UNIVERSITÀ DI PARMA

UNIVERSITY of PARMA

PhD in Biotechnology and Life Science

XXX Cycle

How the molecular analysis helps to understand the domestication and benefits of wheat: characterization of the glutenin subunit composition in a core durum wheat landraces collection.

Coordinator: Prof. Simone Ottonello

Supervisor: Prof. Nelson Marmiroli

Tutor: Dott.ssa Michela Janni

Candidate: Stefania Cadonici

2014/2018

Index

Chapter 1. General introduction 5
1.1 Wheat 5
1.1.1 The importance of wheat 5
1.1.2 Wheat origin, genome and domestication 6
1.1.2.1 Wheat domestication 7
1.1.3 Genetic resources 10
1.1.3.1 Exploitation of genetic resources 11
1.1.3.2 Single Seed Descent (SSD) collection 13
1.1.4 Wheat caryopsis anatomy and composition 15
1.1.4.1 Starch 16
1.1.4.2 Lipids 17
1.1.4.3 Proteins 17
1.2 The Gluten 18
1.2.1 Gluten and its viscoelastic properties 18
1.2.2 Gliadins 20
1.2.3 Glutenins 24
1.2.3.1 Low Molecular Weight Glutenin Subunits (LMW-GS) 25
1.2.3.2 High Molecular Weight Glutenin Subunits (HMW-GS) 27
1.3 Durum wheat quality 30
1.3.1 General concepts 30
1.3.2 Genetic determination of durum wheat quality 31
1.4 Wheat adverse reactions 34
1.4.1 Autoimmune disorders 35
1.4.1.1 Celiac Disease 35
1.4.1.2 Gluten Ataxia (GA) 36
1.4.1.3 Dermatitis herpetiformis (DH) 36

1.4.2 Allergic disorders....37
1.4.2.1 Wheat allergy....37
1.4.3 Not autoimmune and not allergic disorders....38
1.4.3.1 Gluten Sensitivity (GS)....38
Chapter 2. Aim of the project....43
Chapter 3. Materials and Methods....45
3.1 Plant material....45
3.2 HMW-GS protein extraction....45
3.3 Crude protein extract quantification....46
3.4 Electrophoresis....47
3.4.1 SDS-PAGE analysis....47
3.5 Genomic DNA extraction....49
3.5.1 Dellaporta DNA extraction....49
3.5.2 GenElute™ Plant Genomic DNA Miniprep Kit....50
3.5.2.1 DNA quantification....52
3.6 Molecular markers assay for *Glu-A1* and *Glu-B1*....52
3.6.1 PCR conditions and electrophoresis analysis....53
3.7 Purification and sequencing....54
3.7.1 DNA purification from gel....54
3.7.2 DNA purification from PCR solution....55
3.7.3 Sequencing....55
3.8 Statistical analysis....55
3.9 *In vitro* digestion of selected SSD sample....56
3.9.1 Standardized static *in vitro* digestion method....56
3.9.2 Ultra-Performance Liquid Chromatography–Electrospray Ionization Mass Spectrometry (UPLC/ESI-MS) analysis....57
3.9.3 Data processing....58
3.10 Statistical analysis....58

Chapter 4. Results and Discussions 59
4.1 HMW-GS characterization of standard cultivars 59
4.1.1 Proteomic approach 60
4.1.2 PCR marker-based approach 63
4.1.2.1 Analysis with PP1 65
4.1.2.2 Analysis with PP2 65
4.1.2.3 Analysis with PP3 67
4.1.2.4 Analysis with PP4 69
4.1.2.5 Analysis with PP5 70
4.1.2.6 Analysis with PP6 72
4.1.3 Discussions 73
4.2 HMW-GS characterization of SSD genotypes 74
4.2.1 Proteomic approach 76
4.2.2 PCR marker-based approach 80
4.2.2.1 Glu-1 allele diversity 81
4.2.3 Geographic distribution of *Glu-1* alleles 90
4.2.4 Gene-ecology analyses 102
4.2.5 Discussions 105
4.3 *In vitro* digestion SSD genotypes 107
4.3.1 Discussions 114
Chapter 5. Conclusions and future perspectives 116
References 118
Web Sites 139
Acknowledgments 140
List of publications 141

Chapter 1. General introduction

1.1 Wheat

1.1.1 The importance of wheat

Wheat (*Triticum* spp.) is, among cereals, the one of the most important food crop worldwide in terms of production and utilization as nourishment, being second only to rice which provides 21% of the total food calories and 20% of the proteins for more than 4.5 billion people in 94 developing countries (Goutam et al., 2013). According to FAO's estimate wheat had a world production of 715 million Tons in 2013 (FAOstat 2014, http://faostat3.fao.org); out of these, approximately 38 million Tons are represented by the durum wheat production (Kabbaj et al., 2017). As a matter of fact, nowadays, about 95% of the wheat grown worldwide is hexaploid bread wheat (*Triticum aestivum* L.), with most of the remaining 5% being tetraploid durum wheat (*Triticum durum* Desf.; Varzakas et al., 2014, Shewry et al., 2009). The extensive cultivation of wheat is due to several factors; it is the best adapted crop to temperate regions, unlike rice and maize which prefer tropical environments (Gill et al., 2004), it provides high yields and it possesses good nutritional value and technological properties. Wheat is, in fact, one of the major source of carbohydrates (its kernel consists for 80% of starch), minerals and vitamins for humans (Onipe et al., 2015; Koehler and Wieser, 2013), providing about one-fifth of the total food calories and proteins; moreover, it contains soluble and insoluble fiber which are known to exert a healthy role in human nutrition. Furthermore, the importance, both nutritional and, consequently, economic, of this cereal is due to the capability of wheat dough to be processed in a wide range of products, like pasta, bakery products and other wheat-derived goods, thanks to its unique characteristic underpinned by gluten. The Food and Agriculture Organization (FAO) has estimated that to meet the global food demand in 2050, annual world production of crops and livestock will need to be 60 percent higher than it was in 2006 (Elbehri and Food and Agriculture Organization of the United Nations, 2015). About 80 percent of the required increase will need to come from higher yields on an area of land that will not increase much beyond the present level (Gill et al., 2004) and 10 percent from increases in the number of cropping seasons per year (The State of Food and Agriculture: climate change, agriculture and food security, http://www.fao.org/3/a-i6030e.pdf). Hence, new wheat varieties with an increased yield and higher green management are needed. Furthermore, modern food industries require flours with high technological characteristics, like high gluten strength, thus making industrial and nutritional quality main goals for researchers, wheat breeders and farmers alike (Carver, 2009).

1.1.2 Wheat origin, genome and domestication

From a taxonomic point of view, cultivated wheats and their close wild relatives belong to the genus *Triticum* L. of the family *Gramineae* and are members of the tribe *Triticeae*. This is one of the largest and most important tribes in the grass *Poaceae* family, which includes major crop plants such as wheat (*Triticum* spp. L.), oat (*Avena sativa* L.), barley (*Hordeum vulgare* L.), rye (*Secale cereale* L.), maize (*Zea mays* L.) and rice (*Oryza sativa* L.; Dewey, 1984). *Triticeae* tribe contains about 300 species which are characterized by a basic chromosome number x = 7, comprising diploid (2n=2x=14), tetraploid (2n=4x= 28) and hexaploid (2n=6x=42) wheats. Wheat genus *Triticum*, whose centre of origin and differentiation is located in the Middle Est and Transcaucasus region, consists of six species: *Triticum monococcum* L. (AA genome); *Triticum urartu* Tumanian ex Gandilyan (AA genome); *Triticum turgidum* L. (AABB genome); *Triticum timopheevii* (Zhuk.) Zhuk (AAGG genome); *Triticum aestivum* L. (AABBDD genome) and *Triticum zhukovskyi* Menabde & Ericz (AAAAGG genome). Nowadays, the commercial wheats are mainly hexaploid bread wheat, *T. aestivum* (AABBDD genome), and tetraploid durum wheat, *T. durum* (AABB genome), thus containing three and two closely related homeologous genomes, respectively (Peng et al., 2011).

Despite rice, maize and wheat coevolved from a common ancestor about 55-75 million years ago (Kellogg, 2001), they differ greatly in genome size, with *T. aestivum* having the largest genome among agricultural crops; it has a general size of 17 Gb, with each subgenome being approximately 5.5 Gb: 8-fold larger than the one of maize and about 40-fold larger than the one of rice (Gill et al., 2004; Shi and Ling, 2017).

Tetraploid form of current cultivated wheats, whose genome has a general size of 12 Gb, are derived from the spontaneous hybridization between the diploid wild wheat *T. urartu* (AA genome) and an undefined diploid *Aegilops speltoides* species (BB genome), which generated the wild emmer wheat *T. dicoccum* (AABB genome), as shown in **Fig. 1** (Kissing Kucek et al., 2015).

Hexaploid *T. aestivum*, or bread wheat (AABBDD genome; **Fig. 1**) which is the most widely cultivated nowadays, is thought to derive from a spontaneous hybridization of *T. dicoccum* (AABB genome) with *Ae. tauschii* (DD genome), which occurred no more than 8,000 years ago.

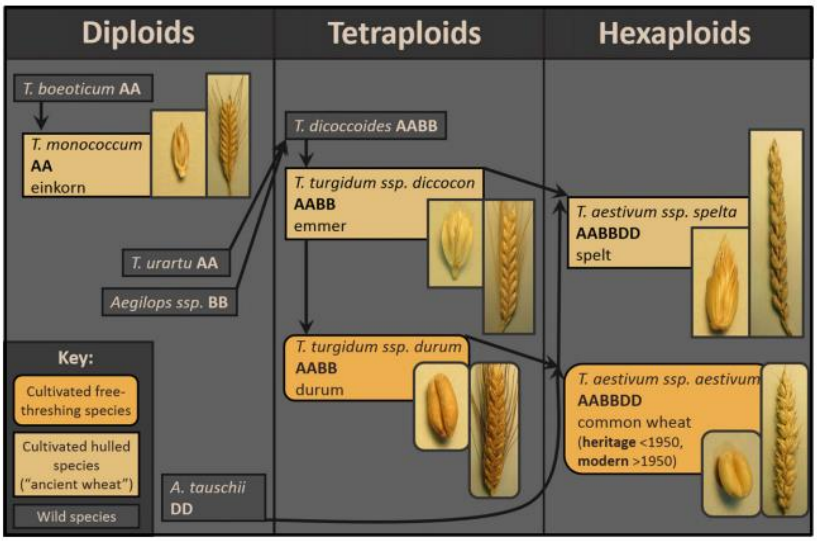

Figure 1: The genealogy of cultivated members of the Triticum family, including various cultivated ancient wheat species, durum wheat, and common wheat (Kissing Kucek et al., 2015).

1.1.2.1 Wheat domestication

The domestication of tetraploid wheat took place about 12000 years ago in the Fertile Crescent, where ancient farmers selected among cultivated forms of wild emmer (*Triticum turgidum* ssp. *dicoccoides*), a naked type that was easier to thresh (*Triticum turgidum* ssp. *Dicoccum*) (Kabbaj et al., 2017). This process was the prerequisite for the evolution of tetraploid wheat (*Triticum turgidum* ssp. *durum*) which, in turn, is considered to be the ancestral form of modern pasta wheat (Jaradat, 2013). The transition from both emmer and spelt to modern durum and bread wheat, respectively, has been associated with the selection of genetic traits that separated them from their wild relatives. In particular, the domestication of wheat has involved, among all, two specific traits. The first one was the development of non-shuttering seed at maturity, which prevented natural seed dispersal and allowed humans to harvest and collect the seed with optimal timing (Jaradat, 2013). This characteristic is important to ensure seed dispersal in natural population and the non-shattering trait is determined by mutations at the *Br* (*brittle rachis*) locus (Nalam et al., 2006; Shewry et al., 2009). The second one was the change from hulled forms, in which the glumes adhere tightly to the grain, to free-threshing naked forms, which arose by a dominant mutant at the Q locus which modified the effects of recessive mutations at the *Tg* (*tenacious glume*) locus (Shewry et al., 2009). The early domesticated forms of einkorn, emmer and spelt were all hulled and not free-threshing, whereas modern forms of tetraploid and hexaploid wheat are free-threshing (**Fig.2**). In addition, size and shape of the wheat grain are independently inherited traits and the process of domestication resulted in a switch from production of a relatively small grain with a long, thin shape to a more uniform larger grain with a short, wide shape (Jaradat, 2013).

Tetraploid wheats are genetically and morphologically diverse and their evolution under domestication has not been fully elucidated. Genetic and archaeological evidence indicates that cultivated emmer evolved from the tetraploid wild emmer in the Fertile Crescent around the 8th millennium BCE (Matsuoka, 2011; Salamini et al., 2002) (**Fig. 2**). Over time, human migration and the spread of agriculture from the Fertile Crescent to and throughout Europe and Asia led to the expansion of the cultivation of naked emmer (Kabbaj et al., 2017). Brandolini et al. (2016) in their DArT (Diversity Array Technology)-Seq- genotyping of 136 einkorn landraces, nine wild einkorn and three *Triticum urartu* reported of a possible spread of agriculture along the trail from Turkey to Bulgaria/Greece via former Yugoslavia to Hungary. The further diffusion of early agriculture from Hungary into central Europe was connected to the Linearbandkeramik culture, as indicated by human ancient DNA profiles from Mesolithic, Neolithic Starcevo and Linearbandkeramik sites (Szécsényi-Nagy et al., 2015). Indeed, Kreuzet al. (2005), working with archaeobotanical remains, showed that the Linearbandkeramik was focussed on einkorn and emmer, and that einkorn was the dominant cereal. A second main path of the spread of agriculture evidenced by Brandolini et al. (2016) followed the Mediterranean coast reaching, finally, Maghreb and Iberia. This last evidence corroborated the conclusions drawn by Oliveira et al. (2011) who genotyped 50 einkorns with a total of 21 nuclear- and chloroplast-derived microsatellites arguing that the history of the Western Mediterranean einkorns (from the Iberian Peninsula and Morocco) was different from the rest of Europe. As a matter of fact, the earliest domesticated wheats were spread, probably via trading by Phoenician merchants, from the Fertile Crescent region, particularly southeastern Turkey, to the West of Mediterranean Basin, where about 75% of the world's durum wheat is produced (Alsaleh et al., 2015), reaching then the Iberian Peninsula (Feldman, 2001).

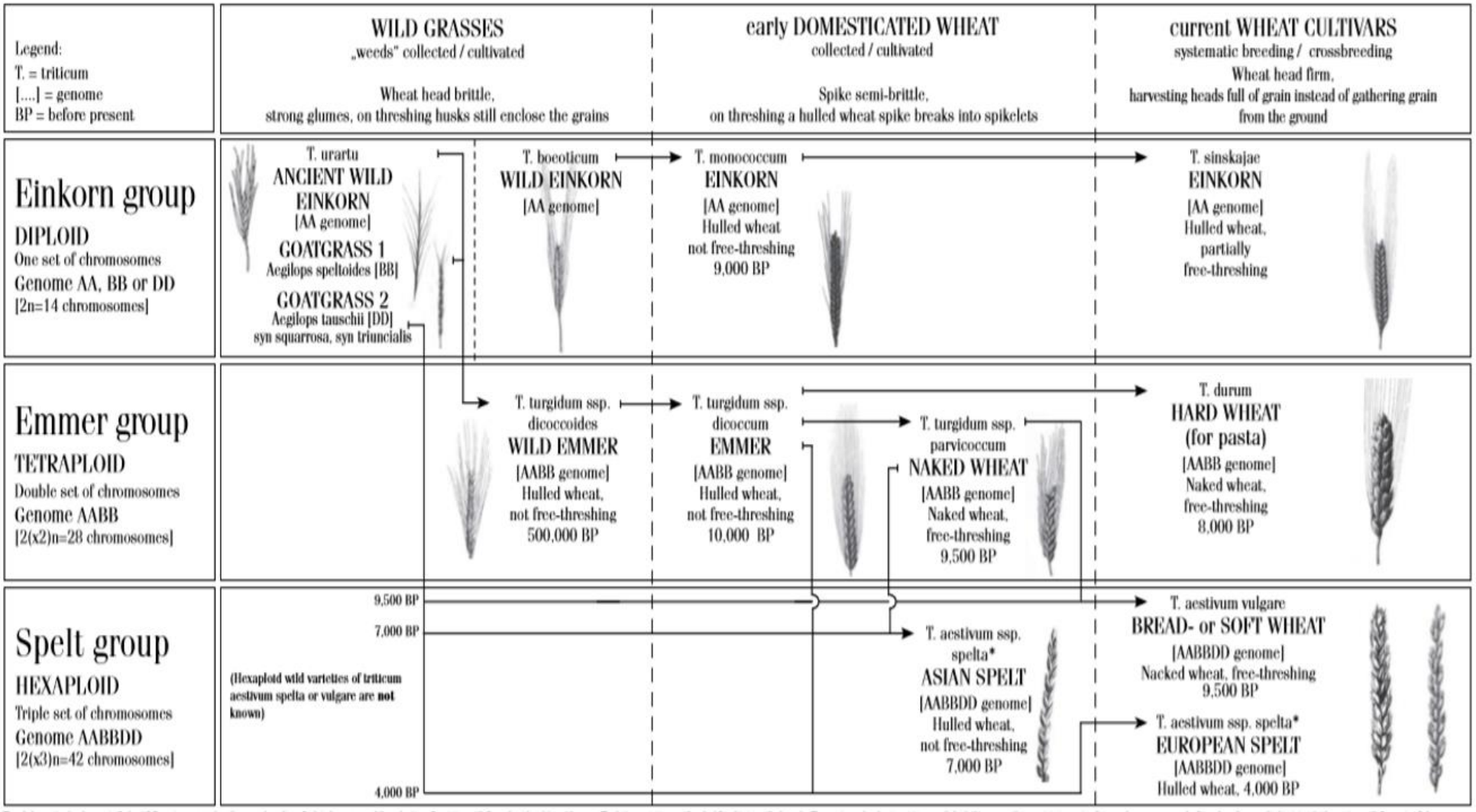

Figure 2: Phylogenetic development of wheat: from wild relatives to current cultivated durum and bread wheats (Wheat Compendium, 2016).

1.1.3 Genetic resources

Plant genetic resources are the biological basis of food security and, directly or indirectly, support the livelihoods of every person on Earth. Plant genetic resources for food and agriculture (PGRFA) consist of diversity of seeds and planting material of traditional varieties and modern cultivars, crop wild relatives and other wild plant species (http://www.fao.org/agriculture/crops/core-themes/theme/seeds-pgr/en/).

Historically the Mediterranean Basin is the most important area of production of durum wheat, where it is usually cultivated under rainfed conditions. This type of climate often entails several environmental stresses on the crop, such as drought stress and a combination of water scarcity and warm temperature, which usually occur during the grain filling period and result in yield reduction but, in most cases, in high grain quality (Nazco et al., 2014).

Autochthonous wheat landraces are defined as traditional varieties, lacking formal crop improvement, with the following features:

- they have been developed by farmers through years of natural and human selection, hence they are associated with small-scale farming (Oliveira et al., 2012);
- they showed a high adaptability to biotic and abiotic stresses (Zeven, 1998). As a matter of fact, their genetic structure is an evolutionary approach to adapt and survive to specific agro-ecological environmental conditions and management practices (Zeven, 1999; Jaradat, 2013);
- they represent an intermediate stage in domestication between wild wheat and elite cultivars. Indeed, allelic variation of genes originally found in the wild, but gradually lost through domestication and breeding, have been recovered only by going back to landraces (Giraldo et al., 2016).

More in general, the term “landrace” referred to a cultivated, genetically heterogeneous variety that has evolved in a certain ecogeographical area and is therefore adapted to the edaphic and climatic conditions and to its traditional management and uses (Casañas et al., 2017).

The process of wild emmer domestication, and the subsequent breeding of domesticated durum and bread wheat, have led to a narrowed scenery of genetic diversity. Indeed, it was estimated that, upon domestication, the initial diversity was reduced by 84% in durum wheat and by 69% in bread wheat (Jaradat, 2013); moreover, with intense focus and overuse of elite plant materials, almost 60 000 plant species in general could be lost by 2025 (Moreta et al., 2015). This phenomenon is commonly known as “genetic erosion”. Nowadays a limited number of elite varieties of durum and common wheat is cultivated, while landraces have been relegated in a second place because of their lower commercial

value (Giraldo et al., 2010). Notwithstanding the main advantages of modern cultivars, such as a greater productivity and higher nutritive values (Siddique et al., 1989; Migliorini et al., 2016), the current yield grain trend is insufficient to meet the rising demand for feeding a world population of 9 billion by 2050 (Friedrich et al., 2014). Moreover, the same varieties are also characterized by a high genetic homogeneity which, in times, results in a greater vulnerability to new biotic and abiotic stresses, as well as to unpredictable environmental conditions (Wang et al., 2017). Consequently, interspecific hybridization between durum élite lines and wild relatives of the *Graminae* family is a promising approach to restore variability in the modern breeding cultivars (Zaïm et al., 2017). Thus, considering both the oncoming climate changes and the increasing of wheat-adverse reactions, it becomes of high priority to boost the rate of genetic improvement, especially as regard traits like tolerance to both biotic and abiotic stresses (Pignone and Hammer, 2013; Jaradat, 2013), technological quality and healthiness, and yielding rate.

1.1.3.1 Exploitation of genetic resources

Progenitors of modern cultivated species can still be found in the centres of origin and several efforts are concentrated on their protection and preservation in germplasm banks (Marmiroli et al., 1999). Most of the exotic material, including landraces, are preserved in gene banks as germplasm and constitute an easily transferable and valuable source of genetic variation for agronomical, morphological, adaptive and quality traits (Nachit, 1992; Moragues et al., 2006 and 2007; Nazco et al., 2014; Giraldo et al., 2016; Ceoloni et al., 2017; Ceoloni et al, 2014). Gene banks are, thus, responsible for developing and maintaining large numbers of collections to ensure the availability of genetic diversity of different species as a public good (Wang et al., 2017). Nowadays, the availability of high-throughput genotyping and phenotyping methods, which provide a large set of data, together with the use of interconnected information network between different academic areas such as plant breeding and biology, could help breeders to search for useful accessions to be used in pre-breeding programs. More precisely, pre-breeding refers to all activities designed to identify desirable characteristics and/or genes from un-adapted materials that cannot be used directly in breeding populations and to transfer these traits to an intermediate set of materials that breeders can use further in producing new varieties for farmers. It is a necessary first step in the use of diversity arising from wild relatives and other unimproved materials. Consequently, a comprehensive characterization of crop germplasm is critical to the optimal improvement of the quality and productivity of crops (Moragues et al., 2007).

The oncoming importance of unlocking the biodiversity of wheat genetic diversity from landraces to face current world's problems have been reported. Lopez et al. (2015) in their review questioned how the diversity can be exploited and made readily available for use by breeders and the scientific community, especially with respect to the field of adaptation to climate change. Sajjad et al. (2011) screened a panel of 500 genotypes of landraces, cultivars and breeding lines of bread wheat considering the grain yield. Top-crosses between wild relatives of durum wheat and elite lines of the ICARDA breeding programs have been made in order to assess their possible demerits for yield drags, to identify their biotic stress response, and to quantify the potentially negative effect of wild alleles on rheological quality (Zaïm et al., 2017) and another investigation was carried out by Konvalina et al. (2017) to evaluate the differences between minority wheat species and common wheat to determine the best rheological characteristics, technological quality as well as correlations between rheological and technological traits.

In this scenery, large-scale genome-wide association analyses is being recently exploited in many plant species thanks to the dramatic reduction in costs of genomic technologies. As a matter of fact, genome-wide association (GWAS) mapping has attracted much interest, mostly in hexaploid bread wheat, as a method complementary to traditional bi-parental mapping to identify novel loci responsible for traits of interest in global germplasm collections (Liu et al., 2017). As a new alternative to traditional linkage analysis, association mapping offers three advantages, (i) increased mapping resolution, (ii) reduced research time, and (iii) analysis of a higher number of alleles (Kushwaha et al., 20017). Indeed, phenotype-genotype relationships in existing germplasm collections or natural populations may be explored with GWAS for hundreds to thousands of accessions simultaneously, hence avoiding the resource intensive development of bi-parental genetic mapping populations (Liu et al., 2017). Association mapping harnesses the genetic diversity of natural populations to potentially resolve complex trait variation to single genes or individual nucleotides. The ability to map QTLs in collections of breeding lines, landraces, or samples from natural populations has great potential for future trait improvement and germplasm security (Zhu et al., 2008). Crossa et al. (2007) were the first to use GWAS investigation methods in wheat to identify rust-resistance genome signals; Sun et al. (2007) revealed new allelic variation distribution for 13 yield-related traits through GWAS studies in 163 bread wheat cultivars. Giraldo et al. (2016) performed association mapping for 18 agro-morphological and grain quality traits in a set of 183 Spanish landraces (*durum*, *turgidum* and *dicoccon*) genotyped with DArT (Diversity Array Technology) markers in order to investigate the genetic and phenotypic diversity, identify marker–trait associations and test the contribution of each subspecies to general diversity of the whole collection and their potential as a source of genetic variation for wheat improvement.

1.1.3.2 Single Seed Descent (SSD) collection

Regarding durum wheat genetic resources, the EURISCO database (http://eurisco.ecpgr.org) reports about 7,000 accessions of traditional varieties and/or landraces on a total of 17,257 accessions. Such a high number of genotypes for breeding studies is a limiting factor: consequently, new approaches are needed to overcome this constraint. A possible way, explored by Pignone and co-workers (2015) was the development of a subset of heterozygotic genotypes, representative of the overall variation, to better explore durum wheat germplasm. The aim was to reach a manageable number of near-homozygous *durum* wheat genotypes to perform in depth the characterization in order to make available to scientific community a set of characters well characterized by appropriate markers to be used in future studies. The tool was the selection through the single seed descent (SSD) approach (**Fig. 3**) (Pignone et al., 2015).

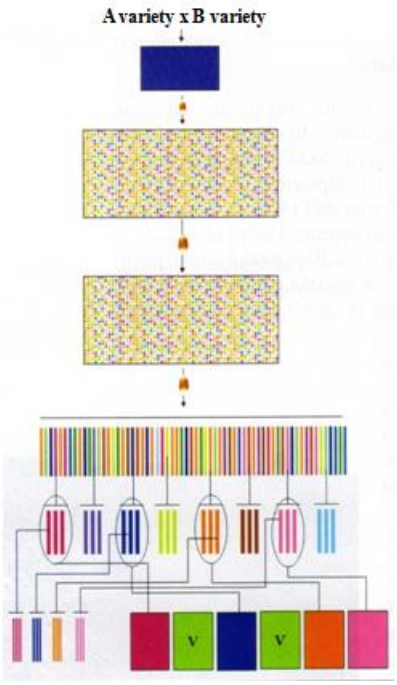

Figure 3: Single seed descent approach, which consists in having a single seed from each plant, bulking the individual seeds, and planting out the next generation. Only one (single) seed is collected from each of the F2 plants and bulked to grow the next (F3) generation. Similar practice is continued till F5 or F6 generation, when the plant would become nearly homozygous and when individual's plants are selected and harvested separately. (Adapted by Barcaccia and Falcinelli, 2012)

Single seed descent method is the modification of bulk breeding schemes, firstly suggested by Goulden (1940) and subsequently modified by Brim (1966), which allows to overcome the problem of natural selection and inadequate sampling of conventional breeding methods and, moreover, allows the early selection of homozygous genotypes. This approach proved to have several advantages such

as increase the breeding efficiency, reduce efforts and maximize the variation in the progeny from a cross between élite lines in its segregating generations (Snape, 1976).

The seed bank of the Institute of Bioscience and Bioresources of the National Research Council of Italy, holds a collection of some 27.000 samples of the genus *Triticum*, among which almost 5.600 are identified as *Triticum durum* (http://ibbr.cnr.it/mgd/). 500 genotypes, representative of 40 countries mainly concentrated in the Mediterranean areas and Horn Africa (**Fig. 4**), have been randomly selected for the constitution of the durum wheat SSD collection (henceforth called SSD collection, Pignone et al., 2015). From each accession one single seed was randomly selected. After planting and growing the derived plants, one seed was chosen from the principle spike of each S1 plant to produce the next generation and this procedure was repeated twice to obtain the S3 generation (**Fig. 3)**. The latter, characterized by high level of homozygosity and genetic stability and, at the same time, by high variability, was then grown under normal field conditions in order to avoid unintentional selection for specific characters (Pignone et al., 2015).

The SSD genotypes were subjected to large scale SNP screening by ILLUMINA Infinium iSelect 90k wheat chip to assess the heterozygosity and the genetic diversity present in the SSD population which was preliminary phenotyped following field growth. These analyses confirmed a high level of genetic variability within the SSD genotypes (Pignone et al., 2015).

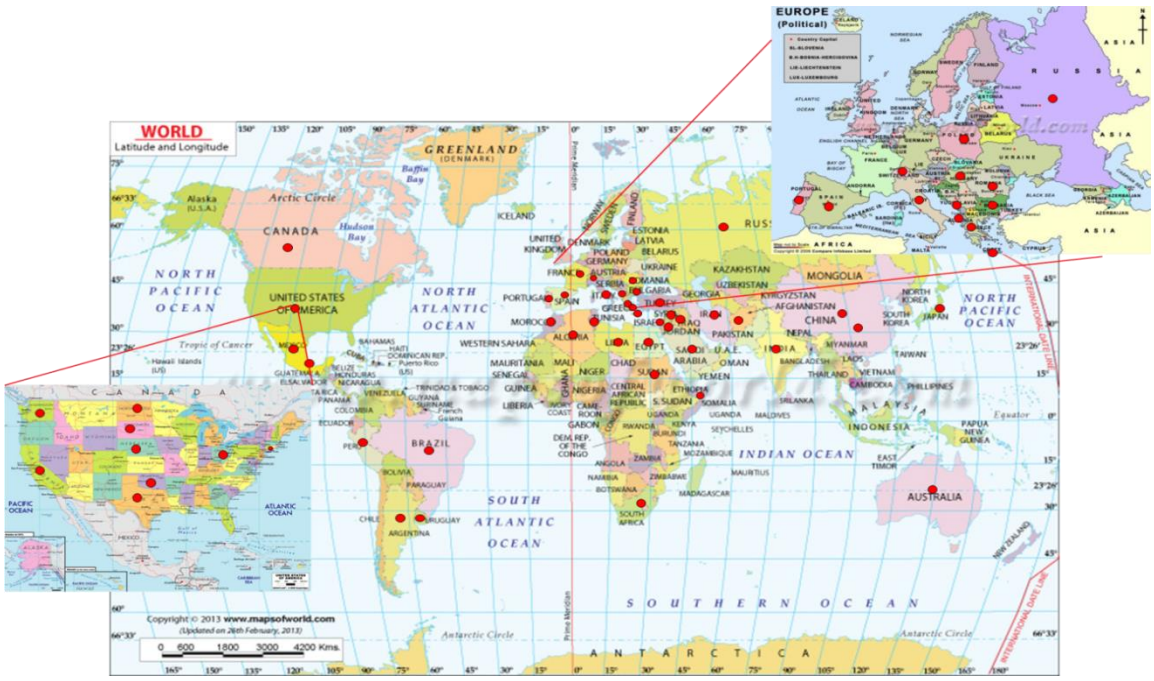

Figure 4: Country of origin of the 500 genotypes included in the SSD core collection (Pignone et al., 2016).

1.1.4 Wheat caryopsis anatomy and composition

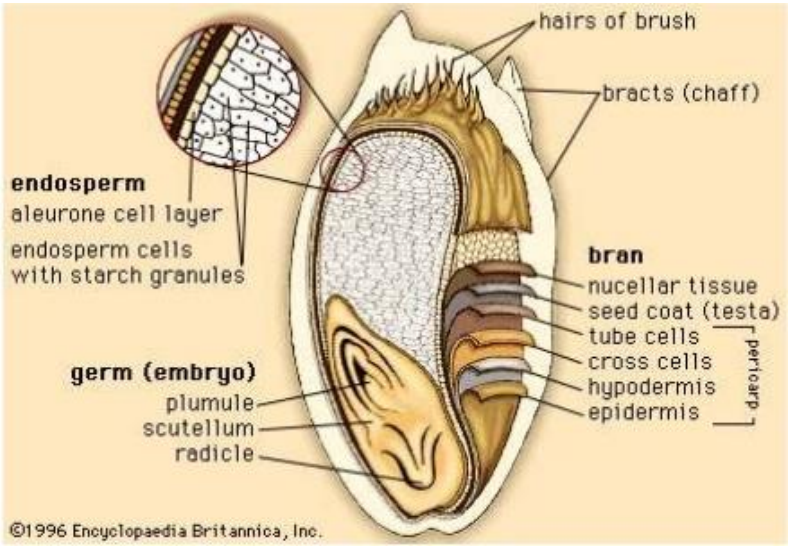

Figure 5: Structure of the durum wheat caryopsis (Šramková et al., 2009).

Wheat caryopsis consists of three distinct anatomical parts: bran, endosperm and germ or embryo.

As shown in **Fig. 5**, the bran is made up by several layers. Pericarp, which surrounds the whole kernel, is the outermost tissue and consists of about 5% of the kernel. The adjacent layer, which is fused with the pericarp, is the seed coat, also called testa. It develops from the integument of the ovule and envelops the endosperm. The aleurone layer, which belongs anatomically to the endosperm, is next to the seed coat and is firmly attached it thanks to the nucellar epidermis. The latter is also called hyaline layer. Of the three parts, bran is a good source of dietary fiber, primarily insoluble; moreover, it contains trace of minerals and a small amount of proteins (Causgrove et al., 2004).

The endosperm, which makes up about 85% of the wheat grain mass, consists of two tissues; starchy endosperm and aleurone. The outer layer is the aleurone; it is composed by cells which are cuboidal in shape and it is important for both the developing seed and the mature plant. This tissue, in fact, accumulates large quantity of oils and lipids that are useful during seed development (Evers and Millart, 2002), it is a site of mineral storage and it may express several pathogen-protective proteins and enzymes, such as amylase and protease, useful the hydrolysis and the provision of nutrients. Starchy endosperm, conversely, is composed by cells that are large, irregularly shaped and which contain starch granules. This tissue occurs as a solid mass occupying the centre of the grain and it is the component with the greatest value as it is a good source of carbohydrates, proteins, iron, vitamine B and soluble fiber (Causgrove et al., 2004).

The embryo, or germ, consists of about 3% of the caryopsis. At grain maturity, the embryo comprises an embryonic axis (shoot, mesocotyl and radicle) and scutellum, which is considered to be

homologous with a cotyledon (Evers and Millart, 2002). It is the most nutritious part of the kernel as it contains the highest concentration of lipids but also vitamin B, vitamin E and minerals (Causgrove et al., 2004).

The kernel is, in cereals, the final site of storage of the photosynthesis products and nitrogen metabolism: as a matter of fact, it contains a limited quantity of proteins (7-15%) and lipids (2-9%) and a high content of carbohydrates (65-75%), mainly as starch.

1.1.4.1 Starch

Starch is the main component of the wheat kernel's endosperm in which it is starch granules are embedded in a matrix of storage proteins (Evers and Millart, 2002). It is the most abundant storage carbohydrate in wheat, representing more than 70% of its dry weight, and it is also the most important polysaccharide in human diet (Lineback and Rasper, 1988; Sestili et al., 2010). Starch is a glucose polymer consisting of two carbohydrates chains, amylose and amylopectin (**Fig. 6**). Amylose is mainly characterized by linear molecules of D-glucose units joined by α-glycosidic bonds (1 → 4) with a polymerization degree of between 10^3 and 10^4 units, while amylopectin is a highly branched polymer of glucose with a α- (1 → 6) bonds frequencies of about 3-4% and with a higher polymerization degree, which is comprised between 10^5 and 10^7 units (Ball et al., 1998; Morell et al., 1995; Rahman et al., 2000; Sestili et al., 2010).

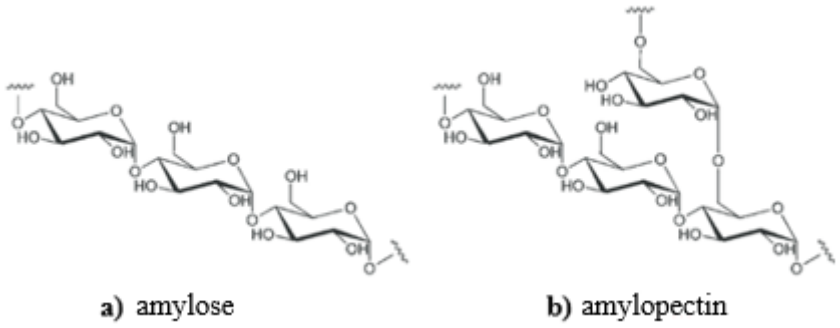

Figure 6: Molecular structure of starch polymer: a) amylose and b) amylopectin.

Usually amylose and amylopectin, within starch, occur in a ratio of 1:3 (Colonna and Buléon, 1992). However, amylose content can vary between species and between different genotypes within species, affecting firstly the amylose/amylopectin ratio and, consequently, also quality traits of wheat flours and their end uses (Zhang 2015, Ferrante, 2006). The genetic factors involved in the establishment of such a ratio have been investigated and the Waxy gene has been identified as a key factor (Yamamori et al., 1994; Rahman et al., 2000; Svihus et al., 2005).

1.1.4.2 Lipids

Lipids are minor constituents of the wheat caryopsis. They derive from membranes, organelles and spherosomes in wheat grain (Goesaert et al., 2005) and vary in physical properties and chemical structures. As a matter of fact, within the caryopsis, there are both polar type lipids (glycolipids and phospholipids) and non-polar type lipids (especially triglycerides). Based on extractability, wheat lipids can be classified as starch lipids, and free or bound non-starch lipids (Pareyt et al., 2011). The latter ones are able to bind flour's proteins during the kneading of the dough, thus playing an important role in defining the rheological characteristics and influencing baking performance. For example, polar lipids (with proteins) increase the gas holding capacity and eventually improve the loaf volume and crumb resilience (Matsoukas and Morrison, 1991; Gan et al., 1995). Generally, it is accepted that a high amount of lipids and a high polar / non-polar lipids ratio exert good effects on bread-making quality (Chung, 1985).

1.1.4.3 Proteins

Storage proteins represent only 8-20% of the mature caryopsis but the unique functional properties of wheat doughs are largely attributable to this fraction. Seed storage proteins were classified for the first time in 1924 by Osborne in four groups, on the basis of their solubility:

- **Albumins,** soluble in water;
- **Globulins,** soluble in dilute saline;
- **Gliadins,** soluble in alcohol water mixtures
- **Glutenins,** soluble in dilute acid or alkali.

Albumins and globulins represent about 20% of the caryopsis' proteins and are mostly located in the casing of the seed, in aleurone cells and germ, while they are relatively less abundant in the endosperm. They have mainly structural and metabolic functions, as they are mostly proteins with enzymatic action. In particular globulins are mostly represented by enzymes with a α and β-amylase activity, thus playing an important role in the processes of dough leavening; these enzymes, in fact, catalyze the hydrolysis of the starch present in the damaged granules, producing substrates needed for the fermentative activity of the yeast (Ferrante, 2006).

Gliadins and glutenins, instead, represent the remaining 80% of the caryopsis' proteins and are only present in the endosperm, where they exert the physiological role of storage proteins. They are commonly named "prolamins" due to their high content of proline and glutamine and they are the major constituents of gluten (Shewry et al., 1986) (**Fig. 7**). Their particular amino acid composition,

in fact, allows storage proteins to be structured, during kneading of the flour with water, in a molecular complex characterized by reticular architecture (the gluten) held together mostly by di-sulphide bonds between cystein residues, but also by non-covalent interactions.

Prolamins can be further classified into three groups (Shewry and Haldford, 2001):

- **Sulphur-poor (S-poor):** comprising ω-gliadins and D-type LMW subunits of glutenin;
- **Sulphur-rich (S-rich):** comprising α, β and γ-gliadins and B- and C-type LMW subunits of glutenin;
- **HMW prolamins:** HMW subunits of glutenin.

Figure 7: Gluten obtained from wheat flour.

1.2 The Gluten

1.2.1 Gluten and its viscoelastic properties

Gluten can be defined as the rubbery mass that remains when the dough is washed to remove starch granules and water-soluble constituents (Wieser, 2007). Since the dry solid, after this action, contains 75-85% proteins and 5-10% lipids, the term "gluten" is usually referred to the mixture of proteins which constitutes itself and which plays a key role in determining the unique baking quality of wheat by conferring water absorption capacity, cohesivity, viscosity and elasticity to dough (Wieser, 2007).

Traditionally, gluten proteins have been divided into the two groups of soluble gliadins and insoluble glutenins, which are present either as monomers or as oligo and polymers (Wrigley and Bietz, 1988). Cysteine residues are responsible of the gluten proteins aggregation, as they allow the establishment of intra- and interchain disulphide bonds. However, also non-covalent interactions, such as hydrogen bonds, and hydrophobic interactions are relevant to this aggregation (Anjum et al., 2007).

Most prolamins share two common structural features. The first one is the presence of distinct regions, or domains, which assume different structures and may have different origins; the second one is the presence of amino acid sequences consisting of repeated blocks based on one or more short peptide motif. These features are responsible for the high proportion of glutamine and proline, while amino acids with charged side groups are poorly represented (Shewry and Haldford, 2001). Both fractions, of gliadins and glutenins, are important contributors to rheological properties of dough. Gliadins have little elasticity and are less cohesive than glutenins, thus contributing mainly to the viscosity and extensibility of the dough system, while glutenins are both cohesive and elastic and are responsible for dough strength and elasticity (Wiser, 2007). A proper mixture of both fractions is essential to impart the viscoelastic properties of dough and the quality of the end product. By the way of example, the production of bread requires dough with a balanced ratio of elasticity and extensibility; the production of biscuits or crackers requires an extremely extensible and little elastic dough, while the production of pasta requires extremely strong doughs (Ferrante, 2006).

Doughs are concentrated systems where shear and tensile forces, imparted by mixing, cause gluten proteins aggregation to form a network (Singh and MacRitchie, 2001). At a molecular level, these types of forces are needed, during the dough development, to extend the large glutenin molecules from their equilibrium conformations. Extended molecules give rise to elastic restoring forces and this high elasticity retards the molecular retraction and maintains the elasticity during resting. Polymer molecules respond to mixing stresses by three main processes: disentaglement, chain orientation and bond breaking. All three can occur during dough mixing. The covalent bonds that are broken are the disulphide bonds between glutenin subunits. Of interest is that the highest stresses occur at the centre of molecules where the probability of chain scission is greater (Singh and MacRitchie, 2001). Schematic mechanisms involved in dough formation are shown in **Fig. 8**. In particular, it has been postulated that disulphide bonds contribute to the process of dough formation through the disulphide-sulphydryl exchange (Lindsay et al., 2000a; Tilley et al., 2001).

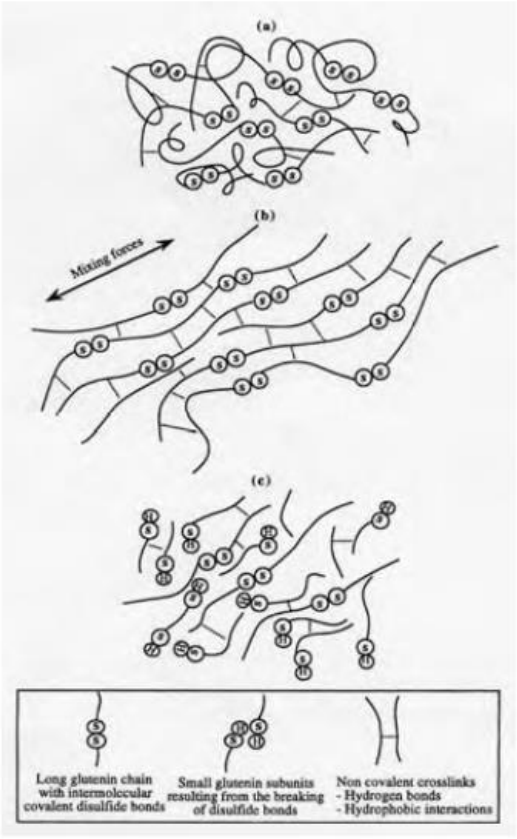

Figure 8: Molecular interpretation of the gluten development: a) beginning of the mixing, b) optimum development, c) overmixing (Létang *et al.*, 1999).

1.2.2 Gliadins

Gliadins account for 40-50% of the total storage proteins (Anderson et al., 1997) and play an important role in determining the functional property of wheat flour, being responsible for the viscous character of wheat dough. Gliadins are present as heterogeneous monomeric polipetides soluble in 70% aqueous alcohol and their molecular weight range from 30 and 70 kDa (Barak et al., 2015). They were initially classified into four groups on the basis of their electrophoretic mobility in A-PAGE (acid-PAGE) at low pH as shown in **Figure 9** (Wieser, 2007): α-gliadins (fastest mobility), β-, γ- and ω-gliadins (slowest mobility).

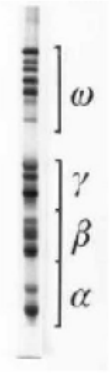

Figure 9: A-PAGE at low pH of wheat gliadins.

However, further analysis on amino acid sequences revealed a structural homology of the *α*- and *β*-gliadins, leading to a new classification that grouped the two types of proteins under one heading, the *α*-type gliadins (Zilic et al., 2011). The monomeric gliadins confer viscosity to dough by interacting with the glutamine residues mainly through noncovalent interactions such as hydrogen bonding, vander Waal's forces, electrostatic and hydrophobic interactions.

In spite of the different mobility in A-PAGE, gliadins show an overall common structure, consisting of a central domain (CD) characterized by repetitive amino acid (AAs) sequences rich in proline (Pro) and glutamine (Gln), and two terminal non-repetitive domains which are hydrophobic and contain most of the ionizable AAs (histidine, arginine and lysine), the latter being present in low levels (Gianibelli et al., 2001). The *α*- and *β*-type gliadins show a similar primary structure consisting of around 250-300 AA residues, with a molecular weight between 30 and 45 kDa. Their sequences are composed by a brief non-repetitive N-terminal domain of 5 residues, a CD of 113-134 AA residues rich in proline and glutamine repetitive sequences (heptapeptide: PQPQPFP and pentapeptide: PQQPY) (Ferranti et al., 2007) and the non-repetitive C-terminal domain of 144-166 residues containing six residues of cysteine, responsible of disulphide bonds that stabilize the protein structure (Shewry and Tatham, 1997), and charged AAs (**Fig. 10a**).

The *γ*-type gliadins show a primary structure similar to the one just reported for α-/β-type, with few differences and molecular weight varies from 30 and 45 kDa. Sequences, in fact, are composed by a non-repetitive N-terminal domain of 12 residues; a highly variable repetitive central domain of 72-161 residues, characterized by only one repetitive sequence (eptapeptide: PQQPFPQ) and a non-repetitive C-terminal domain in which 8 cysteine residues are present (**Fig. 10b**), thus resulting in a higher number of disulphide bonds (Cassidy et al., 1998; Barak et al., 2015).

Finally, *ω*-gliadins have a molecular weight between 44 and 74 kDa and show a general structure completely different from the other types (**Fig. 10c**). As a matter of fact, the primary structure is almost completely constituted by the repetitive central domain, which represents 90-95% of the total protein. Also, the amino acids composition is dissimilar if compared to those of *α*-, *β*- and *γ*-gliadins

in particular the *ω*-types lack cysteine residues, so that there's no possibility of disulphide crosslinks. Furthermore, AA sequences display the highest contents of glutamine, proline and phenylalanine residues (Hisa and Handerson, 2001; Barak et al., 2015), which together account for around 80% of the total composition.

As concerning the secondary structure, spectroscopic studies made by Tahtam and Shewry (1985) pointed out that *α*-/*β*- and *γ*-gliadins contained 30-35% *α*-helix and 10-20% *β*-sheet conformations, while *ω*-gliadins are rich in randomly coiled *β*-turns without detectable *α*-helix or *β*-sheet. Moreover, they reported that ω-types were mainly stabilized by strong hydrophobic interactions, while the other types were stabilized by covalent disulphide bonds and non-covalent hydrogen bonds in their *α*-helix and *β*-sheet.

Figure 10: Schematic representation of the different gliadin group: A) *α*-/*β*-gliadins; B) *γ*-gliadins and C) *ω*-gliadins (Ferranti et al., 2007).

Gliadins are encoded by the complex loci *Gli-1* (*Gli-A1*, *Gli-B1* and *Gli-D1*) and *Gli*-2 (*Gli-A2*, *Gli-B2* and *Gli-D2*), located on the short arms of group 1 and group 6 homologous chromosomes respectively (Wrigley and Shepherd, 1973). Most *γ*- and *ω*-type gliadins are coded by *Gli-1* genes, while all *α*-/*β*-type and some *γ*-type gliadins are coded by the *Gli-2* genes (Ferranti et al., 2007). However, a few gliadin components have been shown to be encoded by additional, dispersed genes (**Fig. 11**). An additional *Gli-B3* locus has been mapped on the short arm of chromosome 1B, between the *Gli-B1* locus and the centromere (Galili and Feldman, 1984; Metakovsky et al., 1986) and, at a comparable position, another one was mapped on the short arm of chromosome 1A (Sobko, 1984), therefore designated Gli-A3. Moreover, other genes coding for gliadin were found on the complex loci *Gli-5* (*Gli-A5*, *Gli-B5* and *Gli-D5*), located on the short arm of group 1 homologous chromosome

(Pogna et al., 1993/1995). Evidence of a novel locus (*Gli-A6*), located distally to *Gli-A1* on chromosome 1A, was reported by Metakovsky et al. (1997), while two new loci, designated *Gli-D4* and *Gli-D5* have been described on the short arm of chromosome 1D (Rodrìguez-Quijano and Carrillo, 1996).

As previously said, gliadins account for about half of the total seed storage proteins, however few of them have been found to influence gluten's quality. More precisely, during formation of the dough, gliadins act as "plasticizer" and promote the viscous flow and extensibility, considered as important rheological characteristics of dough (Barak et al., 2015). Damidaux et al., (1978) found that cultivars having γ-45 gliadin component exhibited stronger gluten as compared to cultivars with γ-42 gliadin component. Components were named depending on their mobility in A-PAGE (Bushuk and Zillman, 1978). Actually, further studies showed that these proteins are markers for quality, but the latter depends, instead, on the presence of different LMW-GS subunits encoded by the *Glu-B3* locus, which is only 2 centimorgan (cM) far from the *Gli-B1* locus. In other words, the relationship between the γ-gliadins 45 and γ-42 and gluten's quality reflects the genetic linkage with the LMW glutenin subunits; γ-gliadins 45 and other gliadins known to exert positive effects on the gluten quality, indeed, are encoded by the *Gli-Blc* locus on the short arm of chromosome 1B, which is closely linked genetically to the allele coding for a group of B type LMW-GS, namely LMW-2, which positive affect gluten strength (Troccoli et al., 2000).

