Journal of Cereal Science 60 (2014) 11–24

Contents lists available at ScienceDirect

Journal of Cereal Science

journal homepage: www.elsevier.com/locate/jcs

Review

Wheat seed storage proteins: Advances in molecular genetics, diversity and breeding applications

Awais Rasheed [a,b], Xianchun Xia [a], Yueming Yan [c], Rudi Appels [d], Tariq Mahmood [b], Zhonghu He [a,e,*]

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

ARTICLE INFO

Article history:
Received 12 November 2013
Received in revised form
28 January 2014
Accepted 29 January 2014
Available online 12 March 2014

Keywords:
Glutenins
Gliadins
Breadmaking quality
Coeliac disease

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.

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,

Abbreviations: HMW-GS, high molecular weight glutenin subunit; LMW-GS, low molecular weight glutenin subunit; SDS-PAGE, sodium dodecyl sulfate polyacrylamide 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.

* Corresponding author. Institute of Crop Science, National Wheat Improvement Center, Chinese Academy of Agricultural Sciences (CAAS), 12 Zhongguancun South Street, Beijing 100081, China. Tel./fax: +86 10 82108547.
E-mail address: zhhecaas@163.com (Z. He).

http://dx.doi.org/10.1016/j.jcs.2014.01.020
0733-5210/

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. Juhász 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 *Glu-B1* and 29 alleles at *Glu-D1* have been reported in bread wheat (McIntosh et al., 2013). The AACC glutenin allele database contains the glutenin compositions of over 8500 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_{max}) and extensibility based on both HMW-GS and LMW-GS alleles (Bekes and Wrigley, 2013).

Genetic diversity in the *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 (*Triticum dicoccoides* and *Triticum turgidum* subsp. *dicoccon*) express the *Glu-Ay* gene (Waines and Payne, 1987), and a Swedish bread wheat cultivar was also found to express a *Glu-Ay* gene (Margiotta et al., 1996). This could be very useful for expanding the narrow allelic diversity at the *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 *Triticum urartu* and *Triticum 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 *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 *Glu-D^{t}1* loci from *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., 2008). 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^l(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 tri-peptide 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