ECRONICON OPEN ACCESS

EC NUTRITION
Research Article

Gluten Alleles and Predicted Dietary Intolerances: Use of a Database of Wheats World-Wide

F Bekes and C W Wrigley*

[1]*FBFD Pty Ltd, Beecroft, NSW, Australia*

[2]*Honorary Professor, Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, St Lucia QLD 4072, Australia*

***Corresponding Author**: C W Wrigley, Honorary Professor, Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, St Lucia QLD 4072, Australia.

Received: June 25, 2017; **Published:** July 25, 2017

Abstract

The gluten composition of over 8,500 wheat varieties world-wide has been catalogued and are available on the web site of AACC International. This catalogue is primarily of value to wheat breeders selecting new varieties having specific grain-quality attributes. In addition, this resource is now potentially available for use in attempts to select wheat genotypes that differ with respect to dietary intolerances, making use of recent knowledge about the relevant aspects of gluten-protein composition and amino-acid sequences.

Keywords: *Gluten Intolerance; Celiac Disease; IgE-Mediated Hypersensitivity; Alleles; Wheat; Genotypes*

Abbreviations

IgE: Immunoglobulin E; G: Genotype: E: Growth Environment: *Glu*: Allele Designation for Genes Coding for Polypeptides of Wheat Glutenin; *Gli*: Allele Designation for Genes coding for Wheat Gliadin Proteins; PSS: Protein Scoring System; LMW-GS: Low-Molecular-Weight Glutenin Subunits; HMW-GS: High-Molecular-Weight Glutenin Subunits

Introduction

Wheat genotypes (varieties) differ considerably in their grain-protein composition and correspondingly in the alleles responsible for the coding for the grains' storage proteins – proteins that become gluten after milling the grain into flour, mixing the flour with water to form a dough, for eventual baking into bread. Wheat genotypes also differ considerably in the respective suitability of their grain for processing into the wide diversity of wheat-based foods in our diet [1,2]. Many of these differences in grain quality relate to variations in grain-protein composition, and thus to the genetic alleles of the wheat genome. The gluten composition of over 8,500 wheat varieties world-wide has been catalogued and are available on the web site of AACC International [3,4], as described below. This catalogue is of prime value to wheat breeders aiming to select new varieties having specific quality attributes. In addition, this resource is available to attempt the selection of wheat genotypes that differ with respect to dietary intolerances, making use of recent knowledge about the relevant aspects of protein composition and amino-acid sequences.

Beneficial outcomes via wheat breeding

The unique visco-elastic properties of wheat dough are largely determined by the amount and composition of the gluten proteins. To a more limited extent, the composition of the gluten proteins also contributes to the various health-related issues caused by wheat gluten in the diet. These issues include celiac disease and other inflammatory conditions, IgE-mediated hypersensitivities and various forms of food intolerance [4].

The amino-acid sequences of gluten proteins that are responsible for triggering responses in sensitive individuals have been identified, showing that they vary in distribution among and between different groups of gluten proteins [2]. The detection of and, in particular, the quantification of gluten proteins, are critical not only due to their direct effects on end-use quality but also for food intolerance reasons.

Grain composition varies between cereal genotypes, thus leading to opportunities in food-allergen research, so that beneficial outcomes with new varieties may be achieved via genotype selection in wheat-breeding programs. These 'beneficial outcomes' have traditionally related to improvements in grain yield and in wheat-processing quality, but improvements with respect to dietary intolerance are also feasible.

Nevertheless, it is difficult to accurately identify the specific proteins that trigger health problems and to elucidate their genotypic frequency, variability and stability. Reasons for this difficulty include the high sequence similarity and multi-species origin of the prolamins in gluten, coupled with limitations in the available methodologies, as reviewed by Haraszi., *et al* [6].

Allergen prediction for specific wheat varieties

High-resolution methods such as mass spectrometry (MS) require accurate quantitative relationships between prolamin peptide biomarkers and the final gluten/prolamin content to relate the detection of peptide mass to their protein sources. These quantitative relationships, however, are difficult to establish due to genotypic and environmental variability.

To assist peptide biomarker searches, epitope mapping, protein selection and medical studies, a database (ProPepper, https://propepper.net) has been developed to contain members of the prolamin superfamily proteins identified from Poaceae species, peptides obtained with single and multi-enzyme *in silico* digestion as well as linear epitopes responsible for wheat-related food disorders [7]. Knowledge of the composition and amounts of epitopes present in a single wheat sample represents a significant gap and the complete grain proteome datasets now available can provide the necessary information to carry out an estimation of allergen prediction for a single cultivar.

