Theor Appl Genet (2010) 121:1239–1251
DOI 10.1007/s00122-010-1384-8

ORIGINAL PAPER

Molecular characterization of the celiac disease epitope domains in α-gliadin genes in *Aegilops tauschii* and hexaploid wheats (*Triticum aestivum* L.)

Zhenze Xie · Congyan Wang · Ke Wang · Shunli Wang · Xiaohui Li · Zhao Zhang · Wujun Ma · Yueming Yan

Received: 6 February 2010 / Accepted: 3 June 2010 / Published online: 17 June 2010

Abstract Nineteen novel full-ORF α-gliadin genes and 32 pseudogenes containing at least one stop codon were cloned and sequenced from three *Aegilops tauschii* accessions (T15, T43 and T26) and two bread wheat cultivars (Gaocheng 8901 and Zhongyou 9507). Analysis of three typical α-gliadin genes (*Gli-At4*, *Gli-G1* and *Gli-Z4*) revealed some InDels and a considerable number of SNPs among them. Most of the pseudogenes were resulted from C to T change, leading to the generation of TAG or TAA in-frame stop codon. The putative proteins of both *Gli-At3* and *Gli-Z7* genes contained an extra cysteine residue in the unique domain II. Analysis of toxic epitodes among 19 deduced α-gliadins demonstrated that 14 of these contained 1–5 T cell stimulatory toxic epitopes while the other 5 did not contain any toxic epitopes. The glutamine residues in two specific ploygluta mine domains ranged from 7 to 27, indicating a high variation in length. According to the numbers of 4 T cell stimulatory toxic epitopes and glutamine residues in the two ploygluta mine domains among the 19 α-gliadin genes, 2 were assigned to chromosome 6A, 5 to chromosome 6B and 12 to chromosome 6D. These results were consistent with those from wheat cv. Chinese Spring nulli-tetrasomic and phylogenetic analysis. Secondary structure prediction showed that all α-gliadins had high content of β-strands and most of the α-helixes and β-strands were present in two unique domains. Phylogenetic analysis demonstrated that α-gliadin genes had a high homology with γ-gliadin, B-hordein, and LMW-GS genes and they diverged at approximate 39 MYA. Finally, the five α-gliadin genes were successfully expressed in *E. coli,* and their expression amount reached to the maximum after 4 h induced by IPTG, indicating that the α-gliadin genes can express in a high level under the control of T_7 promoter.

Communicated by B. Friebe.

Z. Xie, C. Wang, and K. Wang contributed equally to this work.

Z. Xie · C. Wang · K. Wang · S. Wang · X. Li · Z. Zhang · Y. Yan (✉)
Key Laboratory of Genetics and Biotechnology, College of Life Science, Capital Normal University, Beijing 100048, China
e-mail: yanym@hotmail.com

C. Wang
Faculty of Life Science, Langfang Teachers College, Langfang 065000, Hebei, China

W. Ma
Western Australian Department of Agriculture and Food, State Agriculture Biotechnology Centre, Murdoch University, Perth, WA 6150, Australia

Introduction

Wheat is one of the most important crops in the world and is primary used as human staple food. Its flour, which mainly consists of the endosperm part of the wheat grain, can be used to make a variety of food products, such as various breads, noodles, pasta, etc. It contains about 80% of the total protein of the grains, most of these being storage proteins that are classically divided into glutenins and gliadins according to their polymerization properties (Shewry and Halford 2002). Polymeric glutenins consist of the high-molecular-weight glutenin subunits (HMW-GS) and low-molecular-weight glutenin subunits (LMW-GS) while the monomeric gliadins are classified into three groups: α/β-, γ-,and ω-gliadins based on their electrophoretic mobilities

during polyacrylamide gel electrophoresis at low pH (Shewry et al. 2003; Wieser 2007).

The quality of wheat flour is basically determined by the complex storage protein components presented in the wheat endosperm (Shewry and Halford 2002; Ma et al. 2005; Gao et al. 2007). The polymeric glutenins are held together by intermolecular disulphide bonds and form complex gluten polymers. It is known that the visco-elasticity and extensibility properties of gluten dough are mainly determined by glutenins and gliadins, respectively (Ma et al. 2005). Additionally, the proportions of gliadins and glutenin and the size distribution of the glutenin polymers also have important effects on gluten quality (Wieser 2007). The α-gliadins with average molecular weight of 31 kDa, containing 15–30% of the whole proteins of the wheat grain, are most abundant among glutenins and gliadins. Therefore, α-gliadins are the most consumed storage proteins in human life (Chen et al. 2008; Gu et al. 2004; Van Herpen et al. 2006).

It is well known that all storage proteins are encoded by multiple gene families and possess extensive allelic variations (Huo et al. 2007; Bai et al. 2009; Anderson et al. 2009). Previous studies showed that a large number of glutenin and gliadin genes are located in few major chromosomal regions, suggesting that these genes are physically closely linked or clustered (Gao et al. 2007). The gliadin loci are mainly located on the short arms of chromosome 1 (*Gli-1*) and 6 (*Gli-2*) homologous groups (Shewry et al. 2003; Chen et al. 2008). In particular, α-gliadin genes are encoded by the *Gli-2* loci on the chromosome 6A, 6B, and 6D (Anderson et al. 1984). Moreover, α-gliadins appear to be the most abundant multigene families present in each of the three loci in hexaploid common wheats, and the estimated gene copies range from 25–35 (Harberd et al. 1985) to 100 (Okita et al. 1985) or even up to 150 (Anderson et al. 1997) in individual haploid genome. These differences have probably contributed to the duplication and deletion of chromosome segments (D'Ovidio et al. 1991). In addition, Anderson and Greene (1997) speculated that about half α-gliadin genes were inactive or pseudogenes.

The α-gliadins were shown to be the most active protein in triggering celiac disease (CD) owing to their containing several peptides referred to as toxic epitopes that constitute the main toxic components in CD (Arentz-Hansen et al. 2000; Vader et al. 2003; Van Herpen et al. 2006; Vaccino et al. 2009). The toxic epitopes of α-gliadins showed to have genetic diversity, and the quantity and distribution of toxic epitopes can be used to assign α-gliadins to specific chromosome loci (Vader et al. 2003; Van Herpen et al. 2006). According to Van Herpen et al. (2006), the α-gliadin genes derived from chromosome 6A of bread wheat almost invariably contain epitopes glia-α9 (PFPQPQLPY) and glia-α20 (FRPQQPYPQ), but never the intact epitopes glia-α (QGSFQPSQQ) and glia-α2 (PQPQLPYPQ). A number of α-gliadin genes from chromosome 6B do not contain any of the 4 T cell epitopes whereas the α-gliadin genes from chromosome 6D are found to contain all of these T cell epitopes in variable combinations per gene. The A genome codes for a significantly larger average number of glutamine residues in polyglutamine domain I (QI) than the B and D genomes. In polyglutamine domain II (QII), the B genome showed a significantly larger number of glutamine residues than those of the other two genomes.

Heterologous expression, such as in *E. coli*, yeasts and cultured insect cells, is a detection method of gene activity in protein level. In the past years, some HMW-GS and LMW-GS genes have been successfully expressed in *E. coli* and transferred to different wheat cultivars (Galili 1989; Barro et al. 1997; Li et al. 2007, 2008a, b; Zhang et al. 2008; Yan et al. 2009). To date, however, expression of gliadin genes in *E. coli* is seldom reported (Tamás and Shewry 2006). The high-level expression in vitro of gliadin genes is still difficult, and the genetic and biochemical mechanisms remain unknown.

It is generally accepted that *Aegilops tauschii* ($2n = 2x = 14$, DD) is the D genome donor of hexaploid wheat, which is presumed to have arisen from interspecific hybridization between *T. dicoccum* (AABB) and *Ae. tauschii* coupled with subsequent chromosome doubling in southwestern Asia about 8,000–12,000 years ago (Dvorak et al. 1998; Zhang et al. 2008). It has been reported that *Ae. tauschii* accessions contain extensive variations in storage protein compositions, which may provide potential new gene sources for wheat quality improvement (Yan et al. 2003). In order to further explore the gliadin dynamics in nature for wheat improvement, in this work, we have focused on cloning novel α-gliadin full-ORFs from *Ae. tauschii.* A thorough characterization of these cloned genes has been carried.

Materials and methods

Plant materials

Three accessions of *Ae. tauschii*, T15, T26, and T43 were kindly provided by GenBank (Braunschweig, Germany). Two bread wheat cultivars, Zhongyou 9507 and Gaocheng 8901 (*Triticum aestivum* L.), were collected from the Department of Plant Genetics and Breeding, China Agricultural University. The group 6 nulli-tetrasomic lines of Chinese Spring were provided by the Institute of Crop Science, Chinese Academy of Agricultural Science (CAAS).

mRNA extraction and RT–PCR

The mRNA from seed endosperm of 15 days after flowering was isolated by using RNA plant Reagent (Tiangen, Beijing, China). cDNA synthesis was carried out with oligo (dT) from about 100 ng mRNA by using Superscript first-strand synthesis system (Invitrogen, Beijing, China).

