This is the author’s final version of the work, as accepted for publication following peer review but without the publisher’s layout or pagination. The definitive version is available at:

http://dx.doi.org/10.1016/j.jcs.2014.01.020


http://researchrepository.murdoch.edu.au/21682/

Copyright © 2014 Elsevier Ltd.
Wheat seed storage proteins: Advances in molecular genetics, diversity and breeding applications

Awais Rasheed¹², Xianchun Xia¹, Yueming Yan³, Rudi Appels⁴, Tariq Mahmood², Zhonghu He¹⁵*

¹Institute of Crop Science, National Wheat Improvement Center, Chinese Academy of Agricultural Sciences (CAAS), 12 Zhongguancun South Street, Beijing 100081, China
²Department of Plant Sciences, Quaid-i-Azam University, Islamabad 45320, Pakistan
³College of Life Science, Capital Normal University, Beijing 100048, China
⁴Veterinary and Life Sciences, Murdoch University, Perth, WA 6150, Australia
⁵International Maize and Wheat Improvement Center (CIMMYT) China Office, c/o CAAS, 12 Zhongguancun South Street, Beijing 100081, China

*Author for correspondence: Z.H. He, Email: zhhecaas@163.com
Abbreviations

HMW-GS, high molecular weight glutenin subunit; LMW-GS, low molecular weight glutenin subunit; SDS-PAGE, sodium dodecyl sulfate poly-acrylamide gel electrophoresis; HPCE, high performance capillary electrophoresis; RP-HPLC, Reversed-phase high-performance liquid chromatography; 2-DE, two-dimensional electrophoresis; MALDI-TOF-MS, matrix-assisted laser desorption/ionization time-of-flight mass spectrometry; RP-UPLC, reversed-phase ultra performance liquid chromatography; AS-PCR, allele-specific PCR; SNPs, single nucleotide polymorphisms; NILs, Near-isogenic lines; GMP, glutenin macro polymers; CD, coeliac diseases; WDEIA, wheat-dependent exercise induced anaphylaxis; IWGSC, the international wheat genome sequencing consortium; GWAS, genome wide association studies.
Abstract

Wheat seed storage proteins, especially glutenins and gliadins, have unique functional properties giving rise to a wide array of food products for human consumption. The wheat seed storage proteins, however, are also the most common cause of food-related allergies and intolerances, and it has become crucially important to understand their composition, variation and functional properties and interface this knowledge with the grain handling industry as well as the breeders. This review focuses on advances in understanding the genetics and function of storage proteins and their application in wheat breeding programs. These include:

1. The development and validation of high-throughput molecular marker systems for defining the composition and variation of low molecular weight glutenin subunits (LMW-GS) genes and a summary of the more than 30 gene-specific markers for rapid screening in wheat breeding programs;
2. The identification of more than 100 alleles of storage proteins in wild species provide candidate genes for future quality improvement;
3. The documentation of quality effects of individual LMW-GS and HMW-GS for improving end-use quality; and
4. The analysis of α-gliadin genes on chromosomes 6A and 6D with non-toxic epitopes as potential targets to develop less toxic cultivars for people with celiac disease.

Genomic and proteomic technologies that will continue to provide new tools for understanding variation and function of seed storage proteins in wheat are discussed.

Keywords: Bread wheat; Glutenins; Gliadins; Breadmaking quality; Coeliac disease
1. Introduction

Bread wheat (*Triticum aestivum* L.) is one of the most important food crops, with current annual global production of over 680 million tonnes providing approximately one-fifth of the total calorific input of the world population. Meeting the multitude of quality demands for a wide range of products in various countries has to be combined with the needs of an increase in population, constraints of land and water availability, and global climatic changes. The improvements in wheat grain quality has to accompany high yield potential, resistance to biotic and abiotic stresses, and broad adaptation, particularly in environments undergoing the largest effect of climate change. In the most rapidly growing markets of South Asia and China, grain quality improvement has become much more important largely due to rapidly increasing incomes and food diversity. The traditional quality aspects of wheat also need to evolve as new processing technologies are established, and as concerns on health issues increase significantly. The integration of various disciplines such as functional genomics, proteomics, bioinformatics, genetic transformation, breeding and exploitation of new genetic resources, is rapidly promoting our understanding of the genetic and biochemical bases of quality traits in wheat.

Ongoing activities for quality improvement are also capturing information from other crops and model organisms, such as rice and *Arabidopsis*, in order to define genes that underpin the unique quality attributes. The large datasets generated from these responses need to be incorporated into breeding programs in conjunction with high-throughput screening technology in order to efficiently combine traits for wheat cultivars with improved quality.
Wheat seed storage proteins are important quality determinants because they are responsible for dough elasticity and extensibility, and thus for determining the processing qualities in the production of a range of end-products. Biochemical and molecular studies in the past three decades have provided the basis for understanding the genetics, structure and composition of storage proteins (Ma et al. 2009). The development and utilization of functional markers or gene-specific markers for high molecular weight glutenin subunit (HMW-GS), low molecular weight glutenin subunit (LMW-GS), and gliadin alleles have dramatically improved the selection efficiency of breeding materials with desirable genes (Liu et al. 2012). However, most LMW-GS and gliadin genes comprise complex populations of genes (Anderson et al. 2013) with high allelic variation, and thus their contributions to qualities still need to be resolved.

Wheat storage proteins also confer dietary intolerances, such as coeliac diseases (CD) and various allergies (baker’s asthma and wheat-dependent exercise induced anaphylaxis (WDEIA)). Understanding the mechanisms of such diseases from gluten structures and searching for genetic variants conferring reduced intolerances have therefore become crucially important objectives for wheat breeding programs, particularly in developed countries. Juhasz et al. (2012a) proposed a method to define epitope toxicity in wheat proteins, especially the prolamins, by combining database resources, computational tools and proteomic diagnostics. However, much effort is still needed to integrate this information into applied wheat breeding programs.

Previously, Gianibelli et al. (2001) and Shewry et al. (2003a) reviewed the genetics, biochemistry and molecular characterization of glutenin and gliadins in bread wheat, and D’Ovidio and Masci (2004) analyzed in detail the structure, function and genetics of low
molecular weight glutenins. Bonomi et al. (2013) reviewed the structural features of water-insoluble gluten proteins and highlighted the modifications in gluten structure during various processing stages. Similarly, Ribeiro et al. (2013) reviewed the impact of proteomics on the study of gluten proteins and use of transgenesis to improve the quality of gluten. The objectives of this paper are to review the recent advances in defining the nature and properties of wheat storage proteins, focusing on molecular genetics, proteomics, and practical breeding aspects, as well as summarize progress in understanding the health issues of storage proteins.

2. Genetics, molecular characterization and diversity of storage proteins

2.1 HMW-GS

2.1.1 Genetics and polymorphism

HMW-GS accounts for around 12% of the total seed storage protein corresponding to about 1.0-1.7% of the flour dry weight. In the last 20 years, the functional and structural aspects of HMW-GS in relation to dough strength have been defined. HMW-GS genes located at Glu-1 loci on the long arms of homoeologous group 1 chromosomes are named as Glu-A1, Glu-B1 and Glu-D1, respectively (Payne et al. 1980). Each locus produces two subunits of different size, called x-type and y-type subunits (e.g., 1Ax and 1Ay), with comparatively higher and lower molecular weights, respectively. All bread wheat cultivars express 1Bx, 1Dx, and 1Dy subunits while some cultivars express 1By and 1Ax subunits as well. Due to wheat domestication syndrome, the gene encoding the 1Ay subunit usually remains silent in bread and durum wheats. HMW-GS has a lot of allelic variation among wheat germplasm. The Glu-A1 locus has three common allelic variants in bread wheat; however, more than 21 alleles have been documented in different bread
wheat and durum genotypes. More than 69 alleles at \textit{Glu-B1} and 29 alleles at \textit{Glu-D1} have been reported in bread wheat (McIntosh et al. 2013). The AACC glutenin allele database contains the glutenin compositions of over 8,500 wheat genotypes from around the world (http://www.aaccnet.org/initiatives/definitions/Pages/Gluten.aspx). The recent versions of the database also predict the quality of genotype in terms of dough strength (R\text{max}) and extensibility based on both HMW-GS and LMW-GS alleles (Bekes and Wrigley 2013).