Thus, there is now the significant possibility of coupling these two databases, namely, the ProPepper system for allergen prediction together with the listing of gluten-protein alleles for a world-wide collection of thousands of wheat varieties.

Gluten-proteins and wheat-processing qualities

Significantly different dough qualities are required for each of the various products made from wheat flour, such as bread (various types), pasta, pastry, cookies and many other products. The compositions of the gliadin and glutenin fractions of gluten (and thus dough properties) are determined by the interaction of the genetic constitution (G) of the wheat variety involved and its growth environment (E).

On the genotype (variety) side of this G X E interaction, the genes of major importance are the glutenin polypeptides (subunits) which combine by disulfide bonds to form the large aggregates of glutenin macromolecules [8]. The elasticity conferred by the polymeric glutenin is balanced by the gliadin fraction of smaller molecular weight [9].

World-wide wheats and their gluten alleles

The allelic composition of glutenin and gliadin proteins of a wheat sample largely determines its end-use properties. It is the reason that, beyond its key role in fundamental research, the identification of the allelic composition of a sample is essential information in wheat breeding, in the baking industry and possibly also to breed for dietary objectives.

Access to information about the glutenin and gliadin alleles of the world's many wheat varieties has been difficult because it is spread throughout the literature and in private listings. It was the reason that a comprehensive list of allelic data of wheat cultivars bred and cultivated all around the world was collected and prepared for publication as an Appendix in a book published in 2006 by AACC International, under the title "Gliadin and Glutenin: The Unique Balance of Wheat Quality", edited by C W Wrigley, F Békés and W Bushuk [9]. When the large size of this Appendix precluded its inclusion within the covers of a book, the internet permitted its publication.

The AACCI book is still a good source of information about the chemistry, genetics and nomenclature of the three classes of gluten proteins, namely the gliadins and the high- and low-molecular-weight glutenin subunits (HMW-GS and LMW-GS). The gluten-protein database, provided on the AACCI web site, serves as a key collection of allelic data for thousands of cultivars in three separate sections: Gliadins, HMW-GS and LMW-GS. This database continues to be the main website for cereal chemists requiring gluten-protein alleles. Other sources include similar or complementary information: a part of the GrainGenes website [10] and the website of "Genes On-Line" [11].

Gluten-allele catalogue – Version 4

Version 4 of the database, is available on the AACCI web site (http://www.aaccnet.org/initiatives/definitions/Pages/Gluten.aspx). It includes a large number of wheat genotypes (over 8,500). It is a completely re-edited, extended version of the previous three versions in a new format: instead of the simple listing of information in a PDF file, Version 4 is a user-friendly database in which the user can search, screen and utilize specific applications of the glutenin allelic composition such as:

- To assess the genetic potential of a specific variety for processing;
- To determine how readily a specific combination of varieties could be distinguished according to gliadin or glutenin-subunit composition;
- To determine the value of specific genotypes as parent lines to achieve targeted genetic potential for dough quality, applying the well-established Payne-score [12] and/or the Protein Scoring System model of Békés [13].
- To pursue population genetics.
- To explore possibilities of breeding for tolerance to dietary problems.

Version 4 contains two sets of databases (the glutenin and gliadin databases) and related software tools. Both databases contain two text files and an executable program file, downloadable to the user's PC from the AACCI website (http://www.aaccnet.org/initiatives/definitions/Pages/Gluten.aspx). For the convenience of future users, both the glutenin and gliadin program files were written in two alternative versions; one of them is suitable for use with Windows XP and the Windows Vista environment, while the other is for Windows 7 and Windows 8 users.

The Glutenin Database

The glutenin database covers data for over 8,500 cultivars from 80 countries containing 23 alleles of *Glu-A1*, 87 of *Glu-B1*, 25 of *Glu-D1*, 11 of *Glu-A3*, 13 of *Glu-B3* and 12 alleles of *Glu-D3*, derived from 220 publications.

The user can search/screen in the database, using either the name of a cultivar (variety), its country of origin, or allele(s) of any combination of the six glutenin loci. The original source of information for the searched cultivar is available as well as the predicted genetic potential for dough properties, characterized by the Payne score or the Protein Scoring System (PSS) (Figure 1).

