Dipartimento di Scienze Agrarie e Forestali (DAFNE)

Ph.D. course in Plant and Animal Science

XXXI cycle

CLASSICAL AND INNOVATIVE APPROACHES TO STUDY PROTEINS INVOLVED IN ADVERSE REACTION TO WHEAT AND GRAIN QUALITY

AGR-07

PhD Program Coordinator: Prof. Stefania Masci

Ph.D. student: Francesco Camerlengo

Tutors: Prof. Stefania Masci

Prof. Domenico Lafiandra

Ph.D. Francesco Sestili

Academic year 2017-2018

CHAPTER 1:

GENERAL INTRODUCTION

1.1 WHEAT

Wheat is the most cultivated food crop with a harvest area of about 220 million hectares and a world trade greater than all other crops combined (Fig. 1.1).

Nowadays China is the largest producer of wheat in the world followed by India, Russia, United States, Canada and France (http://www.fao.org/faostat/en/#data) (Fig.1.3). In Europe, France is the leading producer right before Ukraine and Germany, whereas Italy is the 8th with 6.966.465 tonnes produced in 2017 (FAOSTAT) (Fig. 1.2). The productivity of wheat ranges from few quintals to more than 10 tons per hectare averaging nearly 3-4 tons per hectare. The wide range of yields is tightly linked to environmental condition, crop management and genetic background.

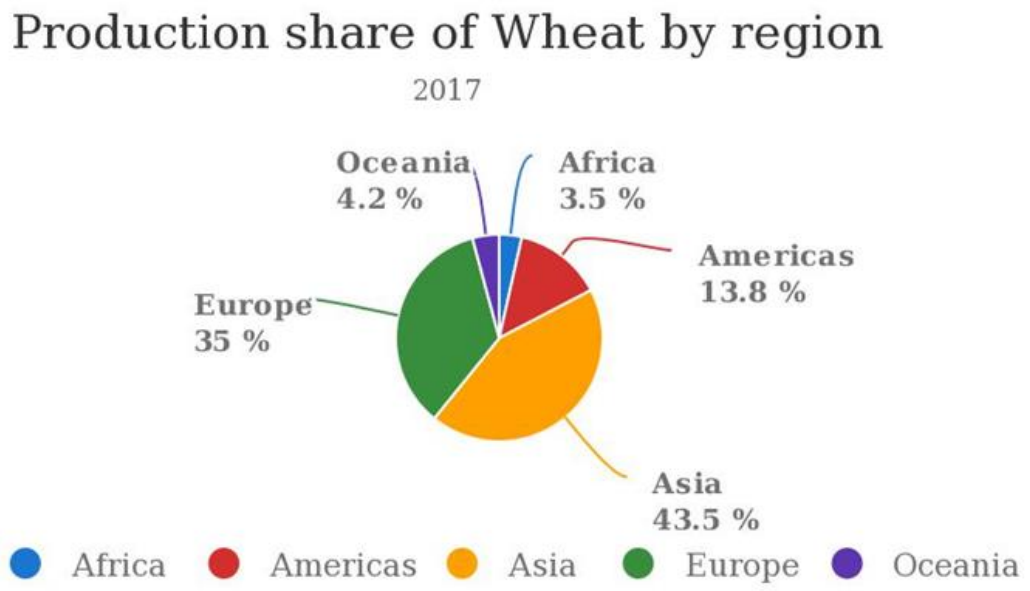

Figure 1.1: Production share of wheat by region *(FAOSTAT data 2017)*

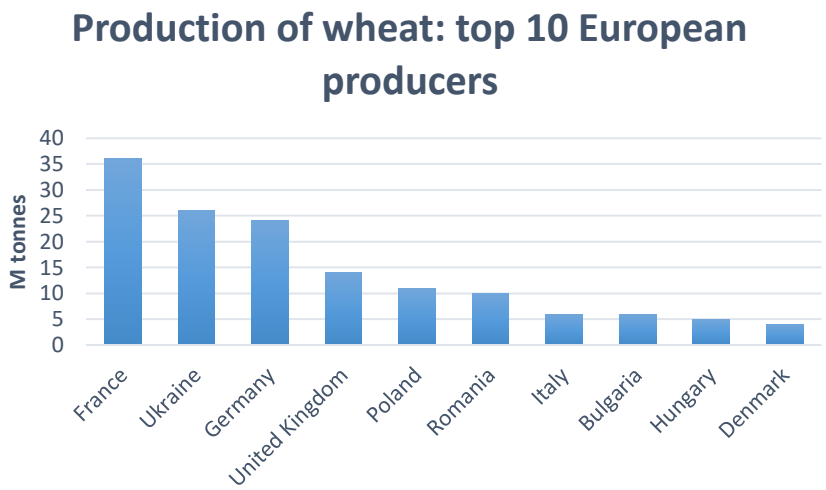

Figure 1.2: Production of wheat in M tonnes: top 10 European producers *(FAOSTAT data 2017)*

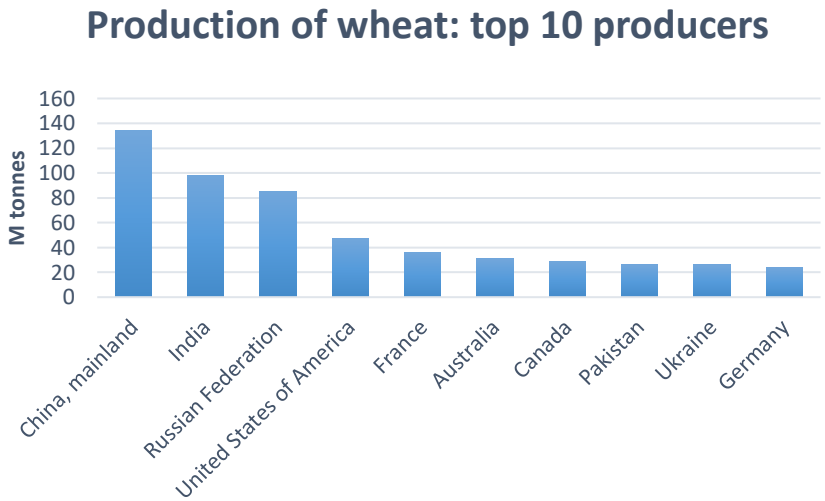

Figure 1.3: Production of wheat in M tonnes: top 10 producers in the world *(FAOSTAT data 2017)*

Wheat and derived foods represent the third food supply for human consumption and the main source of protein, as well as an important source of carbohydrates, dietary fibre and micronutrients, providing 20 % of the global caloric requirements (Figs. 1.4 and 1.5).

Wheat kernels represent a worldwide staple food, processed in a variety of end products, most of which typical of the production area of origin. The dough obtained by the grinding of kernels can be processed in bread, pasta, noodles, baked foods and other products, thanks to the viscoelastic proprieties conferred by gluten proteins.

On the other hand, different pathologies are associated with wheat and derived foods consumption. The majority of wheat related disorders had an increasingly higher incidence in human population

in the last 50 years and many studies have indicated that the protein fraction of kernels is the main cause, although other components seem to be related to specific pathologies.

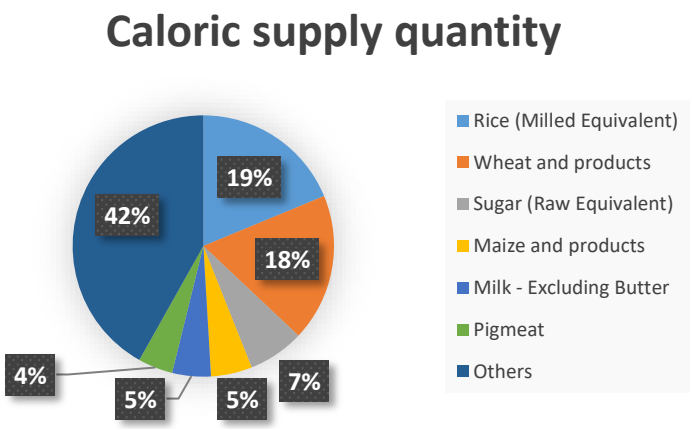

Figure 1.4: Caloric supply quantity. *Kcal//capita/day (FAOSTAT data 2017)*

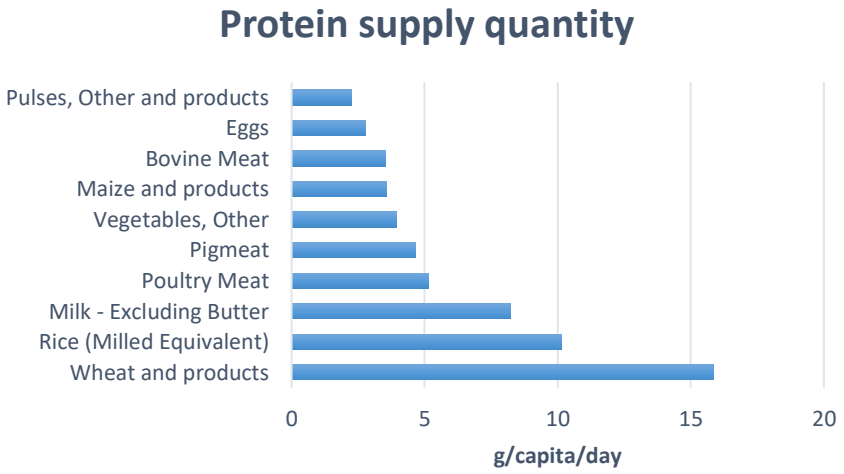

Figure 1.5: Protein supply quantity. *g/capita/day (FAOSTAT data 2017)*

The future scenario, towards which research, agriculture and food technology should be addressed, includes not only crop improvement for higher yields, biotic and abiotic stress resilience, but also the increase of nutritional and healthy value of wheat with a lower environmental impact.

1.2 CULTIVATION AND DIFFUSION

The term wheat is often referred to different crop species of the genus *Triticum*, *Poaceae* family with different levels of ploidy. Its cultivation has spread all over the world since the dawn of agriculture. About 10.000 YA, form the “cradle of agriculture” in the Fertile Crescent (in the

Middle East along Tigris and Euphrates rivers, across today's Syria from Lebanon up to Iran) wheat spread towards Europe and Asia (8.000 – 8500 BP) and then towards Egypt (<8000 BP) under pressure of human domestication of the major staple crops (Fig. 1.6). Wheat arrived in Italy, Sothern France and Spain ca. 7000 BP. Its cultivation was introduced to America in 1529 and to Australia in 1788 (Feldman 2001).

Wheat cultivated species genomes is organised into seven chromosome (1x = 7) and consist of diploids (2n = 2x = 14), tetraploids (2n = 4x = 28) and hexaploids (2n = 6x = 42). The only diploid wheat currently cultivated is einkorn (*T. monococcum* ssp. *monococcum*), but it is poorly grown in parts of Mediterranean basin (Italy and Turkey) mostly for animal feed, although recently it is gaining attention also for human consumption. It is the less cultivated species due to lower yield and high managing costs.

Emmer (*T. turgidum* ssp. *dicoccum*) belong to tetraploid wheats, it was domesticated at the same time of einkorn and today is poorly cultivated in the Middle East and south Asia and in specific Italian regions. The most cultivated tetraploid is durum wheat (*T. turgidum* ssp. *durum*) and accounts for the five % of the total wheat crop. It is grown in the U.S.A., Canada, Russia, India, Italy and some parts of Mediterranean basin. Besides, durum and emmer, other tetraploid wheat subspecies are cultivated, such as *T. turgidum* ssp *turgidum*, *turanicum* and *polonicum*.

Ninety-five % of the total wheat crop is represented by hexaploid bread wheat (T. *aestivum* ssp. *aestivum*), cultivated all over the world. Another hexaploid crop, though somewhat less common than bread wheat, is spelt or hulled wheat (*T. aestivum* ssp. *spelta*), cultivated in Central Europe.

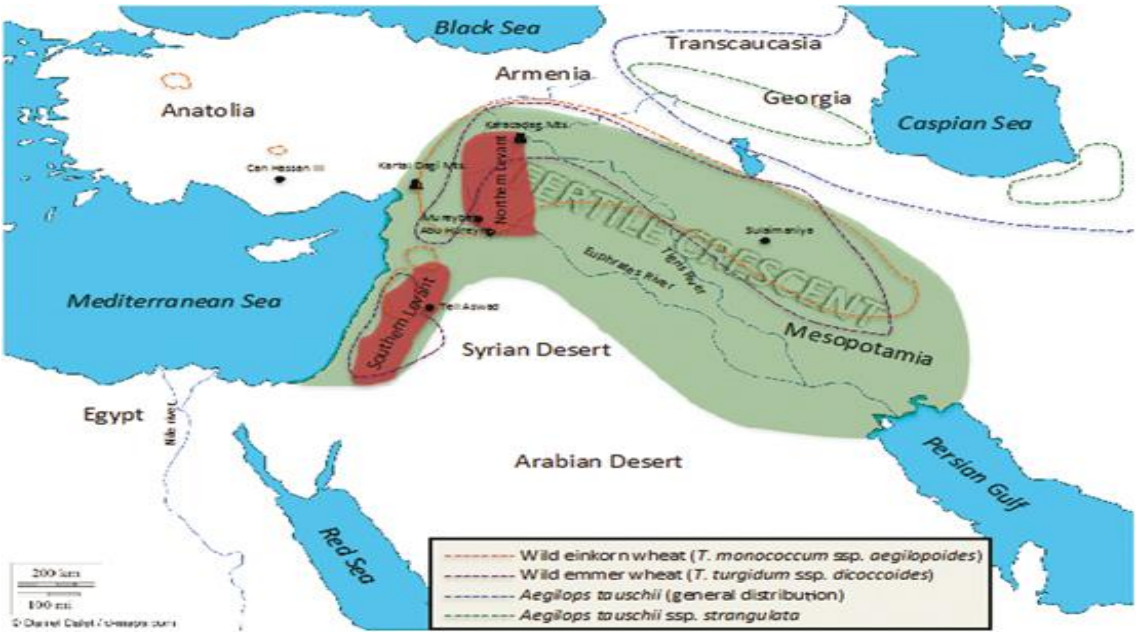

Figure 1.6: Distribution of wild wheats across the Fertile Crescent *(green) (Faris et al 2014)*

Nowadays, durum and bread wheats represent the almost total production of wheat in the world, replacing diploids and early tetraploids wheats. The revolutionary and evolutionary changes contributed to genomic stabilization of polyploids and increased genetic variability, in terms of adaptability, fitness, competitiveness, and colonizing ability (Feldman and Levy 2009).

The dynamic plasticity of the polyploid genome promoted the large diffusion of durum and bread wheat on global scale and conferred the main properties, enabling the realization of a variety of end products, mostly pasta from durum wheat semolina and bread, biscuits and baked foods from bread wheat.

1.2.1 Origin and Evolution

About three Million BP, a common seven-chromosome ancestor gave rise to the diploid progenitors and relatives of modern wheat: the *Triticum* and *Aegilops* taxa.

The A-genome diploids, *T. monococcum* ssp. *aegilopoides* (2n = 2x = 14, AA) and *T. urartu* (2n = 2x = 14, AA), arose from the *Triticum* progenitor, whereas S-genome diploids, *Aegilops Sitopsis* section, and D-genome diploids, *A. tauschii* (2n = 2x = 14, DD), arose from *Aegilops* progenitor.

The first domesticated form is the einkorn diploid wheat (*T. monococcum* ssp. *monococcum*, 2n = 2x = 14, A^mA^m), originated from *T. monococcum* ssp. *aegilopoides* in the northern and eastern part of the Fertile Crescent around 10.000 BP. Soon after, the hybridization of *T. urartu* and a close relative of *Ae. speltoides* (2n = 2x = 14, SS), followed by chromosome doubling, led to the formation of two wild emmer tetraploid wheats, *T. turgidum* ssp. *dicoccoides* about 0.5 MYA, from which the domesticated emmer *T. turgidum* ssp. *dicoccum* (2n = 4x = 28, AA BB) and *T. timopheevii* ssp. *araraticum* (2n = 4x = 28, AA GG) took place, about 9.000-9.500 BP (Feldman 2001, Faris 2014) (Fig. 1.8).

The hybridization of domesticated *T. timopheevii* ssp. *timopheevii* and einkorn wheat gave rise to the hexaploid wheat *T. zhukovskyi* (2n = 6x = 42, A^mA^m AA GG).

The domesticated emmer *T. turgidum* ssp. *dicoccum* is considered the progenitor of modern durum and bread wheat cultivars. Indeed, it was largely cultivated for several thousand years and spread from Fertile Crescent to Anatolia, Mediterranean basin and Europe about 8.000 BP. Around 6.000 BP it was the dominant crop in Egypt, Asia and India, later gradually replaced by durum wheat (Feldman et al. 2001). Durum wheat *T. turgidum* ssp *durum* (2n = 4x = 28, AA BB) formed along with spp. *parvicoccum* around 7.000 BP and became a prominent crop between 3.000 – 2.000 BP.

Other tetraploid wheats originated from domesticated emmer at a later stage and include subspecies *turgidum, turanicum, polonicum* and *carthlicum*.

The domesticated emmer subspecies *T. turgidum* ssp. *dicoccum* led to the formation of hexaploid wheats *T. aestivum* spp. *spelta* and T. *aestivum* spp. *aestivum* (2n = 6x = 42, AA BB DD), hybridizing with *Ae. tauschii* about 8.000 YA (Kihara 1944).

It is certain that *Ae. tauschii* donated the D genome to hexaploid wheats (McFadden and Sears 1946) and most probably the subspecies *strangulata (*Shen et al 2018*)*. The hybridization took place in the area of Iran southeast of the Caspian Sea about 8.000 BP, mostly like giving rise to Asian spelta. Rapidly, spp. *aestivum* evolved and spread to Egypt, Europe and Asia about 6.000 BP. Otherwise, Marcussen and colleagues (2014) supported a model of hybridization between the A and B genomes, giving rise to the D genome ancestor. They estimated that the divergence of the A and B genomes from a common ancestor have occurred almost 7 M BP, and homoploidization occurred between them about 5–6 M BP, giving rise to the D genome (Fig. 1.8).

Other hexaploid wheats originated from *T. aestivum* spp. *spelta* (Asian-like) including subspecies *sphaerococcum*, *macha*, *compactum* and European *spelta.* Much evidence demonstrate that European *spelta* have a separate origin from Asian *spelta*, formed from a cross between *T. aestivum* spp. *compactum*, and domesticate emmer *T. turgidum* ssp. *dicoccum* (Faris 2014).

Different events of polyploidization occurred during the evolution of wild and cultivated wheat, involving multiple sites of the Fertile Crescent and spanned over a long time (Fig. 1.7). Early farmers grew domesticated emmer in mixtures with wild emmer for a long period and hybrids between wild and domesticate tetraploid were very frequent, that led to gene flow increasing genetic variability, formation of subpopulations and reduction of the founder effect. Much less hybridization events occurred between *T. turgidum* and *Ae. tauschii* and between *T. aestivum* and other tetraploids resulting in a larger magnitude of genetic drift (Feldman 2001).

Therefore, wheat domestication was a very slow process, taking place along with allopolyploidization.