Genetic diversity in the \textit{Triticeae} gene pool can provide much information in understanding the variation of HMW-GS, and can also be an important source of genes for quality improvement. Several A-genome wild species and wild tetraploid species (\textit{T. dicoccoides} and \textit{T. turgidum} subsp. \textit{dicoccon}) express the \textit{Glu-Ay} gene (Waines and Payne et al. 1987), and a Swedish bread wheat cultivar was also found to express a \textit{Glu-Ay} gene (Margiotta et al. 1996). This could be very useful for expanding the narrow allelic diversity at the \textit{Glu-A1} locus in both bread wheat and durum wheat where only a limited number of x-type and virtually no y-type subunits are expressed, despite the presence and expression of novel allelic variants in \textit{T. urartu} and \textit{T. monococcum} (Gutierrez et al. 2011). Durum and bread wheat genotypes with four and six subunits, respectively, were also developed by replacing the silenced subunit of \textit{Glu-A1} with expressed ones, and they show an incremental increase in polymeric glutenin quantity, expressed as better dough strength (Alvarez et al. 2009). There are extensive studies on identification and characterization of allelic variation for \textit{Glu-D1} loci from \textit{Ae. tauschii} and D-genome synthetic hexaploids since variation at this locus plays a more significant role in determining dough and end-use product qualities (Xu et al. 2010; Rehman et al. 2010).
Niu et al. (2011) analyzed HMW-GS in *Thinopyrum bessarabicum, Th. intermedium, Lophopyrum elongatum, Aegilops markgrafii* and their addition lines in bread wheat. The genes coding for HMW-GS were identified from several genera of the Triticeae, including *Hordeum, Secale, Taeniatherum, Thinopyrum, Aegilops, Crithopsis,* and *Dasypyrum.* Recently, a superior dough and breadmaking quality allele from *Ae. longissima* was identified by analysis of a Chinese Spring substitution line, CS-1S^{1}(1B) (Wang et al. 2013). These novel HMW-GS alleles may serve as new genetic resources for quality improvement if the potentially positive contributions to processing quality can be confirmed.

2.1.2 Molecular characterization and structural features

The complete coding sequences of 10 HMW-GS alleles including *Ax1, Ax2*, the silent Ay subunit, *Bx7, Bx14, Bx17, By9, Dx2, Dx5, Dy10,* and *Dy12* are known (Forde et al. 1985; Halford et al. 1987, 1992). This has facilitated structure prediction and design of primers to sequence the other alleles used in breeding programs. The x- and y-type subunits share the same structure, signal peptide, N-terminal region, repetitive region and C-terminal region, but x-type subunits are characterized by the presence of a unique tripeptide motif (GQQ), whereas in y-type subunits, the second proline is replaced by a leucine in the GYYPTSPQQ repeat motif (i.e., GYYPTSLQQ). Moreover, the majority of x-type subunits possess four conserved cysteine residues, while y-type subunits usually have seven cysteine residues. It is likely that in HMW-GS, differences in the number of cysteine residues are associated with variation in bread-making quality. For example, *Dx5* associated with superior bread-making quality, has an extra cysteine residue located at the N-terminal part of its repetitive domain (Lafiandra et al. 1993). Similarly, an extra
cysteine residue occurs in the middle of the domain of the Ax2*B subunit, a variant of Ax2*, that shows positive effects on gluten properties (Juhász et al. 2001). In contrast, fewer cysteine residues are present in the N-terminal domains of Bx14 and Bx20 that have negative effects on dough strength (Shewry et al. 2003b). Over-expression of a Bx7 (Bx7OE) subunit resulted in increased dough resistance compared to normal Bx7 gene. It may possibly be due to the insertion of 18 bp in the central repetitive domain without affecting number of cysteine residues (Butow et al. 2003).

Several alleles from Ae. tauschii and D-genome synthetic hexaploids were characterized at the molecular level, including 1Dx1.5' (Lu et al. 2005a), 1Dx2.1' (Lu et al. 2005b), and Dx3' and Dx4' (Wang et al. 2012a). Compared with the 1Dx5 subunit, 1Dx3' has a six amino acid insertion at 146-151 whereas 1Dx4' has a nine amino acid deletion compared with the 1Dx3' subunit. Wang et al. (2012a) also predicted that the secondary structure of the 1Dx3' subunit has a significantly higher α-helix and β-strand contents, indicating it might have positive effects on dough quality. Feng et al. (2011) isolated and characterized a novel chimeric HMW-GS, Dx5', from wheat line W985. This subunit has six helices and one strand, including four helices in the repetitive domain capable of enhancing the dough properties, and possibly having a high potential for quality improvement. In summary, molecular characterization of HMW-GS genes and deduced amino acid structures help to predict the protein structure and any potential value in improving dough properties.

2.1.3 Proteomic and molecular diagnosis

Sodium dodecyl sulfate poly-acrylamide gel electrophoresis (SDS-PAGE) has been widely used for HMW-GS diagnosis. However, limitations involving co-migration of
some subunits, and difficulty in detecting differences in expression levels, led to inaccurate identification of alleles that differ in functional properties. In order to overcome these problems, more powerful separation techniques have been used for HMW-GS separation. For example, high performance capillary electrophoresis (HPCE), a method with various electrophoretic modes, is capable of rapid separation of HMW-GS with high resolution from small samples. HPCE was used to identify new allelic variants of HMW-GS in bread wheat and related species (Li et al. 2006, Gao et al. 2010). Reversed-phase high-performance liquid chromatography (RP-HPLC), that separates proteins based on surface hydrophobicities, provides automated separation, and also allows quantitative analysis of wheat proteins (Gao et al. 2010). Along with the development of wheat proteomics over the past 20 years, two-dimensional electrophoresis (2-DE) and various mass spectrometry techniques, such as matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF-MS), were developed, and these permit high resolution and accurate molecular mass determinations of different protein components and HMW-GS with similar molecular sizes (Zhang et al. 2008a).

Gao et al. (2010) performed a comparative analysis for HMW-GS identification by SDS-PAGE, RP-HPLC, HPCE and MALDI-TOF-MS. Overestimation of molecular mass and low resolution was the main limitation of SDS-PAGE; however, it is still the simplest and cheapest technique suitable for large-scale and high-throughput HMW-GS screening for wheat genotypes in breeding programs, especially when the glutenin composition is clear. MALDI-TOF-MS has several technical advantages including high throughput, high resolution, and accuracy (Zhang et al. 2008a). RP-HPLC and HPCE are considered to be
intermediate between MALDI-TOF-MS and SDS-PAGE. They also show significant
advantages in resolving some HMW-GS, such as Bx14+By15 and Bx20 that cannot be
easily discriminated by SDS-PAGE. HPCE is used only for specific studies when
stringent characterization of novel alleles is required. Both techniques demonstrated
higher resolution and reproducibility than SDS-PAGE, but lower detection power than
MALDI-TOF-MS. Moreover, there are no significant differences in throughput sample;
100 samples can easily be handled per day with all these techniques. Thus, SDS-PAGE is
still the most widely used method, but is rapidly being replaced by PCR methods in
routine breeding programs as DNA is available at any stages of crop development.
MALDI-TOF-MS is an efficient and accurate method, and HPLC and RP-HPLC serve as
complementary methods that can be used to characterize novel alleles.

Advances in molecular biology have enabled protein-based limitations of HMW

glutelin allele resolution to overcome by the use of specific PCR markers. Development
of these markers is based on DNA polymorphisms among the glutenin subunit genes and
once available they can be considered as perfect or functional markers for HMW-GS
alleles. The major advantages are high-throughput capability for assessing different
alleles in breeding materials, and genotyping can be performed during the vegetative
growth stages (Liu et al. 2008). Allele-specific PCR (AS-PCR) markers are available for
the three most common x-type subunits at the Glu-I locus i.e. 1Ax2*, 1Ax1 and 1Ax Null
(Ma et al. 2003; Liu et al. 2008). At Glu-B1, allele specific markers are available for x-
type subunits Bx7, Bx14, Bx17 (Xu et al. 2008), Bx6 (Schwarz et al. 2004), Bx7OE
(Ragupathy et al. 2008; Butow et al. 2003), and y-type subunits By8, By16 and By18
(Lei et al. 2006). At Glu-D1, markers are available to identify 1Dx2, 1Dx5, 1Dy10 and
IDy12 (Liu et al. 2008). AS-PCR provides a fast and efficient tool for screening desirable HMW-GS for maintenance and improvement of quality parameters in breeding programs. However, these markers are not replacements for other detection techniques such as SDS-PAGE, especially in regard to related wild species because these PCR methods are redundant when unknown subunits are present. It is important to validate the presence of specific alleles with diagnostic molecular and biochemical techniques when characterizing wild species and their derivatives.

2.2. LMW-GS

2.2.1 Genetics and polymorphism

LMW-GS represents about 60% of total glutenins and about one third of the storage proteins. The genes encoding LMW-GS, named Glu-A3, Glu-B3 and Glu-D3, are located on the short arms of group 1 chromosomes (Sreeramulu and Singh 1997). In addition, three new loci Glu-2, Glu-4 and Glu-5 (McIntosh et al. 2013) located on chromosomes 1B, 1D and 7D, respectively, have also been reported. Biochemical classification revealed three classes of LMW-GS; B, C and D types based on SDS-PAGE mobility (Jackson et al. 1983), with the B type LMW-GS further divided into three classes, LMW-m, LMW-s and LMW-i based on the first amino acid residue that may be methionine, serine and isoleucine, respectively (Muccilli et al. 2010). Genes at the Glu-A3 locus encode mainly LMWi type subunits (Zhang et al. 2004).

