J. Agr. Sci. Tech. (2014) Vol. 16: 203-215

Molecular Characterization and Phylogenetic Analysis of Novel *α-gliadin* Genes from *Triticum dicoccoides* L.

D. L. Zhang[1], A. L. Gao[1], Y. G. Li[1], Y. R. Su[1], S. B. He[1], and S. P. Li[1]*

ABSTRACT

The quality traits of *Triticum dicoccoides* KU-13441 (*Triticum turgidum* L. var. *dicoccoides* (AABB, 2n= 4x= 28)) were analyzed by mixograph, and the results showed that *T. dicoccoides* KU-13441 had desirable qualities in gluten strength and flour stirring tolerance. Subsequently, seventeen novel full-ORF *α-gliadin* genes and thirty-five pseudogenes were cloned and sequenced from *T. dicoccoides* KU-13441. Among the 17 novel full-ORF *α-gliadin* genes, the putative proteins of *Gli2-TD-44* and *Gli2-TD-46* contained an extra cysteine residue, located in the first nonrepetitive region and *N*-terminal repetitive domain, respectively, rather than in the second nonrepetitive region like other α-gliadins. Prokaryotic expression analysis and western-blotting indicated that these two *α-gliadin* genes could be successfully expressed under the control of T_7 promoter. According to the varying numbers of 4 typical coeliac disease toxic peptides and glutamine residues in the two ploy glutamine domains among the 17 α-gliadins, *Gli2-TD-39*, *Gli2-TD-46* and *Gli2-TD-47* genes were assigned to sub-genome *B* and other 14 genes were assigned to sub-genome *A*. Phylogenetic analysis including two *S-genome* species, *Aegilops longissima* (S^l) and *Aegilops speltoides* (S) revealed that the α-gliadin sequences of the *B* genome in *T. dicoccoides* had closer genetic relationship with those from *Ae. speltoides*. This implies that *Ae. speltoides* might participate in the origin of wheat *B* genome.

Keywords: Prokaryotic expression, Quality traits, Western-blotting, Wild emmer wheat.

INTRODUCTION

Mature wheat grains consist of embryos and endosperms. Gluten, is a heterogeneous mixture of wheat storage proteins present in the endosperm. In wheat gluten, the dominant proteins are prolamins, which are a major determinant of the extensibility and elasticity of dough and are mainly composed of alcohol-soluble (gliadin) and alcohol-insoluble (glutenin) fractions (Naghavi *et al.*, 2013). Gliadins are traditionally categorized into four groups α-, β-, γ- and ω-gliadins, based on their polyacrylamide gel electrophoresis at low pH. Since α- and β-gliadins are closely related in sequences and structures, they are more usually referred to as α-gliadins (Anderson *et al.*, 1997). Among these groups, a-gliadins, with an average molecular weight of 30~45 kDa, are the most abundant gliadins, accounting for 15~30% of proteins in wheat seeds. Therefore, α-gliadins are the most consumed storage proteins by humans (Wieser, 2007).

A feature of α-gliadin is the presence of six conserved cysteine residues, which presumably form three intra-molecular disulfide bonds (Anderson *et al.*, 1997). Interestingly, some α-gliadins contain extra cysteine residues, allowing for the formation of inter-molecular disulfide bonds. Changes in the positions of cysteine residues can affect the patterns of intra- and inter-molecular disulfide bond formation and possibly lead to a positive effect on bread-making quality (Susan *et al.*, 2010). In addition, the α-gliadins and some glutenins from *Triticum* contain several peptides that constitute the main toxic components in coeliac disease (CD). The peptides in α-gliadins mainly harbour 4 CD-

[1] College of Life Science, Henan University, Kaifeng, 475001, Henan, People's Republic of China.
*Corresponding author; e-mail: lisuoping@henu.edu.cn

[DOR: 20.1001.1.16807073.2014.16.1.5.9]

epitopes (Solid *et al.*, 2012), including *DQ2.5-Glia-α1a* (PFPQPQLPY), *DQ2.5-Glia-α2* (PQPQLPYPQ), *DQ2.5-Glia-α3* (FRPQQPYPQ) and *DQ8-Glia-α1* (QGSFQPSQQ), and each of them originates from a specific region of α-gliadins.

Heterologous protein expression in prokaryotic organisms such as *E. coli* is an effective method for the functional characterization of individual proteins. Previously, several high-molecular-weight and low-molecular-weight glutenin subunits (HMW-GS and LMW-GS) genes were expressed in *E. coli* (Li *et al.*, 2008; Yan *et al.*, 2009). To date, however, limited studies focusing on the expression of *gliadin* genes in *E. coli* have been reported since the high-level *in vitro* expression of *gliadin* genes remains difficult (Tamás and Shewry, 2006). The endosperm protein expressed *in vitro* can be added to basic flour, which will help to identify the function of individual *gliadin* gene.

Wild emmer wheat, *Triticum turgidum* L. var. *dicoccoides* (AABB, 2n= 4x= 28) is an annual, predominantly self-pollinated plant with large elongated grains and brittle ears disarticulating at maturity into spikelets. It has many important agricultural characteristics, such as higher resistance to pathogens including stripe rust, stem rust, and powdery mildew (Anikster *et al.*, 2005). In addition, the grains of *T. dicoccoides* possess a higher quality of storage proteins with extensive allelic variations that give rise to a wide range of genotypes (Xu *et al.*, 2004). Thus, *T. dicoccoides* is thought to be an important resource for new genes in wheat quality improvement. Recently, molecular cloning and sequence analysis on the allelic genes of seed storage proteins in this species have been reported (Qi *et al.*, 2006; Li *et al.*, 2008; Jin *et al.*, 2012). Nevertheless, little genetic information of its *α-gliadin* genes has been obtained, and extra cysteine residues that generate possibly positive effects on bread-making quality have not been found.

In the present work, the quality traits of *T. dicoccoides* KU-13441 including gluten strength and flour stirring tolerance were intensively studied by mixograph, and novel *α-gliadin* genes were cloned from *T. dicoccoides* KU-13441 through allele-specific polymerase chain reaction (AS-PCR) to explore the molecular basis of quality traits of *T. dicoccoides* KU-13441.

MATERIALS AND METHODS

Plant Materials

An accession of *T. dicoccoides* KU-13441 was provided by Dr. Takashi ENDO from the National BioResource Project-Wheat, Kyoto University. *Triticum aestivum* Yumai 34 was preserved in Plant Germplasm Resources and Genetic Engineering Laboratory, Henan University, which was planted and harvested in the same condition with *T. dicoccoides* KU-13441 and used as the control of quality analysis.

Quality Trait Analysis

Ground flour was obtained by Quadrumat Junior (Brabender Corp., German). The kneading parameters, mid-line peak time (MPT), mid-line peak height (MPH), angle of descent and 8 min. curve width were measured on a mixograph (National Mfg, Corp., German) through 54240A (AACC method) with ten grams of whole wheat flour.