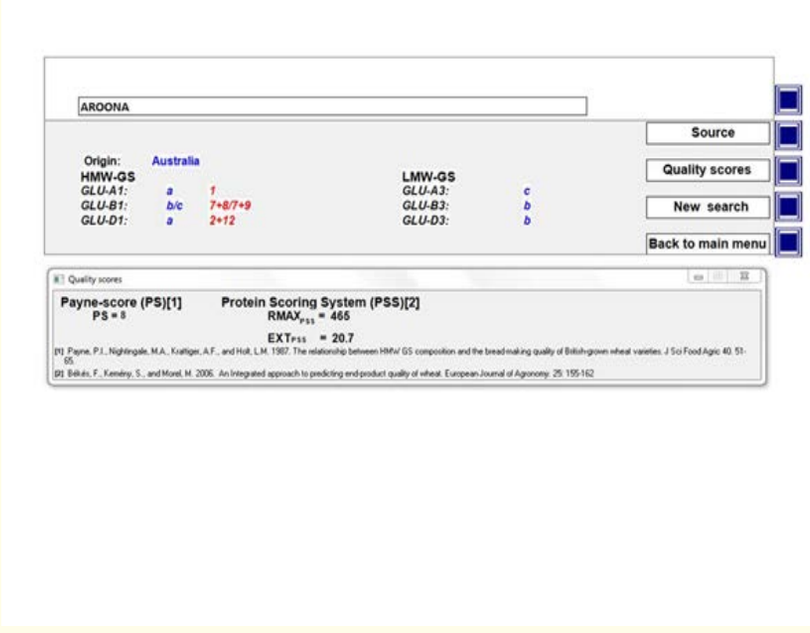

Figure 1: *Result of an enquiry from the glutenin database for the glutenin subunit/allele composition of the variety Aroona.*

Background to the Glutenin Alleles: HMW-GS are coded at the *Glu-1* locus on the long arm of Chromosome 1, while LMW-GS are coded at the *Glu-3* locus on the short arm of Chromosome 1. There are thus six sets of glutenin alleles: *Glu-A1, Glu-B1, Glu-D1* and *Glu-A3, Glu-B3, Glu-D3*. The durum wheats have only the two genomes, A and B; so, there are four sets of glutenin alleles for durum wheats.

The LMW subunits have been more difficult to analyze than the HMW subunits because the bands of the LMW-GS are normally disguised in an SDS gel electrophoretic pattern by the presence of overlapping gliadin proteins. As a result, fewer lists of LMW-subunit composition are reported in the literature. Nevertheless, it has proved possible to accumulate the LMW-subunit composition for many genotypes using electrophoretic techniques, starting with the pioneering studies of Gupta and Jackson [14-16]. The number of publications providing data about the LMW-GS alleles has increased significantly by the application of MALDI-TOF technology in recent years.

Difficulties in Establishing Allelic Composition: Inconsistencies have arisen between published lists, due to differences in the methods used in obtaining composition data and the significant difficulties in interpreting the results of protein fractionation. Early means of identifying protein composition, e.g., gel-electrophoresis, lacked the resolution of more recent methods, e.g., two-dimensional electrophoresis, HPLC, MALDI-TOF and DNA-based molecular-marker techniques.

A fine-tuning of the designation systems for both HMW-GS and LMW-GS alleles is in progress. Good examples for these activities are the excellent work of Espi., *et al.* [17] who clarified the differences among the *Glu-B1b, Glu-B1u* and *Glu-B1*al alleles (i.e., among the Bx7, Bx7* and Bx7^{OE} subunits by molecular methods). International collaborations are proceeding on the systematic comparative analysis of LMW-GS alleles using traditional methods and mass spectroscopy, as well as molecular-marker techniques [18].

The allelic compositions of the wheats listed are provided in the form of the distinct allelic codes (as "*a*", "*b*", … etc.) for the two glutenin loci (*Glu-1* for HMW-GS and *Glu-3* for LMW-GS) for each of the three (A, B and D) or two (A and B) genomes of hexaploid or durum wheat, respectively. In cases of novel alleles when internationally agreed codes are not yet given, we have specified the designation given to the allele in the relevant publication. In the case of HMW-GS alleles, the corresponding protein components (subunits or polypeptides) coded by the allele are also listed. Missing data is indicated as "N.D.". Durum wheats and lines containing the 1B/1R translocation are specifically indicated.

Based on the early electrophoresis-based designations of HMW-GS and LMW-GS alleles, a set of cultivars was selected to act as "standards" for the various subunit mobilities. The list of these cultivars has been extended in Version 4 to include the recently discovered alleles, citing the publication where they were described.