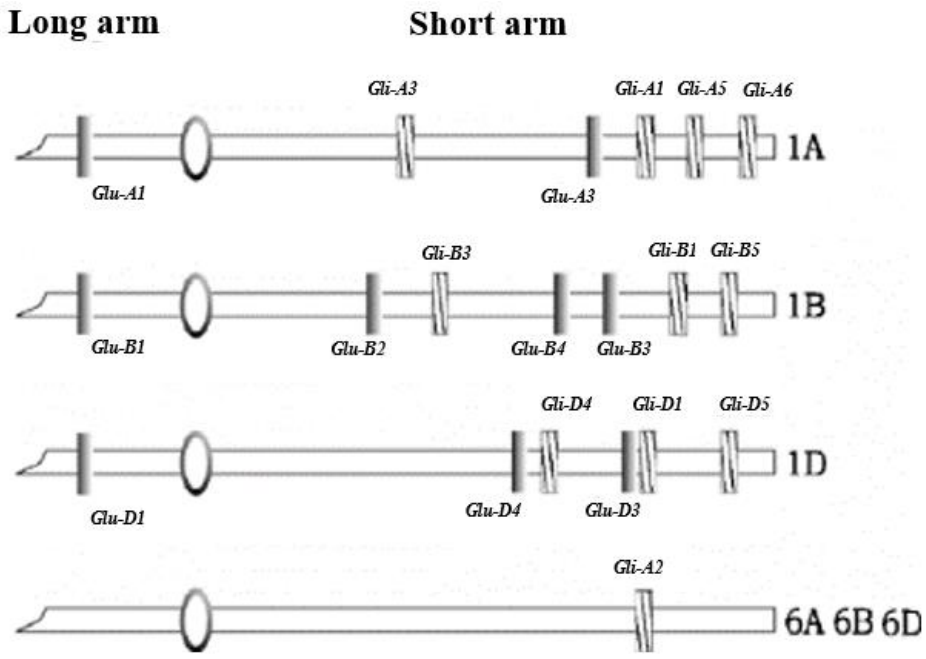

Figure 11: Chromosomal location of major and minor gliadin and glutenin subunit loci in bread wheat (Shewry et al., 2003).

1.2.3 Glutenins

The glutenin fraction is composed by aggregated proteins linked by interchain disulphide bonds, thus constituting polymers, whose molecular weight may vary from 60 000 to several millions Da (Wieser et al., 2006). This ranging of size makes them one of the larger polymers occurring in nature. Moreover, the size of glutenin polymers has been recognized as one of the main determinants of wheat dough properties and baking performance. In particular, the larger dimension of these polymers is responsible for better quality characteristics of dough, such as strength, elasticity and loaf volume. Flour with reduced size of glutenin polymers, instead, shows poor technological quality (Schropp and Wieser, 1996; Wieser, 2007).

After reduction of disulphide bonds, the resulting glutenin subunits show a solubility in aqueous alcohols similar to gliadins. These subunits can be separated on a polyacrylamide gel under denaturing conditions (SDS-PAGE) into two groups on the basis of their electrophoretic mobility (**Fig. 12**): the Low Molecular Weight Glutenin Subunits (LMW-GS) and the High Molecular Weight Glutenin Subunits (HMW-GS). The LMW are the predominant type and the more heterogeneous, while the HMW are represented by a reduced number of subunits, thus being the minor components within the gluten protein family.

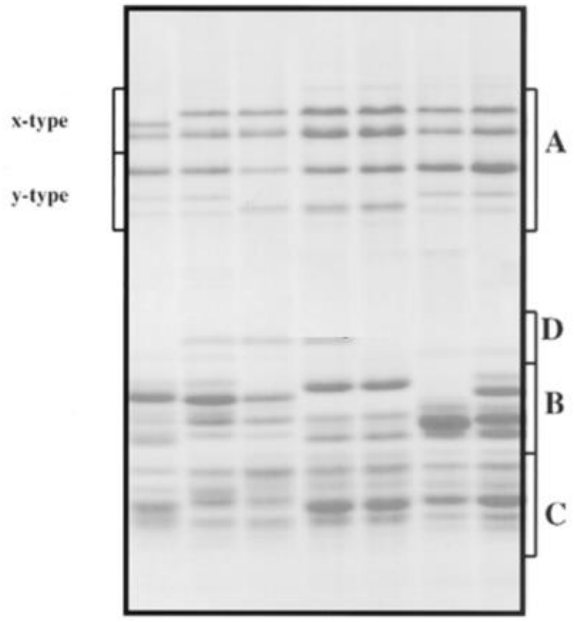

Figure 12: SDS-PAGE of the glutenins fraction after reduction of disulphide bonds. Group A) HMW-GS proteins with x- and y-subunits; Group B) B-, C- and D-type of LMW-GS proteins (Gianibelli et al., 2001).

Glutenin macropolymers are main constituents of gluten and are responsible of dough properties. According to the widely accepted model of gluten structure, cysteine residues present in

HMW-GS are involved in several S-S bonds, which may include: one interchain S-S bond within the N-terminal domain of an x-type subunits, two parallel S-S bonds between the N-termini of y-type subunits, an interchain bond between y-type subunits and LMW glutenin subunits, and a bond linking y-type and x-type subunits in a "head-to-tail" fashion (**Fig. 13**). The molecular basis of a gluten polymer unit is composed by 2 y-type HMW-GS, 4 x-type HMW-GS and around 30 LMW subunits (Wieser, 2007; **Fig. 13 red box**), thus a proper mixture of glutenin components is essential to ensure dough technological properties. Generally, quantitative data on glutenin composition report an HMW-GS/LMW-GS ratio of about 1:2. Glutenin polymer, however, can include more than ten units of this type, with a consequently increase of the HMW/LMW ratio (Wieser, 2007).

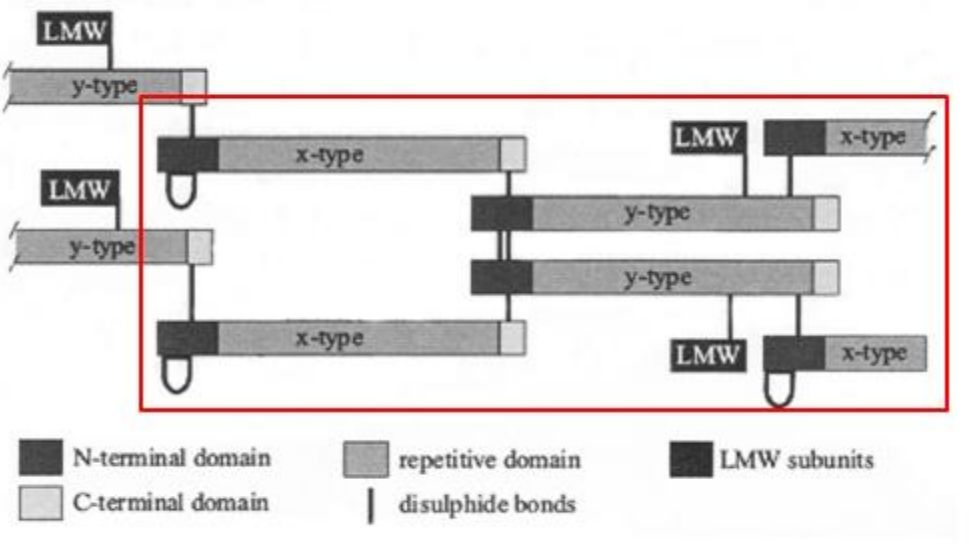

Figure 13: Structural model of gluten which is held together by disulphide bonds (Köhler et al., 1991). The red box highlights the polymer unit in which the HMW-GS form the skeleton and the LMW-GS form the branching.

1.2.3.1 Low Molecular Weight Glutenin Subunits (LMW-GS)

The low molecular weight glutenin subunits (LMW-GS) are an extensive group of proteins which accounts for about one-third of the seed storage proteins and 60% of glutenin proteins, and plays an important role in determining dough properties and the quality of wheat food products (Gupta et al., 1991; Branlard et al., 2001). These proteins are particularly heterogeneous, so that their structure, organization and correlation with quality traits have still to be completely investigated. The reduced subunits were broadly divided into three types, B, C and D by Payne et al. (1979) on the basis of their mobility on SDS-PAGE. Subsequent analyses have revealed that the D-type subunits belong to the S-poor prolamins. This group (D-type) comprises proteins with molecular weight of about 40 kDa, which, on SDS-PAGE migrate slower than B-type, and corresponds to modified ω-gliadins.

Through MALDI mass spectrometry analysis of alkylated and purified proteins of the two groups, in fact, Masci et al. (1993 and 1999) identified similar patterns with three major fluorescent peaks, indicating that the three peptides corresponded to the same part of protein. These results were, thus, consistent with the D-LMW subunits being mutant forms of ω-gliadins, in which a single cysteine residue has been acquired from the D-LMWs as a result of a spontaneous mutation event. Thanks to this cysteine, these proteins take part in the formation of glutenin polymers, making inter-molecular bonds with other subunits. The B-type subunits are the major group in term of quantity and present a molecular weight ranging from 40 to 50 kDa, while the C-type are less abundant and show a molecular weight between 30-40 kDa. In 1984, Kasarda et al. made a detailed study of these two types of subunits by the N-terminal sequencing, which allowed them to identify two different N-terminal sequences for the B-type. These sequences were called LMW-m, LMW-s and LMW-i depending on the first amino acids in the N-terminal sequence (methionine, serine and isoleucine respectively). LMW-s are the most abundant sequences observed, starting with the sequence SHIPGL-, while regarding the LMW-m, three N-terminal variants were found; METSHIPGL-, METSRIPGL- and METSCIPGL (Gianibelli et al., 2001).

All the LMW-GS are composed by 250-300 amino acid residues and, as concerning the primary structure, they show an overall similarity with the structure of gliadins. They have, in fact, a clear two-domain structure, with an N-terminal domain, with a small non-repetitive fraction and a larger repetitive part, and a C-terminal non-repetitive domain (**Fig. 14**) (Gianibelli et al., 2001). The N-terminal repetitive domain accounts for the 30-50% of the total protein, differently from what observed for gliadins or HMW-GS in which this domain is the predominant; it contains one residue of cysteine and its secondary structure is rich in β-turns. The short non-repetitive N-terminal domain, instead, is rich in α-helix and appears to be more compact (Thomson et al., 1992). The C-terminal domain contains seven cysteine residues, at least one of which is unpaired, thus available for intramolecular bonding (Gianibelli et al., 2001).

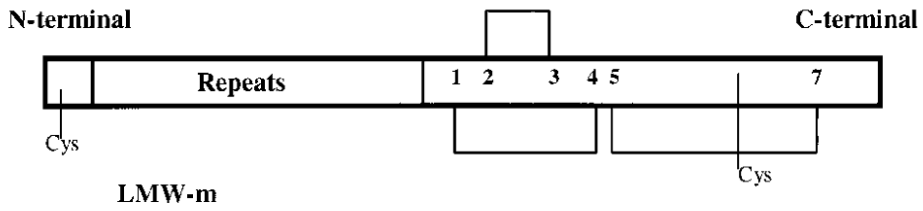

Figure 14: Schematic representation of m-type LMW-GS structure. Free cysteine residues and those formed intramolecular bonds are indicated (Gianibelli et al., 2001).

The B-type LMW-GS, together with some C-type, are controlled by genes at complex loci *Glu-3* (*Glu-A3*, *Glu-B3* and *Glu-D3*), located on the short arm of chromosomes 1as shown in **Fig. 11** (Gianibelli et al., 2001). By screening a collection of 222 hexaploid cultivars, Gupta and Shepherd (1990) detected six different band patterns for the *Glu-A3* locus, nine for the *Glu-B3* locus and five for the *Glu-D3* locus. The D-type subunits, instead, are coded by a separated locus on the chromosome 1B, which is called *Glu-B2* and maps between the *Glu-B1* (locus responsible for HMW-GS genes) and the *Gli-B1*. Some components of the C-type LMW-GS, the ones called "gliadin-like", are encoded on the short arm of chromosome 6 by the same loci *Gli-2*, and this is consistent with these subunits being forms of modified gliadins. Moreover, two LMW-GS of 31 and 32 kDa, which have N-terminal amino acid sequences homologous with those of B- and C-type LMW subunits, have been reported by Sreeramulu and Singh (1997). Genes encoding them (*Glu-D4* and *Glu-D5*) are located on chromosomes 1D and 7D respectively, although their exact location within the chromosomes has not been established.

From a technological point of view, LMW-GS are able to form large aggregates that are related to dough strength and several studies have revealed that the allelic variation at LMW-GS loci is associated with significant differences in dough quality in bread (Gupta et al., 1989, 1994) and durum wheat (Pogna et al., 1990; Ruiz and Carrillo, 1993). What is clearly known is that cysteine residues in the primary structure have two different roles in polymers formation: they can act as "chain extenders", if subunits with two or more cysteine residues that form intermolecular disulphide bonds are considered, or they can act as "chain terminators", if subunits with just one cysteine residue available for intermolecular disulphide bonding are considered. In general, chain extenders are responsible for dough tenacity, defined as the maximum resistance to deformation, while chain terminators seem to have negative effect on dough quality. As previously affirmed, gluten quality is highly affected by the presence of a particular pattern of B type LMW-GS, namely LMW-2 pattern, which are known to be related with good pasta cooking quality and gluten strength (Troccoli et al., 2000).

1.2.3.2 High Molecular Weight Glutenin Subunits (HMW-GS)

HMW glutenin subunits are the less abundant among prolamins components, but are key factors in the process of bread-making as they are the major determinants of gluten elasticity (Tatham et al., 1985) as they appear to promote the formation of larger glutenin polymers. Based on SDS-PAGE analysis, two different groups of HMW subunit have been identified (Bietz and Wall, 1972): one, named x-type, includes protein with lower electrophoretic mobility, with a molecular weight ranging

from 83 to 90 kDa, while y-type, have higher electrophoretic mobility due to their lower molecular weight (67-74 kDa).

By comparing the HMW-GS sequences, it has been demonstrated that they all share a similar three domain structure, with a non-repetitive N-terminal domain of 81 to 104 residues, a repetitive central domain of 481 to 681 residues, and a final non-repetitive C-terminal domain of 42 residues. The central domain, which accounts for 85% of the total protein, is characterized by the tandem repetition of short motives rich in glutamine (nonapeptide: GYYPTSPQQ and hexapeptide: PGQGQQ), conferring elasticity to the molecules. Nonapeptide belonging to the y-type HMW subunits have the second proline replaced by a leucine (GYYPTSLQQ) and the x-type HMW-GS present an additional tripeptide motif (GQQ) (Gianibelli et al., 2001). The N-terminal region presents three to five cysteine residues which provide intermolecular disulphide bonds between HMW-GS and LMW-GS to form protein polymers with sizes that could reach up to tens of millions of Daltons (Shewry et al., 1992; Wrigley et al., 1996). Finally, the C-terminal domain contains one residue of cysteine. Consequently, HMW-GS could present up to six conserved cysteine residues (**Fig. 15**). The most important difference between the x- and y-type subunits lies within the N-terminal domain, as the y-type has an insertion of 18 residues which includes two more adjacent cysteine residues not found in the x-type. The amino acids composition of these proteins has indicated a hydrophilic nature of the central repetitive domain and a hydrophobic nature of the remaining N- and C-terminal domains (Shewry et al., 1989). Both x- and y-type HMW subunits show a β-turn conformation for the central domain (Tatham et al., 1990), while the structure of both terminal regions is predicted as an α-helix (Miles et al., 1991; Shewry et al., 1992).

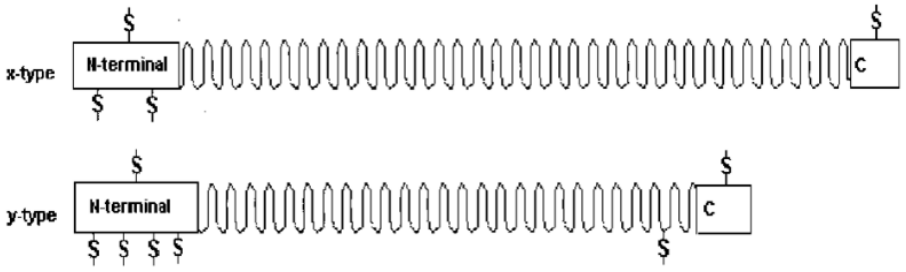

Figure 15: Schematic representation of x- and y-type high molecular weight glutenin subunits (Wieser, 2007).

HMW-GS proteins are encoded by the complex *Glu-1* loci located on the long arm of homologous chromosomes 1 and therefore named *Glu-A1*, *Glu-B1* and *Glu-D1* as shown in **Fig. 11** (Bietz et al., 1975; Payne et al., 1980). Each locus includes two tightly linked genes encoding the HMW-GS x and y type (Shewry et al., 1992). Significant polymorphism in the number and mobility of HMW-GS alleles has been demonstrated for both bread (Lawrence and Shepherd, 1980; Payne et

al., 1980) and durum wheat (Branlard et al., 1989). The range of alleles at these loci was summarized in 1983 by Payne and Lowrence, who identified three alleles at the *Glu-A1* locus, 11 allelic forms at the *Glu-B1* locus and six alleles at the Glu-D1 locus. Originally, a nomenclature scheme based on the mobility on SDS-PAGE of the relative bands, with lower number equating to lower mobility have been proposed (Payne and Lowrence, 1983; Gianibelli et al., 2001, **Fig. 16**). However, with the subsequent identification of new subunits it has been more difficult to follow this logical order; thus, there are some subunits, such as 21, with lower mobility and higher number than the original subunits (Gianibelli et al., 2001). Bread wheats could contain six different HMW-GS but, due to the silencing of some of these genes, they usually possess three to five HMW subunits. Similarly, durum wheats could contain four HMW-GS, which become one to three because of silencing.

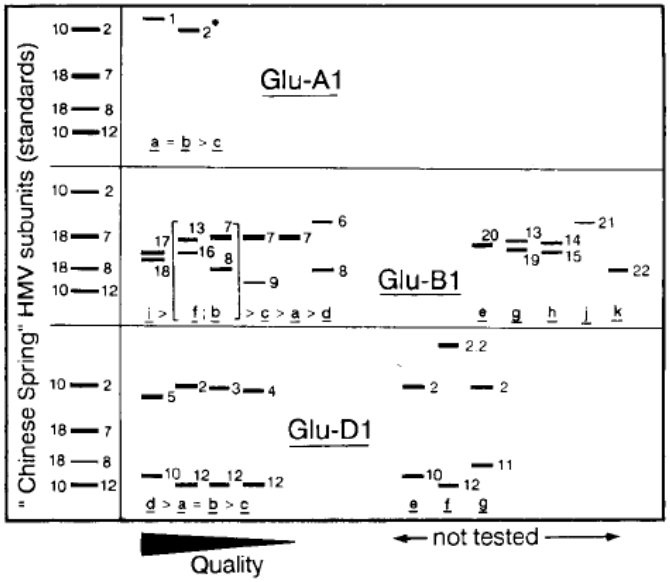

Figure 16. Allelic variation in HMW-GS at three gene loci and corresponding bands' nomenclature, based on SDS-PAGE fractionation. The relationship band-breadmaking quality is indicated. Chinese spring patterns on the left of the box are included for comparison of relative mobilities (Payne et al., 1984).

High Molecular Weight glutenin subunits (HMW-GS) are of considerable interest because of their relationship to bread-making quality for their contribution to dough strength and stabilization (Branlard and Dardevet, 1985). Two features of these protein subunits may play an important role in glutenin elastomers: the number and distribution of disulphide bonds and the properties and interactions of the repetitive domains. Many studies searched for a correlation between HMW-GS and bread-making quality, finding out that the Ax1 and Ax2* at *Glu-A1* locus, together with the Bx7+By8, Bx13+By16, Bx17+By18 and Bx7+By9 alleles at *Glu-B1* locus and the Dx5+Dy10 at the *Glu-D1* locus, impart better quality to common wheat dough (Branlard and Dardevet, 1985; Rasheed et al., 2014). The contribution of the *Glu-B1* alleles to the pasta quality have been represented as the

ranking Bx7+By8>Bx6+By8 (Sisson et al., 2005) with Bx7+By8 linked to a better quality or, Bx17+By18 = Bx13+By19 = Bx7+By8>Bx7>Bx6+By8 when a wider range of alleles was included (Varzakas et al., 2014). On the contrary, allelic combinations such as Bx20+By20 and Bx14+By15 at *Glu-B1* locus were identified as related to the worst dough quality traits (Martinez et al., 2005; Rasheed et al., 2014). Further analysis showed that the N-terminal domain of Bx20 and Bx14 sequences lack of two to three cysteine residues, which are replaced by tyrosine residues (Shewry et al., 2003; Li et al., 2004). Both the conserved cysteine residues and the size of the repetitive domains contribute to the high order structure of HMW glutenins subunits; the former is involved in the formation of inter- or intra-molecular disulphide bonds, while the latter may promote inter-molecular interactions through hydrogen bonding (Shewry et al., 2002).

Nowadays modern cultivars show an HMW-GS allelic variability which is not very large and the cultivars used for high strength wheat breeding are very similar in their HMW glutenin composition. Wheat landraces, instead, have a higher allelic variability and could be very useful for broadening the currently narrow genetic basis of modern cultivars in breeding programs (Giraldo et al., 2016).

1.3 Durum wheat quality

1.3.1 General concepts

Wheat quality is a very broad term, and its definition depends on whether it is being assessed for nutritional or processing purpose; as a matter of fact, the concept of quality assumes different meaning depending either on the step of the production's chain and on the point of view considered (producer or consumer). It is possible to identify different types of quality (Troccoli et al., 2000; Flagella, 2006):

- *agronomic quality*: it is the one of interest for the farmer and it corresponds principally with the productivity. Farmer's concept of quality is closely linked to the need to obtain high yields in order to maximize profit and, in turn, the potential grain yield depends largely on the ability of the varieties to adapt to different environments as nowadays most of the durum wheat growing areas undergo severe weather condition. Thus, the concept of yield stability is closely related with the one of productivity and it encloses also resistance to both biotic and abiotic stresses and the capability to maintain high yields under favorable and un-favorable environments. More precisely, drought, high temperature and terminal stress (i.e. plant death) during grain-filling are major factors affecting durum wheat yield and, consequently, the grain quality;

- *milling quality*: it is the one of interest for the miller and it is mainly represented by high extraction rate, in other words by the proportion of the wheat kernel which is milled into semolina. Other factors that affected the determination of quality for the millers are: the ash content (whose value, in Italy, has to be between 0.9 and 1.2), test weight (which can be a reflection of the soundness of grain and therefore high value may be expected when the grain is undamaged by un-favourable weather conditions) and kernel weight (which reflect the average kernel size, thus larger kernels are expected to have a greater endosperm to bran ratio);
- *technological quality*: it is the one of interest for the processing industry and it is represented by wheat's ability to be processed in different way depending on the final product desired. Generally, the determination of wheat's technological characteristics is carried out through simulation of processing processes and the most important parameters involved in determining the technological quality are: protein content, gluten (as quantity and quality) and determination of technological indices (see paragraph 1.2.2);
- *sensory quality*: it is the one of the consumer, who is concerned about pasta cooking quality as well as the aroma, taste and color. According to the consumer, quality could be defined as a tasty, nutritious and safe product which has the capacity to maintain good texture after cooking and not to become a thick-sticky mass (Troccoli et al., 2000; Varzakaset al., 2014).

1.3.2 Genetic determination of durum wheat quality

Several genes have been identified as affecting wheat quality. In particular, the presence of specific alleles of genes encoding for gliadins or glutenins influences positively or negatively the quality (see paragraph 1.2.2 and 1.2.3).

The rapid development over last few decades of molecular genetics lead to new and more reliable selection criteria, based on genotype. Nowadays, the most applicable and accepted approach, used in breeding programs to assess the protein quality of a certain variety, is based on biochemical and molecular or DNA markers; in particular, the latter ones show several advantages such as being phenotypically neutral, independent from the environment, able to detect polymorphism between individual plants and suitable for high-throughput tests (Lei et al., 2006). The extensive use of molecular markers in plant breeding is due to the possibility of conducting a direct selection of traits that are important for the improvement of a certain crop and, above all, to the capability of detecting these traits at all stages of plant development, even at early stages of the breeding program, so that poor-quality lines are not propagated (Uthayakumaran et al., 2006).

Molecular markers must meet certain requirements in order to be useful and efficient, such as possessing adequate polymorphism, being qualitatively independent to the growth conditions and following a co-dominant pattern of inheritance (Varzakas et al., 2014). They can be classified into three groups: DNA markers based on hybridization (RAPDs), DNA markers based on PCR (i.e. RAPDs, SCARs, SSRs, microsatellites, SST, AFLPs etc.) and DNA markers based on small sequenced fragments (SNPs). Moreover, other types of molecular marker could be useful in detecting genetic polymorphisms; MP-PCR (microsatellite-primed polymerase chain reaction), AP-PCR (arbitrarily primed PCR), AS-PCR (allele-specific PCR) and DAF (DNA amplification fingerprinting) (Varzakas et al., 2014). All these markers have found three main field of applicability in plant breeding and genetics.

The first one is in determining genetic relationship (varietal identification, parental determination exc.) (Liviero et al., 2002; Maestri et al., 2002). The second one involves their use for identification and mapping of loci associated with quantitative and qualitative traits and for monitoring of these loci in breeding programs (marker-assisted-selection MAS). The third one is the use of genetic markers in population genetic studies (Varzakas et al., 2014).

In recent years MAS to improve breeding efficiency has become commonplace in breeding programs with various aims (Bassi et al., 2016) among which the discrimination of different constituents of wheat quality, especially seed storage proteins, thus replacing traditional approaches such as SDS-PAGE electrophoresis. Aa a matter of fact, the latter exhibits several disadvantages: it requires the involvement of toxic compounds (acrylamide and reducing agents), it is time consuming and, talking about the discrimination of glutenin proteins of different varieties, it frequently results in an incorrect interpretation of different alleles due to the same mobility on SDS gels of the corresponding proteins (Gianibelli et al., 2001). MAS, instead, represents an efficient and fast tool to overcome some of the drawbacks in protein-based methods for glutenins and, especially, for HMW-GS allele resolution (Rasheed et al., 2014) and can help avoid the misleading interpretation of the results obtained either by SDS-PAGE analysis or single allelic selection (Ghazy et al., 2012). As Varzakas et al. (2014) stated, MAS provides a much more convenient tool for rapid genetic analyses as it permits single plant selection, DNA can be extracted from any tissue that can be collected at almost all stages of plant growth and multiple markers can be combined in a single screening experiment, enabling breeders to evaluate genotypes at multiple loci using the same template DNA. Among all markers, the allele-specific polymerase chain reaction (AS-PCR) ones are the most useful in detecting genetic polymorphism since they are based on single nucleotide polymorphisms (SNPs) (Liang et al., 2015). The latter are, thus, the most applied in the characterization of HMW-GS (Ahmad, 2000; Butow et al., 2003; Ma et al., 2003; Shwarz et al., 2004; Lei et al., 2006; Liu et al., 2008; Xu et al., 2008;

Raguphaty et al., 2008; Janni et al., 2017) (https//wheat.pw.usda.gov/cgi-bin/GG3/report.cgi?class=marker&name=Glu-B1). Ma et al. have developed simple, rapid, sub-unit specific PCR-based assays to discriminate the predominant alleles *Glu-A1a* (Ax1) and *GluA1b* (Ax2*), *Glu-B1b* (Bx7+By8) and *Glu-B1i* (Bx17+By18), and *Glu-D1d* (Dx5+Dy10) and *Glu-D1a* (Dx2+Dy12) for the A, B and D genome loci respectively for the purpose of enhanced molecular marker-assisted breeding. Similarly, Lei et al. (2006) focused their attention on the development of a set of markers aimed at distinguishing alleles at the *Glu-B1* locus, with the emphasis to the By type genes. As a matter of fact, they provided to the science community one dominant and three co-dominant molecular markers useful to the discrimination of *By8*, *By9*, *By16* and *By20* genes. A co-dominant markers was developed in 2000 by Ahmad to specifically distinguish between *Bx7* and *By17* genes, while Butow et al. (2004) developed two co-dominant markers to discriminate *Bx17*, *Bx7*, *Bx7*^{OE}, *Bx7** and *Bx20* genes. The same *Bx7* and *Bx7** genes were subsequently distinguished by exploiting the molecular marker developed by Espì et al. in 2012. Additional two co-dominant markers were designed by Xu et al. (2008) based on the sequences of *Bx14* and *Bx17* genes, whereas Liang et al. in 2015 developed two sets of SNP-based AS-PCR markers for *By18* gene, with a successful validation on 110 bread wheat varieties with different *Glu-By* alleles. Other types of approaches, instead, are based on the simultaneous usage of different kind of markers to genotype cultivars, investigate the genetic diversity and estimate the genetic relationship. A recent example is the study of Henkrar et al. (2016) who exploited 47 functional markers, ideal for MAS as they are derived directly from the gene conferring the phenotype, and 7 random DNA markers closely linked to 21 loci of the most important targeted traits for breeding, with the final aim to genetically characterize 20 Moroccan genotypes and 19 exotic bread wheats cultivars. Additionally, the same markers were used to assess the diversity analysis and to estimate the genetic relationship among varieties.

Usage of molecular markers have been also extensively exploited to discriminate different LMW-GS genes (Espì et al., 2014). For example, Ikeda et al. in 2002 classified LMW-GS genes of a Japanese soft wheat, Norin 61, into six types on the basis of the distribution of cysteine residues, and twelve groups based on the deduced amino acid sequence identity in the N- and C-terminal conserved domains; subsequently, in 2006, they constructed LMW-GS group-specific PCR primer sets to determine their loci, and identified the proteins corresponding to these groups by determining the N-terminal amino acid sequence of all LMW-GSs of Norin 61 separated by two-dimensional gel electrophoresis (2DE). Also Long et al. (2005) carried out a specific-primers-based PCR assay on Group 1 ditelosomic lines of *Chinese Spring* to identify the chromosome locations of each group of LMW-GS genes; the analysis allowed to successfully assign to specific chromosome arms to five out

of the seven groups of LMW-GS gene. Additionally, led to the identification of two subgroups with high sequence similarity and located on different chromosome arms.

1.4 Wheat adverse reactions

Cereal crops, such as wheat, rye and barley, fulfill a key role for human diet in Western countries and, more recent in Eastern countries as well. However, all these crops are gluten-containing cereals and gluten proteins, ingested or inhalated in the form of raw flour or cooked products, can be responsible for a wide range of adverse reactions. Due to the increasing of their consumption almost all over the world, gluten-related disorders have gradually emerged as an epidemiologically relevant phenomenon with a global prevalence that is estimated around 5% (Elli et al., 2015).

In 2012 Sapone et al. reviewed the spectrum of gluten-related disorders, classifying them accordingly to the immune response involved and proposing a new nomenclature. As shown in **Fig. 17**, three different responses are triggered by the gluten ingestion; one autoimmune, which includes Celiac Disease (CD), Gluten Ataxia and Dermatitis herpetiformis; one allergic, which includes different manifestations of wheat allergy, such as Baker's asthma, food allergy, Wheat-Dependent Exercise-Induced Anaphylaxis (WDEIA) and contact urticaria; one not autoimmune and not allergic, which is known as Gluten Sensitivity (GS) or Non-Celiac-Gluten-Sensitivity (NCGS).

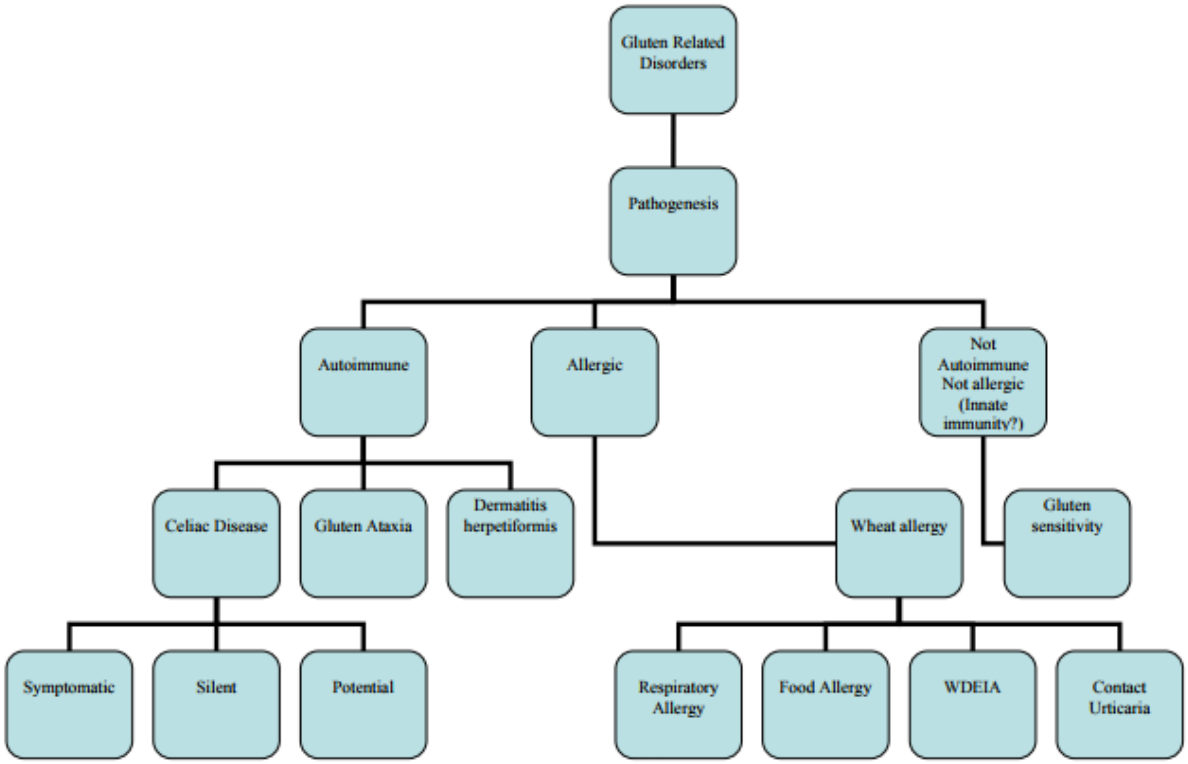

Figure 17: Proposed new nomenclature and classification of gluten-related disorders (from Sapone et al., 2012).

1.4.1 Autoimmune disorders

1.4.1.1 Celiac Disease

Celiac disease (CD) is a chronic, small-intestinal immune-mediated enteropathy initiated by exposure to dietary gluten in genetically predisposed individuals. It is estimated to affect 1% of the general populations, with a considerable proportion of patients remaining undiagnosed and untreated, and it can develop at any age and can affect almost any race (Elli et al., 2015; Leonard et al., 2017). CD patients can display either classical intestinal symptoms, such as chronic diarrhea and weight loss, or non-classical extraintestinal symptoms, such as anemia, osteoporosis and neurological disorders. The incidence of CD is increased in at-risk conditions like family history, IgA deficiency, some genetic syndromes (Down, Turner and William syndromes) and especially type 1 diabetes and thyroiditis (Sapone et al., 2012).

It is well-known that the genetic predisposition plays a crucial role in CD since it is strongly linked with specific human leukocyte antigen (HLA) class II genes, known as HLA-DQ2 and HLA-DQ8, located on chromosome 6p21 (Sapone et al., 2012). The ingestion of gluten in individuals carrying these two specific HLA alleles can elicit a T-cell mediated immune reaction against tissue transglutaminase, leading to mucosal damage and villous atrophy. Gliadins are supposed to be the

active triggers within gluten proteins, since they contain a unique 33-mer gliadin fragment evidenced to be the most immunogenic peptide (Elli et al., 2015); moreover, the latter is resistant to enzymatic degradation by gastric, pancreatic and brush border peptidase. LMW-glutenin proteins presented also several motifs associated with the induction of CD (Mamone et al., 2015). Altered processing by intraluminal enzymes, changes in intestinal permeability and activation of innate immunity seem to precede the activation of the adaptive immune response (Sapone et al., 2012). Hence, gluten epitopes might be subdivided into type's fragments; toxic peptides are able to induce mucosal damage, while immunogenic peptides are able to specifically stimulate HLA-DQ2 or DQ8- restricted T cell lines and T cell clones (Boukid et al., 2017a). Until now, 31 sequences of 9 amino acids have been defined in the gluten (gliadins and glutenins) of wheat and related species (e.g., barley, rye, oat, etc.) as being celiac-triggering peptides, often referred to as celiac "epitopes" (Shewry and Tatham, 2016). The majority of celiac toxic peptides have been identified from *in vitro* studies using peptides cultured with T cell lines or T cell clones derived from the small intestinal mucosa biopsy of CD patients (Carmarca et al., 2012).

1.4.1.2 Gluten Ataxia (GA)

It is defined as an idiopathic sporadic ataxia with positive serological markers of gluten sensitization. Similarly to CD, it is an autoimmune disease triggered by the gluten ingestion in genetically susceptible individuals and it is characterized by damage to the cerebellum thus resulting in ataxia (Sapone et al., 2012; Hadjivassiliou et al., 2008). GA is characterized by insidious onset of predominantly gait ataxia, often associated with symptoms and signs suggestive of peripheral neuropathy. It affects both sex equally and, generally, it appears for the first time at the mean age of 53 (Hadjivassiliou et al., 2008). This disease usually presents with pure cerebellar ataxia or, in rare situations, ataxia in combination with other syndromes or neurological disorder (Sapone et al., 2012.

1.4.1.3 Dermatitis herpetiformis (DH)

Also known as Duhring's disease, Dermatitis herpetiformis is an external manifestation of celiac disease characterised by blisters filled with a watery fluid. Despite its name, DH is not related to herpes virus and even not caused by it : the name means that it is a skin inflammation having an appearance similar to herpes. DH affects 10-15% of people with celiac disease; as a matter of fact, it is recognized that DH is an extraintestinal manifestation of celiac disease as these two conditions occur frequently in the same families or identical twins and the majority of patients with DH have

small bowel mucosal changes compatible with classic or early celiac disease (Salmi et al., 2011). DH affects people of all ages, even if most often appears for the first time between the ages of 30 and 40, and it is somewhat more common in men than women (Celiac.org/celiac-disease/understanding-celiac-disease-2/dermatitis-herpetiformis/). Symptoms normally resolve with a strict, gluten-free diet.

1.4.2 Allergic disorders

1.4.2.1 Wheat allergy

Nowadays, wheat is one of the eight most common triggering factors for immune mediated food allergies which are usually immunoglobulin E (IgE) mediated. These reactions are characterized by a T helper type 2 (Th2) lymphocytic inflammation with predominant Th2 cytokines expression such as interleukin (IL)-4, IL-13 and IL-5. Depending on the route of allergen exposure and the underlying immunologic mechanisms, it is possible to classify wheat allergy (WA) into two cluster of disease: if the IgE mediated reaction is due to ingestion, WA usually refers to *food allergy*, while if the IgE mediated reaction is due to inhalation, WA usually refers to *respiration allergy* (Cianferoni 2016; Sapone et al., 2012).

In turn, food allergy to wheat, which affects approximately 0,4% to 1% of young children (Sicherer and Sampson, 2006; Keet et al., 2009), includes a wide variety of responses such as atopic dermatitis, urticaria, anaphylaxis and the common-known wheat-dependent exercise-induced anaphylaxis (WDEIA). These range of responses may vary between populations and be related to age and symptoms but it has been observed that children have a higher prevalence of food allergy to wheat compared to adults, especially if wheat was introduced after 6 months of age (Poole et al., 2006). Talking about WDEIA, it is a relatively rare, but potentially severe form of food allergy and, more precisely, it is a special form of food-dependent exercise-induced anaphylaxis (FDEIA): indeed, this reaction is specifically due to the combination of wheat intake and cofactors such as exercise. Symptoms are usually elicited by physical activity one to four hours after wheat ingestions, the frequency of WDEIA events vary from patient to patient and this reaction can manifest at any age, with teen and adults without any prior history of food allergy to be the most affected (Du Toit 2007; Sapone et al., 2012). WDEAIA patients typically manifest pruritus, local and generalized urticaria, angioedema, dyspnea, upper respiratory tract obstruction, gastrointestinal symptoms, hypotension, and if extensive, bronchospasm, vascular collapse and anaphylactic shock. High concentration of IgE antibodies specific to ω5-gliadin together with increased concentration of IgA antibodies to total gliadin and ω5-gliadin has been observed (Sapone et al., 2012).

On the other hand, respiratory allergy usually refers to the common-known Baker's asthma reaction. It is one of the most frequently reported forms of occupational asthma in several countries, affecting between 4 and 10% of bakery workers in Europe (Lupi et al., 2014). In particular, it is the first type of occupational asthma in France (Ameille et al., 2003) and the second in the UK (McDonald et al., 2000) and, in most cases, it is preceded by rhinitis. Since this IgE mediated immune reaction is due to the inhalation of wheat flour and dusts, conjunctivitis and skin symptoms may also occur, and symptoms can develop after a latency period which vary from months to years, even decades (Brisman 2002). Baker's asthma affects up to 10-15% of bakers, miller and pastry factory workers (Sapone et al., 2012), while the incidence of rhinitis ranges from 18 to 29% (Cianferoni 2016).

In general, it is possible to affirm that many allergic proteins are involved in WA and the latest update version of the WHO/IUIS Allergen Nomenclature Database describes 21 different well-classified wheat allergens (Elli et al., 2015). Among these, α-Amylase/trypsin inhibitor, the water insoluble ω5-gliadin (Tri a 19) and the nonspecific lipid transfer protein (nsLTP or Tri a 14) have been identified as potent allergens in WDEIA; HMW glutenin subunits seem correlated with the insurgence of contact urticaria; peroxidase, thioredoxins (Tri a 25), serine proteinase inhibitor and several wheat prolamins, such as α-, β-, γ- and ω-gliadins and LMW glutenin subunits, appear to trigger baker's asthma. Although some allergens seem to be mainly associated with specific allergic reaction, accordingly to what aforesaid, it has been demonstrated that there is a clear overlap between the ranges of proteins responsible for different clinical conditions (Inomata et al., 2009; Salcedo et al., 2011).

1.4.3 Not autoimmune and not allergic disorders

1.4.3.1 Gluten Sensitivity (GS)

During the Consensus Conferences in London (2011) and Munich (2012) a group of experts redefined the spectrum of gluten related disease including a new disease which has been called Gluten Sensitivity (GS) (Volta et al., 2013). Based on several contributions on this topic, GS has been defined as a reaction to the gluten ingestion in which not allergic nor autoimmune mechanisms are activated (Holmes 2013). As a matter of fact, GS is defined as a condition in which gastrointestinal and extra-intestinal symptoms are triggered by gluten consumption, in the absence of celiac-specific antibodies and villous atrophy as well as of any allergy related processes; hence, in most cases GS also called Non-Celiac Gluten Sensitivity (NCGS) (Elli et al., 2015). As aforesaid, the clinical presentation of NCGS includes gastrointestinal symptoms like abdominal pain, bloating and bowel irregularity

(diahrrea, constipation, or both), in association with systemic symptoms such the so called "foggy brain", which is described as slowed thinking, memory disturbance or reduced level of alertness, along with headache, joint and muscle pain, fatigue, depression, leg or arm numbness and skin manifestations (eczema or rash) (Leonard et al., 2017). All these symptoms resolve almost totally after gluten withdrawal (Elli et al., 2015). To date, it is still very difficult to diagnose NCGS due to several problems. Firstly, there are no clear serologic or histopathologic criteria for clinicians to confirm the diagnosis of NCGS since confirmed and validated biomarkers are yet to be identified (Elli et al., 2015; Leonard et al., 2017); secondly, symptoms experienced by individuals potentially affected by NCGS in most cases overlap to those of patients affected either by CD or irritable bowel syndrome (IBS) as shown in **Fig. 18**. As a matter of fact, it has been evidenced that NCGS occurs in approximately 30% of individuals suffering from IBS (Brouns et al., 2013) and, at the same time, Fasano et al. (2003) concluded that about 3% of patients with a clinical presentation of IBS were subsequently diagnosed with CD. Hence, NCGS is mainly a "diagnosis of exclusion" made after other wheat related and not-wheat-related disorders have been ruled out.

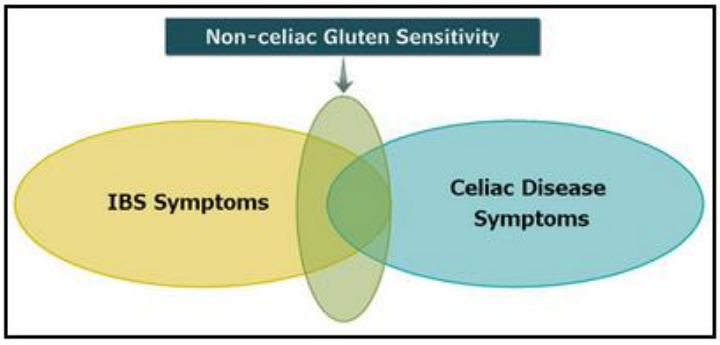

Figure 18: Non-Celiac Gluten Sensitivity (NCGS) can present symptoms that clearly overlap to both Irritable Bowel Syndrome (IBS) and Celiac Disease (CD) (https://gfjules.com/celiac-disease/).

Considering the celiac disease and non-celiac gluten sensitivity patients, two clear differences have been evidenced. Firstly, while all CD patients carry the HLA-DQ2 and HLA-DQ8 alleles, only 50% of the NCGS patients show the same genetic predisposition (Sapone et al., 2012; Elli et al., 2015); secondly, the intestinal permeability in the NCGS is significantly lower than in the CD patients (Holmes 2013). Investigating more thoroughly, it has been observed that the expression of claudin-1 and zonula occludens-1 (ZO-1) is similar in both CD and NCGS conditions, while in the latter one the expression of claudin-4 is significantly higher. Tight junction (tj) and multiple proteins that make up the tj strands, such as occludin, the claudin family and the zonula occludens-1 protein (ZO-1), are known to exert a crucial role in the development of intestinal immunological responsiveness. Moreover, the up-regulation of claudin-4 was associated with an increased expression of toll-like receptor-2 (TLR2) and with a significant reduction of T-regulatory cell marker FoxP3 relative to CD

patients. The reduced small-intestine expression of this T-regulatory cell marker in NCGS has been interpreted in the light of a reduced activation of adaptive immunity relative to CD (Volta et al., 2013). This hypothesis has also been supported by an increased expression in CD patients of different cytokines related to the adaptive immune response, condition which has not been observed in the intestinal biopsies of NCGS patients. All these features, which are used to distinguish NCGS from CD, are summarized in **Table 1**. Taking together the parameters, it is possible to state that CD and NCGS pathologies show a different intestinal mucosal response to gluten ingestion; both share an enhanced innate immune response but, while the immune mechanism underlying CD also include the activation of adaptive immunity, in patients experiencing NCGS no overexpression of adaptive immune markers has been found (Volta et al., 2013).

Table 1: Sum-up of the parameters evaluated to distinguish between celiac disease and non-celiac gluten sensitivity.

PARAMETER	CD	NCGS
Small intestine permeability (lactulose/mannitol)	Increase	Reduction
TJ proteins expression (CLDNs, TJP-1, OCLNs)	Expression's level of CLDN-4 lower than the one in the GS patients	Increase of CLDN-4; ≈ CLDN-1 and ZO-1
Intraepithelial lymphocytes	All patients have HLA-DQ2 and HLA-DQ8 Few numbers of CD3+ IELs	50% of the patients has HLA-DQ2 and/or HLA-DQ8 All have increased number of CD3+ IELs
Expression of cytokines related to the adaptive immune response	Increase of IL-17A, IFN-γ, IL-6 and IL-21	Normal expression's level of IL-17A, IFN-γ, IL-6 and IL-21
Expression of the Toll-like receptors (TLRs)	Increase of TLR1, TLR2 and TLR4	Increase of TLR2; ≈ TLR1 and TLR4

The increasing interest in investigating the pathogenesis and the mechanism underlying NCGS in order to identify and provide specific biomarkers has led to the knowledge that gluten and its proteins may not be the only triggers of NCGS since, as shown in **Fig. 19**, other wheat proteins likely play an important role in causing this disease. For example, ATIs, which are defence-related proteins against pests and parasites by inhibiting their digestive enzymes, accumulated at sufficiently high levels in the grain to also function as storage proteins (Altenbach et al., 2011). Since they appear to strongly resist intestinal degradation in humans and stimulate mucosal cytokine release after feeding *in vivo* (Brouns et al., 2013), ATIs have been suggested as strong activators of inflammation and innate immune reactions in several intestinal and non-intestinal immune disorders, including NCGS, CD and Baker's asthma (Volta et al., 2013).

In addition, many patients who experienced NCGS display also multiple food hypersensitivities, which could be partly related to a diet rich in fermentable oligo-, di- and mono-saccharides and polyols (FODMAPs) since they are poorly absorbed short-chain carbohydrates which caused distension of the intestinal limen with liquid and gas due to their small molecular size and rapid fermentability, thus leading to functional gastrointestinal symptoms (Volta et al., 2013). As a matter of fact, it has been observed that both NCGS patients and IBS patients improve their symptoms after the dietary reduction of FODMAPs (Shepherd et al., 2014). Finally, functional gastrointestinal

symptoms of both NCGS and IBS patients could also be related to food additives (glutamates, benzoates, sulfites and nitrates) which are added to almost all commercial products in order to improve flavor and color and to preserve function.

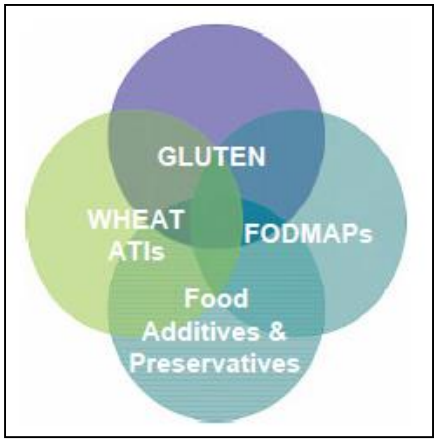

Figure 19: Known triggers of NCGS. Gluten is the primary trigger but also ATIs, FODMAPs and food additives and preservatives have been demonstrated to elicit the innate immune response leading to NCGS (Volta et al., 2013).

Chapter 2. Aim of the project

Five major cereals species (rice, bread and durum wheat, maize and barley) satisfy two thirds of the human food requirements; among these, durum wheat account for about 37 million Tons of the global cereal world's production (Kabbaj et al., 2017). Besides being one of the major source of carbohydrates, minerals and vitamins for humans, and, in addition, to providing about one-fifth of the total food calories and proteins, wheat is the best adapted crop to temperate regions and provides high yields (Gill et al., 2004). Wheat is of crucial importance, both nutritional and, consequently, economic, due to the capability to be processed in a wide range of products, like pasta, bakery products and other wheat-derived goods, thanks to its unique characteristic represented by gluten. FAO has estimated that in order to meet the global food demand in 2050, annual world production of crops and livestock will need to be 60% higher than it was in 2006 (http://www.fao.org/3/a-i6030e.pdf); moreover, about 80% percent of the required increase will need to come from higher yields on an area of land that will not increase much beyond the present level (Gill et al., 2004).

Over times, the process of domestication has led to the development, and subsequent cultivation, of few elite lines with a drastically narrowed genetic diversity: as a matter of fact, it has been estimated that, upon domestication, the initial diversity was reduced by 84% in durum wheat and by 69% in bread wheat (Jaradat, 2013). Concerning durum wheat, the process of "genetic erosion" has been mainly due to the high quality level required by customers and the uniformity of the plant ideotype pursued by breeders, coupled with the diffusion of a relatively small number of outstanding genotypes with proven adaptability and yield potential (Maccaferri et al., 2003). The increase of genetic vulnerability to new biotic and abiotic stresses, as well as to unpredictable environmental conditions, has occurred as a consequence of the narrowing genetic base of the current elite germplasm cultivated (Wang et al., 2017). Moreover, changes in global dietary habits, according to which many countries are experiencing a progressive westernization of diet as well as worldwide diffusion of the Mediterranean diet, which is based on a large number of foods that incorporate gluten, has led to an increasing prevalence of gluten-related disorders (Tovoli et al., 2015). In particular, consumption of wheat is progressively replacing consumption of rice in many countries in North Africa, the Middle East and Asia (Catassi and Cobellis, 2007).

