi, R. Appels and M.K. Morell, 1999. The low-molecular-weight glutenin subunit protein of primitive wheat. II. The genes from A-genome species. Theor. Applied Genet., 98: 126-134.

Long, H., Y.M. Wei, Z.H. Yan, B. Baum, E. Nevo and Y.L. Zheng, 2005. Classification of wheat low-molecular-weight glutenin subunit genes and its chromosome assignment by developing LMW-GS group-specific primers. Theor. Applied Genet., 111: 1251-1259.

Ma, Z.C., Y.M. Wei, H. Long, Z.H. Yan, B. Baum and Y.L. Zheng, 2006. Characterization of low-molecular-weight i-type glutenin subunit genes from diploid wheat in relation to the gene family structure. Mol. Biol., 40: 897-906.

Patacchini, C., 2003. Structural characteristics of low molecular weight glutenin subunits and their influence in determining technological properties of durum wheat. Ph.D Thesis, University of Verona, Italy

Ruiz, M. and J.M. Carrillo, 1995. Separate effects on gluten strength of *Gli-1* and *Glu-3* prolamine genes on chromosomes 1A and 1B in durum wheat. J. Cereal Sci., 21: 137-144.

Sabelli, P. and P.R. Shewry, 1991. Characterization and organization of gene families at the *Gli-1* loci of bread and durum wheat. Theor. Applied Genet., 83: 428-434.

Sambrook, J., E.F. Fritsch and T. Maniatis, 1989. Molecular Cloning: A Laboratory Manual. 1st Edn. Cold Spring Harbour, New York.

Shang, H.Y., Y.M. Wei, H. Long, Z.H. Yan and Y.L. Zheng, 2005. Identification of LMW glutenin like genes from *Secale sylvestre* host. Russ. J. Genet., 41: 1-9.

Shewry, P.R., N.G. Halford and A.S. Tatham, 1992. High molecular weight subunits of wheat glutenin. J. Cereal Sci., 15: 105-120.

Shewry, P.R., A.S. Tatham, F. Barro, P. Barcelo and P. Lazzeri, 1995. Biotechnology of bread-making: unraveling and manipulating the multi-protein gluten complex. Biotechnology, 13: 1185-1190.

Singh, N.K. and K.W. Shepherd, 1988. The structure and genetic control of a new class of disulphide-linked proteins in wheat endosperm. Theor. Applied Genet., 71: 79-92.

Wang, Z.Q., H. Long, Y.L. Zheng, Z.H. Yan, Y.M. Wei and X.J. Lan, 2005. Cloning and analysis of LMW-GS genes from *Triticum aestivum* ssp. *tibetanum* Shao. Act. Genet. Sin., 32: 86-93.

Yue, Y.W., H. Long, Q. Liu, Y.M. Wei, Z.H. Yan and Y.L. Zheng, 2005. Isolation of low-molecular-weight glutenin subunit genes from wild emmer wheat (*Triticum dicoccoides*). J. Applied Genet., 46: 349-355.