In early studies, 20 LMW-GS alleles were identified in 222 bread wheat cultivars from 32 countries (Gupta and Shepherd 1990), including six alleles at Glu-A3, nine at Glu-B3 and five at Glu-D3 locus. McIntosh et al. (2013) list 17, 26 and 11 alleles at the Glu-3 loci. However, this does not include allelic variants reported in wheat wild relatives.
within the Triticeae. Allelic richness at Glu-A3 is relatively low, whereas Glu-B3 shows the highest levels of polymorphism. There is a close genetic linkage between gliadins encoded by Gli-1 loci and LMW-GS encoded by Glu-3 loci (Gianibelli et al. 2001). This linkage helps in diagnosis of several Glu-B3 and Glu-D3 alleles in wheat genotypes. Several gliadins are reliable markers for diagnosing LMW-GS alleles due to their easy detection (Jackson et al. 1996).

Many genes encoding LMW-GS have been isolated and analyzed in cultivated and wild species of the Triticeae. Ae. tauschii, an important species for genetic studies of LMW-GS (Rehman et al. 2008), exhibits greater variation than wheat in the coding sequence of LMW-GS. Other species studied in detail include T. monococcum, T. urartu (Tranquilli et al. 2002), T. turgidum var. dicoccoides (Ciaffi et al. 1993) and var. polonicum (AABB) (Liu and Shepherd 1996), Ae. umbellulata, Ae. comosa, Ae. markgrafii and Ae. speltoides (Li et al. 2010). The variability found for LMW-GS genes in wild relatives indicates a large potential for improving dough properties in wheat. Other studies of Glu-3 genes have extended to more distant relatives, including Elytrigia (Gupta and Shephard 1990), Dasypyrum (Blanco et al. 1991), Hordeum (Atienza et al. 2002), polyploid Aegilops spp. (Li et al. 2008a), Agropyron elongatum (Luo et al. 2005), Crithopsis delileana (Guo et al. 2008), Hordeum chilense, and H. brevisubulatum (Piston et al. 2005). Comparative analyses of nucleotide sequences of LMW-GS revealed some important differences among species. For example, Hordeum chilense and A. elongatum lack the N-terminal regions in the predicted mature proteins (Piston et al. 2005). The potential of LMW-GS from wild species to improve the processing quality
over that of desirable alleles available in current elite wheat germplasm can now be
achieved by facilitating their transfer into bread wheat.

2.2.2 Genomic organization and complexities of LMW-GS
Unlike the two copies of genes in HMW-GS, the exact gene copy numbers of LMW-GS
genes remain unknown, which is a major bottleneck to distinguishing most of the
members in bread wheat. Copy numbers of LMW-GS genes were estimated to range
from 10–20 to 30–40 (Ikeda et al. 2002; D’Ovidio and Masci 2004; Huang and
Cloutier 2008; Dong et al. 2010; Zhang et al. 2011a, b). In the past 10 years, significant
progress has been made in understanding the structural complexities of LMW-GS by
analyzing orthologous sequences in a wide array of wheat genetic resources. BAC library
resources, in particular, provided impetus for identifying and cloning multi-gene families
of LMW-GS (Wicker et al. 2003; Johal et al. 2004; Gao et al. 2007; Huang and
Cloutier, 2008; Dong et al. 2010). Complete ORF sequences for LMW-GS genes in four
wheat cultivars, i.e., Norin 61 (Ikeda et al. 2002, Ikeda et al. 2006), Glenlea (Huang
and Cloutier 2008), Xiaoyan 54 (Dong et al. 2010), and Chinese Spring (Zhang et al.
2011b), and their comparative analyses successfully discriminated members of the
LMW-GS gene family. Ikeda et al. (2002) isolated LMW-GS genes from Norin 61, and
classified them into 12 groups on the basis of sequences from N-, C-terminal domains
and deduced amino acids sequences. Huang and Cloutier (2008) screened a BAC library
to study the genomic organization of LMW-GS and identified 82 positive BAC clones
from Glenlea, and 12 unique active LMW-GS genes were identified and all contained
eight cysteine residues. Dong et al. (2010) dissected the LMW-GS genes in Xiaoyan 54
using BAC library screening, expression profiling, and proteomics analysis. Of 14 unique
LMW-GS genes, 11 active genes were identified, three at *Glu-A3*, two at *Glu-B3* and six at *Glu-D3*. Recently, *Zhang et al.* (2013a) reported that at least 15 LMW-GS genes were present in individual accessions from Chinese core collections, 4-6 of which were located at *Glu-A3*, 3-5 at *Glu-B3*, and 8 at the *Glu-D3* loci. Expression and sequence analysis confirmed the presence of 9-13 active LMW-GS genes in each accession. Thus, it is reasonable to say that the number of active LMW-GS genes in bread wheat is less than 15. A consistent finding among these studies was that *Glu-D3* locus contains a higher number of active genes than the other loci, but the diversity of LMW-GS genes at *Glu-D3* was the lowest (*Zhang et al.* 2013a).

BAC clones derived from the durum wheat cultivar Langdon revealed that five gliadin and gliadin-like genes (pseudo or inactive gliadin genes) and two typical LMW-GS genes were present in a 265 kb BAC contig derived from the short arm of chromosome 1AS, and eight gliadin genes and one typical LMW-GS gene were found in a 140 kb genomic DNA fragment derived from chromosome 1BS (*Gao et al.* 2007). The gliadin genes are more clustered than LMW-glutenin genes which are separated from each other by much larger distances (*Anderson et al.* 2013). The dispersion of LMW-glutenin genes is the result of both insertion of large blocks of repetitive DNA owing to the rapid amplification of retrotransposons and even the presence of genetic loci between them. Moreover, small gene clusters were found in the contigs of each prolamin family. These small clusters were scattered with the longer DNA stretches generally preventing assembly of complete contigs of LMW-GS genes.

### 2.2.3 Molecular characterization and structural features
Several hundreds of LMW-GS gene sequences have been deposited in GenBank and used to predict primary protein structures. This has provided a basis for designing gene- or locus-specific primers for subsequent isolation and cloning. LMW-GS genes at the *Glu-B3* and *Glu-D3* loci in bread wheat were identified using gene-specific primers, and molecular markers were subsequently developed for specific haplotypes (Zhao et al. 2006, 2007a, b). Subsequently, a LMW-GS gene marker system was developed based on conserved sequences and length polymorphisms among the LMW-GS genes and was used to characterize a core collection of Chinese wheats (Zhang et al. 2011a, 2013a). More than 15 genes, including 9 to 13 active genes, were successfully separated from each accession, greatly enhancing our understanding of the molecular structure of LMW-GS. However, very few breeding programs have the facilities (a 3730 DNA analyzer) to survey their germplasm using the marker system, and it will be very difficult to establish associations between these genes and dough quality as well as end-use product quality. At a more applied level, seven allele-specific markers were designed based on single nucleotide polymorphisms (SNPs) among allelic variants at the *Glu-A3* locus (Wang et al. 2010), and 10 allele-specific markers were designed at the *Glu-B3* locus (Wang et al. 2009). These markers were validated and successfully used to characterize LMW-GS composition in Chinese and CIMMYT wheat accessions, and the results were in agreement with those from SDS-PAGE (Jin et al. 2011; Liu et al. 2010). The *Glu-D3* locus displayed less variation and only six active LMW-GS genes were present in all bread wheat cultivars (Zhao et al. 2006, 2007a). These six active LMW-GS genes were extremely conserved among accession in the Chinese core collection (Zhang et al. 2013a). No gene-specific molecular markers have been developed to detect allelic...
variation at the *Glu-D3* locus.

LMW-GS gene sequences are intronless, similar to other prolamin genes. More than 200 nucleotide sequences and cDNA clones from LMW-GS genes have been reported in wheat and wild species (*An* et al. 2006; *Ikeda* et al. 2002; *Long* et al. 2005; *Zhao* et al. 2006, 2007a, b). In most of these cases, only a few LMW-GS genes were isolated from a cultivar due to the complex composition of genes. Collectively, more than 100 LMW-GS genes have been characterized and sequenced from bread wheat, including several partial and pseudo genes (*Cloutier* et al. 2001; *Zhang* et al. 2004). Sequence analysis suggested that the deduced proteins contain a signal peptide, a central repetitive region rich in proline and glutamine, and N and C terminal non-repetitive domains, similar to other prolamins (*Colot* et al. 1989). The N-terminal amino acid sequence of LMW-s type subunits generally start with SHIPGL-, and LMW-m subunits are more variable starting with METSHIPGL-, METSRIPGL- or METSCIPGL- (*Masci* et al. 1993). LMW-i type subunits actually lack an N-terminal region and start directly with the repetitive domain ISQQQQ- being the deduced N-terminal sequence. Besides those three types of subunits, a new LMW-l type subunit was identified in *Aegilops* species with the first amino acid residue being leucine (*Wang* et al. 2011).