Molecular Cloning and Phylogenetic Analysis of *α-gliadin* Genes

Genomic DNA from young leaves was extracted according to the method reported by Yan *et al.* (2009). Based on the alignments of sequences of some *α-gliadin* genes published in GenBank, a pair of primers (P-1: 5′-ATGAAGACCTTTCTCATCCT-3′ and P-2: 5′-TCAGTTAGTACCGAAGATGC-3′) was designed using Primer Premier 5. PCR reactions and molecular cloning program were performed according to the method reported by Chen *et al.* (2008). To ensure the accuracy of the sequence acquired, an enhanced annealing temperature and LA-Tag DNA polymerase (Takara, Japan) with high fidelity were used in the PCR reactions. Each clone was sequenced twice. Sequences of *α-gliadin* gene from *T. dicoccoides* were aligned using Clustal X 1.83. The neighbor-joining phylogenetic distance tree was generated and edited by MEGA 4.0.

[DOR: 20.1001.1.16807073.2014.16.1.5.9]

Prokaryotic Expression and Western Blotting

The cloned *α-gliadin* genes were amplified again by a pair of designed primers P3: 5′-*CGCGGATCCATGAAGACCTTTCTCATCCT*-3′ and P4: 5′-*CCCAAGCTTTCAGTTAGTACCGAAGATGC*-3′ (BamH I and Hind III restriction site is underlined). Expression of proteins in *E. coli* was induced with 1 mM isopropyl β-Δ-thiogalactopyranoside (IPTG) for 6 hours using the cells without IPTG as reference. The separation of the expressed proteins was carried out by SDS-PAGE according to Jin *et al.* (2012). Western blotting was carried out as previously described by Tamás and Shewry (2006). Anti-His-tag mouse monoclonal antibody was used to detect the expressed proteins based on the His-tag sequence present downstream of the cloned gene.

Identification of CD-toxic Peptides and Chromosomal Locations

The CD-toxic peptides in the gliadins were identified according to Cornell and Wills-Johnson (2001). Chromosomal locations of cloned *α-gliadin* genes were determined based on the method of Van Herpen *et al.* (2006).

RESULTS

Quality Traits of Wild Emmer Wheat

T. aestivum Yumai 34 is known for its strong gluten and high quality variety (Xing *et al.*, 2002). In the result of mixograph, the values of mid-line peak time (MPT) and 8 min curve width of *T. aestivum* Yumai 34 were obviously higher than those of *T. dicoccoides* KU-13441, respectively, indicating the better quality of *T. aestivum* Yumai 34 in dough mixing tolerance and elasticity (Figure S1 and Table 1). In contrast, *T. dicoccoides* KU-13441 exhibited higher mid-line peak height (MPH, 56.009), indicating the stronger stirring tolerance than *T. aestivum* Yumai 34. In addition, *T. dicoccoides* KU-13441 showed high gluten strength, as exhibited by its low descent angle (Figure S1 and Table 1). These results revealed that *T. dicoccoides* KU-13441 possessed desirable qualities in gluten strength and flour stirring tolerance, which could be exploited as a potential material in wheat quality breeding.

Isolation and Sequencing of *α-gliadin* Genes

The products with expected sizes (about 900bp) were obtained from the genomic DNA of *T. dicoccoides* KU-13441 by AS-PCR (Figure S2). The purified PCR products were cloned, in which 105 positive clones were selected and sequenced. A total of 52 full length genes with sequence length of 845~957 bp were obtained and temporarily named *Gli2-TD-1~Gli2-TD-52*. Thirty-five genes, *Gli2-TD-1~Gli2-TD-35*, were determined to be pseudogenes (Table S1). Nine of the 35 pseudogenes, *Gli2-TD-1~Gli2-TD-9*, were

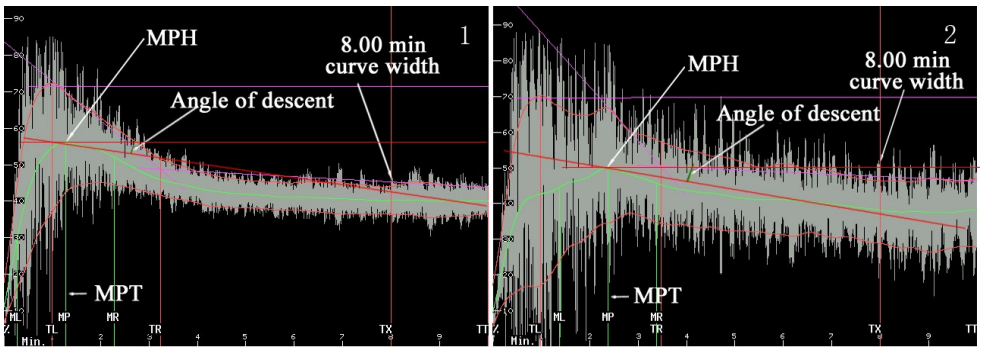

Figure S1. Analysis of the quality effect of *T. dicoccoides* KU-13441 and *T. aestivum* Yumai34 by mixograph. 1: *T. dicoccoides* KU-13441; 2: *T. aestivum* Yumai 34. The flour kneading parameters are represented by the mid-line peak time (MPT), mid-line peak height (MPH), angle of descent and 8 min. curve width.

[DOR: 20.1001.1.16807073.2014.16.1.5.9]

Table1. Comparison of mixographic parameters between *T. dicoccoides* KU-13441 and *T. aestivum* Yumai 34.

Specific name	Mid-line peak time (Min)	Mid-line peak height (cm)	Angle of descent (º)	8min curve width (cm)
T. dicoccoides KU-13441	1.29	56.009	13.8	7.637
T. aestivum Yumai 34	2.38	49.889	18.8	17.572

generated from frame-shift mutation. The other 26 pseudogenes (Gli2-TD-10~Gli2-TD-35) contained at least one in-frame stop codon. Mutations from CAG (Q) to TAG (stop codon) and from CAA (Q) to TAA (stop codon) appeared 9 and 30 times, respectively. Mutations from GAG (E) to TAG (stop codon), TAT (Y) to TAG (stop codon) and TTG (L) to TAG (stop codon) were also observed, indicating the diversity of means in forming pseudogene containing in-frame stop codons.

Seventeen novel full-ORF gene sequences (Gli2-TD-36~Gli2-TD-52) were also obtained (GenBank accession numbers JQ340381~JQ340386, JQ340388 and JQ824846~JQ824855, Table 2). The nucleotide similarities among the 17 gene sequences varied from 81.7 to 99.8%. Their sequence identities with the published *α-gliadin* gene from *Triticeae* ranged from 90 to 98%. Sequence comparison showed that all cloned genes contained typical structural features of previously characterized *α-gliadin* genes, including no introns and ending with stop codon TGA. Since the function of proteins expression cannot be performed on pseudogenes, only the 17 full-ORF genes were further analyzed in the following work.

