Theor Appl Genet (2015) 128:2155–2165
DOI 10.1007/s00122-015-2574-1

ORIGINAL ARTICLE

Molecular characterization of novel LMW-i glutenin subunit genes from *Triticum urartu* Thum. ex Gandil.

Susana Cuesta[1] · Carlos Guzmán[2] · Juan B. Alvarez[1]

Received: 21 July 2014 / Accepted: 23 June 2015 / Published online: 8 July 2015

Abstract
Key message **A high level of genetic diversity was found in LMW-i genes from *Triticum urartu*, resulting in detection of 11 novel alleles. The variability detected could affect gluten quality**.
Abstract Low-molecular weight glutenin subunits are important in determining the viscoelastic properties of wheat dough. *Triticum urartu* Thum. ex Gandil., which is related to the A genome of polyploid wheat, has been shown as a good source of variation for these subunits. The present study evaluated the variability of LMW-i genes in this species. High polymorphism was found in the sequences analysed and resulted in the detection of 11 novel alleles, classified into two sets (Group-I and -II) showing unique SNPs and InDels. Both groups were associated with *Glu-A3-1* genes from common wheat. In general, deduced proteins from Group-II genes possessed a higher proportion of glutamine and proline, which has been previously suggested to be related with good quality. Moreover, there were other changes compared to common wheat. This novel variation could affect dough quality. Additional epitopes for celiac disease were also detected, suggesting that these subunits could be highly reactive. The results showed that *T. urartu* could be an important source of genetic variability for LMW-i genes that could enlarge the genetic pool of modern wheat.

Keywords Celiac disease · Genetic resources · Gluten quality · LMW-i genes · Molecular characterization · *Triticum urartu*

Communicated by A. Graner.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-015-2574-1) contains supplementary material, which is available to authorized users.

✉ Juan B. Alvarez
jb.alvarez@uco.es

[1] Departamento de Genética, Escuela Técnica Superior de Ingeniería Agronómica y de Montes, Edificio Gregor Mendel, Universidad de Córdoba, Campus de Rabanales, CeiA3, 14071 Córdoba, Spain

[2] Wheat Chemistry and Quality Laboratory, Global Wheat Program, International Maize and Wheat Improvement Center (CIMMYT), Texcoco, Mexico

Introduction

The bread-making quality of wheat flour has been widely associated with the endosperm storage proteins (see Wrigley et al. 2006 for reviews). Of these, two main groups, gliadins and glutenins, are the main components of gluten, the protein network that gives the dough its viscoelastic properties (elasticity and extensibility). Both groups differ in their molecular characteristics (Payne 1987). Glutenins are classified into high- (HMWGs) and low-molecular-weight (B-LMWGs and C-LMWGs) subunits. The HMWGs are the polypeptides mainly responsible for bread-making quality in common wheat, whereas this role is largely assumed by the B-LMWGs in durum wheat for pasta-making quality (Wrigley et al. 2006). The HMWGs have been most extensively studied because of their relatively small number and the ease of characterization by SDS-PAGE, whereas the characterization of LMWGs has been more difficult to elucidate due to the complexity of this multigene family (D'Ovidio and Masci 2004). The B-LMWGs are encoded by *Glu-3* loci located on the short arms of chromosomes 1A, 1B and 1D (Singh and Shepherd 1988; Pogna et al. 1990), which are closely linked to

the *Gli-1* loci for γ- and ω-gliadin genes. The B-LMWGs have been divided into three types: LMW-i (isoleucine), LMW-m (methionine) and LMW-s (serine) depending on the first amino acid of the mature protein (D'Ovidio and Masci 2004). The LMW-i are only coded by the *Glu-A3* locus, whereas the LMW-m are synthesised by the three loci and the LMW-s are coded by *Glu-B3* and *Glu-D3*.

The standard LMWG structure consists of four main domains: signal peptide, N-terminal, repetitive and C-terminal. The latter domain is further subdivided into three regions (Cassidy et al. 1998): the cysteine-rich (I), glutamine-rich (II) and highly conserved (III) domains. However, the LMW-i subunits possess a unique structure compared to LMW-m and LMW-s, because the N-terminal domain is missing. The LMWGs usually contain eight cysteine residues: seven in the C-terminal domain, and one in a variable position, in the N-terminal or repetitive domains for the LMW-m and LMW-s or in the C-terminal for the LMW-i (Ikeda et al. 2002). The first (or third for LMW-i) and seventh cysteine residues are involved in inter-molecular disulphide bonds, while the remainder form three intra-molecular disulphide bonds (D'Ovidio and Masci 2004). Consequently, in LMW-i the repetitive domain is precluded in the formation of the inter-molecular disulphide bonds in the gluten polymer. This different structure could lead to functional differences with respect to LMW-m and LMW-s and have a different impact on the viscoelastic properties of dough (Cloutier et al. 2001; Ikeda et al. 2002). Several studies have also associated LMWGs with celiac disease due to the presence of reactive epitopes as for those exhibited in gliadins (see Rasheed et al. 2014 for reviews).

A recent study showed that there were at least 15 LMWGs genes in individual accessions of common wheat (Zhang et al. 2013). The *Glu-A3* locus showed the highest allelic diversity and *Glu-B3* showed moderate diversity, whereas *Glu-D3* was very low. For the *Glu-A3* locus, Wang et al. (2010) identified up to three LMWG genes (*Glu-A3-1*–*Glu-A3-3*), the former encoding LMW-m and the latter two encoding LMW-i. However, Zhang et al. (2013) identified 4–6 genes (two for LMW-m and the rest for LMW-i) suggesting that these last were organized in different haplotypes.

Additional to the evaluation and characterization of the LMWG genes detected in modern wheat, the search for variability to extend the genetic pool is very important for wheat improvement (Jauhar 1993). In this respect, the putative diploid ancestors of the wheat genome could be good sources of useful genes (Srivastava and Damania 1989). As previously mentioned, the LMW-i subunits have been associated with the A genome, whose ancestor has been identified as *Triticum urartu* Thum. ex Gandil. (Dvorak et al. 1993), a wild wheat species ($2n = 2\times = 14$; A^uA^u) of the Fertile Crescent region (Johnson 1975; Miller 1987). Studies by protein separation in SDS-PAGE have shown a wide polymorphism for endosperm storage protein in this species (Rodríguez-Quijano et al. 1997; Lee et al. 1999; Caballero et al. 2008; Martín et al. 2008). In our previous study (Caballero et al. 2008), one broad collection (169 accessions) of *T. urartu* was evaluated for variability of HMWGs and B-LMWGs using SDS-PAGE analysis, and 17 HMWGs and 24 B-LMWGs alleles were detected. Of 17 HMWGs alleles, 12 were molecularly characterized by Alvarez et al. (2013), showing differences between these alleles and those present in common wheat. Of the B-LMWGs alleles, 20 in the collection were considered rare or very rare (frequency ≤5 %).