2.2.4 Proteomic and molecular diagnosis

SDS-PAGE is considered one of the simplest techniques to identify LMW-GS; however, few laboratories in the world have the expertise to conduct it for large numbers of samples in breeding programs. More recently, reversed-phase ultra performance liquid chromatography (RP-UPLC) has been developed to separate LMW-GS in a more effective way (*Yu* et al. 2013). In this system, small diameter columns such as 1.7 μm
columns, can be used and high column performance (100,000-300,000) can be undertaken. Compared with the fastest high-performance liquid chromatography analysis currently available, RP-UPLC analysis is nine times faster in speed, two times higher in resolution, and three times higher in sensitivity, and thus provides a powerful tool for rapid separation and characterization of LMW-GS.

SDS-PAGE, 2-DE, MALDI-TOF-MS and PCR based markers were developed to detect the Glu-3 allelic variation. Liu et al. (2010) compared the four techniques to assess their suitability for use in breeding programs. They indicated that PCR-based markers are the simplest, most accurate, lowest cost technique and therefore recommended it for the identification of Glu-A3 and Glu-B3 alleles in breeding programs. MALDI-TOF-MS showed great potential in separation of LMW-GS, but much more effort is needed before it can be used routinely in breeding program. Considering the Glu-D3 locus, it should be noted that its effect on quality variation among wheat cultivars is very small compared to the Glu-A3 and Glu-B3 loci (Zhang et al. 2012). A total of 17 allele-specific markers for Glu-A3 and Glu-B3 loci have been reported and used, and multiplex PCR protocols have been developed to reduce costs of screening in practical breeding programs (Wang et al. 2009, 2010). Application of functional markers for identification of LMW-GS in various types of wheat germplasm has also been reported by others (Jin et al. 2011; Ram et al. 2011).

A combination of different techniques was required to identify certain alleles of LMW-GS and these combinations are especially useful when characterizing new alleles from genetic resource materials. As more alleles are reported at Glu-A3 and Glu-B3 in bread wheat, more molecular markers will be needed to distinguish them in breeding
germplasm. Liu et al. (2010) recommended a standard set of 30 cultivars to represent all known LMW-GS allelic variants in future studies. Among them, Chinese Spring, Opata 85, Seri 82 and Pavon 76 were recommended as a core set for use in SDS-PAGE gels. Use of the standard cultivar set was recommended to promote and facilitate information sharing on LMW-GS, ultimately enhancing universal quality improvement efforts in wheat. In addition to HMW-GS and LMW-GS, gliadins account for 40-50% of total wheat seed storage proteins and have impacts both on processing and nutritional quality, but their effects on processing quality is less significant than the HMW- and LMW-GS. Gliadins can be separated into α-/β-, γ-, and ω-gliadins in acid-PAGE based on differences in mobility, while α- and β-gliadins are structurally very similar (Kasarda et al. 1984). All ω- and many γ-gliadins are encoded by Gli-1 loci on the short arms of group 1 chromosomes, and all α-, many β-, and some γ-gliadins are encoded by Gli-2 loci on the short arms of homoeologous group 6 chromosomes. In the wheat gene catalogue, 23, 24 and 15 alleles are listed for Gli-A1, Gli-B1 and Gli-D1 in bread wheat and durum wheat, respectively. Similarly, among Gli-2 loci, Gli-A2 encodes 36 alleles, Gli-B2 encodes 47 alleles, and Gli-D1 encodes 31 alleles. The main factors preventing the participation of gliadins in polymeric glutenin structure is lack of free cysteines and the ability to form inter-molecular S-S linkages, hence they have small effect on processing quality. Qi et al. (2009) analyzed 170 γ-gliadin genes from bread wheat and closely related species and found a wide range of amino acid compositions in γ-gliadins, and those γ-gliadins from subgroups SG-10 and SG-12 and γ-gliadins with a short repetitive domain were considered more nutritional. However, compared with glutenins, there has
been little progress during the last 10 years in understanding the structure and function of gliadins. Therefore, no further details will be presented here, although associations of gliadins with health issues will be discussed in a following section of this paper.

3. Integrated genomics, functional analysis and evolution of wheat storage proteins

3.1 Genomic resources and databases to understand prolamin-omics

Recent progress in understanding the genomic organization and in resolving the structural complexities of prolamins are largely due to the establishment of genomics resources and database such as bacterial artificial chromosome (BAC) libraries, availability of saturated physical maps, and expressed sequence tags (ESTs) (Feuillet et al. 2012). The choice of database to be searched is critical, particularly in organisms where a completely annotated genome sequence is not available. In addition, new sequences are constantly being added to gene and protein databases. In the past few years, several hundred new glutenin and gliadin gene sequences were added at the National Center for Biotechnology Information (NCBI) non-redundant database.

A large number of EST collections derived from developing seed cDNA libraries have allowed identification of new classes of prolamin genes and novel seed storage proteins (Anderson et al. 2001; Anderson et al. 2012). The use of large insert BAC libraries permits study of the complex structural organization of prolamin gene families (Gu et al. 2004; Johal et al. 2004). Characterization of wheat BAC clones using an α-gliadin gene probe revealed that positive BAC clones carried one to five copies of the α-gliadin gene per BAC (Gu et al. 2004). In contrast, when the BAC clones were used to characterize LMW-GS genes, no single BAC clone was found to carry more than one copy of the LMW-glutenin gene. Several BACs were needed to provide all seven copies
of a LMW-GS gene in *Ae. tauschii*, the D genome donor of bread wheat (Johal et al. 2004). The cultivar Glenlea BAC library (Nilmelgoda et al. 2003) is an important source for gene discovery, providing the template for genome organization studies of LMW-GS. Eighty two positive clones were identified to have LMW-GS genes, of which 12 were unique and 7 were pseudogenes (Huang and Cloutier 2008). The genomic organization of several LMW-GS and gliadin genes was similarly studied using the BAC library database of durum wheat cultivar Langdon (Gao et al. 2007). The EST databases from bread wheat cultivar Butte 6 also provided substantial information for understanding the structural complexities of α-gliadins (Altenbach et al. 2010a) and γ-gliadins (Altenbach et al. 2010b; Anderson et al. 2013). Now, the available whole genome shotgun draft sequences of *T. urartu* (Ling et al. 2013) and *Ae. tauschii* (Jia et al. 2013) will accelerate deeper and more systematic studies to meet future challenges in understanding wheat processing qualities. Database enhancement by inclusion of cultivar-specific prolamin gene sequences maximizes the number and quality of peptide identifications and increases sequence coverage of storage proteins by proteomic diagnosis. This approach makes it possible to distinguish closely related proteins, to associate individual proteins with sequences of specific genes, and to evaluate proteomic data in a biological context to better address their effect on flour quality.

### 3.2 Gene cloning and functional analysis

Integrated approaches in the modern genomics era have facilitated cloning and functional analysis of new genes for HMW-GS (Butow et al. 2003; Li et al. 2008a; Zhang et al. 2008b), LMW-GS (Jiang et al. 2008; Li et al. 2008b, c), and gliadin (Wang et al. 2012b; Zhu et al. 2010) genes. The strategy for cloning and functional analysis of genes
encoding storage protein is presented in Figure 1. The initial step is usually the identification of the new storage protein that requires validation by multiple proteomic diagnostic platforms. The aforementioned genomic databases provide templates to design primers to isolate full-length open reading frames (ORF) of the encoding genes, which are then analyzed through sequencing and functional investigations. Functional analysis of the encoding genes indicates that storage proteins evolve by unequal crossing over and slip-mismatching (Zhang et al. 2008b). Translation of nucleotide fragments into deduced peptides helps in demonstrating the structural basis of differences between subunits. Similarly, expression of fragments in E. coli is useful for prediction of possible mechanisms of duplication and deletion of large fragments (Zhang et al. 2008b).

3.3 Phylogeny and evolutionary studies in wheat storage proteins

It is thought that tandem gene duplication in globulins gave rise to the prolams (Xu and Messing 2009). The ancestral prolamin gene in wheat progenitors gave rise to the HMW-GS, which constitute the group III prolamins and are unique to the subfamily Pooideae. During evolution of the grasses, prolamin genes were copied and inserted either in tandem or dispersed to different chromosomal locations, which in turn gave rise to further amplifications. Donor copies were either maintained or lost by unequal crossing over (Xu et al. 2012). Genome-wide dispersal gave rise to new groups of prolams, the largest of which is group II, including the LMW-GS, α-, γ-, and ω-gliadins in wheat and the β- and ω-zeins in maize. Group I represents the youngest prolamins that include the α- and ω-zeins in sorghum and maize that are unique to the Panicoideae (Zhang et al. 2013b).

Within the wheat seed prolams, two evolutionary lines resulted in the major prolamin classes (Xu and Messing 2009), with both lines originating separately within...
the super family of related seed proteins. The LMW-GS and gliadins, which share enough
similarities to establish a common origin, emerged from one line. Phylogenetic analysis
showed that α-, γ-, and β-gliadins from barley, and LMW-GS genes from wheat were
closely related, and they differentiated at a relatively late stage. In contrast, ω-gliadins
and HMW-GS genes are separated by a much larger genetic distance compared to other
storage protein genes, indicating that differentiation of these two kinds of genes occurred
very early (approximately 74 million years ago, MYA) (Zhu et al. 2010). Among LMW-
GS, LMW-m type genes could be considered to be the most primitive forms than other
LMW-GS genes, and LMW-s and LMW-i type genes were variant forms among the
LMW-GS gene families (D’Ovidio and Masci 2004).