Hence, the need to identify, select and develop new wheat varieties with an increased yield and higher green management on one side, with high technological quality of the relative flours and healthier impact on humans on the other side, is mandatory for researchers, wheat breeders and farmers alike (Carver, 2009).

In the light of these considerations, two main strategies can be adopted for the establishment of new durum wheat varieties with the aforementioned characteristics. On the one side, novel biotechnologies such as genome editing allow direct transfer of discovered beneficial genes or gene complexes into an elite genetic background, or manipulation of existing genes, in a very rapid and efficient way to obtain expected phenotypes without lengthy backcrossing; on the other side, it could be useful to apply a classical breeding approach exploiting, for example, the reservoir of genetic variability present in the germplasm collections. Landraces, represent a rich source of untapped diversity which, on the one side can be potentially exploited for desired traits and used in breeding programs, and on the other store the traces of men-crops coevolution.

In the present project, an approach based on the exploitation of the available biodiversity was chosen with the aim to characterize for the qualitative, technological and toxicological traits, a subset a 152 genotypes derived by a worldwide germplasm collection of durum wheat, selected by Single Seed Descent (SSD) (Pignone et al., 2015) and, in particular, to identify possible genotypes which could be potentially exploited in future breeding programs.

All samples were investigated with respect to the High Molecular Weight Glutenin Subunits (HMW-GS) composition, since it is well-known that these proteins exert a crucial quality role. Preliminary tests were carried out on a broad set of reference cultivars.

A combined “Omic" approach, in which traditional SDS-PAGE assays in different experimental conditions, at proteomic level, were coupled to the usage of PCR-based molecular markers, at genomic level, was applied in order to precisely identify the *Glu-1* allelic composition.

While making this characterization, the relationship between the glutenin composition and the relative provenance of the germplasm lines was assessed in order to trace wheat evolutionary history, as intended as the history of its diffusion and diversification, since it has been observed that the geographic distribution of the present genetic diversity within a species can be informative about the historical processes that shaped that diversity (Wang et al., 2017). Moreover, a selection of the SSD genotypes within the 152 considered, representative of both common and rare HMW-GS pattern retrieved, has undergone an *in vitro* static digestion to investigate the presence, and the relative amount, of peptides associated with celiac disease.

Chapter 3. Materials and Methods

3.1 Plant material

A subset of 152 genotypes of the SSD core collection described by Pignone et al. (2015) and representatives of 31 countries was kindly provided in frame of the collaboration with the Institute of Bioscience and Bioresources, National Research Council (IBBR-CNR) of Bari (http://ibbr.cnr.it/mgd/) (**Fig. 20**).

In addition, a set of 43 between durum and bread elite cultivars have been selected on the basis of their HMW-GS allelic composition (Wrigley et al., 2006; Fichera et al., 2006; De Vita et al., 2007; Ribeiro et al., 2011; Comastri et al., 2013), and used as standards for the identification of the HMW-GS pattern in the SSD genotypes.

Figure 20: Provenance of the 152 durum wheat landraces analyzed.

3.2 HMW-GS protein extraction

HMW-GS proteins were extracted from the flour of each genotype (SSD and standard cultivars) according the procedure indicated by Singh et al. (1991) which is specific for prolamins, with some modifications (Visioli et al., 2015). This protocol is based on a sequential extraction which exploits

the different solubility of the proteins of interest in specific reagents, thus leading to their separation into fractions (**Fig. 21**). The first step allows the separation of the gliadin fraction from the glutenin fraction and, from the latter one, are then obtained the two enriched fractions of high molecular weight (HMW-GS) and low molecular weight (LMW-GS) components.

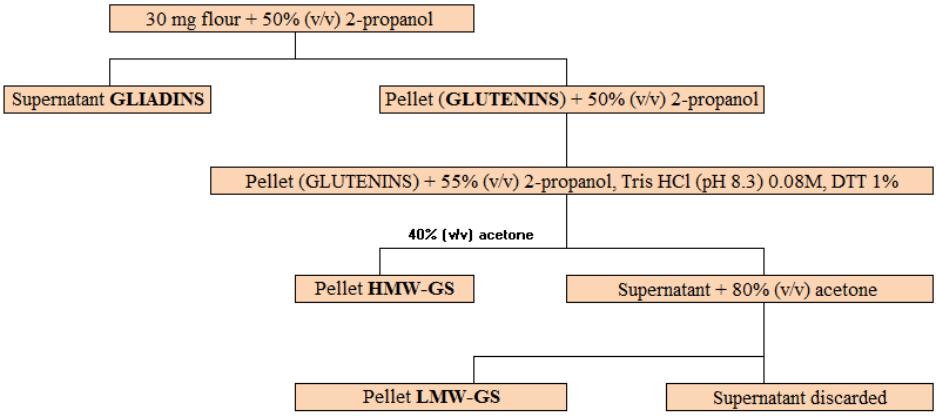

Figure 21: Scheme of the sequential extraction followed to separate the different prolamin fractions.

To remove the gliadin component, fine flour (30 mg) was added by 1.5 mL of 50% (*v/v*) 2-propanol for 20 minutes with continuous mixing at 65°C, following by centrifugation for 5 minutes at 10000 RPM. This step was repeated two more times discarding the supernatant and adding the pellet by a lower volume 2-propanol, from 1.5 mL to 1 mL. At the end of the third time, the pellet, containing the whole GS fraction, was re-suspended into 310 µL of a solution of 55% (*v/v*) 2-propanol, 0,08 M Tris(hydroxymethyl)aminomethane hydrochloric acid (Tris-HCl) pH 8.3 and 1% (*w/v*) dithiothreitol (DTT) as reducing agent, incubated for 30 minutes with continuous mixing at 60 °C and finally centrifuged for 5 minutes at 14000 RPM. Finally, 300 µL of the supernatant, containing both HMW-GS and LMW-GS, was transferred into a new tube, added by 40% (*v/v*) of acetone to precipitate HMW-GS component and centrifuged for 10 minutes at 13000 RPM. The pellet was dried under hood for 15-30 min and retained at -20° C.

3.3 Crude protein extract quantification

The protein content was determined by using the Bradford assay (Bradford, 1976), a colorimetric assay which is based on a shift of absorbance of the Comassie Brilliant Blue G-250 dye; under acidic conditions, the red form of the dye is converted to its blue one in order to bind the assayed protein. The tied form protein-dye has the maximum absorption at 595 nm so that measures are done

at this wavelength; the increase in the absorbance is proportional to the amount of the bound dye and, therefore, to the amount (concentration) of the protein present in the sample.

Protein extracts, containing the HMW-GS fraction, were re-suspended in 200 µL of a solution of (50:50 *v/v*) acetonitrile (ACN) and H_2O with 0.1 % (*v/v*) trifluoroacetic acid (TFA) and then incubated at 65 °C with continuous mixing until the complete solubilization. Following Sigma's manufacturing procedures relative to the utilization of the Bradford reagent, 20 µL of each sample were mixed with 1 mL of the Bradford reagent for 5 minutes at room temperature; samples were then measured through the Varian Cary® 50 Scan Uv-Visible Spectrophotometer at 595 nm. A calibration curve was constructed, by determining the absorbance of standard samples of Serum Albumin Bovine (BSA) at a known concentration, in order to determine sample's concentration. Seven dilutions of BSA, corresponding to the following concentrations: 0.125-, 0.250-, 0.5-, 0.75-, 1.0-, 1.5- and 2.0 mg, were prepared from a stock of 20 mg/mL. After this step of quantifications, samples were dried up in Savant™ SpeedVac™ SPD1010 (Thermo Fisher Scientific, Waltham, MA, USA) at 45 °C and used for SDS-PAGE analyses.

3.4 Electrophoresis

3.4.1 SDS-PAGE analysis

The following solutions were prepared to set up the one-dimensional SDS-PAGE:

- **Laemmli SDS electrophoresis buffer 10X**:
  - 250 mM Tris base
  - 1.92 M Glycine
  - 1% SDS

- **SDS gel loading buffer (LB) 5X**:
  - 250 mM Tris HCl pH 6.8
  - 50% Glycerol
  - 10% SDS
  - 0.1% Bromophenol blue
  - 1:20 β-mercaptoethanol (add fresh to the solution 1X)

- **APS solution 10%**:

- 10% SDS Ammonium persulfate

- **Running gel 12%** (to a final volume of 5 mL):
  - 40% Acrylamide/Bisacrylamide 37.5:1 1.5 mL
  - 1.5 M Tris Hcl pH 8.8 1.3 mL
  - 10% SDS 50 µL
  - 10% APS 50 µL
  - TEMED 3.5 µL
  - Sterile deionized H_2O 2.0965 mL

- **Stacking gel 5%** (to a final volume of 3 mL):
  - 40% Acrylamide/Bisacrylamide 37.5:1 375 µL
  - 1.5 M Tris Hcl pH 6.8 254 µL
  - 10% SDS 30 µL
  - 10% APS 30 µL
  - TEMED 3 µL
  - Sterile deionized H_2O 2.308 mL

- **Fixing solution**:
  - 40% Methanol
  - 7% Acetic acid

The electrophoresis of reduced proteins was performed on Criterion™ Dodeca™ Cell (Biorad, Hercules, CA) by using 7.5% Criterion™ TGX™ Precast Midi Protein Gel (Biorad, Hercules, CA). Dried HMW-GS (2,5 µg) of both SSD genotypes and standard cultivars were re-suspended in different volumes of loading buffer (LB 1X), depending on the corresponding quantification, and boiled 5 minutes at 95 °C, in order to completely denature proteins before loading onto the gels. Unstained SDS-PAGE Molecular Weight Standards, Low Range (Mw 14,400-97,400) Biorad (Hercules, CA) was used for tracing HMW-GS. Electrophoretic separation was carried out using the current resistor Electrophoresis Power Supply EPS_{250} (C.B.S. Scientific Company INC.; Del Mar, California) at 140 V and 200 V in the stacking and running gel respectively, and at 40 mA. The running was stopped as soon as the dye was spilled from gels. Gels were then fixed in 7% (*v/v*) acetic acid and 40% (*v/v*) methanol, stained with Brilliant blue G-colloidal solution (Sigma Aldrich, Milano, Italy) over night and destained in deionized water. Gels were acquired through ChemiDoc™ MP

System (Bio-Rad) fitted in the laboratory of SITEIA.PARMA (Centro Interdipartimentale sulla Sicurezza, Tecnologie e Innovazione Agroalimentare) at Tecnopolo of the University of Parma, and visually analyzed with Image Lab Software 4.0 (Bio-Rad®) for allelic identification using reference varieties.

The electrophoresis analysis was also performed on 12% acrylamide gels, according to Ribeiro et al. (2011), in order to better define the *Glu-B1* subunits. The analysis procedure, in this specific case, consisted of several steps:

- Clamp two glass plates in the casting frames in order to make the casting system;
- Prepare and pour the running gel, in the desired percentage of acrylamide. Add a little volume of cold 2-propanol in order to level gels and let them polymerize for at least 30 minutes;
- Discard the 2-propanol and pour the stacking gels. Insert the well-forming comb and let them polymerize for 15 minutes;
- Prepare samples as previously described before loading them onto gels.

The separation was carried out using the current resistor Electrophoresis Power Supply EPS_{250} (C.B.S. Scientific Company INC.; Del Mar, California) at 140 V for both stacking and running gel and at 40 mA but, unlike the first SDS-PAGE analysis, the running was extended for three hours after the spillage of the dye. Gels were fixed in 7% (*v/v*) acetic acid and 40% (*v/v*) methanol, stained in Coomassie blue R-250 (Biorad, Hercules, CA) over night, destained in deionized water and visually analyzed for allelic identification using reference varieties.

3.5 Genomic DNA extraction

3.5.1 Dellaporta DNA extraction

DNA of the standard cultivars was extracted following Dellaporta protocols with slight modifications (D'Ovidio and Porceddu, 1996). To each sample 612.5 µL of extraction buffer (EB) + 87.5 µL of 10% SDS solution were added.

- **Extraction buffer** (to a final volume of 50 mL):
- 100 mM Tris pH 8
- 5 mL of Tris 1M
- 50 mM EDTA pH 8
- 5 mL of EDTA 0,5 M

- 500 mM NaCl
- 5 mL of NaCl 5M
- 10% SDS

A third of a seed for each genotype (~ 30 mg) was grinded to a very fine powder, by using glossy paper and pestle, and transferred into a 1.5 mL tube. 250 µL out of the 700 µL of the mix EB + SDS solution was added to each tube and homogenized with a little pestle. The remaining 450 µL of the EB + SDS solution was then added, and samples were incubated at 65 °C for 30 minutes, manually mixing every 10 minutes to promote the lysis. At the end of the incubation, 200 µL of Potassium acetate 5M were added; tubes were shaken and placed into ice for at least 30 minutes (maximum one hour) in order to precipitate proteins, polysaccharides and cellular debris. Samples were then centrifuged for 20 minutes at 4 °C and at maximum speed (14000 RPM); supernatants, containing the DNA, were transferred into new sterile tubes and added by 500 µL of 2-propanol at -20 °C. Tubes were shaken to promote the DNA precipitation and centrifuged for 20 minutes at 4 °C and at maximum speed (14000 RPM). Supernatants were discarded, while pellets were washed twice with 400 µL of 70% ethanol at -20 °C. For each washing step, tubes were centrifuged for 5 minutes at maximum speed. Pellets were finally dried under laminar flow hood for 15-20 minutes, re-suspended into 100 µL of sterile deionized water at 50 °C and preserved at -20 °C.

3.5.2 GenElute™ Plant Genomic DNA Miniprep Kit

Sigma's GenElute™ Plant Genomic DNA Miniprep Kit (cat. G2N350) consists of several steps and provides a simple and convenient way to isolate pure DNA from a variety of plant species. It allows to obtain several micrograms of DNA from up to 100 mg of fresh tissue or 10 mg of freeze-dried material rapidly (~ 1 hour). Reagents and equipments provided from the kit are:

- Lysis Solution Part A
- Lysis Solution Part B
- Precipitation Solution
- Binding Solution
- Column Preparation Solution
- Wash Solution Concentrate (it needs to be added of ethanol)
- Elution Solution (10 mM Tris, 1mM EDTA, pH approx 8.0)
- GenElute Filtration Column s in Tubes
- GenElute Nucleic Acid Binding Columns in Tubes
- Collection Tubes, 2 mL capacity.

DNAs of the 152 SSD genotypes considered in the project were extracted using this kit and following steps provided from manufacture procedures:

1) **Disruption of cells:** plant tissue of each of the 152 SSD genotypes was grinded into a fine powder in liquid nitrogen by using 1.5 mL tubes and little pestles, keeping samples on ice for their immediate use;
2) **Lysis of cells:** each tube was added by 350 µL of Lysis Solution A and 50 µL of Lysis Solution B. Tubes were mixed by vortexing and inverting and, then, mixtures were incubated at 65 °C for 10 minutes with occasionally inversion to dissolve the precipitate;
3) **Precipitation of debris:** at the end of the incubation, 130 µL of Precipitation Solution were added to each mixture. Tubes were completely mixed by inversion, placed on ice for 5 minutes and then centrifuged at maximum speed (14000 RPM) for 5 minutes to pellet the cellular debris, proteins and polysaccharides;
4) **Filtration of debris:** supernatants from step 3 were carefully pipetted onto GenElute filtration columns, which were then centrifuged at maximum speed (14000 RPM) for 1 minute in order to remove any cellular debris not removed before. Each filtration column was discarded, while each corresponding tube was retained;
5) **Preparation for the binding:** 700 µL of Binding Solution were added directly to the flow-through liquid from step 4 and columns were mixed thoroughly by inversion;
6) **Preparation of binding column:** GenElute Miniprep Binding Columns were inserted into the provided microcentrifuge tubes, added by 500 µL of the Column Preparation Solution and centrifuged at 12000 RPM for 1 minute. The flow-through liquid was discarded;
7) **Loading of lysate:** 700 µL of the mixtures from step 5 were carefully pipetted into the columns prepared in step 6 and centrifuged at maximum speed (14000 RPM) for 1 minute. The flow-through liquid was discarded and the same procedure was repeated until loading the total lysate from step 5;
8) **First column wash:** the binding columns from step 7 were placed into fresh 2 mL collection tubes, added by 500 µL of the diluted Wash Solution and centrifuged at maximum speed (14000 RPM) for 1 minute. The flow-through liquid was discarded, while the collection tube and the column were retained;
9) **Second column wash:** another 500 µL of diluted Wash Solution were applied to the columns, which were then centrifuged at maximum speed (14000 RPM) for 3 minutes to dry them;
10) **Elution of DNA:** binding columns from step 9 were transferred into fresh 2 mL collection tubes, 100 µL of pre-warmed sterile deionized water were added and columns were centrifuged at maximum speed for 1 minute.

Note: a second elution has been done following the same procedure and using new fresh collection tubes.

Eluted, containing pure genomic DNA, were stored at 4°.

3.5.2.1 DNA quantification

Both the first and the second elution of the SSD genomic DNAs, obtained by using Sigma's GenElute^{TM} Plant Genomic DNA Miniprep Kit, were quantified through electrophoresis.
Agarose gels were prepared at a 1.2% (*w/v*) concentration of agarose in Tris-acetate-EDTA (TAE) buffer 1X. The solution was heated in the microwave oven to completely dissolve the agarose and 10 µL of the Gel Red dye were added before pouring the solution into the tray. 1 µL of each genomic DNA was loaded onto gels and standard samples of 200 ng, 100 ng and 50 ng of genomic DNA from phage λ were used as reference.

3.6 Molecular markers assay for *Glu-A1* and *Glu-B1*

A total of 10 primers pairs were selected from literature to further discriminate between HMW-GS whose electrophoretic profile on SDS-PAGE was ambiguous. Out of these, four primer pairs were useful to further investigate the HMW-GS composition of the SSD subset at *Glu-A1* locus, while the remaining six allowed a deeper insight in the HMW-GS composition at *Glu-B1* locus. The complete set of molecular markers used and their features are listed in **Table 2**.

Table 2: List of molecular markers exploited to better characterize the HMW-GS composition of the 152 SSD genotypes. Alleles discriminated, marker type, molecular marker profile obtained, forward and reverse primers sequences and the literature's references are indicated.

Locus *Glu-1*	Primer pairs	Alleles	Marker type	Molecular Marker profile[a]	Forward and revers primers sequences (5'-3')	References
Glu-A1	Ax_F/ Ax2*_R	Ax1 Ax2*	dominant	no bands 1 band (1319 bp)	ATG ACT AAG CGG TTG GTT CTT GAC CTT GCT CCC CTT GCT TTT	Ma et al. 2003
	Ax1/Ax2*_C Ax1/Ax2*_D	Ax1 Ax2*	co-dominant	1 band (1500 bp) 1 band (1400 bp)	CCA TCG AAA TGG CTA AGC GG GTC CAG AAG TTG GGA AGT GC	Lafiandra et al. 1997
	UMN19_F/R	Ax*Null* Ax1 Ax2*	co-dominant	1 band (362 bp) 1 band (362 bp) 1 band (344 bp)	CGAGACAATATGAGCAGCAAG CTGCCATGGAGAAGTTGGA	Liu et al. 2008
	Ax*Null*_G1_F Ax*Null*_G2_R	Ax*Null* Ax1 Ax2*	dominant	1 band (920 bp) no bands no bands	ACG TTC CCC TAC AGG TAC TA TAT CAC TGG CTA GCC GAC AA	Lafiandra et al. 1997
Glu-B1	Bx_F Bx_R	Bx17 Bx7* no Bx17	co-dominant	1 band (669 bp) 2 bands (650, 750 bp) 2 bands (670, 770 bp)	CGC AAC AGC CAG GAC AAT T AGA GTT CTA TCA CTG CCT GGT	Ma et al. 2003
	ZSBy8_F5 ZSBy8_R5	By8 By8*,By9,By15,By16, By18,By20	dominant	1 band (527 bp) no bands	TTA GCG CTA AGT GCC GTC T TTG TCC TAT TTG CTG CCC TT	Lei et al. 2006
	ZSBy9_F2 ZSBy9_R2	By16 By20 By8,By8*,By9,By15,By18 By18/By26	co-dominant	3 bands (280, 350, 400 bp) no bands 2 bands (280, 350 bp) 1 band (280bp)	GCA GTA CCC AGC TTC TCA A CCT TGT CTT GTT TGT TGC C	Lei et al. 2006; Salmanowicz and Dylewicz 2007
	ZSBy9_aF1 ZSBy9_aR3	By9 By8,By8*,By15,By16, By18,By20	co-dominant	1 band (662 bp) 1 band (707 bp)	TTC TCT GCA TCA GTC AGG A AGA GAA GCT GTG TAA TGC C	Lei et al. 2006
	CauBx752	Bx14 Bx17	co-dominant	1 band (752 bp) 1 band (337 bp)	AGG GGC AGG GAA GAA ACA CT CCA GGC AAC ACA AAT CCA TG	Xu et al. 2008
	Bx7_F Bx7_R	Bx6 Bx7,Bx17	co-dominant	1 band (250 bp) 1 band (220 bp)	CAC TGA GAT GGC TAA GCG CC GCC TTG GAC GGC ACC ACA GG	Schwarz et al. 2004

3.6.1 PCR conditions and electrophoresis analysis

The optimal PCR conditions and the best electrophoresis conditions were tested for each primer pair, according to those reported in literature.

PCR amplifications were performed using the Applied Biosystems™ Veriti™ 96-Well Thermal Cycler. All the reactions were carried out in a final volume of 20 µL, containing GoTaq® Hot Start Master Mix 1X (Promega, WI, USA), 6 pmoles of each primer (Metabion international AG, Planegg/Steinkirchen, Germany; Eurofins Genomics, Ebersberg, Germany) and 1 µL of template DNA. PCRs were set up following the indications provided in the reference papers for each molecular marker (Lafiandra et al., 1997; Ma et al., 2003; Schwarz et al., 2004; Lei et al., 2006; Xu et al., 2008; Liu et al., 2008), sometimes making slight modifications in the thermal cycles or in the melting temperatures.

The amplification products were fractionated by horizontal gel electrophoresis in Tris-Acetate-EDTA (buffer TAE 1X) agarose gel. Depending on the size of the amplicons and in order to obtain the optimal discrimination between bands, gels' concentrations were the following:

- 1.2% for Ax_F/Ax2*_R, Ax1/Ax2*_C/D, Ax*Null*_G1_F/G2_R and UMN19_F/R molecular markers;
- 1.5% for Bx_F/Bx_R, ZSBy8_F5/R5 and ZSBy9_F2/R2 molecular markers;

- 2% for ZSBy9_aF1/aR3 and CauBx752 molecular markers;
- 2.5% for Bx7_F/R molecular marker.

AccuRuler 100 bp DNA RTU ladder (Maestrogen Inc., Hsinchu City 30091, Taiwan), NZYDNA laddr VI (nzytech genes & enzymes, Lisbon, Portugal) and AccuRuler 1 kb DNA RTU ladder (Maestrogen Inc., Hsinchu City 30091, Taiwan) were used as references depending on the amplicon's size.

3.7 Purification and sequencing

3.7.1 DNA purification from gel

PCR products were purified with the NZYGelpure kit (NZYtech, Lisbon, Portugal).
All purification steps and all centrifugations have been carried out at room temperature. In this case the protocol follows ten steps:

1) Excise the DNA fragment from the gel with a clean, sharp scalpel; weigh the gel slice and transfer into a 1.5 mL microcentrifuge tube;
2) Add 300 µL of Binding Buffer for each 100 mg of gel weight;
3) Incubate the tube at 60 °C for 10 min occasionally shaking until agarose is completely dissolved;
4) Add 1 gel volume of 2-propanol to the sample mixing well by pipetting several times;
5) Load the above mixture into the NZYTech spin column placed into a 2 mL collection tube. Centrifuge for 1 minute at maximum speed and discard the flow-through in the collection tube;
6) Add 500 µL of Wash Buffer and centrifuge for 1 minute at maximum speed; the flow-through in the collection tube has to be discarded;
7) Add 600 µL of Wash Buffer and centrifuge for 1 minute at maximum speed; the flow-through in the collection tube has to be discarded again;
8) Centrifuge for 1 minute again to dry the NZYTech spin membrane of residual ethanol;
9) Place the NZYTech spin column into a clean 1.5 mL microcentrifuge tube, add 50 µL of hot deionized water to the centre of the column and incubate at room temperature for 1 minutes. Centrifuge for 1 minute at maximum speed to elute DNA.
10) Store the purified DNA at -20 °C.

3.7.2 DNA purification from PCR solution

All purification steps and all centrifugations should be carried out at room temperature. In this case the protocol follows six steps:

1) Transfer the volume of the reaction mixture into a 1.5 mL microcentrifuge tube and add five volumes of Binding Buffer. Mix by inverting the tube few times and then centrifuge briefly to collect the sample;

2) Add the above mixture to the NZYTech spin column (the maximum loading volume is 700 µL) and centrifuge for 1 minute at maximum speed, then discard the flow-through in the collection tube;

3) Add 600 µL of Wash Buffer and centrifuge for 1 minute at maximum speed; the flow-through in the collection tube has to be discarded again;

4) Centrifuge for 1 minute again to dry the NZYTech spin membrane of residual ethanol;

5) Place the NZYTech spin column into a clean 1.5 mL microcentrifuge tube, add 50 µL of hot deionized water to the centre of the column and incubate at room temperature for 1 minutes. Centrifuge for 1 minute at maximum speed to elute DNA.

6) Store the purified DNA at -20 °C.

3.7.3 Sequencing

An amount of purified PCR fragments corresponding to 100-800 ng were prepared for sequencing using both the GATC (Biotech AG, Cologne, Germany) or Eurofins (Eurofins Genomics, Ebersberg, Germany) sequencing service. The obtained sequences were analyzed using the DNAMan® software (Lynnon Biosoft, Quebec, Canada) and aligned with the available annotated HMW-GS sequences.

3.8 Statistical analysis

Genetic variation at each locus was calculated using the Nei index (Nei, 1973), H, given by the expression $[1 - \sum p_{ij}^2]$, where p_{ij} represented the frequency of the i^{th} allele at the j^{th} locus. Allelic frequencies within the panel were determined from that of the alleles in the individual accessions, and then dividing by the number of accessions (152) (Rasheed et al., 2012).

A hierarchical clustering analysis was performed with software Past 3.18 (Hammer et al., 2001). The clustering method was an unweighted pair-group average (UPMGA), in which clusters are joined based on the average distance between all members in the two groups. The same software package

was used to assess the population structure by means of principal component analysis (PCA), which finds the eigenvalues and eigenvectors of the variance-covariance matrix or the correlation matrix with the Singular Value Decomposition (SVD) algorithm. Clusters generated through both approaches were manually compared by searching for a correlation between genotypes within each cluster.

3.9 *In vitro* digestion of selected SSD sample

A total of 40 SSD genotypes were selected on the basis of their allelic composition for the identification and quantification of gluten peptides (immunogenic and toxic) associated with celiac disease. The selection considered the genotypes carrying alleles commonly known as related to good technological quality (Bx7+By8, Bx6+By8, Bx13+By16) but also the rarest one for which the immunogenic and toxicity properties were not explored so far in the literature. The analyses were performed in collaboration with Professor Stefano Sforza (Food Chemistry Lab., Food and Drug Department, University of Parma). Data were statistically analyzed by using the ANOVA and through principal components analyses (PCA).

3.9.1 Standardized static *in vitro* digestion method

2 g of all SSD samples were grounded for about 1 minute in order to obtain a fine dust. Two replicates for each sample were digested following the standard *in vitro* method of Minekus et al. (2014). Briefly, 0.5 g of the ground wheat sample was incubated 2 min with 0.5 mL simulated saliva containing amylase (75 U/mL of digesta) (SSF, **Table 3**); then, 1 mL of simulated gastric juice containing pepsin (2000 U/mL of digesta) (SGF, **Table 3**) were added and the sample was incubated for 2 h after adjusting the pH to 3. Subsequently, 2 mL of duodenal juice containing pancreatin (100 U trypsin activity/mL of digesta) and bile (10 mmol/L in the total digesta) (SIF, **Table 3**) were added and incubated for 2 h after adjusting the pH to 7. All the digestion steps were carried out at 37 °C under constant gentle mixing. Then, to inactivate the enzymes, the sample was boiled for 10 min at 95°C. After centrifugation (3220*g*, 4°C, 45 min), 295 µL of each sample supernatant was added to 5 µL of internal standard solution (TQQPQQPF(d_5)PQQPQQPF(d_5)PQ, 1.6 mmol L^{-1}).

Table 3: Preparation of stock solutions of simulated digestion fluids. The volumes are calculated for a final volume of 500 mL for each simulated fluid (Minekus et al., 2014).

			SSF		SGF		SIF	
			pH 7		pH 3		pH 7	
Constituent	Stock conc.		Vol. of stock	Conc. in SSF	Vol. of stock	Conc. in SGF	Vol. of stock	Conc. in SIF
	$g\ L^{-1}$	$mol\ L^{-1}$	mL	$mmol\ L^{-1}$	mL	$mmol\ L^{-1}$	mL	$mmol\ L^{-1}$
KCl	37.3	0.5	15.1	15.1	6.9	6.9	6.8	6.8
KH_2PO_4	68	0.5	3.7	3.7	0.9	0.9	0.8	0.8
$NaHCO_3$	84	1	6.8	13.6	12.5	25	42.5	85
NaCl	117	2	—	—	11.8	47.2	9.6	38.4
$MgCl_2(H_2O)_6$	30.5	0.15	0.5	0.15	0.4	0.1	1.1	0.33
$(NH_4)_2CO_3$	48	0.5	0.06	0.06	0.5	0.5	—	—
For pH adjustment								
	$mol\ L^{-1}$		mL	$mmol\ L^{-1}$	mL	$mmol\ L^{-1}$	mL	$mmol\ L^{-1}$
NaOH	1		—	—	—	—	—	—
HCl	6		0.09	1.1	1.3	15.6	0.7	8.4
$CaCl_2(H_2O)_2$ is not added to the simulated digestion fluids, see details in legend								
	$g\ L^{-1}$	$mol\ L^{-1}$		$mmol\ L^{-1}$		$mmol\ L^{-1}$		$mmol\ L^{-1}$
$CaCl_2(H_2O)_2$	44.1	0.3		1.5 (0.75*)		0.15 (0.075*)		0.6 (0.3*)

[a] * in brackets is the corresponding Ca^{2+} concentration in the final digestion mixture.

3.9.2 Ultra-Performance Liquid Chromatography–Electrospray Ionization Mass Spectrometry (UPLC/ESI-MS) analysis

UPLC/ESI-MS analysis was performed accordingly to Prandi et al. (2014). Briefly, the complex mixture obtained from enzymatic cleavage is separated by a RP column (ACQUITY UPLC BEH 300, C18, 1.7 mm, 2.1*150 mm; Waters corp., Milford, MA, USA) in a UPLC/ESI-MS system (Acquity Ultra- performance UPLC with a single quadrupole mass spectrometer; Waters SQD) using a gradient elution. Eluent A is a bi-distilled water solution with 0.1% formic acid (>99%) and acetonitrile (0.2%), and eluent B is an acetonitrile solution with 0.1% formic acid (>99%). Gradient elution was carried out as follows: 0-7 min 100% eluent A; 7-50 min from 100% to 50% eluent A; 50-52.6 min 50% eluent A; 52.6-53 min from 50% to 0% eluent A; 53-58.2 min 0% eluent A; 58.2-59 min from 0% to 100% eluent A; 59-72 min 100% eluent A. The samples are analyzed with UPLC/ ESI-MS in the Full Scan mode. Flow is 0.2 mL/min; analysis time 72 min; column temperature 35°C; sample temperature 18°C; injection volume 2 µL; acquisition time 7-58.2 min; ionization type is positive ions; scan range 100-2000 m/z; capillary voltage 3.2 kV; cone voltage 30 V; source temperature 150°C; desolvation temperature 300°C; cone gas flow 100 l/h; desolvation gas flow 650 l/h.

3.9.3 Data processing

Considering that the first 7 min of chromatographic run were excluded, due to their abundance in salt sugars, the chromatogram might be subdivided into two phases: the first one (from 12 to 35 min), where peptides ranging from Mr 200 to 3600 are eluted, and the second one (after 35 min), where bile and salt are eluted.

The areas of the identified peptides and internal standard TQQPQQPF(d_5)PQQPQQPF(d_5)PQ were integrated with the MassLynx software (Waters, Milford, MA, U.S.A.). The quantification value was obtained as the ratio peptide area/internal standard area multiplied by the moles of internal standard, assuming a response factor of 1. The result is reported on g of whole wheat flour considering the different dilution factors. The identified gluten epitopes were subdivided into two groups: immunogenic peptides (sum of the amounts of the 7 identified immunogenic peptides obtained) and toxic peptides (sum of the amounts of the 3 identified toxic peptides obtained). The sum of immunogenic and toxic peptides is called the total immunogenic-toxic peptides.

3.10 Statistical analysis

For all the peptides identified, the analysis of variance (ANOVA) was performed. Principal component analysis (PCA) was performed based on correlation matrix. The first two principal components were graphically represented in bi-plots. All experimental data were statistically analyzed using the software Past 3.18 (Hammer et al., 2001).

Chapter 4. Results and Discussions

4.1 HMW-GS characterization of standard cultivars

A total of 43 between durum and bread cultivars have been selected, due to their allelic composition at the *Glu-B1* locus, and used as standards for the identification of the HMW-GS patterns in the pool of SSD genotypes. At least 3 cultivars sharing the same allelic combination have been assessed in order to produce more reliable results, with the exception for those carrying the rarest HMW-GS pattern (i.e. GSs Bx7+By15, Bx13+By19, Bx14+By19, Bx23+By18) for whom it has been difficult identifying more than one. The HMW-GS characterization was obtained through a multiple "Omic" approach. In particular, the SDS-PAGE was associated, for those cases where ambiguity remained, to a PCR markers-based approach (ANNEX A, Janni et al., 2017). The complete list of cultivars, their HMW-GS allelic combinations and the institute which kindly provided the seed material, are reported in **Table 4**.

Table 4: List of cultivars used in the assays. The species, the corresponding allelic combination, and the provider institute are specified.

Cultivars name	*Triticum* species	HMW-GS allelic combination			Provider Institute
		Glu-A1	*Glu-B1*	*Glu-D1*	
Abano*&	*Ta*	*null*	Bx7+By9	Dx2+Dy12	CREA, S. Angelo Lodigiano (Lo)
Adraino*&	*Ta*	Ax1	Bx7+By9	Dx5+Dy10	
Apulia&	*Ta*	Ax1	Bx7	Dx2+Dy12	
Argelato*&	*Ta*	*null*	Bx7	Dx2+Dy12	
Arquà&	*Ta*	Ax2*	Bx17+By18	Dx2+Dy12	
Carme*&	*Ta*	Ax1	By18*	Dx2+Dy12	
David*&	*Ta*	Ax1	By18*	Dx1+Dy12	
Emilio Morandi*&	*Ta*	*null*	Bx7*+By8	Dx2+Dy12	
Est Mottin*&	*Ta*	Ax1	Bx6*+By8*	Dx2+Dy12	
Fabiola*	*Ta*	Ax1	Bx17+By18	Dx2+Dy12	
Francia*&	*Ta*	Ax2*	Bx17+By18	Dx2+Dy12	
Firenze*&	*Ta*	*null*	Bx7*+By8	Dx2+Dy12	
Sieve*&	*Ta*	Ax1	Bx6*+By8*	Dx2+Dy12	
Tudest*	*Ta*	*null*	Bx6*+By8*	Dx2+Dy12*	
Virest*	*Ta*	Ax1	Bx6*+By8*	Dx2+Dy12	
Tiberio*	*Ta*	Ax1	Bx7*+By8	Dx5+Dy10	
Abbazia*	*Ta*	Ax1	Bx7*+By8	Dx2+Dy12	
Garibaldino*	*Ta*	Ax1	By18*	Dx2+Dy12	
Isa*&	*Td*	*null*	Bx13+By16	/	CREA, Foggia
Solitario*&	*Td*	*null*	Bx13+By16	/	

&: cultivars analyzed with the proteomic approach
*: cultivars analyzed with the molecular approach

Table 4: Continued.

Cultivars name	*Triticum* species	HMW-GS allelic combination			Provider Institute
		Glu-A1	*Glu-B1*	*Glu-D1*	
Bagudo[&]	*Td*	Ax1	Bx23+By18	/	Univerity of Tràs-os-Montes e Alto Douro, Lisbon, Portugal
Durazio Rijo*[&]	*Td*	*Null*	Bx14+By19	/	
Lobeiro[&]	*Td*	Ax1	Bx13+By19	/	
Marques[&]	*Td*	*Null*	Bx7+By15	/	
Vermelho Fino[&]	*Td*	Ax1	Bx21	/	
Ciccio*[&]	*Td*	*Null*	Bx7+By8	/	IBBR-CNR, Bari
Saragolla*[&]	*Td*	*Null*	Bx6+By8	/	
Svevo [&]	*Td*	*Null*	Bx7+By8	/	
Simeto[&]	*Td*	*Null*	Bx7+By8	/	
Cham*[&]	*Td*	*Null*	Bx7+By8	/	University of Tuscia, Viterbo
Cheyenne*[&]	*Ta*	Ax2*	Bx7*+By9	Dx5+Dy10	
Cadenza*[&]	*Ta*	*Null*	Bx14+By15	Dx5+Dy10	Rothamsted Experimental Station, Harpended, UK
Florida*[&]	*Ta*	Ax1	Bx6+By8	Dx5+Dy10	
Cappelli*	*Td*	*Null*	Bx20+By20	/	Alsia Metapontum Agrobios
Colosseo*[&]	*Td*	*Null*	Bx14+By15	/	University of Bologna
Biensur[&]	*Td*	*Null*	Bx7+By8	/	University of Parma
Appulo*	*Ta*	*Null*	Bx20+By20	*null*	
Capeiti*[&]	*Td*	*Null*	Bx20+By20	/	
Chinese Spring*	*Ta*	*Null*	Bx7+By8	Dx2+Dy12	
Creso*[&]	*Td*	*Null*	Bx6+By8	/	
Dylan*	*Td*	*Null*	Bx6+By8	/	
Liberdur[&]	*Td*	*Null*	Bx20+By20	/	
Ofanto*[&]	*Td*	*Null*	Bx20+By20	/	

[&]: cultivars analyzed with the proteomic approach
*: cultivars analyzed with the molecular approach

4.1.1 Proteomic approach

The proteomic approach involved the optimization of the one dimensional SDS-PAGE assay performed firstly on the standard cultivars reported in **Table 4**, in order to obtain a precise pattern of the migration of different HMW-GS to use as internal markers for the SSD genotypes analyses.

Seven different cultivars (Biensur, Bx7+By8; Saragolla, Bx6+By8; Abano, Bx7+By9; Ofanto, Bx20+By20; Colosseo, Bx14+By15; Isa, Bx13+By16; Francia, Bx17+By18) were considered to fine-tune the best analysis' conditions. The size of each subunits was deduced in 7.5% SDS-PAGE on the basis of the previously reported molecular weight (UPOV, 1994) (**Table 5**). The analyses of the subunit pattern (**Fig. 22**) enabled the distinction for the allelic combinations Bx7+By9, Bx7+By8 and Bx6+By8 (**Fig. 22**, lanes 7, 8 and 9 respectively), but provided a more difficult discrimination for Bx20+By20, Bx14+By15, Bx17+By18 and Bx13+By16 (**Fig. 22,** lanes 3, 4, 5 and 6 respectively).

Table 5: Determined molecular weights for proteins by means of SDS-PAGE analysis. Molecular weights of HMW-GS in the SDS-PAGE refer to UPOV (1994).

SUBUNIT	M.W. SDS-PAGE
Ax2*	108
Ax1	113
Bx6	100
Bx6.1	99
Bx7	98
Bx7^{OE}	-
Bx13	94
Bx14	94
Bx17	89.5
Bx20	94
By8	86
By9	83
By15	91
By16	90
By18	89.5
By19	-
By20	94

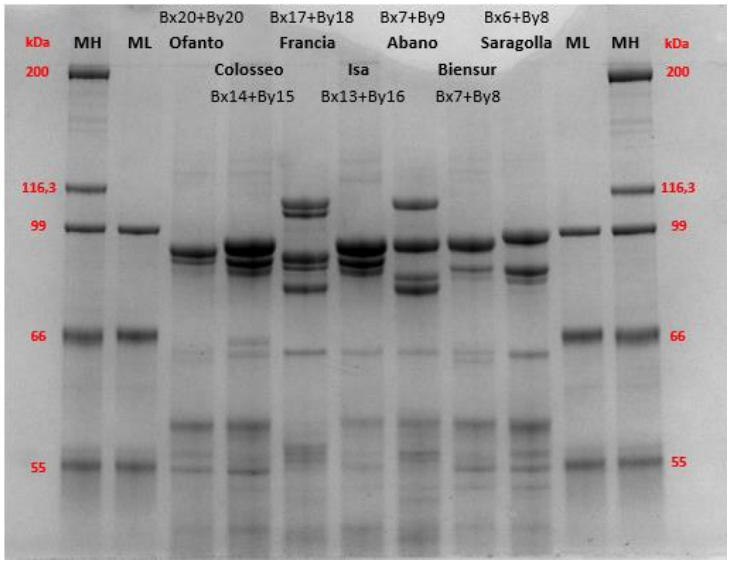

Figure 22: SDS-PAGE of a panel of standard cultivars containing different HMW-GS allelic combinations at *Glu-B1* locus. Lanes 1 and 11, Molecular Weight Standard High range; lanes 2 and 10, Molecular Weight Standard Low range; lane 3, cv. Ofanto (N, Bx20+By20); lane 4, cv. Colosseo (N, Bx14+By15); lane 5, cv. Francia (Ax2*, Bx17+By18, Dx2+Dy12); lane 6, cv. Isa (N, Bx13+By16); lane 7, cv. Abano (N, Bx7+By9, Dx2+Dy12); lane 8, cv. Biensur (N, Bx7+By8); lane 9, cv. Saragolla (N, Bx6+By8).

Notwithstanding the use of proteomic approaches as SDS-PAGE and two-dimensional polyacrylamide gel electrophoresis (2-DE) for the identification of subunit variation at the HMW-GS loci (Payne et al., 1981; Singh et al., 1991), a lack of a clear identification has been reported for those

subunits showing the same mobility on SDS-PAGE, thus leading to an incorrect assignation of different alleles (Gianibelli et al., 2002; Giraldo et al., 2010).

A second round of SDS-PAGE analyses was performed on 12% acrylamide gels, according to Ribeiro et al. (2011). To enlarge the panel of possible variations at the *Glu-B1* locus and to build a reference proteomic map at this locus, an increased number of cultivars was investigated. When available, the number of cultivars carrying rarer allelic combination was increased as internal control (Bx6+By8*, Bx7+By8*, Bx7*+By8*, Bx21, Bx14+By19, Bx23+By18, Bx7+By15 and Bx13+By19, **Table 4**). The composition of HMW-GS at the *Glu-A1* was also investigated.

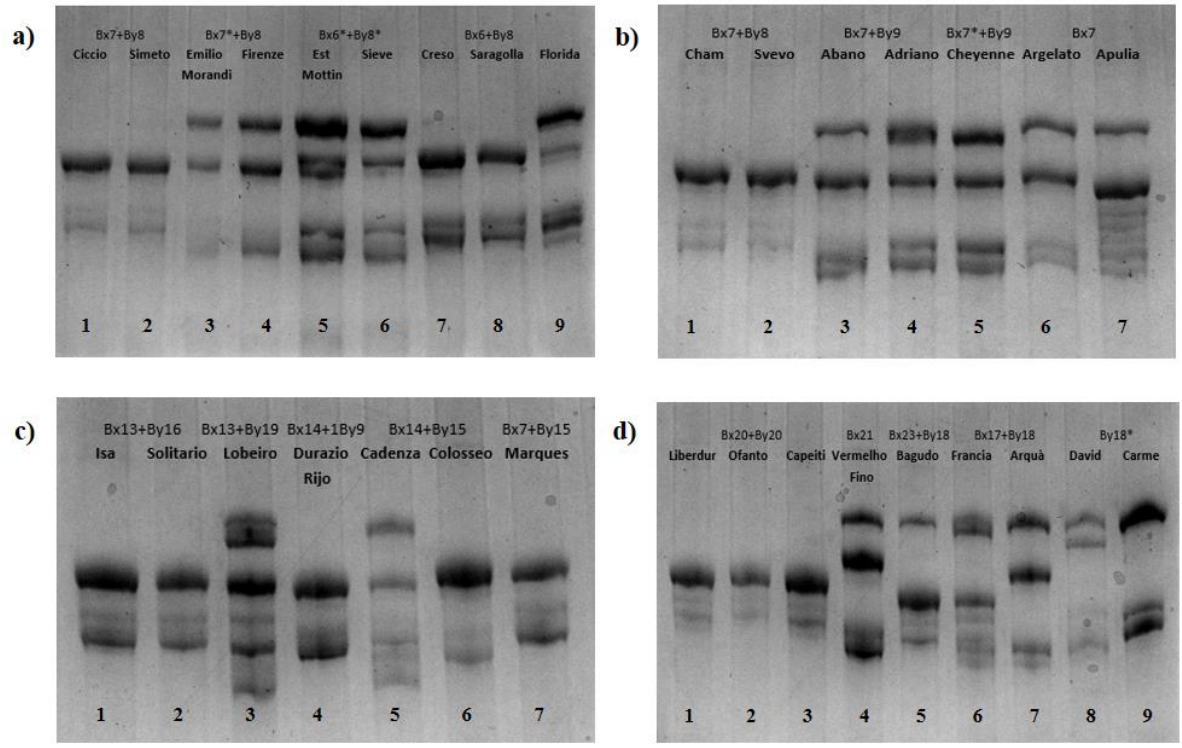

Figure 23: 12% SDS-PAGE of a panel of cultivars carrying different HMW-GS allelic combinations at *Glu-B1* locus. **a)** Ciccio (N, Bx7+By8), Simeto (N, Bx7+By8), Emilio Morandi (N, Bx7*+By8, Dx2+Dy12), Firenze (N, Bx7*+By8, Dx2+Dy12), Est Mottin (Ax1, Bx6*+By8*, Dx2+Dy12), Sieve (Ax1, Bx6*+By8*, Dx2+Dy12), Creso (N, Bx6+By8), Saragolla (N, Bx6+By8) and Florida (Ax1, Bx6+By8, Dx5+Dy10); **b)** Cham (N, Bx7+By8), Svevo (N, Bx7+By8), Abano (N, Bx7+By9, Dx2+Dy12), Adriano (Ax1, Bx7+By9, Dx5+Dy10), Cheyenne (Ax2*, Bx7*+By9, Dx5+Dy10), Argelato (N, Bx7, Dx2+Dy12) and Apulia (Ax1, Bx7, Dx2+Dy12); **c)** Isa (N, Bx13+By16), Solitario (N, Bx13+By16), Lobeiro (Ax1, Bx13+By19), Durazio Rijo (N, Bx14+By19), Cadenza (N, Bx14+By15, Dx5+Dy10), Colosseo (N, Bx14+By15/Bx13+By16) and Marques (N, Bx7+By15); **d)** Liberdur (N, Bx20+By20), Ofanto (N, Bx20+By20), Capeiti (N, Bx20+By20), Vermelho Fino (Ax1, Bx21), Bagudo (Ax1, Bx23+By18), Francia (Ax2*, Bx17+By18, Dx2+Dy12), Arquà (Ax2*, Bx17+By18, Dx2+Dy12), David (Ax1, By18*, Dx1+Dy12) and Carme (Ax1, By18*, Dx2+Dy12).

The new applied conditions led to a more precise discrimination especially of HMW subunits characterized by a similar molecular weight. A clear example is represented by cultivars Lobeiro (HMW-GS Bx13+By19) and Isa and Solitario (HMW-GS Bx13+By16) which share the same x-type subunit but not the y-type. In **Fig. 23** is shown that the difference of mobility between the cultivars

Lobeiro (**Fig. 23**: gel c, lane 3) and Isa and Solitario (**Fig. 23**: gel c, lane 1 and 2 respectively), was appreciable in 12%. Similarly, the mobility of cultivar Lobeiro (HMW-GS Bx13+By19, **Fig. 23**: gel c, lane 3) differed from that of cultivar Durazio Rijo (HMW-GS Bx14+By19, **Fig. 23**: gel c, lane 4). However, the increased SDS gel resolution do not help in distinguishing very similar subunits like Bx7/Bx7* (**Fig. 23**: gel a, lane 1, 2, 3, 4 and gel b) or By8/By8* (**Fig. 23**: gel a). As previously reported, in fact, a small difference in electrophoretic mobility exists between subunits Bx7 and Bx7*, the former one being slightly larger than the latter one (Butow et al., 2004), while subunits By8 and By8* show identical mobility on gel (Lei et al., 2006).

The analyses by SDS PAGE of the set of the so called "Reference cultivars" allow the identification of a proteomic map of HMW-GS subunits for both *Glu-A1* and *Glu-B1*. In order to obtain clear standards to use for the SDS genotypes screening, for those more complex cultivars the PCR marker-based approach has been employed (Lei et al., 2006; Ma et al., 2003; Xu et al., 2008).

4.1.2 PCR marker-based approach

A total of 33 cultivars, evidenced with an asterisk in **Table 4,** were considered in the molecular analysis by using six previously developed PCR-based molecular markers for *Glu-B1* alleles (Chapter 3, **Table 2** and **Table 6)**. The application of this set of molecular markers on a wider set of cultivars confirmed the previously determined specificity but, in some cases, revealed also new banding patterns, with the same molecular marker discriminating more than one specific *Glu-B1* gene (Janni et al., 2017). Also in this case to increase the validity of the results, more than one cultivar for each HMW-GS allelic composition was tested as control. **Fig. 24** showed the results obtained for 16 cultivars.

Table 6: Primers and PCR conditions used.

Primer Pair n°	Primers pair	HMW-GS genes discriminated	Tm	Ref.
PP1	ZSBy8_F5/R5*	*By8*	64°C	
PP2	ZSBy9_F2/R2*	*By16, By20*	62°C	Lei et al., 2006
PP3	ZSBy9_aF1/aR3*	*By9*	59°C	
PP4	Bx_F/ Bx_R*	*Bx17, Bx7**	58°C	Ma et al., 2003
PP5	Bx7_F/Bx7_R*	*Bx6/Bx7*	50°C	Schwarz et al., 2004
PP6	CauBx752*	*Bx14*	53.5°C	Xu et al., 2008

*See Table 2 for detailed primer sequences. The subunits discrimination is based on literature data.