The Glutenin Database

Gliadins are coded at the *Gli-1* and *Gli-2* loci on the short arms of Chromosomes 1 and 6, respectively. There are thus six sets of gliadin alleles: *Gli-A1, Gli-B1, Gli-D1* and *Gli-A2, Gli-B2* and *Gli-D2*. The durum wheats have only the two genomes, A and B; so, there are four sets of gliadin alleles for durum wheats.

The database list of gliadin composition does not include a complete range of reports from the literature even though the gliadins have been more extensively characterized over a longer period of time than the glutenin subunits. Such information is available via the extensive literature references in Chapters 2 to 4 of the AACCI book 'Gliadin and Glutenin: The Unique Balance of Wheat Quality' [19,20]. An obvious difficulty in comparing these diverse sources of information is that there has not been a generally agreed system of nomenclature for gliadin composition for much of the time of that gliadin research has been conducted. A unifying nomenclature, now most generally adopted for gliadins, involves the allocation of alleles for the six loci at *Gli-1* and *Gli-2* for the three genomes (A, B and D).

Data listed in the gliadin database is primarily due to the efforts of one scientist (Dr Metakovsky), using a consistent approach to fractionation and to interpretation [21] using every effort to ensure the authenticity of the samples analyzed. Data of 961 varieties from 31 countries are listed in the database, containing 25 alleles of *Gli-A1,* 24 of *Gli-B1*, 15 of *Gli-D1*, 33 of *Gli-A2*, 52 of *Gli-B2* and 33 alleles of *Gli-D2*. This list has not been published previously in its entirety but parts have been published. Cultivar names originally published in Cyrillic were transformed into Latin lettering in accordance to the rules described by Martynov., *et al* [22]. The gliadin alleles listed were all determined on the basis of the same procedure of acidic gel electrophoresis [23]. As with the glutenin database, the user can search/screen the database, based either on the name of cultivar, on its origin, or on allele(s) of any combination of the six gliadin loci.

The collection of related research data on gliadins produced in the last ten years - produced by other separation/identification techniques such as capillary electrophoresis, reversed-phase HPLC and mass spectroscopy - is in progress and will be available in Version 5 of this database.

Similarly, to the glutenin database, the allelic compositions of the wheats listed are provided in the form of the distinct allelic codes (as "a", "b", etc.) for the two gliadin loci, namely, *Gli-1* and *Gli-2* for each of the three (A, B and D) or two (A and B) genomes of hexaploid or durum wheat, respectively. Missing data is indicated as "N.D.". Durum wheats are specifically indicated.

Based on the early electrophoresis-based designation of gliadin alleles, a set of standard cultivars was selected to serve as "standards" for the various polypeptide mobilities. The list of these cultivars has been extended in Version 4 to include recently discovered alleles, referring to the publication where each was described. So, suitable "standards cultivars" are provided for each of the 182 gliadin alleles appearing in the database.

Biotypes and Contaminants

A difficulty in compiling such a database is the likelihood of polymorphisms within a variety, i.e., more than one version of the allele mix. This may be due to admixtures and contaminants, but more likely they are due to incomplete selection of a pure line. Thus, different grains of one authentic sample may differ in their gluten composition, each grain's composition being an authentic part of the cultivar's constitution.

Such "biotypes" may arise from the original cross that produced the variety under study, being sister lines with alternative alleles that have become fixed in the process of selecting the cultivar. In such cases, the tables indicate that alternative alleles have been detected, e.g., in the form "*a/b*".

On the other hand, such variations may be due to contamination with foreign seed, thus providing wrong information. In such a case, the authenticity of the sample source should be examined. Examination of the composition of the parent lines (if definitely known) may help in establishing that the apparent polymorphism is not due to contamination, but rather that valid biotypes are present. In such cases, where the real biotypes have segregated (such as the work of Lawrence [24] on several Australian varieties containing biotypes), the database contains the data both for a sibling line (e.g., AROONA) and for the pure germplasm biotypes (AROONA A and AROONA B).

Estimation of Dough Quality

Compared to the previous versions of this database, Version 4 has the completely new feature of providing tools to estimate the genetic potential of dough properties for the listed lines based on their glutenin alleles.

Relating protein composition to certain quality traits by statistical means is a frequently used methodology to relate structure-composition to functionality in cereal science [3]. Since its development [12,25], the Payne Score has provided a single-number parameter to estimate the dough strength of a sample as a function of its HMW-GS allelic composition; at the time, this score revolutionized selection for quality in wheat breeding.