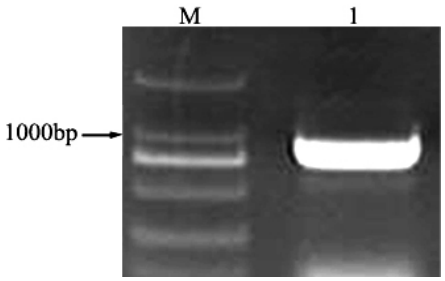

Figure S2. PCR amplification products of *α-gliadin* gene from *T. dicoccoides* KU-13441. M: DNA marker (DL2000); 1: *T. dicoccoides* KU-13441

Molecular Characters of 17 *α-gliadin* Genes

The putative protein sequences were deduced from the seventeen novel genes according to the triplet code rule. *Gli2-TD-36* and *Gli2-TD-37* shared 100% identity in the protein level. High structural similarities between the 17 α-gliadins with the typical wheat α-gliadin (Genbank accession number DQ246446) were found (Figure 1), including a signal peptide with 20 amino acid residues (S), a *N*-terminal repetitive domain(R), the first polyglutamine region (Q1), the first nonrepetitive region (NR1), the second polyglutamine region (Q2), and the second nonrepetitive region (NR2). Multiple sites of variation among the 17 α-gliadins were observed in each region (Table 2). Compared with the other 14 proteins, the mutation from proline (P) to serine (S) appeared in the *N*-terminal of *Gli2-TD-39, Gli2-TD-46* and *Gli2-TD-47*. The "ST" peptide motif was inserted into the *C*-terminal of *Gli2-TD-39, Gli2-TD-48* and *Gli2-TD-49*, different from the other 14 proteins, among which the mutation from isoleucine (I) to asparagines (N) was found in *Gli2-TD-48.* Moreover, the predicted isoelectric points (p*I*) among the 17 α-gliadins exhibited great variations, indicating that the properties of 17 genes were highly diverse.

Interestingly, an extra cysteine residue appeared in *Gli2-TD-44* and *Gli2-TD-46* besides six cysteine residues in the conserved *NR* region of the 17 α-gliadins. Unlike the normal location of C-residues in the *NR2* region of wheat α-gliadin (Chen *et al.*, 2008; Xie *et al.*, 2010), the extra cysteine residues of *Gli2-TD-44* and *Gli2-TD-46* were present in the *NR1* and *R* regions, respectively (Figure 2). For the prokaryotic expression of proteins, recombinant plasmids were constructed containing the full open reading frames of

[DOR: 20.1001.1.16807073.2014.16.1.5.9]

Table S1. Patterns and sites of mutation present in the nucleotide sequences of 35 *α-gliadin* pseudogenes from *T. dicoccoides* KU-13441.

Pseudogene	Pattern of mutation	Site of mutation	Pseudogene	Pattern of mutation	Site of mutation	Pseudogene	Pattern of mutation	Site of mutation
Gli2-TD-1	Deletion A	403	*Gli2-TD-14*	CAA-TAA [a]	91, 616	*Gli2-TD-23*	CAG-TAG [a]	85, 361
Gli2-TD-2	Deletion A	403		TTG-TAG	455		CAA-TAA	364
Gli2-TD-3	Deletion A	403	*Gli2-TD-15*	CAG-TAG [a]	85, 361	*Gli2-TD-24*	TAT-TAG	321
Gli2-TD-4	Deletion T	841		CAA-TAA	364	*Gli2-TD-25*	CAA-TAA [a]	91, 376
Gli2-TD-5.	Deletion T	856	*Gli2-TD-16*	CAA-TAA [a]	121, 235	*Gli2-TD-26*	CAA-TAA [a]	91, 376
Gli2-TD-6	Insertion T	22		CAG-TAG [a]	85, 667	*Gli2-TD-27*	CAA-TAA	355
Gli2-TD-7	Deletion C	705		GAG-TAG	118	*Gli2-TD-28*	CAA-TAA [a]	91, 619
Gli2-TD-8	Deletion A	335	*Gli2-TD-17*	CAG-TAG	157		TTG-TAG	458
Gli2-TD-9	Deletion A	403	*Gli2-TD-18*	CAA-TAA	355	*Gli2-TD-29*	CAG-TAG	157
Gli2-TD-10	CAA-TAA	355	*Gli2-TD-19*	CAA-TAA [a]	91, 616	*Gli2-TD-30*	CAA-TAA	355
Gli2-TD-11	CAA-TAA*	91, 622		TTG-TAG	455	*Gli2-TD-31*	CAA-TAA	421
	TTG-TAG	461	*Gli2-TD-20*	CAA-TAA [a]	91, 619	*Gli2-TD-32*	TAT-TAG	321
Gli2-TD-12	CAA-TAA	355		TTG-TAG	458	*Gli2-TD-33*	CAA-TAA [a]	91, 616
Gli2-TD-13	CAA-TAA	355	*Gli2-TD-21*	CAA-TAA	355		TTG-TAG	455
	CAG-TAG	526	*Gli2-TD-22*	CAA-TAA	355	*Gli2-TD-34*	CAA-TAA	355
Gli2-TD-35	TAT-TAG	321						

[a] Star indicates that double nonsense mutations with the same type were observed in the sequence of one gene.

Gli2-TD-44 and *Gli2-TD-46*. The fusion proteins were about 35~38 kDa (Figure S3). The molecular weight of the objective protein was approximately 33~36 kDa after subtracting 2.16 kDa of the His-tag protein. This molecular weight was consistent with the theoretical molecular weight of *Gli2-TD-44* and *Gli2-TD-46*. Western-blotting results further confirmed that these two genes could be successfully expressed under the control of T_7 promoter (Figure S4).

Identification of Specific CD Epitopes and Chromosomal Location of 17 *α-gliadin* Genes

The 19-residue motif (LGQQQPFPPQQPYPQPQPF) in the *N*-terminal repetitive domains was considered to be active in celiac disease (Wieser, 2001). In addition, short motif (LGQGSFRPSQQN) in the *C*-terminal domain could be associated with adenovirus type 12 infections (Kasarda and D'Ovidio, 1999). Fourteen of the 17 α-gliadins contained the above two specific peptides, except for *Gli2-TD-39*, *Gli2-TD-46* and *Gli2-TD-47* (Figure 1). In addition, based on the alignment of 17 α-gliadin sequences, three α-gliadins (Gli2-TD-39, Gli2-TD-46 and Gli2-TD-47) did not have any of the four types of *CD* epitopes while each of the other 14 α-gliadins contained 1~2 *CD* epitopes (Table 2). Besides, these three α-gliadins had 16~18 more glutamine residues compared with the other 14 α-gliadins in the *Q2* region (Table 2). Chromosomal locations of 17 *α-gliadin* genes were determined based on the method of Van Herpen *et al.* (2006). The α-gliadin derived from the wheat *A* genome almost invariably contains *DQ2.5-Glia-α1a* and *DQ2.5-Glia-α3*, yet has incomplete *DQ2.5-Glia-α2* and *DQ8-Glia-α1*. Most of the α-gliadins that originate from the wheat *B* genome do not contain any of the 4 of *CD* epitopes mentioned above. In addition, Compared with those from wheat *B* genomes, the first polyglutamine region (Q1) derived from the wheat *A* genome contains a significant larger number of glutamine residues. In the second polyglutamine region (Q2), the α-gliadin derived from wheat *B* genome shows a larger number of glutamine residues. Therefore, the three *α-gliadin genes* (Gli2-TD-39, Gli2-TD-46 and Gli2-TD-47) could be assigned to *T. dicoccoides B* genome. Likewise, the other 14 genes could be assigned to *T. dicoccoides* A genome.