The aim of the current study was to molecularly characterize the LMW-i glutenin genes present in eight *Glu-A^{u}3* allelic variants identified in *T. urartu* (*Glu-A3ad*, *Glu-A3af*, *Glu-A3ag*, *Glu-A3ak*, *Glu-A3ao*, *Glu-A3aq*, *Glu-A3au* and *Glu-A3aw*) and to analyse their relationship with those present in common wheat.

Materials and methods

Plant materials

Eight accessions of *T. urartu* previously analysed by SDS-PAGE (Caballero et al. 2008) were used in this study (Table 1). These accessions were obtained from the National Small Grains Collection (Aberdeen, Idaho, USA) and the Institute for Plant Genetics and Crop Plant Research (Gatersleben, Germany).

Protein extraction and electrophoretic analysis and mass spectrometry

Proteins were extracted from single crushed seeds according to the protocol described by Alvarez et al. (2001). Reduced and alkylated glutenin subunits were fractionated by electrophoresis in vertical SDS-PAGE slabs in a discontinuous Tris–HCl–SDS buffer system (pH 6.8/8.8) at a polyacrylamide concentration of 10 % (w/v, C: 1.28). The Tris–HCl/glycine buffer system of Laemmli (1970) was used. Electrophoresis was carried at 30 mA/gel and 18 °C for 45 min after the tracking dye migrated off the gel. Gels were stained overnight with 12 % (w/v) trichloroacetic acid solution containing 5 % (v/v) ethanol and 0.05 % (w/v) Coomassie Brilliant Blue R-250. De-staining was carried out with tap water.

At the same time, cold acetone was added to the same sample supernatants, and then the LMWGs were allowed to precipitate. The samples were then used for matrix-assisted laser desorption ionization time of flight mass

Table 1 LMW-i sequences in *T. urartu*

Accession	Allele[a]	Frequency[b]	NCBI ID	DNA size (bp)	Mr (kDa)		Group	Repetitive size (aa)	Motifs		Glu + Pro (%)	
					Deduced[c]	MALDI[d]			Number	Size	Repetitive	Total
Iraqi accession												
PI 428253	*Glu-A3ao*	vr	KJ780780	1050	N	N	I	–	–	–	–	–
			KJ780787	894	N	N	II	–	–	–	–	–
			KP793238	1059	38.43	38.76	II	162	20	6–14	72.8	53.7
Lebanese accession												
PI 428328	*Glu-A3ag*	r	KJ780782	1056	38.23	37.29	I	149	19	6–9	72.5	52.9
			KJ780784	1059	38.43	39.12	II	162	20	6–14	72.8	53.7
Turkish accessions												
PI 428186	*Glu-A3ak*	r	KJ780778	1047	37.81	38.33	I	149	19	6–9	73.1	52.7
			KJ780785	1062	38.56	38.33	II	162	20	6–14	72.8	53.7
			KJ780786	1059	38.43	38.33	II	162	20	6–14	72.8	53.7
PI 428188	*Glu-A3ad*	r	KJ780779	1050	N	N	I	–	–	–	–	–
			KP793237	1059	38.43	38.26	II	162	20	6–14	72.8	53.7
PI 428191	*Glu-A3au*	vr	KJ780781	1044	37.72	38.91	I	149	19	6–9	72.4	52.3
			KJ780788	1056	38.33	38.91	II	155	19	6–16	74.2	54.7
PI 428225	*Glu-A3aq*	vr	KJ801156	1050	N	N	I	–	–	–	–	–
PI 428255	*Glu-A3af*	r	KJ780783	1050	N	N	I	–	–	–	–	–
TRI 11496	*Glu-A3aw*	vr	KJ801157	1059	38.43	38.54	II	162	20	6–14	72.8	53.7

r rare (≤5 %), *vr* very rare (≤1 %)

[a] McIntosh et al. (2013)

[b] Classification according to frequency in the original collection (Caballero et al. 2008)

[c] *N* pseudogene. Data of the mature proteins without signal peptide

[d] Molecular weight by MALDI–TOF–MS analysis

spectrometry (MALDI–TOF–MS) on an AB Sciex 5800 TOF–TOF apparatus (AB Sciex, Darmstadt, Germany). The matrix used was α-ciano-4-hidroxicinámico (CHCA). The calibration was done with the calibration kit Cal Mix3 (AB Sciex, Darmstadt, Germany). According to the molecular weights of LMWGs obtained from gel electrophoresis and mass spectrometry, the corresponding protein subunit encoded by the studied genes was identified.

DNA extraction and PCR amplification

Genomic DNA was isolated from young leaves of a single plant per accession according to the method of CTAB (Stacey and Isaac 1994). In order to amplify the complete coding region of the LMW-i genes, primers 5′-ATGAAGACCTTCCTCGTCTTT-3′ (Ma et al. 2006) and 5′-TCACACATGACGTTGTGTGAC-3′ (Zhang et al. 2011) were used. PCR amplification of genomic DNA was performed in a volume total of 20 µl containing 50 ng of genomic DNA, 0.3 µM of each primer, 0.4 mM of dNTPs, 1 or 1.5 mM of $MgCl_2$, 1× of reaction buffer and 1 U of *Taq* DNA polymerase (Promega, Madison, WI, USA). The amplification was carried out with a first step of initial denaturation at 94 °C for 3 min followed by 35 cycles of 30 s of denaturation at 94 °C, a step of annealing of 30 s at 58 or 60 °C and then 1.5 min of extension at 72 °C. To finish the process, an extension final step at 72 °C for 10 min was performed. The PCR products (amplicons) were separated by electrophoresis on polyacrylamide gels of 8 % (w/v, C: 1.28 %), stained with ethidium bromide and visualised under UV light.