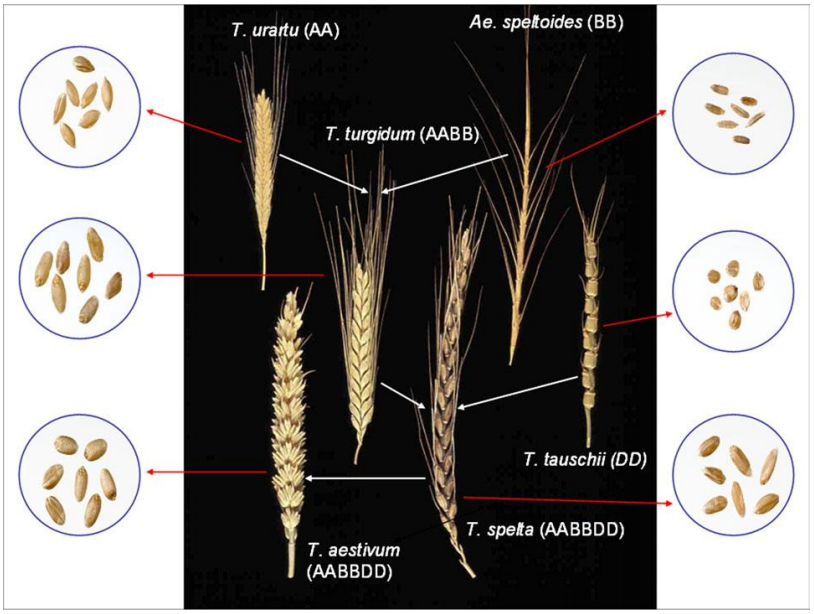

Figure 1.7. ***The evolutionary and genome relationships between cultivated bread and durum wheats and related wild diploid grasses*** *(Shewry 2009)*

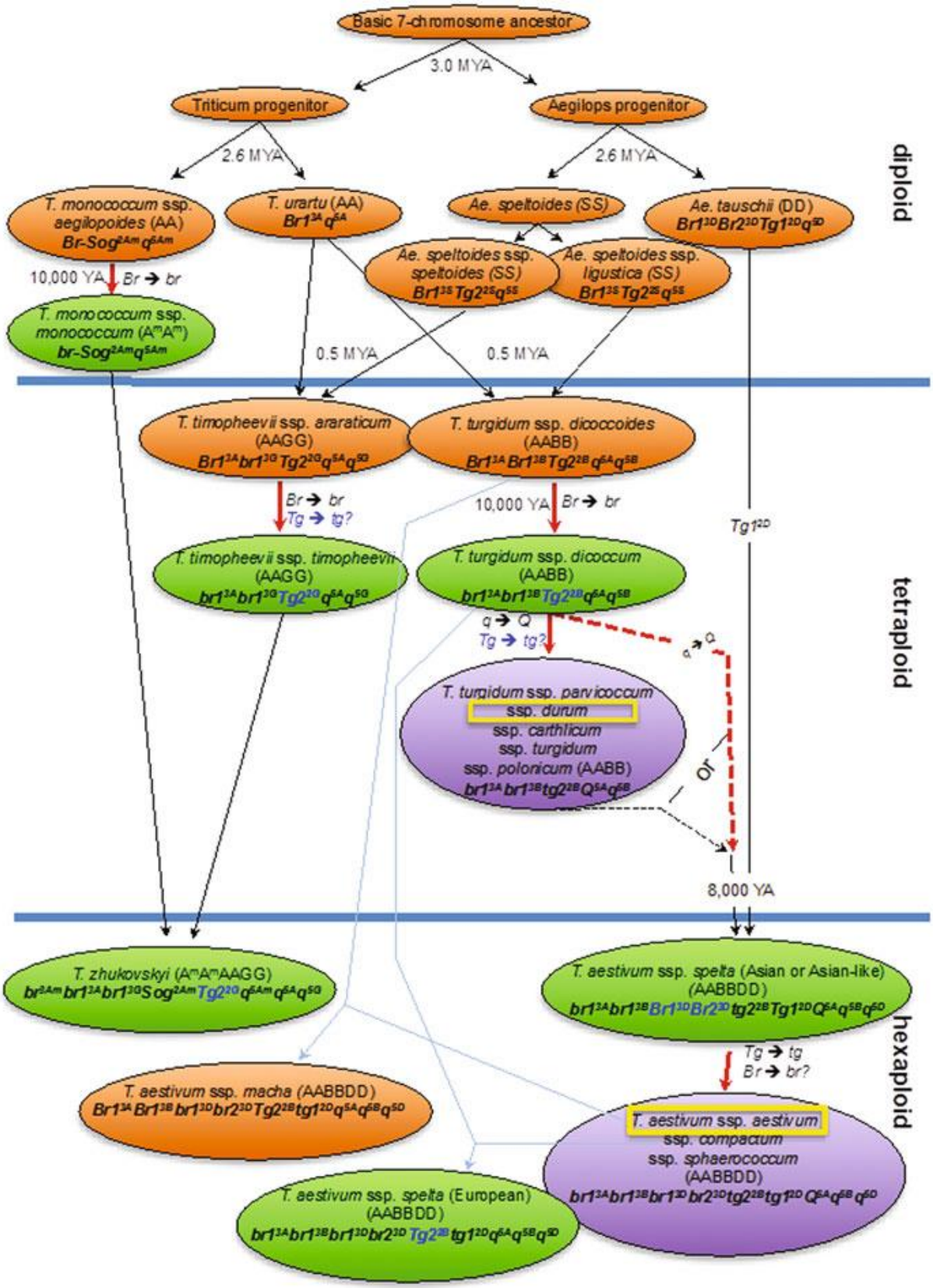

Figure 1.8. ***The evolutionary lineages involving Triticum wheat species*** *(Faris et al 2014).*

1.3 WHEAT BREEDING

The real birth of wheat breeding took place about 10.000 BP, when the gatherer and early farmers started to select plants with interesting traits (Fig. 1.9). Three major genes contributed to the domestication of wheat: *Br* confers brittle rachis, *Tg* confers tenacious glume and *q* confers non free-threshing character. Over thousand years, the early farmers consciously selected the non-brittle rachis, soft glume and free-threshing mutants in the wild wheat population. Indeed, mutations at $Br1^{3A}$ and $Br1^{3B}$ loci occurred during the domestication of *T. monococcum* ssp. *monococcum* and *T. turgidum* ssp. *dicoccum* and led to the loss of seed shattering. The free-threshing character was conferred by mutations at *Tg* and *Q* factor loci. In fact, wheat plants with soft glumes carried mutation at $Tg1^{2D}$ and $Tg2^{2B}$, but they are non-free-threshing. A mutation in the q^{5A} allele led to formation of partially dominant Q^{5A} allele along with $tg1^{2D}$ and $tg2^{2B}$ alleles are necessary to confer the free-threshing character.

The early farmers selected other characters such as grain size, non-dormant seeds, uniform germination, yield, more grains per spike etc. (Feldman 2001). Furthermore, domesticated crop acquired alterations in flowering time, growth habit and other essential traits for wide adaptation to different environments during the spread of wheat from Fertile Crescent to Europe and Asia (Peng et al 2011).

Selection for interesting traits occurred in fields, starting from different genotypes and different species of different ploidy levels over a long period (Feldman 2001). Such sort of selection provided protection against biotic and abiotic stresses ensuring yield stability and increasing genetic variability. The competitive environment gave a selective advantage to taller plants with more tillers and horizontal leaves and gave rise to landraces.

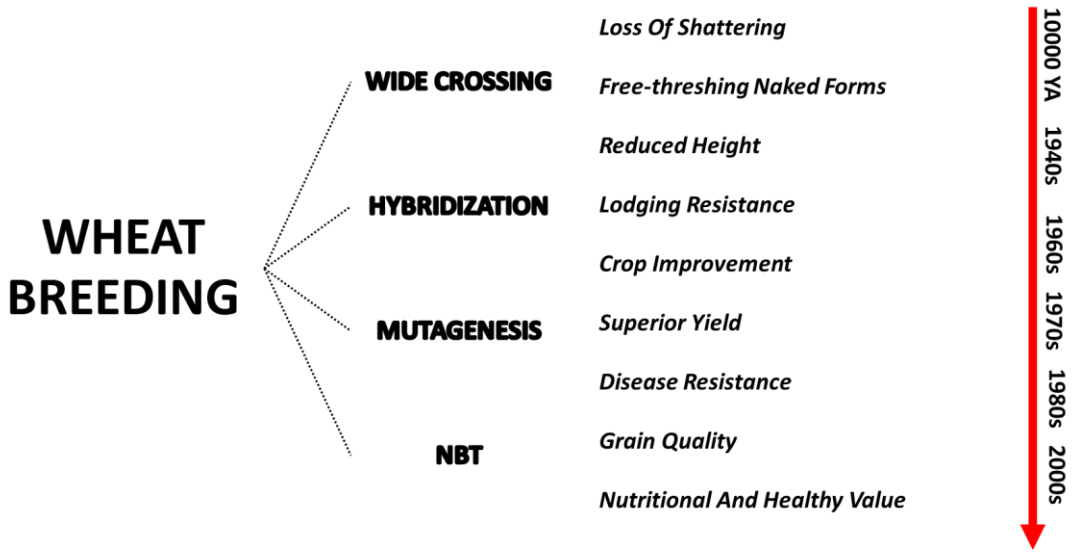

Figure 1.9: schematic representation of breeding techniques and improved traits in wheat.

1.3.1 Conventional Breeding

When the arable land became a limiting factor, farmers selected only for individual genotypes with higher number of grains per unit area creating monotypic crop fields (Fig. 1.9).

Towards the end of XIX century, modern breeding practise for wheat improvement took place in Europe and USA. Increased yield, reduced height, lodging resistance, abiotic and biotic resistance followed by enhanced response to agronomical practise characterised wheat breeding programs since the turn of the last century. The most important achievement of wheat breeding was the introduction of dwarf and semi dwarf genes in wheat world varieties that led to the Green Revolution. In 1935, the variety Norin 10, obtained through crossing of a semi-dwarf Japanese wheat landrace with two American varieties, was released in Japan. The improved semi-dwarf variety Norin 10 contains the *Rht1 (Rht-B1b)* and *Rht2 (Rht-D1b*) genes conferring reduced height to 60-110 cm (Pinthus and Levy 1983). In the course of following years, dwarf and semi-dwarf varieties have been used extensively to introduce, not only reduced plant height and lodging resistance, but also a number of other genes, which increased the number of fertile florets per spikelet, the number of spikelets per head and the number of tillers per plant, resulting in increasing yield (Borlaug 1983).

In the late 1940s, Orville Vogel at Washington State University used Norin 10 to produce high-yielding, semi-dwarf winter wheat varieties. In 1953, Borlaug at CYMMIT began crossing Vogel's semi-dwarf varieties with Mexican varieties. Borlaug's wheat semi-dwarf varieties contributed to development of modern wheat varieties all over the world and especially in South and East Asia.

In Italy, Nazareno Strampelli was the pioneer of wheat breeding. He made the first intervarietal cross in 1900 and, for the first time, introduced dwarfism genes through crossing of Japanese wheats, Italian cultivars and landraces. During the first half of XX century, his bread wheat cultivars (Ardito, Mentana, Villa Glori and Damiano) became the most cultivated in Italy and they spread in several Eastern European countries and particularly in China (Borghi et al 1975). In the 1970s, the reference price for good quality wheats went up addressing the breeding programs towards improving bread-making quality.

Since the end of last century, wheat breeders have focused also on nutritional and healthy values to develop new varieties.

Throughout the XX century to date, wheat breeders around the world mainly used methods and strategies (Fig. 1.9), which can be summarised as follows:

- *Selection within landraces*. It was used to identify cultivars that have superior performance compared with starting population from which they were selected. Two selection methods were mainly adopted by the breeders: *mass selection* and *pure line selection*. Mass selection

is exclusively used today to purify advanced lines, while in the past it was used when landraces were brought to new areas. Pure line selection allows to identify the best individual in a line, thus the best plant-derived progeny is used to develop the new cultivar. These two methods of breeding do not necessary require hybridization, but they are also used for selection after hybridization.

- *Crossbreeding – hybrid wheat*. The most common way of generating genetic variation is to mate (cross) two or more wheat parents that have contrasting genotypes. The practice of hybridization followed by selection as a wheat improvement strategy was initiated in the latter part of the 19th century by Vilmorin (cultivar Dattel) in France in about 1874 and by Wilhelm Rimpau in Germany in 1875 (Bonjeau and Angus 2001). Different strategies of crossing have been developed especially during the first half of the XX century and permitted the creations of the most important modern cultivars. Single way, two way, three way crosses and backcrossing have been widely used to increase genetic variability and to introduce interesting traits in cultivars (Briggs 1938; Dubcovsky 2005).
  Hybrid breeding in wheat offers the possibility to exploit heterosis (hybrid vigour) effect (Kempe and Gils 2011) and to benefit of yield improvement associated with it (Withford et al 2013). Currently, there is a limited number of wheat hybrid cultivars all over the world mainly due to the strong inbreeding nature of wheat closely associated with its naturally autogamous pollination mode (Withford et al 2013; Longin et al. 2012). The major benefits of hybrids compared to other varieties are increased yields, biotic and abiotic stress resistance, and enhanced yield stability (Hallauer et al. 1988).
- *Synthetic wheat*. The first published amphiploid was produced in the late 19th century from a cross between wheat and rye (Wilson 1876) producing *Triticale*. Synthetic hexaploid wheat (SHW) is an artificially created hexaploid wheat obtained through crossing tetraploids wheats (*T. turgidum*) and diploid wild goat grass (*Ae. tauschii*, DD). Embryo rescue and tissue culture are necessary to regenerate triploid organisms that are usually treated with colchicine to double their chromosome number. The first artificial hybridization between tetraploid wheat and *Ae. tauschii*, and hence the first primary synthetic wheat, was reported by McFadden and Sears (1946). Since then, tens of synthetic derivate lines (SDLs) have been registered as cultivars around the world.
- *Mutagenesis and induced translocation*. The use of chemical mutagenesis and radiation treatment allow to induce mutations in the DNA sequence. As of 2007, 200 cultivars of *T. aestivum* and 30 cultivars of *T. turgidum* ssp. *durum* feature a mutation (https://mvd.iaea.org/) (Stephen Baenziger and Ronald M. DePauw 2009). For example, the

mutation of *Ph1* gene (*ph1*) obtained through ethyl methanesulfonate (EMS) treatment and X-rays irradiation allowed many application of chromosome engineering (Sears et al 1993, Wall et al 1971, Gill et al 1993, Sestili et al 2010).

- *In vitro culture - Double haploid*. Doubled haploid technology generates homozygous lines from haploid tissue. The method involves production of plants from haploid tissue and doubling the chromosomes (Guzy- Wróbelskaand and Szarejko 2003). The resultant plant will be completely homozygous and homogeneous. Haploid tissue may be produced by chromosomal elimination in wide crosses or by the direct use of pollen grains, microspores, or ovules as haploid tissue. The chromosome numbers of the haploid plant are doubled by application of the alkaloid colchicine, which interferes with micro tubular activity. Doubled haploid cultivars have been released in a number of countries and some have become dominant cultivars.
- *Molecular breeding*. Started in 1990s with the use of molecular marker, genome mapping and sequencing, it involves the use of modern molecular biology tools such as Marker Assisted Selection (MAS), genomic selection and Genome Wide Association Study (GWAS) (Ortiz et al 2007).

1.3.2 Transgenesis

In the 1970s and 1980s, the advent of biotechnology, particularly of recombinant DNA technology, and the development of *in vitro* tissue culture allowed to modify genomic DNA of wheat through genetic transformation (Dudits et al. 1975; Chin and Scott 1977; Shimada 1978; Shimada and Yamada 1979) (Fig.1.9). Genetic manipulation of wheat has a great potential in functional genomics, but also in plant breeding, since makes it possible the introgression of genes from other species, synthetic genes or any DNA sequence, and enables targeted editing of plant genome to increase genetic variability. In 1992, Vasil and colleagues reported for the first time successful transformation and regeneration of wheat. Current successful examples of transgenic traits include herbicide tolerance, pathogen resistance, insect resistance, abiotic stress tolerance and improved nutritional quality (Vasil, 2007) although no transgenic wheat is on the market up to now.

Different technologies exploiting genetic transformation have been developed such as intragenesis and cisgenesis, RNA interference (RNAi) for gene silencing and genome editing, besides classical transgenesis.

Genetic transformation of wheat require cultures of explants capable to receive exogenous DNA, proper methods to deliver DNA into plant cells and regeneration of fertile plants.

Scutellar cells from immature embryos (10 to 15 days after anthesis) represent the target tissue for wheat transformation. At this developmental stage, the grains are at early–medium milk stage and contain translucent embryos each 1–3 mm in length, which are optimal for transformation (Jones 2005) and for the regeneration of whole plant. Other explants have been used with less success, such as embryogenic callus (Vasil et al., 1992), mature seeds (Miroshnichenko et al. 2011) and immature inflorescence (Rasco-Gaunt et al 1999; Lamacchia et al. 2001), from which microspore or immature pollen grains can be cultured and transformed as recently reported by Bhowmik et al (2018).

In planta transformation methods for wheat have also been reported (Supartana et al., 2006; Risacher et al. 2009; Zale et al. 2009), but they are not widely used yet. In 2018, Hamada and colleagues reported *in planta* genome editing of wheat through biolistic delivery (*in planta* particle bombardament, iPB). The authors argued the application of iPB to transform wheat cultivars recalcitrant to conventional transformation method.

Besides target tissue for transformation, many other factors affect wheat transformation, such as donor plant health and genotype, plasmid vectors, promoters to drive expression of transgenes marker and selectable genes, culture media composition and transformation delivery methods. Among transformation methods, Agrobacterium mediated transformation and biolistic method became the most common ways to deliver DNA into wheat cells, because of their reproducibility and higher efficiency compared to other methods (Blechl and Jones 2009).

Agrobacterium-mediated transformation.

Bacteria containing engineered T_i plasmid are co-cultivated with immature embryos in order to transfer DNA into plant cells (McCullen and Binns 2006). Recombinant DNA can be included in T-region between right and left border of T_i plasmid. In 1997, for the first time, researchers at Monsanto Company (USA) demonstrated that wheat can also be successfully transformed using Agrobacterium (Cheng et al. 1997). Since then many genetic transformation of wheat have been obtained through Agrobacterium mediated transformation.

Agrobacterium mediated gene transfer ensure low copy number integration of the transgene increasing expression stability, no requirement of special equipment and technique, but does not guarantee high transformation efficiency (1%~5%).

Biolistic method.

Also known as microprojectile bombardment, gene gun, or particle gun, the biolistic method consists of coating plasmid DNA onto tungsten or gold particles and to accelerate the coated particles with the force of a pressure pulse of helium gas (Klein et al. 1987), exploding gun powder (Sanford 1988), or electric discharge (McCabe and Christou 1993) to high velocity to pass through plant cell walls and membranes into the nucleus or cytoplasm. The first fertile transgenic wheat plants were made using particle bombardment and were reported 20 years ago (Becker et al. 1994; Nehra et al. 1994; Vasil et al.1992; Weeks et al. 1993). The biolistic method is widely used in many laboratories mostly because it appears less genotype dependent than other methods and it is the most reliable method for the production of fertile transgenic wheat plants.