Coding sequence of α-gliadin genes were amplified by primers designed according to the previously reported sequences: L1 (5′-ATGAAGACCTTTCTCATCCTTG-3′) and R1 (5′-TCAGTTRGTACCRAAGATGCC-3′). Primers were synthesized by Sangon Biotech (Shanghai, China). RT–PCR amplifications were performed in 50 μl reaction volume containing 2.5 U La Taq polymerase (TaKaRa), 60 ng of template DNA, 25 μl 2 × GC buffer II ($MgCl_2$ plus), 0.4 mM dNTP, 0.5 μM of each primer, and making up to 50 μl with ddH_2O. The reaction was carried out in a PTC-100 (MJ Research) using the following protocol: heating lid, pre-denaturation at 94°C for 4 min, cycled 35 times at 94°C for 45 s, 60°C for 1 min and 72°C for 45 s, and a final extension at 72°C for 10 min.

Molecular cloning and sequencing

PCR products were separated by 1.0% agarose gels and the expected fragments were purified from the gels by Gel Extraction Kit (Omega). Subsequently, the purified products were ligated into pGEM-T Easy vector (Promega, Beijing, China) and transformed into competent cells of *E. coli* DH-5α strain. DNA sequencing from three clones was performed by TaKaRa Biotech (Dalian, China).

Sequence comparison, SNPs and InDels identification, and secondary structure prediction

Sequence comparison and identification of single-nucleotide polymorphisms (SNPs) and insertions/deletions (InDels) presented in α-gliadin genes were based on multiple alignments by software Bioedit 7.0. Prediction of secondary structure of deduced amino acid sequences was carried out by PSIPRED server (http://bioinf.cs.ucl.ac.uk/psipred/psiform.html) according to Li et al. (2007).

Identification of toxic epitopes and chromosomal location

The toxic epitopes in the gliadins were identified according to Cornell and Wills-Johnson (2001). Chromosomal location of cloned α-gliadin genes was assigned based on the method of Van Herpen et al. (2006). The results of chromosomal location were further confirmed by using group 6 nulli-tetrasomic lines of Chinese Spring and phylogenetic analysis.

Phylogenetic analysis

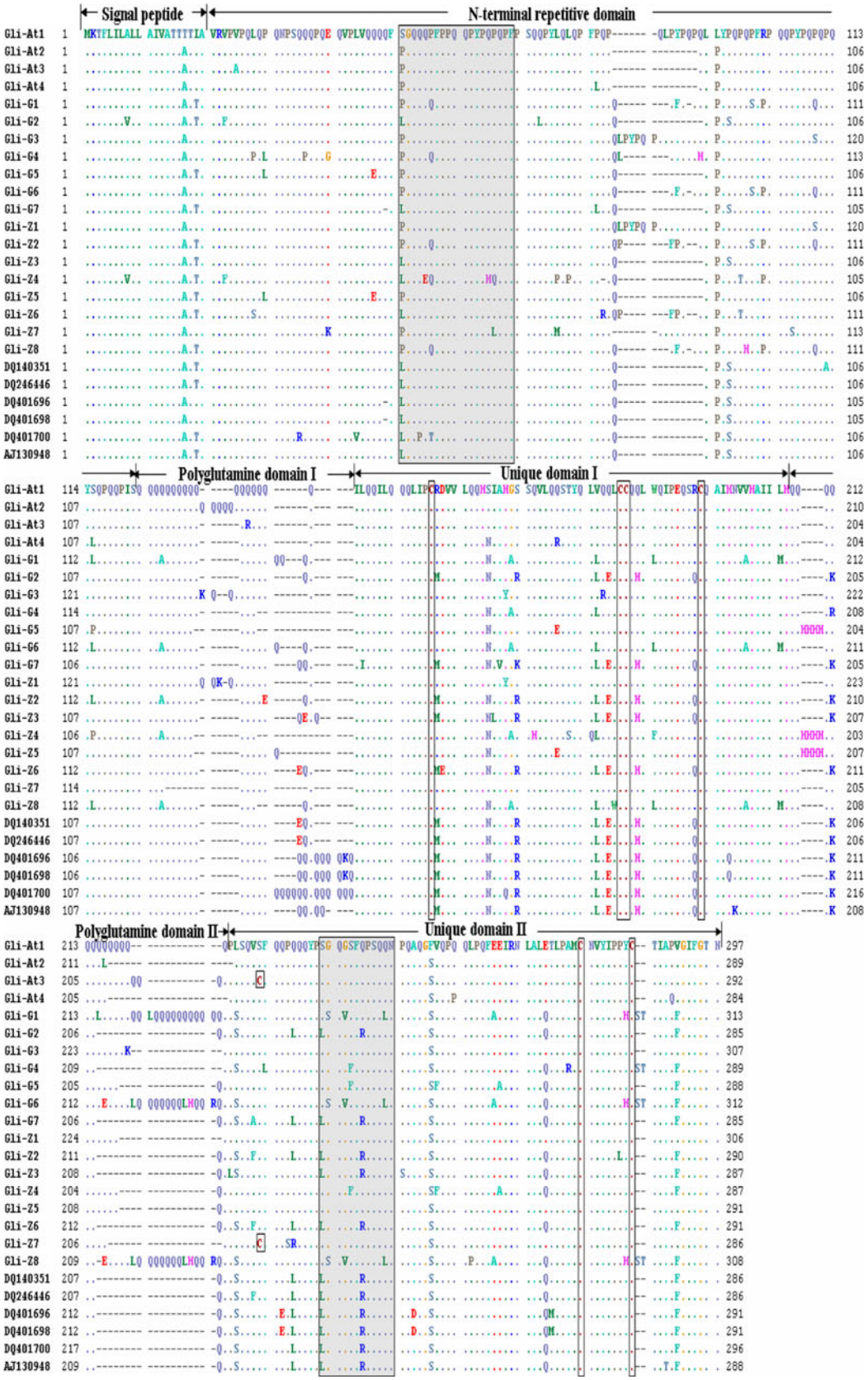
Phylogenetic tree was constructed for the deduced amino acid sequences using software DNAMAN 5.2.2. According to the full-ORF of prolamin genes and other genes from related species, a neighbor-joining tree was constructed by setting the evolutional rate parameter to 6.5×10^{-9} (Allaby et al. 1999) in calculating the divergence times of storage protein genes as well as different genomes by using software MEGA 4 (Tamura et al. 2007).

Results

Molecular characterization of 19 novel α-gliadin genes

A total of 19 full-ORF genes were isolated through RT–PCR amplification, cloning and sequencing. These include 4 from *Aegilops tauschii* accessions T15 (*Gli-At1*), T43 (*Gli-At2*), T26 (*Gli-At3* and *Gli-At4*), 8 from cv. Zhongyou 9507 (*Gli-Z1*, *Gli-Z2*, *Gli-Z3*, *Gli-Z4*, *Gli-Z5*, *Gli-Z6*, *Gli-Z7* and *Gli-Z8*), and 7 from cv. Gaocheng 8901 (*Gli-G1*, *Gli-G2*, *Gli-G3*, *Gli-G4*, *Gli-G5*, *Gli-G6* and *Gli-G7*). In addition to the 19 full-ORF genes, 32 genes from both *Ae. tauschii* and bread wheat cultivars were obtained containning at least one in-frame stop codon. Most of these pseudogenes resulted from C to T substitution, leading to the generation of a TAA or TAG stop codon. Sequence comparison demonstrated that all cloned genes possessed typical structural features of previously characterized α-gliadin genes, i.e., no introns and ending at a stop codon TGA. The multiple alignment of deduced amino acid sequence (Fig. 1) showed that the 19 full-ORF α-gliadin genes had highly similar structures with the 6 typical α-gliadin genes in the GenBank, which were isolated from common wheat and relative species, including accession numbers DQ140351 (*Triticum turgidum* subsp. *dicoccoides*), DQ246446 (*Triticum aestivum*), DQ401696 (*Triticum urartu*), DQ401698 (*Triticum monococcum*), DQ401700 (*Triticum monococcum* subsp. *aegilopoides*) and AJ130948 (*Triticum spelta*). The deduced protein sequences had a common structure model: a short signal peptide with 20 amino acid residues followed by 5 distinct domains: a repetitive domain containing N-terminal with 5 amino acid residues in the beginning, a polyglutamine domain I, a unique domain I, a polyglutamine domain II and a unique domain II containing C-terminal in the end. Apparently, the sizes of α-gliadin genes mainly depend on the repetitive domains and polyglutamine domains. All 19 genes were deposited in GenBank (accession number EF561270 to EF561288).