DNA sequencing analysis

The PCR products were purified using Sureclean (Bioline) and then ligated to pGEM-T (Promega, Madison, WI, USA) and used to transform *Escherichia coli* JM109 competent cells. Thirty colonies of each cloned PCR product were analysed for the presence of LMWGs genes insert. Inserts were amplified with M13 universal primers (binding region adjacent to the insert in the vector) and the PCR products were separated using electrophoresis on polyacrylamide gels of 8 % (w/v, C: 1.28 %). At least three different inserts were sequenced using an ABI Prism 310 Genetic Analyzer (Applied Biosystems, Carlsban, CA, USA). The novel sequences are available from Genbank database.

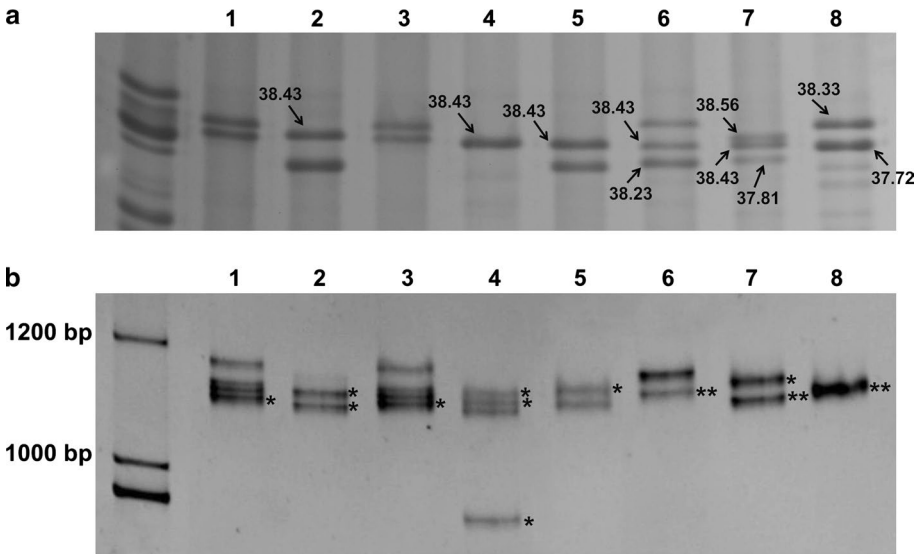

Fig. 1 SDS-PAGE separation of LMW-i subunits (**a**) and PCR amplification of LMW-i glutenin genes (**b**) from *T. urartu*. Lane: *1* PI 428255; *2* PI 428188; *3* PI 428225; *4* PI 428253; *5* TRI 11496; *6* PI 428328; *7* PI 428186; and *8* PI 428191. Each asterisk indicates one amplicon sequenced. Numbers near each protein band indicate deduced molecular weights (kDa) based on the amplicon sequenced information

Data analysis

The sequences obtained were analysed and compared using the Geneious Pro ver. 5.0.3 software (Biomatters Ltd.). Phylogenetic tree was constructed with MEGA5 software (Tamura et al. 2011) using the complete coding sequences obtained together with the sequences of the LMW-i genes identified in common wheat by Zhang et al. (2013), also as the LMW-i gene isolated from common wheat cv. Chinese Spring (NCBI: AY453154; Zhang et al. 2004) and cv. Glenlea (NCBI: AY542896; Cloutier et al. 2001). Six sequences isolated from einkorn (*T. monococcum* L. ssp. *monococcum*; $2n = 2\times = 14$, A^mA^m) were also included in this comparison (NCBI: AY146588-2 and AY146588-3, Wicker et al. 2003; NCBI: DQ307388, DQ307389 and DQ345449, An et al. 2006; and NCBI: DQ234068, Ma et al. 2006). Neighbour-joining cluster with all sequences analysed was generated using the Maximun Composite likelihood method (Tamura et al. 2004) and one bootstrap consensus from 1000 replicates was used (Felsenstein 1985).

DNA analyses were conducted by DnaSP ver. 5.0 (Librado and Rozas 2009) and parameters such as total number of mutations (η), average number of nucleotide differences (k) and number of polymorphic sites (s) were calculated. Nucleotide diversity was estimated as theta (θ), the number of segregating (polymorphic) sites (Watterson 1975), and pi (π), the average number of nucleotide differences per site between two sequences (Nei 1987). Tests of neutrality were performed using Tajima's D statistic (1989).

Results

Isolation and variation of LMWGs genes

Eight of 24 *Glu-A3* alleles described by Caballero et al. (2008) using SDS-PAGE electrophoresis were analysed by PCR amplification: four were catalogued as rare (frequency $\leq$5 %), and the other four were very rare (frequency $\leq$1 %). Six of them (three rare and three very rare) were detected in Turkish accessions, while the other two, one rare and other very rare, were of Lebanese and Iraqi origin, respectively (Table 1). These alleles are formed by several protein components (1–3 bands, Fig. 1a). In the current analysis, the specific amplification of the LMWG genes from these accessions revealed several amplicons (Fig. 1b), although some of them could be LMW-m genes, which were not the main subject of this study, and not LMW-i. Thus, 1–3 LMWG sequences were identified for each accession. In total we sequenced 15 sequences (Table 1), but some of them were the same. As the result, there are 11 different sequences with a size range of 894–1062 bp, among which seven were active genes and four were pseudogenes (KJ780779, KJ780780, KJ780783/KJ801156 and KJ780787) because they had 1–3 in-frame stop codons. The presence of these pseudogenes made it difficult to establish a univocal relation between the subunits observed in SDS-PAGE (Fig. 1a) and the amplicons detected (Fig. 1b). For the *Glu-A^{u}3af* and *Glu-A^{u}3aq* (Fig. 1, lanes 1 and 3), all the LMW-i sequences evaluated were pseudogenes, while for the rest of the alleles the sequences analysed showed very similar sizes (Table 1), being consequently difficult to separate the fragments in