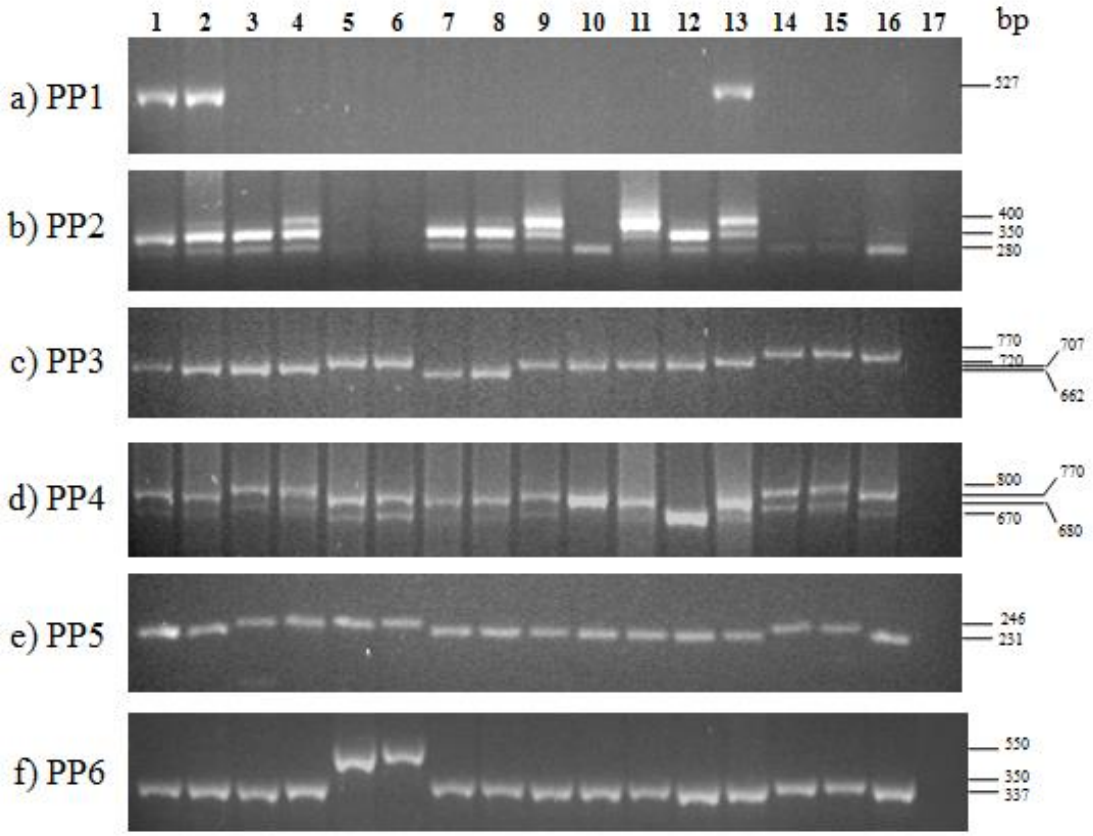

Figure 24: PCR products obtained by using the six selected molecular markers on the panel of 16 durum and bread wheat cultivars (Janni et al., 2017). 1a: PP1 (ZSBy8_F5/R5); 1b: PP2 (ZSBy9_F2/R2); 1c: PP3 (ZSBy9_aF1/aR3); 1d: PP4 (BxF/BxR); 1e: PP5 (Bx7_F/R); 1f: PP6 (CauBx752).
Lane 1, Ciccio (Bx7+By8); 2, Chinese spring (Bx7+By8); 3, Dylan (Bx6+By8); 4, Florida (Bx6+By8); 5, Ofanto (Bx20+By20); 6, Capeiti (Bx20+By20); 7, Adriano (Bx7+By9); 8, Cheyenne (Bx7*+By9); 9, Isa (Bx13+By16); 10, David (By18*); 11, Argelato (Bx7); 12, Francia (Bx17+By18); 13, Emilio Morandi (Bx7*+By8); 14, Est Mottin (Bx6*+By8*); 15, Sieve (Bx6*+By8*); 16, Cadenza (Bx14+By15); 17, negative control.

The sequences of new detected amplicons of primers pair PP2, PP3, PP4 and PP5 were analyzed. Each new band was isolated, purified and sequenced. The available annotated HMW sequences (**Table 7**) were exploited for the alignment, performed with DNAMAN software.

Table 7: Available annotated sequences of HMW genes and their corresponding accession, the NCBI ID is reported.

HMW gene	NCBI ID
Bx6.1	HQ731653
Bx7	X13927
Bx7OE	DQ119142
Bx13	EF540764
Bx14	AY367771
Bx14*	KJ579439
Bx17	AB263219
Bx20	AJ437000
Bx23	AY553933
Bx23*	KF995273
By8	AY245797
By9	X61026
By15	DQ086215
By15*	KJ579440
By16	EF540765
By18	KF430649
By20	LN828972

4.1.2.1 Analysis with PP1

Primer pair 1 (PP1, **Table 6**) was described by Lei et al. (2006) as specific for the *By8* gene: the authors observed a 527 bp DNA fragment in cultivars carrying the *By8* gene, while non-*By8* cultivars lacked the amplification signal. Moreover, the same allowed the discrimination between *By8* and *By8** genes, whose corresponding protein subunits are difficult to identify with SDS-PAGE due to their identical electrophoretic mobility (**Fig. 23**: gel a) (Lei et al., 2006). Hence, this primer pair has been considered as reference molecular marker for *By8* in many research papers (Motawei et al., 2008; Goutam et al., 2015; Ghazy et al., 2012; Salmanowicz and Dylewicz, 2007).

In the panel of cultivars tested, the expected amplicon was observed in those carrying the Bx7+By8 and Bx7*+By8 allelic combination (Ciccio, Chinese Spring and Emilio Morandi respectively; **Fig. 24a**; lane 1, 2 and 13) but not in any of those carrying the Bx6+By8 (**Fig. 24a**; lane 3 and 4). The discrimination between By8 and By8* allele was also verified at a molecular level (**Fig. 24a**; lane 1, 2 ,13, 14 and 15). A possible difference in the *By8* gene sequence in those genotypes carrying the Bx6+By8 or Bx7+By8 alleles could not be excluded as also previously reported by Patacchini and co-workers (2001) who supposed the presence of structural differences, at protein level, between subunits 1By8 belonging to the two allelic combinations aforesaid.

4.1.2.2 Analysis with PP2

Primer pair 2 (PP2, **Table 6**) was developed to selectively amplify the *By16* gene by Lei et al. (2006) on the basis of the 45 bp deletion in the *By16* gene and on the amplification of the repetitive

domain of the *By* gene. A complex pattern would be expected as indicated in **Table 2** (Chapter 3) (Lei et al., 2006; Motawei, 2008; Ghazy et al., 2012; Goutam et al., 2015).

The specificity of PP2 for the *By16* gene (3 bands) and *By20* (no bands) was confirmed as shown in **Fig. 24b** (lane 9 and lanes 5 and 6 respectively). In addition, new banding patterns were observed: a single 280 bp fragment was detected in the *By18** cultivar David (**Fig. 24b**: lane 10) (and confirmed in Carme) as previously detected also in *Triticale* (Salmanowicz and Dylewicz, 2007). The same profile was observed also in Cadenza (Bx14+By15) (**Fig. 24b**: lane 16). Finally, the faint band of 280 bp retrieved in cultivars carrying the Bx6*+By8* allelic combination (**Fig. 24b**: lane 14 and 15) was interpreted as a non-specific signal and the unexpected pattern observed for cultivars Florida (Bx6+By8), Argelato (Bx7) and Emilio Morandi (Bx7*+By8) (**Fig. 24b**: lanes 4, 11 and 13 respectively), which showed the three bands for *By16*, might be the result of a non-specific primers annealing.

Sequence analysis of the 280 bp amplicon retrieved in cvs. David and Cadenza (**Fig. 25a**) did not allow the assignment of the band to a specific subunit. The sequence obtained for David (By18*) showed four SNPs (SC30, **Table 8**) which, in the end, revealed a 100% homology with the *By15** gene (KY579440), while the specific band in Cadenza (Bx14+By15) showed three SNPs in common with other *By* sequences and one unique SNP (A1212G) (**Fig. 25b** and SC31, **Table 8**).

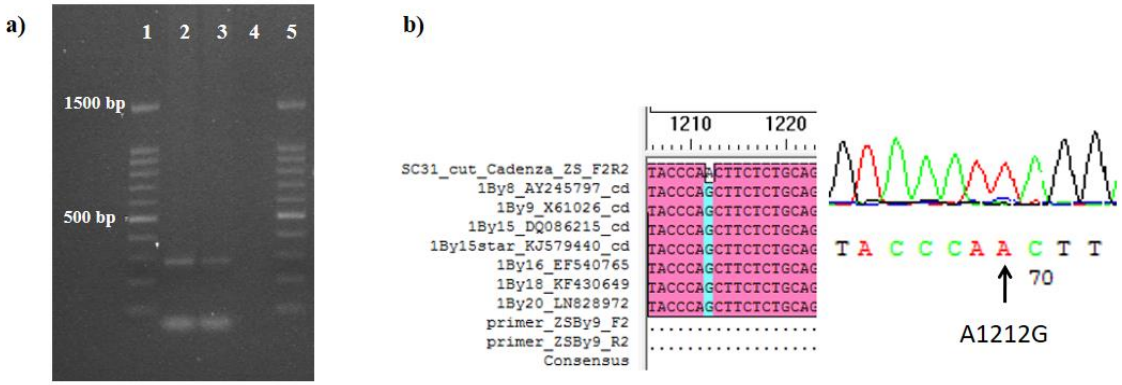

Figure 25: **a**) Purified amplicons of cvs. David and Cadenza obtained by using PP2 (ZSBy9_F2/R2). Lanes 1 and 5, AccuRuler 100 bp Plus DNA Ladder (Maestrogen); lane 2, David (By18*); lane 3, Cadenza (Bx14+By15); lane 4, negative control. **b**) Portion of the alignment and of the corresponding electropherogram showing the unique SNP retrieved the amplicon of cv. Cadenza by using PP2.

Table 8: Summary of the SNPs retrieved in the amplicons of David (SC30) and Cadenza (SC31) amplified with primer pair 2.

Sequence	Genotype	SNP	Feature
SC30	David	C1165T	shared by seq. *By15*, *By15** and *By20*
		T1205C	shared by seq. *By15**
		A1313G	
		C134T	shared by seq. *By15*, *By15** and *By20*
SC31	Cadenza	C1165T	shared by seq. *By15*, By15* and By20
		G1173C	shared by seq. By15 and By20
		A1212G	unique of sequence SC31
		C1345T	shared by seq. By15, By15* and By20

These results, together with the lack of the entire *By18** sequence in the database did not allow any specific assignment for the molecular marker, indicating that, when a high variability in the composition of alleles at the *Glu-B1* locus is present in the tested genotypes, the use of PP2 for the screening may result in the misidentification of genotypes.

4.1.2.3 Analysis with PP3

Primer pair 3 (PP3, **Table 6)** was developed to discriminate the *By9* gene from other *By* genes, on the basis of the 45 bp size difference in the amplicons produced (Lei et al., 2006). A 662 bp amplicon was produced specifically for the *By9* gene, while the other *By* cultivars produced a 707 bp fragment (Lei et al., 2006; Motawei et al., 2008).

By testing the panel of standard cultivars selected, the expected fragment of 662 bp was observed in *By9* cultivars Adriano and Cheyenne (Bx7+By9 and Bx7*+By9 respectively) (**Fig. 24c**: lanes 7 and 8). A new amplicon of 720 bp was observed in Ofanto and Capeiti, carrying the Bx20+By20 allelic combinations. Similar results were found in *Triticale*, where a 750 bp band was observed in cultivars carrying the *By20** gene (Salmanowicz and Dylewicz, 2007). Moreover, cultivars Est Mottin and Sieve (Bx6*+By8*) gave a unique fragment of approximately 770 bp (**Fig. 24c**: lanes 14 and 15), which was then confirmed also in the cultivar Virest (**Table 4**).

Sequence analysis of the 720 bp band retrieved in Ofanto and Capeiti (Bx20+By20) showed a total of 10 SNPs and a 45 bp insertion (SC22/SC23, **Table 9**) which revealed a 100% identity with the *By20* sequence (LN828972), thus confirming that the primer pair 3 can assign the By20 subunit. The corresponding Bx6*+By8* sequence, instead, showed a more complex picture, with the presence of a total of 13 SNPs. Out of these, ten are shared by other *By* sequences, while 3 are unique in the

Bx6*+By8* cultivar Est Mottin (**Fig. 26** and SC24/SC25, **Table 10**). However, despite the presence of these 3 unique SNPs the, the lack of the *By8** genomic sequence in the database did not allow any specific assignment to be made for this allele. Two additional cultivars (Tudest and Virest) carrying the Bx6*+By8* allelic combination were tested with PP3 molecular marker in order to provide more reliable data. Since the results obtained were the same (**Fig. 27**), it is possible to hypothesize that PP3 is also suitable for identifying the By8* allele. This last result, however, needs further analyses.

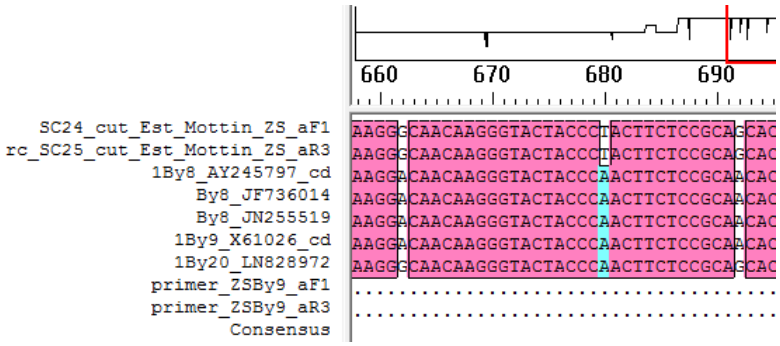

Figure 26: Portion of the alignment of the Est Mottin sequence (Bx6*+By8*, SC24/25) with all the other By sequences available. The alignment evidenced the unique SNP T680A out of the three retrieved.

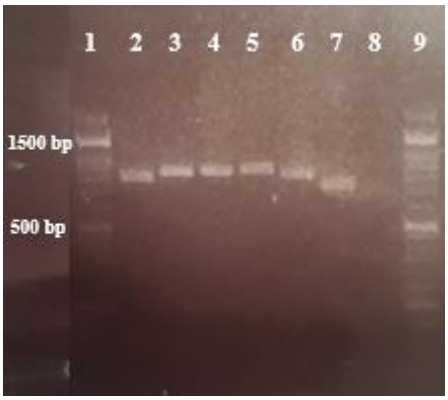

Figure 27: Amplifications products obtained by testing PP3 (ZsBy9_aF1/aR3) on four different cultivars carrying the Bx6*+By8* allelic combination. Lanes 1 and 9, AccuRuler 100 bp Plus DNA Ladder (Maestrogen); lane 2, Tudest; lane 3, Virest; lane 4, Est Mottin; lane 5, Sieve; lane 6, Capeiti (Bx20+By20); lane 7, Adriano (Bx7+By9); lane 8, negative control.

Table 9: Summary of the SNPs retrieved in the amplicons of Ofanto (SC22/23) amplified with primer pair 3.

Sequences	Genotype	SNP	Features
SC22/23	Ofanto	C590T	shared by seq. By15, By15*, By16 and By20
		G693A	
		G663A	shared by seq. By15, By15* and By20
		C752T	
		G1134A	
		A797G	shared by seq. By15 and By20
		G814C	
		A886G	
		C1007G	
		G1174C	
		/	45bp insertion shared by seq. By15 and By20

Table 10: Summary of the SNPs retrieved in the amplicons of Est Mottin (SC24/25) amplified with primer pair 3.

Sequence	Genotype	SNP	Feature
SC24/SC25	Est Mottin	C589T	shared by seq. By15, By15*, By16 and By20
		G692A	
		G662A	shared by seq. By15, By15* and By20
		C751T	
		G1133A	
		A885G	shared by seq. By15 and By20
		C1006G	
		G1173C	
		T680A	unique of sequence SC24/SC25
		G727A	
		A1178G	
		G796A	shared by all the sequences considered with the exception of By20
		C813G	
		/	45bp insertion shared by seq. By15 and By20

4.1.2.4 Analysis with PP4

Primer pair 4 (PP4, **Table 6**) was developed by Ma et al. (2003) to specifically distinguish between the *Bx17* gene and the non-*Bx17*, giving one or two bands respectively (Chapter 3, **Table 2**). Different fragment sizes for cultivars carrying the non-*Bx17* genes were reported by other authors; in particular, two fragments of 650 and 750 bp for the *Bx7** gene and two of 670 and 770 bp for any other *Bx* gene (Butow et al., 2003; Ghazy et al., 2012).

The present research confirmed the previously-reported specificities, as shown for cv. Francia (Bx17+By18) which possesses the expected 670 bp fragment (**Fig. 24d**: lane 12) and for Cheyenne (Bx7*+By9) with a 650 and 750 bp fragments (**Fig. 24d**: lane 8). In addition to the conventional banding patterns, a single fragment of 770 bp was observed for the cv. David (**Fig. 24d**: lane 10)**,** then confirmed in cv. Carme, carrying the By18* allele.

To confirm the hypothesis of a *By18** specificity for the PP4 primers set, the 770 bp amplicon of David and Carme was purified (**Fig. 28a**), sequenced and aligned with a selection of *Bx* and *By* annotated alleles. The analysis showed a 100% sequence identity with the *Bx23** and *Bx14** genes (ID KF995273 and KJ579439) due to the presence in the sequence of two SNPs (A1731G and G1736A) that are specific for the aforesaid genes (**Fig. 28b**). However, the lack of an annotated genomic *By18** sequence did not allow any hypothesis of specificity for the By18*, Bx14* and Bx23 alleles to be formulated using this primer pair. Nevertheless, the same 770 bp band was confirmed also in cultivar Garibaldino which carries the By18*, suggesting a possible specificity for this allele.

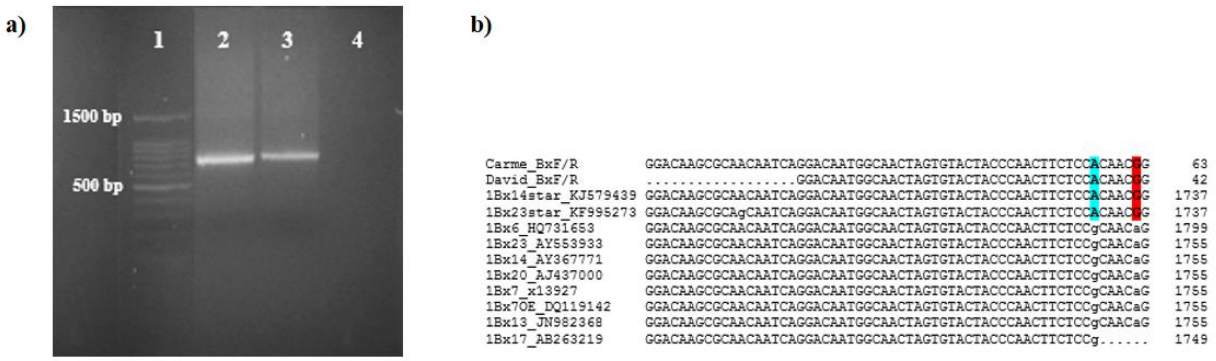

Figure 28: **a**) Purified amplicons of cvs. Carme and David obtained by using PP4 (Bx_F/Bx_R). Lanes 1, AccuRuler 100 bp Plus DNA Ladder (Maestrogen); lane 2, Carme (By18*); lane 3, David (By18*); lane 4, negative control. **b**) Portion of the alignment of Carme and David sequences with all the Bx sequences annotated. The two SNPs specific of the amplicons analyzed, and which evidenced a 100% homology with the Bx14* and Bx23* genes, are indicated (A1731G is highlighted in blue; G1736A is highlighted in red).

4.1.2.5 Analysis with PP5

The co-dominant primer pair 5 (PP5, **Table 6**) was developed by Schwarz et al. (2004) on the basis of a 15 bp insertion observed for the Bx6 allele and absent in all other Bx alleles, thus allowing the specific discrimination of the *Bx6* gene. The authors reported that the PP4 profile consists of two bands of 246 bp for the *Bx6* cultivars and 231 bp for the non-*Bx6* cultivars (Chapter 3, **Table 2**).
The tested cultivars confirmed the high specificity of this primer pair for discriminating the *Bx6* gene, since the expected amplicon of 246 bp was retrieved in cvs. Dylan and Florida (Bx6+By8, **Fig. 24e**: lanes 3 and 4). Moreover, the same pattern was also observed for the *Bx20* genotypes (Ofanto and Capeiti, Fig. 26e: lanes 5 and 6) and for the *Bx6** Est Mottin and Sieve (**Fig 24e**: lanes 14 and 15), then confirmed also in cultivar Virest (**Table 4**). To our knowledge, these features of PP4 were never previously reported.

Amplicons of Ofanto and Est Mottin were purified (**Fig. 29a**) and underwent to sequence analysis, which revealed for both few SNPs (three for *Bx20* and four for *Bx6**) shared with many

other *Bx* sequences (SC28 and SC29 respectively, **Table 11**). Moreover, the presence of the 15 bp insertion also in the mRNA *Bx20* sequence (AJ437000) (**Fig. 29b** and **c**) make consistent the capability of PP5 to distinguish also genotypes carrying the Bx20+By20 and Bx6*+By8* allelic combinations; however, as the Bx6* sequence is not available in the database so far, the hypothesis of a specific additional assignment for the PP5 cannot be formulated.

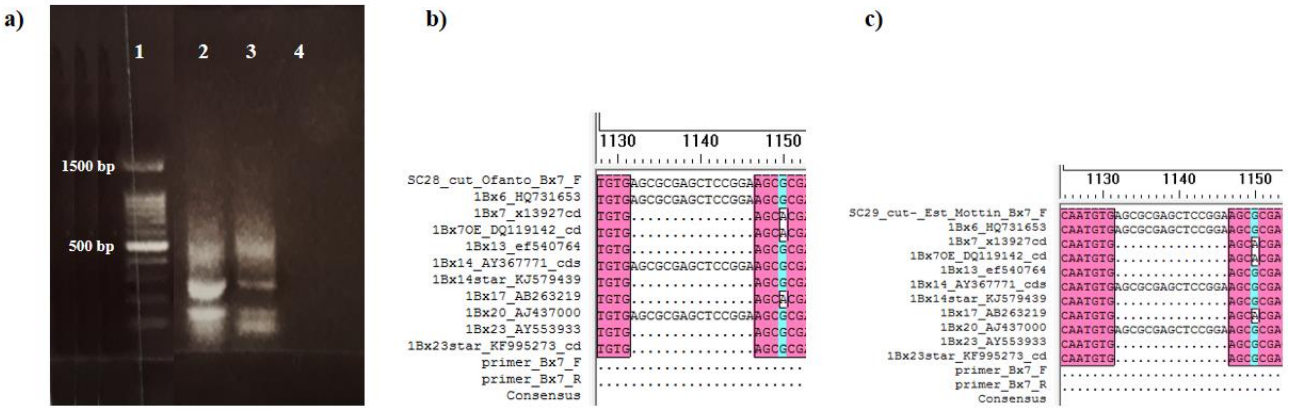

Figure 29: **a**) Purified amplicons of cvs. Ofanto and Est Mottin obtained by using PP5 (Bx7_F/Bx7_R). Lanes 1, AccuRuler 100 bp Plus DNA Ladder (Maestrogen); lane 2, Ofanto (Bx20+By20); lane 3, Est Mottin (Bx6*+By8*); lane 4, negative control. **b**) and **c**) Portion of the alignment with all the Bx sequences annotated which highlights the 15 bp insertion shared by both Ofanto (B) and Est Mottin (C) sequences.

Table 11: Summary of the SNPs retrieved in the amplicons Ofanto (SC28) and Est Mottin (SC29) amplified with primer pair 5.

Sequences	Genotype	SNP	Features
SC28	Ofanto	A1083G	shared by seq. Bx14 and Bx20
		A1165G	
		A1210G	
		/	15 bp insertion shared by seq. Bx14 and Bx20
SC29	Est Mottin	G1083A	shared by all the sequences considered with the exception of Bx14 and Bx20
		G1165A	
		G1210A	
		G1241A	shared by seq. Bx6 and Bx13
		/	15 bp insertion shared by seq. Bx14 and Bx20

4.1.2.6 Analysis with PP6

Finally, primer pair 6 (PP6, **Table 6**) was developed by Xu et al. (2008) to distinguish between *Bx14* and *Bx17* genes at the *Glu-B1* locus. A fragment of 752 bp for *Bx14* and a fragment of 337 bp for *Bx17* were expected from the analyses of the standard cultivars. Using the indicated conditions in all cultivars only the *Bx17* band of 337 bp was detected, including in the cv. Cadenza (Bx14+By15, **Fig. 24f**: lane 16). The result was confirmed by testing three additional cultivars carrying the Bx14+By15 allelic combination (**Fig. 30a**). Moreover, from the analysis, PP6 showed a single 550 bp band when applied to *Bx20* cultivars (**Fig. 24f**: lanes 5 and 6) never previously reported. This result was verified on a total of six *Bx20* cultivars (**Fig. 30b**).

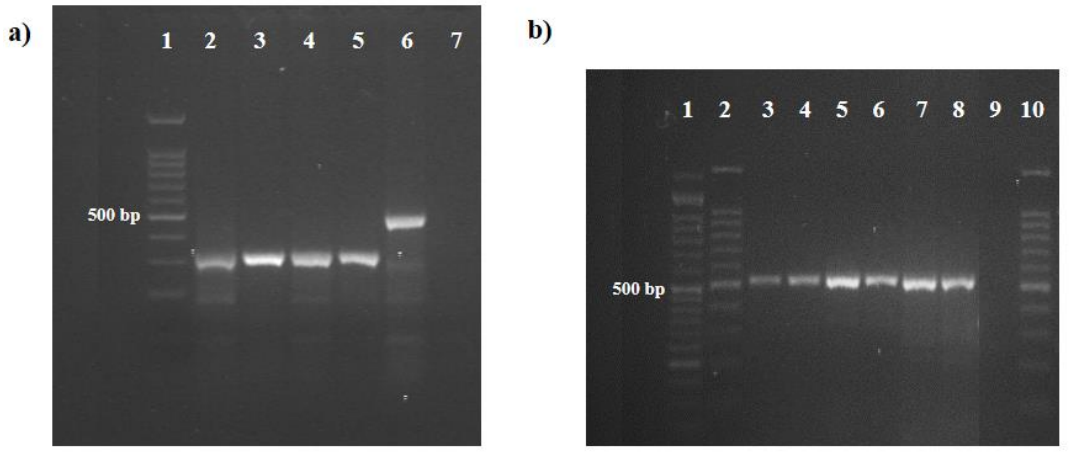

Figure 30: **a**) Amplification products obtained by testing PP6 (CauBx752) and four different cultivars carrying the Bx14+By15 allelic combination. Lane 1, AccuRuler 100 bp Plus DNA Ladder (Maestrogen); lane 2, Colosseo'; lane 3; Colosseo''; lane 4, Cadenza; lane 5, Durazio Rijo; lane 6; Ofanto (Bx20+By20); lane 7; negative control.
b) PCR products of six different cultivars Bx20+By20 amplified with PP6 (CauBx752). Lane 1, NZYDNA Ladder VI (NzyTech); lanes 2 and 10, AccuRuler 100 bp Plus DNA Ladder (Maestrogen); lane 3, Appulo; lane 4, Liberdur; lane 5, Cappelli; lane 6, Trinacria; lane 7, Ofanto; lane 8, Capeiti; lane 9, negative control.

The analysis of the 550 bp band observed in cv. Ofanto, showed a 100% identity with a portion of the promoter sequence of the *Bx14* allele (AY367771) that might be shared also by the *Bx20* gene for which the promoter sequence has not been deposited. However, the same specificity was not observed in the *Bx14* genotypes, and did not lead to any hypothesis of further specificity for the PP6 marker, though results highlighted that a wide number of cultivars should be taken in consideration when using this molecular marker to avoid allele misidentification.

4.1.3 Discussions

The results here presented, showed that the combination of two approaches, the proteomic and the genomic one, have been successfully applied in combination to determine the composition at the *Glu-B1* locus of wheat cultivars. In particular, the usage of both SDS-PAGE analyses and PCR marker-based assays led to a more precise discrimination of the allelic banding pattern, allowing to overcome allele misidentification. PCR-based molecular markers, indeed, represent an efficient and fast tool to overcome some of the drawbacks in protein-based methods for HMW-GS allele resolution (Rasheed et al., 2014), like the frequently incorrect interpretation of alleles sharing similar mobility on SDS gels (Gianibelli et al., 2001; Giraldo et al., 2010). On the basis of the available HMW-GS sequences, a large number of PCR-based DNA molecular markers have been developed and applied (D'Ovidio et al., 1995; Lafiandra et al., 1997; Ma et al., 2003; Lei et al., 2006); a survey of the molecular markers available for the *Glu-B1* locus is reported in **ANNEX A**, **Table 3**. Notwithstanding the large number of markers developed, the allele specificity has been verified on specific and limited numbers of test cultivars selected on a case by case basis and representative of the expected HMW-GS subunits at the *Glu-B1* locus (Ma et al., 2003; Lei et al., 2006).

In the first part of the present project, a large number of reference cultivars, carrying thirteen different HMW-GS allelic combinations, have been analyzed through the aforesaid "Omic" approaches, leading to the development of a wider reference proteomic map potentially useful for researchers working on wheat quality and in Marker Assisted Selection (MAS) programs to readily detect desired traits in wheat populations or elite cultivars. In addition to reviewing the molecular markers developed for HMW-GS allelic composition at the *Glu-B1* locus, PCR analysis, together with the consequent sequencing analysis suggest additional putative specificity for *Glu-B1* alleles for the molecular markers tested which, to our knowledge, were never reported before. In recent years, however, several HMW genes at the *Glu-B1* locus have been deposited and annotated and analysis of the alignment within the x and y type reveals a high level of DNA or mRNA sequence similarity between them, thus explaining the complex pattern obtained in the application of the *Glu-B1* molecular markers. By sequencing the obtained amplicons, indeed, it was possible to add just a new assignment for the PP3 marker (ZSBy9_aF1/aR3) for the *By20* allele, which is of particular importance in analysis of the glutenin composition in durum wheat landrace collections, where the Bx20+By20 alleles are often very frequent (Bellil et al., 2014; Moragues et al., 2006). Several of the markers tested produced a specific band for Bx20+By20, suggesting that the coding sequence of this gene may share several haplotypes. The identification of haplotypes is useful for the identification of

regions of the genome associated with traits of interest or candidate genes but also for targeting the development of specific molecular markers for MAS (Varshney et al., 2006).

For all the other markers, since the observed polymorphisms were in common with many other HMW-GS genes, it was not possible to hypothesize any further specificity. An example of the complexity of the link between molecular marker banding pattern and the allele assignment is reported for the PP4 marker (Bx_F/Bx_R) for which the 770 bp band obtained in two *By18** cultivars (Carme and David, **Fig. 24d**) has been sequenced and analyzed (**Fig. 28a**). The sequence analyses revealed a 100% sequence similarity with *Bx14** and *Bx23** alleles which share with By18* two specific substitutions (A1731G and G1736A, **Fig. 28b**). The absence of any *By18** DNA sequence deposited, however, did not allow any specificity to be assigned and only suggests that *By18** also shares the same SNPs. Nevertheless, clear indications of reproducible banding patterns obtained by using the selected molecular markers PP3 (ZsBy9_aF1/aR3) and PP6 (CauBx752) have been shown (**Fig. 27** and **Fig. 30** respectively) which, to our knowledge, were never previously reported.

Hence, these results reported for the set of reference cultivars provided a selected guidance for the interpretation of the results when a similar approach is used on durum wheat populations and cultivars panels with high variability at the *Glu-B1* locus.

4.2 HMW-GS characterization of SSD genotypes

The seed bank of IBBR-CNR held in Bari holds a collection of about 27000 samples of the genus *Triticum* and more of 5600 samples are identified as *Triticum durum*. Out of these, 500 accessions have been randomly selected to create a handy subset (SSD core collection). The Single Seed descent approach have been applied for the material multiplication and selection in order to obtain a set of material with little heterozygosity, potentially representative of the overall variation present in the germplasm and considering geographical origin of each genotype, trying to retain a high level of representativeness of all agro-climate regions (Pignone et al., 2015).

The germplasm panel (so on called SSD collection) considered in the project for the HMW-GS characterization comprised a subset of 152 genotypes of *Triticum durum* genotypes representatives of 31 countries (**Table 12**).

Table 12: Countries of origin of the 152 durum wheat landraces used to develop the germplasm set.

Country of origin	Number of entries
Afghanistan	1
Algeria	6
Saudi Arabia	1
Azerbaijan	1
Bosnia Herzegovina	1
Bulgaria	1
Cyprus	1
Crete	7
Egypt	4
Ethiopia	10
France	2
Japan	1
Jordan	1
Greece	16
India	3
Iran	9
Iraq	13
Italy	13
Yugoslavia	1
Libya	2
Morocco	9
Perù	1
Portugal	2
Romania	1
Russia	1
Syria	3
Spain	5
Tunisia	17
Turkey	6
Ukraine	2
USA-ND/MN/WA	11

Based on the results obtained for the reference cultivars, an identical approach was performed on the SSD collection. The approach relies on the application first of the proteomic screening, and only on those unclear or undetermined genotypes, the molecular approach was applied.

This approach led to the identification of three different alleles at *Glu-A1* locus, while a greater variability, with a total of fifteen alleles, was observed at *Glu-B1* locus (**Table 15**). At the *Glu-A1* locus the allelic composition was only contributed by x-type subunits such as Ax1, Ax2* and *null*, which are controlled by *Glu-A1a, -A1b* and *-A1c* alleles respectively, whereas at *Glu-B1,* eleven known alleles (*a*, *b*, *an*, *d*, *e*, *f*, *1g*, *h*, *z*, *al* and *ak*) were represented, along with four which have not yet been allocated an allelic designation (GS combinations Bx14+By19, Bx14+By20, Bx7+By19 and

Bx6+By8*). There were six x-type (7, 6, 20, 13, 14 and 7*) and six y-type (8, 20, 16, 15, 19 and 8*) GSs.

4.2.1 Proteomic approach

A first round of SDS-PAGE (7.5% precast polyacrylamide gels) analyses was performed on the 152 entries of the SSD to identify the SSD carrying the most common subunits (in particular at the *Glu-B1* locus) and, then, the analyses were sharpened by increasing the SDS-PAGE gel concentration to distinguish the most putative rare subunits. A total of twelve gels were run (Criterion™ Dodeca™ Cell, Biorad, Hercules, CA). Examples of the protein profiles are reported in **Fig. 31.**

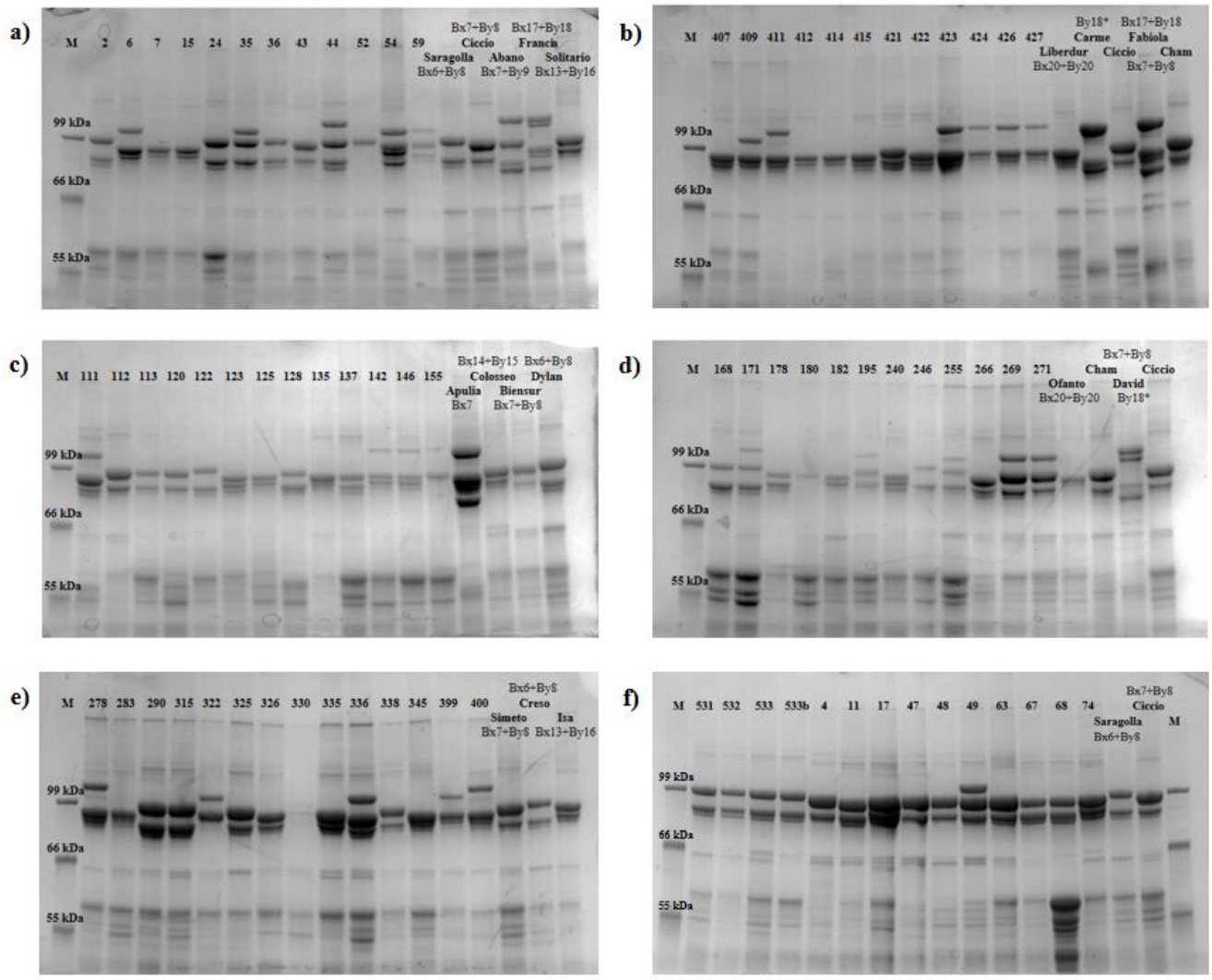

Figure 31: SDS-PAGE run on the SSD genotypes on 7.5% precast polyacrylamide gels. A Molecular Weight Standards Low Range (BioRad) and two to five standard cultivars were loaded onto each gel. The expected range for the *Glu-A1* subunits was between 100 and 113 kDa, while the one for the *Glu-B1* subunits was between 89 and 100 kDa.

Glu-A1 locus encoded one of two x-type subunits (1, 2*) or for a null allele, reflecting the presence of, respectively, *Glu-A1a, -A1b* and *-A1c*. Based on the first analysis with SDS-PAGE, 107 entries (70.4%) were typed as carriers of *Glu-A1c*, while the other 45 carried either *Glu-A1a* or *-A1b*. In particular, 27 genotypes (17.8%) were assigned as *Glu-A1a* and 18 (11.8%) as *Glu-A1b* (**Table 13**). The *Glu-B1* locus, instead, showed a greater variability in the SSD collection. The most common glutenin subunit (GS) combinations, such as Bx7+By8, Bx6+By8 and Bx20+By20, were readily identifiable and clearly assigned by SDS-PAGE. More precisely, out of the 152 genotypes considered, 28 (18.4%) were assigned Bx6+By8, 43 (28.3%) as Bx7+By8and 42 (27.6%) as Bx20+By20. Examples of SSD genotypes carrying the Bx6+By8 GS are shown in **Fig. 31a** (entries 2, 24, 35 and 44), in **Fig. 31c** (entry 122) and in **Fig. 31d** (entries 168 and 171); examples of typical Bx7+By8 GS mobility were retrieved in SSD 43 (**Fig. 34a**), in SSDs 112, 113 and 120 (**Fig. 31c**), in SSDs 269 and 271 (**Fig. 31d**) and in many others as shown in **Fig. 31f** (entries 4, 11, 17, 47, 48, 49, 63, 67 and 68). Examples of genotypes carrying the Bx20+By20 GS are retrievable in **Fig. 31b** (entries 407, 409, 411, 412, 414, 415, 422, 423, 424, 426 and 427). All the aforesaid allelic combinations displayed an electrophoretic mobility perfectly overlapping with the reference cultivars.

The remaining 39 entries (25.7%) showed undeterminable patterns (for example SSDs 123, 125, 137, 142 and 146, **Fig. 31c**). Moreover, a number of SSD genotypes showed complex pattern characterized by four to five bands, thus being more similar to those of common wheat even if they were verified being durum wheat (i.e. SSD 54; **Fig. 31a**).

Table 13: Allele frequency at the two *Glu-1* loci identify through SDS-PAGE analyses.

Locus	Allele	Subunit	Number of genotypes	Frequency (%)
Glu-A1	*a*	Ax1	27	17.8
	b	Ax2*	18	11.8
	c	Null	107	70.4
GluB1	*b*	Bx7+By8	43	28.3
	d	Bx6+By8	28	18.4
	e	Bx20+By20	42	27.6
	unknown		39	25.7

Thus, further analyses were performed with SDS-PAGE (12% polyacrylamide gels; Ribeiro et al., 2011) including appropriate reference cultivars. Nevertheless, this approach was not conclusive allowing to group SSD genotypes on the basis of similar migration patterns but do not sort out the precise HMW-GS composition. **Fig. 32** and **33**).

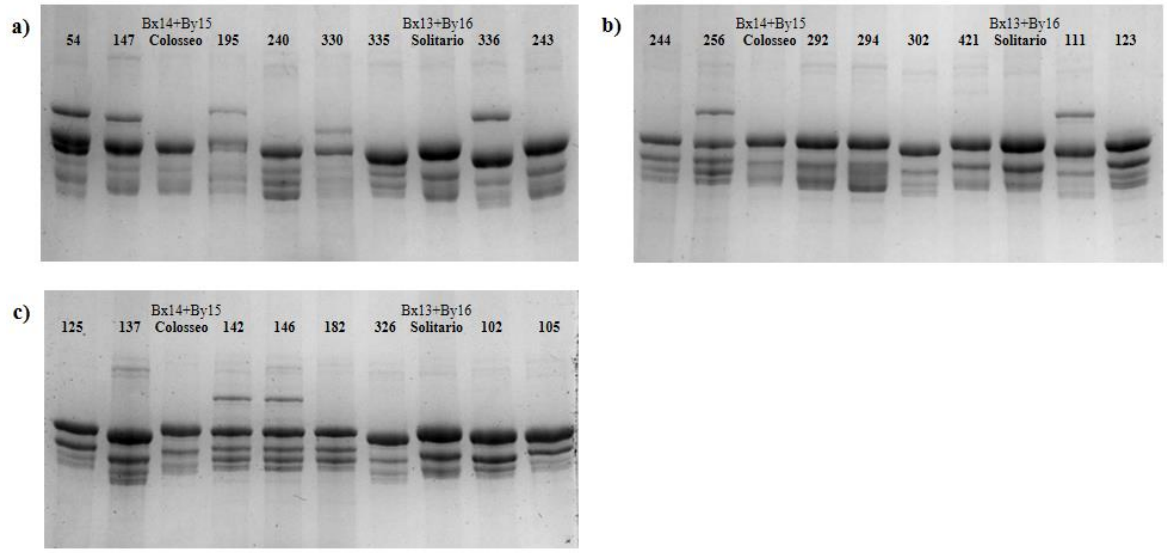

Figure 32: SDS-PAGE of SSDs with uncertain HMW allelic combination on 12% polyacrylamide gels. Cultivars Colosseo (Bx14+By15) and Solitario (Bx13+By16) were used for the comparison.

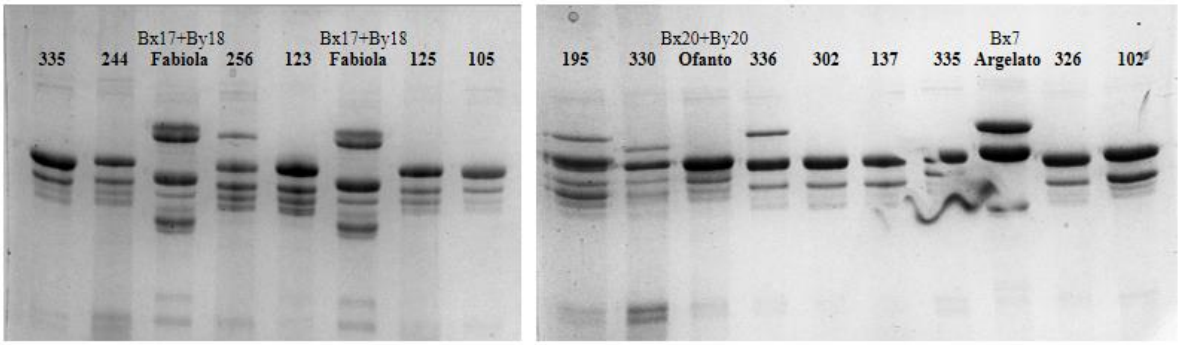

Figure 33: SDS-PAGE of SSDs with uncertain HMW allelic combination on 12% polyacrylamide gels. Cultivars Fabiola (Bx17+By18), Ofanto (Bx20+By20) and Argelato (Bx7) were used for the comparison.

A set of cultivars previously characterized for particular HMW-GS subunits such as Marques (Bx7+By15) or Durazio Rijo (Bx14+By19) (**Table 4**) were used for the following analyses. A perfect identity between SSDs 173 and 343 with cultivar Marques or between SSD 348 and cultivar Durazio Rijo was detected (**Fig. 34a**), thus assigning those entries as GSs Bx7+By15 and Bx14+By19 respectively. Similarly, SSDs 102, 178 (**Fig. 34b**) and SSDs 266, 302 ,326, 336 (**Fig. 34c**) were tagged as Bx14+By19. Moreover, in some cases, a perfect match between the SSD genotypes and more than one cultivars pattern was observed assigning for example SSDs 451 and 453 (**Fig. 34b**) as Bx7 and By19 subunit.

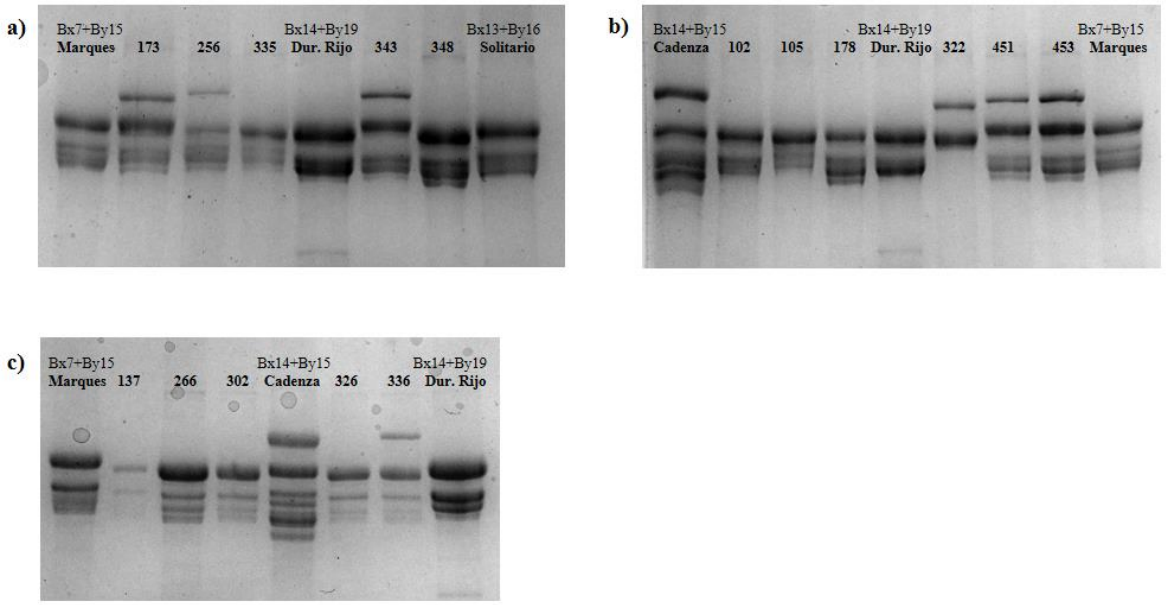

Figure 34: SDS-PAGE of SSDs with uncertain HMW allelic combination on 12% polyacrylamide gels. Cultivars Marques (Bx7+By15), Durazio Rijo (Bx14+By19), Solitario (Bx13+By16) and Cadenza (Bx14+By15) were used for the comparison.

As control the SSDs previously assigned as Bx6+By8, Bx7+By8 and Bx20+By20 were also separated on 12% SDS-PAGE gel (**Fig. 35**). Allelic combination Bx6+By8 and Bx20+By20 were confirmed (examples are shown in **Fig. 35a** and **35c**) in all entries with the exception for three genotypes (i.e. SSD 96, **Fig. 35c**). When the Bx7+By8 SSD genotypes were analyzed a mismatch with previously assignations occurred. In particular, a total of thirteen genotypes showed a pattern not overlapping with those of the corresponding cultivar Ciccio or Simeto, either for one or for both subunits (i.e. SSDs 173, 298, 343 and 17, **Fig. 35c** red box).

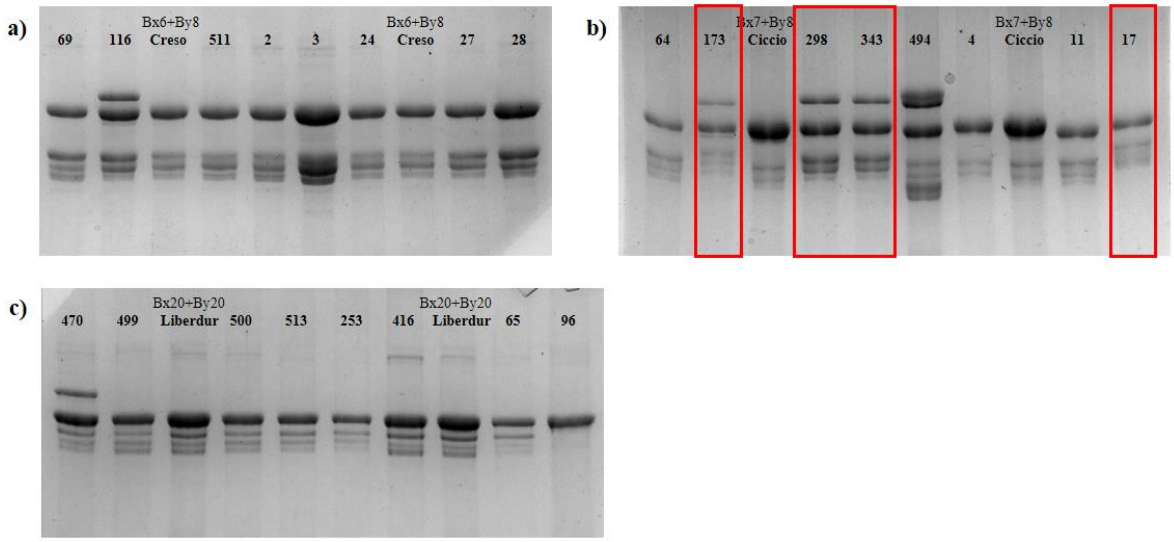

Figure 35: SDS-PAGE on 12% polyacrylamide gels of SSDs previously assigned as Bx6+By8, Bx7+By8 or Bx20+By20. **a**) Examples of SSDs confirmed as carriers of the GS allelic combination Bx6+By8. Cultivar Creso was used for the comparison; **b**) Examples of the validation of SSDs carrying the Bx7+By8 GS allelic combination. Cultivars Ciccio and Simeto (data non shown) were used for the comparison. Red Box indicates genotypes showing patterns not consistent with those of the corresponding reference cultivars, such as SSDs 173, 298, 343 and 17; **c**) Examples of SSDs confirmed as carriers of the Bx20+By20 GS allelic combination. Cultivar Liberdur and Ofanto (data not shown) were used for the comparison. A total of three SSDs exhibited just one subunit.

Thanks to the collaboration with the IBBA-CNR (Milan), the set of germplasm was analyzed through the Lab-on-a-chip technology (Marchetti et al., 2011; Balázs et al., 2012; Živančev et al., 2013), obtaining the same results above described for each of the SSD assigned as Bx7+By8, Bx6+By8 and Bx20+By20 (Janni et al., 2018 submitted). Discrepancy were, instead, observed for the remaining genotypes.