To further test the above results of chromosomal location, a phylogenetic tree was constructed based on the sequences of the 17 α-gliadins from this study and those of the known A^u genome of *Triticum urartu*

[DOR: 20.1001.1.16807073.2014.16.1.5.9]

Table 2. Molecular characters and chromosomal locations of 17 cloned full-ORF *α-gliadin* genes from *T. dicoccoides* KU-13441.

Gene	Accession number	*N*-terminal sequence	*C*-terminal sequence	Predicted molecular weight (kD)	Predicted pI	Four typical CD epitopes				Gglutamine residues (Q1)	Glutamine residues (Q2)	Chomosomal location
						DQ2.5-Glia-α1a	*DQ2.5-Glia-α2*	*DQ2.5-Glia-α3*	*DQ8-Glia-α1*			
Gli2-TD-36	JQ340381	*VRVPVPQLQPQ*	*CTIAPFGIFGTN*	34.09	6.98	+	-	+	-	27	7	6A
Gli2-TD-37	JQ340382	*VRVPVPQLQPQ*	*CTIAPFGIFGTN*	34.09	6.98	+	-	+	-	27	7	6A
Gli2-TD-38	JQ340383	*VRVPVPQLQPQ*	*CTIAPFGIFGTN*	33.15	8.24	-	-	+	-	20	6	6A
Gli2-TD-39	JQ340384	*VRVSMPQLQLQ*	*CSTTIAPFGIFGTN*	35.94	6.26	-	-	-	-	17	23	6B
Gli2-TD-40	JQ340385	*VRVPVPQLQPQ*	*CTIAPFGIFGTN*	32.77	6.99	+	-	+	-	21	7	6A
Gli2-TD-41	JQ340386	*VRVPVPQLQPQ*	*CTIAPFGIFGTN*	34.08	6.98	+	-	+	-	27	7	6A
Gli2-TD-42	JQ340388	*VRVPVPQLQPQ*	*CTIAPFGIFGTN*	33.83	6.98	+	-	+	-	25	7	6A
Gli2-TD-43	JQ824846	*VRVPVPQLQPQ*	*CTIAPFGIFGTN*	32.84	8.24	-	-	+	-	18	6	6A
Gli2-TD-44	JQ824847	*VRVPVPQLQPQ*	*CTIAPFGIFGTN*	32.62	8.20	+	-	+	-	20	7	6A
Gli2-TD-45	JQ824848	*VRVPVPQLQPQ*	*CTIAPFGIFGTN*	32.75	6.56	+	-	+	-	20	7	6A
Gli2-TD-46	JQ824849	*VRVSMPQLQLQ*	*CTIAPFGIFGTN*	35.92	6.56	-	-	-	-	17	23	6B
Gli2-TD-47	JQ824850	*VRVSMPQLQLQ*	*CTIAPFGIFGTN*	35.90	6.26	-	-	-	-	17	23	6B
Gli2-TD-48	JQ824851	*VRVPVPQLHPQ*	*CSTTIAPFGNFGTN*	32.76	7.67	+	-	+	-	21	6	6A
Gli2-TD-49	JQ824852	*VRVPVPQLQPQ*	*CSTTIAPFGIFGTN*	32.85	8.24	-	-	+	-	18	6	6A
Gli2-TD-50	JQ824853	*VRVPVPQLQPQ*	*CTIAPFGIFGTN*	32.70	8.24	-	-	+	-	17	6	6A
Gli2-TD-51	JQ824854	*VRVPVPQLQPQ*	*CTIAPFGIFGTN*	32.84	8.24	-	-	+	-	18	6	6A
Gli2-TD-52	JQ824855	*VRVPVPQLQPQ*	*CTIAPFGIFGTN*	32.85	8.57	-	-	+	-	18	5	6A

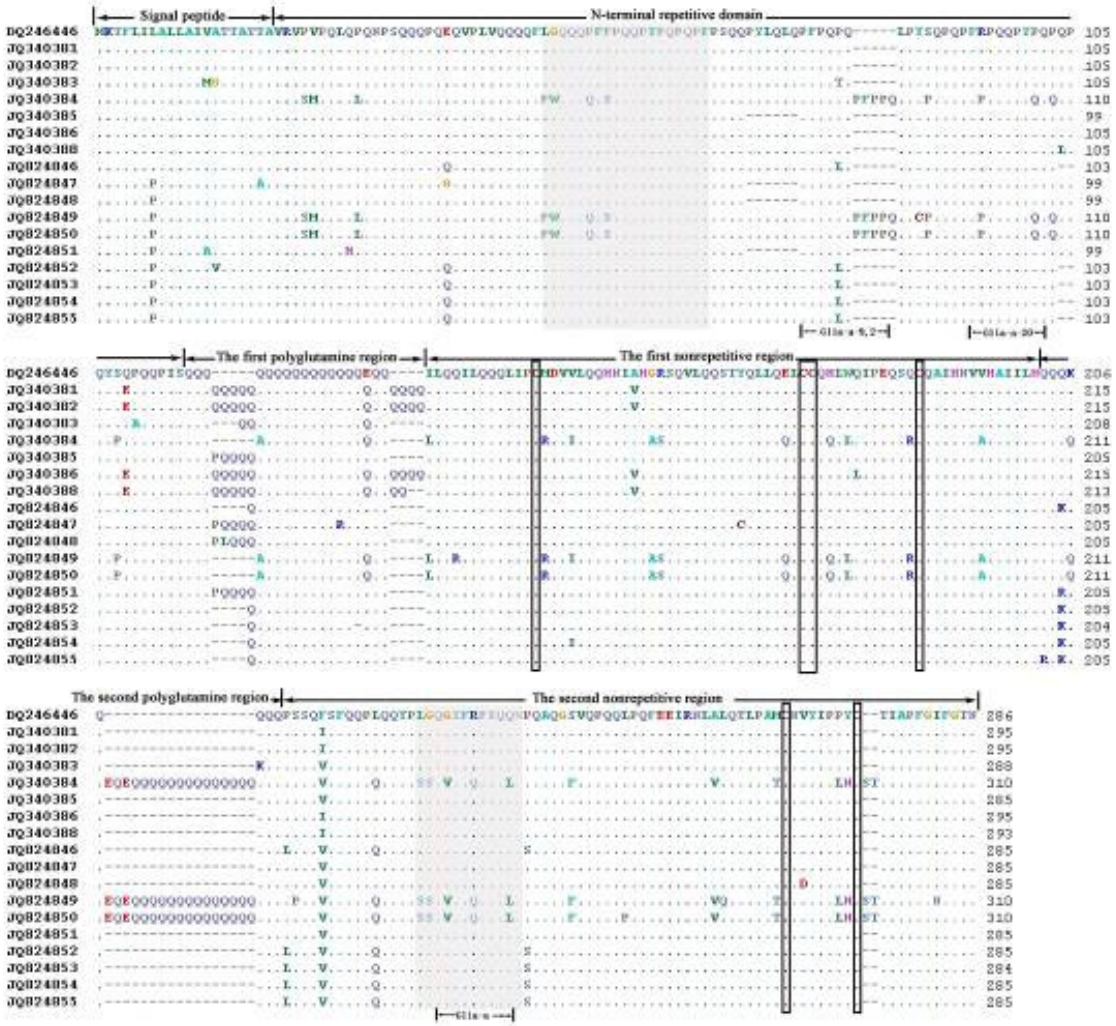

Figure 1. Multiple sequence alignment of 18 α-gliadins. DQ246446 derives from common wheat, and the other 17 sequences are obtained in this study. Dash represents deletion of amino acids residues. Dot indicates the identical amino acid residues. Black frame shows the positions of cysteine residues. Gray segment of the aligned sequences shows the position of the peptides, which have activity in celiac disease (in the *N*-terminal domain) and are associated with adenovirus type 12 infections (in the *C*- terminal domain).