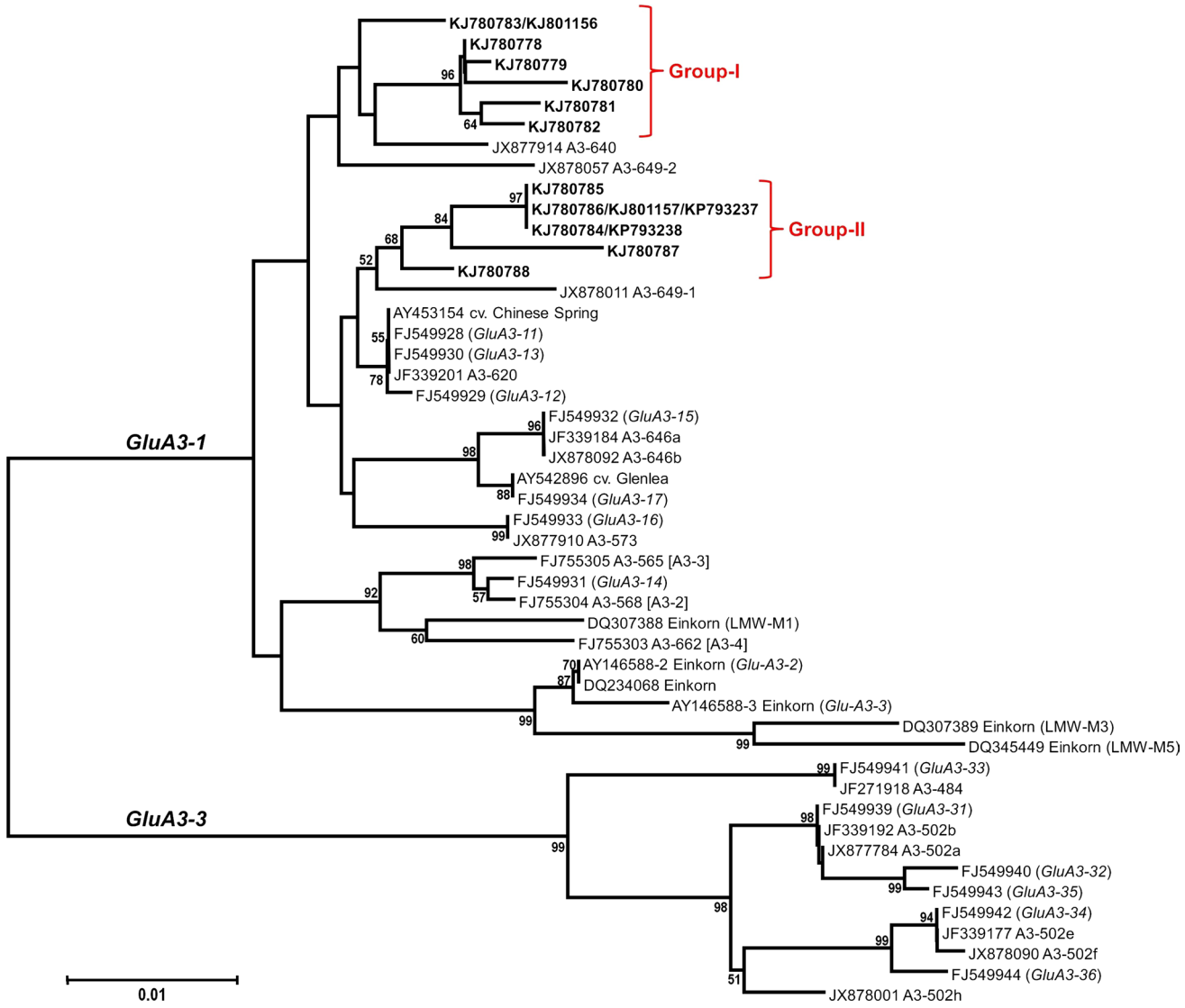

Fig. 2 Neighbour-joining tree based on the maximum composite likelihood method of LMW-i gene sequences detected in *T. urartu* accessions (*bold*), together with the previous sequences described in einkorn and common wheat. The denomination used by these last authors appears between *brackets*. *Numbers* in nodes indicate bootstrap estimates from 1000 replications

the gel. For this reason, some sequences have been associated to one unique band as shown in Fig. 1b. The deduced amino acid sequences of all had an isoleucine as the first residue in the mature protein (codon 61-63) and were classified as LMW-i genes.

The 11 genes obtained here were analysed together with other LMW-i genes from einkorn and common wheat to evaluate the relationships between them (Fig. 2). The sequences were grouped into two clusters that corresponded to the *GluA3-1* and *GluA3-3* genes described by Wang et al. (2010). There were different groups within each cluster. The *GluA3-1* gene had up to three sets: the first one associated with *A3-640/A3-649-2* genes described by Zhang et al. (2013); the second with the *A3-573/A3-620/A3-646a/A3-646b/A3-649-1* genes, and the third with *A3-565/A3-568/A3-662*—named in a previous study as *A3-2*, *A3-3* and *A3-4*, respectively (Dong et al. 2010). The other cluster (*GluA3-3*) showed two groups associated with *A3-484* and *A3-502*, respectively. The *T. urartu* sequences were grouped in two sets that we named Group-I and -II inside the *GluA3-1* cluster (Fig. 2). Group-I of *T. urartu* appeared associated with the *A3-640/A3-649-2* sequences, whereas Group-II showed more similarity with the *A3-573/A3-620/A3-646a/A3-646b/A3-649-1* sequences. All LMW-i einkorn sequences appeared clearly separated from the *T. urartu* sequences (Fig. 2). Five of them formed a differentiated cluster, while the other one (LMW-M1) was associated

Table 2 DNA polymorphism and test statistics for selection of 11 sequences from *T. urartu*

	Signal peptide	Repetitive	C-terminal			Complete
			I	II	III	
Group-I sequences						
Region	1–60	61–507	508–732	733–891	892–1056	1–1056
Size	60	447	225	159	165	1056
η	1	15	5	3	1	25
k	0.53	5.13	1.67	1.00	0.33	8.67
s	1	14	5	3	1	24
SS	0	4	1	2	1	8
NSS	1	11	4	1	0	17
h	2	5	5	4	2	6
$\theta \times 10^{-3}$	7.3	13.7	9.7	8.9	2.7	10.1
$\pi \times 10^{-3}$	8.9	11.5	7.4	6.8	2.0	8.3
D	0.850	−1.347	−1.337	−1.233	−0.933	−1.313
Group-II sequences						
Region	1–60	61–552	553–777	778–939	940–1083	1–1083
Size	60	492	225	162	144	1083
η	1	7	3	6	2	19
k	0.40	3.20	1.40	2.60	0.80	8.40
s	1	7	3	6	2	19
SS	1	3	2	4	0	10
NSS	0	4	1	2	2	9
h	2	3	3	3	3	3
$\theta \times 10^{-3}$	8.0	9.8	7.4	20.0	6.7	10.3
$\pi \times 10^{-3}$	6.7	9.4	7.2	18.1	5.6	9.5
D	−0.816	−0.338	−0.175	−0.668	−0.972	−0.582
Overall sequences						
Region	1–60	61–552	553–777	778–939	940–1104	1–1104
Size	60	492	225	162	165	1104
η	2	18	9	10	6	45
k	0.51	5.74	2.62	3.49	2.87	15.24
s	2	17	9	10	6	44
SS	1	6	4	5	3	19
NSS	1	12	5	5	3	26
h	3	8	8	7	5	9
$\theta \times 10^{-3}$	11.4	17.4	15.8	23.7	14.2	17.2
$\pi \times 10^{-3}$	8.5	17.3	13.4	24.2	20.0	17.4
D	−0.778	−0.293	−0.622	−0.096	1.593	−0.039

η total number of mutations, k average number of nucleotide differences, s number of polymorphic sites, *SS* synonymous substitutions, *NSS* non-synonymous substitutions, h number of haplotypes, θ Watterson's estimate, π nucleotide diversity and D Tajima's estimate D test

with the A3-662 sequence described by Zhang et al. (2013).