The second evolutionary line of wheat prolamins contains only the HMW-GS, and
likely arose as a tandem duplication of a globulin gene, with one of the resulting genes
evolving into a HMW-GS (Kong et al. 2004). A subsequent tandem duplication of a
chromosome fragment containing an ancestral HMW-glutenin and globulin genes
resulted in the x- and y-type HMW-GS. The two HMW-glutenin genes remained as two
conserved genes, whereas the gliadins and LMW-glutenin genes radiated into large multi-
gene families with highly variable copy numbers of member. All HMW-GS genes
encoded by Glu-1 loci from different Triticum species are homologous, and x-type alleles
show greater variation than the y-type alleles. Therefore, both x- and y-type subunit genes
could have diverged from an ancestral sequence prior to separation of the wheat
genomes, and the discrepancy is probably due to relatively rapid variation of the x-type
genes. Network analysis of Glu-D1-1 HMW-GS from Ae. tauschii demonstrated common
ancestors of the other Glu-D1-1 alleles in an associated star-like phylogeny, suggesting
that there were at least four independent origins of bread wheat (Zhang et al. 2008b). In addition to unequal homologous recombination, duplication and deletion of large fragments occurring in \textit{Glu-D-1-1} and \textit{Glu-3} alleles were attributed to illegitimate recombination (Zhang et al. 2008b; Li et al. 2008; Wang et al. 2011). The time of the duplication event of x- and y-type subunits could extend to 17 MYA, suggesting that during the origin of cultivated wheat the A and B genomes did not evolve simultaneously, and that they might have been introgressed into hexaploid at different times and rates (Blatter et al. 2004). By analyzing haplotype variation at the \textit{Glu-D1} locus, Dong et al. (2013) established that the \textit{Glu-D1d} allele was likely already differentiated in ancestral hexaploid wheat and \textit{T. spelta} around 10,000 years ago and was subsequently transmitted and domesticated.

4. Quantifying prolamin allelic effects on processing qualities

The quality effects of HMW-GS and LMW-GS have been well studied to facilitate incorporation of desirable quality genes into bread wheat germplasm. Usually two types of populations have been used to study the effects of glutenin alleles on dough properties and end-use qualities: viz., mapping populations developed from biparental crosses of parents having different glutenin alleles (recombinant inbred lines and doubled haploid lines), and non-structured populations, general collections of cultivars, biotypes from a cultivar, and breeding lines. Near-isogenic lines (NILs) differing by single glutenin alleles in a common genetic background were developed and they proved to be the best genetic stocks to study effects of single alleles on quality. In some recent studies, an array of NILs derived in cultivar Aroona were used to determine quality effects (Jin et al. 2012; Zhang et al. 2012). Similarly, glutenin identification platforms that are more
sophisticated and high-throughput (AS-PCR markers and MALDI-TOF) greatly
improved such studied (Liu et al. 2010).

4.1 Individual allelic effects and interactions of glutenin on processing quality
The allelic effects of HMW-GS on dough properties and bread-making quality are well
investigated because the alleles are relatively easy to identify. It is generally agreed that
\textit{Glu-D1} has more profound effects on processing quality than \textit{Glu-A1} and \textit{Glu-B1} (Zhang
et al. 2009). Among HMW-GS, x-type subunits are much more important than y-type
subunits and their contributions are much larger (Wieser and Kieffer 2001). The
presence and amount of subunit $Dx5$ is particularly significant due to its special disulfide
structure.

Previously, Payne (1987) proved that the subunits \textit{Ax1} or \textit{Ax2*} at \textit{Glu-A1}, \textit{Bx7+By8},
\textit{Bx17+By18}, and \textit{Bx13+By16} at \textit{Glu-B1}, and \textit{Dx5+Dy10} at \textit{Glu-D1} were desirable
subunits for bread-making quality. Several groups later confirmed this in different
germplasm pools (Brönneke et al. 2000; He et al. 2005). Recently, ion beam-induced
mutants of bread wheat cultivar Xiaoyan 81 were used to understand the effect of
mutations at \textit{Glu-1} loci on gluten functionality (Yang et al. 2014). Based on the glutenin
macro polymers (GMP) contents and elastic and viscous moduli, the \textit{Glu-1} loci can be
ranked as \textit{Glu-D1}>\textit{Glu-B1}>\textit{Glu-A1}. According to the contributions of individual glutenin
subunits to dough strength and pan bread making quality, different glutenin subunits at
each of the three loci were ranked as: \textit{1}>\textit{2}>\textit{N} at \textit{Glu-A1}, \textit{7+8≥13+16>17+18=7+9} at \textit{Glu-
B1}, and \textit{5+10≥2+12>4+12} at \textit{Glu-D1} (He et al. 2005; Wrigley et al. 2009). However,
much more work is needed to understand the associations between HMW-GS and
qualities of Chinese/Asian products, such as steamed bread and noodles. Jin et al. (2012)
found that variation in subunits at Glu-A1 had a significant effect on the shape of steamed bread, while Glu-B1 made significant contribution to stress relaxation of steamed bread.

LMW-GS also contributes significantly to dough strength and extensibility. At Glu-A3, the ranking of alleles for dough strength is Glu-A3d > Glu-A3b > Glu-A3c > Glu-A3f > Glu-A3a > Glu-A3e, whereas ranking for dough extensibility is slightly different, viz., Glu-A3c > Glu-A3b ≥ Glu-A3f > Glu-A3e (Zhang et al. 2012). There are controversies regarding different kinds of populations or collections. For example, Cane et al. (2008) reported that Glu-A3e was correlated with inferior dough resistance and extensibility, whereas Zheng et al. (2009) found that Glu-A3e was a favorable allele for dough-mixing properties. The contribution of Glu-A3 was the highest among all LMW-GS loci, especially Glu-A3f which had a strong positive effect on end-use quality against the backgrounds of all HMW-GS combinations. He et al. (2005) reported that Glu-A3d was slightly better than others for dry white Chinese noodle quality, and Glu-B3j associated with the 1B/1R translocation, had strong negative effects on all quality traits. At Glu-B3, alleles were ranked Glu-B3b = Glu-B3d = Glu-B3g > Glu-B3h > Glu-B3a > Glu-B3c for dough strength, whereas ranking for dough extensibility was Glu-B3i > Glu-B3f = Glu-B3g > Glu-B3h > Glu-B3a > Glu-B3b > Glu-B3d (Zhang et al. 2012). At Glu-D3, the effect of alleles on most traits were not reported; alleles were ranked Glu-D3d = f > Glu-D3e > Glu-D3a = Glu-D3c = Glu-D3b with respect to dough strength (Cornish 2007; Zhang et al. 2012). It is generally agreed that the effects of Glu-D3 alleles on processing qualities are less significant in comparison with those at Glu-A3 and Glu-B3.

4.2 Effects of gliadins on processing quality
The complex genetic structure, linkage with LMW-GS and inheritance in blocks (not individual alleles) are the main bottlenecks to dissecting the functional properties of individual gliadin genes in wheat. Cysteine residues in gliadins facilitate interchain cross links with glutenins that influence flour characteristics. Moreover, certain proteins with amino acid sequences very similar to gliadins, contain an extra cysteine residue that enables them to be incorporated into the glutenin polymer group (D’Ovidio and Masci 2004). These proteins are structurally similar to gliadins, but functionally similar to LMW-glutenin subunits. It has been hypothesized that these proteins serve as chain terminators of the glutenin polymer thereby limiting its size and influencing the quality of the flour. Gliadin proteins contribute little to resistance and extension in dough and are mainly related to dough cohesiveness (Uthayakumaran et al. 2000). The gliadin fraction reportedly contributes to viscosity properties. Piston et al. (2011) analyzed transgenic lines with γ-gliadins silenced by RNAi technology. Mixograph testing revealed that silencing increased the contents of other gluten proteins and quality parameters were thereby slightly affected. Similar findings were obtained when γ-gliadins were silenced in three different cultivars, and it was concluded that reduction in γ-gliadins had no direct effect on mixing and bread-making properties, and the compensatory effects on the synthesis of the other prolamins may result in stronger dough with improved over-mixing resistance.