Comparative analysis of 25 deduced α-gliadin subunits demonstrated that, in the N-terminal repetitive domains, Gli-G3 and Gli-Z1 had an extra LPYPQPQ insertion


Signal peptide
N-terminal repetitive domain
Polyglutamine domain I
Unique domain I
Polyglutamine domain II
Unique domain II

◀ **Fig. 1** Mutliple alignment of the deduced amino acid sequences of 25 alpha-gliadin genes. *Black frames* show the position of cysteine residues. *Dashes* represent the deletions. The *gray segments* of the aligned sequences show the position of the peptides, which had activity in celiac disease (in the N-terminal domain) and were associated with adenovirus type 12 infections (in the C-terminal domain)

compared to Gli-At1, Gli-G4, and Gli-Z7. In the unique domains, 6 highly conserved cysteine residues were presented in all subunits; however, both Gli-At3 and Gli-Z7 contained an extra cysteine residue instead of serine in the unique domain II.

Three typical α-gliadin genes, *Gli-At4*, *Gli-G1,* and *Gli-Z4* were used to detect SNPs and InDels variations by comparison with 9 other α-gliadin genes reported previously (Table 1). A total of 8, 39 and 33 SNPs were detected in *Gli-At4*, *Gli-G1,* and *Gli-Z4*, respectively. An insertion (TCGACC) at the position 937–942 was presented in *Gli-G1* while a deletion (TTGCAACAACAACAACA GCAGCAACTGCAACAA) at position 676–708 and an insertion (TCGACC) at the position 937–942 were identified in *Gli-Z4*. In addition, there were 6, 19, and 18 non-synonymous substitutions in *Gli-At4*, *Gli-G1,* and *Gli-Z4*, respectively.

Identification of specific toxic epitopes related to some diseases and chromosomal location of cloned α-gliadin genes

Toxicity epitopes of celiac disease were showed to link to the presence of PSQQ and QQQP motifs in the N-terminal repetitive domains (Cornell and Wills-Johnson 2001). As shown in Fig. 1, the N-terminal repetitive domains of Gli-G2, Gli-G7, Gli-Z3, Gli-Z4, and Gli-Z6 subunits contained a 19-residue motif (LGQQQPFPPQQPYPQPQ PF) that was considered to have activity in celiac disease (Wieser 2001). In addition, another short motif (LGQG SFRPSQQN) was presented in the C-terminal domain of Gli-G2, Gli-G7, Gli-Z2, Gli-Z3, and Gli-Z6 subunits, which could be associated with adenovirus type 12 infections (Kasarda and D'Ovidio 1999).

The number of 4 T cell stimulatory toxic epitopes and glutamine residues in the two ploygltamine domains in the 19 α-gliadins were identified and listed in Table 2. Considerable variations in toxic epitope compositions existed in the 19 α-gliadins. Four previously reported celiac disease toxic epitopes, viz. glia-α2 (PQPQLPYPQ), glia-α9 (PFPQPQLPY), glia-α20 (FRPQQPYPQ), and glia-α (QGSFQPSQQ. Cornell and Wills-Johnson 2001), were

Table 1 The positions of SNPs and InDels identified between *Gli-At4*, *Gli-G1*, *Gli-Z4,* and other nine α-gliadin genes

α-gliadin genes	23	63	67	144	163	167	195	227–244	245	257	259	266	292	298	305
Gli-At4	C	**C**	G	**G**	C	C	G	----------	**T**	T	C	C	C	C	G
Gli-G1	C	T	G	**G**	C	**A**	G	----------	C	**C**	**T**	**T**	C	**T**	**C**
Gli-Z4	**T**	T	**T**	A	**G**	**A**	**T**	----------	C	**C**	**T**	**T**	**A**	C	**C**
Other 9 α-gliadin genes	C	T	G	A	C	C	G	TGCAG/ACTGCAACCATTTC	C	T	C	C	C	C	G

α-gliadin genes	329	343	344	345	354	372	375	376	377	378	466	501	527	537	540	548	554	561	563
Gli-At4	C	T	**T**	G	**G**	G	G	C	A	G	T	A	G	A	T	**G**	C	G	T
Gli-G1	**A**	**C**	**T**	**A**	A	**A**	**A**	**G**	**C**	**A**	T	**C**	**C**	A	T	A	C	G	T
Gli-Z4	C	**C**	C	G	A	**A**	**A**	**G**	**C**	**A**	**C**	**C**	**C**	**T**	**A**	A	**G**	**A**	**A**
Other 9 α-gliadin genes	C	T	C	G	A	G	G	C	A	G	T	A	G	A	T	A	C	G	T

α-gliadin genes	564	576	585	593	594	613	638	652	666	669	672	675	676–708
Gli-At4	G	G	G	G	G	C	T	C	A	A	A	A	----------
Gli-G1	**A**	G	**A**	**T**	G	**A**	**C**	**A**	**C**	**C**	**C**	**C**	----------
Gli-Z4	G	**A**	G	**T**	**T**	C	T	C	A	A	A	A	**TTGCAACAACAACAACAGCAGCAACTGCAACAA**
Other 9 α-gliadin genes	G	G	G	G	G	C	T	C	A	A	A	A	----------

α-gliadin genes	747	753	777	790	797	800	807	818	828	831	840	841	846	851	869	872	918	931	937–942	953
Gli-At4	G	G	A	G	G	C	G	A	G	C	T	G	G	**C**	A	A	C	T	------	**A**
Gli-G1	**A**	G	**G**	**A**	**T**	C	G	**T**	**A**	**T**	**C**	G	**A**	A	**C**	A	C	**C**	**TCGACC**	C
Gli-Z4	G	**A**	A	G	G	**T**	**A**	A	G	C	T	**T**	G	A	A	**C**	**G**	T	------	C
Other 9 α-gliadin genes	G	G	A	G	G	C	G	A	G	C	T	G	G	A	A	A	C	T	------	C

Horizontal dashes indicate the deletions of nucleotide. Other nine α-alpha-gliadin genes include DQ140351 (Qi et al. 2006), DQ245447 (Jia et al. 2006), DQ246446, DQ246447, DQ246448, DQ401696, DQ401698, DQ401700, and AJ130948 (Kasarda and D'Ovidio 1999)

Table 2 Number of 4 T cell stimulatory toxic epitopes and glutamine residues in the two polyglutamine domains presented in 19 full-ORF α-gliadin genes from *Ae. tauschii* and bread wheat cultivars and their chromosomal locations

α-gliadin genes	Gliα-α	Gliα-α2	Gliα-α9	Gliα-α20	QI	QII	Chromosomal location
Gli-At1	1	1	1	1	17	13	6D
Gli-At2	1	1	1	1	22	7	6D
Gli-At3	–	1	1	1	16	16	6D
Gli-At4	–	–	–	1	16	8	6D
Gli-G1	–	–	–	–	19	27	6B
Gli-G2	–	–	1	1	17	8	6A
Gli-G3	1	2	1	1	20	13	6D
Gli-G4	–	1	1	1	13	7	6D
Gli-G5	–	1	1	1	12	16	6D
Gli-G6	–	–	–	–	18	27	6B
Gli-G7	–	–	–	1	18	8	6D
Gli-Z1	1	2	1	1	21	11	6D
Gli-Z2	–	–	–	–	17	8	6B
Gli-Z3	–	–	1	1	19	8	6A
Gli-Z4	–	–	–	–	12	16	6B
Gli-Z5	1	1	1	1	15	16	6D
Gli-Z6	–	–	–	1	18	8	6D
Gli-Z7	1	1	1	–	10	9	6D
Gli-Z8	–	–	–	–	15	27	6B

found in most α-gliadins. In total, 14 α-gliadins contained 1–5 toxic epitopes while 5 (Gli-G1, Gli-G6, Gli-Z2, Gli-Z4 and Gli-Z8) had no toxic epitopes.

It is clear that the 4 α-gliadin genes from *Ae. tauschii* were located on the *Gli-D2* locus of D genome. The number of the 4 T cell stimulatory toxic epitopes and glutamine residues in the two ploygluatamine domains (Table 2) were used to determine the chromosomal locations of the 15 α-gliadin genes isolated from the 2 bread wheat cultivars. The deduced protein sequences of both *Gli-G2* and *Gli-Z3* genes contained glia-α9 and glia-α20, the number of glutamine residues in polyglutamine domain I was larger, and there was no presence of glia-α or glia-α2. Based on these facts, they could be assigned to the chromosome 6A. Similarly, 5 α-gliadin genes (*Gli-G1*, *Gli-G6*, *Gli-Z2*, *Gli-Z4,* and *Gli-Z8*) were assigned to chromosome 6B because their amino acid sequences do not contain any 4 T cell epitopes and had a significantly larger number of glutamine residues in the polyglutamine domain II. The remaining 8 genes (*Gli-G3*, *Gli-G4*, *Gli-G5*, *Gli-G7*, *Gli-Z1*, *Gli-Z5*, *Gli-Z6,* and *Gli-Z7*) contained at least one T cell epitopes, making them to be assigned to chromosome 6D. In order to test the reliability of these assigning strategies, the 4 α-gliadin genes (*Gli-At1*, *Gli-At2*, *Gli-At3* and *Gli-At4*) from *Ae. tauschii* were also applied with these assigning methods and the results showed that they were all assigned to the *Gli-D2* locus on chromosome 6D. Furthermore, analysis of nulli-tetrasomic lines of Chinese Spring as well as phylogenetic tree (see below) also confirmed the validity of chromosomal locations described above (Fig. 2).