Therefore, the unassigned SSDs underwent the molecular analyses.

4.2.2 PCR marker-based approach

Entries characterized by *Glu-B1* allelic combination not clearly detectable by proteomic were further analyzed at a molecular level. In addition, the Bx7+By8 genotypes were included to verify a possible mismatch between subunit By8 and By8* at a protein level. The set of six PCR-based molecular markers mentioned in **Table 6** and, four primer pairs targeting the *Glu-A1* genes were used. The complete set of markers used is listed in **Table 14**.

Table 14: Primers and PCR conditions used.

Primer Pair n°	Primers pair	HMW-GS genes discriminated	Tm	Ref.
PP1	ZSBy8_F5/R5*	*By8*	64°C	
PP2	ZSBy9_F2/R2*	*By16, By20*	62°C	Lei et al., 2006
PP3	ZSBy9_aF1/aR3*	*By9*	59°C	
PP4	Bx_F/ Bx_R*	*Bx17, Bx7**	58°C	Ma et al., 2003
PP5	Bx7_F/Bx7_R*	*Bx6/Bx7*	50°C	Schwarz et al., 2004
PP6	CauBx752*	*Bx14*	53.5°C	Xu et al., 2008
PP7	Ax_F/Ax2*_R	*Ax1*	58°C	Ma et al., 2003
PP8	Ax1/Ax2*_C/D	*Ax1, Ax2**	62°C	Lafiandra et al., 1997
PP9	UMN19_F/R	*Ax1, Ax2**	60°C	Liu et al., 2008
PP10	Ax*Null*_G1/G2	*AxNull*	62°C	Lafiandra et al., 1997

*See Table 2 for detailed primer sequences. The subunits discrimination is based on literature data.

4.2.2.1 *Glu-1* allele diversity

For discrimination between the *Glu-A1a* and *Glu-A1b* genes, the three primer pairs used (PP7, PP8 and PP9, **Table 14**) gave same results. Out of the 45 genotypes tested, 10 were identified as carriers of the *Ax2** gene, while the remaining 35 revealed typical *Ax1* band. In **Fig. 36**, examples of the results obtained by testing PP9 on the SSD genotypes are reported; genotypes highlighted in green showed the *Ax2** amplicon of 344 bp, while those in white showed the *Ax1* band of 362 bp. Moreover, the use of PP10 (**Table 14**) on SSD genotypes randomly selected, allowed to confirm them as characterized by the presence of the *AxNull* allele, obtaining, for everyone, the expected 920 bp fragment (data not shown).

The genotyping exercise slightly increased the frequency of *Glu-A1a* at the expense of *Glu-A1b* over what had been deduced from the SDS-PAGE analysis. Such differences in results are consistent with what observed by Lafiandra et al. (1997) who revealed, by SDS-PAGE analysis of primitive cultivars and landraces of durum and bread wheats, the presence of few allelic variants at *Glu-A1* locus which are difficult to assign as *Ax1* or *Ax2** on the basis of their electrophoretic mobility.

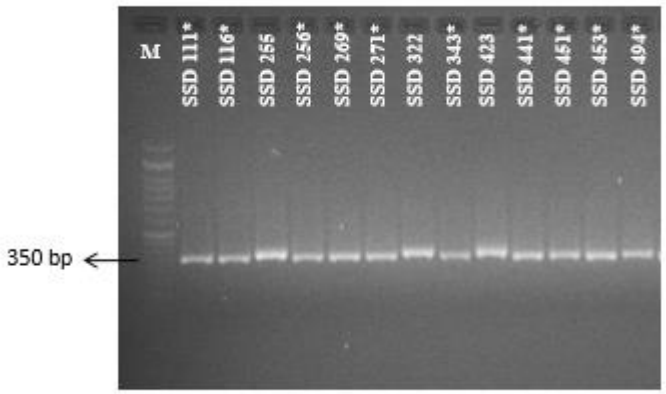

Figure 36: PCR products of SSD genotypes amplified with PP9 (UMN19_F/R); genotypes carrying the *Ax2** gene are those indicated with an asterisk, while theose without the asterisk carried the *Ax1* gene. NZYDNA ladder VI was used as marker.

A summary of the alleles identified at *Glu-A1* locus and their relative frequencies is shown in **Table 15**. In over 70.39% of the SSD population tested, the null allele was the most frequent followed by Ax1 (23.02%) and Ax2* (6.57%)**.** The same order in terms of allele frequencies (null> 1> 2*) is in agreement with previous studies of landraces and old varieties of durum wheat (Moragues et al., 2006; Nazco et al., 2012). The overall diversity at this locus, as expressed by H was 0.45, significantly lower than the one observed at the *Glu-B1* locus (**Tab. 15**). With the exception of the null allele, 45 (29.59%) among the total 152 SSD genotypes had either Ax1 or Ax2* alleles which are known to impart better quality to wheat flour, even if their contribution is lower than those of proteins coded by the *Glu-B1* locus (Rasheed et al., 2012). As concerning the allele diversity at *Glu-B1* locus, 47 of the 48 entries were genotyped using six molecular markers (PP1 to PP6, **Table 14**). Only for one genotype (SSD 111, **Table 16**) was not possible to determine the composition at the *Glu-B1* locus.
The application of PP1 (**Fig. 37**) and PP4 on the SSD genotypes previously assigned as Bx7+By8 GS (*Glu-B1b*) showed that eight, out of 45, carried the Bx7+By8* allele (*Glu-B1al*), while two carried the Bx7*+By8* allele (*Glu-B1ak*) (**Table 15**). Additionally, by performing a cross-analyses with PP2, PP3 (**Fig. 38a** and **38b**, respectively) and PP6 separately, two rare subunits were detected (GS Bx14+By20). Finally, the application firstly of PP1 and, then, of PP5 to the same genotypes led to the identification of one SSD characterized by the GS Bx6+By8*.

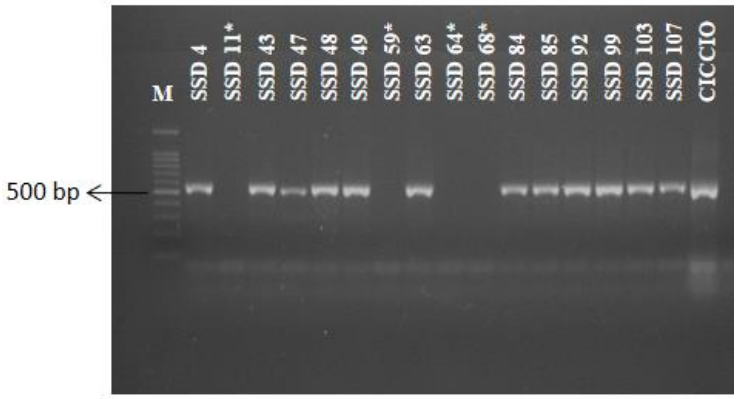

Figure 37: Examples of PCR products obtained by testing the assigned Bx7+By8 entries with PP1. AccuRuler 100bp DNA ladder was used as marker. SSD 11, 59, 64 and 68 (highlighted with an asterisk) lacked the By8 typical band.

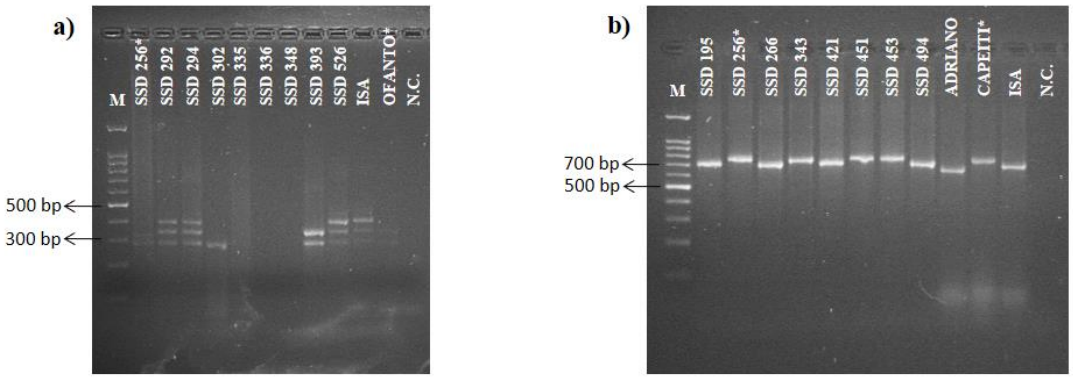

Figure 38: Examples of PCR products obtained by testing the uncertain genotypes with **a**) PP2 and **b**) PP3. AccuRuler 100bp DNA ladder was used as marker in both analyses. Cultivars Isa (Bx13+By16) and Ofanto (Bx20+By20), gel a, and cultivars Adriano (Bx7+By9), Capeiti (Bx20+By20) and Isa (Bx13+By16), gel b, were used as references. An example of entry showing the typical Bx20+By20 patter either with PP2 and PP3 is highlighted with an asterisk, finally revealing the Bx14+By20 GS.

The complete list of the 152 entries and the set of HMW-GSs detected at both *Glu-A1* and *Glu-B1* loci through the combined "Omic" approach described, is reported in **Table 16**.

Table 15: Allele frequency and genetic diversity at the two *Glu-1* loci.

Locus	Allele	Subunit	Number of genotypes	Frequency (%)	H (Nei's index)
Glu-A1	*a*	Ax1	35	23.02	
	b	Ax2*	10	6.57	**0.45**
	c	Null	107	70.39	
GluB1	*a*	Bx7	3	1.97	
	b	Bx7+By8	35	23.02	
	an	Bx6	1	0.65	
	d	Bx6+By8	28	18.42	
	e	Bx20+By20	47	30.92	
	f	Bx13+By16	9	5.92	
	h	Bx14+By15	1	0.65	
	-	Bx14+By19	8	5.26	
	-	Bx14+By20	2	1.31	**0.803**
	-	Bx7+By19	3	1.97	
	z	Bx7+By15	1	0.65	
	al	Bx7+By8*	8	5.26	
	ak	Bx7*+By8*	2	1.31	
	-	Bx6+By8*	1	0.65	
	1g	Bx14	1	0.65	
	abnormal	Bx7+By8; Bx20+By20	1	0.65	
	undetermined		1	0.65	

- Alleles not annotated

Table 16. Allelic status at the *Glu-A1* and *Glu-B1* loci of the 152 lines germplasm set.

SSD genotypes	**HMW-GS**	
	Glu-A1	*Glu-B1*
2	*null*	Bx6+By8
3	*null*	Bx6+By8
4	*null*	Bx7+By8
6	Ax1	Bx20+By20
7	*null*	Bx20+By20
11	*null*	Bx7+BY8*
15	*nnll*	Bx20+By20
17	*null*	Bx13+By16
24	*null*	Bx6+By8
27	*null*	Bx6+By8
28	*null*	Bx6+By8
32	*null*	Bx6+By8
33	Ax1	Bx20+By20
34	*null*	Bx6+By8
35	Ax1	Bx6+By8
36	*null*	Bx6+By8
41	*null*	Bx6+By8
43	*null*	Bx7+By8
44	Ax1	Bx6+By8
47	*null*	Bx7+By8
48	*null*	Bx7+By8
49	*null*	Bx7+By8
52	*null*	Bx6
54	Ax1	**Bx20+By20; Bx7+By8**
59	Ax1	Bx7+By8*
63	*null*	Bx7+By8
64	*null*	Bx7+By8*
65	*null*	Bx20+By20
66	*null*	Bx7+By8*
67	*null*	Bx7+By8*
68	*null*	Bx7+By8*
69	*null*	Bx6+By8
70	*null*	Bx13+By16
74	*null*	Bx13+By16
83	*null*	Bx6+By8
84	Ax1	Bx7+By8
85	*null*	Bx7+By8
92	*null*	Bx7+By8
96	*null*	Bx7
99	*null*	Bx7+By8
102	*null*	Bx14+By19
103	*null*	Bx7+By8
105	*null*	Bx14+By15
107	*null*	Bx7+By8

Table 16: Continued.

SSD genotypes	HMW-GS	
	Glu-A1	*Glu-B1*
109	*null*	Bx7+By8
111	Ax2*	**ND**
112	*null*	Bx7+By8
113	*null*	Bx7+By8
114	*null*	Bx7+By8
115	*null*	Bx7+By8
116	Ax2*	Bx6+By8
118	*null*	Bx7+By8
120	*null*	Bx7+By8
122	*null*	Bx6+By8
123	*null*	Bx20+By20
124	*null*	Bx6+By8
125	*null*	Bx20+By20
128	*null*	Bx7+By8*
131	*null*	Bx7+By8
135	*null*	Bx20+By20
136	Ax1	Bx20+By20
137	*null*	Bx14+By19
142	Ax1	Bx20+By20
146	Ax1	Bx20+By20
147	Ax1	Bx13+By16
155	*null*	Bx7
158	*null*	Bx7+By8
168	*null*	Bx6+By8
171	Ax1	Bx6+By8
173	Ax1	Bx7+By15
178	*null*	Bx14+By19
180	*null*	Bx7
182	*null*	Bx20+By20
185	Ax1	Bx20+By20
188	Ax1	Bx7+By8
191	*null*	Bx7+By8
195	Ax1	Bx7+By8
240	*null*	Bx20+By20
243	*null*	Bx13+By16
244	*null*	Bx20+By20
246	*null*	Bx6+By8
253	*null*	Bx20+By20
255	Ax1	Bx7+By8
256	Ax2*	Bx14+By20
262	*null*	Bx20+By20
264	*null*	Bx6+By8
266	*null*	Bx14+By19
269	Ax2*	Bx7+By8

Table 16: Continued.

SSD genotypes	HMW-GS	
	Glu-A1	*Glu-B1*
271	Ax2*	Bx7+By8
278	Ax1	Bx20+By20
283	*null*	Bx20+By20
288	Ax1	Bx7+By8
290	*null*	Bx6+By8*
292	*null*	Bx13+By16
294	*null*	Bx13+By16
298	Ax1	Bx6+BY8
302	*null*	Bx14+By19
303	*null*	Bx20+By20
308	*null*	Bx6+By8
315	*null*	Bx7+By8*
322	Ax1	Bx14
325	*null*	Bx7+By8
326	*null*	Bx14+By19
328	*null*	Bx7+By8
330	Ax1	Bx20+By20
335	*null*	Bx14+By20
336	Ax1	Bx14+By19
338	*null*	Bx7+By8
343	Ax2*	Bx7+By19
345	*null*	Bx20+By20
348	*null*	Bx14+By19
350	*null*	Bx7+By8
393	*null*	Bx7*+By8*
397	Ax1	Bx20+By20
399	Ax1	Bx20+By20
400	Ax1	Bx20+By20
407	*null*	Bx20+By20
409	Ax1	Bx20+By20
411	Ax1	Bx20+By20
412	*null*	Bx20+By20
414	*null*	Bx20+By20
415	*null*	Bx20+By20
416	*null*	Bx20+By20
421	*null*	Bx13+By16
422	*null*	Bx20+By20
423	Ax1	Bx20+By20
424	Ax1	Bx20+By20
426	Ax1	Bx20+By20
427	Ax1	Bx20+By20
431	Ax1	Bx20+By20
432	Ax1	Bx20+By20
441	Ax2*	Bx20+By20

Table 16: Continued.

SSD genotypes	HMW-GS	
	Glu-A1	*Glu-B1*
443	*null*	Bx7+By8
447	*null*	Bx20+By20
451	Ax2*	Bx7+By19
453	Ax2*	Bx7+By19
457	*null*	Bx7+By8
459	*null*	Bx6+By8
467	*null*	Bx20+By20
470	Ax1	Bx20+By20
487	*null*	Bx20+By20
494	Ax2*	Bx7*+By8*
499	*null*	Bx20+By20
500	*null*	Bx20+By20
511	*null*	Bx6+By8
513	*null*	Bx20+By20
526	*null*	Bx13+By16
531	*null*	Bx6+By8
532	*null*	Bx6+By8
533	*null*	Bx6+By8
533bis	*null*	Bx6+By8
534	*null*	Bx7+By8

ND Not Determined

The Bx20+By20 (*Glu-B1e*; 30,9%) was the most frequent allele identified within the SSD genotypes, followed by the Bx7+By8 (*Glu-B1b*, 23,0%) and the Bx6+By8 (*Glu-B1d*, 18,4%): over 70% of the entries carried one of these three alleles. The most frequently encountered minor alleles were the 13+16 (*Glu-B1f*, 5,9%), and the Bx7+By8* (*Glu-B1al*, 5,3%). The non-assigned allele encoding the GS combination Bx7+By19 was carried by three entries, while the GS combinations Bx14+By20 and Bx7*+By8* (*Glu-B1ak*) were both present in two entries; finally, the GS combinations Bx7+By15 (*Glu-B1z*), Bx14+By19, Bx14+By15 (*Glu-B1h*) and Bx6+By8* were each identified in just a single entry. Five entries produced an x-type but not a y-type *Glu-B1* subunit, of which three carried the GS Bx7 (*Glu-B1a*), one produced the GS Bx14 (*Glu-B1g*) and one carried the GS Bx6 (*Glu-B1an*) (**Table 15**). Finally, one entry of the whole subset analyzed (0,65%) was heterogeneous/heterozygous at the *Glu-B1*, carrying both Bx6+By8 (*Glu-B1d*) and Bx20+By20 (*Glu-B1e*) allelic combinations (**Table 15**), consistently with reported abnormality at *Glu-B1* locus found in wheat landraces (Ribeiro et al., 2011; Zheng et al., 2011). The diversity present at *Glu-B1* was considerably greater than at *Glu-A1* (0,80 versus 0,45).

In general, the Hs indices at the two loci observed in the SSD lines were higher than those found by Bellil et al. (2014) (0.03 and 0.67 for the *Glu-A1* and *Glu-B1* loci respectively), while they were similar to those reported by Ribeiro et al. (2011) for old Portuguese varieties, with a slightly lower value for the *Glu-A1* locus.

An overall representation of the *Glu-A1* and *Glu-B1* composition for the 152 SSD is reported in **Fig. 39a** and **39b** respectively.

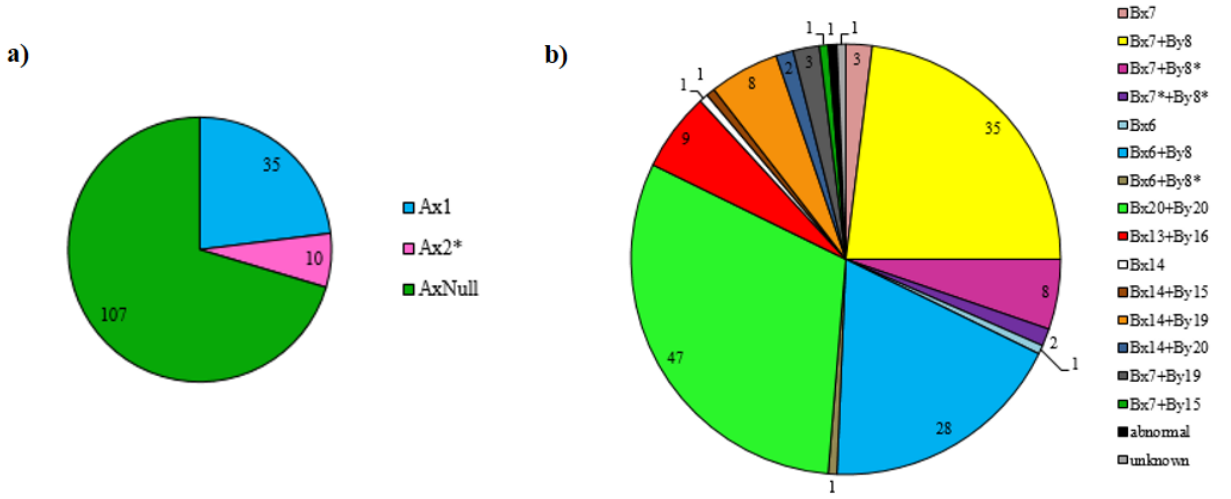

Figure 39: *Glu-1* locus composition of the 152 durum wheat lines **a)** *Glu-A1* and **b)** *Glu-B1*. The number in the pie charts indicates the entries within each category.

A summary of the HMW-GS composition at both *Glu-A1* and *Glu-B1* loci of the SSD genotypes and the corresponding frequencies, are presented in **Table 17**. Twenty-seven different combinations were represented across the collection. The four combinations *null*, Bx7+By8 (*Glu-A1c/Glu-B1b*; 28/152, 18,4%), *null*, Bx20+By20 (*Glu-A1c/Glu-B1e*; 26/152, 17,1%), *null*, Bx6+By8 (*Glu-A1c/Glu-B1d*; 23/152, 15,1%) and Ax1, Bx20+By20 (*Glu-A1a/Glu-B1e*; 20/152, 13,2%) predominated. Eight entries carried the combination *null*, Bx13+By16 (*Glu-A1c/Glu-B1f*), seven carried *null*, Bx7+By8* (*Glu-A1c/Glu-B1al*) seven carried *null* (*Glu-A1c*), *Glu-B1* GS Bx14+By19, five carried Ax1, Bx7+By8 (*Glu-A1a/Glu-B1b*), four carried *null*, Bx6+By8 (*Glu-A1a/Glu-B1d*), three carried *null*, Bx7 (*Glu-A1c/Glu-B1a*), three carried Bx7+By19 and two carried Ax2*, Bx7+By8 (*Glu-A1b/Glu-B1b*). The other 15 GS combinations were each carried by just one entry.

Some of the most representative combinations are characterized by rare alleles at *Glu-B1* locus; among these 5,26% is represented by the Bx13+By16 GS, 4,6% by both the Bx7+By8* and Bx14+By19 GSs and 1,97% by the Bx7+By19 GS (**Table 17**). Such remarks were expected since the

germplasm lines analyzed are principally composed by old varieties and landraces, on which no selection in terms of qualitative traits or productivity has been made.

Table 17: The frequency of combined *Glu-A1* and *Glu-B1* genotype within the 152 durum wheat lines.

***Glu-A1* and *Glu-B1* Alleles**	***Glu-A1* and *Glu-B1* Subunits**	**# Genotypes**	**Frequencies %**
c, b	(*null*, Bx7+By8)	28	18.42
c, e	(*null*, Bx20+By20)	26	17.1
c, d	(*null*, Bx6+By8)	23	15.13
a, e	(Ax1, Bx20+By20)	20	13.15
c, f	(*null*, Bx13+By16)	8	5.26
c, al	(*null*, Bx7+By8*)	7	4.6
c, -	(*null*, Bx14+By19)	7	4.6
a, b	(Ax1, Bx7+By8)	5	3.29
a, d	(Ax1, Bx6+By8)	4	2.63
c, a	(*null*, Bx7)	3	1.97
b, -	(Ax2*, Bx7+By19)	3	1.97
b, b	(Ax2*, Bx7+By8)	2	1.31
a, al	(Ax1, Bx7+By8*)	1	0.66
b, ak	(Ax2*, Bx7*+By8*)	1	0.66
c, ak	(*null*, Bx7*+By8*)	1	0.66
c, an	(*null*, Bx6)	1	0.66
b, d	(Ax2*, Bx6+By8)	1	0.66
c, -	(*null*, Bx6+By8*)	1	0.66
b, e	(Ax2*, Bx20+By20)	1	0.66
a, f	(Ax1, Bx13+By16)	1	0.66
c, h	(*null*, Bx14+By15)	1	0.66
a, 1g	(Ax1, Bx14)	1	0.66
a, -	(Ax1, Bx14+By19)	1	0.66
c, -	(*null*, Bx14+By20)	1	0.66
b, -	(Ax2*, Bx14+By20)	1	0.66
a, z	(Ax1, Bx7+By15)	1	0.66
a, d+e	(Ax1, Bx6+By8/Bx20+By20)	1	0.66
Undetermined		1	0.66

4.2.3 Geographic distribution of *Glu-1* alleles

The geographic distribution and the relative frequencies of the various *Glu-A1* and *Glu-B1* alleles across the set of 152 entries are displayed in **Fig. 40**, **Fig. 41** and **Table 18**. With respect to *Glu-A1*, the most frequently encountered allele in almost all countries of origin (the exceptions were Afghanistan, Peru, Bulgaria and Saudi Arabia) was *Glu-A1c*, which was present in 107 of the 152

entries. Within the 152 entries, the highest frequencies of the *Glu-A1c* allele were observed among entries originating from Tunisia (9.9%) followed by Italy (8.6%), USA (6.6%), Ethiopia (5.9%), Iraq (5.3%) and Morocco (5.3%), **Table 18**). The *Glu-A1a* allele, was detected in 35 of the 152 entries and was particularly frequent in material from S.E. European (Greece: 5.3%, Crete: 2.0%, Turkey: 2.0%), Iberian Peninsula (3.0%) or N. African (Egypt and Algeria: 2.0%, Tunisia: 1.3%) provenance. The *Glu-A1b* allele was represented in eight entries originating from S.W. Asia (Iraq and Iran, each 2.6%) and two from Greece (1.3%). With respect to *Glu-B1*, the most frequent allele was *Glu-B1e* (47/152 entries): this allele was concentrated in materials originating from S.E. Europe (Greece, Bulgaria, Cyprus), S.W. Asia (Turkey, Jordan) and India. S.W. Asia (Iraq, Iran, Syria, Saudi Arabia, Azerbaijan), India and N. Africa (Morocco, Tunisia, Libya and Egypt) provided the majority of the second most common allele *Glu-B1b* (35/152). These same areas featured a substantial level of diversity at *Glu-B1*, with relative high frequencies of *Glu-B1al* (4.6%), *Glu-B1f* (3.3%) and GS Bx14+By19 (1.3%). The *Glu-B1d* allele (28/152) was most strongly associated with a N. African or N. American provenance and was not represented at all among entries originating from around the Black Sea (Turkey and Bulgaria). The *Glu-B1f* allele, although globally rare, was relatively common in entries derived from N. Africa and S. Europe, while the alleles GS Bx14+By19, GS Bx14+By20 and GS Bx7+By19 were encountered in material from S.W. Asia and India (respectively 2,6%, 1,3% and 2,0%). N. American and Italian materials were dominated by carriers of the three high frequency alleles *Glu-B1e*, *Glu-B1b* and Glu-*B1d*.

Table 18: Allele frequency at *Glu-A1* and *Glu-B1* for entries grouped by provenance.

Locus	Allele	Total[$] (152)	Afghanistan[#] (Tot. 1)	Total[$] (152)	Algeria[#] (Tot. 6)	Total[$] (152)	Saudi Arabia[#] (Tot. 1)	Total[$] (152)	Azerbaijan[#] (Tot. 1)	Total[$] (152)	Bosnia Herzegovina[#] (Tot. 1)
Glu-A1											
	Ax1	0.65	100	1.97	50	0.65	100	/	/	/	/
	Ax2*	/	/	/	/	/	/	/	/	/	/
	Null	/	/	1.97	50	/	/	0.65	100	0.65	100
Glu-B1											
	Bx7	/	/	/	/	/	/	0.65	100	/	/
	Bx7+By8	0.65	100	0.65	16.67	0.65	100	/	/	/	/
	Bx6	/	/	/	/	/	/	/	/	/	/
	Bx6+By8	/	/	2.63	66.66	/	/	/	/	/	/
	Bx20+By20	/	/	0.65	16.67	/	/	/	/	/	/
	Bx13+By16	/	/	/	/	/	/	/	/	/	/
	Bx14+By15	/	/	/	/	/	/	/	/	/	/
	Bx14+By19	/	/	/	/	/	/	/	/	0.65	100
	Bx14+By20	/	/	/	/	/	/	/	/	/	/
	Bx7+By8*	/	/	/	/	/	/	/	/	/	/
	Bx7*+By8*	/	/	/	/	/	/	/	/	/	/
	Bx6+By8*	/	/	/	/	/	/	/	/	/	/
	Bx7+By19	/	/	/	/	/	/	/	/	/	/
	Bx7+By15	/	/	/	/	/	/	/	/	/	/
	Bx14	/	/	/	/	/	/	/	/	/	/
	Abnormal	/	/	/	/	/	/	/	/	/	/
	Unknown	/	/	/	/	/	/	/	/	/	/

[$]: Frequencies of the alleles considering the total 152 SSD genotypes analyzed.
[#]: Percentage of the single allele within the entries of the count

Table 18: Continued.

Locus	Allele	Total[$] (152)	Bulgaria[#] (Tot. 1)	Total[$] (152)	Cyprus[#] (Tot. 1)	Total[$] (152)	Crete[#] (Tot. 7)	Total[$] (152)	Egypt[#] (Tot. 4)	Total[$] (152)	Ethiopia[#] (Tot. 10)	Total[$] (152)	France[#] (Tot. 2)	Total[$] (152)	Japan[#] (Tot. 1)
Glu-A1															
	Ax1	0.65	100	/	/	1.97	42.86	1.97	75	0.65	10	/	/	/	/
	Ax2*	/	/	/	/	0.65	14.28	/	/	/	/	/	/	/	/
	Null	/	/	0.65	100	1.97	42.86	0.65	25	5.92	90	1.31	100	0.65	100
Glu-B1															
	Bx7	/	/	/	/	/	/	/	/	/	/	/	/	/	/
	Bx7+By8	/	/	/	/	/	/	/	/	2.63	40	/	/	/	/
	Bx6	/	/	/	/	/	/	0.65	25	/	/	/	/	/	/
	Bx6+By8	/	/	/	/	/	/	/	/	1.31	20	/	/	/	/
	Bx20+By20	0.65	100	0.65	100	3.94	85.71	/	/	1.31	20	0.65	50	/	/
	Bx13+By16	/	/	/	/	0.65	14.29	/	/	0.65	10	/	/	0.65	100
	Bx14+By15	/	/	/	/	/	/	/	/	/	/	/	/	/	/
	Bx14+By19	/	/	/	/	/	/	/	/	0.65	10	0.65	50	/	/
	Bx14+By20	/	/	/	/	/	/	/	/	/	/	/	/	/	/
	Bx7+By8*	/	/	/	/	/	/	0.65	25	/	/	/	/	/	/
	Bx7*+By8*	/	/	/	/	/	/	/	/	/	/	/	/	/	/
	Bx6+By8*	/	/	/	/	/	/	/	/	/	/	/	/	/	/
	Bx7+By19	/	/	/	/	/	/	/	/	/	/	/	/	/	/
	Bx7+By15	/	/	/	/	/	/	0.65	25	/	/	/	/	/	/
	Bx14	/	/	/	/	/	/	/	/	/	/	/	/	/	/
	abnormal	/	/	/	/	/	/	0.65	25	/	/	/	/	/	/
	unknown	/	/	/	/	/	/	/	/	/	/	/	/	/	/

[$]: Frequencies of the alleles considering the total 152 SSD genotypes analyzed.
[#]: Percentage of the single allele within the entries of the country.

Table 18: Continued.

Locus	Allele	Total[$] (152)	Jordan[#] (Tot.1)	Total[$] (152)	Greece[#] (Tot. 16)	Total[$] (152)	India[#] (Tot. 3)	Total[$] (152)	Iran[#] (Tot. 9)	Total[$] (152)	Iraq[#] (Tot. 13)	Total[$] (152)	Italy[#] (Tot. 13)	Total[$] (152)	Yugoslavia[#] (Tot. 1)
Glu-A1															
	Ax1	/	/	5.26	50	/	/	/	/	0.65	7.7	/	/	/	/
	Ax2*	/	/	0.65	6.25	/	/	2.63	44.44	2.63	30.76	/	/	/	/
	Null	0.65	100	4.6	43.75	1.97	100	3.29	55.56	5.26	61.54	8.55	100	0.65	100
Glu-B1															
	Bx7	/	/	/	/	1.31	66.67	/	/	/	/	/	/	/	/
	Bx7+By8	/	/	0.65	6.25	/	/	2.63	44.44	3.95	46.15	1.97	23.08	0.65	100
	Bx6	/	/	/	/	/	/	/	/	/	/	/	/	/	/
	Bx6+By8	/	/	/	/	/	/	/	/	0.65	7.7	3.29	38.46	/	/
	Bx20+By20	0.65	100	9.21	87.5	0.65	33.33	0.65	11.11	/	/	2.63	30.77	/	/
	Bx13+By16	/	/	/	/	/	/	/	/	/	/	0.65	7.69	/	/
	Bx14+By15	/	/	/	/	/	/	/	/	0.65	7.7	/	/	/	/
	Bx14+By19	/	/	/	/	/	/	1.31	22.22	0.65	7.7	/	/	/	/
	Bx14+By20	/	/	/	/	/	/	0.65	11.11	0.65	7.7	/	/	/	/
	Bx7+By8*	/	/	/	/	/	/	/	/	/	/	/	/	/	/
	Bx7*+By8*	/	/	0.65	6.25	/	/	/	/	/	/	/	/	/	/
	Bx6+By8*	/	/	/	/	/	/	/	/	/	/	/	/	/	/
	Bx7+By19	/	/	/	/	/	/	0.65	11.11	1.31	15.38	/	/	/	/
	Bx7+By15	/	/	/	/	/	/	/	/	/	/	/	/	/	/
	Bx14	/	/	/	/	/	/	/	/	/	/	/	/	/	/
	abnormal	/	/	/	/	/	/	/	/	/	/	/	/	/	/
	unknown	/	/	/	/	/	/	/	/	0.65	7.7	/	/	/	/

[$]: Frequencies of the alleles considering the total 152 SSD genotypes analyzed.
[#]: Percentage of the single allele within the entries of the country.

Table 18: Continued.

Locus	Allele	Total[$] (152)	Libya[#] (Tot. 2)	Total[$] (152)	Morocco[#] (Tot. 9)	Total[$] (152)	Peru[#] (Tot. 1)	Total[$] (152)	Portugal[#] (Tot. 2)	Total[$] (152)	Romania[#] (Tot. 1)	Total[$] (152)	Russia[#] (Tot.1)	Total[$] (152)	Syria[#] (Tot. 3)
Glu-A1															
	Ax1	/	/	0.65	11.11	0.65	100	0.65	50	/	/	/	/	0.65	33.33
	Ax2*	/	/	/	/	/	/	/	/	/	/	/	/	/	/
	Null	1.31	100	5.26	88.89	/	/	0.65	50	0.65	100	0.65	100	1.31	66.67
Glu-B1															
	Bx7	/	/	/	/	/	/	/	/	/	/	/	/	/	/
	Bx7+By8	/	/	0.65	11.11	/	/	/	/	/	/	0.65	100	1.97	100
	Bx6	/	/	/	/	/	/	/	/	/	/	/	/	/	/
	Bx6+By8	0.65	50	0.65	11.11	0.65	100	/	/	/	/	/	/	/	/
	Bx20+By20	/	/	1.31	22.22	/	/	1.31	100	/	/	/	/	/	/
	Bx13+By16	/	/	0.65	11.11	/	/	/	/	/	/	/	/	/	/
	Bx14+By15	/	/	/	/	/	/	/	/	/	/	/	/	/	/
	Bx14+By19	/	/	/	/	/	/	/	/	/	/	/	/	/	/
	Bx14+By20	/	/	/	/	/	/	/	/	/	/	/	/	/	/
	Bx7+By8*	0.65	50	2.63	44.45	/	/	/	/	/	/	/	/	/	/
	Bx7*+By8*	/	/	/	/	/	/	/	/	0.65	100	/	/	/	/
	Bx6+By8*	/	/	/	/	/	/	/	/	/	/	/	/	/	/
	Bx7+By19	/	/	/	/	/	/	/	/	/	/	/	/	/	/
	Bx7+By15	/	/	/	/	/	/	/	/	/	/	/	/	/	/
	Bx14	/	/	/	/	/	/	/	/	/	/	/	/	/	/
	abnormal	/	/	/	/	/	/	/	/	/	/	/	/	/	/
	unknown	/	/	/	/	/	/	/	/	/	/	/	/	/	/

[$]: Frequencies of the alleles considering the total 152 SSD genotypes analyzed.
[#]: Percentage of the single allele within the entries of the country.

Table 18: Continued.

Locus	Allele	Total[$] (152)	Spain[#] (Tot. 5)	Total[$] (152)	Tunisia[#] (Tot. 17)	Total[$] (152)	Turkey[#] (Tot. 6)	Total[$] (152)	Ukraine[#] (Tot. 2)	Total[$] (152)	USA[#] (Tot. 11)
Glu-A1											
	Ax1	1.97	60	1.31	11.76	1.97	50	/	/	0.65	9.1
	Ax2*	/	/	/	/	/	/	/	/	/	/
	Null	1.31	40	9.87	88.24	1.97	50	1.31	100	6.58	90.9
Glu-B1											
	Bx7	/	/	/	/	/	/	/	/	/	/
	Bx7+By8	1.31	40	0.65	5.89	/	/	1.31	100	1.97	27.3
	Bx6	/	/	/	/	/	/	/	/	/	/
	Bx6+By8	/	/	5.26	47.05	0.65	16.67	/	/	2.63	36.3
	Bx20+By20	1.31	40	1.31	11.75	1.97	50	/	/	1.97	27.3
	Bx13+By16	0.65	20	1.97	17.64	/	/	/	/	/	/
	Bx14+By15	/	/	/	/	/	/	/	/	/	/
	Bx14+By19	/	/	0.65	5.89	0.65	16.67	/	/	/	/
	Bx14+By20	/	/	/	/	/	/	/	/	/	/
	Bx7+By8*	/	/	0.65	5.89	/	/	/	/	0.65	9.1
	Bx7*+By8*	/	/	/	/	/	/	/	/	/	/
	Bx6+By8*	/	/	0.65	5.89	/	/	/	/	/	/
	Bx7+By19	/	/	/	/	/	/	/	/	/	/
	Bx7+By15	/	/	/	/	/	/	/	/	/	/
	Bx14	/	/	/	/	0.65	16.67	/	/	/	/
	abnormal	/	/	/	/	/	/	/	/	/	/
	unknown	/	/	/	/	/	/	/	/	/	/

[$]: Frequencies of the alleles considering the total 152 SSD genotypes analyzed.
[#]: Percentage of the single allele within the entries of the country.

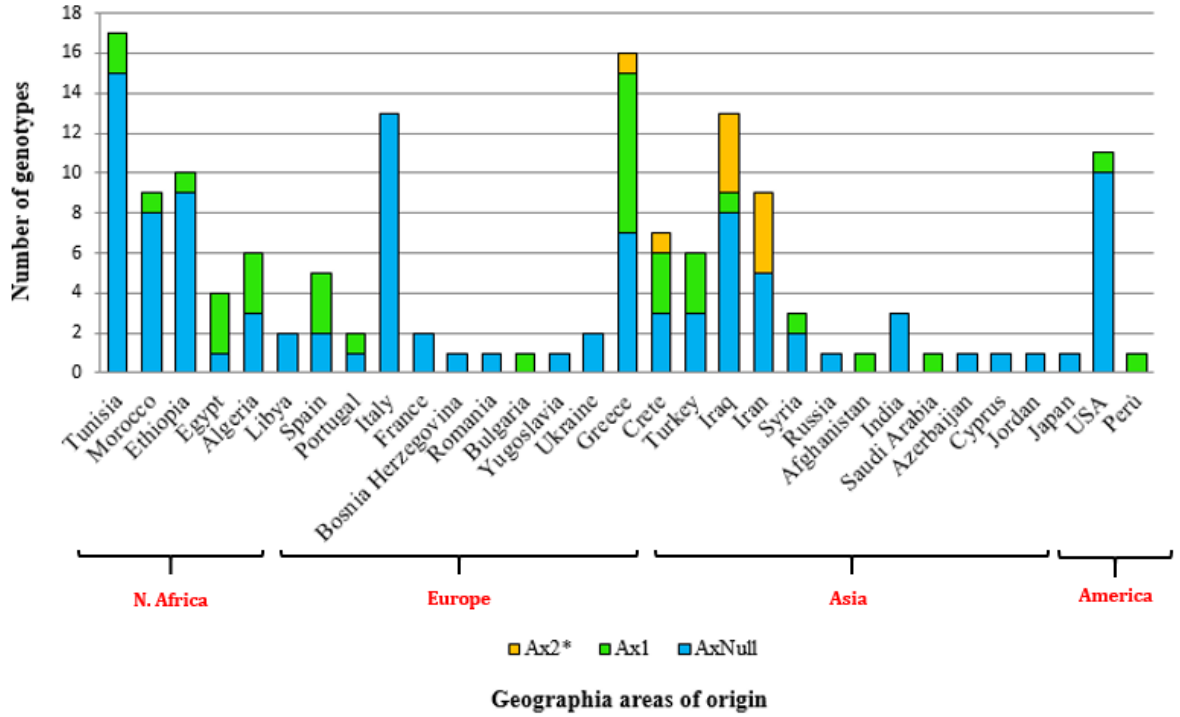

Figure 40: *Glu-A1* status and the geographic provenance of the germplasm entries. Countries are grouped according to the corresponding continent.

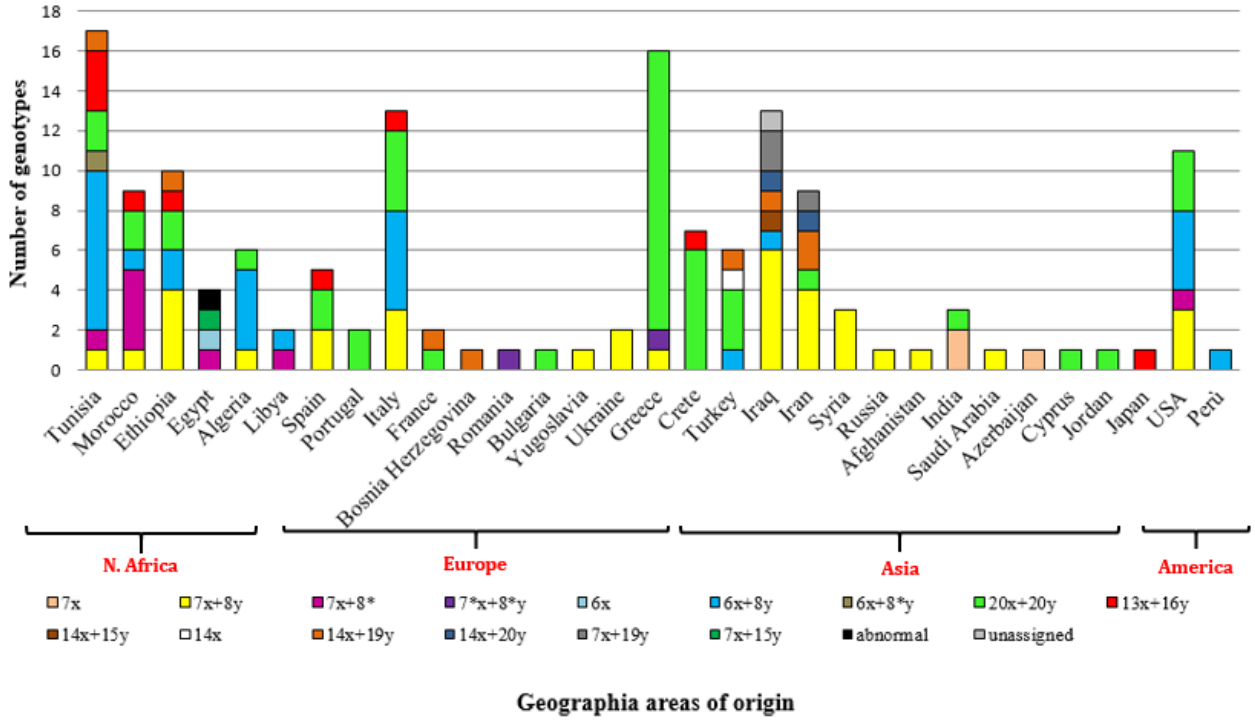

Figure 41: *Glu-B1* status and the geographic provenance of the germplasm entries. Countries are grouped according to the corresponding continent.

Since the geographic distribution of the present genetic diversity within a species can be informative about the historical processes that shaped the genetic diversity (Wang et al., 2017), a cluster analysis (hierarchical clustering and a principal component analysis (PCA)) was performed This cluster analysis showed that Greece clustered alone, with the presence of all three alleles albeit with a small number of Ax2 * genotypes (**Fig. 42a**). The same profile, although located in a different branch of the cluster, probably due to the lower number of entries, is shared by Crete. Italy is grouped in the same cluster with Tunisia, as both share the almost exclusive presence of the null allele, as well as Iran and Iraq, which are clustered together since representative of Ax2* genotypes; as a matter of fact, the latter ones represent the 44.44% and the 30.77% respectively. America, Morocco and Ethiopia clustered together due to the large number of Ax*Null* genotypes, followed by a small number of Ax1 genotypes. Finally, Algeria, Turkey, Egypt and Spain are grouped since they share for the 50% of the entries the *Null* allele and for the remaining 50% the Ax1 allele. Similar clustering was obtained by analyzing data belonging to the *Glu-B1* locus (**Fig. 42b**). From this analyses Greece genotypes forms a separate providing the highest number of genotypes carrying the Bx20+By20 allele. Italy, America and part of N. Africa (Tunisia and Algeria), are grouped together for the general predominance of the Bx6+By8 GS, followed by GSs Bx7+By8 and Bx20+By20. Italy and Tunisia also share the Bx13+By16 GS. Within the same cluster, however, Tunisia diverges from the others since it exhibits the greatest number of *Glu-B1* allelic variants. Also in this grouping, Crete is clustered separately from Greece and at the opposite end of the dendrogram, despite the high number of entries Bx20+By20. This is probably due to the lower number of germplasm lines deriving from this area (tot. 7) compared to those originating from Greece (tot. 16). Iraq, Iran, Ethiopia, Morocco, and Tunisia are the countries showing the greatest genetic variability. In particular, Iraq and Iran present the largest number of rare alleles (GSs Bx14+By20, Bx14+By19, and Bx7+By19). Moreover, the GS Bx14+By19 is also identified in Tunisia and Ethiopia in conjunction with other rare alleles such as Bx7+By8* and Bx6+By8*. Finally, a high number of countries are located in the central clusters of the dendrogram probably for the limited number of entries in the collection from Yugoslavia, Russia, Afghanistan, Saudi Arabia Portugal, Bulgaria, Cyprus, Jordan Spain, Syria and Ukraine.

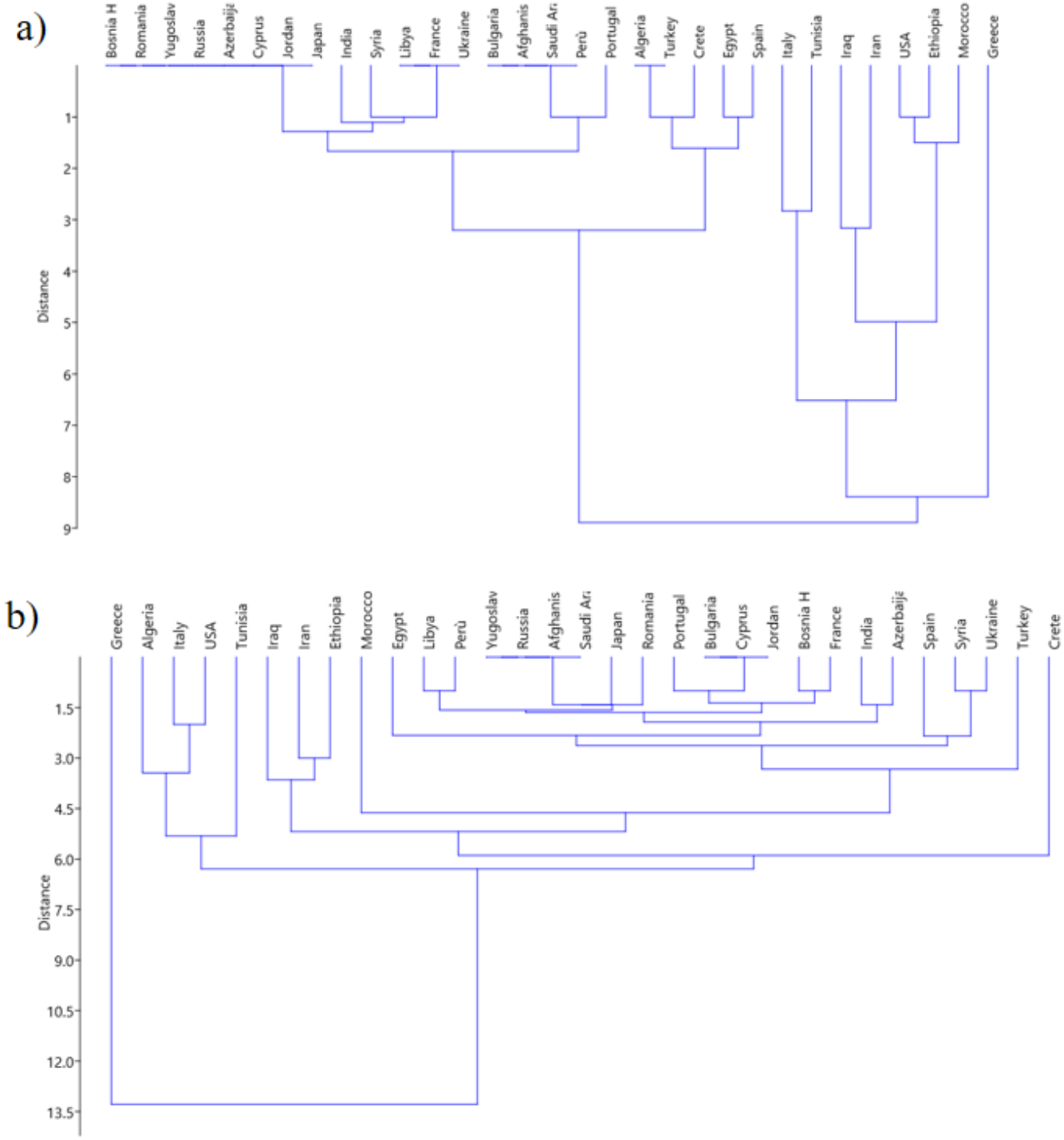

Figure 42: Dendrogram showing the relationships between the allelic composition of the germplasm lines and their provenance, based on unweighted pair-group average (UPMGA) method: **a)** as concerning the *Glu-A1* locus; **b)** as concerning the *Glu-B1* locus.

The PCA analysis was also performed in order to assess the individual differences.

In the case of *Glu-A1* locus (**Fig. 43**), the first two components account for 95% of the variance. The first coordinate accounting for 81.5% of the variance, is clearly discriminating the SSD carrying the *Null* allele. Tunisia and Italy (15 and 13 SSD AxNull, respectively) were located on the far right of the plot, immediately followed by USA (10), Ethiopia (9), Iraq and Morocco (8). The second

coordinate, which accounts for 13.5% of the variance, cluster the SSD having the Ax1 allele; indeed, Greece, which showed a total of 8 SSD out of 16 with this allelic combination, was located at the highest end of the y axis.

In the PCA analysis relative to the *Glu-B1* locus the first two components represent almost 76% of the variance. The first coordinate accounting for 49.6% of the variance represent the number of SSD carrying the Bx20+By20 allele and clearly cluster Greece SSDs among all entries (**Fig. 44**, green circle). The second coordinate, accounting for 26.3% of the variance, cluster the SSD carrying the Bx6+By8 allele. Tunisia was clustered alone while Italy (5), USA and Algeria (4) and Ethiopia (2) group together. (**Fig. 44**, purple and brown circle, respectively). In general, the PCA confirms the UPGMA analyses.