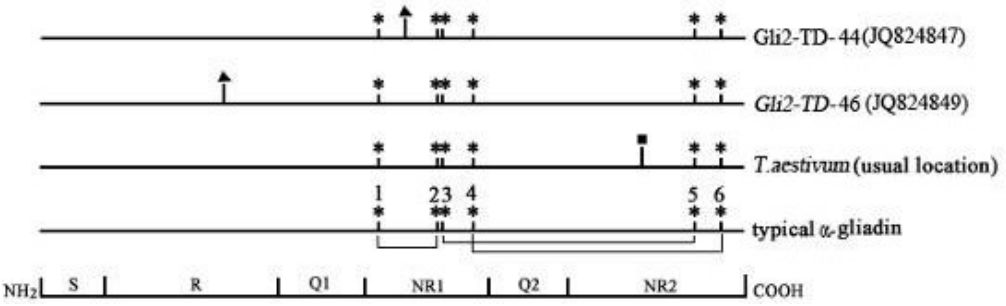

Figure 2. Distribution of cysteine residues in typical α-gliadins and putative α-gliadins identified in this study. Triangles: The cysteine residues involved in inter-molecular *S*–S bonds in this study; Square: The cysteine residues usually involved in inter-molecular *S–S* bonds in common wheat; Stars: The cysteine residues involved in intra-molecular S–S bonds in previous reports, 1–6: Conserved cysteine residue position in typical a-gliadin.

[DOR: 20.1001.1.16807073.2014.16.1.5.9]

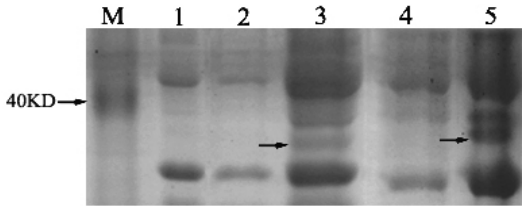

Figure S3. Analysis of induced expression products of *Gli2-TD-44* and *Gli2-TD-46* by SDS-PAGE.
M: Blue plus^{TM} protein marker; 1: Protein expressed from pET-28a(+) plasmid, induced by IPTG; 2,4: Protein expressed from the recombinant plasmids *PET(28a)-TD-44* and *PET(28a)-TD-46* without induction; 3,5: Protein expressed from the recombinant plasmids *PET(28a)-TD-44* and *PET(28a)-TD-46* induced by IPTG, Arrow indicates the expressed products of *Gli2-TD-44* and *Gli2-TD-46.*

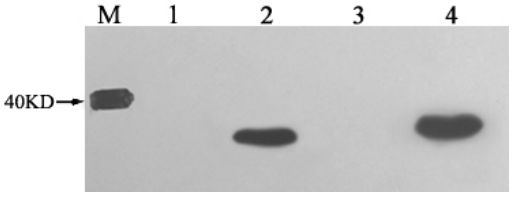

Figure S4. Analysis of the expression products of Gli2-*TD-44* and Gli2-*TD-46* by western blotting.
M: Blue plusTM protein marker; 1, 3: Recombinant plasmid PET(28a)-*TD-44* and PET(28a)-*TD-46* without induction, 2, 4: Recombinant plasmid PET(28a)-*TD-44* and PET(28a)-*TD-46* with induction.

(Genbank accession numbers JF927571, JF927573, JF927575 and JF927577) and *B* genome of wheat (Genbank accession numbers K02068, K03075, M11073 and U51303) (Figure S5). *Gli2-TD-39*, *Gli2-TD-46* and *Gli2-TD-47* derived from *T. dicoccoides* and 4 *α-gliadin* genes of the wheat *B* genome belong to one group, indicating that the 3 genes have a close genetic relationship with the wheat *B* genome. This result further verifies the conclusion mentioned above that *Gli2-TD-39, Gli2-TD-46* and *Gli2-TD-47* originate from the *T. dicoccoides B* genome. The other 14 genes of *T. dicoccoides* and 4 *α-gliadin* genes of the A^u genome of *T. urartu* belong to the other group, implying a close genetic relationship of these 14 genes with the *A* genome.

Phylogenetic Analysis of *α-gliadin* Gene

To explore the origin of the wheat *B* genome, another phylogenetic tree was constructed with the sequences of 17 α-gliadins obtained in this study and those from all of the species of *Aegilops* registered in the Genbank (Figure 3), including part of the α-gliadin of *Aegilops speltoides*, *Aegilops longissima*, *Aegilops tauschii*, *Aegilops comosa*, *Aegilops uniaristata*, *Aegilops umbellulata*, *Aegilops kotschyi* and *Aegilops markgrafii*. Three α-gliadins originate from the *T. dicoccoides B* genome (Gli2-TD-39, Gli2-TD-46 and Gli2-TD-47) grouped with the sequences from *Ae. speltoides* (bootstrap value 73%), indicating that the α-gliadin sequence of the *T. dicoccoides B* genome has a close genetic relationship with that of *Ae. speltoides*. Comparatively, the α-gliadin sequences of other species from *Aegilops* exhibited low homology to those of the *T. dicoccoides B* genome. The above results provided theoretical evidence for the participation of *Ae. speltoides* in the origin of wheat *B* genome.

DISCUSSION

The number of *α-gliadin* genes is highly variable among wheat and its ancestors, ranging from 25 to 150 copies per haploid genome because the duplication of *α-gliadin* genes and base mutation happened during evolutionary process. However, only a few *α-gliadin* genes have been detected by protein electrophoresis. Many base mutations in the *α-gliadin* genes are nonsense or frame-shift mutations, causing 50% of *α-gliadin* genes to be pseudogenes (DuPont *et al.*, 2004). In this study, the transition of *C* to *T* leads to the premature appearance of a stop codon in 26 pseudogenes (Gli2-TD-10~Gli2-TD-35), which may be responsible for the nonsense mutation, since the sequences of *α-gliadin* gene contain abundant glutamine codons (CAG and CAA). In addition, 9 pseudogenes (Gli2-TD-1~Gli2-TD-9) contained frame-shift mutations (1 base lost in most cases).