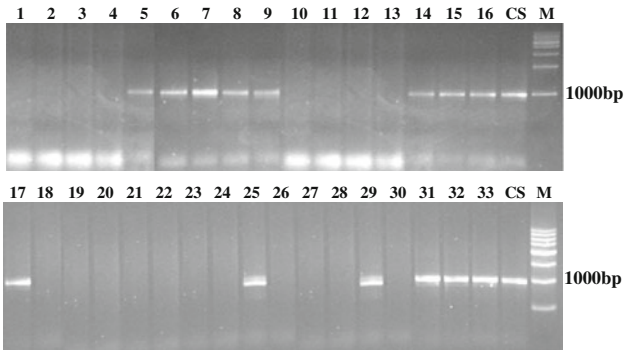


Fig. 2 Chromosomal location of 19 α-gliadin genes by using group 6 nulli-tetrasomic lines of Chinese Spring (CS). *1* N6AT6B (*Gli-G2*). *2* N6AT6D (*Gli-Z3*). *3* N6BT6A (*Gli-G1*). *4* N6BT6D(*Gli-G6*). *5* T15. *6* T26. *7* T43. *8* Zhongyou 9507. *9* Gaocheng 8901. *10* N6BT6A (*Gli-Z2*). *11* N6BT6D (*Gli-Z4*). *12* N6BT6D (*Gli-Z8*). *13* N6DT6A (*Gli-At1*). *14* N6BT6A (*Gli-G2*). *15* N6DT6A (*Gli-Z3*). *16* N6DT6B (*Gli-G6*). *17* N6AT6D (*Gli-At1*). *18* N6DT6B (*Gli-At2*). *19* N6DT6A (*Gli-At3*). *20* N6DT6B (*Gli-At4*). *21* N6DT6B (*Gli-G3*). *22* N6DT6A (*Gli-G4*). *23* N6DT6B (*Gli-G5*). *24* N6DT6A (*Gli-G7*). *25* N6BT6D (*Gli-At2*). *26* N6DT6B (*Gli-Z1*). *27* N6DT6A (*Gli-Z5*). *28* N6DT6B (*Gli-Z6*). *29* N6AT6D (*Gli-G3*). *30* N6DT6A (*Gli-Z7*). *31* N6AT6D (*Gli-G4*). *32* N6BT6D (*Gli-G5*). *33* N6AT6D (*Gli-Z1*)

Table 3 Secondary construct predictions of 19 deduced α-gliadins

α-gliadins	Secondary structure	Account (%)	Number	Distribution				
				A	B	C	D	E
Gli-At1	α-helix	19.5	6	0	0	4	0	2
	β-strand	5.8	5	0	0	0	1	4
Gli-At2	α-helix	23.4	6	0	0	3.5	0.5	2
	β-strand	5.6	5	0	0	0	0	5
Gli-At3	α-helix	24.3	6	0	0	4	0	2
	β-strand	5.9	5	0	0	0	0	5
Gli-At4	α-helix	23.5	5	0	0	3.5	0.5	1
	β-strand	5.7	5	0	0	0	0	5
Gli-G1	α-helix	12.3	5	0	0	2.5	0.5	2
	β-strand	7.9	7	0	0	0	1	6
Gli-G2	α-helix	24.9	6	0	0	4	0	2
	β-strand	7.6	5	0	0	1	0.5	3.5
Gli-G3	α-helix	23.0	9	0	0	4	0	5
	β-strand	6.3	2	0	0	0	0	2
Gli-G4	α-helix	24.9	5	0	0	3.5	0.5	1
	β-strand	10.0	5	0	0	0	0	5
Gli-G5	α-helix	24.6	6	0	0	4	0	2
	β-strand	5.2	5	0	0	0	0	5
Gli-G6	α-helix	17.1	5	0	0	3	0	2
	β-strand	7.5	6	0	0	0	0	6
Gli-G7	α-helix	23.4	5	0	0	2.5	0.5	2
	β-strand	5.7	4	0	0	0	0	4
Gli-Z1	α-helix	29.4	6	0	0	4	0	2
	β-strand	4.5	5	0	0	0	0	5
Gli-Z2	α-helix	25.9	7	1	0	3.5	0.5	2
	β-strand	7.0	6	0	0	0	0	6
Gli-Z3	α-helix	25.8	5	0	0	2.5	0.5	2
	β-strand	5.6	5	0	0	0	0	5
Gli-Z4	α-helix	24.0	6	0	0	4	0	2
	β-strand	5.6	5	0	0	0	0	5
Gli-Z5	α-helix	24.0	6	0	0	4	0	2
	β-strand	5.5	5	0	0	0	0	5
Gli-Z6	α-helix	21.4	5	0	0	2.5	0.5	2
	β-strand	8.5	4	0	0	0	0	4
Gli-Z7	α-helix	16.5	5	0	0	2.5	0.5	2
	β-strand	3.4	3	0	0	0	0	3
Gli-Z8	α-helix	18.7	6	0	0	4	0	2
	β-strand	9.7	8	0	0	1	0	7

A N-terminal domain

B Polyglutamine domain I

C Unique domain I

D Polyglutamine domain II

E Unique domain II

Secondary structure prediction of deduced α-gliadins

The secondary structure of the mature protein subunits of the 19 deduced α-gliadins were predicted by the PSIPRED server (Table 3). In general, the positions of α-helix and β-strand of the α-gliadins were relatively conservative, which were mainly presented in the unique domains and C-terminal domain. All α-gliadins showed to have high content of β-strands. Particularly, more α-helixes and β-strands occurred in unique domain II. Of the 19 α-gliadins, only one α-helix of Gli-Z2 was presented in the N-terminal domain and no secondary structures were found in the polyglutamine domain I. The numbers of α-helix and β-strand in the 19 α-gliadins ranged from 5 to 9 (17.1–29.4%) and 2 to 8 (6.3–9.7%), respectively. In the polyglutamine domain II, 9 α-gliadin subunits (Gli-At2, Gli-At4, Gli-G1, Gli-G4, Gli-G7, Gli-Z2, Gli-Z3, Gli-Z6, and Gli-Z7) had half α-helixes, while only Gli-At1 and Gli-G2 contained one and half β-strands, respectively. It is obvious that the numbers and distributions of the α-helix and β-strand are not evenly distributed in the different mature α-gliadins.

Phylogenetic analysis among α-gliadin genes and other cereal storage protein gene family

A homology tree was constructed for the deduced amino acid sequences of the 19 full-ORF α-gliadin genes obtained in this work and 18 GenBank, prolamin genes from common wheat, related species and barley, viz. 12 α-gliadin genes: AJ130948 (Kasarda and D'Ovidio 1999); AJ133605, AJ133609, and AJ133610 (Arentz-Hansen et al. 2000); DQ401696, DQ234066, DQ234067, DQ246447, DQ246448, K03074, U51303, and X02539 (Sumner-Smith et al. 1985); 2 ω-gliadin genes: AF280605 and AY667097; a γ-gliadin gene M11077 (Okita et al. 1985); and 3 barley hordein genes: B-hordein AY695368, C-hordein X60037, and D-hordein D82941.

As shown in Fig. 3, 37 prolamin genes were clustered into four subgroups, corresponding to D-hordeins, ω-gliadins and C-hordeins, γ-gliadins and B-hordeins, and α-gliadins. The α-gliadins were closely related to γ-gliadins and B-hordeins with 53% homology. The ω-gliadins and C-hordeins had 76% homology while γ-gliadins and B-hordeins shared 79% homology. The D-hordeins was isolated from other prolamins and shared only 36% homology.