4.3 Quality effects of “interaction and relative proportion” of different gluten fractions

Both HMW- and LMW-GS are important factors in determining dough properties and end use quality; it is likely that HMW-GS are more important for dough strength, and
LMW-GS more important for dough extensibility. Actually, interactions between
glutenins and gliadins, and between HMW-GS and LMW-GS are considered as important
variables (Bekes et al. 2006). Therefore, it is not a simple matter of recommending a
desirable combination of HMW-GS and LMW-GS for quality improvement although
selection of desirable subunits/alleles will generally benefit breeding efforts for quality
improvement. On a qualitative basis, ranking of glutenin fractions have been well
described by Jin et al. (2012) and Zhang et al. (2012). Based on data collected from
global publications, Bekes et al. (2006) reported that for dough strength, the interaction
among glutenin fractions can be ranked as HMW-HMW>HMW-LMW>LMW-LMW,
while for gluten extensibility, HMW-LMW>LMW-LMW>HMW-HMW. Similarly, the
contribution of individual loci for gluten strength was ranked as Glu-D1>Glu-B1>Glu-
A1>Glu-A3=Glu-B3>Glu-D3, whereas for gluten extensibility, the ranking was Glu-
A3>Glu-B3>Glu-D3>Glu-A1=Glu-B1=Glu-D1. Overall, the contributions of protein
alleles can be summarized in a pictorial model modified from Ikeda and Takata (2012)
in Figure 2 by plotting the gluten strength effect of individual glutenin alleles in breeding
implications.

In quantitative terms, the relative proportion of gluten fractions (glutenins and
gliadins) and their sub-fractions (x- and y-type HMW-GS) are important determinants of
bread-making quality. There is still a debate as to the role of the various protein classes
on bread-making parameters such as dough properties, loaf volume, and crumb
characteristics. For example, gliadin proteins were reported to be closely related to loaf
volume in some studies (Khatkar et al. 2002), but in other studies gliadin proteins had
no significant effects on loaf volume, and glutenin proteins were the major components
responsible for loaf volume (Uthayakumaran et al. 2002). Zhang et al. (2007b) also found that gliadin/glutenin ratios showed significant and negative associations with dough properties and Chinese steamed bread quality. Alteration of this ratio had a complex effect on the dough extensibility, indicating the different rheological properties of glutenins and gliadins (Uthayakumaran et al. 2000).

HMW-GS/LMW-GS weight ratios vary between 0.18 and 0.74, suggesting the amount of LMW-GS (55 to 85%) may vary from only slightly higher to an almost six fold that of HMW-GS (15 to 45%) (see review from Veraverbeke and Declour 2002). HMW-GS/LMW-GS ratios determined dough extensibility (Zhang et al. 2007a) and were correlated with farinograph development time and stability (Zhang et al. 2007b). Synthetic dough comprising homopolymers of HMW-GS were found to be stronger than that those composed of LMW-GS homopolymers (Beasley et al. 2002). The mixing time, peak resistance, maximum resistance to extension, and loaf height increased with increases in HMW/LMW ratio, whereas resistance to breakdown and extensibility decreased with the incorporation of glutenin subunits (Uthayakumaran et al. 2000). The contribution of different glutenin subunits to the molecular properties of dough is directly related to the polymeric structure of the glutenin protein. Both variations in structure and in relative amounts of HMW-GS correlated strongly with dough strength, whereas alteration in LMW-GS and gliadin composition affected dough extensibility.

5. Health related aspects of storage proteins

Wheat protein is one of the most extensively utilized proteins in the human diet, and it is also one of the most common causes of food-related allergens and intolerances. The terms wheat-related intolerance (sensitivity) and wheat allergy are often used interchangeably. Wheat allergy refers to the negative reaction where symptoms appear
rapidly following exposure to macro-molecules such as proteins. Several recent reviews address different aspects of the role of storage proteins in allergies and their clinical and pharmaceutical applications. Tatham and Shewry (2008) reviewed the cereal proteins responsible for dietary and respiratory allergies in relation to their structural and biological properties. Sapone et al. (2012) summarized the current knowledge about the main forms of gluten reactions which may be allergic and autoimmune (coeliac disease, dermatitis herpetiformis, and gluten ataxia). They also outlined the pathological, clinical and epidemiological differences and proposed new nomenclature and classifications for gluten related disorders. Mamone et al. (2011) presented an in-depth review on the proteomics of allergens in wheat storage proteins and their derived products. Juhasz et al. (2012a) proposed a method to identify the “epitope toxicity” by analyzing the number and composition of epitopes on a single protein. The use of knowledge about the composition and level of epitopes present in a cultivar is restricted due to the absence of an efficient prediction methodology. In this regard, the “epitope toxicity” value obtained by utilizing wheat proteome data is a significant measure and can be applied to wheat improvement and food industry for development of gluten-free or less toxic products. Moreover, the published genome of Brachypodium distachyon, a diploid wild grass with low prolamin content, can be effectively utilized to study allergen potential of prolamin and non-prolamin proteins (Juhasz et al. 2012b). Given the recent publication of these reviews, we will only provide an overview of toxic prolamin-related epitopes in different wheat genetic resources. A summary of prolamin-related toxic epitopes in wheat is given in Table 1. Allelic variation for toxic prolamin epitopes is also reviewed in order to develop strategies to address these disorders in wheat.
5.1 Coeliac disease

Coeliac disease (CD) is the most important disease resulting from gluten intolerance. It is an intestinal T cell-mediated disease caused by an uncontrolled immune response to gluten, which contains several peptides that constitute the main toxic component in CD (van de Wal et al. 1998). It causes a medium to severe chronic inflammatory condition that is characterized by injury to the lining of the small-intestine on exposure to the gluten. CD affects 0.5-2.0% of the human population (Rewers 2005). In CD patients, CD4+ T cells are present in the lamina propria that secrete interferon-gamma upon recognition of gluten-derived peptides bound to HLA-DQ2 or HLA-DQ8 molecules present on antigen presenting cells. Strikingly, most of the gluten peptides implicated in CD require modification by the enzyme transglutaminase before they can bind to the disease-predisposing HLA-DQ molecules and trigger T cell responses (Koning 2003). In addition to the adaptive CD4+ T cell response to gluten, CD is characterized by upregulation of IL-15, an intraepithelial T cell infiltrate expressing the NKG2D receptor, and the overexpression of a ligand for NKG2D (MICA) (Meresse et al. 2004).

A variety of studies have been conducted to identify the gluten peptides responsible for CD. These peptides have been identified in α-, ω-, and γ-gliadins (van de Wal et al. 1998; Arentz-Hansen et al. 2002; van Herpen et al. 2006) as well as in HMW and LMW glutenins (Kooy-Winkelaar et al. 2011; Bodd et al. 2012). The α-gliadins are considered to be the most relevant components of gluten causing CD (Vader et al. 2003; van Herpen et al. 2006; Vaccino et al. 2009). Alpha-gliadin genes on chromosome 6A in bread wheat almost invariably contain epitopes glia-α9 and glia-α20, but not the intact epitopes glia-α or glia-α2. Genes on chromosome 6B do not contain...
any of the four T cell epitopes, and genes on chromosome 6D contain all of these T cell epitopes in variable combinations (van Herpen et al. 2006). Therefore, α-gliadins produced by chromosomes 6A and 6D could be the major targets for reducing CD through genetic improvement of wheat.

The identification of gluten peptides involved in CD is not yet complete. A large variation was observed for the amounts of CD4 T cell stimulatory peptides present in α- and γ-gliadins and glutenins among diploid, tetraploid, and hexaploid wheat accessions (Molberg et al. 2005; Spaenij-Dekking et al. 2005; Mitea et al. 2010). If such differences are genetically determined, this would provide strategies for selection and breeding of wheat cultivars suitable for consumption by CD patients. van Herpen et al. (2006) analyzed several wheat species with A-, B- and D-genomes and showed that epitopes are distributed non-randomly. Complete sequence analysis of toxic epitopes in T. monococcum led to the conclusion that it is very difficult to breed wheats lacking toxic epitopes in their storage proteins (Vaccino et al. 2009). In fact, it seems hardly feasible to create new genotypes lacking all 17 harmful peptides, present in proteins produced by different loci. Likewise, silencing of the genes giving rise to immunostimulatory sequences via targeted mutagenesis (Vader et al. 2003) also appears somewhat unrealistic.

The number of T cell-stimulatory epitopes (glia- α2/α9) found in Ae. tauschii were more comparable to A and B genome species (Xie et al. 2010). A number of α-gliadin genes from S+ (Ae. speltoides) and B genome species did not contain any of the 4T cell stimulatory epitopes (Spaenij-Dekking et al. 2005). Additionally, large variations in glutamine residues in the two polyglutamine domains are also present in α-gliadin genes.
from *Ae. tauschii*, suggesting a useful resource for reducing gliadin toxicity levels in wheat breeding programs. Ongoing work on wheat genetic resources and genomics aims to identify specific variations that might be helpful in developing functional markers for incorporation of less toxic epitopes (Xie et al. 2010). Li et al. (2012, 2013) identified specific SNPs and InDels in the four T-cell stimulatory toxic epitopes and the two polyglutamine domains which could be used as effective molecular markers for screening α-gliadin genes with no toxic epitopes. These variations could be of importance for consideration in the transfer of potentially valuable genes from *Aegilops* species to bread wheat.