1.3.3 New Breeding Techniques

New breeding techniques are emerging rapidly from advances in genomic research and for the application in crop improvement (Fig. 1.9). They enable precise, targeted and reliable changes in the genome and have significant potential for the sustainable intensification of agriculture and food security. Unlike chemical or radiation-induced mutagenesis, often traditionally used as a basis for crop improvement, the new breeding techniques do not create multiple, unknown, unintended mutations throughout the genome.

The development of new techniques in plant breeding have not led to the replacement of the older methods. At the moment, the use of all available technologies is essential for plant breeding. Conventional breeding techniques, transgenesis and new plant breeding techniques are essential components that plant breeders should adopt for crop improvement. During the last 20 years new biotechnological techniques and especially new plant breeding techniques have been developed.

The new breeding techniques currently commonly adopted include:

- *Cisgenesis and Intragenesis*. They include the transfer of gene(s) from same or closely related species.
- *Genome/Gene editing (Targeted mutagenesis).* It includes specific target modification mediated by meganuclease, zinc-finger nuclease, TALEN (transcription activator-like effector nuclease) and CRISPR-Cas tools.
- Transient introduction of recombinant DNA, for example *oligonucleotide-directed mutagenesis, agro-infiltration* and *RNA-induced DNA methylation gene silencing*
- *Reverse breeding*. It implies the use of double haploid technology to produce homozygous line starting from heterozygous plant and avoiding back-crossing and selection.

- *Synthetic Genomics*. Large functional DNA molecules can now be synthesised efficiently and quickly without using any natural template.

For several of these techniques, the resultant plant product is free from foreign genes and would not be distinguishable from the product generated by conventional breeding techniques. However, the definition of GMOs remains the same as in 1990, but it does not reflect the state of art of modern breeding technologies. Crops produced using some of these new plant breeding techniques cannot be distinguished from their conventionally counterparts and therefore there are claims that they should be exempted from the GMO legislation.

1.3.3.1 Gene Editing

In contrast to the transgenic approach, which leads to random insertions and different phenotypes, genome-editing methods produce defined mutants, thus becoming a potent tool in functional genomics and crop breeding (Fig. 1.10).

The first generation genome editing techniques includes:

- *Meganuclease*. They are specific DNA-cleaving enzymes that recognize target sequences on DNA ranging from 12 to 40 base pairs. Some meganucleases have been engineered and contain a non-specific nuclease domain fused with a sequence-specific DNA binding domain. Such engineered nucleases can precisely cleave the targeted gene and the breaks can be repaired through DNA repair mechanisms of the cell (Gaj et al. 2013).
- *Zinc Finger Nuclease (ZFN).* Zinc fingers are zinc ion-regulated small protein motifs that bind to DNA in a sequence-specific manner. Each zinc finger module sequence recognizes a 3-bp DNA and the fused nuclease cut the DNA. Therefore, unlike meganucleases, multiple zinc finger modules could be assembled into a larger complex to achieve higher DNA binding specificity (Gaj et al. 2013).
- *Transcription Activator-Like Nuclease (TALEN).* TALENs are engineered transcription activator-like effector (TALE) domain repeats, which can recognize one single base on DNA, fused with the FokI nuclease domain necessary for target genome editing(Stephens and Barakate 2017). TALENs show higher target binding specificity and versatility compared to ZFNs thanks to single-base recognition system. TALENs have been used for genome editing in plants including *Arabidopsis* (Cermak et al., 2011), rice (Li et al., 2012), tobacco (Zhang et al. 2013) and *Brachypodium* (Shan et al., 2013).

ZFN and TALEN were the dominant genome editing tools until the rise of Clustered Regularly Interspaced Palindromic Repeats (CRISPR) and Crispr associated protein (Cas), thereafter Jinek et

al (2012) demonstrated that CRISPR biological immunity can be exploited as genome engineering tool (Fig. 1.10). Since 2013, second-generation genome editing techniques such as CRISPR/Cas9 and its derivatives, replace the first-generation genome editing tools. For the first time ever, researchers had an extremely flexible tool that could be easily guided to target any location in the genome by simply designing a short synthetic guide RNA (sgRNA). Due to high editing efficiency and ease of use, researchers from diverse fields quickly adopted CRISPR technology for various genome-targeting purposes.

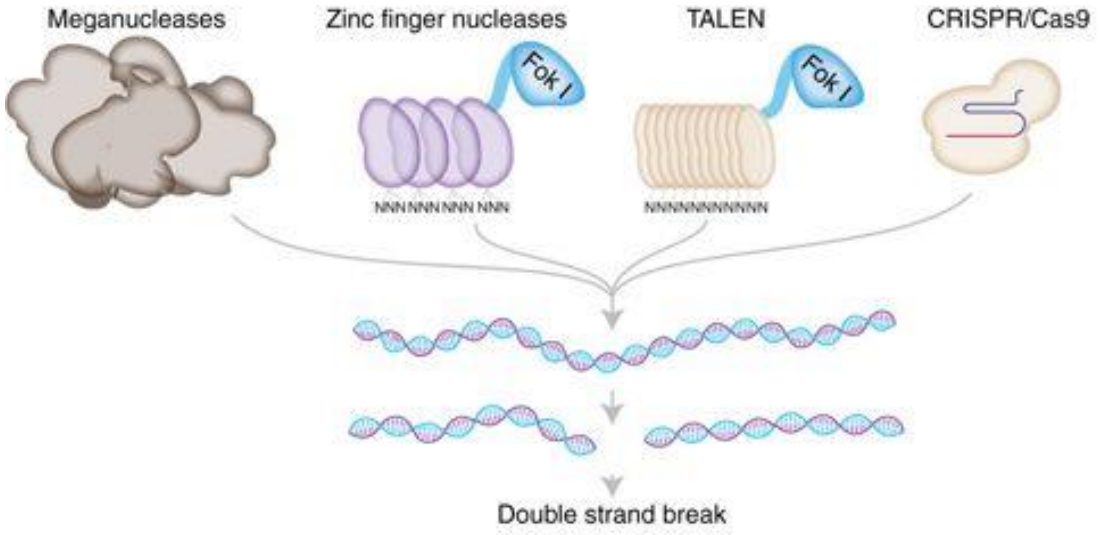

Figure 1.10: The basic working principle of major genome-editing technologies *(Adli et al 2018).*

1.3.3.1.1 *CRISPR-Cas*

CRISPR-Cas enables genome editing introducing targeted DNA double-strand breaks (DSBs) and exploiting cellular DNA repair mechanisms. It requires the use of an engineered endonuclease guided to a specific position on the genome by sgRNAs.

CRISPR-Cas systems are part of the adaptive immune system of bacteria and archaea (Fig. 1.11). They allow the cleavage of invading DNA in a sequence-dependent manner, ensuring protection against viruses. The integration of short fragments of the invading DNA (*spacers*) into the CRISPR array ensure the heritable immunity of bacteria. The CRISPR array consists of identical repeats (referred to as *crispr RNA, crRNA*), and unique spacers (referred to as *protospacers*) located in a genomic locus also containing *trans-activating crRNA (tracrRNA)* and a series of genes encoding endonuclease Cas proteins. The entire array is transcribed as a single mRNA (*pre-crRNA*), under the direction of a promoter located in the leader sequence, and it is partially complementary to *tracrRNA*. The RNase III, which recognise RNA duplex, cleaves the *dsRNA* realising *crRNA/tracrRNA* complexes (Fig 1.11). These complexes activate and guide Cas proteins to target specific genomic locus introducing DSBs. A prerequisite for cleavage is the presence of a

conserved protospacer-adjacent motif (PAM) downstream of the target DNA, which usually has the sequence 5′-NGG-3 for SpCas9. The presence of a PAM proximal to the acquired spacer and its absence in the CRISPR array facilitates robust immunity while averting autoimmune targeting of the CRISPR array (Hille et al 2018).

On the basis of Cas genes and the nature of interference complex, CRISPR-Cas systems have been divided in two classes, six types and several subgroups. Class 1 CRISPR-Cas systems include type I, III and IV, commonly found in *Archea* and employ multi-Cas protein complexes for interference.

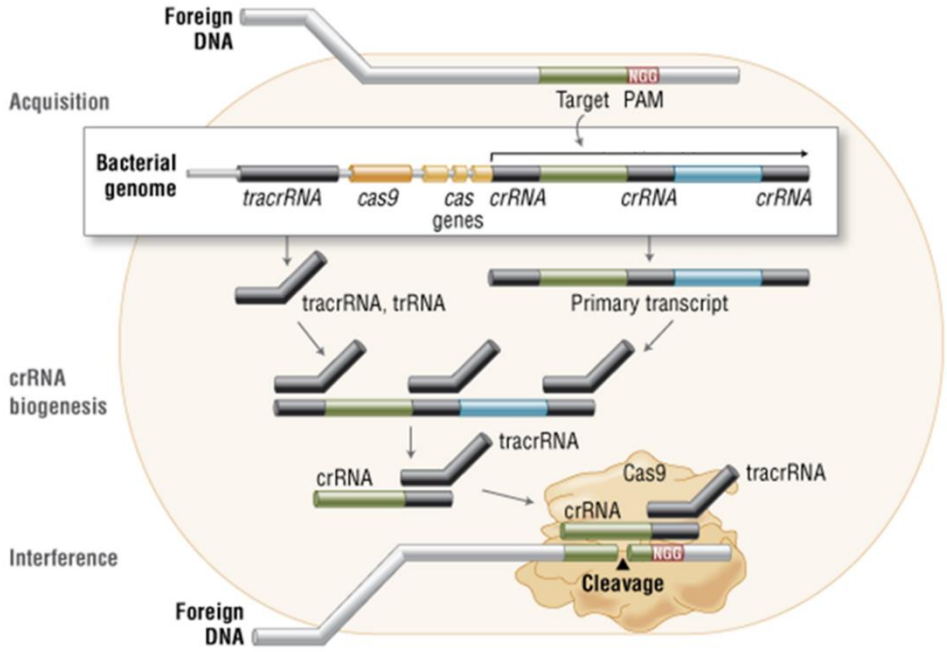

Figure 1.2: Bacterial Adaptive Immunity. *CRISPR loci organization on bacterial genome, processing of crRNA/tracrRNA complexes and Cas9 in vivo activity (https://international.neb.com/tools-and-resources/feature-articles/crispr-cas9-and-targeted-genome-editing-a-new-era-in-molecular-biology)*

Class 2 includes types II, V and VI which acting through one endonuclease protein. The most widely type used as genome editing tool is type II Cas9 from *Streptococcus pyogenes* (SpCas9), because of its highly efficiency and its simple PAM sequence (5'-NGG-3').

Many CRISPR-Cas system are continually being discovered and show differences in their size, PAM requirements and substrate preferences, such as Cpf1 proteins (class 2 type V endonuclease) from *Acid aminococcus sp* and *Lachnospiraceae bacterium* that began widely used (Zetsche et al 2015, Fonfara et al 2016). In a recent paper published in Nature, Liu et al. demonstrate the DNA cleavage characteristics of CasX (Liu et al 2019)

The Cas9 protein consists of a bi-lobed architecture including adjacent active sites and two nucleic acid binding grooves: a large recognition (REC) lobe and a small nuclease (NUC) lobe that are

connected by a helix bridge (Fig 1.12a). The REC lobe contributes to the specific function of Cas9, whereas the NUC lobe includes a protospacer-adjacent motif (PAM)-interacting domain (PI) and two nuclease domains (RuvC-like domains and HNH domain) each of which cut one DNA strand (Nishimasu et al 2014) (Fig 1.12b). The Cas9 is activated when combined with the tracrRNA/crRNA complex or sgRNA at its REC lobe. When Cas9 recognize the complementary spacer sequence adjacent to PAMs on double strand DNA, the HNH nuclease domain cleaves the gRNA–DNA hybrid, while RuvC cleaves the other strand to form a double-strand break (DSB) (Song et al 2016).

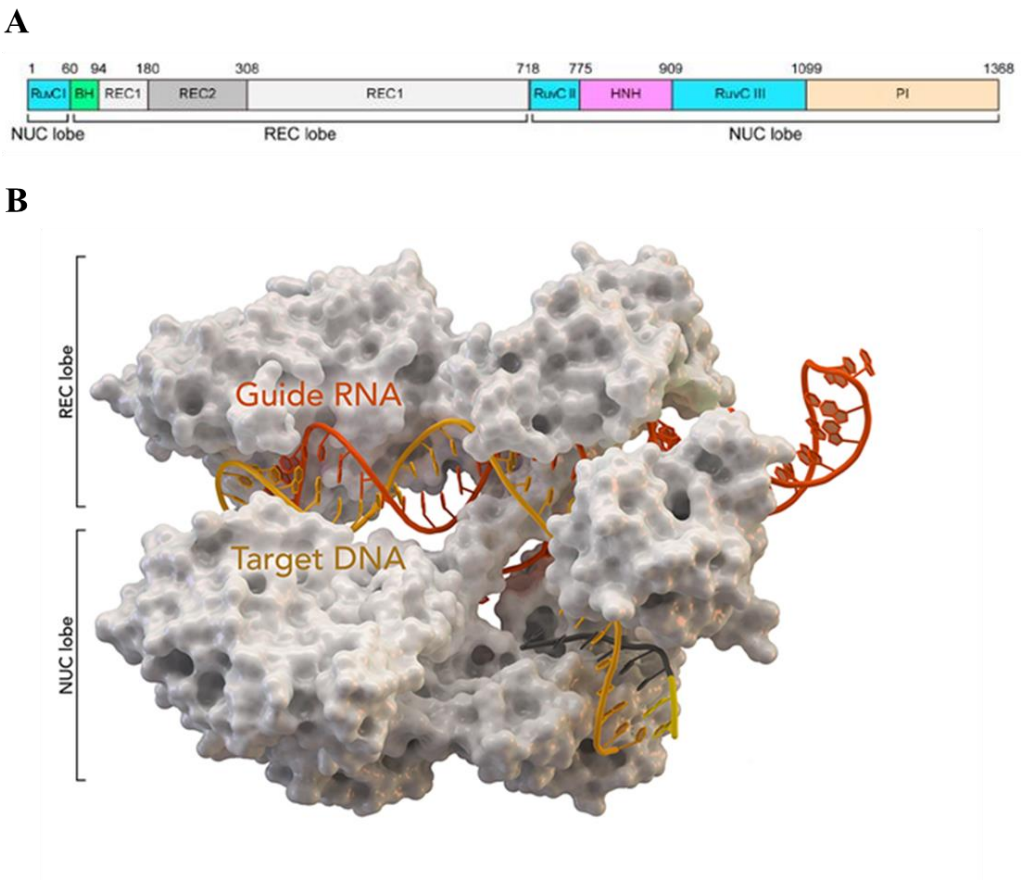

Figure 1.3: Cas 9 structure. *(A) Domain organization of the type II-A Cas9 protein from S. pyogenes)). (B) 3D structure of the type II-A Cas9 protein from S. pyogenes (https://it.wikipedia.org/wiki/Cas9#/media/File:Cas9_5AXW.png)*

The rise of CRISPR as breeding tool began in 1987 when Ishino et al identified CRISPR arrays for the first time in the *Escherichia coli*. Many years passed before the understanding of their biological role. In the 2007, Barrangou et al showed the role of CRISPR arrays in adaptive immunity when combined with Cas genes, starting from the evidence of the spacers as homologous to viral and plasmid sequences (Bolotin et al. 2005; Mojica et al. 2005; Pourcelet al. 2005). Definitely, CRISPR

mechanism based on RNA-mediated DNA targeting was demonstrated in the following years by different authors (Brouns et al. 2008; Deltcheva et al. 2011; Gameau et al. 2010; Marraffini and Sontheimer 2008). Jinek et al. in the 2012 showed that a chimeric sgRNA, which exerts the same function of crRNA and transcrRNA complex, could be achieved by the fusion of crRNA and tracrRNA and changing 20 nucleotides in the crRNA, which confer targeting specificity. Since then, hundreds of papers were published demonstrating the versatility and the ease of use of CRISPR-Cas system as breeding tool in several organisms, including higher plants. CRISPR/Cas9-based genome editing in plants were reported for the first time in the 2013 (Feng et al. 2014; Li et al. 2013; Nekrasov et al. 2013; Shan et al. 2013). At the present CRISPR-Cas genome editing have been used in many important crops including maize, wheat, barley, soybean, tomato, sweet orange etc.

Different genome modifications can be achieved through CRISPR-Cas depending on DNA repair mechanism adopted by the cell to repair DSBs; non-homologous end joining (NHEJ) and homologous recombination (HR). The NHEJ occurs more often than HR during DSBs repair (Ray and Lager 2002), and causes random insertions or deletions (indels), which can result in frameshift mutations and gene-knockout. Alternatively, when a template with regions of homology to the sequence surrounding the DSB is available, the DNA damage can be repaired by HR, and this mechanism can be exploited to achieve precise gene modifications or gene insertions (Fig. 1.13).

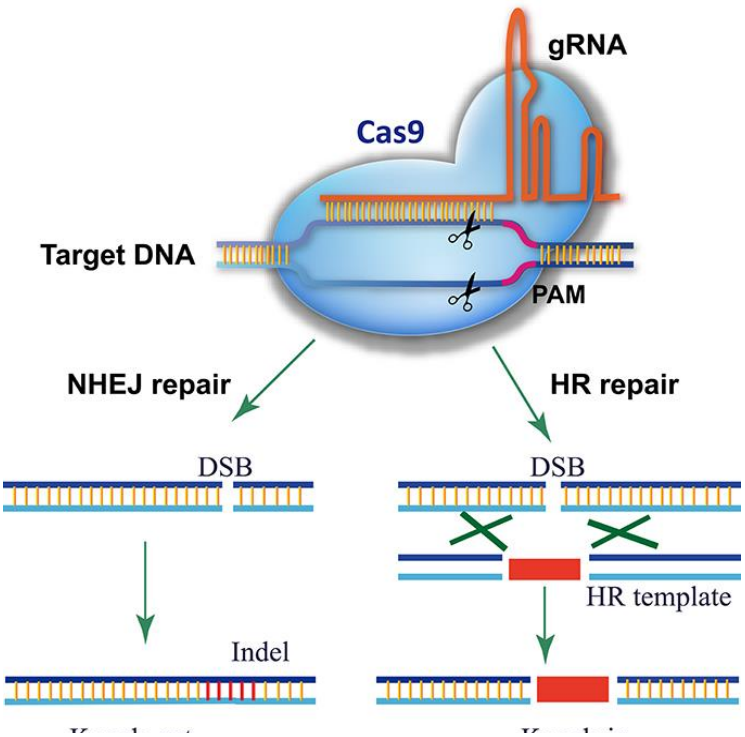

Figure 1.13: DNA repair mechanism. *Double strand break introduced by Cas mechanisms of DNA repair adopted by the cell (Razzaq and Masood 2018).*

Recent advances in the understanding of CRISPR systems and Cas proteins along with the need to improve editing efficiency, targeting fidelity, packaging and delivering into different cell type, allowed to extend CRISPR-Cas tool frontiers through re-engineering Cas proteins. The major goal remains the reduction of off-target effects. Introduction of point mutation in RuvC and HNH nuclease domains, inactivating catalytic activity, can be exploited to modulate Cas activity. A point mutation in one of the two catalytic domains results in a nickase enzyme (D10A and H840A SpCas9) which combined with paired guide RNAs can improve cleavage specificity. Simultaneous point mutation in both domains results in catalytically inactive dead Cas9 (dCas9) with complete loss of DNA cleavage activity. The dCas9 fused to transcription activation or repression domains allows transcriptional repression or activation. Furthermore, various functional protein domains can be loaded on dCas9 and exploited for different purposes. For example, a dCas9 fusion with the histone acetyltransferase p300 could activate target gene expression (Hilton et al. 2015), while a fusion with the histone demethylase LSD1 led to repression of target gene expression (Kearns et al. 2015) and even dCas9-GFP fusion protein has been used to study chromatin organization. (Chromatin imaging) (Chen et al 2013). O'Connel et al in the 2014 demonstrated that CRISPR-Cas can be used for RNA recognition and cleavage implementing RNA research in transcriptomic and allowing virus enhanced resistance.