[DOR: 20.1001.1.16807073.2014.16.1.5.9]

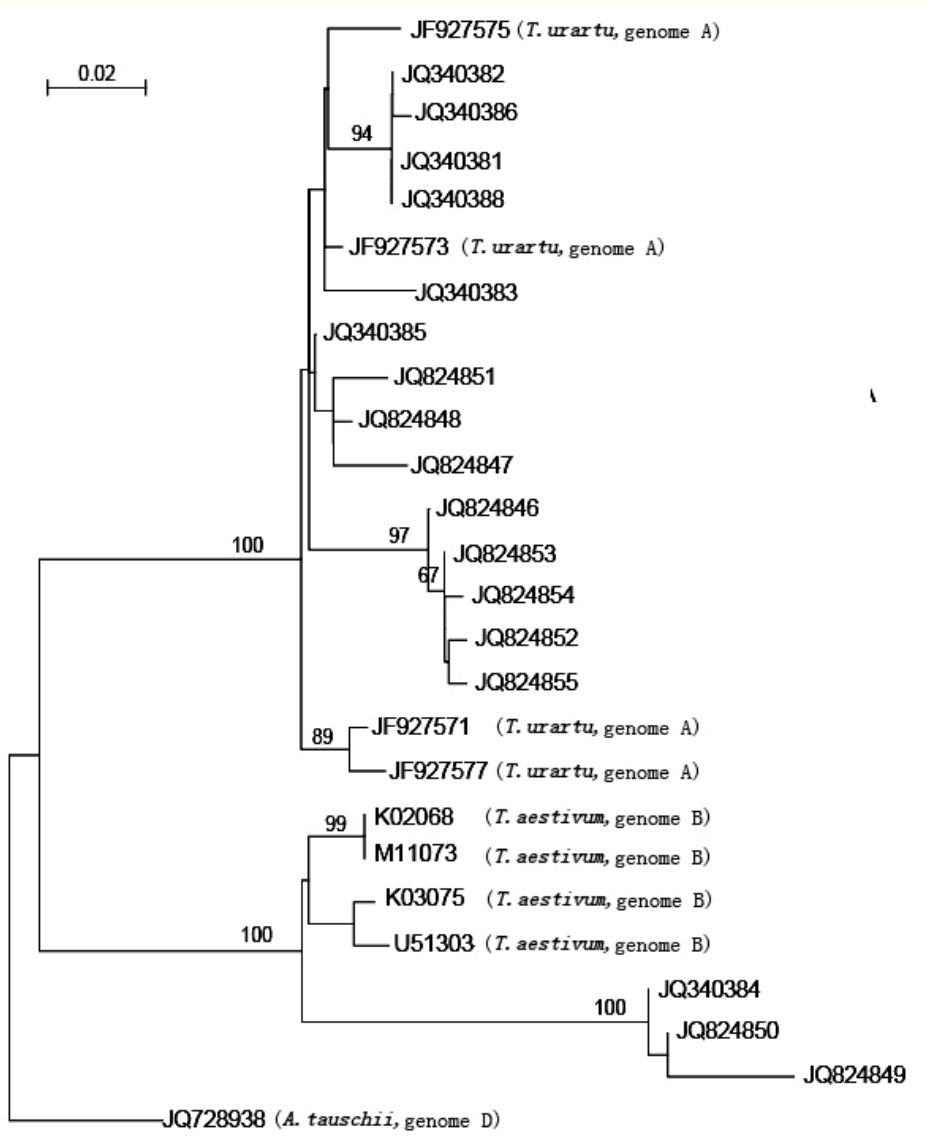

Figure S5. Phylogenetic analysis of the amino acid sequences of the 17 full-ORF *α-gliadin* genes from this study and 8 *prolamin* genes from common wheat and *T. urartu*. Genbank accession numbers JF927571, JF927573, JF927575 and JF927577 come from *T. urartu*. Genbank accession numbers K02068, K03075, M11073 and U51303 come from common wheat. The α-gliadin of *A.tauschii* (Genbank accession number JQ728938) is used as an outgroup.

The extra cysteine residues have been found in some prolamins subunit (Huang *et al.*, 2010). Moreover, it has been reported that the increase in the number of cysteine residues favors the formation of inter-molecular disulfide bonds, promoting the combination of gliadin and glutenin polymer (Susan *et al.*, 2010). In the present work, the extra cysteine residue in *Gli2-TD-46* appeared in the *R* region and the one in *Gli2-TD-44* appeared in the *NR1* region, which were different from the normal position in *NR2* region of wheat α-gliadin. This might be one of the reasons why *T. dicoccoides* KU-13441 has desirable qualities in gluten strength and flour stirring tolerance. In addition, prokaryotic expression of the above two genes indicates that they might possess potential function in *T.*

[DOR: 20.1001.1.16807073.2014.16.1.5.9]