### 5.2 Baker’s asthma

Baker’s asthma is an occupational hypersensitivity disease that results from the inhalation of flour and dust during grain processing. It has been recognized since Roman times when slaves who handled flour and dough were required to wear masks, and was described by Ramazzini in about 1700. Bakers’ asthma is now recognized as one of the most common types of occupational asthma. For example, it is recognized as the second most common type by the UK Health and Safety Executive (www.hse.gov.uk/asthma/bakers.htm). Earlier studies indicated that multiple allergens were present, with the water-soluble albumins being particularly reactive with the IgE fractions from patients’ sera and it is now well established that grain α-amylase inhibitors (CM proteins) are the major proteins responsible for baker’s asthma as reviewed by Tatham and Shewry (2008). A few reports suggest that wheat prolamin is partially responsible for this disease (Mittag et al. 2004). Sandiford et al. (1997) used immunoblotting to identify IgE against a range of wheat gluten proteins. All of the patients’ sera tested, showed IgE binding to α-
gliadins, 21 of 24 to fast \(\omega\)-gliadin, 18 of 24 to \(\beta\)-gliadin, 12 of 24 to \(\gamma\)-gliadin and 15 of 24 to slow \(\omega\)-gliadin. Competitive binding studies with salt-soluble proteins also showed reduced binding to purified \(\alpha\)- and \(\gamma\)-gliadins (but not to \(\beta\)- or \(\omega\)-gliadins), indicating the presence of common epitopes. Western blotting of fractions from total gliadins, glutenins and albumins-1-globulins showed reactions with IgE in 15 patients with baker’s asthma. Nine sera reacted with albumins-1-globulins, two with gliadins and seven with glutenins (Mittag et al. 2004). Therefore, more focused studies will be needed to dissect the role of wheat glutenins and gliadins from the other proteins to better devise wheat improvement strategies, following the same route as for CD.

5.3 Wheat-dependent exercise-induced anaphylaxis

Wheat-dependent exercise-induced anaphylaxis (WDEIA) is a form of allergic reaction jointly induced by the ingestion of a causative food and subsequent physical exercise. The first case of exercise-induced anaphylaxis (EIA) was reported by Sheffer and Austen (1980). Wheat is the most widely reported food to be associated with EIA. Patients with WDEIA display a range of clinical symptoms from generalized urticaria to severe allergic reactions including anaphylaxis.

\(\alpha\)-gliadin protein detected by immunoblotting had the major allergen domain (Palouso et al. 1999), but Morita et al. (2003) subsequently established that \(\omega\)5-gliadins (Tri a 19) were the major allergen in WDEIA. Later, synthetic peptides were used to identify seven important epitopes (Table 1) within the primary sequence of \(\omega\)5-gliadin (Matsuo et al. 2004, 2005). Four (QQIPQQQ, QQFPQQQ, QQSPEQQ and QQSPQQQ) of seven epitopes were dominant. Due to the non-availability of full-length sequences,
they searched a wheat EST library for ω5-gliadin-like sequences and identified four
sequences with GenBank accession numbers BE590637, BQ608902, BQ895830 and
BQ245835. Later, Matsuo et al. (2005) identified four further IgE-binding epitope
sequences, QQFHQQQ, QSPEQQQ, YQQYPQQ and QQPPQQ, in three patients with
WDEIA using a recombinant ω-gliadin. Mutational analysis of the QQIPQQQ and
QQFPQQQ peptides indicated that the amino acids at positions Gln 1, Pro 4, Gln 5, Gln 6
and Gln 7 were critical for IgE binding. A similar finding was reported by Battais et al.
(2005) on two immune-dominant epitopes (QQQLPQQQ and QQFPQQQ) on ω5-
gliadin.

Another important finding was identification of allergen domains in HMW-GS that
also reacted with IgE from patients with WDEIA (Morita et al. 2003). This was
confirmed by Matsuo et al. (2005) through establishment of the IgE-binding epitopes of
wheat HMW-GS using synthetic peptides. They also identified three epitopes (QQPGQ,
QQPGQQGQQ and QQSGQQGQ) from the repetitive domain. Genome-level studies on the
variability of these allergens in wheat and related species are still needed in order to
bridge a significant gap in our current knowledge-base.

6. Future challenges

Improvement of processing qualities for various products and reduction of the adverse
health effect of storage proteins are becoming more important. We need to develop more
sophisticated tools and to upgrade the resolution of current technologies in order to move
forward. Although progress has been made in this context, HMW-GS at Glu-1 and
LMW-GS at Glu-3 explain only about 60% of the variability in dough properties and
end-use qualities. Therefore, the identification of other major QTLs or loci is needed for a better understanding of wheat processing qualities. For example, the presence of a further class of storage proteins, named Avenin-like proteins identified by Kan et al. (2006), could also contribute to the functional properties of wheat flour. The international wheat genome sequencing consortium (IWGSC) efforts and genome wide association studies (GWAS) with modern marker technologies such as genotyping by sequencing and SNPs will be crucially important. Completion of the wheat DNA sequencing project will greatly enhance our capacities in gene discovery, gene cloning and marker development. Similarly, research efforts on classes of storage proteins such as lunasin (Mitchell et al. 2013) and triticin (Shailaja et al. 2002) need to be extended to validate their effects and to determine how they might be manipulated for grain quality improvement. For processing quality, quality consistency over locations and seasons is equally important to quality performance per se. It is expected that proteomics will play a leading role in understanding quality consistency and variation in different environments.

Apart from recent achievements on certain aspects related to wheat storage proteins, there are many challenging tasks that need to be covered for a better understanding and achievement of applied outputs. The first task is to resolve remaining problems on the identification of ambiguous glutenin alleles. Different diagnostic techniques, difficulties in sharing standard germplasm, differences in growing conditions to evaluate quality contributions by specific alleles and linkage/epistatic effects are the main problems in this regard. A proposal is underway to form an International Gluten Working Group to resolve such problems. Liu et al. (2010) provide an excellent example of what can be achieved through collaboration by providing a common platform for...
resolving LMW-GS related issues. Such types of integrated activity with more
coordination will be helpful in addressing other difficulties. Genetic resources also need
to be improved to facilitate the dissection of quality effects of individual glutenin and
gliadin alleles. The Aroona-derived NILs and durum cultivar Langdon derived D-genome
synthetic hexaploids (Xu et al. 2010) have proven to be the ideal genetic resources for
studies targeting the properties of particular protein alleles.

Studies on T cell stimulatory epitopes in all types of gluten proteins are relatively
new, and given that the role of the innate immune system is only beginning to be
understood, it may be premature to start breeding non-toxic wheat cultivars or cultivars
expressing lower levels of toxic polypeptides. However, there is a need to develop
functional markers for the most toxic epitopes, as these can be exploited in selection for
less toxic epitopes in wheat breeding programs.
Table 1. Allergen domains in wheat seed storage proteins