Finally, base-editing technologies have been developed in order to convert a single base into another without DNA DSBs (second-generation genome-editing tools).

Notably, it was also shown that CRISPR allows high-efficiency multiplex genome engineering by using multiple gRNAs with different sequences targeting different loci simultaneously (Cong et al. 2013; Mali et al. 2013, Adli 2018).

CRISPR-Cas technology also represents a leap forward in transgene-free modified organisms. Transient expression of sgRNAs and Cas9, through ribonucleotide protein (RNP) complexes, or segregation of sgRNA/Cas9 transgenes in later generations allow to obtain transgene-free progeny with desired modifications.

1.4 WHEAT KERNEL

In the wheat kernel, the pericarp and seed coat are tightly linked. This type of fruit is given the botanical term of caryopsis. The grain is usually between 5 and 9 mm in length, weighs between 35 and 50 mg. The plants accumulate products of photosynthesis and nitrogen metabolism in the grain, which contains a large amount of carbohydrates (about 65-80 %), mainly in the form of starch, a good amount of proteins (7-18 %), 2 -9 % of lipids and small amounts of vitamins and

micronutrients. The wheat grain (Fig. 1.14) contains three distinct parts that are separated during the milling process to produce flour (Surget and Barron 2005):

- *Germ*. It is about 2-3% of the kernel dry weight. The germ is often separated from flour in milling because the fat content limits flour's shelf-life. It is rich in proteins (25%) and lipids (8-13%) and contain rather high level (4, 5%) of minerals (Zinc, Iron, copper and magnesium). It also represent an important source of E vitamin (Cornell 2003).
- *Bran*. It comprises about 13-17% of the kernel dry weight and consists of several layers ranging from pericarp and seed coat (external layer) to aleurone layer (inner layer). This part of grain is separated from flour during milling process but it is rich of water-insoluble fibre, minerals and B vitamins (Posner and Hibbs 2005; Delcour and Hoseney 2010). The aleurone layer is rich in proteins and enzymes, which play a vital role in the germination process.
- *Endosperm*. It makes up about 80-85% of the dry kernel weight. The endosperm without the aleurone layer is referred to as mealy or starchy endosperm. Its cells are rich in starch granules (60-70 % of the dry kernel weight) and also contain fats (1, 5%) and proteins (13%): albumins, globulins and the major proteins of the gluten complex, glutenins and gliadins (proteins that will form the gluten at dough making). The contents of minerals (ash) and of dietary fibres are low, 0, 5% and 1, 5%, respectively (Belderok et al. 2000; Stone and Morell 2009).

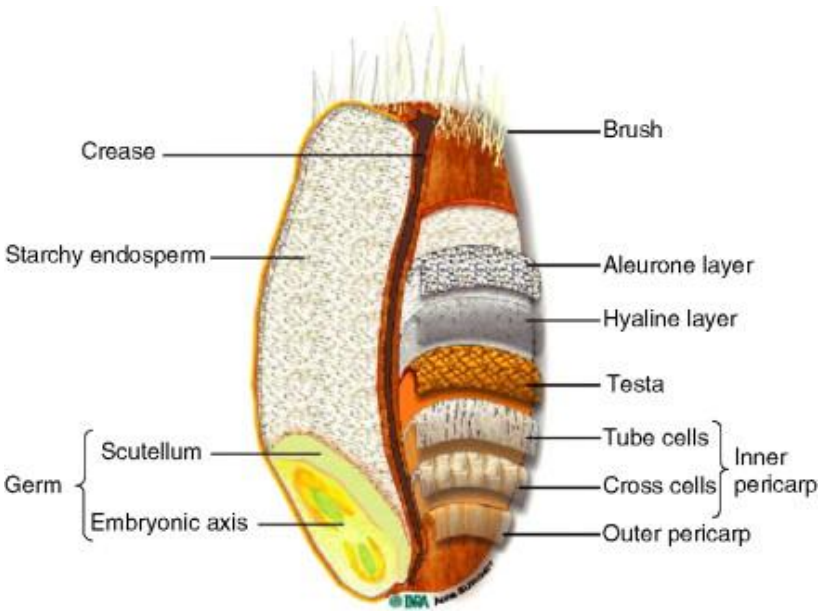

Figure 1.14. Wheat kernel structure

Caryopsis hardness is a key factor in the end use of wheat flour determination. Indeed, the hardness of the seed influences the level of damaged starch during the milling and has a strong effect on the

level of water absorption of the flour. Bread wheat varieties may be classified as “hard” and “soft”, whereas durum wheat varieties exists as only “hard” wheat, though recently soft durum wheat have been obtained (Morris and Beecher 2012). The locus *Ha* on chromosome 5D encodes for three enzymes involved in the determination of soft or hard phenotype (puroindoline a, b and Grain Softness Protein, GSP-1) (Bhave e Morris, 2008).

1.4.1 Proteins

The protein content of wheat grains may vary between 10% and 18% of the total dry matter. Wheat proteins are classified according to their extractability and solubility in various solvents (Fig. 1.15). This classification is based on the classic work of T.D. Osborne (1907) at the turn of the last century. In his procedure, sequential extraction of ground wheat grain results in the three different protein fractions: albumins and globulins, gliadins and glutenins. Since their high content in proline and glutamine, gliadins and glutenins are referred to as prolamins and classified as *gluten forming* proteins, whereas albumins and globulins should be considered *non-gluten forming* proteins. An

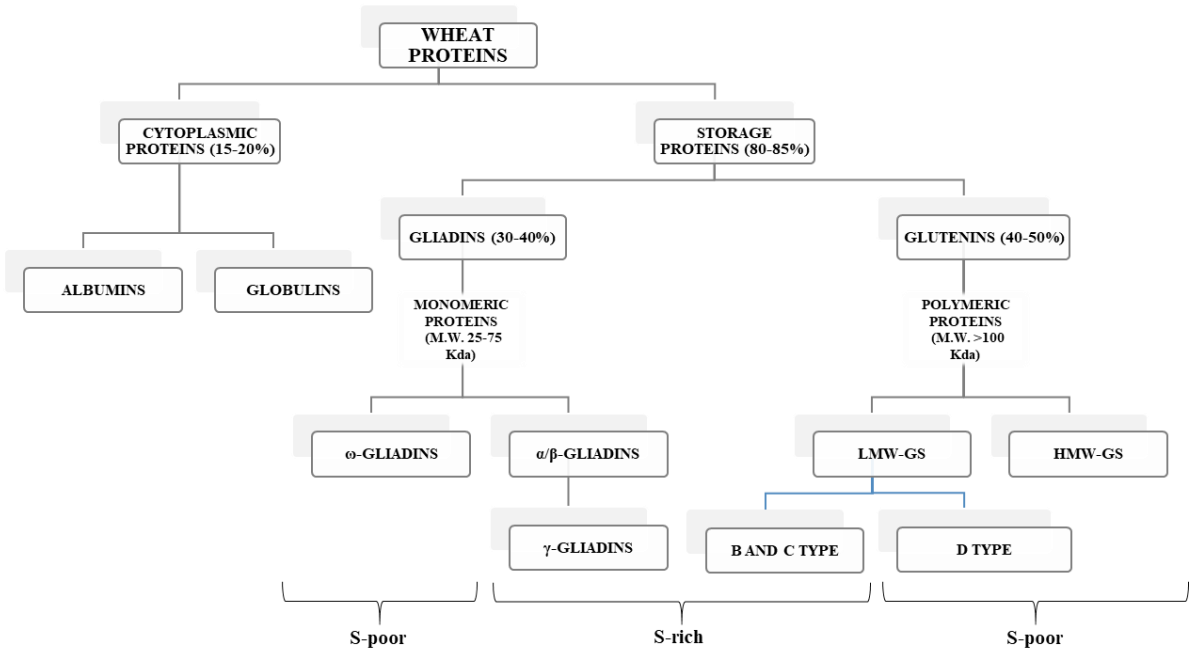

Figure 1.4: Classification of wheat endosperm proteins on the basis of solubility, as described by Osborne (1907).

alternative method of classification, used for all *Triticeae* prolamins (wheat, barley and rye), is based on sulphur content of proteins: S-rich proteins, S-poor proteins and high molecular weight protein subunits (HMW-GS) (Kreis et al. 1985; Shewry and Tatham 1999).

1.4.1.1 Albumins and globulins

Albumin are water-soluble proteins, globulins are proteins soluble in salt solutions and together they are the smallest wheat proteins in size. They represent about 15-20% of the wheat flour proteins and accumulate in the aleurone layer and scutellum. They include a large number of enzymes with α/β – amylase activity, serpins, and other physiologically active proteins. These predominant albumins and globulins serve as nutrient reserves for the germinating embryo. They also help in protecting embryo from insects and pathogens before germination (Dupont and Altenbach, 2003).

Because of their higher lysine and methionine contents as compared to the rest of the proteins in the wheat grain, albumins and globulins have the best-balanced amino acid composition (Lasztity 2017).

1.4.1.2 Gliadins

Gliadins are monomeric proteins soluble in alcohol/water mixtures (Anderson and Greene 1997). They are accumulated into the endosperm as storage proteins and participate to gluten polymeration.

Their molecular weight range between 20 – 60 KDa and they contain cysteine residues (*S-rich* proteins) involved in intra-chain disulphide bonds. According to their electrophoretic mobility at low pH in acid polyacrylamide gel electrophoresis (A-PAGE), they are classified as α-gliadins, β-gliadins, γ-gliadins e ω-gliadins (Fig. 1.16) (Shewry et al. 1986, Woychik et al. 1961).

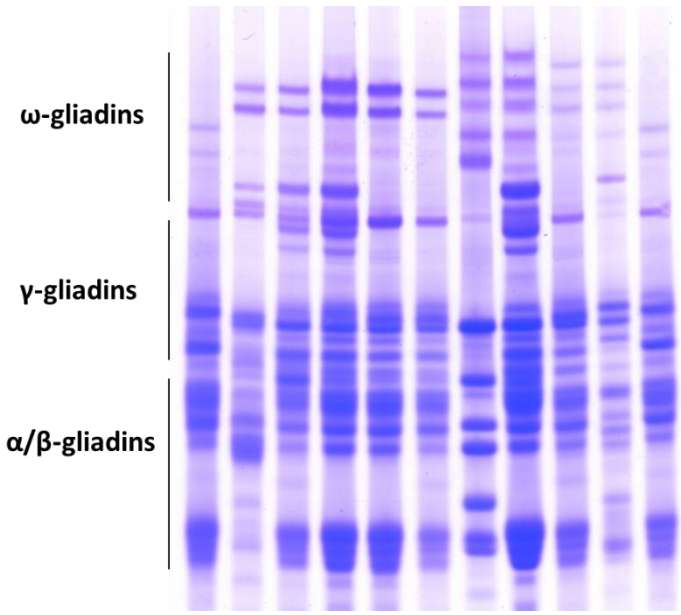

Figure 1.5: A-PAGE. *Electrophoretic mobility of gliadins extracted from durum and bread wheat.*

Despite their different electrophoretic mobility (MW 30-45 KDa), α- and β- gliadins show similar primary structure (Kasarda et al 1987) consists of three domains: N-terminal domain, repetitive domain (R-domain) made of Proline and Glutamine repeated motifs and C-terminal domain containing six cysteine residues and most of charged aminoacids (Fig. 1.20) (Bartels et al. 1986; Anderson et al. 1984; Rafalski et al. 1984; Kasarda et al. 1987).

The molecular weight (30-45 KDa) and the structure of γ-gliadins is similar to and consists of three separate domains: N-terminal domain, repetitive domain rich in proline and glutamine and C-terminal domain with eight cysteine residues and the most of charged aminoacids (Bartels *et al.* 1986).

The structure of ω-gliadins is completely different from α-, β- and γ- gliadins. They have higher molecular weight, ranging between 40 – 60 KDa and consists of long repetitive domain containing glutamine, glutamic acid, proline and phenylalanine residues and short C-terminal domain (Fig. 1.20). They are deprived of cysteine residues (S-poor proteins) so that they form no inter or intra chain bond (Hsia and Anderson 2001; Hsia and Anderson, 2013).

Cysteine residues of α-, β- and γ- gliadins can be found in conserved positions and in equal number (6 for α- gliadins and 8 for γ- gliadins). They are involved in intra-chain disulphide bonds resulting in coiled structures, able to establish only secondary interaction with other gluten proteins (Delcour et al. 2012; Bonomi et al. 2013). Gliadins with an additional cysteine residue due to a point mutation can form intermolecular bonds enabling the integration of the modified gliadins into the glutenin polymers acting as chain terminators (Camerlengo et al 2017).

Gliadins are encoded by multigene families located at complex loci on the homeologous chromosome of groups 1 and 6 (Fig. 1.17). The α/β gliadins are controlled by genes located at the *Gli-2* loci (*Gli-A2, Gli-B2* e *Gli-D2*) on the short arms of the chromosomes of group 6A, 6B and 6D. It is estimated that the average number of copies of α-gliadin genes per haploid genome ranges from 25–35 to 100–150 copies; this large number has been generated by duplication, deletion events and retrotransposon insertion (D'Ovidio et al. 1991; Okita et al. 1985; Anderson et al. 1997). The ω- and γ- gliadins are encoded by genes at the three complex locus *Gli-A1*, *Gli-B1* and *Gli-D1* on the short arms of the homeologous chromosomes of group 1 and on other minor loci such as *Gli-3*, *Gli-5* and *Gli-6* (Fig. 1.17).

Many gliadin genes are considered inactive or pseudogenes (most of them are α- gliadin genes) due to point mutation of CAA or CAG codons in TAA or TAG stop codons (Anderson et al 1997; Xie et al. 2010; Van Herpen et al. 2006).

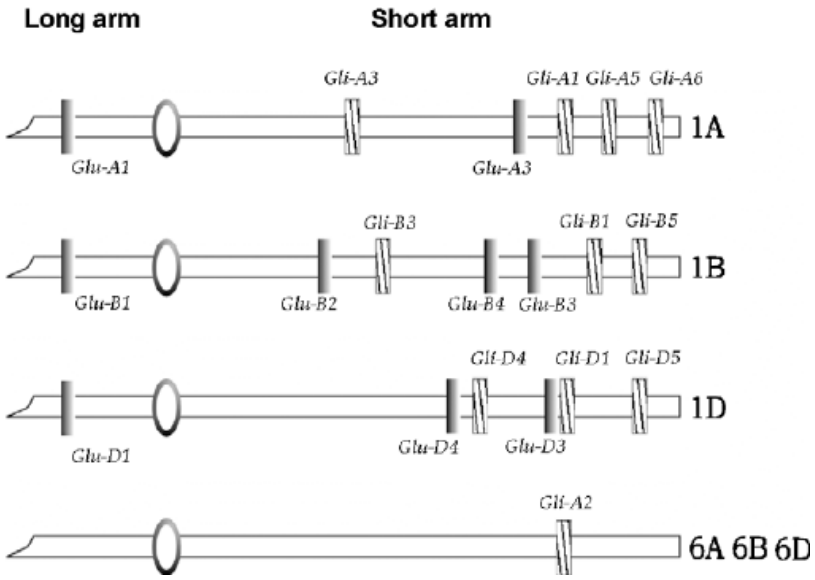

Figure 1.6: Chromosome distribution of gliadin and glutenin gene loci.

1.4.1.3 Glutenins

Glutenins are polymeric proteins extractable in acid solution. They form the largest protein polymers in nature with molecular weight ranging from 60 KDa to tenth millions (Wrigley 1996; Wieser 2007), affecting baking performance of flour dough (Schropp and Wieser 1996; Weegels et al. 1996; Gras et al. 2001). Glutenins appear to be largely responsible for gluten elasticity (MacRitchie 1992; Wieser 2007).

Following the reduction of intra and intermolecular disulphide bond with reducing agents, glutenins may be separated on polyacrylamide gel electrophoresis (SDS-PAGE) in two subunit classes: High molecular weight subunits (HMW-GS) ranging between 70-90 KDa and low molecular weight subunits ranging between 30-45 KDa (Wieser 2007). The glutenins subunits has also been further classified into four groups: HMW-GS belong to subgroup A and LMW-GS belong to B, C and D subgroup (Fig. 1.18) (Shewry and Tatham 1990; Wrigley 1996; Gianibelli et al. 2001). Payne and Corfield in 1979 reduced and separated high molecular weight gliadins on SDS-PAGE and observed similar mobilities of D and C subunits of LMW-GS. The C and D groups of LMW-GS are highly similar in sequence to individual α- /β-type and γ-gliadins, respectively, and are considered to be derived from these components by mutations resulting in the presence of additional cysteine residues, which are able to form interchain disulfide bonds (D'Ovidio and Masci 2004). The B-type LMW subunits form a discrete group and no closely related gliadin components have been identified (Shewry et al 2003).

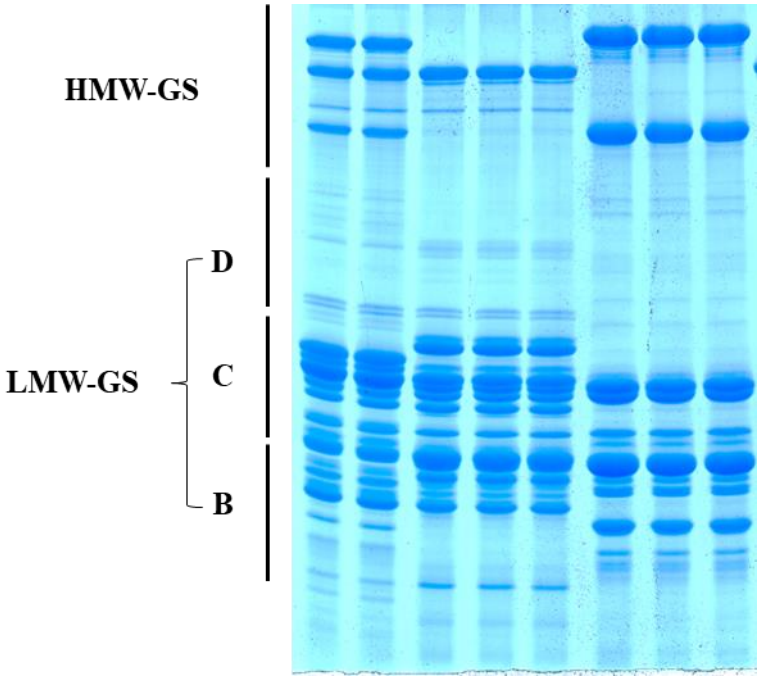

Figure 1.7: SDS-PAGE. *Electrophoretic mobility of glutenins extracted from durum and bread wheat.*

Although HMW-GS play the most important role in forming the gluten polymer, therefore in bread making quality, they represent no more than 10 % of the total kernel proteins (Shewry et al., 1992).