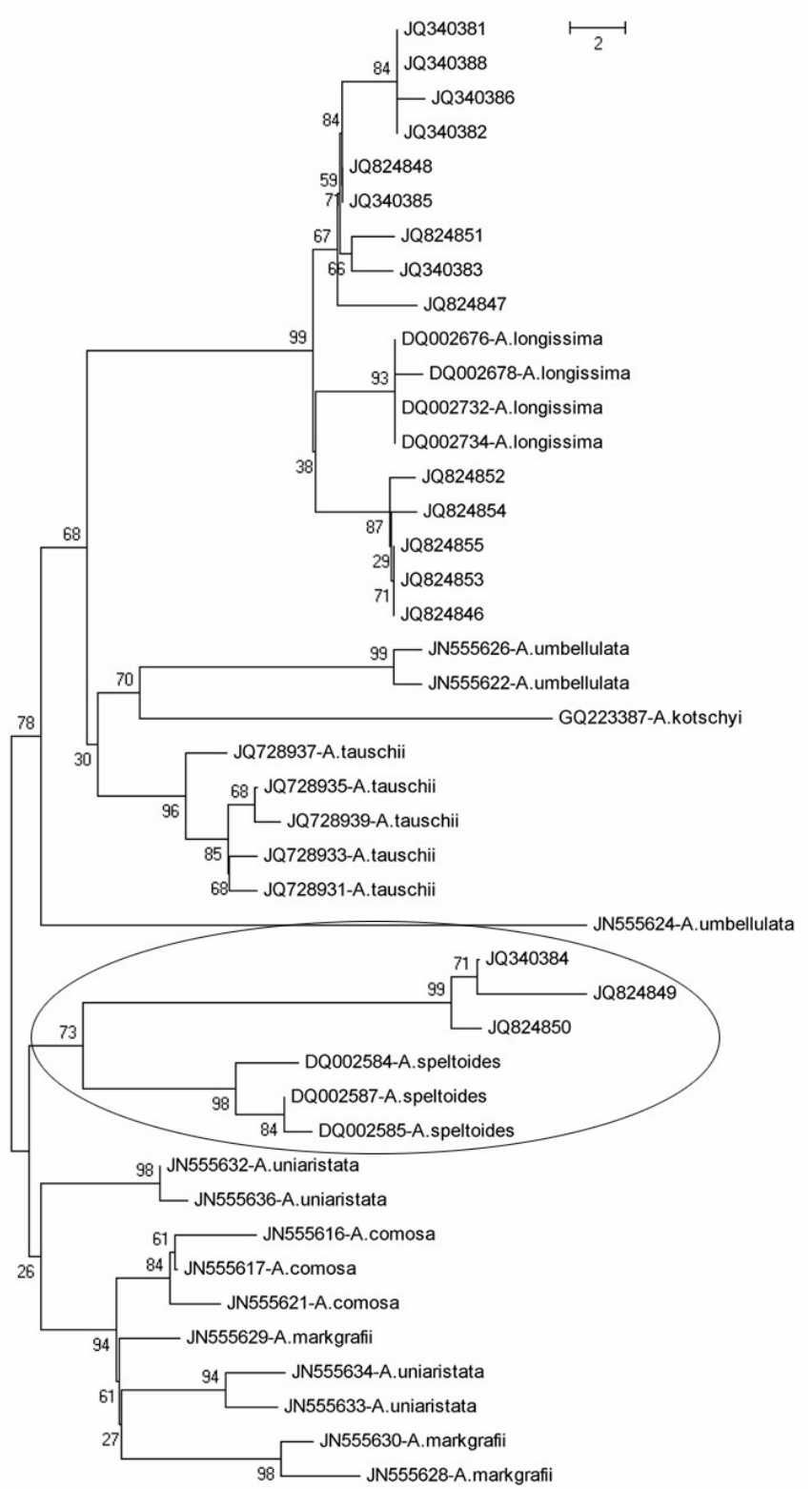

Figure 3. The evolutionary relationships based on the amino acid sequences of 17 *α-gliadin* genes from *T. dicoccoides* and those of *α-gliadin* genes from other species of *Aegilops* .

[DOR: 20.1001.1.16807073.2014.16.1.5.9]

dicoccoides, which provide novel candidate genes for wheat quality improvement. The origin of the *B* genomes of wheat is an important research field. Up to now, there are two prevailing viewpoints on this including monophyletic and polyphyletic origin. The polyphyletic origin view believes that the *B* genome is a recombined genome derived from two or more diploid *Aegilops* species based on a high level of differentiation in the *B* genome (Liu *et al.*, 2003). On the contrary, recent studies suggest that the *B* genome might originate from one certain species from *Aegilops Sitopsis,* supporting the monophyletic origin view (Haider, 2013). Haider (2012) has conducted a phylogenetic analysis of chloroplast DNA between *T. aestivum* and 8 different *Aegilops* species using cleaved amplified polymorphic sequence (CAPS) and sequencing on 28 chloroplast loci. The result indicated that *Ae. speltoides* might be the chloroplast donor to bread wheat and the donor of its *B* genome. The alignment of homogeneous storage protein genes has been widely applied in revealing the evolutionary relationships of *Triticeae* species (Guo *et al.*, 2011). Since the *α-gliadin* gene sequences of *Ae. sharonensis, Ae. bicornis* and *Ae. searsii* had not been registered in the NCBI database, the α-gliadin sequences of all of the species of *Aegilops* including *Ae. speltoides*, *Ae. longissima* and those from *T. dicoccoides* were compared in this work. It was found that *Gli2-TD-39, Gli2-TD-46* and *Gli2-TD-47* originating from the *T. dicoccoides B* genome formed a relatively strong group with those from *Ae. Speltoides*, whereas these three sequences and those from *Ae. longissima* shared a rather low homology. It is well known that *T. dicoccoides* is the ancestor of the *A* and *B* chromosomes of the common wheat (Ishii *et al.*, 2001). The present work implies that, out of the two distinct *S* genomes analyzed here (S^1 and S), at least *Ae. speltoides* (S) may have taken part in the evolution of the wheat *B* genome.

ACKNOWLEDGEMENTS

This work was supported by NSFC (No.31271713) and National Science and Technology Project "863" (No.2011BAD07B01-1, No. 2012AA101105). We also thank the National BioResource Project-Wheat, Kyoto University, for providing *T. dicoccoides* KU-13441.

REFERENCES

1. Anderson, O. D., Litts, J. C. and Greene, F. C. 1997. The *α-gliadin* Gene Family. I. Characterization of Ten New Wheat *A-gliadin* Genomic Clones, Evidence for Limited Sequence Conservation of Flanking DNA, and Southern Analysis of Gene Family. *Theor. Appl. Genet.,* **95:** 50-58.
2. Anikster, Y., Manisterski, J., Long, D. L. and Leonard, K. J. 2005. Leaf Rust and Stem Rust Resistance in *Triticum dicoccoides* Populations in Israel. *Plant. Dis.,* **89:**55-62.
3. Chen, F. G., Xu, C. H., Chen, M. Z., Wang, Y. H. and Xia, G. M. 2008. A New *α-gliadin* Gene Family for Wheat Breeding: Somatic Introgression Line II-12 Derived from *Triticum aestivum* and *Agropyron elongatum*. *Mol. Breed,* **22:** 675-685.
4. Cornell, H. J. and Wills-Johnson, G. 2001. Structure-activity Relationships in Coeliac-toxic Gliadin Peptides. *Amino Acids,* **21:** 243-253.
5. DuPont, F. M., Vensel, W., Encarnacao, T., Chan, R. and Kasarda, D. D. 2004. Similarities of Omega Gliadins from *Triticum urartu* to Those Encoded on Chromosome *1A* of Hexaploid Wheat and Evidence for Their Post-translational Processing. *Theor. Appl. Genet.,* **108:** 1299-1308.
6. Guo, Z. F., Long, X. Y., Dong, P., Wei, Y. M., Bai, L. P., Dang, X. X., Wan, H. L., Zhang, L. J. and Zheng, Y. L. 2011. Molecular Cloning of Novel *α-gliadin* Genes from *Crithopsis delileana* and the Evolution Analysis with Those from Triticeae. *Genes. Genom.,* **33:** 155-161.
7. Haider, N. 2012. Evidence for the Origin of the *B* Genome of Bread Wheat Based on Chloroplast DNA. *Turk. J. Agric. For.,* **36:** 13-25.
8. Haider, N. 2013. The Origin of the *B* Genome of Bread Wheat (*Triticum aestivum* L.). *Russ. J. Genet.,* **49:** 263-274.
9. Huang, Z., Long, H., Wei, Y.M., Qi, P. F., Yan, Z. H. and Zheng, Y. L. 2010. Characterization and Classification of γ-gliadin Multigene Sequences from *Aegilops*

[DOR: 20.1001.1.16807073.2014.16.1.5.9]