<table>
<thead>
<tr>
<th>Protein</th>
<th>Epitope sequence</th>
<th>Toxic</th>
<th>Genome assigned</th>
<th>Allergy</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>HMW-GS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DQ8-glut-H1</td>
<td>QGYYPYTPSQ</td>
<td>CD</td>
<td></td>
<td></td>
<td>van de Wal et al. 1998</td>
</tr>
<tr>
<td>Peptide-1</td>
<td>QQPGQ</td>
<td>A, B, D</td>
<td></td>
<td>WDEIA</td>
<td>Matsu et al. 2005</td>
</tr>
<tr>
<td>Peptide-2</td>
<td>QQPGQGQ</td>
<td>A, B, D</td>
<td></td>
<td>WDEIA</td>
<td>Matsu et al. 2005</td>
</tr>
<tr>
<td>Peptide-3</td>
<td>QQSGQGQ</td>
<td>A, B, D</td>
<td></td>
<td>WDEIA</td>
<td>Matsu et al. 2005</td>
</tr>
<tr>
<td>LMW-GS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>glutenin-17</td>
<td>PFSEQEQQPV</td>
<td>CD</td>
<td></td>
<td></td>
<td>Stepniak et al. 2005; Vader et al. 2003</td>
</tr>
<tr>
<td>glutenin-156</td>
<td>FSQQQESP</td>
<td>CD</td>
<td></td>
<td></td>
<td>Stepniak et al. 2005; Vader et al. 2003</td>
</tr>
<tr>
<td>α-gliadins</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>glia-α</td>
<td>QGSFQPSQQ</td>
<td>Yes</td>
<td>A, B, D, S, U, M, C, N</td>
<td>CD</td>
<td>van Herpen et al. 2006</td>
</tr>
<tr>
<td>glia-α2</td>
<td>PQPQLPYPPQ</td>
<td>Yes</td>
<td>M, U, C, N</td>
<td>CD</td>
<td>Li et al. 2012</td>
</tr>
<tr>
<td>glia-α9</td>
<td>FFPQPQLPY</td>
<td>Yes</td>
<td>A, B, D</td>
<td>CD</td>
<td>Xie et al. 2010</td>
</tr>
<tr>
<td>glia-α20</td>
<td>FRPQPYPQ</td>
<td>Yes</td>
<td>A</td>
<td>CD</td>
<td>Vaccino et al. 2009</td>
</tr>
<tr>
<td>glia-γ1</td>
<td>PQSFPEQQQ</td>
<td>CD</td>
<td></td>
<td></td>
<td>Sjöström et al. 1998</td>
</tr>
<tr>
<td>glia-γ2</td>
<td>QPEQPAQL</td>
<td>CD</td>
<td></td>
<td></td>
<td>Qiao et al. 2005; Vader et al. 2002</td>
</tr>
<tr>
<td>glia-γ3</td>
<td>QPEQPYPPQ</td>
<td>CD</td>
<td></td>
<td></td>
<td>Arentz-Hansen et al. 2002</td>
</tr>
<tr>
<td>glia-γ4a</td>
<td>SQPSEQFPQ</td>
<td>CD</td>
<td></td>
<td></td>
<td>Arentz-Hansen et al. 2002</td>
</tr>
<tr>
<td>glia-γ4b</td>
<td>PQPSEQFPQ</td>
<td>CD</td>
<td></td>
<td></td>
<td>Arentz-Hansen et al. 2002</td>
</tr>
<tr>
<td>glia-γ4c</td>
<td>QPOQFPFPQ</td>
<td>CD</td>
<td></td>
<td></td>
<td>Qiao et al. 2005</td>
</tr>
<tr>
<td>glia-γ4d</td>
<td>QPOQFPFCQ</td>
<td>CD</td>
<td></td>
<td></td>
<td>Qiao, unpublished</td>
</tr>
<tr>
<td>glia-γ5</td>
<td>QPFPFPFPQ</td>
<td>CD</td>
<td></td>
<td></td>
<td>Arentz-Hansen et al. 2002</td>
</tr>
<tr>
<td>Peptide</td>
<td>Enzyme</td>
<td>Reference</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------</td>
<td>--------</td>
<td>-----------------</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>QQIPQQQ</td>
<td>WDEIA</td>
<td>Matsuo et al. 2004</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>QQFPQQQ</td>
<td>WDEIA</td>
<td>Matsuo et al. 2004</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>QQSPEQQ</td>
<td>WDEIA</td>
<td>Matsuo et al. 2004</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>QQSPQQQ</td>
<td>WDEIA</td>
<td>Matsuo et al. 2004</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>QQLPQQQ</td>
<td>WDEIA</td>
<td>Matsuo et al. 2004</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>QQYPQQQ</td>
<td>WDEIA</td>
<td>Matsuo et al. 2004</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PYPP</td>
<td>WDEIA</td>
<td>Matsuo et al. 2004</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>QQFHQQQ</td>
<td>WDEIA</td>
<td>Matsuo et al. 2005</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>QSPEQQQ</td>
<td>WDEIA</td>
<td>Matsuo et al. 2005</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>YQQYPQQ</td>
<td>WDEIA</td>
<td>Matsuo et al. 2005</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>QQPPQQ</td>
<td>WDEIA</td>
<td>Matsuo et al. 2005</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>QQQLPQQQ</td>
<td>WDEIA</td>
<td>Battais et al. 2005</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>QQQFPQQQ</td>
<td>WDEIA</td>
<td>Battais et al. 2005</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 1. Strategy for cloning and functional analysis of novel glutenin and gliadin genes
Figure 2. Breeding implications of glutenin alleles for specific wheat-based products; Y-AXIS represents *Glu-1* alleles and X-AXIS represents *Glu-3* allele effects on dough strength at different grain protein levels (modified from Ikeda and Takata 2012)
Acknowledgements

The authors are grateful to Prof. R. A. McIntosh at Plant Breeding Institute, University of Sydney, and Dr. Xiaofei Zhang at University of Minnesota, for critical reviews of this manuscript. This study was supported by International Collaboration Projects from the Ministry of Science and Technology (2013DFG30530) and China Agriculture Research System (CARS-3-1-3). We also acknowledge the Higher Education Commission (HEC) in Pakistan for providing an overseas fellowship to the first author under the International Research Support Initiative Program (IRSIP).
References


Butow, B.J., Ma, W., Gale, K.R., Cornish, G.B., Rampling, L., Larroque, O., Morell, M.K., Bekes, F., 2003. Molecular discrimination of Bx7 alleles demonstrates that a highly expressed high-molecular-weight glutenin allele has a major impact on wheat flour dough strength. Theoretical and Applied Genetics 107, 1524–1532.


Halford, N.G., Forde, J., Anderson, O.D., Greene, F.C., Shewry, P.R., 1987. The nucleotide and deduced amino acid sequence of an HMW glutenin subunit gene from chromosome 1B of bread wheat (Triticum aestivum L.) and comparison with those of genes from chromosomes 1A and 1D. Theoretical and Applied Genetics 75, 117–126.

encoded by chromosome 1A of bread wheat (*Triticum aestivum* L.) indicates quantitative effects on grain quality. Theoretical and Applied Genetics 83, 373-378


dimensional electrophoresis and the chromosomal localisation of their controlling

Jackson, E.A., Morel, M.H., Sontag Strohm, T., Branlard, G., Metakovsky, E.V.,
Redaelli, R., 1996. Proposal for combining the classification systems of alleles of Gli-
1 and Glu-3 loci in bread wheat (Triticum aestivum L.). Journal of Genetics and

Jia, J.Z., Zhao, S.C., Kong, X.Y., Li, Y.R., Zhao, G.Y., He, W.M., Appels, R., Pfeifer,
M., Tao, Y., Zhang, X.Y., Jing, R.L., Zhang, C., Ma, Y.Z., Gao, L.F., Gao, C.,
Spannagl, M., Mayer, F.X., Li, D., Pan, S.K., Zheng, F.Y., Hu, Q., Xia, X.C., Li,
Middleton, C., Quan, Z.W., Liu, G.M., Wang, J., Yang, H.M., Liu, X., He, Z.H.,

Jiang, C.X., Pei, Y.H., Zhang, Y.Z., Li, X.H., Yao, D.N., Yan, Y.M., Hsam, S.L.K.,
glutenin subunit genes from Aegilops longissima, Triticum dicoccoides and T.

Jin, H., Yan, J., Pena, R.J., Xia, X.C., Morgounov, A., Han, L.M., Zhang, Y., He, Z.H.,
2011. Molecular detection of high- and low-molecular weight glutenin subunit genes
in common wheat cultivars from 20 countries using allele-specific markers. Crop and
Pasture Science 62,746–754.


Lei, Z.S., Gale, K.R., He, Z.H., Gianibelli, C., Larroque, O., Xia, X.C., Butow, B.J., Ma, W., 2006. Y-type gene specific markers for enhanced discrimination of high-


SDS-PAGE, 2-DE, MALDI-TOF-MS and PCR in common wheat. BMC Plant Biology 10, 124.


Ma, W., Anderson, O., Kuchel, H., Bonnardeaux, Y., Collins, H., Morell, M.K.,
Muehlbauer, G.J. (Eds.), Genetics and Genomics of the Triticeae, Springer New
York, pp, 611–652.


Margiotta, B., Urbano, M., Colaprico, G., Johansson, E., Buonocore, F., D’Ovidio, R.,
Lafiandra, D., 1996. Detection of y-type subunit at the Glu-A1 locus in some Swedish

glutelin subunits: N-terminal sequences and evidence for the presence of cysteine.
Cereal Chemistry 70, 581–585.

Matsuo, H., Kohno, K., Morita, E., 2004. Molecular cloning, recombinant expression and
IgE-binding epitope of omega-5 gliadin, a major allergen in wheat-dependent

Matsuo, H., Kohno, K., Niihara, H., Morita, E., 2005. Specific IgE determination to
epitope peptides of ω-5 gliadin and high molecular weight glutenin subunit is a useful
tool for diagnosis of wheat-dependent exercise-induced anaphylaxis. The Journal of
Immunology 175, 8116–8122.


Pistón, F., Gil-Humanes, J., Rodríguez-Quijano, M., Barro, F., 2011. Down-regulating γ-gliadins in bread wheat leads to non-specific increases in other gluten proteins and has no major effect on dough gluten strength. PloS One 6, e24754.


Shewry, P.R., Gilbert, S., Savage, A., Tatham, A., Wan, Y.F., Belton, P., Wellner, N., D’ovidio, R., Bekes, F., Halford, N., 2003b. Sequence and properties of HMW subunit 1Bx20 from pasta wheat (Triticum durum) which is associated with poor end use properties. Theoretical and Applied Genetics 106, 744-750.


related to the magnitude and breadth of gluten-specific T cell responses. Proceedings of National Academy of Sciences USA 100, 12390–12395.


Figure legends:

**Figure 1.** Strategy for cloning and functional analysis of novel glutenin and gliadin genes

**Figure 2.** Breeding implications of glutenin alleles for specific wheat-based products; Y-AXIS represents *Glu-1* alleles and X-AXIS represents *Glu-3* allele effects on dough strength at different grain protein levels (modified from Ikeda and Takata 2012)
Highlights

- 30 allele-specific markers for characterization of glutenins have been developed.
- More than 100 novel alleles of storage proteins have been identified from wild species.
- Quality effects of individual glutenin alleles have been well documented.
- α-gliadin genes on chromosome 6A and 6D will be useful to develop less toxic cultivars.
- Future prospects of genomics and proteomics technologies are discussed.