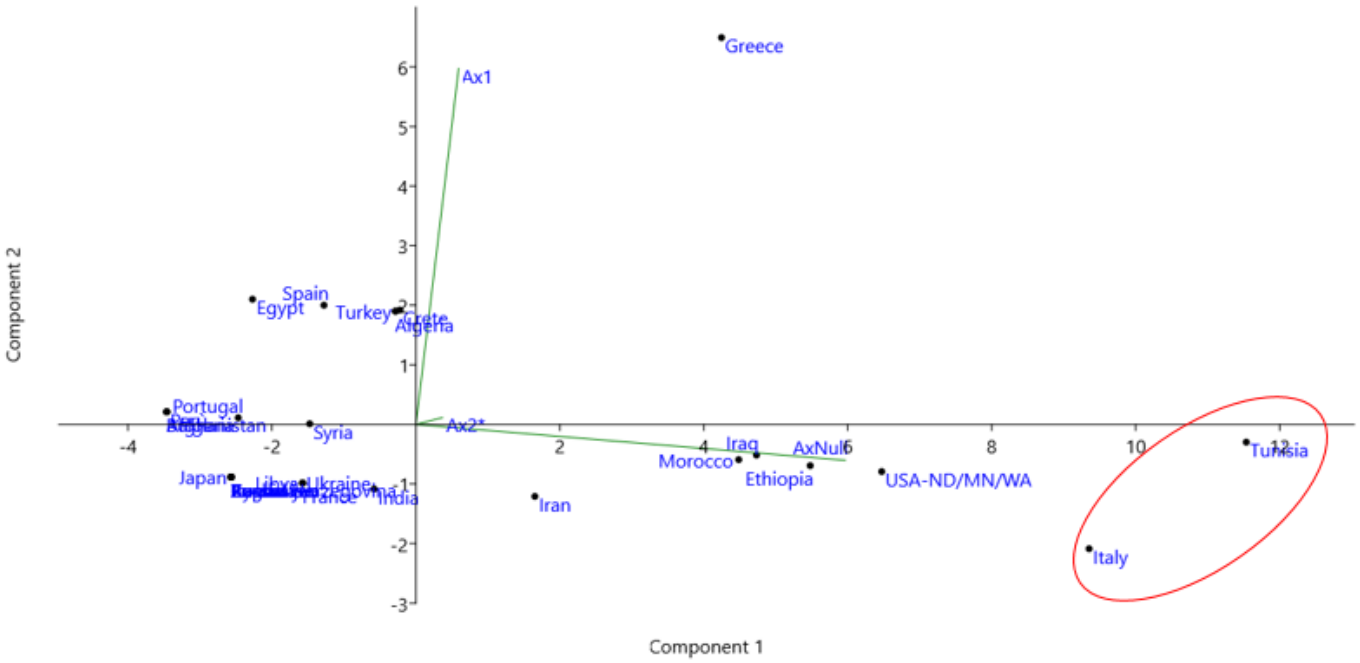

Figure 43: PCA Plot for the *Glu-A1* locus. Each point corresponds to a specific country.

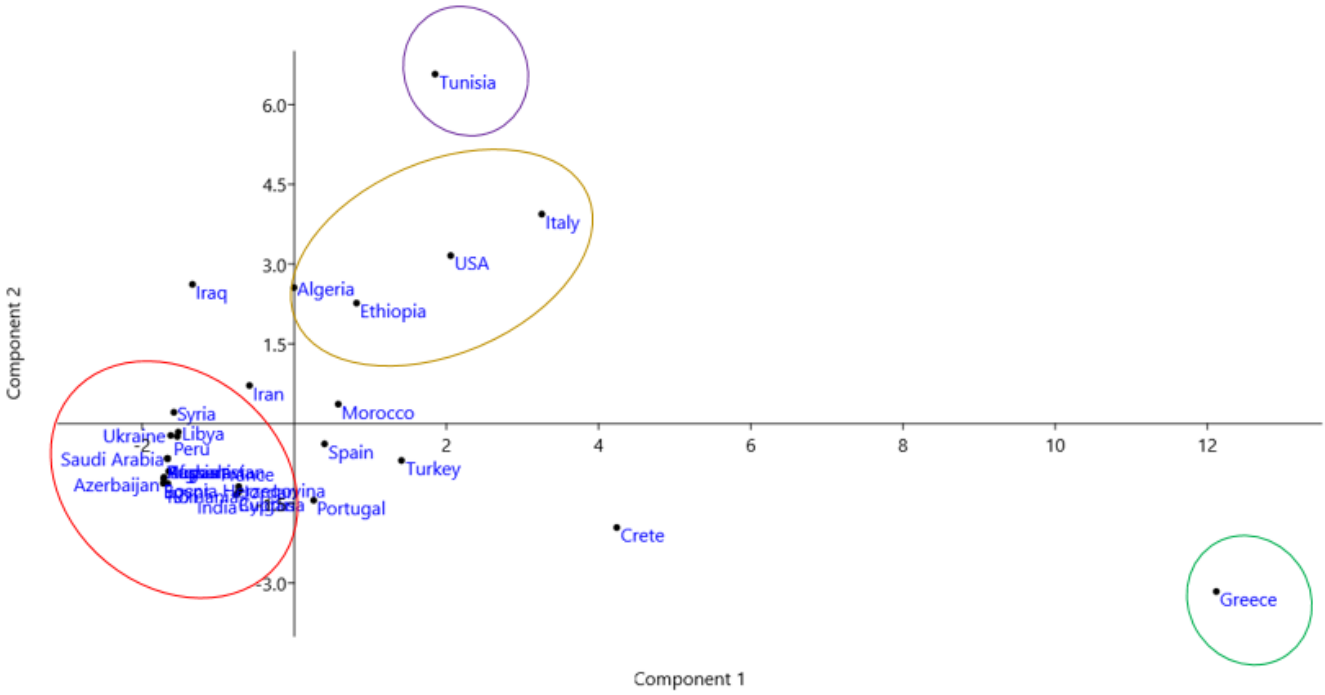

Figure 44: PCA Plot for the *Glu-B1* locus. Each point corresponds to a specific country.

4.2.4 Gene-ecology analyses

The characterization of the HMW-GS allelic combination in the germplasm material considered, which revealed a substantial level of diversity, demonstrated how such a highly heritable trait can serve as a means to trace the diffusion and diversification of a crop species.

The A genome locus *Glu-A1* featured three alleles, of which the null allele was by far the most common, followed by the allele encoding subunit Ax1; the third allele, responsible for subunit Ax2*, was represented in only ten lines. The same ranking with respect to allele frequency has been noted by both Branlard et al. (1989) in their characterization of 502 durum wheat coming from 23 countries and by Moragues et al. (2006) in a study of 63 durum landraces sourced from the Mediterranean Basin. The predominance of the *null* allele (*Glu-A1c*) has been retrieved in all the germplasm collections so far investigated, thus being confirmed in a number of other surveys (Bellil et al., 2014; Branlard et al., 2003; Ribeiro et al., 2011). The null allele was also the most frequent one when the entries were grouped according to provenance, and was the only allele recovered among the Italian entries (**Fig. 45a**). The Ax1 allele (*Glu-A1b*) was relatively frequent in N. African and S.W. Asian material; it was also found in the single Bulgarian genotype analyzed (**Fig. 45a**), an observation contrasting Moragues et al. (2006). A frequency imbalance could be explained where the flour is also destined for bread making, because the presence of the null allele was correlated with dough extensibility (Branlard et al., 2003). The presence of the Ax2* allele (*Glu-A1b*) has been associated with improved performance for some other dough quality parameters (SDS-sedimentation value and mixogram score), although this conclusion was reached on the basis of a rather small number of test entries (Raciti et al., 2003; Sissons, 2008). Anyway, these considerations should be considered when selecting other genotypes for further analyses.

More extensive variation was present at the *Glu-B1* locus, where 15 alleles were detected; the Moragues et al. (2006) study identified 14 *Glu-B1* alleles, while the Branlard et al. (1989) one found ten. The most frequent alleles at this locus were Bx20+By20, Bx7+By8 and Bx6+By8 (*Glu-B1e, Glu-B1b* and *Glu-B1d* respectively) a ranking consistent with that recorded for a set of 45 Algerian durum wheat landraces and old cultivars (Cherdouh et al., 2005); moreover, the predominance of the Bx20+By20 allele (*Glu-B1e*) has featured strongly in several other germplasm collections (Bellil et al., 2014; Moragues et al., 2006; Nazco et al., 2014). The frequency of some of the minor alleles like Bx13+By16 (*Glu-B1f*), Bx7+By8* (*Glu-B1al*) and GS Bx14+By19 *(*~5%) was comparable between the present germplasm set and that studied by Cherdouh et al. (2005). Both Bx6+By8 (*Glu-B1d*, present in 28 of the 152 lines) and Bx14+By15 (*Glu-B1h*, one line) alleles have been associated with the dough quality parameters SDS sedimentation value and resistance breakdown value (Brites and

Carrillo, 2001), while according to Branlard et al. (2003), Bx6+By8 (*Glu-B1d*) is also beneficial in terms of biscuit making quality. The high frequency of the Bx7+By8 allelic combination (*Glu-B1b*, 23/152 entries) may similarly derive from its association with strong gluten and good pasta quality (Nazco et al., 2014). According to Sissons et al. (2005), the ranking of *Glu-B1* alleles based on their contribution to pasta quality is Bx7+By8>Bx20>Bx6+By8 (*-B1b>-B1e>-B1d*), an ordering adjusted by Varzakas et al. (2014) to take into account less common alleles to Bx17+By18; Bx13+By19; Bx7+By8>Bx7≥Bx6+By8 (*–B1i>B1g>-B1b>-B1a>-B1d*). The locus was polymorphic in materials originating from the Fertile Crescent, as well as from N. Africa and Ethiopia (**Fig. 45b**). The general preponderance of the Bx20+By20 allele (*Glu-B1e*) has been noted by other researchers (Moragues et al., 2006; Nazco et al., 2014), although curiously it is somewhat less ubiquitous in Iberian germplasm (Moragues et al., 2006). The Bx13+By16 allele (*Glu-B1f*), seen in the African material, was not represented among the Fertile Crescent lines, while some other alleles such as GS Bx7+By19, 7 (*Glu-B1a*), GS Bx14+By20 and Bx14+By15 (*Glu-B1h*) were present in the latter, but not in the former set of germplasm (**Fig. 45b**). The N. African (but not the Ethiopian) lines included representatives of Bx7+By8* allele (*Glu-B1al*), while the allele encoding subunit Bx14 (*Glu-B1-1g*) showed the opposite pattern. GS Bx14+By19 was detected in the Fertile Crescent and Ethiopian material, but not in the N. African germplasm. The rare Bx7*+By8* allele (*Glu-B1ak*) was found only in material with a Romanian or a Greek provenance (**Fig. 45b**). A Greek presence in what is now Romania has been dated back as far as the 7th century BCE (Tsetskhladze, 2008). At the same time, the evidence is that one of the main routes by which agricultural know-how entered Europe during Neolithic times passed through the Balkans, with Greece representing one of the first European sites where agriculture was adopted (Ammerman and Cavalli-Sforza, 1984; Sonnante et al., 2009). This is consistent to what observed by Oliveira et al. (2011) in their phylogeography study of einkorn landraces, who deduced that different einkorn germplasm genotypes may have been introduced from Greece. Moreover, evidence from archaeology and genetic analysis of emmer landraces suggest Puglia, in South Italy, as the region where the first agricultural communities appeared in the Italian Peninsula (Isaac et al. 2010), possibly seafaring from Greece or the Balkans. In conclusion, interestingly, the gene ecology analyses performed in this thesis, made on the basis of phenotypic data, like proteomic data, perfectly reflects the history of wheat diffusion from its center of origin and tells us the history of wheat movements in the migration routes.

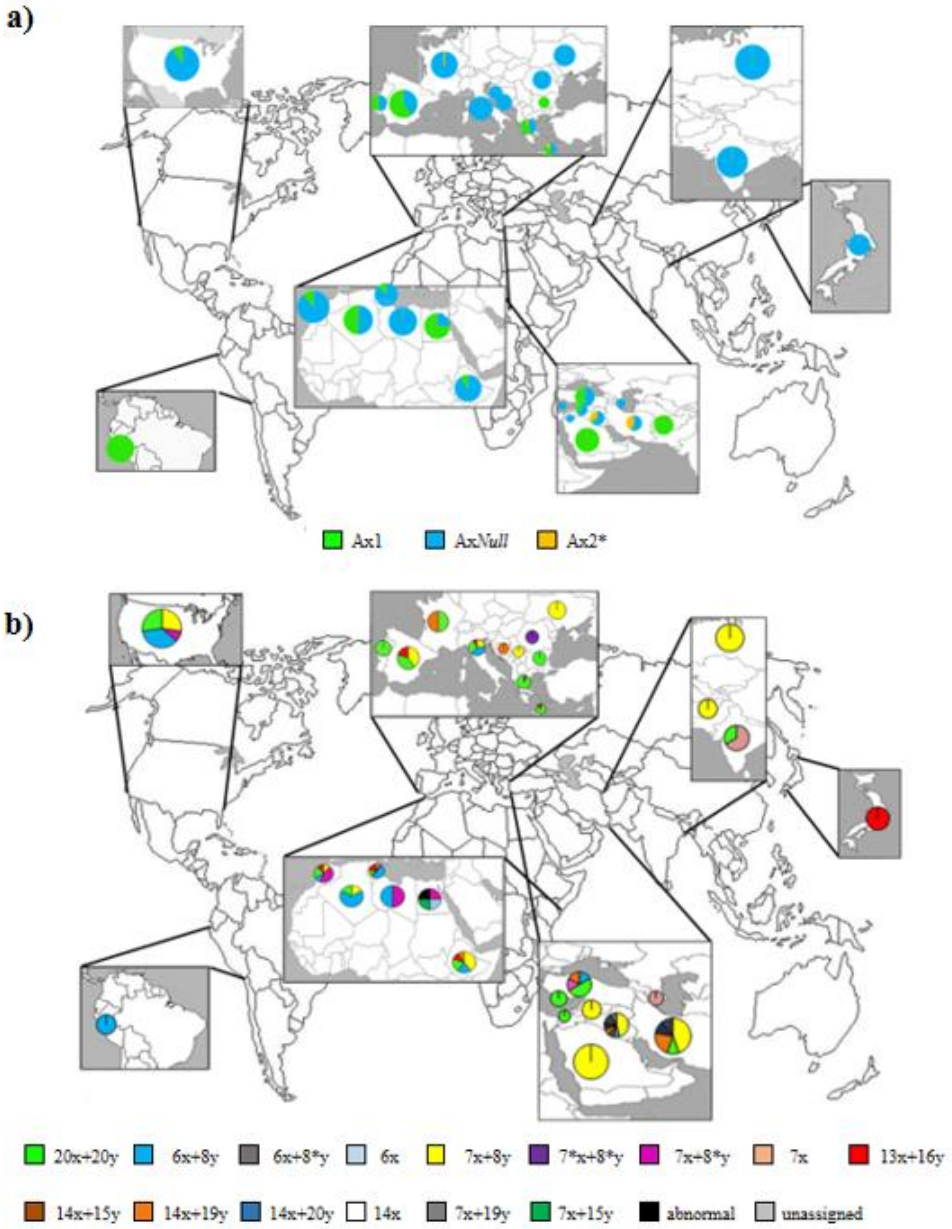

Figure 45: *Glu-1* status and the geographic provenance of the germplasm entries for **a)** *Glu-A1* and **b)** *Glu-B1*. The breakdown of the pie charts indicates the entries within each category.

4.2.5 Discussions

Despite modern agriculture increased food production, one person out of nine is still undernourished. Breeders achieved a linear increase at an average of 32 million metric tons per year, however the challenge of higher yields and good quality is still a main issue considering that the global population is projected to exceed 9 billion by 2050 (Wang et al., 2017). Moreover, due to the current climate changes, which have increased the frequency of unpredictable extreme events, the development and adoption of stress resilient crop cultivars is imperative to ensure sustainability of agricultural production and global food security (Wang et al., 2017). Radical changes in the agricultural management processes for the most common commodities is mandatory (Zivy et al., 2015). With intense focus and overuse of elite plant materials, 75% of the genetic diversity of domestic agricultural crops has disappeared and the narrow genetic base of elite germplasm is responsible of the increasing genetic vulnerability to new biotic and abiotic stresses, as well as to unpredictable environmental conditions (Wang et al., 2017). Germplasm materials represent an important source of genetic variability which can be explored and exploited for novel traits which could enable in better addressing the aforesaid goals.

Nowadays, a combination of multiple "Omics" approaches to characterize available germplasm collection is highly recommended since a characterization based on genomics, providing a list of potentially useful genes, together with proteomics and metabolomics can lead to the understanding of the molecular network interactions operating in a given genotype (Zivy et al., 2015). The rapid developing of high-throughput genotyping, biotechnology and phenotyping methods, together with advances in statistical methods and modelling, provide opportunities to address various challenges for utilization of germplasm lines, allowing the identification and exploitation of major genes, minor alleles, or haplotypes affecting complex traits (Wang et al., 2017). In addition, the application of a multidisciplinary approach that link the phenotype (the proteome) to the genotype (the alleles) enables to draw some conclusions on the history and migration routes that affected wheat domestication and adaptation.

In the present project, we characterize a set of 152 durum wheat landraces at both the protein and DNA level for their composition in HMW-GS; moreover, a wide selection of elite varieties has been included in the analyses. The results obtained confirmed that the combination of SDS-PAGE and molecular markers is still necessary for the characterization of the composition in HMW-GS of wheat genotypes to avoid allele misidentification and that the use of PCR markers allows at least the identification of new haplotypes.

The observed diversity patterns at the two *Glu-1* loci are largely consistent with the idea that durum wheat diversified in three distinct geographical locations, namely the Fertile crescent, N. Africa and the highlands of Ethiopia (Kidane et al., 2017; Pecetti et al., 1992; Porceddu et al., 1973). The present data corroborate this latter idea, since in both cluster analysis (**Fig. 42 a** and **b**) and PCA (**Fig. 43** and **44**) Iraq, Iran and Ethipia on the one hand and Tunisia, Algeria (and Italy) on the other hand, are always grouped together probably on the basis of the high number of genotypes sharing the same allele, but, at the same time, the same countries are also those with the greatest genetic variability. In addition, they support the proposed history of the spread of wheat cultivation across the Mediterranean Basin (Nazco et al., 2014). The materials originating from the northern and southern shores of the Mediterranean shared a greater degree of genetic similarity than they did with materials of S.W. Asian provenance; as a matter of fact, in the hierarchical cluster analysis at both *Glu-A1* and *Glu-B1* loci (**Fig. 42a** and **42b**, respectively), Italy always clustered with Tunisia and Algeria. Also the PCA analysis confirmed this evidence, grouping together the aforesaid countries in the same area of the plots (**Fig. 43**, red circle; **Fig 44**, purple and brown circle). Thus, the presented data imply, as suggested by Moragues et al. (2006), that wheat was likely brought to southern Italy from N. Africa. The rather rare allele 7 (*Glu-B1a*) was restricted, as similarly noted by Moragues et al. (2006), to India and S.W. Asia, which suggests an independent expansion of wheat cultivation eastwards from the Fertile Crescent. Trading relationships between N. Africa and Europe were undoubtedly encouraged by the geopolitical stability associated with the expansion of the Roman empire. By the beginning of the first millennium, wheat trade from North Africa to Rome started during Roman Republican period (509 - 27 BCE). At times of the Empire (27 BCE – 476 CE), after the transformation of Egypt in a Roman province with full status and the reconstruction of Cartago, North Africa became the largest wheat supplier of Rome (Garnsey, 1983). According to Scarascia Mugnozza (2005), one consequence of the occupation of Ethiopia by Italy during the first half of the 20th century was the import of Italian durum wheat germplasm, but the marked differentiation between Italian and Ethiopian landraces exposed by genetic diversity analyses implies that the two gene pools share very little common ancestry (Kabbaj et al., 2017). It is interesting to notice that the presence of the allele Bx14+By19 in both the Horn of Africa and the Fertile Crescent might support the Biblical and Quranic tale of the Queen of Sheba. The tradition and some interpretations of the Bible (I Kings 10:13) imply the queen being in love with King Solomon and returning to her country bringing up her child by Solomon. Bible stories also tell of the ships of Ophir, about the Israelites traveling to the Queen of Sheba's land (Hirschberg and Cohen, 2007). The legend might refer to actual exchange of goods, including seeds for sowing, between Israel and the Horn of Africa prior to the destruction of the Temple of Solomon (586 BCE).

The object study demonstrates how the characterization within a germplasm collection of a specific quality trait gives information not only on alleles diversity but also reflects the domestication routes and allele migration that occurs in the last 10,000 years. This is not surprising since the gluten protein fraction severely affect the dough properties, which is the main characteristics that advantage wheat over other temperate crops (Shewry, 2009). Wheat and other cereals became staple food since the early Neolithic era, being consumed mainly as a porridge together with other cereals or legumes (Sonnante et al., 2009). Remains at Ohalo II, dating between 19,000 and 11,000 YBP, show that grinding was a common practice and some flat stone indicate that probably bread cooking was already a common technology (Gremillion, 2011). That bread was not leavened, looking similar to many sorts of flatbreads still produced nowadays worldwide, but cooking made more nutrients available. Leavened bread needs the coincidence of three factors: high gluten flour, fermenting microbia and an oven: this coincidence occurred in ancient Egypt around 4000 BCE. Ovens and fermented foods were already known in earlier times, the key to bread was wheat with the proper gluten quality (Gremillion, 2011). Leavened bread diffused in the Mediterranean, and during Roman times bread consumption together with wheat cultivation increased significantly (Zhou et al., 2014).

It is interesting to notice that the centers of diversity and diffusion so far described strongly correlate to the use of wheat. In the Mediterranean cous cous, pasta and leavened bread making are based on specific gluten properties. The middle-eastern/Horn of Africa areas, instead, are characterized by the use of flatbreads, often cooked in clay ovens or on flat stones, as in the Hebrew tradition of matzah (מַצָּה) or the pocket pita bread.

In conclusion, the presented data demonstrate that phenotypic traits, at least the ones with high heritability and with low levels of interaction with external stimuli, may be as useful as genetic ones to trace crop evolutionary history, as intended as the history of its diffusion and diversification.

4.3 *In vitro* digestion SSD genotypes

The analyses of the presence of ten peptides associated with celiac disease (CD) in the SSD panel was performed, preliminary, on a limited number of SSD genotypes (40) selected, so far, on the basis of their relative HMW-GS allelic combination. Three samples with the same HMW pattern have been considered, if available (**Table 19**).

Table 19: List of SSD genotypes selected for the identification of gluten peptides associated with celiac disease. The corresponding HMW-GS allelic combination is shown.

SSD genotypes	HMW-GS allelic combination
92	Bx7+By8
109	Bx7+By8
112	Bx7+By8
195	Bx7+By8
269	Bx7+By8
325	Bx7+By8
59	Bx7+By8*
64	Bx7+By8*
66	Bx7+By8*
315	Bx7+By8*
393	Bx7*+By8*
494	Bx7*+By8*
96	Bx7
155	Bx7
180	Bx7
2	Bx6+By8
44	Bx6+By8
69	Bx6+By8
116	Bx6+By8
511	Bx6+By8
290	Bx6+By8*
52	Bx6
17	Bx13+By16
70	Bx13+By16
292	Bx13+By16
294	Bx13+By16
302	Bx14+By19
326	Bx14+By19
336	Bx14+By19
348	Bx14+By19
322	Bx14
256	Bx14+By20
451	Bx7+By19
453	Bx7+By19
173	Bx7+By15
244	Bx20+By20
253	Bx20+By20
278	Bx20+By20
415	Bx20+By20
416	Bx20+By20

The Minekus et al. (2014) method was adopted to imitate durum wheat digestion: the identified gluten peptides related to CD are presented in **Table 20**.

Table 20: List of the identified gluten peptides from wheat related to CD[a].

Code	Identified peptides related to CD	Protein		
	total immunogenic peptides (IP)		**Rt (min)**	**m/Z**
IP1	TQQPQQPFPQ	γ-gliadin	18.33	1198.8/600.3
IP2	SQQPQQPFPQPQ	γ-gliadin	17.03	1409.9/705.5
IP3	QAFPQQPQQPFPQ	γ-gliadin	22.44	1540.9/771.4
IP4	TQQPQQPFPQQPQQPFPQ	γ-gliadin	23.19	1075.6/717.6
IP5	PQTQQPQQPFPQFQQPQQPFPQPQQP	γ-gliadin	25.43	1034.4/1551.3/776
IP6	FPQQPQLPFPQQPQQPFPQPQQPQ	γ-gliadin	18.47	1429.3/953
IP7	QQPQQPFPQPQQTFPQQPQLPFPQQPQQPFP	γ-gliadin	29.47	1228.8/921.8
	total toxic peptides (TP)			
TP1	LQPQNPSQQQPQ	α-gliadin	14.33	1392.9/697
TP2	RPQQPYPQPQPQ	α-gliadin	15.04	732.5
TP3	LQPQNPSQQQPQEQVPL	α-gliadin	22.06	980.6

[a]: IP, immunogenic peptide; TP, toxic peptide.

After the *in vitro* digestion, the panel of peptides which survived was characterized by liquid chromatography- mass spectrometry (UPLC/ESI-MS) (**Fig. 46**) and then statistically analyzed.
The identified peptides are grouped into two categories, namely immunogenic (triggering the adaptive immunity) or toxic (triggering the innate immunity) peptides (Boukid et al., 2017b), and derived mainly from gliadins (γ-gliadins and α-gliadins). These peptides have been reported to stimulate CD4+ T lymphocytes selectively isolated from small intestinal mucosa of CD patients; recently also glutenins have been included in these groups (Mamone et al., 2015). As a matter of fact, the sequences of these epitopes are rich in glutamine and proline, which make them more resistant to gastrointestinal digestion and, consequently, induce the immune response in the case of CD subject (Lauret and Rodrigo, 2013). All the identified immunogenic peptides were exclusively γ-gliadins; moreover, all these immunogenic sequences contained a sequence motif QQPQQPFPQ, which has been identified related to the response DQ2.5-glia-γ4c (Janssen et al., 2015) abut toxic peptides identified in the digestates were mostly α-gliadins. Within these peptides, the sequence QQQP was observed in TP1 and TP3, while the sequence PSQQ was only found in TP1 and TP3 (**Table 20**). The latter two sequences were exclusively identified in the toxic peptides and absent in the non-toxic ones (Cornell and Stelmasiak, 2016).

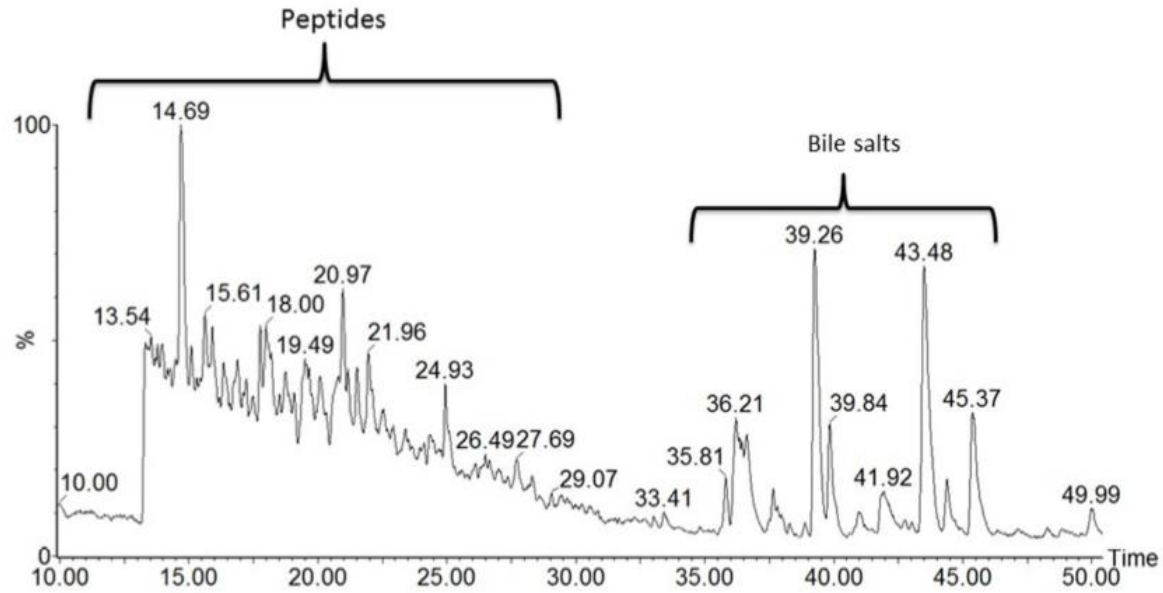

Figure 46: UPLC/ESI-MS chromatogram of a wheat sample.

Gluten fulfills two opposite key roles for humans: it is at the base of wheat technological properties but, unfortunately, it also has relevant immunological implications. During gluten digestion in the stomach, some resistant peptides are formed able to trigger celiac disease in genetically disposed patients. The high germplasm biodiversity could be an efficient strategy for the identification of genotypes carrying after digestion, a lower number of celiac disease epitopes (Prandi et al., 2017).

The analysis of 40 SSD genotypes did not show a high specificity in the peptide content within the SSD. All the SSD showed the presence of all the peptides analysed with the exception of four genotypes (SSD 52, 326, 278 and 415) in which one out of the ten peptides was absent (IP7 TP 10 IP3 and IP5, **Table 21**). In general gluten epitopes that derived from γ-gliadin were more abundant than those from α-gliadin, in agreement with the results of Prandi et al. (2014) and Boukid et al. (2017a) perfomed at different steps of the pasta production chain for three elite varieties (Svevo, Saragolla and Meridiano) and Tunisian durum wheats, respectively. Generally, IP4 and IP5 resulted the more abundant peptides in all the genotypes tested, while TP10 resulted as the less present in the majority of the cases.

Table 21: Summary of the values (ppm) of each immunogenic peptide (IP from 1 to 7) and toxic peptide (TP from 8 to 10) retrieved in the SSD genotypes analyzed. The total immunogenic amount (TI), total toxic amount (TP) and total immunogenic and toxic amount (TIT) for each sample is also reported. SSD genotypes are grouped on the basis of the relative HMW-GS allelic combination possessed.

HMW-GS	Genotype	IP1	IP2	IP3	IP4	IP5	IP6	IP7	TP1	TP2	TP3	TI	TT	TIT
7+8	**ssd 92**	131	156	124	723	577	173	345	279	179	79	2229	537	2766
	ssd 109	179	170	87	480	285	292	149	304	231	28	1642	563	2205
	ssd 112	178	167	120	637	420	381	269	375	489	63	2172	927	3099
	ssd 195	145	171	138	785	629	215	353	333	346	78	2436	757	3193
	ssd 269	148	158	104	587	427	212	277	316	249	43	1913	608	2521
	ssd 325	127	110	90	421	296	140	131	253	171	26	1315	450	1765
7+8*	**ssd 59**	121	56	84	607	463	68	153	169	108	33	1552	310	1862
	ssd 64	147	170	64	432	274	280	190	247	303	21	1557	571	2128
	ssd 66	157	183	116	777	633	389	583	333	650	204	2838	1187	4025
	ssd315	169	194	78	434	290	269	177	239	162	17	1611	418	2029
7*+8*	**ssd 393**	141	185	56	474	320	182	161	263	310	37	1519	610	2129
	ssd 494	118	67	28	613	322	125	136	100	297	90	1409	487	1896
7	**ssd 96**	77	72	53	482	296	88	102	98	45	13	1170	156	1326
	ssd 155	140	119	65	386	255	191	97	225	147	14	1253	386	1639
	ssd 180	150	151	99	622	426	167	236	203	143	25	1851	371	2222
6+8	**ssd 2**	138	158	111	664	437	211	295	315	280	51	2014	646	2660
	ssd 44	134	158	105	654	468	188	324	297	190	68	2031	555	2586
	ssd 69	157	135	91	629	501	304	331	330	472	111	2148	913	3061
	ssd 116	104	139	72	646	598	82	533	55	268	95	2174	418	2592
	ssd 511	191	233	116	776	570	224	342	276	206	53	2452	535	2987
6+8*	**ssd 290**	184	208	142	747	596	395	427	357	405	162	2699	924	3623
6	**ssd 52**	184	159	103	686	462	168	1	237	201	30	1763	468	2231
13+16	**ssd 17**	133	185	142	797	608	196	500	323	334	85	2561	742	3303
	ssd 70	223	251	109	618	400	356	289	228	152	24	2246	404	2650
	ssd 292	144	159	114	879	771	269	734	115	556	294	3070	965	4035
	ssd 294	172	178	116	728	502	325	324	341	538	87	2345	966	3311
14+19	**ssd 302**	118	124	56	461	341	199	248	197	306	58	1547	561	2108
	ssd 326	216	173	118	727	487	180	0	296	156	25	1901	477	2378
	ssd 336	126	126	72	503	300	107	106	182	120	14	1340	316	1656
	ssd 348	121	135	66	506	300	85	116	253	214	17	1329	484	1813
14	**ssd 322**	121	82	66	520	271	73	110	147	82	10	1243	239	1482
14+20	**ssd 256**	173	188	124	821	602	189	374	403	251	71	2471	725	3196
7+19	**ssd 451**	155	158	109	706	448	129	226	263	113	46	1931	422	2353
	ssd 453	124	98	54	565	474	132	275	126	242	91	2181	459	2640
7+15	**ssd 173**	163	86	82	752	413	77	97	185	85	13	1670	283	1953
20+20	**ssd 244**	194	192	147	1061	974	385	855	251	880	390	3808	1521	5329
	ssd 253	141	215	85	640	471	254	259	294	281	78	2065	653	2718
	ssd 278	147	156	52	552	358	88	219	140	50	0	1572	190	1762
	ssd 415	130	310	19	99	4	159	137	362	352	58	858	772	1630
	ssd 416	151	150	64	666	511	343	305	202	397	129	2918	728	3646

Only in six SSD (66, 69, 244, 290, 292 and 416, **Table 21**) the toxic peptide 10 showed a value higher than 100 ppm, while only in SSD 415 the IP4 and IP5 revealed low ppm values (99 and 4, respectively, **Table 21**). The total amount of the immunogenic peptides (TI) varies within the rangeof 858 ppm (SSD 415) and 3808 ppm (SSD 244) while the total of the toxic peptides (TP) varies between 156 (SSD 96) and 1521 ppm (SSD 244). The SSD with the lower total immunogenic and toxic (TIT) value was the SSD 96 (1326 ppm), while the one showing the greater value was the SSD 244 (5329 ppm) (**Table 21**).

A schematic representation of the results obtained from the *in vitro* digestion analysis is reported in **Fig. 47**.

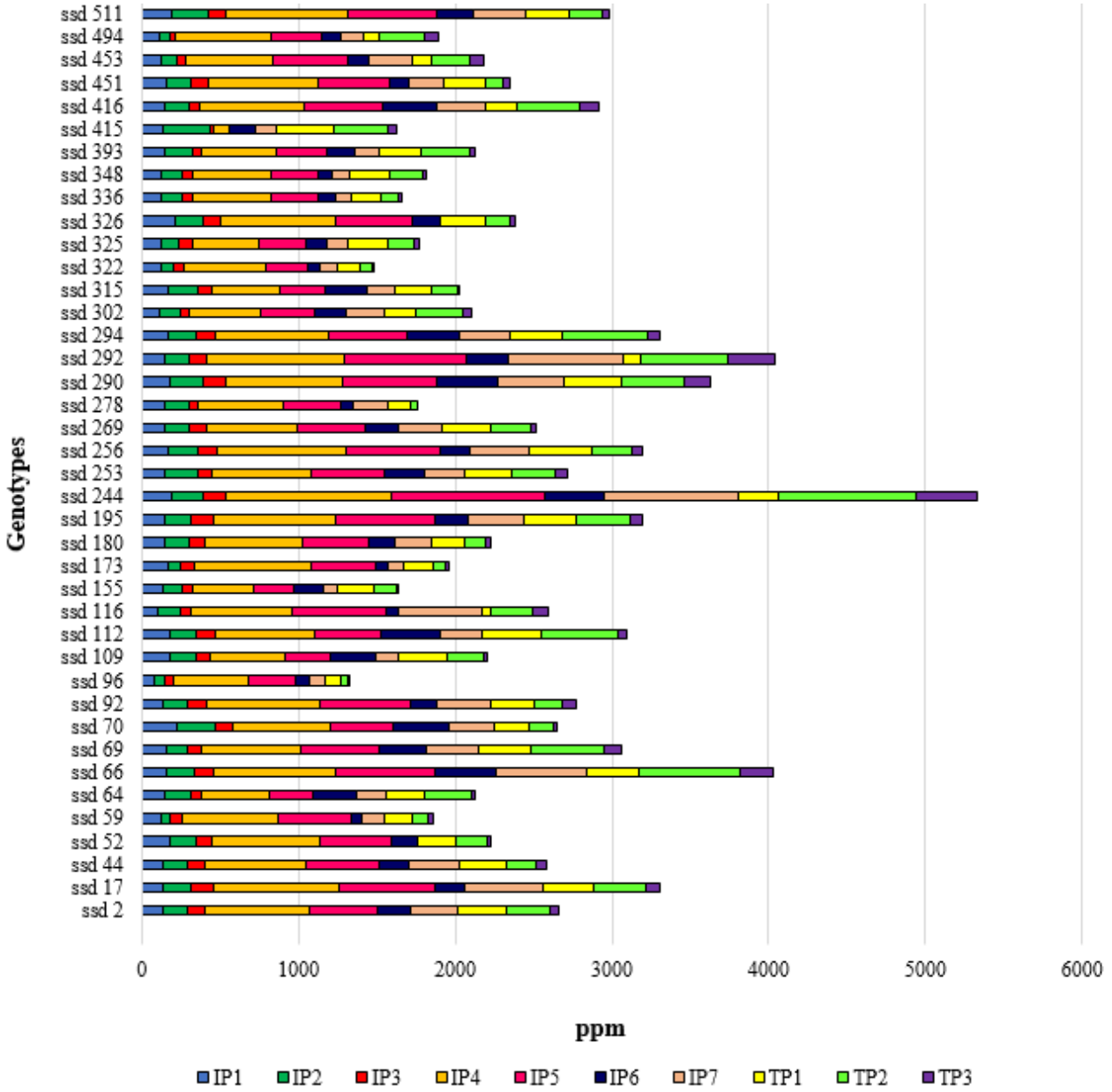

Figure 47: Total immunogenic and toxic (TIT) peptides retrieved in the 40 SSD genotypes digested.

The influence of genotype on CD-related peptides content was investigated statistically by performing ANOVA in which each genotype was compared to all the others among the 40 tested. The results did not detect a particular genotype as significant different from all the others; however, five SSDs revealed a potential immunotoxic significantly different ($P \leq 0.05$) from, at least, ten of the other genotypes considered in the analysis. SSD 96 differed from the 40% of the samples considered, being, as already affirmed, the one with the lowest ppm value of TIT, followed by SSD 244 which differed from the 32.5% of the other genotypes due the higher level of toxicity. Moreover, SSDs 66, 290 and 322 were evidenced in the ANOVA as significantly different from the 27.5%, 25% and 22.5%, respectively. SSDs 66 and 290 revealed high value of the immunogenic peptides as almost all the genotypes tested but, additionally, also the value of the three toxic peptides resulted very high.

A PCA was performed to enable an overview of immunogenic and toxic peptides associated with celiac disease and genotypes. The first two components of the PCA (**Fig. 48**) explained about 97.64% of the total variation with PC1 accounting for 93.498% and PC2 accounting for 4.15%. The first component (x axis) was related to the immunogenic peptide 7 and total immunogenic and toxic amount; accordingly, SSD 244 (**Fig. 48**, red circle), followed by SSDs 292 and 66 (**Fig. 48**, purple circle) were gathered on the extreme right side of the plot. The second component (y axis) was principally related to TP2, total toxic amount, IP4, IP5 and total immunogenic amount. Notably, SSDs 96 and 322, mentioned above for the results they produced in the ANOVA test, are located in the lower left corner of the plot (**Fig. 48**, brown circle), highlighting their low amount of CD-associated peptides content with respect to the other 38 SSD analyzed. Moreover, SSD 415 is the only one located in the upper left corner of the plot (**Fig. 48**, green circle) probably because it is the one with the lowest total immunogenic amount and, more precisely, it is the only one who revealed a completely different tendency concerning IP4 and IP5; as a matter of fact, in contrast to the general trend, in the SSD 415 these two peptides showed irrelevant values.

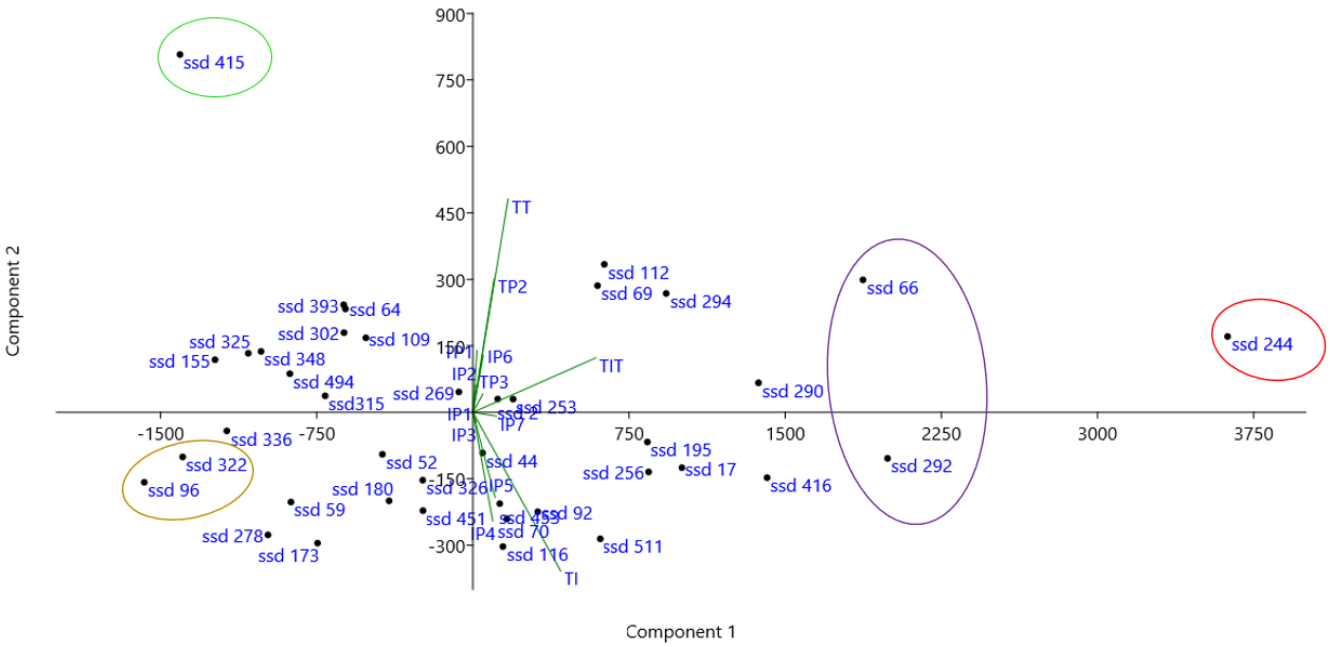

Figure 48: Biplot of principal components analysis of durum wheat epitopes associated with celiac disease.

PCA analyses were performed considering the CD-related peptides content and the provenance of the 40 genotypes, but it did not reveal a particular correlation between the two variables. Same results were obtained when combining in the PCA the CD-related peptides content and the HMW-GS allelic combination retrieved in the samples tested.

4.3.1 Discussions

Gluten represents an essential good for humans due to its capability to determine the viscoelastic properties of dough in bread making and in the production of wheat derived products. However, it has also important immunological implications, especially for those individuals genetically predisposed (Goesaert et al., 2005). The varietal selection undertaken by breeders over time with the aim to achieve the desired rehological properties has led to a decrease in the genetic biodiversity of wheat varieties present nowadays on the market (Prandi et al., 2017). Thus, the greater biodiversity of germplasm material could offer a new starting point to identify genotypes characterized by high technological properties, as the presence of HMW-GS allelic combnation known to exert good effects on gluten viscoelasticity, and encoding a lower number of coeliac disease epitopes. This idea is supported by the results presented by van den Broeck et al. (2010): in their study of 36 modern varieties and 50 landraces, it was demonstrated that, in general, the toxicity of modern wheat has increased. In another study conducted by Shalk et al. (2017), the amount of the 33-mer peptide from α-gliadin, known to be the most immune dominant gluten peptide, was determine in 57 samples of different heat species from around the world, including hexaploid common wheat and spelt, tetraploid durum wheat and emmer and diploid einkorn in order to make a precise assessment of the importance of this peptide associated with CD. The corresponding results pointed out the complete absence of the 33-mer peptide in durum wheat and emmer cultivars (genome AABB) as well as in diploid einkorn cultivars (genomeAA), which both do not contain the D genome. Hence, authors concluded that this absence could be explained by the fact that this peptide is encoded by genes located in the *Gli-2* locus on chromosome 6D, which is missing in the genotypes aforementioned. Nowadays, not only wheat consumption as wheat flour and wheat-based products is increasing worldwide, but also wheat gluten is increasingly applied as an additive in a wide and growing variety of processed foods and in other products, including medicines (Day et al. 2006; Maltin et al., 2009). Hence, even if landraces will require further breeding to increase their agronomical and food-technological value, considering a worldwide occurrence of 1% of CD patients and a high frequency (85–90%) of undiagnosed individuals, some agronomic drawbacks may be acceptable (van den Broeck et al., 2010).

In terms of peptide production during gastrointestinal digestion, the outcome of the *in vitro* digestions for the 40 SSDs analyzed was quite similar. SSD 96 showed a $P \leq 0.05$ with respect to 16 out of the 40 SSD, thus differentiating from 40% of the samples. No correlation between the CD-related peptides content and the geographic provenance af the genotypes has been detected by performing PCA and, similarly, the investigation of a possible link between the immunotoxic peptides

content and the HMW-GS allelic combinations of the SSD considered did not produce relevant considerations. These results need further analyses widening the sample number, since only few genotypes for each country were preliminary analyzed.

Chapter 5. Conclusions and future perspectives

This project deals with the characterization of a wide panel of durum wheat landaraces and its characterization for HMW-GS composition at the *Glu-A1* and *Glu-B1* loci.

The strategy applied in the project relies on the application of a combined approach to obtain a deep map of the HMW-GS composition.

The results obtained in the project, indeed, are of high and different value: i) a methodological value, since the results obtained on elite varieties, as well as on a panel of 152 landraces, highlight the need of performing a combined Omic approach to determine, as precisely as possible, the composition of HMW-GS in the durum wheat kernel; ii) a breeding value, since the identification of landraces with increased quality characteristics and with a history of climate adaptation could be applied in future breeding programmes; iii) a healthy value, since the high variability encountered within the SSD collection in terms of immune and toxic peptides could lead to the health improvement. Moreover, the deep characterization of a large set of durum and bread wheat varieties for HMW-GS composition represent a practical guide for those researchers and breeders interested in HMW composition of modern and old wheat varieties.

The application of a combined "Omic" approach was highly efficient in revealing both known HMW-GS patterns, but also rare alleles and new combinations within them that need further analyses to assess their contribution in technological properties.

Moreover, increasing the number of cultivar tested, a new specificity for the ZSBy9_aF1/aR3 molecular marker was discovered, allowing the clear identification of *By20*. The combined analyses performed on the germplasm material, revealed, as expected, a high level of genetic variability, since a total of eight uncommon HMW-GS allelic combinations, in addition to the more diffuse ones, were observed within the genotypes. These results corroborate the usefulness of germplasm collections, as reservoir variability to be explored and exploited for wheat breeding. Moreover, this work offers new insights in the diffusion and diversification of a crop species. The validity of the presented work lay on the consideration that a phenotypic trait, as the composition in HMW-GS in wheat endosperm, reflects the history of the wheat domestication and migrations routes. The link between the HMW-GS composition retrieved and the geographic origin of the 152 SSD genotypes gave information about the domestication routes and allele migration that occurs in the last 10,000 years, corroborating between all the results presented, the hypothesis that wheat was likely brought to southern Italy from North Africa.

To test whether the exploitation of wheat landraces can have positive health effects, an *in vitro* digestion that produce peptides containing immunogenic and toxic sequences was preliminary

performed on 40 SSDs genotypes. These analyses did not show significant differences in the content of CD-related peptides within the panel selected, but at least allowed the identification of one genptype, SSD 96, that differed from the 40% of the sample. In addition, SSD 244 differs from 32.5% of the genotypes.

In the light of these consideration, and since its HMW-GS pattern could be considered good from a technological point of view, it could be interesting to further analyze SSD 96 for its rheological parameters and to extend the analyses on the entire panel of landraces available.

References

Ahmad, M. (2000). Molecular marker-assisted selection of HMW glutenin alleles related to wheat bread quality by PCR-generated DNA markers. *Theoretical and Applied Genetics, 101*:892–896.

Alsaleh, A., Shehzad, B., Derya, M., Azrak, M., Kilian, B., Özkan, H., & Nachit, M. (2014). Genetic Linkage Map of Anatolian Durum Wheat Derived from a Cross of Kunduru-1149 × Cham1. *Plant Molecular Biology Reporter*.

Altenbach, S. B., Vensel, W. H., & Dupont, F. M. (2011). The spectrum of low molecular weight alpha-amylase/protease inhibitor genes expressed in the US bread wheat cultivar butte 86. *BMC Research Notes, 4*:242.

Ameille J., Pauli, G., Calastreng-Crinquand, A., Vervloët, D., Iwatsubo, Y., Popin, E., … Kopferschmitt-Kubler, M. C., and the corresponding members of the ONAP (2003). Observatoire National des Asthmes Professionnels. Reported incidence of occupational asthma in France, 1996-99: the ONAP programme. *Occupational and Environmental Medicine, 60*: 136-41.

Ammerman, A. J., & Cavalli-Sforza, L.L. (1984). *The Neolithic Transition and the Genetics of Populations in Europe*. Princeton University Press, Princeton, New Jersey.

Anderson, O. D., Litts, J. C., & Greene, F. C. (1997). The α-gliadin gene family: 1. Characterization of ten new wheat α-gliadin genomic clones, evidence for limited sequence conservation of flanking DNA, and southern analysis of the gene family. *Theoretical and Applied Genetics, 95*:50–58.

Anjum, F. M., Khan, M. R., Din, A., Saeed, M., Pasha, I., & Arshad, M. U. (2007). Wheat gluten: High molecular weight glutenin subunits - Structure, genetics and relation to dough elasticity. *Journal of Food Science, 72*(3): R56-R63.

Baláza, G., Tömösközi, S., Harasztos, A., Németh, V., Tamás, Á., Morgounov, A., … Békés, F. (2012). Advantages and Limitations of Lab-on-a-chip Technique in the Analysis of Wheat Proteins. *Cereal Research Communications, 40*(4):562-572.

Ball, S.G., Van de Wal, M.H.B.J., & Visser, R.G.F. (1998). Progress in understanding the biosynthesis of amylose. *Trends in Plant Science*, 3:462-467.

Barcaccia, G., & Falcinelli, M., (2012). Genetica e genomica. Vol. 2: Miglioramento genetico. Liguori Ed. EAN: 9788820737429.

Barak, S., Mudgil, D., & Khatkar, B. S. (2015). Biochemical and functional properties of wheat gliadins: a review. *Critical Reviews in Food Science and Nutrition*, *55*:357-368.

Bassi, F. M., Bentley, A. R., Charmet, G., Ortiz. R., & Crossa, J. (2016). Breeding schemes for the implementation of genomic selection in wheat (*Triticum* spp.). *Plant Science*, *242*:23-36.

Bellil, I., Hamdi, O., & Khelifi, D. (2014). Diversity of five glutenin loci within durum wheat (Triticum turgidum L. spp. durum (Desf.) Husn.) germplasm grown in Algeria. *Plant Breeding*, *133*:179-183.