Since the significant success of using the Payne Score in breeding programs as a single number to estimate dough strength from HMW-GS allelic composition, there have been several attempts to also involve the LMW-glutenin alleles in similar mathematical formulas [13,26-28]. Using sophisticated statistical approaches, the Wheat Simulator [28], and the Protein Scoring System (PSS) [13] can describe the effects of both HMW- and LMW-GS alleles on dough strength and extensibility, both individually and considering pair-wise interactions among the alleles.

Applications of the models indicate that the approach of relating allelic composition to quality attributes is possible with careful data selection and applying robust mathematical tools. One of the most important conceptual findings from evaluating such models is the realization that, because of the large contribution of allele-allele interactions, the different allelic combinations (rather than the individual glutenin alleles) should be targeted in breeding situations to develop new lines with desired quality attributes predicted from the bread-making potential of flour samples.

The genetic potential of a line, with a certain combination of alleles at the six glutenin-coding loci, can meaningfully be predicted where both the contribution of the individual alleles and their pair-wise interactions play equally important roles [27,29,30].

Further applications of the models relate to the milling industry for predicting the bread-making potential of blended mill streams by modelling the non-linear characteristics of quality parameters [31,32].

On-going Development of the Databases

The databases are "open" for new data, for corrections and for general suggestions. Please contact the authors at firinc47@gmail.com or c.wrigley1@uq.edu.au. Such contributions will be built into the program and acknowledged. We hope that the database and its search facilities will continue to be useful in achieving its aims of providing the tools for all sectors of the industry to improve wheat-quality attributes.

Conclusions

The combined use of the AACCI gluten allele, genome sequence and allergen databases, prediction methodology, and cereal chemistry results in a better understanding of the level of toxicity present in the end-products from wheat flour. The workflow presented in the

review of Juhász., *et al.* [33] provides information about the number and distribution of epitopes for a single protein or a protein fraction. Also, epitopes present in the highest frequency and for the most harmful proteins can be identified. The "epitope toxicity" value obtained in this way is a significant research output from the analysis of large datasets that can be applied to the food industry.

Acknowledgements

This article originally appeared in Cereal Foods World (CFW) and was revised and adapted for republication here with the permission of AACC International http://aaccipublications.aaccnet.org/journal/cfw.

Bibliography

1. Juhász A., *et al*. "Wheat proteins". Applied Food Protein Chemistry, Ed. Z. Ustunol, Wiley-Blackwell, Hoboken, NJ, USA (2013): 219-304.

2. Shewry PR and Tatham AS. "Improving wheat to remove celiac epitopes but retain functionality". *Journal of Cereal Science* 67 (2016): 12-21.

3. Bekes F and Wrigley CW. "Gluten alleles and predicted dough-quality for wheats world-wide: A great resource, free on the AACCI web site". *Cereal Foods World* 58.6 (2013): 325-328.

4. Bekes F and Wrigley CW. "Gluten Alleles and Predicted Dough Quality for Wheat Varieties Worldwide. Web-published gluten protein database" (2013).

5. Sapone A., *et al*. "Spectrum of gluten-related disorders: Consensus on new nomenclature and classification". *BMC Medicine* 10 (2012): 13.

6. Haraszi R., *et al*. "Analytical methods for detection of gluten in food—method developments in support to the legislations on labelling of foodstuffs". *Journal of AOAC International* 94.4 (2011): 1006-1025.

7. Juhász A., *et al*. "ProPepper: a curated database for identification and analysis of peptide and immune-responsive epitope composition of cereal grain protein families". *Database* (2015): bav100.

8. Weiser H., *et al*. "The polymeric glutenins". Gliadin and glutenin. The unique balance of wheat quality. Eds. C.W. Wrigley, F. Békés and W. Bushuk, AACCI Press, St Paul, Min., USA (2006): 213-240.

9. Wrigley CW., *et al*. "Gluten: A balance of gliadin and glutenin". Gliadin and glutenin. The unique balance of wheat quality. Eds. C.W. Wrigley, F. Békés and W. Bushuk, AACCI Press, St Paul, Min., USA (2006): 3-32.

10. Graybosch R. "Endosperm storage proteins. 3. Glutenins" (1994).

11. Martinov SP and Faberova I. "Pedigree and identified alleles of genes" (2006).

12. Payne PI. "Genetics of wheat storage proteins and effect of allelic variation on bread-making quality". *Annual Review of Plant Physiology and Plant Molecular Biology* 38 (1987): 141-153.