In order to further investigate the evolutionary relationships among different storage protein genes, a neighbor-joining tree was constructed with the deduced amino acid sequences of 7 typical α-gliadin genes obtained in this work and 16 previously characterized storage protein genes (Fig. 4). An estimation of divergence time (million years ago, MYA) among the 23 genes was shown in Table 4. As shown in Fig. 4, the 23 genes from common wheat and related species were apparently clustered into 5 clades, viz. HMW-GS and D-hordein genes, ω-gliadin and C-hordein genes, γ-gliadin genes, B-hordein and LMW-GS genes, and α-gliadin genes, which were generally consistent with the results from phylogenetic tree described above. The HMW-

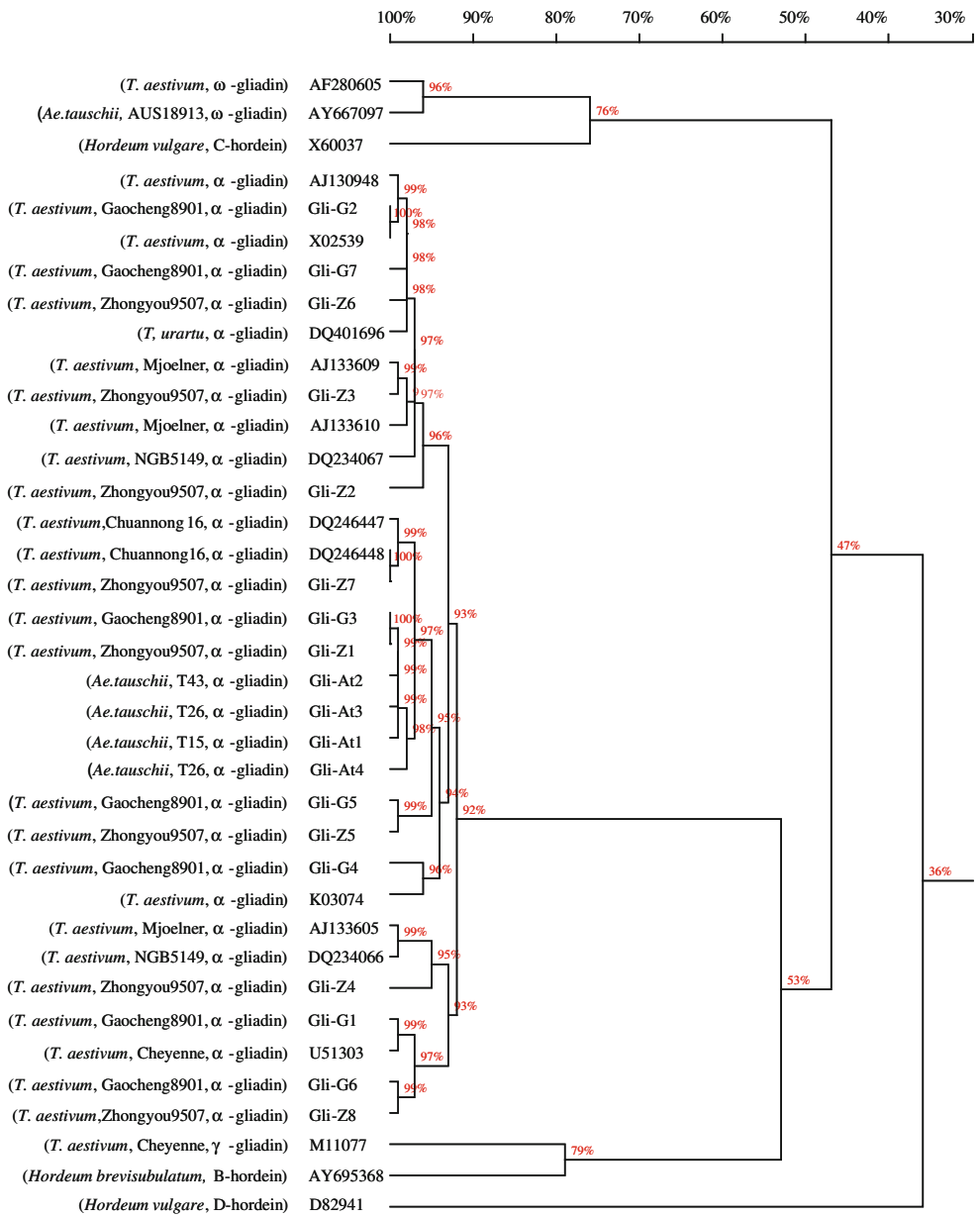


Fig. 3 Homology tree based on the coding regions of deduced amino acid sequence of 19 cloned alpha-gliadin genes and 18 other prolamin genes from related species. The suffixes of GenBank accession numbers indicate the different types of genes

GS and D-hordein genes were divergent at 17.00 MYA, and their divergence with other storage genes occurred about 69 MYA. The α-gliadin genes, classified into an independent clade indicated by broken line ellipse, diverged with the near clade (γ-gliadin, B-hordein and LMW-GS genes) at approximately 39 MYA. In addition, relatively late divergence between ω-gliadin and C-hordein, LMW-GS and B-hordein genes occurred about 18 and 25 MYA, respectively. The divergent times estimated between individual storage protein genes (Table 4) displayed similar results, and the α-gliadin genes from the *Gli-D2* locus of *Ae. tauschii* and bread wheat cultivars diverged more recently, generally less than 5 MYA.

Discussion

Allelic variations, molecular structures and functions of the α-gliadin genes

The gliadin storage protein genes were shown to have extensive allelic variations, which were closely related to

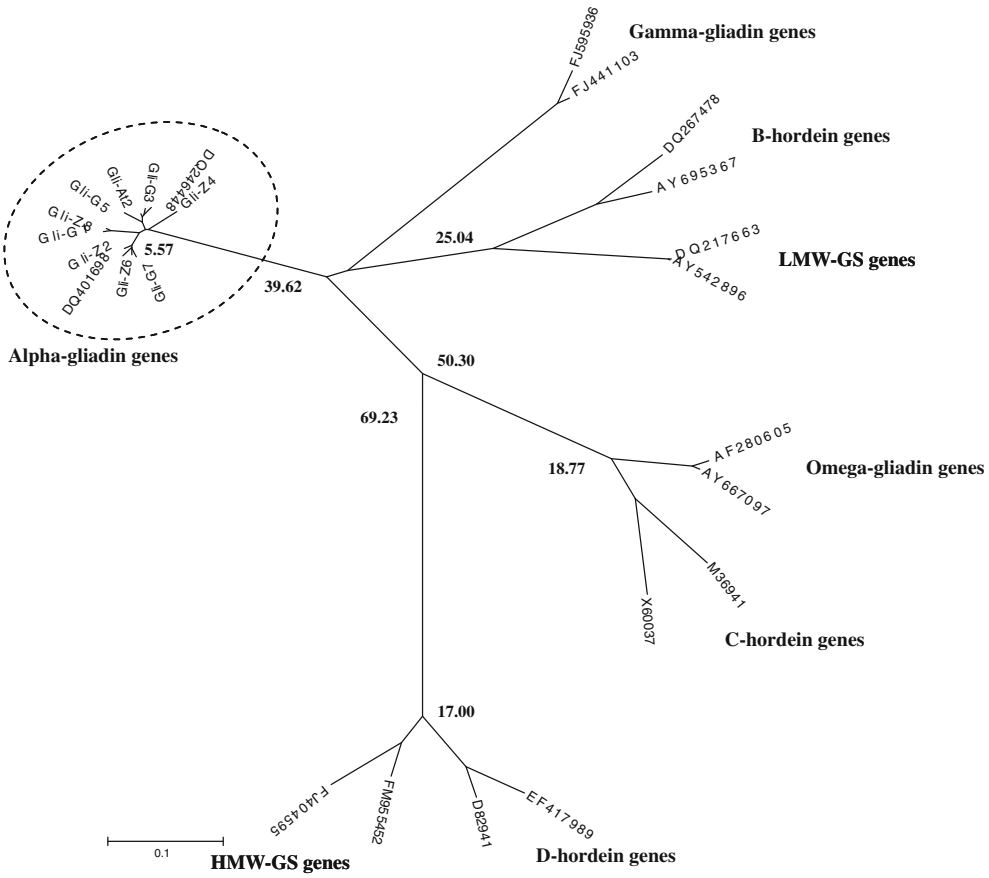


Fig. 4 Neighbor-joining tree based on the deduced amio acid sequence of 9 cloned α-gliadin genes (*Gli-At2*, *Gli-G1*, *Gli-G3*, *Gli-G5*, *Gli-G7, Gli-Z2*, *Gli-Z4*, *Gli-Z6* and *Gli-Z8*) and 16 other related genes (α-glidin DQ246448, DQ401698, ω-gliadin AF280605, AY667097, γ-gliadin FJ441103, FJ595936, B-hordein DQ267478, AY659367, C-hordein M36941, X60037, D-hordein EF417989, D82941, LMW-GS DQ217663, AY542896 and HMW-GS FM955452, FJ404595. The divergent times were specified by the numbers, and the unit was million years ago (MYA)

flour quality (Metakovsky et al. 1997; Wang et al. 2008). A total of 111 gliadin alleles, including 19 α-gliadins among 360 common wheat cultivars have been identified in the past (Metakovsky 1991). The estimated number of α-gliadin genes ranges from 25 to 150 copies in different wheat genotypes. However, approximately 50% (Anderson and Greene 1997), even 87% (Van Herpen et al. 2006) of the α-gliadin genes are pseudogenes, mainly resulting from C to T change. In the current work, more than 60% of separated α-gliadin genes were pseudogenes and C to T substitution accounted for about 70% of the causal. In plants, as much as 20% of the total residues can be methylated and methylation at the 5-position of cytidine is the most common modified DNA base. This could result in C to T substitution in a higher percentage (Anderson and Greene 1997).