All Group-I sequences had a deletion of 45 bp inside the repetitive domain along with a 21-bp insertion in C-terminal domain III. For Group-II, there were up to three different types. One of them had a deletion of 6 bp in the repetitive domain, whereas the others had two deletions (3 and 24 bp) inside the same domain. A third type was characterized by a long deletion of 177 bp covering the end of the repetitive domain and the beginning of the C-terminal domain.

Comparison of the complete sequences showed the highest level of polymorphism in Group-I sequences, with 25 mutations at 24 polymorphism sites and 17 non-synonymous changes (Table 2). For Group-II sequences, there were 19 polymorphic sites with nine non-synonymous

changes (Table 2). For all sequences (Group-I and -II), the total number of polymorphism sites and mutations were higher than those of individual groups, with 44 and 45, respectively; of those polymorphic sites 26 were non-synonymous changes (Table 2). The repetitive domain showed the largest degree of variation for all sequences and for each individual group, followed by C-terminal II for overall sequences and the Group-II alleles. For the Group-I sequences, the second most variable domain was the C-terminal I, with five mutations and polymorphism sites. The most conserved domain was the C-terminal III for overall sequences.

Assessment of the nature and function of the seed storage proteins determined that these genes were evolutionarily neutral. In this respect, estimation of the nucleotide diversity of the DNA sequences, obtained here by two statistics—pi (π) and theta (θ)—suggested that this diversity was associated with a drift–mutation balance, consistent with a neutral equilibrium shown by a non-significant Tajima's D test (Table 2).

Deduced amino acid sequence analysis

The size of deduced mature proteins of Group-I was 327-331 amino acid residues with the deduced molecular weights of 37.72–38.23 kDa, while those of Group-II were 331–332 amino acid residues with the deduced molecular weights of 38.33–38.56 kDa (Table 1). The relation between these deduced proteins and the bands detected by SDS-PAGE (Fig. 1a) was established by the further analysis with MALDI–TOF–MS. As shown in Fig S1, the mass spectrometric data detected apparent protein peaks of each accession with molecular weights consistent with that from deduced protein (Table 1). Only in two cases (lane 7: PI 428186; and lane 8: PI 428191), the MALDI–TOF–MS analysis did not show a clear discrimination among the different subunits. Both profiles showed one unique peak associated with three or two subunits evaluated, respectively (Fig. S1b). The small differences observed between both Mr data could be the result of post-translational modification as other authors have reported (Lauriere et al. 1996; An et al. 2006).

The additional analysis of deduced sequences showed that the length of the repetitive domain was 155–162 residues in Group-II and 149 in Group-I sequences. The Group-I sequences had 19 repeat motifs in this domain with a range of 6–9 residues, whereas the Group-II had large motifs with up to 14 residues, with the exception of KJ780788 with a motif of 16 residues. This last sequence had only 19 motifs, while all other Group-II sequences had 20 (Table 1).

Analysis of these repeat motifs showed the presence in these sequences of peptides (Glt-156: PFSQQQQSPF, and Glt-17: PFSQQQQQ), considered as stimulating epitopes of T-cells in celiac disease by Vader et al. (2002). The Glt-156 peptide was detected twice in all these sequences, with the exception of KJ780788 which was detected once. In addition, the Glt-17 peptide was found once in overall sequences, but KJ780782 did not present this motif (Fig. S2).

The glutamine and proline contents in mature protein were higher in Group-II than in Group-I, with means of 53.9 and 52.6 %, respectively; this rate was conserved when only the repetitive domain was analysed, with 73.1 and 72.7 %, respectively. The highest glutamine and proline contents in overall sequences were for the KJ780788 sequence, with 53.4 % for mature protein and 73.0 % for the repetitive domain; and the lowest was for KJ780781, with 52.4 and 72.4 %, respectively (Table 1).

In the current study, two LMW-i subunits previously detected in two common wheat cultivars (Chinese Spring and Glenlea) were used to compare amino acid sequences. The AY453154 sequence of cv. Chinese Spring showed a high similarity with the AY542896 sequence of cv. Glenlea. Both sequences showed nine amino acid changes: six in the repetitive domain and one in each C-terminal, together with two InDels in the repetitive domain and one in C-II terminal (Fig. S3). Comparison of the deduced mature protein of the *T. urartu* sequences and these two LMW-i subunits is shown in Table 3.

The first two positions indicated appeared as one InDel inside the AY453154 sequence, with four of the other 17 changes similar between the last sequence and the overall *T. urartu* sequences. There was no change in the cysteine backbone (eight residues) characteristic of these LMW-glutenin subunits. However, there were some changes surrounding the cysteine residues: Arg234 → Gln and Arg342 → Thr in overall subunits, and Ala213 → Val in the KJ780781 sequence.

Proline and glutamine residues are important in the maintenance of LMWG structures. Five replacements affected proline in the subunits of Group-I. In three positions (7, 24 and 287), the existing proline was changed for another residue (Ala, Ser and Gln, respectively), while the other two replacements led to a proline residue in position 63. In two positions (78 and 301), glutamine was changed for another residue (Table 3). For Group-II, only two changes affected proline residues: Pro287 → Gln and Ser112 → Pro, while there were changes in glutamine residues in four positions: 12, 21, 178 and 299 (Table 3).