13. Békés F., *et al*. "An Integrated approach to predicting end-product quality of wheat". *European Journal of Agronomy* 25.2 (2006): 155-162.

14. Gupta RB. "Genetic control of LMW glutenin subunits in bread wheat and association with physical dough properties". PhD Thesis. University of Adelaide, Glen Osmond, South Australia (1989).

15. Gupta RB and Shepherd KW. "Two-step one-dimensional SDS-PAGE analysis of LMW subunits of glutenin. I. Variation and genetic control of the subunits in hexaploid wheats". *Theoretical and Applied Genetics* 80.1 (1990): 65-74.

16. Jackson EA., *et al*. "Proposal for combining the classification systems of alleles of Gli-1 and Glu-3 loci in bread wheat". *Journal of Genetics and Plant Breeding* 50 (1996): 321-336.

17. Espi A., *et al*. "A PCR-based method for discriminating between high molecular weight glutenin subunits Bx7 and Bx7* in Triticum aestivum L". *Plant Breeding* 131.5 (2012): 571-573.

18. Liu L., *et al*. "Comparison of low molecular weight glutenin subunits identified by SDS-PAGE, 2-DE, MALDI-TOF-MS and PCR in common wheat". *BMC Plant Biology* 10 (2010): 124.

19. Metakovsky EV., *et al*. "Gliadins of common wheat: Polymorphism and genetics". Gliadin and glutenin. The unique balance of wheat quality. Eds. C W Wrigley, F Békés and W Bushuk, AACCI Press, St Paul, Min., USA, (2006): 35-84.

20. Branlard GP and Metakovsky EV. "Some Gli alleles related to common wheat dough quality". Gliadin and glutenin. The unique balance of wheat quality. Eds. CW Wrigley, F Békés and W Bushuk. AACCI Press, St Paul, Min., USA (2006): 115-140.

21. Metakovsky EV. " Gliadin allele identification in common wheat. II. Catalogue of gliadin alleles in common wheat". *Journal of Genetics Breeding* 45 (1991): 325-344.

22. Martinov SP., *et al*. "Catalogue genealogies and gene alleles identified in 31,000 cultivars and lines of wheat". Research Institute of Crop Protection, Prague, Czech Republic (1992).

23. Metakovsky EV and Novoselskaya AY. "Gliadin allele identification in common wheat. 1. Methodological aspects of the analysis of gliadin pattern by one-dimensional polyacrylamide gel electrophoresis". *Journal of Genetics Breeding* 45 (1991): 317-324.

24. Lawrence GJ. "The high-molecular-weight glutenin subunit composition of Australian wheat cultivars". *Australian Journal of Agricultural Research* 37 (1986): 125-133.

25. Payne PI., *et al*. "Relationship between HMW GS composition and measures of the bread-making quality of wheat varieties grown in Spain". *Journal of Cereal Science* 7 (1988): 229-236.

26. Gupta RB., et al. "Prediction of physical dough properties from glutenin subunit composition in bread wheats: correlation studies". Cereal Chemistry 68 (1991): 328-333.

27. Cornish GB., *et al*. "Prediction of dough properties for bread wheats". Gliadin and Glutenin. The Unique Balance of Wheat Quality. Eds. C.W. Wrigley, F. Békés and W. Bushuk. AACCI Press: St. Paul, MN (2006): 243-280.

28. Eagles HA., *et al*. "Estimation and utilisation of glutenin gene effects from the analysis of unbalanced data from wheat breeding programs". *Australian Journal of Agricultural Research* 53 (2002): 1047-1057.

29. Branlard G., *et al*. "Allelic diversity of HMW and LMW GS and omega-gliadins in French bread wheat". *Genetic Resources and Crop Evolution* 50.7 (2003): 669-679.

30. Baracskai I., *et al*. "A Retrospective Analysis of HMW and LMW Glutenin Alleles". *Cereal Research Communications* 39 (2011): 226-237.

31. Békés F. "New aspects in quality related wheat research: II: New methodologies for better quality wheat (Review)". *Cereal Research Communications* 40 (2012): 307-333.

32. Békés F. "Studying the protein-protein interactions and functional properties of the wheat storage proteins in a gluten-free model system". *Agro Food Industry Hi Tech* 22 (2012): 1-7.

33. Juhász A., *et al*. "The epitopes in wheat proteins for defining toxic units relevant to human health". *Functional and Integrative Genomics* 12.4 (2012): 585-598.

Volume 10 Issue 2 July 2017