So far, secondary structure features of α-gliadins have not been reported. In this study, we predicted the secondary structures of the 19 full-ORF α-gliadins and demonstrated that the α-helixes and β-strands of α-gliadins were mainly presented in the unique domain II, while the N-terminal domain and polyglutamine domain I had few or no secondary structures. Furthermore, the numbers and percentage of α-helix and β-strand were highly variable among the 19 α-gliadins. Our results suggested that the unique domain II in the α-gliadins could be the most important domain for the formation of gluten macropolymer in conjunction with HMW and LMW glutenins through inter-molecular disulfide bonds.

The α-helixes and β-strands, mainly presented in the C-terminal domain, are the key factors for the formation of gliadin secondary structures. More α-helixes and β-strands could be beneficial for increasing viscosity and elasticity of doughs. Therefore, the high variations in the numbers of α-helixes and β-strands among α-gliadins will result in a high variation in dough physical qualities, making it possible to achieve desirable wheat qualities through screening α-gliadins based on their secondary structural features. Additionally, a long repetitive domain and an extra cysteine residue present in Gli-At3 and Gli-Z7 could have a positive influence on flour quality (An et al. 2006; Li et al. 2008a). Two polyglutamine domains encoded by microsatellite-like sequences are characteristically presented in

Table 4 Estimation of divergence time (MYA) among 9 obtained α-gliadin genes and 16 other genes

No.	Genes	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1	*Gli-At2*	–	1.00	0.54	0.77	1.00	1.00	1.08	1.00	1.08	0.62	1.00	1.92	2.00	2.08	2.08	1.85	1.92	2.00	1.92	1.85	1.85	1.92	1.92	1.92	1.92
2	*Gli-G1*	5.92	–	1.08	1.08	1.08	0.92	1.08	1.08	0.54	1.00	1.15	1.85	1.92	2.00	2.00	1.85	1.92	2.08	2.00	1.85	1.85	1.92	1.92	1.92	1.92
3	*Gli-G3*	1.54	6.54	–	0.85	1.00	1.00	1.15	1.00	1.08	0.62	1.00	1.92	1.92	2.08	2.08	1.92	1.92	2.08	1.92	1.85	1.85	2.00	1.92	1.92	1.92
4	*Gli-G5*	3.08	6.54	3.69	–	1.08	1.08	1.00	1.00	1.08	0.85	1.08	1.92	1.92	2.08	2.08	1.92	1.92	2.00	1.92	1.85	1.85	1.92	1.92	1.92	1.92
5	*Gli-G7*	5.23	6.77	5.46	6.15	–	0.69	1.23	0.62	1.08	1.08	0.77	2.00	2.00	2.08	2.08	1.92	1.92	2.08	2.00	1.85	1.85	1.92	1.92	1.92	1.92
6	*Gli-Z2*	5.08	5.08	5.23	5.92	2.15	–	1.15	0.54	0.92	1.08	0.77	1.92	1.92	2.08	2.00	1.85	1.85	2.00	1.92	1.85	1.85	2.00	1.92	1.92	1.92
7	*Gli-Z4*	5.92	6.77	6.54	5.08	8.08	7.23	–	1.15	1.15	1.08	1.23	1.92	1.92	2.08	2.00	1.92	1.92	2.08	2.00	1.85	1.85	1.92	1.92	1.92	1.92
8	*Gli-Z6*	5.08	6.38	5.23	5.92	1.77	1.31	7.69	–	1.00	1.08	0.69	1.92	1.92	2.08	2.08	1.92	1.92	2.00	1.92	1.85	1.85	1.92	1.92	1.92	1.92
9	*Gli-Z8*	6.38	1.31	6.15	6.77	6.38	4.62	7.00	5.92	–	1.08	1.08	1.92	1.92	2.00	2.00	1.85	1.92	2.00	1.92	1.85	1.85	2.00	1.92	1.92	1.92
10	*DQ246448*	1.77	6.77	2.00	3.92	6.15	5.92	6.54	5.92	6.77	–	1.08	1.92	1.92	2.08	2.08	1.92	1.92	2.00	1.92	1.85	1.85	1.92	1.92	1.92	1.92
11	*DQ401698*	5.08	7.00	5.23	6.38	2.62	2.62	8.08	2.15	6.54	5.92	–	1.92	2.00	2.08	2.00	1.92	1.85	2.00	1.92	1.85	1.85	2.00	1.92	1.92	1.92
12	*FJ595936*	31.31	31.31	31.15	32.00	32.85	32.46	32.23	32.46	31.54	32.23	32.46	–	1.00	2.00	2.08	2.00	2.00	2.00	2.00	1.85	1.92	1.92	1.85	1.92	2.00
13	*FJ441103*	30.23	30.46	30.46	30.92	32.00	31.54	30.69	31.54	30.69	31.15	31.31	5.23	–	2.00	2.00	2.00	2.00	2.00	1.92	1.85	1.92	1.92	1.92	2.00	2.00
14	*AF280605*	35.54	34.85	34.85	35.92	36.38	35.54	34.38	36.62	34.38	35.31	36.38	39.69	38.15	–	0.69	1.85	1.85	1.69	1.69	1.92	2.00	1.92	2.00	1.92	1.92
15	*AY667097*	35.31	34.15	34.62	35.31	36.15	35.31	33.77	36.38	33.77	35.08	36.15	39.00	37.46	2.38	–	1.85	1.85	1.69	1.69	2.00	2.00	1.92	2.00	1.92	1.92
16	*DQ267478*	32.00	32.46	32.00	33.31	33.08	32.00	32.46	32.85	31.77	33.08	33.08	37.69	36.15	41.38	41.62	–	1.46	2.08	1.92	1.92	1.92	2.08	2.00	1.69	1.77
17	*AY695367*	30.92	31.54	31.15	30.92	31.54	30.46	31.31	31.31	30.92	32.23	31.54	37.00	35.31	39.92	40.77	12.08	–	2.08	2.00	1.92	1.92	2.00	2.00	1.69	1.77
18	*M36941*	38.38	38.54	38.15	38.77	39.00	38.38	37.23	39.46	38.15	38.54	39.92	40.31	39.92	18.85	18.38	42.92	42.31	–	1.62	2.00	2.00	2.08	2.08	2.00	2.00
19	*X60037*	39.23	39.23	38.77	39.69	39.00	38.77	37.00	39.23	38.77	39.69	39.23	42.54	41.85	19.08	19.92	44.08	41.38	17.31	–	1.92	2.00	2.08	2.08	1.92	1.92
20	*EF417989*	44.23	44.92	44.69	45.38	44.46	44.46	44.23	44.23	44.92	45.38	44.08	48.46	49.54	44.23	43.62	45.62	48.00	43.15	44.08	–	1.38	1.62	1.69	2.00	2.00
21	*D82941*	44.08	44.46	43.85	44.08	43.85	43.85	44.08	43.62	44.08	44.92	43.38	46.69	47.54	42.08	41.62	45.15	46.00	43.15	43.62	9.46	–	1.77	1.62	2.00	1.92
22	*FM955452*	41.85	42.08	42.31	42.31	41.62	41.23	42.31	41.62	42.08	42.77	41.62	47.54	47.31	41.00	40.77	44.23	46.23	41.23	41.85	18.00	18.15	–	1.46	2.00	2.00
23	*FJ404595*	44.46	45.15	44.92	44.92	44.08	44.08	44.69	44.46	44.69	45.38	44.23	49.08	48.23	40.08	39.92	44.69	46.46	40.08	41.00	18.62	17.54	11.15	–	2.00	2.00
24	*DQ217663*	31.54	33.54	32.23	33.08	33.08	32.69	33.54	33.08	33.31	32.46	33.54	32.85	31.54	37.23	36.85	23.69	23.00	40.31	41.23	46.46	45.62	47.15	45.62	–	0.31
25	*AY542896*	31.54	33.54	32.23	33.08	32.69	32.69	33.54	33.08	33.31	32.46	33.54	32.85	31.54	36.85	36.38	23.92	23.23	39.92	40.77	46.69	45.77	47.15	45.15	0.46	–

Below the diagonal, divergent time of genes; above the diagonal, standard error

the α-gliadins, and their amino acid residues ranged from 7 to 27 in this work (Table 2). A high proportion of glutamine residues in the repeats appeared to be helpful for improving visco-elasticity properties of doughs since intermolecular interactions of the large numbers of glutamine side chains act as good hydrogen bond donors and acceptors (Masci et al. 2000).