Section *Sitopsis*. *Cereal. Res. Commun.,* **38:** 1-14.
10. Ishii, T., Mori, N. and Ogihara, Y. 2001. Evaluation of Allelic Diversity at Chloroplast Microsatellite Loci among Common Wheat and Its Ancestral Species. *Theor. Appl. Genet.,* **103:** 896-904.
11. Jin, M., Xie, Z. Z., Ge, P., Li, J., Jiang, S. S., Subburaj, S., Li, X. H., Zeller, F. J., Hsam, S. L. and Yan, Y. M. 2012. Identification and Molecular Characterization of HMW Glutenin Subunit 1By16* in Wild Emmer. *J. Appl. Genet.,* **53:** 249-258.
12. Kasarda, D. D. and D'Ovidio, R. 1999. Deduced Amino Acid Sequence of an *Alpha-gliadin* Gene from Spelt Wheat (Spelta) Includes Sequences Active in Celiac Disease. *Cereal Chem.,* **76:** 548-551.
13. Li, X. H., Wang, A. L., Xiao, Y. H., Yan, Y. M., He, Z. H., Appels, R., Ma, W. J., Hsam, S. L. K. and Zeller, F. J. 2008. Cloning and Characterization of a Novel Low Molecular Weight Glutenin Subunit Gene at the *Glu-A3* Locus from Wild Emmer Wheat (*Triticum turgidum* L. var. *dicoccoides*). *Euphytica,* **159:** 181-190.
14. Liu, B., Segal, G., Rong, J. K. and Feldman, M. 2003. A Chromosome-specific Sequence Common to the *B* Genome of Polyploid Wheat and *Aegilops searsii*. *Plant. Syst. Evol.,* **241:** 55-66.
15. Naghavi, M. R., Ranjbar, M., Hassani, M. H., Aghaee, M. J. and Bamneshin, M. 2013. Characterization of Iranian Accessions of *Aegilpos crassa* Boiss. Using Flow Cytometry and Protein Analysis. *J. Agr. Sci. Tech.,* **15:** 811-818.
16. Qi, P. F., Yue, Y. W., Long, H., Wei, Y. M., Yan, Z. H. and Zheng, Y. L. 2006. Molecular Characterization of *α-gliadin* Genes from Wild Emmer Wheat (*Triticum dicoccoides*). *DNA Sequence,* **17:** 415-421.
17. Saitou, N. and Nei, M. 1987. The Neighbor-joining Method: A New Method of Reconstructing Phylogenetic Trees. *Mol. Biol. Evol.,* **4:** 406-425.
18. Sollid, L. M., Qiao, S. W., Anderson, R. P., Gianfrani, C. and Koning, F. 2012. Nomenclature and Listing of Celiac Disease Relevant Gluten T-cell Epitopes Restricted by HLA-DQ Molecules. *Immunogenetics*, **64**:455-460.
19. Susan, B. A., William, H. V. and Frances, M. D. 2010. Integration of Transcriptomic and Proteomic Data from a Single Wheat Cultivar Provides New Tools for Understanding the Roles of Individual Alpha Gliadin Proteins in Flour Quality and Celiac Disease. *J. Cereal Sci.,* **3:** 1-9.
20. Tamás, L. and Shewry, P. R. 2006. Heterologous Expression and Protein Engineering of Wheat Gluten Proteins. *J. Cereal Sci.,* **43:** 259-274.
21. Van Herpen, T. W. J. M., Goryunova, S. V., van der Schoot, J., Mitreva, M., Salentijn, E., Vorst, O., Schenk, M. F., van Veelen, P. A., Koning, F.**,** van Soest, L. J. M., Vosman, B., Bosch, D., Hamer, R. J., Gilissen, L. J. W. J. and Smulders, M. J. M. 2006. *Alpha-gliadin* Genes from the *A, B,* and *D* Genomes of Wheat Contain Different Sets of Celiac Disease Epitopes. *BMC Genom.,* **7:** 1.
22. Wieser, H. 2001. Comparative Investigations of Gluten Proteins from Different Wheat Species I. Qualitative and Quantitative Composition of Gluten Protein Types. *Eur. Food Res. Technol.,* **213:** 183-186.
23. Wieser, H. 2007. Chemistry of Gluten Proteins. *Food Microbiol.*, **24:** 115-119.
24. Xie, Z. Z., Wang, C. Y., Wang, K., Wang, S. L., Li, X. H., Zhang, Z., Ma, W. J. and Yan, Y. M. 2010. Molecular Characterization of the Celiac Disease Epitope Domains in *α-gliadin* Genes in *Aegilops tauschii* and Hexaploid Wheats (*Triticum aestivum* L.). *Theor. Appl. Genet.,* **121:** 1239-1251.
25. Xing, T. M., Huang, L. X. and Dong, Y. N. 2002. Main Characteristics and High Yield Cultivation Technology fromYumai 34. *J. Triticeae Crops,* **22:** 101-102.
26. Xu, S. S., Khan, K., Klindworth, D. L., Faris, J. D. and Nygard, G. 2004. Chromosomal Location of Genes for Novel Glutenin Subunits and Gliadins in Wild Emmer Wheat (*Triticum turgidum* L. var. *dicoccoides*). *Theor. Appl. Genet.,* **108:** 1221-1228.
27. Yan, Y. M., Jiang, Y., An, X. L., Peia, Y. H., Li, X. H., Zhang, Y. Z., Wang, A. L., He, Z., Xi, X., Bekes, F. and Ma, W. 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.

[DOR: 20.1001.1.16807073.2014.16.1.5.9]

شناسایی ملکولی و تحلیل تبارزایی ژن های نول(جدید)آلفا گلیادین از *Triticum dicoccoides* L.

د. ل. ژانگ، ا. ل. گااو، ی. گ. لی، ی. ر. سو، س. ب. هی، و س. پ. لی

چکیده

صفات کیفیتی *Triticum dicoccoides* KU-13441 (*Triticum turgidum* L. var. dicoccoides (AABB, 2n=4x=28) با دستگاه نمودارمخلوط (mixograph) تجزیه و تحلیل شد و نتایج نشان داد که *Triticum dicoccoides* KU-13441 کیفیت های مطلوبی از نظر استحکام و قدرت گلوتن و تحمل همزنی آرد دارد. سپس، هفده ژن نول و full-ORF آلفا گلیادین و ۳۵ شبه ژن از *T. dicoccoides* KU-13441 همسانه سازی و توالی شدند. در میان ۱۷ ژن نول و full-ORF آلفا گلیادین پروتئین های putative شامل Gli2-TD-44 و Gli2-TD-46 دارای یک سیستئین باقیمانده اضافی بودند که به ترتیب در اولین ناحیه غیر تکراری و دامنه تکراری پایانه N قرار داشتند که از موقعیت سیستئین باقیمانده در آلفا گلیادین ها ی دیگر که در ناحیه تکراری دوم قرار دارند متفاوت بود. نتایج تجزیه شبه هسته داری تظاهر و لکه گذاری وسترنی حکایت از آن داشت که این دو ژن آلفا گلیادین می توانند تحت کنترل راه انداز T_7 با موفقیت بیان شوند. مطابق اعداد متغیر از ۴ پپتید سمی بیماری سیلیاک و باقیمانده گلوتامین در دو دامنه پلی گلوتامین، از میان ۱۷ ژن آلفا گلیادین، ژن های Gli2-TD-46 ,Gli2-TD-39 و Gli2-TD- ۴۷ به ژنوم فرعی(زیر ژنوم) B تخصیص داده شدند و ۱۴ ژن دیگر به ژنوم فرعی A اختصاص یافتند. تحلیل تبار شناسی دو گونه از ژنوم- S شامل *Aegilops longissima* (Sl) و *Aegilops speltoides* (S) آشکار کردند که توالی آلفاگلیادین ژنوم B در *T. dicoccoides* رابطه ژنتیکی نزدیکی با توالی Ae. Speltoides دارد. این امر حاکی از آن است که Ae. Speltoides احتمالا در مبداء ژنوم B گندم شرکت دارد.

[DOR: 20.1001.1.16807073.2014.16.1.5.9]