Bietz, J. A., & Wall, J. S. (1972). Wheat gluten subunits: Molecular weights determined by sodium sulfate-polyacrylamide gel electrophoresis. *Cereal Chemistry Journal*, *49*:416-430.

Bietz, J. A., Shepherd, K. W., & Wall, J. S. (1975). Single-kernel analysis of glutenin: Use in wheat genetics and breeding. *Cereal Chemistry Journal*, *52*:513-532.

Boukid, F., Prandi, B., Sforza, S., Sayar, R., Seo, Y. W., Mejri, M., & Yacoubi, I. (2017a). Understanding the effects of genotype, growing year and breeding on Tunisian durum wheat allergenicity. 2. The celiac disease case. *Journal of Agricultural and Food Chemistry*, *65*(28);5837-5846.

Boukid, F., Prandi, B., Buhler, S., & Sforza, S. (2017b). Effectiveness of germination on protein hydrolysis as a way to reduce adverse reactions to wheat. *Journal of Agricultural and Food Chemistry*, *65*:9854-9860.

Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of micrograms quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry*, *72*(1-2):248-254.

Brandolini, A., Volanmte, A., & Heun, M. (2016). Geographic differentiation of domesticated einkorn wheat and possible Neolithic migration routes. *Heredity*, *117*:135-141.

Branlard, G., & Dardevet, M. (1985). Diversity of grain protein and bread wheat quality: II. Correlation between molecular weight subunits of glutenin and flour quality characteristics. *Journal of Cereal Science*, *3*:345-354.

Branlard, G., Autran, J. C., & Monneveux, P. (1989). High molecular weight glutenin subunit in durum wheat (*T. durum*). *Theoretical and Applied Genetics*, *78*:353-358.

Branlard, G., Dardevet, M., Saccomano, R., Lagoutte, F., & Gourdon, J. (2001). Genetic diversity of wheat storage proteins and bread wheat quality. *Euphytica*, *119*:59–67.

Branlard, G., Dardevet, M., Amiour, N., & Igrejas, G. (2003). Allelic diversity of HMW and LMW glutenin subunits and omega-gliadins in French bread wheat (*Triticum aestivum* L.). *Genetic Resources and Crop Evolution*, *50*:669-679.

Brim, C. A. (1966). A modified pedigree method of selection in soybeans. *Crop Science*, *16*:220.

Brisman, J. (2002). Baker's Asthma. *Occupational and Environmental Medicine*, *59*:498-502.

Brites, C., & Carrillo, J. M. (2001). Influence of high molecular weight (HMW) and low molecular weight (LMW) glutenin subunits controlled by Glu-1 and Glu-3 loci on durum wheat quality. *Cereal Chemistry*, *78*(1):59-63.

Brouns, F. J. P. H., Van Buul, V. J., & Shewry, P. R. (2013). Does wheat make us fat and sick? *Journal of Cereal Science*, *58*:209-215.

Bushuk, W. & Zillman, R. R. (1978). Wheat cultivars identification by gliadin electrophoregrams. I. Apparatus method and nomenclature. *Canadian Journal of Plant Science*, *58*:505–515.

Butow, B. J., Ma, W., Gale, K. R., Cornish, G. B., Rampling, L., Larroque, O., … Bekes, F. (2003). Molecular discrimination of Bx7 alleles demonstrates that over expression has a major impact on wheat flour dough strength. *Theoretical and Applied Genetics*, *107*:1524–1532.

Butow, B.J., Gale, K.R., Ikea, J., Luhàsz, A., & Bedö, Z. (2004). Dissemination of the high expressed Bx7 glutenin subunit (*Glu-B1al* allele) in wheat as revealed by novel PCR markers and RP-HPLC. *Theoretical and Applied Genetics*, *109*:1525-1535.

Camarca, A., del Mastro, A., & Gianfrani, C. (2012). Repertoire of gluten peptides active in celiac disease patients: perspectives for translational therapeutic applications. *Endocrine, Metabolic, & Immune Disorders - Drug Targets*, *12*:207e219.

Carver, B. F. (2009). *Wheat Science and Trade*. Wiley-Blackwell.

Casañas, F., Simó, J., Casals, J., & Prohens, J. (2017). Toword and evolved concept of landrace. *Frontiers in Plant Science*, *8*:145.

Cassidy, B. G., Dvorak, J., & Anderson, O. D. (1998). The wheat low molecular-weight glutenin genes: Characterization of six new genes and progress in understanding gene family structure. *Theoretical and Applied Genetics*, *96*:743–750.

Catassi, C., & Cobellis, G. (2007). Coeliac disease epidemiology is alive and kicking, especially in the developing world. *Digestive and Liver Disease*, *39*:908-910.

Causgrove, P., Causgrove, L., & Peterson, B. (2004). Wheat and Flour Testing Methods – a guide to under standard wheat and flour quality. Wheat Marketing Center. Portland, OR.

Cherdouh, A., Khelifi, D., Carrillo, J. M., & Nieto-Taladriz, M. T. (2005). The high and low molecular weight glutenin subunit polymorphism of Algerian durum wheat landraces and old cultivars. *Plant Breeding*, *124*(4):338-342.

Chung, O.K., (1985). Lipid-protein interaction in wheat flour, dough, gluten and protein fractions. *Cereal Foods World*, *31*:242-256.

Cianferoni, A. (2016). Wheat allergy: diagnosis and management. *Journal of Asthma and Allergy*, *9*:13-25.

Ceoloni, C., Kuzmanović, L., Forte, P., Gennaro, A., & Bitti, A. (2014). Targeted exploitation of gene pools of alien Triticeae species for sustainable and multi-faceted improvement of the durum wheat crop. *Crop and Pasture Science*, *65*(1):96-111.

Ceoloni, C., Kuzmanović, L., Ruggeri, R., Rossini, F., Forte, P., Cuccurullo, A., & Bitti, A. (2017). Harnessing genetic diversity of wild gene pools to enhance wheat crop production and sustainability: Challenges and opportunities. *Diversity*, *9*(4):55.

Colonna, P., & Buléon, A. (1992). New insights on starch structure and properties. In *Cereal chemistry and technology: a long past and a bright future*. In *Proceedings of the 9th International Cereal and Bread Congress*, pp. 25-42, Paris, France, Institut de Recherche Technologique Agroalimentaire des Céréales (IRTAC).

Comastri, A., Gullì, M., Visioli, G., Campioli, D., Imperiale, D., & Marmiroli, N. (2013). Analisi proteomiche nei semi di cereali a paglia. Libro degli Atti del IX Convegno 9th AISTEC (Associazione Italiana di Scienza e Tecnologia dei Cereali). Congress, Book of Abstract pp.113-117. Bergamo, Italy.

Cornell, H. J., & Stelmasiak, T. (2016). The significance of key amino acid sequences in the digestibility and toxicity of gliadin peptides in celiac disease. *International Journal of Celiac Disease*, *4*:113−120.

Crossa, J., Burgueño, J., Dreisigacker, S., Vargas, M., Herrera-Foessel, S. A., Lillemo, M., … Ortiz, R. (2007). Association analysis of historical bread wheat germplasm using additive genetic covariance of relatives and population structure. *Genetics*, *177*:1889–1913.

Damidaux, R., Autran, J. C., Grignac, P., & Feillet, P. (1978). Evidence of relationships useful for breeding between the electrophoretic pattern of gliadins and the viscoelastic properties of the gluten in *Triticum durum* Desf. C.R. Hebd. Seances *Acad. Sci. Ser. D. Sci. Nat*, *287*:701–704.

Day, L., Augustin, M. A., Batey, I. L., & Wrigley, C. W. (2006). Wheat-gluten uses and industry needs. *Trends in Food Science & Technology*, *17*:82–90.

De Vita, P., Li Destri Nicosia, O., Nigro, F., Platani, C., Riefolo, C., Di Fonzo, N., & Cattivelli, L. (2007). Breeding progress in morpho-physiological, agronomical and qualitative traits of durum wheat cultivars released in Italy during the 20th century. *European Journal of Agronomy*, *26*:39-53.

Dewey, D.R. (1984). The genomic systems of classification as a guide to intergeneric hybridization with the perennial Triticeae. In: Gustafson, J. P. (Ed), *Gene manipulation in plant improvement*. (pp. 209-279). Plenum press, NY.

D'Ovidio, R., Masci, S., & Porceddu, E. (1995). Development of a set of oligonucleotide primers specific for genes at the Glu-1 complex loci of wheat. *Theoretical and Applied Genetics, 91*:189-194.

D'Ovidio, R, & Porceddu, E. (1996). PCR-based assay for detecting 1B-genes for low molecular weight glutenin subunits related to gluten quality properties in durum wheat. *Plant Breeding, 115*:413-415.

Du Toit, G. (2007). Food-dependent exercise-induced anaphylaxis in childhood. *Pediatric Allergy and Immunology, 18*(5):455–463.

Elbehri, A., and Food and Agriculture Organization of the United Nations (2015). Climate change and food systems: global assessments and implications for food security and trade.

Elli, L., Branchi, F., Tomba, C., Villalta, D., Norsa, L., Ferretti, F., … Bardella, M. T. (2015). Diagnosis of gluten related disorders: Celiac disease, wheat allergy and non-celiac gluten sensitivity. *World Journal of Gastroenterology, 21*(23):7110-7119.

Espì, A., Giraldo, P., Rodriguez-Quijano, M., & Carrilo, J. M. (2012). A PCR-based method for discriminating between high molecular weight glutenin subunits Bx7 and Bx7* in *Triticum aestivum* L. *Plant Breeding, 131*:571-573.

Espì, A., Rodriguez-Quijano, M., Vazquez, J. F., Carrillo, J. M., & Giraldo, P. (2014). Molecular characterization of Glu-B3 locus in wheat cultivars and segregating populations. *Journal of Cereal Science, 60*:374-381.

Evers, Y., & Millart, S. (2002). Cereal grain structure and development: some implication for quality. *Journal of Cereal Science, 36*:261-284.

Fasano, A., Berti, I., Gerarduzzi, T., Not, T., Colletti, R. B., Drago, S., … Horvath, K. (2003). Prevalence of celiac disease in at-risk and not-at-risk groups in the United States: a large multicenter study. *Archives of Internal Medicine, 163*:286–92.

Feldman, M. (2001). Origin of cultivated wheat. In: Bonjean, A. P., & Angus, W. J. (Eds), *The World Wheat Book. A History of Wheat Breeding*. (pp. 3–56). Paris, France: Lavoisier Publishing.

Ferrante, P. (2006). *Contributo a una migliore conoscenza del genoma e del proteoma di endosperma del frumento attraverso espressione eterologa e studi funzionali di singoli peptidi* (PhD thesis). Univerità degli Studi della Tuscia e di Viterbo.

Ferranti, P., Mamone, G., Picariello, G.& Addeo, F. (2007). Mass spectrometry analysis of gliadins in celiac disease. *Journal of Mass Spectrometry, 42*:1531–1548.

Fichera C., Sciacca F., Blanco C., Di Silvestro, S., & Palumbo, M. (2006). Molecular characterization of durum wheat cv. "Sant'Agata" by AFLP in fluorescence. *Journal of Genetic and Breeding, 60*: 00-00.

Flagella, Z. (2006). Qualità nutrizionale e tecnologica del frumento duro. *Italian Journal of Agronomy*, *1*:203-239.

Friedrich, C., Longin, H., & Reif, J. C. (2014). Redesigning the exploitation of wheat genetic resources. *Trends in Plant Science*, *19*(10):631-636.

Ghazy, A. I., Zanouny, A. I., Moustafa, K. A., & Al-Doss, A. A. (2012). Molecular screening of high molecular weight glutenin genes in spring bread wheat genotypes in Saudi Arabia. *Journal of Food, Agriculture and Environment*, *10*(1):157-161.

Galili, G., & Feldman, M. (1983). Genetic control of endosperm proteins in wheat. *Theoretical and Applied Genetics*, *66*:77-86.

Gan, Z., Ellis, P. R., & Schofield, J. D. (1995). Mini review: gas cell stabilization and gas retention in wheat bread dough. *Journal of Cereal Science*, *21*:215-230.

Garnsey, P. (1983). Grain for Rome. In Garnsey, P., Hopkins, K. & Whittaker, C. R. (Eds.), *Trade in the Ancient Economy*. (pp. 118-130). University of California Press.

Gianibelli, M. C., Larroque, O. R., Mac Ritchie, F., & Wrigley, W. (2001). Biochemical, genetic and molecular characterization of wheat glutenin and its component subunits. *Cereal Chemistry Journal*, *78*:635-646.

Gianibelli, M. C., Echaide, M., Larroque, O. R., Carrillo, J. M., & Dubcovsky, J. (2002). Biochemical and molecular characterization of Glu-1 loci in Argentinean wheat cultivars. *Euphytica*, *128*(1):61–73.

Gill, B. S., Appels, R., Botha-Oberholster, A. M., Buell, C. R., Bennetzen, J. L., Chalhoub, B., … Sasaki, T. (2004). A Workshop Report on Wheat Genome Sequencing: International Genome Research on Wheat Consortium. *Genetics*, *168*:1087–1096.

Giraldo, P., Rodriguez-Quijano, M., Simon, C., Vazquez, J. F., & Carrillo, J. M. (2010). Allelic variation in HMW glutenins in Spanish wheat landraces and their relationship with bread quality. *Spanish Journal of Agricultural Research*; *8*(4):1012-1023.

Giraldo, P., Royo, C., González, M., Carrillo, J. M., & Ruiz, M. (2016). Genetic Diversity and Association Mapping for Agromorphological and Grain Quality Traits of a Structured Collection of Durum Wheat Landraces Including subsp. *durum*, *turgidum* and *diccocon*. PlosOne, *11*(11):e0166577.

Goesaert, H., Brijs, C., Veraverbeke, W. S., Courtin, C. M., Gebruers, K., & Delcour, J. A. (2005). Wheat constituents: how they impact bread quality, and how to impact their functionality. *Trends Food Science & Technology*, *16*:12-30.

Goulden, C.H. (1939). Problems in plant selection. In *Proceedings of the 7th International Genetical Congress*, pp. 132–133, Edinburgh, Scotland.

Goutam, U., Kukreja, S., Tiwari, R., Chaudhury, A., Gupta, R. K., Dholakia, B. B., & Yadav, R. (2013). Biotechnological approaches for grain quality improvement in wheat: present status and future possibilities. *Australian Journal of Crop Science*, *7*(4):469-483.

Goutam, U., Tiwari, R., Gupta, R. K., Kukreja, S. & Chaudhury, A. (2015). Allelic variations of functional markers for high molecular weight glutenin genes in Indian wheat (*Triticum aestivum* L.) cultivars and their correlation with bread loaf volume. *Indian Journal of Plant Physiology*, *20*(1):97-102.

Gremillion, K. J. (2011). *Ancestral Appetites: Food in Prehistory*. Cambridge University Press, Cambridge, UK, Mar 2011, pp. 182.

Gupta, R. B., Singh, N. K., & Shepherd, K. W. (1989). The cumulative effect of allelic variation in LMW and HMW glutenin subunits on dough properties in the progeny of two bread wheats. *Theoretical and Applied Genetics*, *77*:57-64.

Gupta, R. B., & Shepherd, K. W. (1990). 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*:65-74.

Gupta, R. B., Bekes, F., & Wrigley, C. W. (1991). Prediction of physical dough properties from glutenin subunit composition in bread wheats - correlation studies. *Cereal Chemistry*, *68*:328-333.

Gupta, R. B., & MacRitchie, F. (1994). Allelic variation at glutenin subunit and gliadin loci, Glu-1, Glu-3 and Gli-1 of common wheats. II. Biochemical basis of the allelic effects on dough properties. *Journal of Cereal Science*, *19*:19-29.

Hadjivassiliou, M., Sanders, D. S., Woodroofe, N., Williamson, C., & Grunewald R. A. (2008). Gluten Ataxia. *The Cerebellum*, 494-498.

Hammer, O., Harper, D. A. T., & Ryan, P. D. (2001). PAST: Paleontological Statistic software package for education and data analysis. *Palaeontologia Electronica*, *4*(1):9.

Henkrar, F., El-Haddoury, J., Ouabbou, H., Bendaou, N., & Udupa, S. M. (2016). Genetic characterization of Moroccan and the exotic bread wheat cultivars using functional and random DNA markers linked to the agronomic traits for genomic-assisted improvement. *3 Biotech*, *6*:97.

Hirschberg, H. Z., & Cohen, H. J. (2007). ARABIA. In: Skolnik, F., Berenbaum, M. (Eds), *Encyclopaedia Judaica* (2nd ed.). (vol 2 pp. 295) Thomson Gale.

Hisa, C. C., & Anderson, O. D. (2001). Isolation and characterization of wheat ω-gliadin genes. *Theoretical and Applied Genetics*, *103*:37-44.

Holmes, G. (2013). Non coeliac gluten sensitivity. *Gastroenterology and Hepatology From Bed to Bench*, *6*(3):115-119.

Ikeda, T. M., Nagamine, T., Fukuoka, H., & Yano, H. (2002). Characterization of new low-molecular-weight glutenin subunit genes in wheat. *Theoretical and Applied Genetics*, *104*:680–687.

Ikeda, T., Araki, E., Fujita, Y., Yano, H., (2006). Characterization of low-molecular weight glutenin subunit genes and their protein products in common wheats. *Theoretical and Applied Genetics*, *112*, 327e334.

Inomata, N. (2009). Wheat allergy. *Current Opinion in Allergy and Clinical Immunology*, *9*:238-243.

Isaac, A. D., Muldoon, M., Brown, K. A., & Brown, T. A. (2010). Genetic analysis of wheat landraces enables the location of the first agricultural sites in Italy to be identified. *Journal of Archaeological Science*, *37*:950-956.

Janni, M., Cadonici, S., Pignone, D., & Marmiroli, N. (2017). Survey and new insights in the application of PCR-based molecular markers for the identification of HMW-GS at the *Glu-B1* locus in durum and bread wheat. *Plant Breeding*, *136*(4):467-473.

Janni, M., Cadonici, S., Bonas, U., Grasso, A., Dahab, A. A. D., Visioli, G., … & Marmiroli, N. Gene-ecology of durum wheat HMW glutenin reflects their diffusion from the centre of origin. (submitted).

Janssen, G., Christis, C., Kooy-Winkelaar, Y., Edens, L., Smith, D., van Veelen, P., & Koning, F. (2015). Ineffective degradation of immunogenic gluten epitopes by currently available digestive enzyme supplements. *PlosOne*, *10*(6): e0128065.

Jaradat, A. A. (2013). Wheat landraces: A mini review. *Emirates Journal of Food and Agriculture*, *25*(1):20-29.

Junker, Y., Zeissig, S., Kim, S. J., Barisani, D., Wieser, H., Leffler, D. A., … Freitag, T. L. (2012). Wheat amylase trypsin inhibitors drive intestinal inflammation via activation of toll-like receptor 4. *Journal of Experimental Medicine*, *209*:2395-2408.

Kabbaj, H., Sall, A. T., Al-Abdallat, A., Geleta, M., Amri, A., Filali-Maltouf, A., … Bassi, F. M. (2017). Genetic Diversity within a Global Panel of Durum Wheat (*Triticum durum*) Landraces and Modern Germplasm Reveals the History of Alleles Exchange. *Frontiers in Plant Science*, *8*:1277.

Keet, C. A., Matsui, E. C., Dhillon, G., Lenehan, P., Paterakis, M., & Wood, R. A. (2009). The natural History of wheat allergy. *Annals of Allergy, Asthma & Immunology*, *102*: 410-415.

Kellogg, E. A., (2001). Evolutionary history of the grasses. *Plant Physiology*, *125*:1198–1205.

Kidane, Y. G., Mancini, C., Mengistu, D. K., Frascaroli, E., Fadda, C., Pè, M. E., & Dell'Acqua, M. (2017). Genome Wide Association Study to Identify the Genetic Base of Smallholder Farmer Preferences of Durum Wheat Traits. *Frontiers in Plant Science*, *8*:1230.

Kissing Kucek, L., Veenstra, L. D., Amnuaycheewa, P., & Sorrells, M. E. (2015). A grounded guide to gluten: how modern genotypes and processing impact wheat sensitivity. *Comprehensive Reviews in Food Science and Food Safety*, *14*:1-18.

Köhler, P., Belitz, H. D., & Wieser, H. (1991). Disulphide bonds in wheat gluten: isolation of a cysteine peptide from glutenin. *Z. Lebensum Unters Forsch. Journal*, *192*: 234–239.

Köehler, P. & Wieser, H. (2013). Chemistry of Cereal Grains. In Gobbetti, M., & Gänzle, M. (Eds.), *Handbook on Sourdough Biotechnology*. Chapter 2. Springer Science+Business Media. New York.

Konvalina, P., Grausgruber, H., Tran Dang, K., Vlasek, O., Capouchova, I., Sterba, Z., … Kopecky, M. (2017). Rheological and Technological Quality of Minor Wheat Species and Common Wheat. Wheat Improvement, Management and Utilization, chapter 12:255-273.

Kreuz, A., Marinova, E., Schäfer, E., & Wiethold, J. (2005). A comparison of early Neolithic crop and weed assemblages from the Linearbandkeramik and the Bulgarian Neolithic cultures: differences and similarities. *Vegetation History and Archaeobotany*, *14*:237-258.

Kushwaha, U. K. S., Mangal, V., Bairwa, A. K., Adhikari, S., Ahmed, T., Bhat, P., … & Singh, N. K. (2017). Association mapping, principles and techniques. *Journal of Biological and Environmental Engineering*, *2*(1):1-9.

Lafiandra, D., Tucci, G. F., Pavoni, A., Turchetta, T., & Margiotta, B. (1997). PCR analysis of x-and y-type genes present at the complex Glu-A1 locus in durum and bread wheat. *Theoretical and Applied Genetics*, *94*:235–240.

Lauret, E., & Rodrigo, L. (2013). Celiac disease and autoimmune-associated conditions. BioMed Research International, *2013*,127589.

Lawrence, G. J., & Shepherd, K. W. (1980). Variation in glutenin protein subunits of wheat. *Australian Journal of Biological Sciences*, *33*:221-233.

Lei, Z. S., Gale, K. R., He, Z. H., Gianibelli, C., Larroque, O., Xia, X. C., … Ma, W. (2006). Y-type gene specific markers for enhanced discrimination of high-molecular weight glutenin alleles at the Glu-B1 locus in hexaploid wheat. *Journal of Cereal Science, 43:*94-101.

Leonard, M. M., Sapone, A., Catassi, C., & Fasano, A. (2017). Celiac Disease and Nonceliac Gluten Sensitivity, A Review. *Clinical Review & Education, 318*(7):647-656.

Létang, C., Piau, M., & Verdier, C. (1999). Characterization of wheat flour-water doughs. Part I: rheometry and microstructure. *Journal of Food Engineering,* 41:121-132.

Li, Q. Y., Yan, Y. M., Wang, A. L., An, X. L., Zhang, Y. Z., Hsam, S. L. K., & Zeller, F. J. (2006). Detection of HMW glutenin subunit variations among 205 cultivated emmer accessions (*Triticum turgidum* ssp. *dicoccum*). *Plant Breeding, 125*:120-124.

Li, W., Wan, Y., Liu, Z., Liu, K., Liu, X., Li, B., … Wang, D. (2004). Molecular characterization of HMW glutenin subunit allele 1Bx14: further insights into the evolution of Glu-B1-1 alleles in wheat and related species. *Theoretical and Applied Genetics, 109*:1093-1104.

Liang, X., Zhen, S., Han, C., Wang, C., Li, X., Ma, W., & Yan, Y. (2015). Molecular characterization and marker development for hexaploid wheat-specific HMW glutenin subunit 1By18 gene. *Molecular Breeding, 35*:221.

Lindsay, M. P., & Skerritt, J. H. (2000). Immunocytochemical localisation of gluten proteins uncovers structural organization of glutenin macropolymer. *Cereal Chemistry Journal, 77*:360-369.

Lineback, D. R., & Rasper, V. F., (1988). Wheat carbohydrates. In Pomeranz, Y. (Ed.), *Wheat: chemistry and technology*. (pp. 277-372). AACC International, Inc., St. Paul, MN.

Liu, S., Chao, S., & Anderson J. A. (2008). New DNA markers for high molecular weight glutenin subunit in wheat. *Theoretical and Applied Genetics, 118*:177-183.

Liu, W., Maccaferri, M., Chen, X., Laghetti, G., Pignone, D., Pumphrey, M., &Tuberosa, R. (2017). Genome-wide association mappingreveals a rich genetic architecture of stripe rust resistance loci in emmer wheat (*Triticum turgidum* ssp. *dicoccum*). *Theoretical and Applied Genetics, 130*:2249-2270.

Liviero, L., Maestri, M., Gullì, M., Nevo, E., & Marmiroli, N. (2002). Ecogeographic adaptation and genetic variation in wild barley, application of molecular markers targeted to environmentally regulated genes. *Genetic Resources and Crop Evolution, 00*:1-0.

Lopes, M. S., El-Basyoni, I., Baenziger, P. S., Singh, S., Royo, C., Ozbek, K., … Vikram, P. (2015). Exploiting genetic diversity from landraces in wheat breeding for adaptation to climate change. *Journal of Experimental Botany, 66*(12):3477-3486.

Lupi, R., Masci, S., Rogniaux, H., Tranquet, O., Brossard, C., Lafiandra, D., … Larré, C. (2014). Assessment of the allergenicity of soluble fraction from GM and commercial genotypes of wheat. *Journal of Cereal Science, 60*:179-186.

Ma, W., Zhang, W., & Gale, K. R. (2003). Multiplex-PCR typing of high molecular weight glutenin alleles in wheat. *Euphytica, 134*:51-60.

Maccaferri, M., Sanguineti, M. C., Donini, P., & Tuberosa R. (2003). MIcrosatellite analysis reveals a progressive widening of the genetic basis in the elite durum wheat germplasm. *Theoretical and Applied Genetics, 107*:783-797.

Maestri, E., Malcevschi, A., Massari, A., & Marmiroli, N. (2002). Genomic analysis of cultivated barley (Hordeum vulgare) using sequence-tagged molecular markers. Estimates of divergence based on RFLP and PCR markers derived from stress-responsive genes, and simple-sequence repeats (SSRs). *Molecular Genetics and Genomics*, *267*:186-201.

Maltin, V., Charabaty, A., & Mangione, R. (2009). Medications: a hidden source of gluten. *Practical Gastroenterology*, *33*:32–36+38.

Mamone, G., Nitride, C., Picariello, G., Addeo, F., Ferranti, P., & Mackie, A. (2015). Tracking the Fate of Pasta (T. Durum Semolina) Immunogenic Proteins by in Vitro Simulated Digestion. *Journal of Agricoltural and Food Chemistry*, *63*:2660-2667.

Marchetti-Deschmann, M., Lehner, A., Peterseil, V., Sövegjarto, F., Hochegger, R., & Allmaier, G. (2011). Fast wheat variety classification by capillary gel electrophoresis-on-a-chip after single-step one-grain high molecular weight glutenin extraction. *Analytical and Bioanalytical Chemistry*, *400*:2403-2414.

Marmiroli, N., Maestri, E., Liviero, L., Massari, A., Malcevschi, A., & Monciardini, P. (1999). Application of genomics in assessing biodiversity in wild and cultivated barley. *Molecular Ecology*, *8*:S95-S106.

Martinez, M. C., Ruiz, M., & Carrillo, J. M. (2005). Effects of different prolamin alleles on durum wheat quality properties. *Journal of Cereal Science*, *41*:123-131.

Masci, S., Lafiandra, D., Porceddu, E., Lew, E. J. L., Tao, H. P., & Kasarda, D. D. (1993). D glutenin subunits: N-Terminal sequences and evidence for the presence of cysteine. *Cereal Chemistry Journal*, *70*:581-585.

Masci, S., Egorov, T. A., Ronchi, C., Kuzmicky, D. D., Kasarda, D. D., & Lafiandra, D. (1999). Evidence for the presence of only one cysteine residue in the D type low-molecular-weight subunits of wheat glutenin. *Journal of Cereal Science, 29*:17-25.

Matsoukas, N. P., & Morrison, W. R. (1991). Breadmaking quality of 10 Greek bread wheats: 2. Relationships of protein, lipid and starch components to baking quality. *Journal of the Science of Food and Agriculture, 55*:87-101.

Matsuoka, Y. (2011). Evolution of polyploid *Triticum* wheats under cultivation: the role of domestication, natural hybridization and allopolyploid speciation in their diversification. *Plant Cell Physiology*, *52*:750–764.

McDonald, J. C., Keynes, H. L., & Meredith, S. K. (2000). Reported incidence of occupational asthma in the United Kingdom 1989–1997. *Occupational and Environmental Medicine*, *57*: 823-9.

Metakovsky, E. V., Akhmedov, M. G., & Sozinov, A. A. (1986). Genetic analysis of gliadin-coding genes reveals gene clusters as well as single remote genes. *Theoretical and Applied Genetics, 73*:278-285.

Metakovsky, E. V., Chernakov, V. M., Upelniek, V. P., Redaelli, R., Dardevet, M., Branlard, G., & Pogna, N. E. (1997). Minor ω-gliadin-coding loci on chromosome 1A of common wheat: a revision. *Journal of Genetics and Breeding, 50*:277-286.

Migliorini, P., Spagnolo, S., Torri, L., Arnoulet, M., Lazzerini, G., & Ceccarelli, S. (2016). Agronomic and quality characteristics of old, modern and mixture wheat varieties and landraces for organic bread chain in diverse environments of northern Italy. *European Journal of Agronomy, 79*:131-141.

Miles, M. J., Carr, H. J., MacMaster, T. C., I'Anson, K. J., Belton, P. S., Morris, V. J., … Tatham, A. S. (1991). Scanning tunneling microscopy of a wheat seed storage protein reveals details of an unusual supersecondary structure. Proceedings of the National Academy of Sciences, *88*:68-71. USA.

Minekus, M., Alminger, M., Alvito, P., Balance, S., Bohn, T., Bourlieu, C., … Brodkorb, A. (2014). A standardised static *in vitro* digestion method suitable for food – an international consensus. *Food and Function, 5*:1113-1124.

Motawei, M. I. (2008). Molecular weight glutenin alleles at *Glu-B1* locus. Genes, Genomes and Genomics, *Global Science Books*; *2*(1):53-56.

Moragues, M., Zarco-Hernandez, J., Moralejo, M. A., & Royo, C. (2006). Genetic diversity of glutenin protein subunits composition in durum wheat landraces [*Triticum turgidum* ssp. *turgidum* convar. *durum* (Desf.) MacKey] from the Mediterranean basin. *Genetic Resources and Crop Evolution, 53*:993–1002.

Moragues, M., Moralejo, M., Sorrells, M. E., & Royo, C. (2007). Dispersal of durum wheat [*Triticum turgidum* L. ssp. *turgidum* convar. *durum* (Desf.) MacKey] landraces across the Mediterranean basin assessed by AFLPs and microsatellites. *Genetic Resources and Crop Evolution, 54*:1133-1144.

Morell, M.K., Rahman, S., Abrahams, S.L., & Appels, R. (1995). The biochemistry and molecular biology of starch synthesis in cereals. *Australian Journal of Plant Physiology, 22*:647-660.

Moreta, D. E., Mathur, P. N., Van Zonneveld, M., Amaya, K., Arango, J., Gomez Selvaraj, M., & Dedicova, B. (2015). Current Issues in Cereal Crop Biodiversity. *Biotechnological Applications of Biodiversity, 147*:1-35.

Nachit, M. M. (1992). *Durum Breeding for Mediterranean Drylands of North Africa and West Asia*. *Wheat Special Report 9*. Centro Internacional de Mejoramiento de Maız y Trigo (CIMMYT),

Mexico.Nakamura, H. (2000). Allelic variation at high-molecular-weight glutenin subunit Loci, *Glu-A1*, *Glu-B1* and *Glu-D1*, in Japanese and Chinese hexaploid wheats. *Euphytica*, *112*:187-193.

Nalam, V. J., Vales, M. I., Watson, C. J. W., Kianian, S. F., & Riera-Lizarazu O. (2006.) Map-based analysis of genes affecting the brittle rachis character in tetraploid wheat (Triticum turgidum L.). *Theoretical and Applied Genetics*, *112*:373–381.

Nazco, R., Villegas, D., Ammar, K., Peña, R. J., Moragues, M., & Royo, C. (2012). Can Mediterranean durum wheat landraces contribute to improved grain quality attributes in modern cultivars? *Euphytica*, *185*:1–17.

Nazco, R., Peña, R. J., Ammar, K., Villegas, D., Crossa, J., Moragues, M., & Royo, C. (2014). Variability in glutenin subunit composition of Mediterranean durum wheat germplasm and its relationship with gluten strength. *Journal of Agricultural Science*, *152*:379-393.

Nei, M. (1973). Analysis of Gene Diversity in Subdivided Populations. *Proceedings of the Natural Academy of Sciences* USA, *70*(12), Part I:3321-3323.

Oliveira, H. R., Jounes, H., Leigh, F., Lister, D. L., Jones, M. K., & Peña-Chocarro, L. (2011). Phylogeography of einkorn landraces in the Mediterranean basin and Central Europe: population structure and cultivation history. *Archaeological and Anthropological Science*, *3*:327.

Oliveira, H. R., Campana, M. G., Jones, H., Hunt, H. V., Leigh, F., Redhouse, D. I., … Jones, M. K. (2012). Tetraploid Wheat Landraces in the Mediterranean Basin: Taxonomy, Evolution and Genetic Diversity. *PlosOne*, *7*(5):e37063.

Onipe, O. O., Jideani, A. I. O., & Beswa, D. (2015). Composition and functionality of wheat bran and its application in some cereal food products. *International Journal of Food Science and Technology*, *50*:2509–2518.

Osborne, T.B. (1924). *The vegetable proteins* (2th ed). Longmans, Green and Co., London.

Pareyt, B., Finnie, S., Putseys, J., & Delcour, J. (2011). Lipids in bread making: sources, interactions, and impact on bread quality. *Journal of Cereal Science*, *54*:266-279.

Patacchini, C., Masci, S., & Lafiandra, D. A. (2001). New method to purify y-type high-molecular weight glutenin subunits from wheat. *Journal of Cereal Science*, *34*(2):135-139.

Payne, P. I., Corfield, K. G., & Blackman, J. A. (1979). Identification of a high-molecular-weight subunit of glutenin whose presence correlates with bread-making quality in wheats of related pedigree. *Theoretical and Applied Genetics*, *55*:153.

Payne, P. I., Law, C. N., & Mudd, E. E. (1980). Control by homoeologous group 1 chromosomes of the high-molecular-weight subunits of glutenin, a major protein of wheat endosperm. *Theoretical and Applied Genetics*, *58*:113-120.

Payne, P. I., Corfield, K. G., Holt, L. M., & Blackman, J. A. (1981). Correlation between the inheritance of certain high-molecular weight subunits of glutenin and bread making quality in progenies of six crosses of bread wheat. *Journal of the Science of Food and Agricolture*, *32*(1):51-60.

Payne, P. I., & Lawrence, G. J. (1983). Catalogues of alleles for the complex gene loci Glu-A1, Glu-B1 and Glu-D1 which code for high molecular weight subunits of glutenin in hexaploid wheat. *Cereal Research Communications*, 11(1):29-35.

Payne, P. I., Holt, L. M., Jackson, E. A., & Law, C. N. (1984). Wheat storage proteins: their genetics and their potential for manipulation by plant breeding. *Philosophical Transactions of the Royal Society of London. Series B*, *304*:359-371.

Pecetti, L., Annicchiarico, P., & Damania, A. B. (1992). Biodiversity in a germplasm collection of durum wheat. *Euphytica*, *60*(3):229-238.

Peng, J. H., Sun, D., & Nevo, E. (2011). Wild emmer wheat, *Triticum dicoccoides*, occupies a pivotal position in wheat domestication process. *Australian Journal of Crop Science*, *9*:1127-1143.

Pignone, D., & Hammer, K. (2013). In: Kole, C. (Ed), *Genomics and breeding for climate-resilient crops*. (vol 1. pp 9–26). Springer, Berlin Heidelberg.

Pignone, D., De Paola, D., Rapanà, N., & Janni, M. (2015). Single seed descent: a tool to exploit durum wheat (*Triticum durum* Desf.) genetic resources. *Genetic Resources and Crop Evolution*, *62*:1029-1035.

Pignone D., De Paola D., Rapanà N., & Janni M. (2016). Capturing wild relative and landrace diversity for crop improvement using a new selection tool to exploit genetic resources in durum wheat. In Maxted, N., Dulloo, M. E., & Ford-Lloyd, B. V. (Eds), Enhancing crop genepool use: capturing wild relative and landrace diversity for crop improvement, Chapter 6: 47-53.

Piotto, B., Giacanelli, V., & Ercole, S. (2010). La conservazione ex situ della biodiversita delle specie vegetali spontanee e coltivate in Italia: stato dell'arte, criticita e azioni da compiere. Manuali e linee guida, 54. ISPRA—Istituto Superiore per la protezione e la ricerca ambientale, Roma. ISBN 978-88448-0416-9.

Pogna, N. E., Autran, J. C., Mellini, F., Lafiandra, D., & Feillet, P. (1990). Chromosome 1B-encoded gliadins and glutenin subunits in durum wheat: genetics and relationship to gluten strength. *Journal of Cereal Science*, *11*:15-34.

Pogna, N. E., Metakovsky, E. V., Readelli, R., Raineri, F., & Dachkevitch, T. (1993). Recombination mapping of *Gli-5*, a new gliadin-coding locus on chromosome 1A and 1B in common wheat. *Theoretical and Applied Genetics*, *87*:113-121.

Pogna, N. E., Metakovsky, E. V., Readelli, R., Dachkevitch, T., & Cernacov, V. M. (1995). The group 1 chromosomes of wheat contain several loci coding for gliadins. In Li, Z. S., & Xin, Z. Y. (Eds), Proceedings of the 8th International Wheat Genetic Symposium, pp 137-140. China, Agriculture Scientech Press, Beijing, China.

Poole, J. A., Barriga, K., Leung, D. Y., Hoffman, M., Eisenbarth, G. S., Rewers, M., & Norris, J. M. (2006). Timing of initial exposure to cereal grains and the risk of wheat allergy. *Pediatrics*, *117*(6): 2175-2182.

Porceddu, E., Perrino, P., & Olita, G. (1973). Preliminary information on an Ethiopian wheat germplasm collection mission. In Scarascia Mugnozza, G. T. (Ed), *Proceedings of the Symposium on Genetic and Breeding of Durum Wheat*, pp 181-200.

Prandi, B., Mantovani, P., Galaverna, G., & Sforza, S. (2014). Genetic and environmental factors affecting pathogenicity of wheat as related to celiac. *Journal of Cereal Science*, *59*:62-69.

Prandi, B., Tedeschi, T., Folloni, S., Galaverna, G., & Sforza, S. (2017). Peptides from gluten digestion: a comparison between old and modern wheat varieties. *Food Research International*, *91*:92-102.

Raciti, C. N., Doust, M. A., Lombardo, G. M., Boggini, G., & Pecetti, L. (2003). Characterization of durum wheat mediterranean germplasm for high and low molecular weight glutenin subunits in relation with quality. *European Journal of Agronomy*, *19*:373-382.

Ragupathy, R., Naeem, H. A., Reimer, E., Lukow, O. M., Sapirstein, H. D., & Cloutier, S. (2008). Evolutionary origin of the segmental duplication encompassing the wheat *Glu-B1* locus encoding the overexpressed Bx7 (Bx7OE) high molecular weight glutenin subunit. *Theoretical and Applied Genetics*, *116*:283–296.

Rahman, S., Li, Z., Batey, I., Cochrane, M.P., Appels, R., & Morell, M.K. (2000). Genetic alteration of starch functionality in wheat. *Journal of Cereal Science*, *31*:91-110.

Rasheed, A., Mahmood, T., Gul Kazi, A., Ghafoor, A., & Mujeeb-Kazi, A. (2012). Allelic Variation and Composition of HMW-GS in Advanced Lines Derived from D-genome Synthetic Hexaploid/Bread Wheat (*Triticum aestivum* L.). *Journal of Crop Science and Biotechnology*, *15*(1):1-7.

Rasheed, A., Xia, X., Yan, Y., Appels, R., Mahmood, T., & He, Z. (2014). Wheat seed storage proteins: Advance in molecular genetics, diversity and breeding applications. *Journal of Cereal Science*, *60*:11-24.

Ribeiro, M., Carvalho, C., Carnide, V., Guedes-Pinto, H., & Igrejas, G. (2011). Towards allelic diversity in the storage proteins of old and currently growing tetraploid and hexaploid wheats in Portugal. *Genetic Resources and Crop Evolution*, *58*:1051-1073.

Rodriguez-Quijano, M., & Carrillo, J. M. (1996). Linkage map of prolamin loci Gli-D4 and Gli-D5 in hexaploid wheat. *Plant Breeding*, *115*:189-191.

Ruiz, M., & Carrillo, J. M. (1993). Linkage relationships between prolamin genes on chromosomes 1A and 1B of durum wheat. *Theoretical and Applied Genetics*, *87*:353-360.

Sajjad, M., Khan, S. H., & Salam Khan, A. (2011). Exploitation of Germplasm for Grain Yield Improvement in Spring Wheat (*Triticum aestivum*). *International Journal of Agriculture & Biology*, *13*:695-700.

Salamini, F., Özkan, H., Brandolini, A., Schafer-Pregl, R., & Martin, W. (2002). Genetics and geography of wild cereal domestication in the Near East. *Nature Reviews Genetic*, *3*:429–41.

Salcedo, G., Quirce, S., & Diaz-Perales, A. (2011). Wheat allergens associated with Baker's asthma. *Journal of Investigational Allergology and Clinical Immunology*, *21*:81-92.

Salmanowicz, B. P., & Dylewicz, M. (2007). Identification and characterization of high-molecular-weight glutenin genes in Polish triticale cultivars by PCR-based DNA markers. *Journal of Applied Genetic*, *48*(4):347-357.

Salmi, T. T, Hervonen, K. Kautiainen, H., Collin, P., & and Reunala, T. (2011). Prevalence and incidence of dermatitis herpetiformis: a 40-year prospective study from Finland. *British Journal of Dermatology*, *165*:354-359.

Sapone, A., Bai, J. C., Ciacci, C., Dolinsek, J., Green, P. H. R., Hadjivassiliou, M., … Fasano, A. (2012). Spectrum of gluten-related disorders: consensus on new nomenclature and classification. *BMC Medicine*, *10*:13.

Scarascia Mugnozza, G. T. (2005). The Contribution of Italian Wheat Geneticists: From Nazareno Strampelli to Francesco D'Amato. Rome, Accademia Nazionale delle Scienze.

Schalk, K., Lang, C., Wieser, H., Koehler, P., & Scherf, K. A. (2017). Quantitation of the immunodominant 33-mer peptide from α-gliadin in wheat flours by liquid chromatography tandem mass spectrometry. *Scientific Reports*, *7*:45092.

Shi, X., & Ling, H. Q. (2017). Current advances in genome sequencing of common wheat and its ancestral species. *The Crop Journal*, *6*:15-21.

Schropp, P. & Wieser, H. (1996). Effects of High Molecular Weight Subunit of Glutenin on rheological properties of wheat gluten. *Cereal Chemistry*, *73*(3):410-413.

Schupann, D., Junker, Y., Zevallos, V., & Wieser, H. (2012). Cereal triggers of innate immune activation. In: Koehler, P. (Ed.), *25th Meeting Working Group on Prolamin Analysis and Toxicity*. (pp. 93e96). German Research Centre for Food Chemistry, Germany.

Schwarz, G., Felsenstein, F. G., & Wenzel, G. (2004). Development and validation of a PCR-based marker assay for negative selection of the HMW glutenin allele Glu-B1-1d (Bx-6) in wheat. *Theoretical and Applied Genetics, 109*:1064-1069.

Sestili, F., Janni, M., Doherty, A., Botticella, E., D'Ovidio, R., Masci, S., … Lafiandra, D. (2010). Increasing the amylose content of durum wheat through silencing of the SBEIIa genes. *BMC Plant Biology*, *10*:144.

Shepherd, S. J., Halmos, E., & Glance, S. (2014). The role of FODMAPs in irritable bowel syndrome. *Current Opinion in Clinical Nutrition and Metabolic Care, 17*:605-609.

Shewry, P. R., Tatham, A. S., Forde, J., Kreis, M., & Miflin, B. J. (1986). The classification and nomenclature of wheat gluten proteins: a reassessment. Journal of Cereal Science, 4:97-106.

Shewry, P. R., Halford, N. G., & Tatham, A. S. (1989). The high molecular weight subunits of wheat, barley and rye: Genetics, molecular biology, chemistry and role in wheat gluten structure and functionality. In Miflin, B. J. (Ed), *Oxford Surveys of Plant and Molecular Cell Biology*. (Vol. 6, pp 163-219). Oxford University Press: London.

Shewry, P. R., Halford, N. G., & Tatham, A. S. (1992). The high molecular weight subunits of wheat glutenin. *Journal of Cereal Science, 15*:105-120.

Shewry, P. R., Napier, J. A., & Tatham, A. S. (1995). Seed Storage Proteins: Structures 'and Biosynthesis. *The Plant Cell*, *7*:945-956.

Shewry, P.R., & Tatham, A.S. (1997). Disulphide Bonds in Wheat Gluten Proteins. *Journal of Cereal Science*, *25*:207–227.

Shewry, P. R., & Halford, N. G. (2002). Cereal seed storage proteins: structures, properties and role in grain utilization. *Journal of Experimental Botany*, *53*(370):947-958.

Shewry, P. R., Halford, N. G., Belton, P. S., & Tatham, A. S. (2002). The structures and properties of gluten: an elastic protein from wheat grain. *Philosophical Transactions of the Royal Society of London B, 357*:133–142.

Shewry, P., Halford, N., & Lafiandra, D. (2003). Genetics of wheat gluten proteins. *Advances in Genetics*, vol *49*, chapter 3.

Shewry, P. R. (2009). Wheat. *Journal of Experimental Botany*, *60*(6):1537-1553.

Shewry, P., & Tatham, A. (2016). Improving wheat to remove coeliac epitopes but retain functionality. *Journal of Cereal Science*, *67*:12−21.

Sicherer, S. H, & Sampson, H. A. (2006). Food allergy. *Journal of Allergy and Clinical Immunology*, *117*(2 Suppl Mini-Primer):S470-5.

Siddique, K. H. M., Belford, R. K., Perry, M. W., & Tennant, D. (1989). Growth, development and light interception of old and modern wheat cultivars in a Mediterranean-type environment. *Australian Journal of Agricultural Research*, *40*:473-487.

Singh, H., & MacRitchie, F. (2001). Application of polymer science to properties of gluten. *Journal of Cereal Science*, *33*:231-243.

Singh, N. K., Shepherd, K. W., & Cornish, G. B. (1991). A simplified SDS-PAGE procedure for separating LMW subunit of glutenin. *Journal of Cereal Science*, *14*:203-208.

Sissons, M. J., Ames, N. P., Hare, R. A., & Clarke, J. M. (2005). Relationship between glutenin subunit composition and gluten strength measurements in durum wheat. *Journal of the Science of Food and Agriculture*, *85*:2445-2452.

Sissons, M. J. (2008). Role of Durum Wheat Composition on the Quality of Pasta and Bread. *Food*, *2*(2):75-90.

Snape, J. W. (1976). A theoretical comparison of diploidized haploid and single seed descent populations. *Heredity*, *36*:275–277.

Sobko, T. A. (1984). Identification of a new locus which controls the synthesis of alcohol-soluble protein in the endosperm of winter common wheat. *Journal of Agricultural Science*, (Kiev) *7*:78-80.

Sonnante, G., Hammer, K., & Pignone, D. (2009). From the cradle of agriculture a handful of lentils: history of domestication. *Rendiconti Lincei*, *20*(1):21-37.

Šramková, Z., Gregová, E., & Šturdìk, E. (2009). Chemical composition and nutritional quality of wheat grain. *Acta Chimica Slovaca*, *2*(1):115-138.

Sreeramulu, G., & Singh, N. K. (1997). Genetic and biochemical characterization of novel low molecular weight glutenin subunits in wheat (*Triticum aestivum* L.). *Genome*, *40*:41-48.

Sun, C., Zhang, F., Yan, X., Zhang, X., Dong, Z., Cui, D., & Chen, F. (2017). Genome-wide association study for 13 agronomic traits reveals distribution of superior alleles in bread wheat from the Yellow and Huai Valley of China. *Plant Biotechnology Journal*, *15*:953–969.

Svihus, B., Uhlen, A. K., & Harstad, O. M. (2005). Effect of starch granule structure, associated components and processing on nutritive value of cereal starch: A review. *Animal Feed Science and Technology*, *122*:303–320.

Szécsényi-Nagy, A., Brandt, G., Keerl, V., Jakucs, J., Haak, W., Möller-Rieker, S., … Alt, K. F. (2015). Tracing the genetic origin of Europe's first farmers reveals insights into their social organization. Proceedings B Biological Science, rspb.royalsocietypublishing.org,

282(1805):20150339.

Tatham, A. S., & Shewry, P. R. (1985). The conformation of wheat gluten proteins: The secondary structures and thermal stabilities of α-, β-, γ- and ω-Gliadins. *Journal of Cereal Science*, *3*:103-113.

Tatham, A. S., Miflin, B. J., & Shewry, P. R. (1985). The beta-turn conformation in wheat gluten proteins: Relationship to gluten elasticity. *Cereal Chemistry Journal*, *62*:405-442.

Tatham, A. S., Drake, A. F., & Shewry, P. R. (1990). Conformational studies of synthetic peptides corresponding to the repetitive region of the high molecular weight (HMW) glutenin subunits of wheat. *Journal of Cereal Science*, *11*:189-200.

Thomson, N. H., Miles, M. J., Tatham, A. S., & Shewry, P. R. (1992). Molecular images of cereal proteins by STM. *Ultramicroscopy*, *4244*:1118-1122.

Tilley, K. A., Benjamin, R. E., Bagorogoza, K. E., Okot-Kotber, B. M., Prakash, O., & Kwen, H. (2001). Tyrosine cross-links: molecular basis of gluten structure and function. *Journal of Agriculture Food Chemistry*, *49*:2627-2632.

Tovoli, F., Masi, C., Giudetti, E., Negrini, G., Paterini, P., & Bolondi, L. (2015). Clinical and diagnostic aspects of gluten related disoders. *World Journal of Clinical Cases*, *3*(3):275-284.

Troccoli, A., Borrelli, G. M., De Vita, P., Fares, C., & Di Fonzo, N. (2000). Durum wheat quality: a multidisciplinary concept. *Journal of Cereal Science, 32*:99-113.

Tsetskhladze, G. R. (2008). Greek colonization: an account of Greek colonies and other settlements overseas. Koninklijke Brill NV, Leiden, The Netherlands.

Uthayakumaran, S., Listiohadi, Y., Baratta, M., Batey, I. L., & Wrigley, C. W. (2006). Rapid identification and quantitation of high-molecular-weight glutenin subunits. *Journal of Cereal Science*, *44*(1):34-39.

UPOV (1994). Wheat-guidelines for the conduct of tests for distinctness, homogeneity and stability. TG/3/11 + Corrections. International Union for the Protection of New Varieties of Plants, Geneva.

van den Broeck, H. C., de Jong, H. C., Salentijn, E. M. J., Dekking, L., Bosch, D., Hamer, R. J., … Smulders, M. J. M. (2010). Presence of celiac disease epitopes in modern and old hexaploid wheat varieties: Wheat breeding may have contributed to increased prevalence of celiac disease. *Theoretical and Applied Genetics*, *121*:1527–1539.