The HMW-GS subunits consists of three domains: short N-terminal domain, a repetitive domain of glutamine and proline ranging from 400 to 700 residues and C-terminal domain. On the basis of their repetitive motifs and the number and position of cysteine residue, HMW-GS can be classified in x-type subunits and y-type subunits. The x-type subunits contain four cysteine residues (three in N-terminal domain and one in C-terminal domain) and tripeptide, hexapeptide, and nonapeptide motifs, while the y-type subunits contain seven cysteine residues (five in N-terminal domain and two in C-terminal domain) and hexa- and nonapeptides motifs (Shewry et al. 2003). Two tightly linked genes at each homeologous *Glu-1* locus on the long arms of chromosome 1A, 1B and 1D encode for x-type and y-type subunits (Fig. 1.17) (Payne et al., 1984; Gianibelli et al., 2001). Because of the silencing of some genes, only three to five HMW subunit genes are expressed in different bread wheat cultivars. All cultivars contain 1Bx, 1Dx and 1Dy subunits, some cultivars also contain 1By or 1Ax and 1Ay subunits, (Payne et al 1983). When only one subunit is expressed by the *Glu-B1* or *Glu-A1* loci, this is always x type. A similar situation occurs at the *Glu-A1* and *Glu-B1* loci in durum wheat; 1Ax and 1Bx subunit are always expressed but sometimes 1Ay subunit may be expressed (Waines and Payne 1987; Levy et al. 1988). The silencing of some of the genes could be caused by the presence of a transposon-like insertion in the coding region (Harberd et al. 1986).

Allelic variation in expressing subunits at *Glu-1* loci were observed in last decades (Lawrence and Shepherd 1980; Payne and Lawrence 1983) resulting in proteins subunits with different mobility on gel electrophoresis and with different effects on flour bread-making proprieties (Fig. 1.19) (Lawrence et al. 1988; Payne and Seekings 1996; Lafiandra et al. 2000).

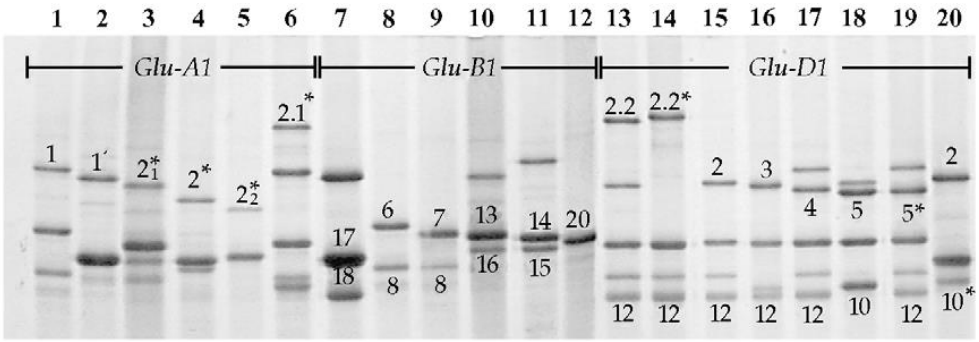

Figure 1.8: SDS–PAGE of allelic HMW glutenin subunits present at the three Glu-1 loci *(Shewry et al 2013).*

The typical B-type LMW-GS (35-45 KDa) have essentially similar structure of S-rich prolamins (α/β-, γ-gliadins and B-type and D-type): a short N-terminal domain followed by a repetitive domain and three C-terminal domains (Fig. 1.20). B-type LMW-GS contain six conserved cysteines as α/β- gliadins and additional cysteine residues in the N-terminal domain and C-terminal domain, which may form intermolecular bonds (Shewry and Tatham 1997). The C-type subunits (25-35 KDa) include α/β- gliadins that contain an additional cysteine residue involved in intermolecular bond. The D-type LMW-GS include ω, which have different structure and do not form disulphide bonds (Fig. 1.20).

Glu-3 loci on the short arms of chromosomes 1A, 1B and 1D encodes for B-type LMW and are tightly linked with *Gli-1* loci (encoding most of the ω- and γ-gliadins). Large allelic variation has been reported at the three *Glu-3* loci in bread wheat (Masci et al. 1995). Minor loci encoding B-type and C-type LMW, referred to as *Glu-B2*, *Glu-D4* and *Glu-D5*, were reported (Shewry and Halford 2003).

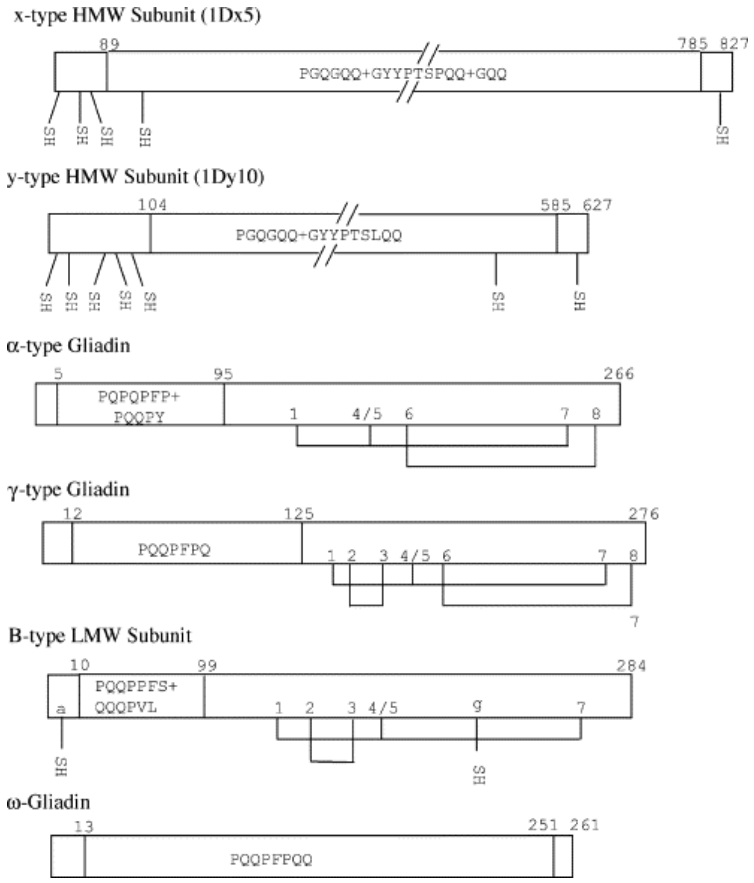

Figure 1.20: Domain structure of the typical wheat gluten proteins *(Shewry et al 2013).*

1.4.2 Starch

Starch is the most abundant carbohydrate in the wheat grain. It ranges between 60 % and 75 % of the total dry weight of the grain and accumulate in form of granules; large lenticular granules (25 - 40µm) and small spherical granules (5 – 10 µm) (Belderok et al. 2000). Starch consists of two types of glucose polymers: amylose and amylopectin. Amylose is a linear polymer with α-(1, 4) - linkages, showing a degree of polymerization of 10^3 – 10^4 glucose units. The degree of polymerization of amylopectin reaches 10^7 glucose units and α-(1, 4) linked glucose polymers are connected by α-(1, 6)-linkages, causing a larger and branched structure.

Normal wheat starch typically contains 20–30% amylose and 70–80% amylopectin (Konik-Rose et al., 2007). Modifying starch composition and amylose/amylopectin balance starch proprieties can be adapted for novel end use (Stone e Morell, 2009; Sestili et al., 2014).

1.4.3 Lipids

Lipids (3-4 % of the whole grain) are present mainly in the germ part of the seeds even though small quantities are present in the endosperm and bran. Their ability to associate with proteins and with starch due their amphipathic nature enable the formation of inclusion complexes. About 70%

of total lipids are non-polar lipids and polar lipids such as glycolipids and phospholipids represent the remaining quantity (Sugawara and Miyazawa 2001). High level of linoleic acid (C18:2) with lower amounts of palmitate (C16:0) and oleate (C18:1) were reported in wheat grain (Morrison 1978).

1.4.4 Fibre and Micronutrients

The grain is a significant source of dietary fibres and micronutrients, especially minerals and vitamin B and E. Vitamin A is a precursor of β-carotene and other carotenoids, which is also synthesized in wheat as well as in green and yellow vegetables. Carotenoids provide colour to wheat-based products, especially pasta from durum wheat. Much of the micronutrient content resides in the bran and aleurone layer.

Mineral content of wheat kernel includes: Iron, Zinc, Selenium, Manganese, Phosphorus, copper and their amount depends on the mineral content of the soil where the crop is grown and folate.

Dietary fibre is defined as the polysaccharide components of plants, which are indigestible, by enzymes in the human gastrointestinal tract (Bermink, 1994). It has divided in two main components: Soluble dietary fibres, which are soluble in water and includes pectic substances and hydrocolloids, and insoluble dietary fibres, which are insoluble in water and includes cellulose, hemicellulose and lignin. The fibre content of whole-grain wheat ranges from 12-15% of the dry weight and they are located especially in the bran. Arabinoxylans (AX) and (1→3), (1→4)-β-glucans are major components of wheat endosperm cell walls. Soluble fibre such as (1→3, 1→4)-β-D-glucan (referred to as β-glucan), has been shown to have immunostimulating activity as well as effects on the glycaemic, insulin, and cholesterol responses to foods (Dalmo and Bøgwald 2008).

The greatest portion of dietary fibres and micronutrients is removed during the milling process, when bran is separated from the endosperm. Therefore, the increasing awareness of the potential benefits of high-fibre diets has promoted a growing interest for the consumption of whole-grain breads.

1.5 WHEAT RELATED DISORDERS

Several disorders are associated with wheat derived food products consumption; in fact, it is considered one of the most common allergens source in human diet along with eggs, milk, soya and fish. Different studies highlighted that all wheat grain components (gluten proteins, α-amylase/trypsin inhibitors, oligosaccharides, disaccharides, monosaccharides and polyols) may be involved in triggering different pathologies.

Wheat related disorders might be classified into:

- *Autoimmune disorders*. They include celiac disease (CD), gluten-induced ataxia and dermatitis herpetiformis. The prevalence of CD is approximately 1% in the North American and European populations and it is substantially increasing in prevalence. The pathogenesis of CD is characterized by gluten, especially gliadin fraction that is the most important environmental factor in the pathogenesis of CD, in genetically predisposed people. In CD patients, gluten ingestion and partial digestion of gliadins and glutenins peptides cause an inappropriate T cell mediated immune response, causing self-perpetuating inflammatory damage of the small bowel and other organs, with variable combination of gluten-dependent clinical manifestations, in genetically predisposed subjects.
- *Wheat allergy*. It is an IgE-mediated food allergy and causes skin, respiratory and gastrointestinal tract involvement and in the most severe cases, anaphylaxis. This may also occur as a wheat-dependent exercise-induced anaphylaxis (WDEIA) due to the combination of wheat ingestion before physical activity. Baker's asthma and rhinitis, caused by inhalation of raw wheat flour, depends also on an IgE-mediated reaction, as demonstrated by Houba et al. (1998). The most involved allergens are α-amylase/trypsin inhibitors, but gliadins have been included as important allergens in the pathogenesis of wheat allergy.
- *Neither immune nor allergic disorders*. It concern a clinical condition in which patients show intestinal and extra-intestinal symptoms after ingestion of wheat food products and in which allergic or autoimmune mechanisms have been excluded. Among this clinical condition, there is non-celiac wheat sensitivity (NCWS) and seems to be triggered by α-amylase/trypsin inhibitors and maybe FODMAPs (fructans, galactans, fructose and polyols) (Mansueto et al 2019; Sherf 2019).

REFERENCES

Adli, M. (2018). The CRISPR tool kit for genome editing and beyond. *Nature communications*, *9*(1), 1911.

Anderson, O. D., & Greene, F. C. (1997). The α-gliadin gene family. II. DNA and protein sequence variation, subfamily structure, and origins of pseudogenes. *Theoretical and applied genetics*, *95*(1-2), 59-65.

Anderson, O. D., Litts, J. C., & Greene, F. C. (1997). The α-gliadin gene family. I. Characterization of ten new wheat α-gliadin genomic clones, evidence for limited sequence conservation of flanking DNA, and southern analysis of the gene family. *Theoretical and Applied Genetics*, *95*(1-2), 50-58.

Anderson, O. D., Litts, J. C., Gautier, M. F., & Greene, F. C. (1984). Nucleic acid sequence and chromosome assignment of a wheat storage protein gene. *Nucleic acids research*, *12*(21), 8129-8144.

Baenziger PS, DePauw RM, 2009 Wheat Breeding: Procedures and Strategies, pp. 275–308. In Wheat Science and Trade Book, edited by Brett F. Carver. Wiley-Blackwell.

Bartels, D., Altosaar, I., Harberd, N. P., Barker, R. F., & Thompson, R. D. (1986). Molecular analysis of γ-gliadin gene families at the complex Gli-1 locus of bread wheat (T. aestivum L.). *Theoretical and Applied Genetics*, *72*(6), 845-853.

Becker, D., Brettschneider, R., & Lörz, H. (1994). Fertile transgenic wheat from microprojectile bombardment of scutellar tissue. *The Plant Journal*, *5*(2), 299-307.

Belderok, B., Mesdag, J., Mesdag, H., & Donner, D. A. (2000). *Bread-making quality of wheat: a century of breeding in Europe*. Springer Science & Business Media.

Belderok, B., Mesdag, J., Mesdag, H., & Donner, D. A. (2000). *Bread-making quality of wheat: a century of breeding in Europe*. Springer Science & Business Media.

Bermink, M. R. (1994). Fibre analysis. In S. S. Nielson (Ed.), Introduction to the chemical analysis of foods (pp. 169–180). Boston, MA, USA: Jones and Bartlett Publishers Inc.

Bhave, M., & Morris, C. F. (2008). Molecular genetics of puroindolines and related genes: allelic diversity in wheat and other grasses. *Plant molecular biology*, *66*(3), 205-219.

Bhowmik, P., Ellison, E., Polley, B., Bollina, V., Kulkarni, M., Ghanbarnia, K., ... & Kagale, S. (2018). Targeted mutagenesis in wheat microspores using CRISPR/Cas9. *Scientific reports*, *8*.

Blechl, A. E., & Jones, H. D. (2009). Transgenic applications in wheat improvement. *Wheat Science and Trade*, 397-435.

Bolotin, A., Quinquis, B., Sorokin, A., & Ehrlich, S. D. (2005). Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology*, *151*(8), 2551-2561.

Bonjean, A. P., & Angus, W. J. (2001). *The world wheat book: a history of wheat breeding*. Lavoisier Publishing.

Bonomi, F., Iametti, S., Mamone, G., & Ferranti, P. (2013). The performing protein: Beyond wheat proteomics?. *Cereal Chemistry*, *90*(4), 358-366.

Borghi, B., Boggini, G., & Corino, L. (1975). BREEDING FOR QUALITY IN COMMON WHEAT. I o: EARLY SELECTION IN F 2 AND F 3 GENERATIONS. *Cereal Research Communications*, 205-214.

Borlaug, N. E. (1983). Contributions of conventional plant breeding to food production. *Science*, *219*(4585), 689-693.

Briggs, F.N. 1938. The use of the backcross in crop improvement. Am. Nat. 72:285–292.

Brouns, S.J., Jore, M.M., Lundgren, M., Westra, E.R., Slijkhuis, R.J., Snijders, A.P., Dickman, M.J., Makarova, K.S., Koonin, E.V., and van der Oost, J. (2008). Small CRISPR RNAs guide antiviral defense in prokaryotes. Science 321, 960–964.

Camerlengo, F., Sestili, F., Silvestri, M., Colaprico, G., Margiotta, B., Ruggeri, R., ... & Lafiandra, D. (2017). Production and molecular characterization of bread wheat lines with reduced amount of α-type gliadins. *BMC plant biology*, *17*(1), 248.

Cermak, T., Doyle, E. L., Christian, M., Wang, L., Zhang, Y., Schmidt, C., ... & Voytas, D. F. (2011). Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic acids research*, *39*(12), e82-e82.

Chen, B., Gilbert, L. A., Cimini, B. A., Schnitzbauer, J., Zhang, W., Li, G. W., ... & Huang, B. (2013). Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell*, *155*(7), 1479-1491.

Cheng, M., Fry, J. E., Pang, S., Zhou, H., Hironaka, C., Duncan, D. R., Conner, T. W., and Wan, Y. 1997. Genetic transformation of wheat mediated by Agrobacterium tumefaciens. Plant Physiol. 115: 971–980.

Chin, J. C., and Scott, K. J. 1977. Studies on the formation of roots and shoots in wheat Callus cultures. Ann. Bot. 41: 473–481.

Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., ... & Zhang, F. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science*, *339*(6121), 819-823.

Cornell H (2003) In: Cauvain SP (ed) Bread Making: Improving Quality. Woodhead Publishing, Cambridge

D'Ovidio, R., Lafiandra, D., Tanzarella, O.A., Anderson, O.A., Greene, F.C. (1991). Molecular characterization of bread wheat mutants lacking the entire cluster of chromosome 6A-controlled gliadin components. *Journal of Cereal Science*, Vol. 14, pp. 125–129.

Dalmo, R. A., & Bøgwald, J. (2008). ß-glucans as conductors of immune symphonies. *Fish & Shellfish Immunology*, *25*(4), 384-396.

Delcour, J. A., & Hoseney, R. C. (2010). Principles of cereal science and technology, AACC International. *Inc., St. Paul, MN, USA*, 229-235.

Delcour, J. A., Joye, I. J., Pareyt, B., Wilderjans, E., Brijs, K., & Lagrain, B. (2012). Wheat gluten functionality as a quality determinant in cereal-based food products. *Annual review of food science and technology*, *3*, 469-492.

Deltcheva, E., Chylinski, K., Sharma, C.M., Gonzales, K., Chao, Y., Pirzada, Z.A., Eckert, M.R., Vogel, J., and Charpentier, E. (2011). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature 471, 602–607.

D'Ovidio, R., & Masci, S. (2004). The low-molecular-weight glutenin subunits of wheat gluten. *Journal of cereal science*, *39*(3), 321-339.

Dubcovsky, J. 2005. The Wheat CAP. Applied wheat genomics [Online]. Available at http://maswheat.ucdavis.edu/ (verified 24 Mar. 2008).

Dudits, D., Nemet, G., and Haydu, Z. 1975. Study of callus growth and organ formation in wheat (Triticum aestivum) tissue cultures. Can. J. Bot. 53: 957–963.