In addition to these amino acid changes, some important InDels mainly in the repetitive and C-III domains were detected in these sequences of *T. urartu* (Fig. S3). All sequences showed one extended deletion in the repetitive domain, with 38 residues in Group-I sequences and 23 in Group-II. Furthermore, the KJ780788 sequence

Table 3 Amino acid comparison of two LMW-i subunits from common wheat and the Group-I and -II sequences detected in *T. urartu*

	Repetitive domain								
Position[a]	7	12	21	24	27	44	63	78	112
AY453154 (cv. Chinese spring)	–	–	Gln	Pro	Ser	Leu	Pro	Gln	Pro
AY542896 (cv. Glenlea)	Pro	Gln	Gln	Pro	Ser	Ser	Leu	Gln	Ser
Group I									
KJ780778	Ala					Ala	Pro		–
KJ780781	Ala			Ser	Leu	Ala	Pro		–
KJ780782	Ala				Leu	Ala	Pro	His	–
Group-II									
KJ780784/KP793238		Glu	Lys				Pro		Pro
KJ780785		Glu	Lys				Pro		Pro
KJ780786/KJ801157/KP793237		Glu	Lys				Pro		Pro
KJ780788		Glu					Pro		Pro

	C-I terminal			C-II terminal				C-III terminal		
Position[a]	178	213	234	256	287	299	301	324	339	342
AY453154 (cv. Chinese Spring)	Gln	Ala	Gln	Ile	Gln	Gln	Gln	His	Thr	Thr
AY542896 (cv. Glenlea)	Gln	Ala	Arg	Ile	Pro	Gln	Gln	His	Thr	Arg
Group I										
KJ780778			Gln	Val	Gln		Glu		Asn	Thr
KJ780781		Val	Gln	Val	Gln		Glu		Asn	Thr
KJ780782			Gln	Val	Gln		Glu		Asn	Thr
Group-II										
KJ780784/KP793238	Arg		Gln		Gln	Leu		Tyr		Thr
KJ780785	Arg		Gln		Gln	Leu		Tyr		Thr
KJ780786/KJ801157/KP793237	Arg		Gln		Gln	Leu		Tyr		Thr
KJ780788			Gln		Gln	Leu				Thr

[a] The positions are determined by the mature protein of the AY542896 sequence (LMWG-50) isolated by Cloutier et al. (2001)

presented one additional deletion (PPFSQQQQ) between residues 171 and 178 of the reference sequence (AY542896). For Group-I, two insertions of one glutamine residue were detected: one in the repetitive domain (residues 110–111 of AY542896) and other in the C-II domain (residues 276–277). This first insertion was also detected in all Group-II sequences, while the latter was not found in the KJ780788 sequence, which had a unique deletion (HGTFLQP) in the C-III domain (Fig. S3).

Discussion

The low variation of some traits of interest in wheat breeding has suggested the possibility of searching for variation in wheat relatives. Among them, the species identified as putative donors of the three genomes present in common wheat could be the main candidates, because they are included in the primary gene pool and their crossing with wheat does not require special techniques.

Numerous studies have suggested that *T. urartu* is the species donor of the A genome (Dvorak et al. 1993), an event that likely occurred ~0.5 million years ago (Huang et al. 2002; Dvorak and Akhunov 2005). Among the traits related to wheat quality that could be transferred from this wild diploid wheat to modern wheat are seed storage proteins (glutenins and gliadins). A particular characteristic of the A genome is the presence of a LMWG type (LMW-i) not detected in the other wheat genomes. These glutenin subunits are exclusively synthesised by genes of the *GluA3* locus, consequently *T. urartu* is the natural source of these subunits. In this study, seven genes and four pseudogenes of this LMWG type were detected and sequenced in eight *T. urartu* accessions.

Recent studies suggested that the LMWG genes in the A genome are synthesised by two genes for the LMW-m glutenin subunits and 2–4 genes for the LMW-i glutenin subunits (Zhang et al. 2013). According to previous classifications (Wang et al. 2010), the LMW-m subunits are encoded by the *GluA3-2* gene, and LMW-i by the *GluA3-1* and *GluA3-3*. The reconciliation of this classification with

the data of Zhang et al. (2013) and data of the present study suggest that the *GluA3-2* gene of Wang et al. (2010) corresponds with the m_{AD} cluster defined by Zhang et al. (2013), which was formed by two genes.

The *GluA3-3* gene could be also constituted by two genes: one most frequent formed by all variants of the *A3-502* gene (Zhang et al. 2013), together with all alleles of this gene described by Wang et al. (2010), without the *GluA3-33* allele that appeared associated with the other gene (*A3-484*), which is less frequent (Zhang et al. 2013). This latter sub-group of the *GluA3-3* gene formed the i_A-5 haplotype according to Zhang et al. (2013), together with one sub-group of the *GluA3-1* gene constituted by the *A3-565/A3-568/A3-662* genes, previously described as *A3-2*, *A3-3* and *A3-4* by Dong et al. (2010). In this last group could also be included the LMW-i genes found in *T. monococcum* L. ssp. *monococcum* by Wicker et al. (2003) and Ma et al. (2006). This suggests that this group could have a different origin to that indicated by these authors, and this should be studied in the future.

Similar to the other groups found, the LMW-i haplotype (i_A) could be generally formed by up to three genes: one or two for *Glu-A3-1* and one for *GluA3-3*, being the two sequence sets of *T. urartu* obtained in this study associated with both genes of *GluA3-1*. Nucleotide diversity was high, which could have two main causes: the neutral nature of these genes to evolution (Gepts 1990), and the role of these proteins in plant biology where they are mainly a source of amino acids. These high values are frequent in wild species because lack of selection pressure has generated no advantage of one variant over others. However, in cultivated species, this effect is counteracted by unconscious selection on the part of farmers of the allelic variants due to their function in quality properties of desirable food products. This narrows genetic variability by genetic drift processes (Gepts 1990).

Six of the eight accessions evaluated in this study showed several active LMW-i genes and the relation between their deduced proteins and the bands detected by SDS-PAGE could be established. However, for the other two accessions active LMW-i genes were not found among the PCR products obtained in current study. Some of these PCR products were identified as LMW-m genes (results not shown), which open the possibility of the expressed proteins of these accessions could be LMW-m, although no LMW-m protein encoded has been identified so far.