Varshney, R.K., Hoisington, D.A., & Tyagi, A. K. (2006). Advances in cereal genomics and applications in crop breeding. *Trends Biotechnology*, *24* (11):490-499.

Varzakas, T., Kozub, N., & Xynias, I. N. (2014). Quality determination of wheat: genetic determination, biochemical markers, seed storage proteins - bread and durum wheat germplasm. *Journal of the Science of Food and Agriculture*, *94*:21819-2829.

Visioli, G., Comastri, A., Imperiale, D., Paredi, G., Faccini, A., & Marmiroli, N. (2015). Gel-based and gel-free analytical methods for the detection of HMW-GS and LMW-GS in wheat flour. *Food Analytical Methods*, *9*:469-474.

Volta, U., Caio, G., Tovoli, F., & De Giorgio, R. (2013). Non-celiac gluten sensitivity: questions still to be answered despite increasing awareness. *Cellular and Molecular Immunology*, *10*:383-392.

Wang, C., Hu, S., Gardner, C., & Lübberstedt, T. (2017). Emerging Avenues for Utilization of Exotic Germplasm. *Trends in Plant Science*, *22*(2): 624-637.

Wieser, H. (2007). Chemistry of gluten proteins. *Food Microbiology*, *24*:115-119.

Wrigley, C. W., & Shepherd, K. W. (1973). Electrofocusing of grain proteins from wheat genotypes. *Annals of the New York Academy of Sciences*, *209*:154–162.

Wrigley, C. W., & Bietz, J. A. (1988). Proteins and amino acids. In Pomeranz, Y. (Ed), Wheat – Chemistry and Technology. (vol 1: pp. 159-275). American Association of Cereal Chemistry, St. Paul.

Wrigley, C. W. (1996). Giant proteins with flour power. *Nature*, *381*:738-739.

Wrigley, C. W., Bekes, F., Cavanagh, C. R., & Bushuk, W. (2006). The Gluten Composition of Wheat Cultivars and Genotypes, PART II. Composition table for the HMW subunits of Glutenin. Retrieved from http://www.accnet.org

Xu, Q., Xu, J., Liu, C. L., Chang, C., Wang, C. P., You, M. S., … Liu, G. T. (2008). PCR-based markers for identification of HMW-GS at Glu-B1x loci in common wheat. *Journal of Cereal Science*, *47*:394-398.

Yamamori, M., Nakamura, T., Endo, T. R., & Nagamine, T. (1994). Waxy protein- deficiency and chromosomal location of coding genes in common wheat. *Theoretical and Applied Genetics*, *89*:179-184.

Zaïm, M., El Hassouni, K., Gamba, F., Filali-Maltouf, A., Belkadi, B., Sourour, A., … Bassi, F. (2017). Wide crosses of durum wheat (Triticum durum Desf.) reveal good disease resistance, yield stability, and industrial quality across Mediterranean sites. *Field Crop Research, 214*:219-227.

Zeven, A. C. (1998). Landraces: a review of definitions and classifications. *Euphytica, 104*(2):127-39.

Zeven, A.C. (1999). The traditional inexplicable replacement of seed and seedware of landraces and cultivars: a review. *Euphytica*, *110*:181-191.

Zhang, Y. (2015). *Investigations of Gluten Proteins from Functional and Historical Perspectives* (PhD Thesis). A Oregon State University.

Zheng, W., Peng, Y., Ma, J., Appels, R., Sun, D., & Ma, W. (2011). High frequency of abnormal high molecular weight glutenin alleles in Chinese wheat landraces of the Yangtze-River region. *Journal of Cereal Science*, *54*:401-408.

Zhou, W., Therdthai, N., & Hui, Y. H. (2014). Introduction to Baking and Bakery Products. In Zhou, W., Hui, Y. H., De Leyn, I., Pagani, M. A., Rosell, C. M., Selman, J. D., Therdthai, N, (Eds.), *Bakery Products Science and Technology* (2nd ed). (pp. 3-16). John Wiley & Sons, Ltd.

Zhu, C., Gore, M., Buckler, E. S., & Yu, J. (2008). Status and Prospects of Association Mapping in Plants. *The Plant Genome*, *1*(1):5-20.

Zilic S., Barac, M., Pesic, M., Dodig, D., & Ignjatovic-Micic, D. (2011). Characterization of proteins from grain of different bread and durum wheat genotypes. *International Journal of Molecular Science*, *12*:5878–5894.

Živančev, D. R., Nikolovski, B. G., Torbica, A. M., Mastilović, J. S., & Dukić, N. H. (2013). Lab-on-a-chip method uncertainties in determination of high-molecular-weight glutenin subunit. *Chemical Industry & Chemical Engineering Quarterly*, *19*(4):553-561.

Zivy, M., Wienkoop, S., Renaut, J., Pinheiro, C., Goulas, E., & Carpentier, S. (2015). The quest for tolerant varieties: The importance of integrating “omics” techniques to phenotyping. *Frontiers in Plant Science*, *6*:448.

Web Sites

Celiac Disease Foundation: https://celiac.org/celiac-disease/understanding-celiac-disease-2/dermatitis-herpetiformis/

EURISCO database: http://eurisco.ecpgr.org

FAO, Seeds and Plant Genetic Resources: http://www.fao.org/agriculture/crops/core-themes/theme/seeds-pgr/en/

FAO, The State of Food and Agriculture: climate change, agriculture and food security: http://www.fao.org/3/a-i6030e.pdf

FAOSTAT: http://faostat3.fao.org

Gf jules: https://gfjules.com/celiac-disease/

GrainGenes Marker Report, Glu-B1: https://wheat.pw.usda.gov/cgi-bin/GG3/report.cgi?class=marker&name=Glu-B1

Seed Bank of the Institute of Bioscience and Bioresources of the National Research Council of Bari: http://ibbr.cnr.it/mgd/

Acknowledgments

The author would thank the Chemical and Life Science and Environmental Sustainability Department of Parma for providing financial support and all the equipment and facilities used in this project. I gratefully thank Prof. Nelson Marmiroli as my supervisor for giving me the opportunity to conduct my PhD on a very interesting and present-day topic and for giving me useful advices during this period. I would like to thank my tutor, Dr. Michela Janni of IBBR CNR of Bari (IT) for her guidance and, above all, for demonstrating me with her daily behavior how a good researcher has to be and what has to do. I would also like to thank Dr. Aldo Ceriotti and Dr. Aldo Grasso of IBBA of Milan (IT) for the collaboration in analyzing data and for the precious discussions which have deepen the knowledge about my project, as well as Domenico Pignone of IBBR CNR of Bari (IT). I'd like to expresso my gratitude to Dr. Stefano Sforza, of Food and Drug Department, University of Parma, and to his team for hosting me in their laboratory and for giving me the opportunity to conduct a final and crucial part of my project. A special thanks to CREA-SCV Sant'Angelo Lodigiano, University of Tuscia-Viterbo, Rothamsted Research Experimental Station, UK, and CREA-Foggia for providing the genetic material useful for the project. Finally, the author would like to thank for financially supporting part of the research described the projects FOODINTEGRITY (EU-FP7 No. 613688), FACCE-SURPLUS project INTENSE and Smart Wheat project, POS FESR 2014-2020 – action 1.2.2.

To all the Professors, Researchers, Post-doc, technician and PhD Students of my group Prof. Elena Maestri, Dr. Giovanna Visioli, Prof. Mariolina Gullì, Dr. Roberta Ruotolo, Dr. Marina Caldara, Dr. Caterina Agrimonti, Urbana Bonas, Dr. Aliosha Malcevschi, Prof. Annamaria Sanangelantoni, Dr. Marta Marmiroli, Dr. Davide Imperiale, Dr. Sara Graziano, Dr. Luca Pagano, Dr. Francesca Mussi, Dr. Laura Paesano, Dr. Graziella Pira, Dr. Valentina Gallo, Dr. Valentina Buffagni, Dr. Riccardo Rossi, Dr. Giacomo Lencioni, Dr. Serena Pantalone and Dr. Jacopo Magnani, a heartfelt thanks for both the scientific and human support that I've been always able to perceive.

In the end, a special thanks to my husband Alessandro, to my parents and to my great friend, Francesco, who have been at every moment an important support.

List of publications

The scientific productions carried out during my PhD include: one national conference proceeding related to the molecular analysis applied in the research project presented in this PhD thesis; a paper published on Plant Breeding and a submitted paper.

The publications are listed and attached following:

- CADONICI S., JANNI M., PIGNONE D., MARMIROLI N.
  Survey and new insights in the application of PCR-based molecular markers for the identification of HMW-GS ate the *Glu-B1* locus in durum and bread wheat. Poster Communication Abstract – 7.13 Proceedings of the LX SIGA Annual Congress. *Catania, Italy – 13/16 September, 2016*. ISBN 978-88-904570-6-7.

- JANNI, M., CADONICI, S., PIGNONE, D., MARMIROLI, N. (2017). Survey and new insights in the application of PCR-based molecular markers for identification of HMW-GS at the *Glu-B1* locus in durum and bread wheat. *Plant Breeding*, *136*(4):467-473. (**ANNEX A**)

- JANNI, M., CADONICI, S., BONAS, U., GRASSO, A., DAHAB, A. A. D., VISIOLI, G., PIGNONE, D., CERIOTTI, A., MARMIROLI, N. Gene-ecology of durum wheat HMW glutenin reflects their diffusion from the center of origin. (Submitted).

DiBA
Dipartimento di Agricoltura Alimentazione e Ambiente

SOCIETÀ ITALIANA GENETICA AGRARIA

crea

ITALIAN SOCIETY OF AGRICULTURAL GENETICS

PROGRAMME

POSTER LIST

Catania, 13th - 16th September 2016

Proceedings of the LX SIGA Annual Congress
Catania, Italy – 13/16 September, 2016
ISBN **978-88-904570-6-7**

Poster Communication Abstract – 7.13

SURVEY AND NEW INSIGHTS IN THE BAPPLICATION OF PCR-BASED MOLECULAR MARKERS FOR THE IDENTIFICATION OF HMW-GS AT THE *Glu-B1* LOCUS IN DURUM AND BREAD WHEAT

CADONICI S.*, JANNI M.**, PIGNONE D.**, MARMIROLI N.*

*) Department of Life Sciences, University of Parma, Parco Area delle Scienze 11/A, 43124 Parma (Italy)
**) Institute of Bioscience and Bioresources (IBBR), National Research Council (CNR), Via Amendola 165/A, 70126 Bari (Italy)

gluten, HMW-GS, molecular markers, Glu-B1 locus, Triticum

Wheat is one of the major crops grown and consumed in human diet thanks to its nutritional value and technological properties. Modern food industries require flours with high technological characteristics, like high gluten strength, for the production of pasta, bakery products and other wheat-derived goods, thus making high dough quality and technological properties main goals for wheat breeders. These features are conferred by gluten polymer; in particular, High Molecular Weight glutenin subunits (HMW-GS) are of considerable interest because of their relationship to bread-making quality and since they contribute to strengthen and stabilize dough. HMW-GS are encoded by the complex loci *Glu-1* of the homologous chromosomes 1A and 1B where each locus presents two tightly linked genes coding for two different protein subunits, the *x*-type and the *y*-type, of higher and lower molecular weight respectively. The identification of allelic composition at the *Glu-B1* locus is very important to wheat quality improvement, allowing the investigation of their correlation with gluten strength and quality. Recently several PCR based molecular markers to tag specific HMW glutenin genes encoding *Bx* and *By* subunits have been developed. In this research paper a survey of the molecular markers developed for the HMW-GS at *Glu-B1* locus is presented; 17 durum and bread wheat test cultivars have been used with a panel of 6 molecular markers for the identification of the *Glu-B1* alleles. New insights in the discrimination of HMW-GS alleles at the *Glu-B1* locus are presented.

Received: 27 September 2016 | Accepted: 19 May 2017

DOI: 10.1111/pbr.12506

ORIGINAL ARTICLE

WILEY Plant Breeding

Survey and new insights in the application of PCR-based molecular markers for identification of HMW-GS at the *Glu-B1* locus in durum and bread wheat

Michela Janni[1,2,*] | Stefania Cadonici[3*] | Domenico Pignone[1] | Nelson Marmiroli[3]

[1]Institute of Bioscience and Bioresources (IBBR), National Research Council (CNR), Bari, Italy

[2]Institute of Materials for Electronics and Magnetism, Parma, Italy

[3]Department of Chemistry, Life Sciences and Environmental Sustainability, University of Parma, Parma, Italy

Correspondence
Michela Janni, Institute of Bioscience and Bioresources (IBBR), National Research Council (CNR), Bari, Italy.
Email: michela.janni@ibbr.cnr.it

Funding information
POS-FER, Grant/Award Number: 2014-2020; ERA-NET FACCE SURPLUS, Grant/Award Number: ID169; Progetto Bandiera Interomics, Grant/Award Number: 2014-2020; RGV-FAO 4 - MIPAF

Communicated by R. McIntosh

Abstract

Wheat, among all cereal grains, possesses unique characteristics conferred by gluten; in particular, high molecular weight glutenin subunits (HMW-GS) are of considerable interest as they strictly relate to bread-making quality and contribute to strengthening and stabilizing dough. Thus, the identification of allelic composition, in particular at the *Glu-B1* locus, is very important to wheat quality improvement. Several PCR-based molecular markers to tag-specific HMW glutenin genes encoding *Bx* and *By* subunits have been developed in recent years. This study provides a survey of the molecular markers developed for the HMW-GS at the *Glu-B1* locus. In addition, a selection of molecular markers was tested on 31 durum and bread wheat cultivars containing the By8, By16, By9, Bx17, Bx6, Bx14 and Bx17 *Glu-B1* alleles, and a new assignation was defined for the ZSBy9_aF1/R3 molecular marker that was specific for the By20 allele. We believe the results constitute a practical guide for results that might be achieved by these molecular markers on populations and cultivars with high variability at the *Glu-B1* locus.

KEYWORDS
high molecular weight glutenin, *Triticum aestivum*, *Triticum durum*

1 | INTRODUCTION

Wheat is one of the most important cereals worldwide in terms of production and utilization as a staple food. It is also a source of energy for industry, protein and dietary fibre in human nutrition and is the most widely grown cereal in the world in terms of land surface. It is widely accepted that the amount and composition of gluten proteins are responsible for qualitative differences between durum and bread wheat, in terms of pasta and bread-making properties (Lafiandra, Sanguineti, Maccaferri, & Deambrogio, 2007). Gluten consists mainly of glutenins, responsible for the elasticity of dough, and gliadins, which provide extensibility (Gianibelli, Larroque, Mac Ritchie, & Wrigley, 2001). Glutenin polymers are composed of high molecular weight glutenins (HMW-GS) and low molecular weight glutenins (LMW-GS) linked by disulphide bonds, which greatly influence the quality characteristics of wheat (Shewry, Halford, & Tatham, 1992). In particular, HMW-GS play an important role in affecting viscoelastic properties by determining the size and distribution of glutenin polymers (Lafiandra et al., 2007). HMW-GS are encoded by genes at the *Glu-1* loci present on the long arms of homoeologous group 1 chromosomes (Payne & Lawrence, 1983). Each *Glu-1* locus (*Glu-A1*, *Glu-B1* and *Glu-D1)* has two tightly linked HMW-GS genes, which encode two subunits, one larger and one smaller, defined as x-type and y-type, respectively, and identified after separation of proteins by sodium dodecyl sulphate–polyacrylamide gel electrophoresis (SDS-PAGE) under denaturing conditions (Payne & Lawrence, 1983; Shewry et al., 1992). A close relationship has been demonstrated between HMW-GS composition and bread-making quality (Branlard & Dardevet, 1985).

Until now, SDS-PAGE on seed proteins has been routinely used to analyse and discriminate a large number of HMW-GS alleles

In memory of Prof. Renato D'Ovidio.
*These authors contributed equally to this work

(Shewry et al., 1992) but may result in incorrect interpretation of allelic differences (Gianibelli et al., 2001).

PCR-based DNA markers provide an additional tool to overcome these constraints and also allow genotyping at early growth stages. On the basis of the available HMW-GS sequences, a large number of PCR-based DNA molecular markers have been developed and applied (D'Ovidio, Masci, & Porceddu, 1995; Lafiandra, Tucci, Pavoni, Turchetta, & Margiotta, 1997; Lei et al., 2006; Ma, Zhang, & Gale, 2003). Notwithstanding the large number of markers developed, the allelic specificity has been verified on specific and limited numbers of test cultivars (cvs) selected on a case by case basis and representative of the expected HMW-GS subunits at the *Glu-B1* locus (Lei et al., 2006; Ma et al., 2003). When applied to different species such as triticale (x*Tricosecale)*, a wider discrimination of HMW-GS alleles at the *Glu-B1* locus was observed (Salmanowicz & Dylewicz, 2007).

In the present research, we used a set of molecular markers specifically developed for the *Glu-B1* locus on 31 test cultivars of durum and bread wheat containing a specific set of *Glu-B1* alleles. The PCR analysis revealed new banding patterns for many of the primer pairs and some of them, as far as we know, were not previously reported. Our results confirmed that a combination of SDS-PAGE and molecular markers is still necessary for HMW-GS characterization in wheat genotypes to avoid allele misidentification, but the use of PCR markers allows at least the identification of new haplotypes. In addition, a survey of all molecular markers available for the *Glu-B1* locus is presented as a useful tool for researchers working on wheat quality.

2 | MATERIALS AND METHODS

2.1 | Plant material

On the basis of the *Glu-B1* composition reviewed in Wrigley, Bekes, Cavanagh, and Bushuk (2006), 31 cultivars (Table 1), representing the most common HMW-GS subunits, were used in this study. Nineteen bread wheat cultivars and twelve durum wheat cultivars were analysed for their allele composition at the *Glu-B1* locus.

TABLE 1 Cultivars used in the study.

Cultivar name	Species	HMW-GS subunits at the *Glu-B1* locus
Argelato	*Ta*	Bx7
David	*Ta*	By18*
Carme	*Ta*	By18*
Garibaldino	*Ta*	By18*
Chinese Spring	*Ta*	Bx7+By8
Ciccio	*Td*	Bx7+By8
Cham	*Td*	Bx7+By8
Abbazia	*Ta*	Bx7*+By8
Emilio Morandi	*Ta*	Bx7*+By8
Firenze	*Ta*	Bx7*+By8
Tiberio	*Ta*	Bx7*+By8
Abano	*Ta*	Bx7+By9
Adriano	*Ta*	Bx7+By9
Cheyenne	*Ta*	Bx7*+By9
Creso	*Td*	Bx6+By8
Dylan	*Td*	Bx6+By8
Florida	*Ta*	Bx6+By8
Langdon	*Td*	Bx6+By8
Saragolla	*Td*	Bx6+By8
Est Mottin	*Ta*	Bx6*+By8*
Sieve	*Ta*	Bx6*+By8*
Virest	*Ta*	Bx6*+By8*
Isa	*Td*	Bx13+By16
Solitario	*Td*	Bx13+By16
Cadenza	*Ta*	Bx14+By15
Colosseo	*Td*	Bx14+By15
Durazio Rijo	*Td*	Bx14+By19
Fabiola	*Ta*	Bx17+By18
Francia	*Ta*	Bx17+By18
Capeiti	*Td*	Bx20+By20
Ofanto	*Td*	Bx20+BY20

Ta, Triticum aestivum; Td, Triticum durum
HMW subunit compositions are based on Wrigley et al. (2006 Part II)

2.2 | DNA extraction, PCR conditions and sequencing

Genomic DNA was extracted from 10 mg of leaf tissue following the protocol indicated in D'Ovidio, Tanzarella, and Porceddu (1992). Amplification of genes encoding the HMW-GS at the *Glu-B1* locus was performed using HMW-GS gene-specific primers (Table 2). PCR was carried out on an Applied BiosystemVeriti™ 96-well thermal cycler with GoTaq® Hot Start Mastermix and GoTaq® Colorless Mastermix (Promega, WI, USA) following the manufacture's indications with slight modifications to the melting temperature (Tm; Table 3). PCR contains a final volume of 20 μl containing 1 unit of GoTaq® Hot Start polymerase (Promega, WI, USA), 1X PCR buffer, 6 pmol of each primer (Metabion International AG, Planegg/Steinkirchen, Germany), 200 mM of each dNTP and 1 μl DNA. PCR conditions were set as indicated by Lei et al. (2006), Ma et al. (2003), Schwarz, Felsenstein, and Wenzel (2004) and Xu et al. (2008). Amplified products were separated on agarose gels in Tris-acetate–EDTA buffer (1X TAE). An AccuRuler 100-bp ladder (Maestrogen Inc., Hsinchu City 30091, Taiwan) was used as a molecular size marker.

Amplicons that differed from those reported in the literature were isolated, purified and sequenced with NZYGelpure kit (NZYtech, Lisbon, Portugal) and GATC (Biotech AG, Cologne, Germany) sequencing service.

Alignments were made with DNAMAN (Lynnon Biosoft) software using the available annotated sequences (*Bx6.1*, HQ731653; *Bx7*, X13927, *Bx7*^{OE}, DQ119142, *Bx13*, EF540764; *Bx14*, AY367771;

TABLE 2 Primers used and PCR conditions

Primer designation	Name[a]	HMW-GS genes discriminated	Tm °C
PP1	ZSBy8_F5/R5*	*By8*	64
PP2	ZSBy9_F2/R2*	*By16*, *By20*	62
PP3	ZSBy9_aF1/aR3*	*By9*	59
PP4	Bx_F/Bx_R*	*Bx17*, *Bx7**	58
PP5	Bx7_F/Bx7_R*	*Bx6/Bx7*	50
PP6	CauBx752*	*Bx14*	53.5

[a]See Table 3 for detailed primer sequences. Subunit discrimination is based on the literature.

*Bx14**, KJ579439; *Bx17*, AB263219; *Bx20*, AJ437000; *Bx23*, AY553933; *Bx23**, KF995273; *By8*, AY245797; *By9*, X61026; *By15*, DQ086215; *By15**, KJ579440; *By16*, EF540765; *By18*, KF430649; *By20*, LN828972).

3 | RESULTS

Six previously developed PCR-based molecular markers for *Glu-B1* alleles were tested on a panel of 31 test cultivars with high variability in composition of HMW alleles at the *Glu-B1* locus (Table 1). These showed previously determined specificities, but revealed also new banding patterns. The new bands were validated by sequencing. For the PP3 marker, a new assignment for the By20 allele was confirmed, but for most of them it was not possible to assign any further allele specificity due to the presence of several polymorphisms shared with more than one *Glu-B1* genomic sequence.

3.1 | *By8* gene-specific primers

Primer pair 1 (PP1, Table 2) was described by Lei et al. (2006) as specific for the *By8* gene. The expected 527 bp amplicon for *By8* was observed in the test cultivar panel (Figure 1a) carrying the Bx7+By8 allele but not in any of those carrying the Bx6+By8 combination. The discrimination between the By8 and By8* alleles was also verified (Figure 1a). A possible difference in the *By8* gene sequence in genotypes carrying the Bx6+By8 or Bx7+By8 alleles could not be excluded as also previously hypothesized at the protein level by Patacchini, Masci, and Lafiandra (2001).

3.2 | *By16*- and *By20*- or null gene-specific markers

Primer pair 2 (PP2, Table 2) was developed to selectively amplify the *By16* gene by Lei et al. (2006) on the basis of the 45-bp deletion in the *By16* gene and on amplification of the repetitive domain of the *By* gene. A complex pattern would be expected as indicated in Table 3 (Ghazy, Zanouny, Moustafa, & Al-Doss, 2012; Goutam, Tiwari, Gupta, Kukreja, & Chaudhury, 2015; Lei et al., 2006; Motawei, 2008).

The specificity of PP2 for *By16* (three bands) and *By20* (no bands) was confirmed, and in addition, new banding patterns were observed (Figure 1b). A single 280-bp fragment was detected in *By18** cv. David (and confirmed in cv. Carme) as previously detected also in *triticale* (Salmanowicz & Dylewicz, 2007). The same profile was also observed in cv. Cadenza (14+15). Sequence analysis of the 280-bp amplicon retrieved in David and Cadenza did not allow assignment of the band to a specific subunit (data not shown). The sequence obtained for David (*By18**) showed 100% homology with the *By15** gene (KY579440), whereas the specific band in Cadenza (Bx14+By15) showed three SNPs in common with other *By* sequences (data not shown) and one unique SNP (A1212G). These results, together with the lack of the entire *By18** sequence in the database, did not allow any specific assignment for the molecular marker.

These results indicate that when high variability in the composition of alleles at the *Glu-B1* locus is present in the tested genotypes, the use of PP2 for the screening may result in misidentification of the genotype.

3.3 | *By9*-specific primers

Primer pair 3 (PP3, Table 2) was developed to discriminate *By9* from other *By* genes, on the basis of the 45 bp size difference in the amplicons produced (Lei et al., 2006). A 662-bp amplicon was produced specifically for the *By9* gene, while the other *By* cultivars produced a 707-bp fragment (Lei et al., 2006; Motawei, 2008).

The *By9* cv. Adriano and Cheyenne (7+9 and 7*+9, respectively) had the expected 662-bp fragment (Figure 1c). A new amplicon of 720 bp was present in cv. Ofanto and Capeiti, carrying the Bx20+By20 combination. Similar results were reported for triticale, where a 750-bp band was observed in cultivars carrying the *By20** gene (Salmanowicz & Dylewicz, 2007). Moreover, cv. Virest, Est Mottin and Sieve (6*+8*) gave a unique fragment of approximately 770 bp (Figure 1c). The sequence of the 720-bp band showed a 100% identity with the *By20* sequence (LN828972) confirming that the PP3 pair can assign the By20 subunit (data not shown). The corresponding 6*+8* sequence showed the presence of ten SNPs shared by other *By* sequences. Despite the presence of three unique SNPs in the 6*+8* cultivars, the lack of the *By8** genomic sequence in the database did not allow any specific assignment to be made for this allele. However, we tested four cultivars carrying the 6*+8* subunits and obtained the same result indicating that PP3 is also suitable for identifying the *By8** allele. This last result, however, needs further confirmation.

3.4 | *Bx17*-specific primers

Primer pair 4 (PP4, Table 2) was developed by Ma et al. (2003) to specifically distinguish between *Bx17* and non-*Bx17* genotypes, giving one or two bands, respectively (Table 3). Different fragment sizes for cultivars carrying non-*Bx17* genes were reported by other authors; in particular, two fragments of 650 and 750 bp for the *Bx7** gene and two of 670 and 770-bp for any other *Bx* gene (Butow et al., 2003; Ghazy et al., 2012).

TABLE 3 Survey of PCR-based molecular markers for the *Glu-B1* locus

Primer pair	*Glu-B1* subunit allele	Marker profile	Forward and reverse primer sequences (5′-3′)	References
Bx_F/Bx_R	*Bx17* *Bx7** *No Bx17*	1 band (669 bp) 2 bands (650, 750 bp) 2 bands (670, 770 bp)	CGC AAC AGC CAG GAC AAT T AGA GTT CTA TCA CTG CCT GGT	Ma et al. (2003)
ZSBy8_F5 ZSBy8_R5	*By8* *By8**,*By9*,*By15*,*By16*, *By18*,*By20*	1 band (527 bp) No band	TTA GCG CTA AGT GCC GTC T TTG TCC TAT TTG CTG CCC TT	Lei et al. (2006)
ZSBy9_F2 ZSBy9_R2	*By16* *By20* *By8*,*By8**,*By9*,*By15*,*By18* *By18*/*By26*	3 bands (280, 350, 400 bp) No band 2 bands (280, 350 bp) 1 band (280 bp)	GCA GTA CCC AGC TTC TCA A CCT TGT CTT GTT TGT TGC C	
ZSBy9_F7 ZSBy9_R6	*By9* *By8*,*By8**,*By15*,*By18* *By16*,*By20*,*ByNull*	2 bands 3 bands No band	TAC CCA GCT TCT CAG CAG TTG TCC CGA CTG TTG TGG	
ZSBy9_aF1 ZSBy9_aR3	*By9* *By8*,*By8**,*By15*,*By16*, *By18*,*By20*	1 band (662 bp) 1 band (707 bp)	TTC TCT GCA TCA GTC AGG A AGA GAA GCT GTG TAA TGC C	
CauBx752	*Bx14* *Bx17*	1 band (752 bp) 1 band (337 bp)	AGG GGC AGG GAA GAA ACA CT CCA GGC AAC ACA AAT CCA TG	Xu et al. (2008)
CauBx642	*Bx14* *Bx17*	1 band (642 bp) 1 band (534 bp)	GGG CAA TCG GGG TAC TTC C CCC TTG TCT TGG CTG TTG TC	
Bx7F/Bx7R	*Bx7* *Bx7**	1 band (182 bp) 1 band (164 bp)	CAA CTT CTT CAC AGC AGT CTA AAG GTG GCA AAG GCG CA	Espì, Giraldo, Rodriguez-Quijano, and Carrillo (2012)
Bx7_F/Bx7_R	*Bx6* *Bx7*,*Bx17*	1 band (250 bp) 1 band (220 bp)	CAC TGA GAT GGC TAA GCG CC GCC TTG GAC GGC ACC ACA GG	Schwarz et al. (2004)
SNP_F/SNP_R	*By18* *By8*	1 band (527 bp) No band	TTA GCG CTA AGT GCC GTC C TTG TCC TAT TTG CTG CCC TT	Liang et al. (2015)
F1/TR1	*By18* *No By18*	2 bands (284, 365 bp) No band	CAA CAA AAC GGG CGT TGT ACC ACG GTT TGC TC	
TF1/TR2	*By18* *No By18*	2 bands (284, 365 bp) No band	CAA CAA AAC GGG CGT TGT CAA CAA AAC GGG CGT TGT	
TaBAC1215C06_F517 TaBAC1215C06_R964	*Bx7* *Bx7** *Bx7oe*	No band No band 1 band (447 bp)	ACG TGT CCA AGC TTT GGT TC GAT TGG TGG GTG GAT ACA GG	Ragupathy et al. (2008)
TaBAC1215C06_F24671 TaBAC1215C06_R25515	*Bx7* *Bx7** *Bx7oe*	No band No band 1 band (884 bp)	CCA CTT CCA AGG TGG GAC TA TGC CAA CAC AAA AGA AGC TG	
Bx7F_428 Bx7R_693	*Bx7** *Bx7oe*	No band 1 band (1,116 bp)	CAA CAA CTT GTG GGG GCC TT GCG CTT AGC CAT CTC AGT GAA C	Radovanovic and Cloutier (2003)
BxFp/BxR	*Bx17* *Bx7* *Bx7oe*	5 bands 7 bands 7 bands	CAA GGG CAA CCA GGG TAC AGA GTT CTA TCA CTG CCT GGT	Butow, Gale, Ikea, Luhàsz, and Bedö (2004)
BxF_MARBxR_MAR	*Bx20* *Bx7oe* *Bx7*	1 band (800 bp) 1 band (563 bp) 1 band (520 bp)	CCT CAG CAT GCA AAC ATG CAG C CTG AAA CCT TTG GCC AGT CAT GTC	Butow et al. (2004)
Bx7_G7/Bx7_G8	*Bx6* *Bx7*	1 band (~2,500 bp) 1 band (2,373 bp)	ATG GCT AAG CGC CTG GTC CT TGC CTG GTC GAC AAT GCG TCG CTG	Anderson and Green (1989)
By8_YAN_F2 By8_YAN_R2	*By8*	1 band (~2,100 bp)	ATG GCT AAG CGG TTG GTC CT TCA CTG GCT AGC CGA CAA TG	Yan et al. (2009)

Our research confirmed previously reported specificities, as shown for cv. Francia (17+18), which possesses the expected 670-bp fragment (Figure 1d) and for Cheyenne (7*+9) with 650- and 750-bp fragments.

In addition to the conventional banding pattern, we observed a single fragment of 770 bp in cv. David (Figure 1d) and cv. Carme (data not shown), carrying the By18* allele. To confirm the hypothesis of a *By18** specificity for the PP4 primer set, the 770-bp band of

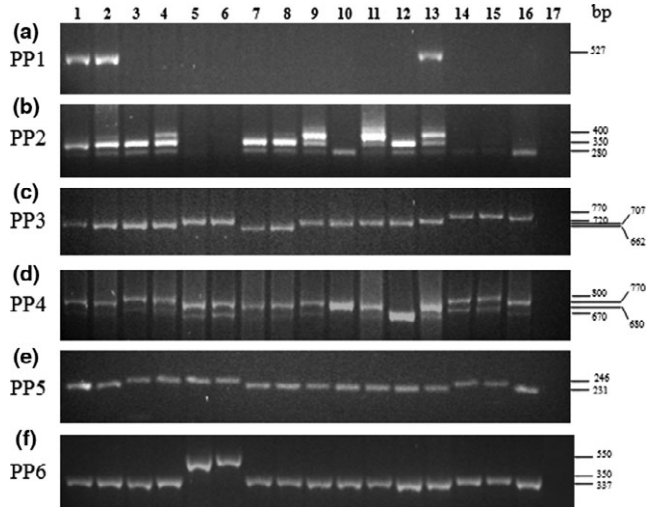

FIGURE 1 Markers profile of 16 wheat cultivars. (a): PP1 (ZSBy8_F5/R5); (b): PP2 (ZSBy9_F2/R2); (c): PP3 (ZSBy9_aF1/aR3); (d): PP4 (BxF/BxR); (e) PP5 (Bx7_F/R); (f): PP6 (CauBx752). Lanes: 1, Ciccio (Bx7+By8); 2, Chinese spring (Bx7+By8); 3, Dylan (Bx6+By8); 4, Florida (Bx6+By8); 5, Ofanto (Bx20+By20); 6, Capeiti (Bx20+By20); 7, Adriano (Bx7+By9); 8, Cheyenne (Bx7*+By9); 9, Isa (Bx13+By16); 10, David (By18*); 11, Argelato (Bx7); 12, Francia (Bx17+By18); 13, Emilio Morandi (Bx7*+By8); 14, Est Mottin (Bx6*+By8*); 15, Sieve (Bx6*+By8*); 16, Cadenza (Bx14+By15); 17, negative control

David and Carme (*By18**) was purified, sequenced and aligned with a selection of *Bx* and *By* annotated alleles (see Materials and Methods) showing a 100% sequence identity with the *Bx23** and *Bx14** genes (ID KF995273 and KJ579439, data not shown). However, lack of an annotated genomic *By18** sequence does not allow any hypothesis of specificity for the By18*, Bx14* and Bx23 alleles to be formulated using this primer pair.

Nevertheless, the same 770-bp band was confirmed in cv. David, Carme and Garibaldino that carry *By18**, suggesting a possible specificity for this allele.

3.5 | *Bx6*-specific primers

The panel of test cultivars was also screened with the co-dominant primer pair 5 (PP5, Table 2) specific for the *Bx6* gene, developed by Schwarz et al. (2004) on the basis of a 15-bp insertion observed for the Bx6 allele and absent in all other Bx alleles. They reported that the PP4 profile consists of two bands of 246 bp for *Bx6* cultivars and 231 bp for non-*Bx6* cultivars (Figure 1e).

In our experiments, the tested cultivars confirmed the high specificity of this primer pair for discriminating the *Bx6* gene. Moreover, a band of 246 bp was also observed for *Bx20* (Ofanto and Capeiti) and *Bx6** (Virest, Est Mottin and Sieve) genotypes, which, to our knowledge, was not previously reported (Figure 1e). These results are consistent with the presence of the 15-bp insertion also in the *Bx20* mRNA sequence (AJ437000), but as the Bx6* sequence is so far not available in databases, the hypothesis of a specific additional assignment for PP5 cannot be formulated.

3.6 | *Bx14*- and *Bx17*-specific primers

Xu et al. (2008) developed the primer pair 6 (PP6, Table 2) to distinguish between *Bx14* and *Bx17*. A 752-bp fragment for *Bx14* and 337 bp for *Bx17* were expected from analyses of the test cultivars. Using the indicated conditions on all cultivars, only the *Bx17* band of 337 bp was detected, including cv. Cadenza (Bx14+By15, Figure 1f). Although not previously reported, PP6 showed a single 550-bp band when applied to six *Bx20* cultivars (Figure 1f). Analyses of the 550-bp band showed a 100% identity with a portion of the promoter sequence of the *Bx14* allele (AY367771) that might be shared also by the *Bx20* gene for which the promoter sequence has not been deposited. However, the same specificity was not observed in *Bx14* genotypes and did not lead to any hypothesis of further specificity for the PP6 marker, although results highlighted that a large number of cultivars should be considered when using this molecular marker to avoid allele misidentification.

4 | DISCUSSION

Until now, the composition of the *Glu-B1* locus in many commercial cultivars was assigned by the SDS-PAGE method that frequently results in an incorrect interpretation of different alleles due to similar mobilities on SDS gels (Gianibelli et al., 2001). However, the consistency of this approach, which relies on protein sequencing, has rarely been tested. The possibility of a mismatch between the assignation of a specific subunit to a genotype on the basis of SDS-PAGE mobility and through DNA sequence analysis should be considered. Thus, the apparent molecular weights from SDS agarose gel electrophoresis are not themselves sufficient to make conclusions on the identities of two HMWs; DNA or protein sequence is also required (Anderson & Green, 1989). DNA molecular markers (MMs) have been of great use for plant breeding, and marker-assisted selection (MAS) is one of the key features in gene identification and selection of new genotypes.

The case of seed proteins is particularly intriguing genetically because the phenotype (of a specific seed mutant) is also detected at the molecular level. Nevertheless, the advantage of PCR-based analyses over SDS-PAGE for the identification of cultivars with a particular HMW allelic combination has been reported (Ahmad, 2000; Lei et al., 2006; Rasheed et al., 2014; Xu et al., 2008). MAS can be applied at an early developmental stage; the results are independent of the environment and are suitable for high-throughput testing (Lei et al., 2006).

In recent years, several HMW genes at the *Glu-B1* locus have been deposited and annotated and analyses of alignment within x and y types reveal a high level of DNA or mRNA sequence similarity between them, thus explaining the complex pattern obtained in the application of the *Glu-B1* molecular markers.

By sequencing the amplicons, it was possible to add a new assignment for the ZSBy9_aF1/aR3 marker for the By20 allele, but for all the other markers, the observed polymorphisms were shared with other HMW-GS genes, so it was not possible to hypothesize

any further specificity. PCR-based molecular markers represent an efficient and fast tool to overcome some of the drawbacks in protein-based methods for HMW-GS allele resolution (Rasheed et al., 2014), and different results might be achieved especially when high variability at the *Glu-B1* locus is expected in the cultivars or populations tested.

An example of the complexity of the link between molecular marker banding pattern and the allele assignment is reported for the PP4 marker for which the 770-bp band obtained in two *By18** cultivars has been sequenced and analysed. Sequence analyses revealed a 100% sequence similarity with the *Bx14** and *Bx23** alleles that share two specific substitutions (A1731G and G1736A) with *By18**. The absence of deposited *By18** DNA sequence, however, did not allow specificity to be assigned and only suggests that *By18** also shares the same SNPs. However, the use of this genomic tool to detect desired traits in wheat populations or in elite cultivars is widely used in MAS in conventional breeding and also represents one of the main pillars in more modern approaches such as genomic selection (Bassi, Bentley, Charmet, Ortiz, & Crossa, 2016).

Our results highlight that genomic and proteomic tools are complementary to overcome allele misidentification when possible. If this is not possible, then a suitably large number of control cultivars should be used to confirm the specificity of markers on a case-by-case basis.

However, clear indications of reproducible banding patterns not previously reported and obtained using nominated molecular markers have been shown in the current work. In addition, a new *By20* specificity for the ZSBy9_aF1/aR3 primer set developed by Lei et al. (2006) was demonstrated by sequencing.

This is of particular importance in analysis of the glutenin composition in durum wheat landrace collections, where the Bx20+By20 alleles are often very frequent (Bellil, Hamdi, & Khelifi, 2014; Moragues, Zarco-Hernandez, Moralejo, & Royo, 2006). Several of these markers produced a specific band for Bx20+By20, suggesting that the coding sequence of this gene may share several haplotypes. The identification of haplotypes is useful for the identification of regions of the genome associated with traits of interest or candidate genes but also for targeting the development of specific molecular markers for MAS (Varshney, Hoisington, & Tyagi, 2006).

The present contribution, in addition to reviewing the molecular markers developed for HMW-GS allelic composition at the *Glu-B1* locus, suggests additional putative specificity for *Glu-B1* alleles among the molecular markers tested.

ACKNOWLEDGEMENT

The authors thank CREA-SCV Sant'Angelo Lodigiano, University of Tuscia-Viterbo, Rothamsted Research Experimental Station, UK, and CREA-Foggia for provision of durum and bread wheat seeds.

CONFLICT OF INTEREST

No conflict of interest exits in the submission of this manuscript, and the manuscript is approved by all authors for publication.

REFERENCES

Ahmad, M. (2000). Molecular marker-assisted selection of HMW glutenin alleles related to wheat bread quality by PCR-generated DNA markers. *Theoretical and Applied Genetics*, *101*, 892–896.

Anderson, O. D., & Green, F. C. (1989). The characterization and comparative analysis of high-molecular-weight glutenin genes from genomes A and B of a hexaploid bread wheat. *Theoretical and Applied Genetics*, *77*, 689–700.

Bassi, F. M., Bentley, A. R., Charmet, G., Ortiz, R., & Crossa, J. (2016). Breeding schemes for the implementation of genomic selection in wheat (*Triticum* spp.). *Plant Science*, *242*, 23–36.

Bellil, I., Hamdi, O., & Khelifi, D. (2014). Diversity of five glutenin loci within durum wheat (*Triticum turgidum* L. spp. durum (Desf.) Husn.) germplasm grown in Algeria. *Plant Breeding*, *133*, 179–183.

Branlard, G., & Dardevet, M. (1985). Diversity of grain protein and bread wheat quality: II. Correlation between molecular weight subunits of glutenin and flour quality characteristics. *Journal of Cereal Science*, *3*, 345–354.

Butow, B. J., Gale, K. R., Ikea, J., Luhàsz, A., & Bedö, Z. (2004). Dissemination of the high expressed Bx7 glutenin subunit (*Glu-B1al* allele) in wheat as revealed by novel PCR markers and RP-HPLC. *Theoretical and Applied Genetics*, *109*, 1525–1535.

Butow, B. J., Ma, W., Gale, K. R., Cornish, G. B., Rampling, L., Larroque, O., … Bekes, F. (2003). Molecular discrimination of Bx7 alleles demonstrates that over expression has a major impact on wheat flour dough strength. *Theoretical and Applied Genetics*, *107*, 1524–1532.

D'Ovidio, R., Masci, S., & Porceddu, E. (1995). Development of a set of oligonucleotide primers specific for genes at the Glu-1 complex loci of wheat. *Theoretical and Applied Genetics.*, *91*, 189–194.

D'Ovidio, R., Tanzarella, O. A., & Porceddu, E. (1992). Isolation of an alpha- type gliadin gene from *Triticum durum* Desf and genetic polymorphism at the Gli-2 loci. *Journal of Genetics and Breeding*, *46*, 41–48.

Espì, A., Giraldo, P., Rodriguez-Quijano, M., & Carrillo, J. M. (2012). A PCR-based method for discriminating between high molecular weight glutenin subunits Bx7 and Bx7* in *Triticum aestivum* L. *Plant Breeding*, *131*, 571–573.

Ghazy, A. I., Zanouny, A. I., Moustafa, K. A., & Al-Doss, A. A. (2012). Molecular screening of high molecular weight glutenin genes in spring bread wheat genotypes in Saudi Arabia. *Journal of Food, Agriculture and Environment*, *10*, 157–161.

Gianibelli, M. C., Larroque, O. R., Mac Ritchie, F., & Wrigley, C. W. (2001). Biochemical, genetic, and molecular characterization of wheat glutenin and its component subunits. *Cereal Chemistry*, *78*, 635–646.

Goutam, U., Tiwari, R., Gupta, R. K., Kukreja, S., & Chaudhury, A. (2015). Allelic variations of functional markers for high molecular weight glutenin genes in Indian wheat (*Triticum aestivum* L.) cultivars and their correlation with bread loaf volume. *Indian Journal of Plant Physiology*, *20*, 97–102.

Lafiandra, D., Sanguineti, M. C., Maccaferri, M., & Deambrogio, E. (2007). Molecular markers and QTL analysis for grain quality improvement in wheat. In R. K. Varshney & R. Tuberosa (Eds.), *Genomics-assisted crop improvement*. Vol. 2. Genomics Applications in Crops (pp. 25–50).Dordrecht: Springer.

Lafiandra, D., Tucci, G. F., Pavoni, A., Turchetta, T., & Margiotta, B. (1997). PCR analysis of x-and y-type genes present at the complex Glu-A1 locus in durum and bread wheat. *Theoretical and Applied Genetics*, *94*, 235–240.

Lei, Z. S., Gale, K. R., He, Z. H., Gianibelli, C., Larroque, O., Xia, X. C., … Ma, W. (2006). Y-type gene specific markers for enhanced discrimination of high-molecular weight glutenin alleles at the Glu-B1 locus in hexaploid wheat. *Journal of Cereal Science*, *43*, 94–101.

Liang, X., Zhen, S., Han, C., Wang, C., Li, X., Ma, W., & Yan, Y. (2015). Molecular characterization and marker development for hexaploid

wheat-specific HMW glutenin subunit 1By18 gene. *Molecular Breeding*, *35*, 1–16.

Ma, W., Zhang, W., & Gale, K. R. (2003). Multiplex-PCR typing of high molecular weight glutenin alleles in wheat. *Euphytica*, *134*, 51–60.

Moragues, M., Zarco-Hernandez, J., Moralejo, M. A., & Royo, C. (2006). Genetic diversity of glutenin protein subunits composition in durum wheat landraces [*Triticum turgidum* ssp. turgidum convar. durum (Desf.) MacKey] from the Mediterranean basin. *Genetic Resources and Crop Evolution*, *53*, 993–1002.

Motawei, M. I. (2008). Molecular weight glutenin alleles at *Glu-B1* locus. *G3*, *2*, 53–56.

Patacchini, C., Masci, S., & Lafiandra, D. (2001). New method to purify y-type high-molecular weight glutenin subunits from wheat. *Journal of Cereal Science*, *34*, 135–139.

Payne, P. I., & Lawrence, G. J. (1983). Catalogue of alleles for the complex loci Glu-A1, Glu-B1 and Glu-D1 which coded for high-molecular-weight subunits of glutenin in hexaploid wheat. *Cereal Research Communications*, *11*, 29–35.

Radovanovic, N., & Cloutier, S. (2003). Gene-assisted selection for high molecular weight glutenin subunits in wheat doubled haploid breeding programs. *Molecular Breeding*, *12*, 51–59.

Ragupathy, R., Naeem, H. A., Reimer, E., Lukow, O. M., Sapirstein, H. D., & Cloutier, S. (2008). Evolutionary origin of the segmental duplication encompassing the wheat *GLU-B1* locus encoding the overexpressed Bx7 ($Bx7^{OE}$) high molecular weight glutenin subunit. *Theoretical and Applied Genetics*, *116*, 283–296.

Rasheed, A., Xia, X., Yan, Y., Appels, R., Mahmood, T., & He, Z. (2014). Wheat seed storage proteins: Advance in molecular genetics, diversity and breeding applications. *Journal of Cereal Science*, *60*, 11–24.

Salmanowicz, B. P., & Dylewicz, M. (2007). Identification and characterization of high-molecular-weight glutenin genes in Polish triticale cultivars by PCR-based DNA markers. *Journal of Applied Genetics*, *48*, 347–357.

Schwarz, G., Felsenstein, F. G., & Wenzel, G. (2004). Development and validation of a PCR-based marker assay for negative selection of the HMW glutenin allele Glu-B1-1d (Bx-6) in wheat. *Theoretical and Applied Genetics*, *109*, 1064–1069.

Shewry, P. R., Halford, N. G., & Tatham, A. S. (1992). High molecular-weight subunits of wheat glutenin. *Journal of Cereal Science*, *15*, 105–120.

Varshney, R. K., Hoisington, D. A., & Tyagi, A. K. (2006). Advances in cereal genomics and applications in crop breeding. *Trends in Biotechnology*, *24*, 490–499.

Wrigley, C. W., Bekes, F., Cavanagh, C. R., & Bushuk, W. (2006). *The gluten composition of wheat cultivars and genotypes, PART II. Composition table for the HMW subunits of glutenin*. [electronic publication] http://www.aaccnet.org/grainbin/gluten_gliadin.asp. AACC International. www.accnet.org

Xu, Q., Xu, J., Liu, C. L., Chang, C., Wang, C. P., You, M. S., … Liu, G. T. (2008). PCR-based markers for identification of HMW-GS at *Glu-B1x* loci in common wheat. *Journal of Cereal Science*, *47*, 394–398.

Yan, Y. M., Jiang, Y., An, X. N., Pei, Y. H., Li, X. H., Zhang, Y. Z., … Ma, W. (2009). Cloning, expression and functional analysis of HMW glutenin subunit *1By8* gene from Italy pasta wheat (*Triticum turgidum* L. ssp. *durum*). *Journal of Cereal Science*, *50*, 398–406.

How to cite this article: Janni M, Cadonici S, Pignone D, Marmiroli N. Survey and new insights in the application of PCR-based molecular markers for identification of HMW-GS at the *Glu-B1* locus in durum and bread wheat. *Plant Breed*. 2017;136:467–473. https://doi.org/10.1111/pbr.12506

Gene-ecology of durum wheat HMW glutenin reflects their diffusion from the center of origin

Janni M.[1,2*&], Cadonici S.[3&], Bonas U. [3], Grasso A.[4], Dahab A. A. D.[4], Visioli G.[3], Pignone D.[1], Ceriotti A.[4], Marmiroli N.[3,5]

[1]Institute of Bioscience and Bioresources (IBBR), National Research Council (CNR), Via Amendola 165/A, 70126 Bari, Italy

[2]Institute of Materials for Electronics and Magnetism (IMEM), National Research Council (CNR), Parco Area delle Scienze 37/A 43124 Parma

[3]Department of Chemistry, Life Sciences and Environmental Sustainability, University of Parma, Parco Area delle Scienze, 11/A, 43124 Parma, Italy

[4]National Research Council (CNR), Institute of Agricultural Biology and Biotechnology (IBBA), Via Bassini 15, 20133 Milano, Italy

[5]Regione Emilia-Romagna (IT) SITEIA. PARMA Technopole

*Corresponding author michela.janni@ibbr.cnr.it

[&]These authors equally contributed to the work

Key words: wheat germplasm, SDS-PAGE, Lab-on-a-chip, wheat domestication, migration routes

Running title: HMW-GS variation in durum wheat landraces

Abstract

The production of many food items processed from wheat grain relies on the use of high gluten strength flours. As a result, about 80% of the allelic variability in the genes encoding the glutenin proteins has been lost in the shift from landraces to modern cultivars. Here, the allelic variability in the genes encoding the high molecular weight glutenin subunits (HMW-GSs) has been characterized in 152 durum wheat lines developed from a set of landraces. The allelic composition at the two *Glu-1* loci (*Glu-A1* and *-B1*) was obtained at both the protein and the DNA level. The former locus was represented by three alleles, of which the null allele *Glu-A1c* was the most common. The *Glu-B1* locus was more variable, with nine alleles represented, of which *Glu-B1b* (HMW-GSs 7+8), *-B1d* (6+8) and *-B1e* (20+20) were the most frequently occurring. The composition of HMW-GSs has been used to make inferences regarding the diffusion and diversification of durum wheat. The relationships of these allelic frequencies with their geographical distribution within the Mediterranean basin is discussed in terms of gene-ecology.