DuPont, F. M., & Altenbach, S. B. (2003). Molecular and biochemical impacts of environmental factors on wheat grain development and protein synthesis. *Journal of cereal science*, *38*(2), 133-146.

Faris, J. D. (2014). Wheat domestication: Key to agricultural revolutions past and future. In *Genomics of plant genetic resources* (pp. 439-464). Springer, Dordrecht.

Feldman, M., & Levy, A. A. (2009). Genome evolution in allopolyploid wheat—a revolutionary reprogramming followed by gradual changes. *Journal of Genetics and Genomics*, *36*(9), 511-518.

Feldman, M., 2001 The origin of cultivated wheat, pp. 1–56. In The Wheat Book, edited by A. Bonjean, and W. Angus. Lavoisier Tech & Doc, Paris.

Feng, Z., Mao, Y., Xu, N., Zhang, B., Wei, P., Yang, D. L., ... & Zeng, L. (2014). Multigeneration analysis reveals the inheritance, specificity, and patterns of CRISPR/Cas-induced gene modifications in Arabidopsis. *Proceedings of the National Academy of Sciences*, *111*(12), 4632-4637.

Fonfara, I., Richter, H., Bratovič, M., Le Rhun, A., & Charpentier, E. (2016). The CRISPR-associated DNA-cleaving enzyme Cpf1 also processes precursor CRISPR RNA. *Nature*, *532*(7600), 517.

Gaj, T., Gersbach, C. A., & Barbas III, C. F. (2013). ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends in biotechnology*, *31*(7), 397-405.

Garneau, J.E., Dupuis, M.-E` ., Villion, M., Romero, D.A., Barrangou, R., Boyaval, P., Fremaux, C., Horvath, P., Magada´ n, A.H., and Moineau, S. (2010). The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature 468, 67–71.

Gianibelli, M. C., Larroque, O. R., MacRitchie, F., & Wrigley, C. W. (2001). Biochemical, genetic, and molecular characterization of wheat glutenin and its component subunits. *Cereal Chemistry*, *78*(6), 635-646.

Gill, K.S., B.S. Gill, T.R. Endo, and Y. Mukai. 1993. Fine physical mapping of Ph1, a chromosome-pairing regulator gene in polyploid wheat. Genetics 134:1231–1236.

Gras, P. W., Anderssen, R. S., Keentok, M., Békés, F., & Appels, R. (2001). Gluten protein functionality in wheat flour processing: a review. *Australian Journal of Agricultural Research*, *52*(12), 1311-1323.sh

Guzy-Wróbelska, J., and I. Szarejko. 2003. Molecular and agronomic evaluation of wheat doubled haploid lines obtained through maize pollination and anther culture methods. Plant Breed. 122:305–313.

Hallauer AR, Russell WA, Lamkey KR (1988) Corn breeding. In: Sprague GF, Dudley JW (eds) Corn and corn improvement, 3rd edn. American Society Agronomy, Madison, pp 463–564.

Hamada, H., Liu, Y., Nagira, Y., Miki, R., Taoka, N., & Imai, R. (2018). Biolistic-delivery-based transient CRISPR/Cas9 expression enables in planta genome editing in wheat. *Scientific reports*, *8*(1), 14422.

Harberd, N. P., Bartels, D., & Thompson, R. D. (1985). Analysis of the gliadin multigene loci in bread wheat using nullisomic-tetrasomic lines. *Molecular and General Genetics MGG*, *198*(2), 234-242.

Hille, F., Richter, H., Wong, S. P., Bratovič, M., Ressel, S., & Charpentier, E. (2018). The biology of CRISPR-Cas: backward and forward. *Cell*, *172*(6), 1239-1259.

Hilton, I. B., D'ippolito, A. M., Vockley, C. M., Thakore, P. I., Crawford, G. E., Reddy, T. E., & Gersbach, C. A. (2015). Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nature biotechnology*, *33*(5), 510.

Houba, R., Doekes, G., & Heederik, D. (1998). Occupational respiratory allergy in bakery workers: a review of the literature. *American journal of industrial medicine*, *34*(6), 529-546.

Hsia, C. C., & Anderson, O. D. (2001). Isolation and characterization of wheat ω-gliadin genes. *Theoretical and Applied Genetics*, *103*(1), 37-44.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A., & Charpentier, E. (2012). A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. *Science*, *337*(6096), 816-821.

Jones, H. D. (2005). Wheat transformation: current technology and applications to grain development and composition. *Journal of Cereal Science*, *41*(2), 137-147.

Kasarda, D. D., Adalsteins, A. E., & Laird, N. F. (1987). Gamma-Gliadins with alpha-type structure coded on chromosome 6B of the wheat (Triticum aestivum L.) cultivar'Chinese Spring'. In *Proceedings of the 3rd International Workshop on Gluten Proteins, Budapest, Hungary, May 9-12, 1987/editors, R. Lasztity, F. Bekes*. Singapore: World Scientific, c1987..

Kearns, N. A., Pham, H., Tabak, B., Genga, R. M., Silverstein, N. J., Garber, M., & Maehr, R. (2015). Functional annotation of native enhancers with a Cas9–histone demethylase fusion. *Nature methods*, *12*(5), 401.

Kempe, K., & Gils, M. (2011). Pollination control technologies for hybrid breeding. *Molecular Breeding*, *27*(4), 417-437.

Kihara, H. (1944). Discovery of the DD-analyser, one of the ancestors of Triticum vulgare. *Agric. Hort.*, *19*, 13-14.

Klein, T. M., Wolf, E. D., Wu, R., & Sanford, J. C. (1987). High-velocity microprojectiles for delivering nucleic acids into living cells. *Nature*, *327*(6117), 70.

Konik-Rose, C., Thistleton, J., Chanvrier, H., Tan, I., Halley, P., Gidley, M., ... & McMaugh, S. (2007). Effects of starch synthase IIa gene dosage on grain, protein and starch in endosperm of wheat. *Theoretical and Applied Genetics*, *115*(8), 1053.

Kreis, M., Forde, B. G., Rahman, S., Miflin, B. J., & Shewry, P. R. (1985). Molecular evolution of the seed storage proteins of barley, rye and wheat. *Journal of molecular biology*, *183*(3), 499-502.

Lafiandra, D., Masci, S., D'Ovidio, R., & Margiotta, B. (2000). The genetics of wheat gluten proteins: an overview. In *Wheat gluten. Proceedings of the 7th International Workshop Gluten 2000, Bristol, UK, 2-6 April 2000* (pp. 3-10). Royal Society of Chemistry.

Lamacchia, C., Shewry, P. R., Di Fonzo, N., Forsyth, J. L., Harris, N., Lazzeri, P. A., ... & Barcelo, P. (2001). Endosperm-specific activity of a storage protein gene promoter in transgenic wheat seed. *Journal of Experimental Botany*, *52*(355), 243-250.

Lasztity, R. (2017). *The chemistry of cereal proteins*. Routledge.

Lawrence, G. J., & Shepherd, K. W. (1980). Variation in glutenin protein subunits of wheat. *Australian Journal of Biological Sciences*, *33*(2), 221-234.

Lawrence, G. J., MacRitchie, F., & Wrigley, C. W. (1988). Dough and baking quality of wheat lines deficient in glutenin subunits controlled by the Glu-A1, Glu-B1 and Glu-D1 loci. *Journal of Cereal Science*, *7*(2), 109-112.

Levy, A. A., Galili, G., & Feldman, M. (1988). Polymorphism and genetic control of high molecular weight glutenin subunits in wild tetraploid wheat Triticum turgidum var. dicoccoides. *Heredity*, *61*(1), 63.

Li, J. F., Norville, J. E., Aach, J., McCormack, M., Zhang, D., Bush, J., ... & Sheen, J. (2013). Multiplex and homologous recombination–mediated genome editing in Arabidopsis and Nicotiana benthamiana using guide RNA and Cas9. *Nature biotechnology*, *31*(8), 688.

Li, T., Liu, B., Spalding, M. H., Weeks, D. P., & Yang, B. (2012). High-efficiency TALEN-based gene editing produces disease-resistant rice. *Nature biotechnology*, *30*(5), 390.

Liu, J. J., Orlova, N., Oakes, B. L., Ma, E., Spinner, H. B., Baney, K. L., ... & Al-Shayeb, B. (2019). CasX enzymes comprise a distinct family of RNA-guided genome editors. *Nature*, *566*(7743), 218.

Longin CFH, Mühleisen J, Maurer HP et al (2012) Hybrid breeding in autogamous cereals. Theor Appl Genet 125:1087–1096.

MacRitchie, F. (1992). Physicochemical properties of wheat proteins in relation to functionality. In *Advances in food and nutrition research* (Vol. 36, pp. 1-87). Academic Press.

Mali, P., Yang, L., Esvelt, K. M., Aach, J., Guell, M., DiCarlo, J. E., ... & Church, G. M. (2013). RNA-guided human genome engineering via Cas9. *Science*, *339*(6121), 823-826.

Mansueto, P., Soresi, M., Iacobucci, R., La Blasca, F., Romano, G., D'Alcamo, A., & Carroccio, A. (2019). Non-celiac wheat sensitivity: a search for the pathogenesis of a self-reported condition. *Italian Journal of Medicine*, *13*(1), 15-23.

Marcussen, T., Sandve, S. R., Heier, L., Spannagl, M., Pfeifer, M., Jakobsen, K. S., ... & International Wheat Genome Sequencing Consortium. (2014). Ancient hybridizations among the ancestral genomes of bread wheat. *science*, *345*(6194), 1250092.

Marraffini, L.A., and Sontheimer, E.J. (2008). CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. Science 322, 1843–1845.

Masci, S., Lew, E. J. L., Lafiandra, D., Porceddu, E., & Kasarda, D. D. (1995). Characterization of low-molecular-weight glutenin subunits in durum wheat by RP-HPLC and N-terminal sequencing. *Cereal Chemistry*, *72*(1), 100-104.

McCabe, D., & Christou, P. (1993). Direct DNA transfer using electric discharge particle acceleration (ACCELL™ technology). *Plant Cell, Tissue and Organ Culture*, *33*(3), 227-236.

McCullen, C. A., & Binns, A. N. (2006). Agrobacterium tumefaciens and plant cell interactions and activities required for interkingdom macromolecular transfer. *Annu. Rev. Cell Dev. Biol.*, *22*, 101-127.

McFadden, E. S., & Sears, E. R. (1946). The origin of Triticum spelta and its free-threshing hexaploid relatives. *Journal of Heredity*, *37*(3), 81-89.

Miroshnichenko, D. N., Poroshin, G. N., & Dolgov, S. V. (2011). Genetic transformation of wheat using mature seed tissues. *Applied biochemistry and microbiology*, *47*(8), 767-775.

Mojica, F. J., García-Martínez, J., & Soria, E. (2005). Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *Journal of molecular evolution*, *60*(2), 174-182.

Morris, C. F., & Beecher, B. S. (2012). The distal portion of the short arm of wheat (Triticum aestivum L.) chromosome 5D controls endosperm vitreosity and grain hardness. *Theoretical and applied genetics*, *125*(2), 247-254.

Morrison, W. R., (1978). Wheat lipid composition. Cereal Chemistry., Vol 55, pp. 548-558

Nehra, N. S., Chibbar, R. N., Leung, N., Caswell, K., Mallard, C., Steinhauer, L., ... & Kartha, K. K. (1994). Self-fertile transgenic wheat plants regenerated from isolated scutellar tissues following microprojectile bombardment with two distinct gene constructs. *The Plant Journal*, *5*(2), 285-297.

Nekrasov, V., Staskawicz, B., Weigel, D., Jones, J. D., & Kamoun, S. (2013). Targeted mutagenesis in the model plant Nicotiana benthamiana using Cas9 RNA-guided endonuclease. *Nature biotechnology*, *31*(8), 691.

Nishimasu, H., Ran, F. A., Hsu, P. D., Konermann, S., Shehata, S. I., Dohmae, N., ... & Nureki, O. (2014). Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell*, *156*(5), 935-949.

Okita, T. W., Cheesbrough, V., & Reeves, C. D. (1985). Evolution and heterogeneity of the alpha-/beta-type and gamma-type gliadin DNA sequences. *Journal of Biological Chemistry*, *260*(13), 8203-8213.

Ortiz, R., Trethowan, R., Ferrara, G. O., Iwanaga, M., Dodds, J. H., Crouch, J. H., ... & Braun, H. J. (2007). High yield potential, shuttle breeding, genetic diversity, and a new international wheat improvement strategy. *Euphytica*, *157*(3), 365-384.

Osborne, T. B. (1907). *The proteins of the wheat kernel* (No. 84). Carnegie institution of Washington.

Payne, P. I., & Corfield, K. G. (1979). Subunit composition of wheat glutenin proteins, isolated by gel filtration in a dissociating medium. *Planta*, *145*(1), 83-88.

Payne, P. I., & Seekings, J. A. (1996, September). Characterisation of Galahad-6, Galahad-7 and Galahad-8, isogenic lines that contain only one HMW glutenin subunit. In *Proceedings of the 7th International Gluten Workshop"(CW Wrigley, ed.)* (pp. 14-17).

Payne, P. I., Holt, L. M., & Lawrence, G. J. (1983). Detection of a novel high molecular weight subunit of glutenin in some Japanese hexaploid wheats. *Journal of Cereal Science*, *1*(1), 3-8.

Payne, P. I., Jackson, E. A., Holt, L. M., & Law, C. N. (1984). Genetic linkage between endosperm storage protein genes on each of the short arms of chromosomes 1A and 1B in wheat. *Theoretical and Applied Genetics*, *67*(2-3), 235-243.

Peng, J. H., Sun, D., & Nevo, E. (2011). Domestication evolution, genetics and genomics in wheat. *Molecular Breeding*, *28*(3), 281.

Pinthus, M. J., & Levy, A. A. (1983). The relationship between the Rht 1 and Rht 2 dwarfing genes and grain weight in Triticum aestivum L. spring wheat. *Theoretical and Applied Genetics*, *66*(2), 153-157.

Posner, E. S., & Hibbs, A. N. (2005). Wheat Flour Milling, 2nd. *Ed. AACC, St. Paul, Minnesota, USA P*, *489*.

Pourcel, C., Salvignol, G., & Vergnaud, G. (2005). CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology*, *151*(3), 653-663.

Rafalski, J. A., Scheets, K., Metzler, M., Peterson, D. M., Hedgcoth, C., & Söll, D. G. (1984). Developmentally regulated plant genes: the nucleotide sequence of a wheat gliadin genomic clone. *The EMBO journal*, *3*(6), 1409-1415.

Rasco-Gaunt, S., Riley, A., Barcelo, P., & Lazzeri, P. A. (1999). Analysis of particle bombardment parameters to optimise DNA delivery into wheat tissues. *Plant Cell Reports*, *19*(2), 118-127.

Ray, A., & Langer, M. (2002). Homologous recombination: ends as the means. *Trends in plant science*, *7*(10), 435-440.

Razzaq, A., & Masood, A. (2018). CRISPR/Cas9 System: A Breakthrough in Genome Editing. *Mol Biol*, *7*(210), 2.

Risacher, T., Craze, M., Bowden, S., Paul, W., & Barsby, T. (2009). Highly efficient Agrobacterium-mediated transformation of wheat via in planta inoculation. In *Transgenic Wheat, Barley and Oats* (pp. 115-124). Humana Press.

Sanford, J. C. (1988). The biolistic process. *Trends in biotechnology*, *6*(12), 299-302.

Scherf, K. A. (2019). Immunoreactive cereal proteins in wheat allergy, non-celiac gluten/wheat sensitivity (NCGS) and celiac disease. *Current Opinion in Food Science*.

Schropp, P., & Wieser, H. (1996). Effects of high molecular weight subunits of glutenin on the rheological properties of wheat gluten. *Cereal chemistry (USA)*.

Sears, R.G., T.J. Martin, and J.P. Shroyer. 1993. Arlin hard white wheat. Kans. St. Univ. Coop. Ext. Serv. Publ. L- 882, Manhattan.

Sestili, F., Botticella, E., Bedo, Z., Phillips, A., & Lafiandra, D. (2010). Production of novel allelic variation for genes involved in starch biosynthesis through mutagenesis. *Molecular Breeding*, *25*(1), 145.

Sestili, F., Botticella, E., & Lafiandra, D. (2014). TILLING for improved starch composition in wheat. In *Genomics of plant genetic resources* (pp. 467-487). Springer, Dordrecht.

Shan, Q., Wang, Y., Li, J., Zhang, Y., Chen, K., Liang, Z., ... & Gao, C. (2013). Targeted genome modification of crop plants using a CRISPR-Cas system. *Nature biotechnology*, *31*(8), 686.

Shan, Q., Wang, Y., Li, J., Zhang, Y., Chen, K., Liang, Z., ... & Gao, C. (2013). Targeted genome modification of crop plants using a CRISPR-Cas system. *Nature biotechnology*, *31*(8), 686.

Shen, L., Luo, G., Song, Y., Song, S., Li, Y., Yang, W., ... & Zhang, A. (2018). Low molecular weight glutenin subunit gene composition at Glu-D3 loci of Aegilops tauschii and common wheat and a further view of wheat evolution. *Theoretical and Applied Genetics*, *131*(12), 2745-2763.

Shewry, P. R. (2009). Wheat. *Journal of experimental botany*, *60*(6), 1537-1553.

Shewry, P. R., & Tatham, A. S. (1990). The prolamin storage proteins of cereal seeds: structure and evolution. *Biochemical journal*, *267*(1), 1.

Shewry, P. R., & Tatham, A. S. (1997). Disulphide bonds in wheat gluten proteins. *Journal of Cereal Science*, *25*(3), 207-227.

Shewry, P. R., & Tatham, A. S. (1999). The characteristics, structures and evolutionary relationships of prolamins. In *Seed Proteins* (pp. 11-33). Springer, Dordrecht.

Shewry, P. R., D'Ovidio, R., Lafiandra, D., Jenkins, J. A., Mills, E. C., & Békés, F. (2009). Wheat grain proteins. *Wheat: chemistry and technology*, (Ed. 4), 223-298.

Shewry, P. R., Halford, N. G., & Lafiandra, D. (2003). Genetics of wheat gluten proteins. In *Advances in genetics* (Vol. 49, pp. 111-184). Academic Press.

Shewry, P. R., Halford, N. G., & Tatham, A. S. (1992). High molecular weight subunits of wheat glutenin. *Journal of cereal science*, *15*(2), 105-120.

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*(2), 97-106.

Shimada, T. 1978. Plant regeneration from callus induced from wheat embryos. Jpn. J. Genet. 53: 371–374.

Shimada, T. and Yamada, Y. 1979. Wheat plants regenerated from embryo cell cultures. Jpn. J. Genet. 54:379–385.

Song, G., Jia, M., Chen, K., Kong, X., Khattak, B., Xie, C., ... & Mao, L. (2016). CRISPR/Cas9: a powerful tool for crop genome editing. *The crop journal*, *4*(2), 75-82.

Stephens, J., & Barakate, A. (2017). Gene editing technologies–ZFNs, TALENs, and CRISPR/Cas9.