Both *T. urartu* groups showed clear differences inside their sequences, mainly in the presence of exclusive SNPs and InDels. These InDels appeared mainly in the repetitive domain, whose structure of tandem motifs could generate duplication or deletion of one or various motifs by slippage during replication (Cassidy and Dvorak 1991). These duplications or deletions may cause repetitive domains to be generally large, which could be associated with dough quality as suggested by Masci et al. (1998, 2000). A large repetitive domain provides a greater number of glutamines available for inter-molecular interactions through hydrogen bonds that could strengthen the gluten network, thus increasing dough elasticity. Furthermore, SNP variations could also be a source of functional changes, particularly those substitutions affecting the structure of LMWGs (Tanaka et al. 2005; An et al. 2006). Although the repetitive domain of all these sequences showed deletions that make these sequences are smaller than those in common wheat, in general, all deduced proteins from Group-II genes possess a longer repetitive domain than for Group-I. Furthermore, several SNPs detected in these *T. urartu* genes, mainly the proline and glutamine substitutions, could affect gluten strength because this alters protein structure and so affect its elasticity.

Another important feature of LMWGs is the presence of eight cysteine residues, which are important in forming the inter- and intra-molecular disulphide bonds that determine the structure of the gluten polymer (D'Ovidio and Masci 2004). These cysteines were conserved in the 11 LMW-i genes detected in *T. urartu*. However, positions around the cysteine residues are also important in the formation of disulphide bonds and appear to be highly conserved (Masci et al. 1998). Substitutions at this level were found in the present study. Two changes could affect three cysteine residues in all alleles, while the KJ780781 allele could affect five of eight cysteine residues. This replacement could affect protein structure and consequently that of the gluten network.

Although celiac disease has been mainly associated with the gliadins, some authors have found reactive epitopes in the HMWG and LMWG- subunits (van de Wal et al. 1998; Vader et al. 2002). Two epitopes described by Vader et al. (2002) were found in the LMW-i genes detected in *T. urartu* accessions of the present study. Curiously, due to the InDel present in the repetitive domain, these proteins showed one additional motif not present in the LMW-i subunits of common wheat used as a reference, what means that probably could be more reactive. This suggests that although the quality genetic pool of common wheat could be increased by introgression of these proteins, the new materials have no advantage in possible use by celiac patients. On the contrary, some LMW-m subunits detected in wheat do not present these epitopes (results not shown). Although further studies should be carried out, this suggests that the LMW-i subunits could be related to celiac disease of wheat flour and the LMW-m subunits are not.

In conclusion, *T. urartu* was shown to be an important source of novel variation for LMW-i genes. All 11 novel genes detected were associated with *Glu-A3-1* genes and showed differences from those in common wheat, which

could lead to functional changes. Therefore, these alleles might be useful in wheat breeding for quality improvement. Further studies are required to evaluate the effect of the novel alleles on the quality of modern wheat and their relationship to celiac disease.

Author contribution statement CG and JBA conceived and designed the study. SC performed the experiments. All authors analysed the data and wrote the paper. All of them have read and approved the final manuscript.

Acknowledgments This research was supported by Grant AGL2010-19643-C02-01 from the Spanish Ministry of Economy and Competitiveness, co-financed with European Regional Development Fund (FEDER) from the European Union. The first author is grateful to the Spanish Ministry of Economy and Competitiveness (FPI programme) and European Social Fund for a predoctoral fellowship. We thank to the National Small Grain Collection (Aberdeen, ID, USA) and the Institute for Plant Genetics and Crop Plant Research (Gatersleben, Germany) for supplying the analysed material. Also, we thank to Dr. E. Chicano (Proteomic Unit, IMIBIC, Córdoba, Spain) for the MALDI-TOF–MS analysis.

Compliance with ethical standards

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

Ethical standards The authors declare that the experiments comply with the current laws of Spain.

References

Alvarez JB, Martín A, Martín LM (2001) Variation in the high molecular weight glutenin subunits coded at the *Glu-H^{ch}1* locus in *Hordeum chilense*. Theor Appl Genet 102:134–137

Alvarez JB, Gutierrez MV, Guzman C, Martin LM (2013) Molecular characterisation of the amino- and carboxyl-domains in different *Glu-A1x* alleles of *Triticum urartu* Thum. ex Gandil. Theor Appl Genet 126:1703–1711

An X, Zhang Q, Yan Y, Li Q, Zhang Y, Wang A, Pei Y, Tian J, Wang H, Hsam SLK, Zeller FJ (2006) Cloning and molecular characterization of three novel LMW-i glutenin subunit genes from cultivated einkorn (*Triticum monococcum* L.). Theor Appl Genet 113:383–395

Caballero L, Martín MA, Alvarez JB (2008) Allelic variation for the high- and low-molecular-weight glutenin subunits in wild diploid wheat (*Triticum urartu*) and its comparison with durum wheats. Aust J Agric Res 59:906–910

Cassidy BG, Dvorak J (1991) Molecular characterization of a low-molecular-weight glutenin cDNA clone from *Triticum durum*. Theor Appl Genet 81:653–660

Cassidy BG, Dvorak J, Anderson OD (1998) The wheat low-molecular-weight glutenin genes: characterization of six new genes and progress in understanding gene family structure. Theor Appl Genet 96:743–750

Cloutier S, Rampitsch C, Penner GA, Lukow OM (2001) Cloning and expression of a LMW-i glutenin gene. J Cereal Sci 33:143–154

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

Dong L, Zhang X, Liu D, Fan H, Sun J, Zhang Z, Qin H, Li B, Hao S, Li Z, Wang D, Zhang A, Ling HQ (2010) New insights into the organization, recombination, expression and functional mechanism of low molecular weight glutenin subunit genes in bread wheat. PLoS One 5:e13548

Dvorak J, Akhunov ED (2005) Tempos of gene locus deletions and duplications and their relationship to recombination rate during diploid and polyploid evolution in the *Aegilops-Triticum* alliance. Genetics 171:323–332

Dvorak J, Terlizzi Pd, Zhang H-B, Resta P (1993) The evolution of polyploid wheats: identification of the A genome donor species. Genome 36:21–31

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

Gepts P (1990) Genetic diversity of seed storage proteins in plants. In: Brown AHD, Clegg MT, Kahler AL, Weir BS (eds) Plant population genetics, breeding and genetic resources, 1st edn. Slinauer Associattes, Sunderland, pp 64–82

Huang S, Sirikhachornkit A, Su X, Faris J, Gill B, Haselkorn R, Gornicki P (2002) Genes encoding plastid acetyl-CoA carboxylase and 3-phosphoglycerate kinase of the *Triticum/Aegilops* complex and the evolutionary history of polyploid wheat. Proc Natl Acad Sci USA 99:8133–8138