The α-gliadins generally contain six cysteine residues with conserved positions in two unique domains, which could form three intra-molecular disulfide bonds and result in a smaller and more compactly folded globular protein (Khatkar et al. 2002). In this study, both Gli-At3 and Gli-G7 had an extra cysteine residue in the unique domain II. Changes in the numbers and positions of cysteine residues might affect the pattern of disulfide bond formation, and strongly correlate with flour quality. An odd number of cysteine residues tended to join the disulfide cross-linked gluten matrix, with a positive effect on dough quality (Kasarda et al. 1984; Shewry et al. 2003). Thus, two specific polyglutamine domains present in the α-gliadins, especially the polyglutamine domain II with some α-helixes and β-strands, and the presence of extra cysteine residues could play important roles in the formation of gluten macropolymer, and consequently insert positive effects on wheat dough quality. A high level of genetic variation in the α-gliadin loci is expected to be valuable sources for wheat quality improvement.

Phylogenetics and evolution of the α-gliadin gene family

Recent research in the genomic organization and evolutionary mechanism of the *Gli-2* loci have revealed that abundant retrotransposons, such as *Sabrina*, *Fatima,* and *Wis*-related retroelements, are presented in the *Gli-2* regions, which are considered to participate in shaping the *Gli-2* regions during the genome evolution and to be primarily responsible for the genome expansion of wheat and other cereals (Gu et al. 2004). The α-gliadin genes showed to be not evenly distributed in the *Gli-2* region, which were separated by numerous retrotransposons and their distance varied significantly. It is suggested that the *Gli-2* regions were not stable, and underwent considerable dynamic changes during wheat evolution process, mainly resulting from the frequent gene amplification/deletion, single-base change and rapid amplification and insertion of retrotransposons (Anderson and Greene 1997; Gu et al. 2004). More recently, duplication and deletion of large fragments occurring in *Glu-D-1-1* alleles and the generation of a LMW-GS chimeric gene were found to attribute to illegitimate recombination in addition to unequal homologous recombination (Li et al. 2008b; Zhang et al. 2008). The characteristics of genome organization and gene structure of the *Gli-2* loci, especially the separated gene arrangement and the presence of repetitive sequences and microsatellite-like sequences encoding two polyglutamine domains, would facilitate the occurrence of illegitimate recombination events.

It was shown that the α-gliadin genes were divergent earlier than the separation of the various wheat genomes (Gu et al. 2004). Our results demonstrated that the α-gliadins were closely related to γ-gliadin, LMW-GS, and barley B-hordein genes, and diverged at approximately 39.62 MYA. On the contrary, a distant evolutionary relationship existed between α-gliadins and ω-gliadin, C and D-hordein, and HMW-GS genes. The α-gliadin genes from *Gli-D2* loci of *Ae. tauschii* and hexaploid common wheat were divergent more recently (Table 4) and displayed a high homology (Fig. 3), being consistent with the previous result that *Ae. tauschii* is the D genome donator of *T. aestivum* (Dvorak et al. 1998; Yan et al. 2003). The cultivated wheat (AABBDD) was considered to arise by hybridization between *T. dicoccum* (AABB) and *Ae. tauschii* (DD), and the recent origin of the AABBDD wheat is no earlier than 9,000 years ago (Allaby et al. 1999). Recent investigation demonstrated that the hybridization event occurred at least four times (Zhang et al. 2008). It could deduce that the divergent time of α-gliadin genes at *Gli-D2* loci is much earlier than these hybridizations.

The phylogenetic analysis showed that the 19 cloned full-ORF α-gliadin genes had high homology (92%) with the 12 other α-gliadin genes, particularly closely related to γ-gliadin and barley B-hordein genes. In addition, 19 α-gliadins were divided into three subgroups, generally corresponding to the A, B, and D chromosome assignment. For instance, the 10 genes assigned to chromosome 6D and the 4 to chromosome 6B were clustered into two separate subgroups. This further confirmed the results of chromosomal assignment based on the polyglutamine domains and 4 T cell stimulatory toxic epitopes compositions.

Toxic epitopes and their application in chromosomal location and wheat quality improvement

To date, two major types of allergy related to wheat gliadins have been well characterized, viz. bakers' asthma and wheat-dependent exercise-induced anaphylaxis (WDEIA), with the ω_5-gliadins being the major protein group (Tatham and Shewry 2008). It was known that gluten-specific T cell responses in the small intestine were highly related to the inflammatory response in CD. Some specific α-gliadins as well as other gluten peptides could bind to HLA-DQ2/8 and induce lamina propria CD4 T cell responses, and consequently result in damage of the small intestine mucosa (Vader et al. 2002). Recent investigations have shown that the α-gliadins from diploid *Triticum* species

form distinct groups, and the number of the 4 T cell stimulatory epitopes (glia-α, glia-α2, glia-α9 and glia-α20) appears to be genome specific (Van Herpen et al. 2006). In the present work, based on the numbers and distributions of the toxic epitopes, the 19 full ORF α-gliadin genes were successfully assigned to 6A, 6B, and 6D genomes. In particular, 8 α-gliadin genes from bread wheat cultivars and the 4 from *Ae. tauschii* were assigned to *Gli-D2* locus of the chromosome 6D. These results were further confirmed by the analysis of group 6 nulli-tetrasomic lines of Chinese Spring (Fig. 2) and phylogenetic tree (Fig. 3), strongly suggesting that this assigning method for the α-gliadin genes is reliable.

The toxic epitopes present in the α-gliadin genes appeared to be genetically diverse (Vader et al. 2003; Van Herpen et al. 2006). More presence of T cell-stimulatory epitopes (glia-α2/α9) was found in the D genome species compared to A and B genome species; a number of α-gliadin genes from S^s (*T. speltoides*) and B genome species did not contain any of the 4 T cell stimulatory epitopes (Spaenij-Dekking et al. 2005). In the present study, the number of the 4 T cell-stimulatory epitopes of α-gliadin genes from 6A, 6B, and 6D genomes was found to be 4, 0, and 41, respectively (Table 2), which is consistent with the previous findings. Additionally, large variations of glutamine residues in the two ploygltamine domains were also found to present in different α-gliadin genes (Table 2). This could be useful for reducing the gliadin toxicity in wheat breeding program, especially for the D genome. In the past years, many useful genes have been found in *Ae. tauschii*, including different resistance and tolerance genes (Cox et al. 1995) as well as glutenin genes related to quality (Yan et al. 2003; Zhang et al. 2006). These genes had been introgressed and utilized in wheat cultivar improvement programs via crosses between bread wheat or tetraploid species and *Ae. tauschii* (Hsam et al. 2001). It is expected that the SNPs and InDels could be sources of molecular markers for reducing the α-gliadin content of the D genome in transferring valuable genes from *Ae. tauschii* to bread wheat.

Acknowledgments This research was financially supported by grants from the National Natural Science Foundation of China (30830072, 30771334), the Chinese National Basic Research Program (2009CB118303), National 863 project (2006AA10Z186), and Beijing Natural Scientific foundation and the Key Developmental Project of Science and Technology of Beijing Municipal Commission of Education (KZ200910028003).

References

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

An XL, Zhang Q, Yan YM, Li QY, Zhang YZ, Wang AL, Pei YH, Tian JZ, Wang HB, Hsam SL, 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

Anderson OD, Greene FC (1997) The α-gliadin gene family. II. DNA and protein sequence variation, subfamily structure, and origins of pseudogenes. Theor Appl Genet 95:59–65

Anderson OD, Litts JC, Gautier MF, Greene FC (1984) Nucleic acid sequence and chromosome assignment of a wheat storage protein gene. Nucleic Acids Res 12:8129–8144

Anderson OD, Litts JC, Greene FC (1997) The α-gliadin gene family. I, characterization of ten new wheat α-gliadin genomic clones, evidence for limited sequence conservation of flanking DNA, and Southern analysis of the gene family. Theor Appl Genet 95:50–58

Anderson OD, Gu YQ, Kong K, Lazo GR, Wu J (2009) Structure of the omega-gliadin gene family. Funct Integ Genomics 9:397–410

Arentz-Hansen EH, McAdam SN, Molberg O, Kristiansen C, Sollid LM (2000) Production of a panel of recombinant gliadins for the characterisation of T cell reactivity in coeliac disease. Gut 46:46–51

Bai W, Cai A, Rz Z, Li A, Huo N, Li S, Gu YQ, Yabut L, Jia J, Qi Y (2009) Novel microRNAs uncovered by deep sequencing of small RNA transcriptomes in bread wheat (*Triticum aestivum* L.) and *Brachypodium distachyon* (L.) Beauv. Functional and Integrative Genomics 9:532–541