Stone, B., Morell, M.K., (2009). Carbohydrates. In: Khan K., Shewry P.R. (Eds.). Wheat: chemistry and technology. AACC International, Inc., St. Paul, MN., pp. 299-362.

Sugawara, T., & Miyazawa, T. (2001). Beneficial effect of dietary wheat glycolipids on cecum short-chain fatty acid and secondary bile acid profiles in mice. *Journal of nutritional science and vitaminology*, *47*(4), 299-305.

Supartana, P., Shimizu, T., Nogawa, M., Shioiri, H., Nakajima, T., Haramoto, N., ... & Kojima, M. (2006). Development of simple and efficient in planta transformation method for wheat (Triticum aestivum L.) using Agrobacterium tumefaciens. *Journal of bioscience and bioengineering*, *102*(3), 162-170.

SURGET, A., & BARRON, C. (2005). Histologie du grain de blé. *Industries des céréales*, (145), 3-7.

van Herpen, T. W., Goryunova, S. V., van der Schoot, J., Mitreva, M., Salentijn, E., Vorst, O., ... & Vosman, B. (2006). Alpha-gliadin genes from the A, B, and D genomes of wheat contain different sets of celiac disease epitopes. *BMC genomics*, *7*(1), 1.

Vasil, V., Castillo, A. M., Fromm, M. E., & Vasil, I. K. (1992). Herbicide resistant fertile transgenic wheat plants obtained by microprojectile bombardment of regenerable embryogenic callus. *Bio/technology*, *10*(6), 667.

Waines, J. G., & Payne, P. I. (1987). Electrophoretic analysis of the high-molecular-weight glutenin subunits of Triticum monococcum, T. urartu, and the A genome of bread wheat (T. aestivum). *Theoretical and Applied Genetics*, *74*(1), 71-76.

Wall, A.M., R. Riley, and M.D. Gale. 1971. The position of a locus of chromosome 5B in Triticum aestivum affecting homoeologous meiotic pairing. Genet. Res. 18:329–339.

Weegels, P. L., Hamer, R. J., & Schofield, J. 1. (1996). Functional properties of wheat glutenin. *Journal of Cereal Science*, *23*(1), 1-17.

Weeks, J. T., Anderson, O. D., & Blechl, A. E. (1993). Rapid production of multiple independent lines of fertile transgenic wheat (Triticum aestivum). *Plant Physiology*, *102*(4), 1077-1084.

Whitford R, Fleury D, Reif JC et al (2013) Hybrid breeding in wheat: technologies to improve hybrid wheat seed production. J Exp Bot 64:5411–5428.

Wieser, H. (2007). Chemistry of gluten proteins. *Food microbiology*, *24*(2), 115-119.

Wilson, A.S. 1876. On wheat and rye hybrids. Trans. Proc. Bot. Soc. Edinb. 12:286–288.

Woychik, J. H., Boundy, J. A., & Dimler, R. J. (1961). Starch gel electrophoresis of wheat gluten proteins with concentrated urea. *Archives of Biochemistry and Biophysics*, *94*(3), 477-482.

Wrigley, C. W. (1996). Giant proteins with flour power. *Nature*, *381*(6585), 738.

Xie, Z., Wang, C., Wang, K., Wang, S., Li, X., Zhang, Z., ... & Yan, Y. (2010). Molecular characterization of the celiac disease epitope domains in α-gliadin genes in Aegilops tauschii and hexaploid wheats (Triticum aestivum L.). *Theoretical and applied genetics*, *121*(7), 1239-1251.

Zale, J. M., Agarwal, S., Loar, S., & Steber, C. M. (2009). Evidence for stable transformation of wheat by floral dip in Agrobacterium tumefaciens. *Plant cell reports*, *28*(6), 903-913.

Zetsche, B., Gootenberg, J. S., Abudayyeh, O. O., Slaymaker, I. M., Makarova, K. S., Essletzbichler, P., ... & Koonin, E. V. (2015). Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell*, *163*(3), 759-771.

Zhang, Y., Zhang, F., Li, X., Baller, J. A., Qi, Y., Starker, C. G. & Voytas, D. F. (2013). Transcription activator-like effector nucleases enable efficient plant genome engineering. *Plant physiology*, *161*(1), 20-27.

CHAPTER 2:

PRODUCTION AND MOLECULAR CHARACTERIZATION OF BREAD WHEAT LINES WITH REDUCED AMOUNT OF α-TYPE GLIADINS

Camerlengo et al. BMC Plant Biology (2017) 17:248
DOI 10.1186/s12870-017-1211-3

BMC Plant Biology

RESEARCH ARTICLE Open Access

Production and molecular characterization of bread wheat lines with reduced amount of α-type gliadins

Francesco Camerlengo[1], Francesco Sestili[1], Marco Silvestri[2], Giuseppe Colaprico[3], Benedetta Margiotta[3], Roberto Ruggeri[1], Roberta Lupi[1], Stefania Masci[1] and Domenico Lafiandra[1*]

Abstract

Background: Among wheat gluten proteins, the α-type gliadins are the major responsible for celiac disease, an autoimmune disorder that affects about 1% of the world population. In fact, these proteins contain several toxic and immunogenic epitopes that trigger the onset of the disease. The α-type gliadins are a multigene family, encoded by genes located at the complex *Gli-2* loci.

Results: Here, three bread wheat deletion lines (Gli-A2, Gli-D2 and Gli-A2/Gli-D2) at the *Gli-2* loci were generated by the introgression in the bread wheat cultivar Pegaso of natural mutations, detected in different bread wheat cultivars. The molecular characterization of these lines allowed the isolation of 49 unique expressed genes coding α-type gliadins, that were assigned to each of the three *Gli-2* loci.
The number and the amount of α-type gliadin transcripts were drastically reduced in the deletion lines. In particular, the line Gli-A2/Gli-D2 contained only 12 active α-type gliadin genes (−75.6% respect to the cv. Pegaso) and a minor level of transcripts (−80% compared to cv. Pegaso). Compensatory pleiotropic effects were observed in the two other classes of gliadins (ω- and γ-gliadins) either at gene expression or protein levels.
Although the comparative analysis of the deduced amino acid sequences highlighted the typical structural features of α-type gliadin proteins, substantial differences were displayed among the 49 proteins for the presence of toxic and immunogenic epitopes.

Conclusion: The deletion line Gli-A2/Gli-D2 did not contain the 33-mer peptide, one of the major epitopes triggering the celiac disease, representing an interesting material to develop less "toxic" wheat varieties.

Keywords: Gluten, α-gliadins, Celiac disease, Bread wheat, Deletion lines

Background

With a surface of 220 M of hectares, wheat represents the most widely cultivated crop and a major component of the human diet. This is the result of its agronomic adaptability, ease of storage, nutritional value and ability of its flour to produce a large variety of palatable and satisfying foods [1]. It is firmly established that functional properties of doughs, necessary to realize different types of end products are mainly related to gluten proteins. The main constituents of the gluten are gliadins and glutenins that represent the storage proteins, accumulated in the endosperm during the graining phase and used by the growing embryo during the first phases of germination. Gliadins and glutenins constitute about 80% of the total proteins and are classified as "prolamins" because of their high content in proline and glutamine [2]. Glutenins are polymeric proteins, whose constituents, high and low molecular weight glutenin subunits (HMW-GS and LMW-GS, respectively), are held together by intermolecular disulfide bonds [3, 4]. Genes encoding these subunits are present at the *Glu-1* and *Glu-3* loci, present on the long and short arm of the homeologous group 1 chromosomes, respectively [4]. Differently, gliadins are monomeric proteins, classified

* Correspondence: lafiandr@unitus.it
[1]Department of Agriculture and Forestry Sciences, University of Tuscia, 01100 Viterbo, Italy
Full list of author information is available at the end of the article

in four groups, known as α-, β-, γ- and ω-gliadins, on the basis of their decreasing mobility in electrophoresis at acid pH [5]. Subsequent comparisons of amino acid and DNA sequences have demonstrated that the separation of gliadins in four groups is not correct as α- and β-gliadins share similar sequences leading Kasarda et al. [6] to suggest that, from a structural point of view, gliadins could be subdivided into three groups: α-, γ- and ω-type.

The genes encoding ω- and γ-type gliadins are located at three complex *Gli-1* loci (*Gli-A1*, *Gli-B1* e *Gli-D1*) present on the short arms of the homeologous chromosomes of group 1, whereas α-type are controlled by genes located at the *Gli-2* loci (*Gli-A2*, *Gli-B2* e *Gli-D2)* on the short arms of the homeologous chromosomes of group 6 [7, 8].

Within the same *Gli-1* and *Gli-2* loci, genes encoding for mutant form of gliadins are also present [9, 10]. They are highly similar in sequence to individual α-, γ- and ω-type gliadins, and are probably derived from these components by point mutations resulting in the presence of an additional cysteine residue. The extra cysteine residue is capable to form intermolecular disulfide bonds resulting in incorporation of the modified gliadin into the glutenin polymers where it can act as chain terminator and limiting their molecular weight distribution [3, 11].

It is estimated that the average number of copies of α-gliadin genes per haploid genome ranges from 25–35 to 100–150 copies; this large number has been generated by duplication, deletion events and retrotransposon insertion [12–14].

At least half of α-type gliadin genes are considered inactive or pseudogenes because they show premature stop codons in the coding sequences [15–17]. From a quantitative point of view the proportions of the different classes of gliadins fluctuate considerably among different varieties, although the α-type gliadins are generally present in larger amount, followed by the γ- and ω-type gliadins [18].

Several studies have shown that peptides derived from incomplete digestion of all gluten proteins, but in particular those of the α-type gliadins, are the main responsible for the onset of celiac disease [17, 19, 20]. Celiac disease is a permanent food disorder caused by an abnormal autoimmune response to dietary gluten peptides that occurs in genetically susceptible individuals carrying T-cell with HLA-DQ2 or -DQ8 receptors [21–23]. The autoimmune component is demonstrated by serologic autoantibodies such as serum anti-tissue transglutaminase (tTG) and anti-endomysial antibodies (EMA) [24]. The partially digested peptides contain epitopes that are involved in the disease processes in different manner and they are accordingly classified into two groups: toxic and immunogenic epitopes. Toxic peptides can induce mucosal damage when added in culture to duodenal endoscopic biopsy or if administered in vivo, whereas the peptides defined immunogenic stimulate HLA-DQ2- or DQ8-restricted T cell clones isolated from jejunal mucosa or peripheral blood of coeliac patients [25]. However, it is not excluded that some peptides can perform both functions, toxic and immunogenic.

Several studies have demonstrated that the large genetic diversity existing at the different loci encoding gliadin components is reflected in a large variation in the number of T-cell–stimulatory peptides present in different durum and bread wheat accessions as well as in diploid progenitors of polyploid wheats [17, 26–28]. In the past years, several lines and cultivars have been identified in durum and bread wheat, characterized by the deletion of genes associated to the *Gli-1*, *Gli-2* and *Glu-1* loci [29, 30].

In this paper, we report the production of bread wheat lines with reduced amount of α-type gliadins, and the characterization of their corresponding genes, with the aim to produce wheat varieties with a reduced content in immunogenic and toxic peptides.

Methods

Plant material and experimental site

Bread wheat lines lacking gliadin components at the three homeologous *Gli-1* loci and two lines without the gliadins associated at the *Gli-A2* and *Gli-D2* loci, previously identified in different bread wheat cultivars, have been initially used to develop the lines used in this work (Table 1). In order to have the single deletion lines in the same genetic background, they were crossed and backcrossed four times with the Italian bread wheat cultivar Pegaso. The gliadins present in the seeds obtained after each backcross were analyzed by one dimensional polyacrylamide gel electrophoresis in acidic conditions (A-PAGE) and those lacking the gliadins, associated at one of the five above mentioned loci, were used for the next step of backcrossing. BC_4F_6 seeds were used for all the analyses. The single deletion lines obtained from the backcross with the bread wheat cv Pegaso were crossed between them, thus, besides single Gli-A2, or Gli-D2 deletion lines, also double Gli-A2/Gli-D2 deletion lines were isolated.

The three BC_4F_6 deletion lines and the cultivar Pegaso were grown in open field during 2014/2015 growing

Table 1 List of deletion mutant lines at the *Gli-1* and *Gli-2* loci

Locus	Cultivar	Country	Reference
Gli-A1	Saratovskaja 29	USSR	Redaelli et al. [30]
Gli-B1	Spada	Italy	Lafiandra et al. [31]
Gli-D1	Darius	France	Branlard et al. [32]
Gli-A2	Reader	USA	Lafiandra et al. [29]
Gli-D2	Saratovskaja 29	USSR	Redaelli et al. [30]

season at the Experimental Farm of the University of Tuscia, located in Viterbo, Italy (lat. 42°26' N, long. 12° 04' E, altitude 310 m a.s.l.). The experimental site is characterized by a Mediterranean climate with mean annual maximum and minimum temperatures of 19 and 8 °C, respectively, and annual rainfall of 743 mm.

Electrophoretic analysis of gliadins

Monomeric gliadins were extracted from single kernel and fractionated by A-PAGE as described by Pflüger et al. [33]. Two-dimensional (two-pH) polyacrylamide gel electrophoresis was carried out as described by Lafiandra and Kasarda [34].

Isolation of genes coding α-type gliadin proteins expressed in grain

Genes coding α-type gliadins were amplified by Reverse transcriptase-PCR. Total RNA was extracted from immature kernels (24 Day Post Anthesis, DPA) using the Spectrum Plant Total RNA kit (Sigma-Aldrich, St. Louis, USA) and following manufacturer's instructions. One μg of RNA constituted the template for the synthesis of cDNA, based on the QuantiTect Reverse Transcription Kit (Qiagen, Hilden, Germany). Each PCR reaction were performed in a volume of 20 μL containing 10 μL of GoTaq Green Master Mix (Promega), 0.5 μM of each primer (L1 and R1, Table 2) and 1 μL of cDNA. PCR conditions were: an initial denaturation at 95 °C for 5 min, followed by 35 cycles at 95 °C for 1 min, 61 °C for 30 s,72 °C for 1 min and a final extension at 72 °C for 5 min. PCR amplicons were checked on 1% agarose gel electrophoresis.

PCR amplicons were recovered from gel by NucleoSpin® Gel and PCR Clean-up kit (Macherey-Nagel, Duren, Germany) and the purified products were ligated into pGEM-T Easy vector (Promega, Madison, WI, USA). JM109 Competent Cells (Promega, Madison, WI, USA) were transformed with the ligation reaction. Transformed cells were plated onto LB medium plus Ampicillin containing X-Gal and IPTG (Sigma-Aldrich,) for blue/white screening.

White colonies were inoculated in LB medium plus Ampicillin, and plasmids were extracted with the NucleoSpin® Plasmid kit (Macherey-Nagel, Duren, Germany). The presence of the insert was verified by digesting about one μg of each recombinant plasmid for 2 h with 5 units of EcoRI restriction enzyme (Promega, Madison, WI, USA). Digested plasmids were checked on an agarose gel (0.8%).

DNA sequencing was performed with the universal primers T7 and SP6 by Eurofins Genomics (Ebersberg, Germany).

Full length sequences of α-gliadin genes were reconstructed by overlapping the fragments with the bioinformatic program Geneious R8. Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/) was used to align and compare the deduced amino acid sequences.

Phylogenetic analysis

The phylogenetic tree was generated by Neighbor Joining method using MEGA7 software package (http://www.megasoftware.net/) [39].

Table 2 List of primer pairs used to isolate the α-type gliadins and in RT-PCR analysis

Primer name	Sequence	Reference
L1	5'-ATGAAGACCTTTCTCATCCTTG	Xie et al. [35]
R1	5'-TCAGTTRGTACCRAAGATGCC	
RT a-gli F	5'-AGACCTTTCTCATCCTTGCC	Li et al. [36]
RT a-gli R	5'-TGTACCAATGGAACTTGCTCT	
G1Gli*1	5'-ACAGGTGAACCCATGCAAGAATTT	Piston et al. [37]
G1Gli*2	5'-TGCATGATGATGGAATGTATGATGG	
G2Gli*1	5'-TCATTCCCCCAACAACAACGG	Piston et al. [37]
G2Gli*2	5'-AGGTTTGCATTGTTGCAAGAGGAT	
G3 Gli*1	5'-GCAAATCCTGGTGCCACTGTCTCAA	Piston et al. [37]
G3 Gli*2	5'-TGGGACATACACGTTGCACATGGTT	
G4 Gli*1	5'-GATCCTGCGGCCACTATTTCAGCTC	Piston et al. [37]
G4 Gli*2	5'-CAGGTGGCACATACACGTTGCACAT	
QF18	5'-AAGGCAAGCAAGCAGTAG	Altenbach et al. [38]
QR18	5'-GATTGTTGAGGTGATTGTAGC	
QF21	5'-CAACCACCACAACAATTC	Altenbach et al. [38]
QR21	5'-TTACATCTCTTCATTTCATAGG	

Quantification of gliadin transcripts

Real-time PCR was performed on total RNA using primer pairs specific for α-, γ and ω-gliadins (Table 2). The analyses were performed in the CFX 96 Real-Time PCR Detection System device (Bio-Rad Hercules, CA, USA). The reactions were carried out in a final volume of 15 μL, consisting in 7.5 μL SsoAdvUniver SYBR GRN SMX (Bio-Rad Hercules, CA, USA), 0.5 μM of each primer and 1 μL of cDNA, following the protocol below: 94 °C for 30 s and 40 cycles at 94 °C for 5 s, 60 °C for 30 s and melt curve 65–95 °C with 0.5 °C increment 5 s/step. Three technical replicates per biological sample and three independent plants per genotype were analyzed. The quantification analysis was performed as described in Sestili et al. [40].

Results

Separation of gliadins by one- and two-dimensional electrophoresis

The utilization of the three different deletion lines, without the *Gli-1* loci (*Gli-A1*, *Gli-B1*, *Gli-D1*) and the two without the *Gli-2* loci (*Gli-A2* and *Gli-D2*), all backcrossed four times with the bread wheat cv Pegaso, have allowed the isolation of five single deletion lines. The five lines obtained differ for the absence of different clusters of gliadin components encoded by genes present at the *Gli-1* or *Gli-2* locus, whereas the remaining gliadin genes are the same as those present in Pegaso.

One dimensional electrophoretic separations of gliadin components present in the bread wheat cultivar Pegaso and derived single and double deletion lines (Gli-A2, Gli-D2 and Gli-A2/Gli-D2) are reported in Fig. 1. Three independent single seeds were analyzed for each deletion line. Comparison of the electrophoretic separations of gliadins present in Pegaso, with single and double deletion lines, allowed to assign gliadin components to the three different *Gli-2* loci. Although the gliadins were extracted from single kernels for each genotype and the same volume was loaded on the gel, an increase of intensity of the other gliadin fractions was observed in the deletion lines and this was more marked in the double deletion genotype (Fig. 1).

Two dimensional electrophoretic separations have a higher resolving power of a complex mixture of proteins as is the case for the gliadin fraction present in wheat. Using the two dimensional separation at two different pH [34] a more complete chromosomal assignment of the gliadins associated at the *Gli-2* loci was possible (Fig. 2). Though the deletion line involving the *Gli-B2* locus was not available, components associated at this locus were also identified, after assignment of gliadin components at the three *Gli-1* loci (data not shown), along with those associated at the *Gli-A2* and *Gli-D2* loci.