Ikeda TM, Nagamine T, Fukuoka H, Yano H (2002) Identification of new low-molecular-weight glutenin subunit genes in wheat. Theor Appl Genet 104:680–687

Jauhar PP (1993) Alien gene transfer and genetic enrichment of bread wheat. In: Damania AB (ed) Biodiversity and wheat improvement. ICARDA-A Wiley Sayce Publication, Aleppo, pp 103–119

Johnson BL (1975) Identification of the apparent B-genome donor of wheat. Can J Genet Cytol 17:21–39

Laemmli UK (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227:680–685

Lauriere M, Bouchez I, Doyen C, Eynard L (1996) Identification of glycosylated forms of wheat storage proteins using two-dimensional electrophoresis and blotting. Electrophoresis 17:497–501

Lee YK, Bekes F, Gupta R, Appels R, Morell MK (1999) The low-molecular-weight glutenin subunit proteins of primitive wheats. I. Variation in A-genome species. Theor Appl Genet 98:119–125

Librado P, Rozas J (2009) DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451–1452

Ma ZC, Wei YM, Long H, Yan ZH, Baum B, Zheng YL (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

Martín MA, Martín LM, Alvarez JB (2008) Polymorphisms at the *Gli-A^{u}1* and *Gli-A^{u}2* loci in wild diploid wheat (*Triticum urartu*). Euphytica 163:303–307

Masci S, D'Ovidio R, Lafiandra D, Kasarda DD (1998) Characterization of a low-molecular-weight glutenin subunit gene from bread wheat and the corresponding protein that represents a major subunit of the glutenin polymer. Plant Physiol 118:1147–1158

Masci S, D'Ovidio R, Lafiandra D, Kasarda DD (2000) A 1B-coded low-molecular-weight glutenin subunit associated with quality in durum wheats shows strong similarity to a subunit present in some bread wheat cultivars. Theor Appl Genet 100:396–400

McIntosh RA, Yamazaki Y, Dubcovsky J, Rogers WJ, Morris C, Appels R, Xia XC (2013) Catalogue of gene symbols for wheat. http://www.shigen.nig.ac.jp/wheat/komugi/genes/macgene/2013/GeneSymbol.pdf. Accessed 20 June 2015

Miller TE (1987) Systematic and evolution. In: Lupton FGH (ed) Wheat breeding: its scientific basis. Chapman & Hall, London, pp 1–30

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York

Payne PI (1987) Genetics of wheat storage proteins and the effects of allelic variation on bread-making quality. Ann Rev Plant Physiol 38:141–153

Pogna PE, Autran JC, Mellini F, Lafiandra D, Feillet P (1990) Chromosome 1B-encoded gliadins and glutenin subunits in durum wheat: genetics and relationship to gluten strength. J Cereal Sci 11:15–34

Rasheed A, Xia X, Yan Y, Appels R, Mahmood T, He Z (2014) Wheat seed storage proteins: advances in molecular genetics, diversity and breeding applications. J Cereal Sci 60:11–24

Rodríguez-Quijano M, Nieto-Taladriz MT, Carrillo JM (1997) Variation in B-LMW glutenin subunits in Einkorn wheats. Genet Resour Crop Evol 44:539–543

Singh NK, Shepherd KW (1988) Linkage mapping of genes controlling endosperm proteins in wheat. 1. Genes on the short arms of group-1 chromosomes. Theor Appl Genet 75:628–641

Srivastava JP, Damania AB (1989) Use of collections in cereal improvement in semi-arid areas. In: Brown AHD, Frankel OH, Marshall DR, Williams JT (eds) The use of plant genetic resources. Cambridge University Press, Cambridge, pp 88–104

Stacey J, Isaac P (1994) Isolation of DNA from plants. In: Isaac PG (ed) Methods in molecular biology: protocols for nucleic acid analysis by non-radioactive probes. Humana Press, Totawa, pp 9–15

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585–595

Tamura K, Nei M, Kumar S (2004) Prospects for inferring very large phylogenies by using the neighbor-joining method. Proc Nat Acad Sci USA 101:11030–11035

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28:2731–2739

Tanaka H, Shimizu R, Tsujimoto H (2005) Genetical analysis of contribution of low-molecular-weight glutenin subunits to dough strength in common wheat (*Triticum aestivum* L.). Euphytica 141:157–162

Vader W, Kooy Y, van Veelen P, de Ru A, Harris D, Benckhuijsen W, Peña S, Mearin L, Drijfhout JW, Koning F (2002) The gluten response in children with celiac disease is directed toward multiple gliadin and glutenin peptides. Gastroenterology 122:1729–1737

van de Wal Y, Kooy YMC, van Veelen PA, Peña SA, Mearin LM, Molberg Ø, Lundin KEA, Sollid LM, Mutis T, Benckhuijsen WE, Drijfhout JW, Koning F (1998) Small intestinal T cells of celiac disease patients recognize a natural pepsin fragment of gliadin. Proc Natl Acad Sci USA 95:10050–10054

Wang L, Li G, Peña RJ, Xia X, He Z (2010) Development of STS markers and establishment of multiplex PCR for *Glu-A3* alleles in common wheat (*Triticum aestivum* L.). J Cereal Sci 51:305–312

Watterson GA (1975) On the number of segregating sites in genetical models without recombination. Theor Popul Biol 7:256–276

Wicker T, Yahiaoui N, Guyot R, Schlagenhauf E, Liu Z-D, Dubcovsky 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

Wrigley CW, Bekes F, Bushuk W (eds) (2006) Gliadin and glutenin: the unique balance of wheat quality. AACC International Press, St. Paul

Zhang W, Gianibelli MC, Rampling LR, Gale KR (2004) Characterisation and marker development for low molecular weight glutenin genes from *Glu-A3* alleles of bread wheat (*Triticum aestivum* L). Theor Appl Genet 108:1409–1419

Zhang X, Liu D, Jiang W, Guo X, Yang W, Sun J, Ling H, Zhang A (2011) PCR-based isolation and identification of full-length low-molecular-weight glutenin subunit genes in bread wheat (*Triticum aestivum* L.). Theor Appl Genet 123:1293–1305

Zhang X, Liu D, Zhang J, Jiang W, Luo G, Yang W, Sun J, Tong Y, Cui D, Zhang A (2013) Novel insights into the composition, variation, organization, and expression of the low-molecular-weight glutenin subunit gene family in common wheat. J Exp Bot 64:2027–2040