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

Chen F, Xu C, Chen M, Wang Y, Xia G (2008) A new alpha-gliadin gene family for wheat breeding: somatic introgression line II-12 derived from *Triticum aestivum* and *Agropyron elongatum*. Mol Breeding 22:675–685

Cornell HJ, Wills-Johnson G (2001) Structure-activity relationships in coeliac-toxic gliadin peptides. Amino Acids 21:243–253

Cox TS, Sears RG, Bequette RK, Martin TJ (1995) Germplasm enhancement in winter wheat × *Triticum tauschii* backcross populations. Crop Sci 35:913–919

D'Ovidio R, Lafiandra D, Tanzarella OA, Anderson OD, Greene FC (1991) Molecular characterization of bread wheat mutants lacking the entire cluster of chromosome 6A controlled gliadin components. J Cereal Sci 14:125–129

Dvorak J, Luo MC, Yang ZL, Zhang HB (1998) The structure of the *Aegilops tauschii* genepool and the evolution of hexaploid wheat. Theor Appl Genet 97:657–670

Galili G (1989) Heterologous expression of a wheat high molecular weight glutenin gene in *Escherichia coli*. Proc Natl Acad Sci USA 86:7756–7760

Gao SC, Gu YQ, Wu JJ, Coleman-Derr D, Huo NX, Crossman C, Jia JZ, Zuo Q, Ren ZL, Anderson OD, Kong XY (2007) Rapid evolution and complex structural organization in genomic regions harboring multiple prolamin genes in the polyploid wheat genome. Plant Mol Biol 65:189–203

Gu YQ, Crossman C, Kong XY, Luo MC, You FM, Coleman-Derr D, Dubcovsky J, Anderson OD (2004) Genomic organization of the complex alpha-gliadin gene loci in wheat. Theor Appl Genet 109:648–657

Harberd NP, Bartels D, Thompson RD (1985) Analysis of the gliadin multigene loci in bread wheat using nullisomic-tetrasomic lines. Mol Gen Genet 198:234–242

Hsam SLK, Kieffer R, Zeller FJ (2001) Significance of *Aegilops tauschii* glutenin genes on breadmaking properties of wheat. Cereal Chem 78:521–525

Huo N, Lazo GR, Vogel JP, You F, Ma Y, Hayden DM, Coleman-Derr D, Hill T, Dvorak J, Anderson OD, Luo M, Gu YQ (2007) Characterizing the nuclear genome structure of *Brachypodium distachyon* using BAC end sequences. Funct Integr Genomics 8:135–147

Jia J, Fu J, Zheng J, Zhou X, Huai J, Wang J, Wang M, Zhang Y, Chen X, Zhang J, Zhao J, Su Z, Lv Y, Wang G (2006) Annotation and expression profile analysis of 2073 full-length cDNAs from stress-induced maize (*Zea mays* L.) seedlings. Plant J 48:710–727

Kasarda DD, 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

Kasarda DD, Okita TW, Bernardin JE, Baecker PA, Nimmo CC, Lew EJ, Dietler MD, Greene FC (1984) Nucleic acid (cDNA) and amino acid sequences of alpha-type gliadins from wheat (*Triticum aestivum*). Proc Natl Acad Sci USA 81:4712–4716

Khatkar BS, Fido RJ, Tatham AS, Schofield JD (2002) Functional properties of wheat gliadins. II. Effects on dynamic rheological properties of wheat gluten. J Cereal Sci 35:307–313

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

Li X, Wang A, Xiao Y, Yan Y, He Z, Appels R, Ma W, Hsam SLK, Zeller FJ (2008a) 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

Li X, Ma W, Gao L, Zhang Y, Wang A, Ji K, Wang K, Appels R, Yan Y (2008b) A novel chimeric low-molecular-weight glutenin subunit gene from the wild relatives of wheat *Aegilops kotschyi* and *Ae. juvenalis*: evolution at the *Glu-3* Loci. Genetics 180:93–101

Ma W, Appels R, Bekes F, Larroque O, Morell MK, Gale KR (2005) Genetic characterisation of dough rheological properties in a wheat doubled haploid population: additive genetic effects and epistatic interactions. Theor Appl Genet 111:410–422

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

Metakovsky EV (1991) Gliadin allele identification in common wheat. II. Catalogue of gliadin alleles in common wheat. J Genet Breed 45:325–344

Metakovsky EV, Annicchiarico P, Boggini G, Pogna NE (1997) Relationship between gliadin alleles and dough strength in Italian bread wheat cultivars. J Cereal Sci 25:229–236

Okita TW, Cheesbrough V, Reeves CD (1985) Evolution and heterogeneity of the alpha-/beta-type and gamma-type gliadin DNA sequences. J Biol Chem 260:8203–8213

Qi PF, Yue YW, Long H, Wei YM, Yan ZH, Zheng YL (2006) Molecular characterization of alpha-gliadin genes from wild emmer wheat (*Triticum dicoccoides*). DNA Seq 17:415–421

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

Shewry PR, Halford NG, Lafiandra D (2003) Genetics of wheat gluten proteins. Adv Genet 49:111–184

Spaenij-Dekking L, Kooy-Winkelaar Y, van Veelen P, Drijfhout JW, Jonker H, van Soest L, Smulders MJ, Bosch D, Gilissen LJ, Koning F (2005) Natural variation in toxicity of wheat: potential for selection of nontoxic varieties for celiac disease patients. Gastroenterology 129:797–806

Sumner-Smith M, Rafalski JA, Sugiyama T, Stoll M, Soll D (1985) Conservation and variability of wheat alpha/beta-gliadin genes. Nucleic Acids Res 13:3905–3916

Tamás L, Shewry PR (2006) Heterologous expression and protein engineering of wheat gluten proteins. J Cereal Sci 43:259–274

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. Mol Biol Evol 24:1596–1599

Tatham A, Shewry P (2008) Allergens to wheat and related cereals. Clin Experim Allergy 38:1712–1726

Vaccino P, Becker H-A, Brandolini A, Salamini F, Kilian B (2009) A catalogue of *Triticum monococcum* genes encoding toxic and immunogenic peptides for celiac disease patients. Mol Gen Genomics 281:289–300

Vader W, Kooy Y, Van Veelen P, De Ru A, Harris D, Benckhuijsen W, Pena 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

Vader W, Stepniak D, Kooy Y, Mearin L, Thompson A, van Rood JJ, Spaenij L, Koning F (2003) The HLA-DQ2 gene dose effect in celiac disease is directly related to the magnitude and breadth of gluten-specific T cell responses. Proc Natl Acad Sci USA 100:12390–12395

Van Herpen T, Goryunova SV, Van der Schoot J, Mitreva M, Salentijn E, Vorst O, Schenk MF, Van Veelen PA, Koning F, Van Soest LJM, Vosman B, Bosch D, Hamer RJ, Gilissen L, Smulders MJM (2006) Alpha-gliadin genes from the A, B, and D genomes of wheat contain different sets of celiac disease epitopes. BMC Genomics 7:1–13

Wang AL, Gao LY, Li XH, Zhang YZ, He ZH, Xia XC, Zhang Y, Yan YM (2008) Characterization of two 1D-encoded omega-gliadin subunits closely related to dough strength and pan bread-making quality in common wheat (*Triticum aestivum* L.). J Cereal Sci 47:528–535

Wieser H (2001) Comparative investigations of gluten proteins from different wheat species. III. N-terminal amino acid sequences of α-gliadins potentially toxic for coeliac patients. Eur Food Res Technol 213:183–186

Wieser H (2007) Chemistry of gluten proteins. Food Microbiol 24:115–119

Yan Y, Hsam SLK, Yu J, Jiang Y, Zeller FJ (2003) Allelic variation of the HMW glutenin subunits in *Aegilops tauschii* accessions detected by sodium dodecyl sulphate (SDS-PAGE), acid polyacrylamide gel (A-PAGE) and capillary electrophoresis. Euphytica 130:377–385

Yan Y, Jiang Y, An XL, Pei YH, Li XH, Zhang YZ, Wang AL, He Z, Xia X, Bekes F, 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

Zhang Y, Li Q, Yan Y, Zheng J, An X, Xiao Y, Wang A, Pei Y, Wang H, Hsam SL, Zeller FJ (2006) Molecular characterization and phylogenetic analysis of a novel glutenin gene (*Dy10.1*^{t}) from *Aegilops tauschii*. Genome 49:735–745

Zhang Y, Li X, Wang A, An X, Zhang Q, Pei Y, Gao L, Ma W, Appels R, Yan Y (2008) Novel x-type high-molecular-weight glutenin genes from *Aegilops tauschii* and their implications on the wheat origin and evolution mechanism of Glu-D1–1 proteins. Genetics 178:23–33