A total of seventeen gliadins were found associated to the *Gli-2* loci, seven to the *Gli-A2*, six to the *Gli-D2* and

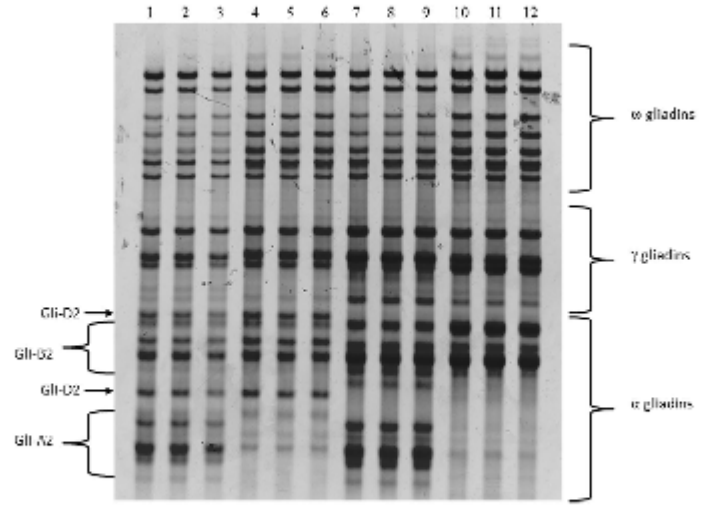

Fig. 1 A-PAGE separation of gliadin fractions. Proteins were extracted from kernels of Pegaso (1–3), Gli-A2 deletion line (4–6); Gli-D2 deletion line (7–9) and Gli-A2/Gli-D2 double deletion line (10–12)

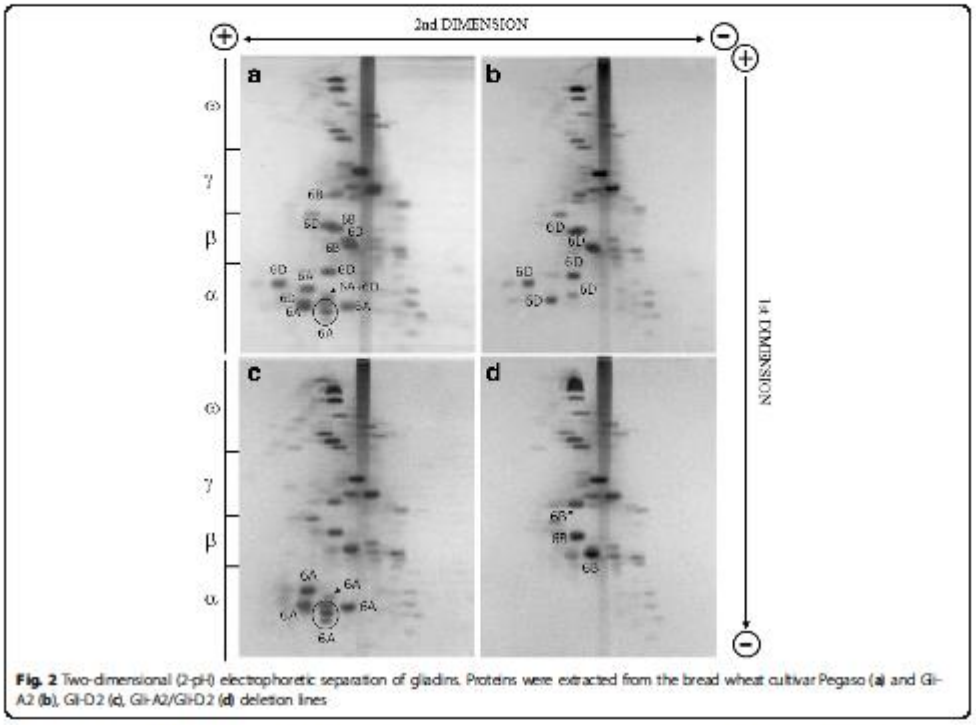

Fig. 2 Two-dimensional (2-pH) electrophoretic separation of gliadins. Proteins were extracted from the bread wheat cultivar Pegaso (**a**) and Gli-A2 (**b**), Gli-D2 (**c**), Gli-A2/Gli-D2 (**d**) deletion lines

three to the *Gli-B2* locus. It is very likely that these numbers are an underestimation since it has been reported that single protein spots could derive from different genes because of their close similarity [7].

Isolation and genetic localization of α-type gliadins expressed in the bread wheat cv Pegaso

The transcripts of α-type gliadin genes were isolated by Reverse transcriptase-PCR on immature caryopsis (24 DPA) of the bread wheat cv Pegaso and derived deletion lines at the *Gli-A2*, *Gli-D2* and *Gli-A2/Gli-D2* loci. Full-length ORFs were amplified using specific oligonucleotides (L1 and R1) designed on highly conserved regions previously described by Xie et al. [35]. The PCR products were checked on agarose gel (data not shown) and ranged from 800 to 950 base pairs. After purification, the amplicons were cloned into pGEM T-easy vector. In total, 381 sequences of α-type gliadins were identified and sequenced. In particular, 104 sequences were isolated from the bread wheat cv Pegaso; 101, 107 and 69 were isolated from the deletion lines Gli-A2, Gli-D2 and Gli-A2/Gli-D2, respectively. Several chimeric genes, probably produced during PCR amplification, and some pseudogenes were observed and discarded. Blast analysis on NCBI database highlighted that the sequences isolated from the deletion lines Gli-A2 and Gli-D2 had a high similarity with genes coding α-type gliadins from *Triticum aestivum*, *Triticum urartu*, *Aegilops speltoides* and *Aegilops tauschii* species; while the sequences deriving from the double deletion line and associated to the *Gli-B2* locus showed the higher similarity with genes coding α-type gliadins isolated from *Aegilops speltoides*; this is not unexpected, as it has been suggested that an unknown close relative of *Aegilops speltoides* donated the B genome to polyploid wheats [41].

After removing either redundant genes and chimeric/pseudogenic sequences, 49 unique genes were isolated from the cv Pegaso and submitted to the EMBL database (from the acc. n° LT627562 to LT627610). Considering together the data obtained by the sequencing of the four genotypes, all the full-length sequences were assigned to one of the homoeologous *Gli-2* loci: 18 sequences to the *Gli-A2*, 12 to the *Gli-B2* and 19 to the *Gli-D2*. Alignment of the isolated α-type-gliadins is reported in Additional file 1: Figure S1. All α-type gliadin genes isolated lack introns and their size range from 846 to 960 bp.

Analysis of amino acidic sequences of the α-type gliadins isolated from the bread wheat cv Pegaso

The 49 deduced amino acidic sequences were aligned through the bioinformatic tool Clustal Omega (EMBL-EBI) (Aditional file 2: Figure S2). All the sequences had the typical features of α-type gliadins, characterized from a signal peptide (P) and five domains: 1) N-terminal domain (R1); 2) Polyglutamine domain 1 (QR1); 3) Unique domain 1 (NR1); 4) Polyglutamine domain 2 (QR2); 5) C-terminal unique domain (NR2). In Additional file 3: Table S1 the number of amino acids present in each domain of the 49 α-type gliadins isolated from the bread wheat cv Pegaso is reported. Size of the deduced proteins ranges from 281 to 319 amino acids, whereas the molecular weight ranges from 30.4 to 34.7 kDa. The signal peptide is strongly conserved in size (20 aa) and composition in all the sequences. Similarly the domains NR1 and NR2 resulted highly conserved among the different genomes and ranged from 68 to 69 and 76 to 78 aa, respectively.

The α-type gliadins assigned to the *Gli-B2* locus, with the exception of Gli-B2–1, Gli-B2–8 and Gli-B2–9, are larger in size compared to those assigned to the other two loci. In particular, their size (including the signal peptide) ranged from 312 to 319 aa; differently, α-type gliadins assigned to the *Gli-A2* and *Gli-D2* loci ranged from 283 to 294 and from 282 to 309 aa, respectively (Aditional file 3: Table S1; Additional file 2: Figure S2). These differences were mainly due to the length of the two polyglutamine domains (QR1 and QR2) and the N-terminal domain (R1); in fact, QR1 was highly polymorphic among the different α-type gliadins and ranged from 18 to 29 aa for the *Gli-A2* locus, from 21 to 33 for the *Gli-B2* and from 12 to 25 for the *Gli-D2*. Similarly, the QR2 domain was different in length for the proteins associated to the *Gli-B2* (7–33) and *Gli-D2* (8–18) loci, while was identical in size (8 aa) for the α-type-gliadins associated to the *Gli-A2* locus (Additional file 3: Table S1).

InDels polymorphism was also observed in the R1 domain, except for the sequences associated to the *Gli-A2* locus, that were highly conserved (92 aa) (Additional file 3: Table S1). This domain was higher in size in the α-type gliadins associated to the *Gli-D2* locus compared to those of the other two loci. In particular, a peptide of seven amino acids (QPQLPYP), present from one to three times in sixteen genes out of the nineteen possible, is responsible for this difference (Additional file 2: Figure S2).

All the 18 sequences associated to the *Gli-A2* locus resulted more conserved among them than those of the other two loci; in particular the size of the 4 domains (R1, NR1, QR2 and NR2) was identical in the 18 α-type-gliadins; only the QR1 varied in size (from 18 to 29 aa). Differently for the proteins associated to the other two loci, the NR1 and NR2 domains were conserved, while R1 and QR2, as above reported, changed in length among the different sequences (Additional file 3: Table S1).

The position and number of cysteine residues (in total 6) were highly conserved in the deduced proteins except for the six sequences (Gli-B2–1, Gli-D2–4, Gli-D2–5, Gli-D2–6, Gli-D2–7 and Gli-D2–8) which showed the presence of an additional cysteine residue in the NR2 domain. The extra cysteine is the result of a point mutation in the codon TCC, coding for the amino acid serine, to TGC. The polymorphism was detected in the same position in all sequences above cited (Additional file 2: Figure S2).

Phylogenetic analysis of α-type gliadins isolated from the bread wheat cv Pegaso

The phylogenetic relationships among the 49 α-type gliadins were investigated on the basis of the predicted amino acidic sequences. This analysis showed that it is possible to distinguish the α-type gliadins on the basis of their genomic origin (Fig. 3). In fact, the phylogenetic tree revealed that the deduced proteins were grouped in three large clusters (I, II and III): all the sequences associated at the *Gli-A2* locus were assigned to the cluster I; whereas, those associated at the *Gli-B2* and *Gli-D2* loci were assigned to the cluster II and III, respectively. Only the sequence Gli-B2–1 did not cluster with other sequences in the phylogenetic tree, but it was located at an intermediate position between the clusters I and II.

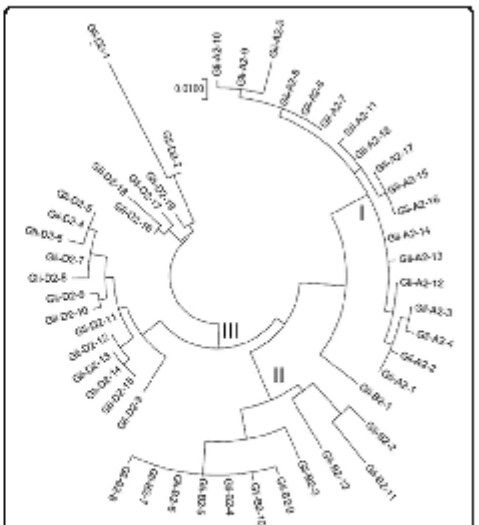

Fig. 3 Phylogenetic relationships among the α-type gliadins isolated from the cultivar Pegaso. The tree was constructed based on neighbor-joining analysis using the entire coding regions

Analysis of gene expression of the ω-, γ- and α-type gliadins

To investigate at molecular level the reduction of α-type gliadin transcripts and to confirm the possible pleiotropic effects on the other classes of gliadins observed at protein level (Fig. 1), a Real Time PCR experiment was carried out on mRNA extracted from 24 DPA immature kernels of Pegaso, the two single deletion lines and the double one. Seven primer pairs were used: one specific for genes encoding α-type gliadins, four for γ-gliadins (G1, G2, G3, G4) [37] and two for ω-gliadins (Q18, Q21) [38]. Data shown in Fig. 4 highlighted that the single deletion lines (Gli-A2 and Gli-D2) had a significant reduction of the expression of α-type gliadins, that ranged from 25% in the Gli-A2 line to 62% in the Gli-D2 one. A more drastic reduction (80%) was observed in the double deletion line Gli-A2/Gli-D2. These results indicated that the genes located on the locus *Gli-D2* were more expressed or in a higher number of copies compared to the *Gli-A2* and *Gli-B2* loci.

The primer pairs, designed by Piston et al. [37], were used to quantify the levels of the four classes of γ-gliadin transcripts in immature grains. No changes were detected in the single deletion line Gli-A2; differently, these genes were up-regulated (up to 6 fold) in the other two deletion lines (Gli-D2 and Gli-A2/Gli-D2) (Fig. 4). A significant increase, but less evident, was also observed for the gene expression of the ω-gliadins in the genotype Gli-A2/Gli-D2, whereas no significant differences were observed in the single deletion lines (Fig. 4). All together, these data confirmed the existence of a compensatory mechanism as a consequence of the α-type gliadins down-regulation, in particular, when the locus *Gli-D2* was absent.

Analysis of toxic and immunogenic peptides present in the 49 α-gliadins expressed in bread wheat cv Pegaso

The presence of toxic and immunogenic epitopes was investigated in the isolated α-gliadin genes. The analysis was focused on seven toxic epitopes previously described [25] and on five dominant immunogenic epitopes [36] (Additional file 4: Table S2 and Additional file 5: Table S3). Six toxic peptides, α-Glia (31–43), α-Glia (31–49), α-Glia (31–55), α-Glia (44–55), α-Glia (51–70) and α-Glia (56–75), were found in the N-terminal repetitive domain; whereas, the epitope α-Glia (206–217) was localized in the C-terminal domain. The position of the seven toxic epitopes was highly conserved, but a specificity in epitope occurrence on α-type gliadins assigned to the different genomes was observed. In particular the toxic peptides α-Glia (31–43) was specific of six and sixteen α-type gliadins present at the *Gli-B2* and *Gli-D2* loci, respectively. Similarly, the epitope α-Glia (31–49) was not detected in the proteins associated to the *Gli-A2* and was observed in fourteen α-type gliadins (five associated to the *Gli-B2* and nine to the *Gli-D2*).

The peptides α-Glia (31–55) and α-Glia (206–217) were specific for the D and A genome, respectively. The first one was present in nine sequences; the last one was

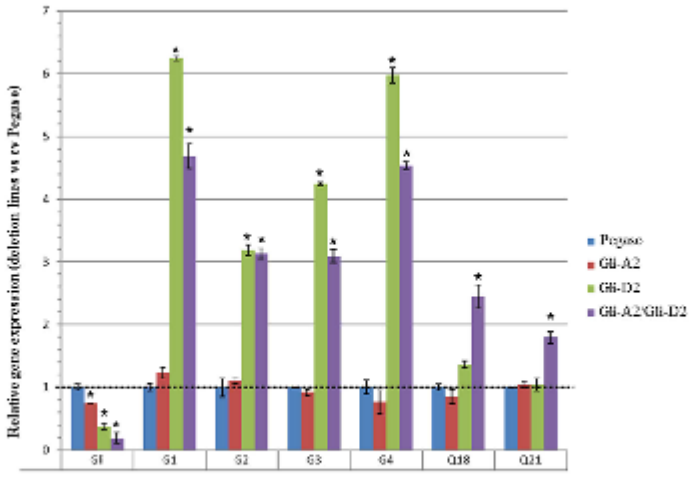

Fig. 4 Analysis of relative gene expression of α- (Gl), γ- (G1, G2, G3 and G4) and ω-type (Q18 and Q21) gliadin transcripts in immature grain of bread wheat cv Pegaso and derived Gli-2 deletion lines. Dotted line indicates the relative transcription value of the control (cv. Pegaso). Each bar corresponds to the standard error, along with an asterisk to indicate where the deletion line values differed significantly ($P < 0.05$) from that of the wild type

detected in all the α-type gliadins of the *Gli-A2* locus. Two epitopes (α-Glia (44–55) and α-Glia (51–70)) were detected in all the three *Gli-2* loci, although they were more abundant in the α-type gliadins of the *Gli-A2* and *Gli-D2* loci compared to those of *Gli-B2*. α-Glia (56–75) was characteristic of five α-type gliadins of the *Gli-D2* locus: it was not observed in any sequences assigned to the *Gli-A2* and *Gli-B2* loci (Additional file 4: Table S2).

The presence of the five immunogenic epitopes (Glia-α, Glia-α2, Glia-α9, Glia-α20 and 33-mer), able to specifically stimulate HLA-DQ2- or DQ8-restricted T cell clones, was investigated in deduced protein sequences of the 49 α-gliadins isolated from the cultivar Pegaso (Additional file 5: Table S3).

Glia-α9 and Glia-α20 were the most abundant, being identified in 30 and 25 sequences, respectively. Noteworthy, these epitopes were present only in one sequence produced from the *Gli-B2* locus. The Glia-α2 peptide was typical of α-gliadins associated at the D genome, being completely absent in those associated with the *Gli-A2* and *Gli-B2* loci. In particular, it was detected in 13 sequences out of 19 associated with the *Gli-D2* locus.

Gli-α was present in 24 sequences (13 associated with the D genome and 11 with the B genome): this epitope was not detected in α-gliadins deriving from the *Gli-A2* locus.

The 33-mer immunogenic peptide, that is considered the principal responsible of the development of the gluten dependent lesion since it is the most immunogenic powerful inducer [42], was completely absent in the α-gliadins assigned to *Gli-A2* and *Gli-B2* loci; whereas it was detected in four proteins assigned to *Gli-D2* locus.

The comparison of deduced amino acid sequences between Pegaso and derived Gli-2 deletion lines highlighted that these latter had a reduced number of toxic and immunogenic epitopes (Table 3). This reduction was either quantitative or qualitative: the deletion lines had less α-type gliadin proteins (31 in the Gli-A2 genotype; 30 in Gli-D2 and 12 in Gli-A2/Gli-D2) compared to Pegaso (49); moreover, as some epitopes were specific of one of the three *Gli-2* loci, each deletion lines had lost some of these peptides. More in detail, the immunogenic peptides Glia-α2 and 33-mer and the toxic peptides α-Glia (31–55) and α-Glia (56–75) were completely absent in the Gli-D2 and Gli-A2/Gli-D2 deletion lines (Table 3). Similarly, the toxic peptide α-Glia (206–217) associated to the *Gli-A2* locus was not detected in the α-type gliadins present in the two genotypes lacking this locus (Gli-A2 and Gli-A2/Gli-D2 deletion lines) (Table 3). In conclusion, the total number of the immunogenic and toxic epitopes detected in the Gli-A2/Gli-D2 deletion line was, respectively, 13 and 17, drastically reduced if compared to the cultivar Pegaso, where 96 immunogenic and 111 toxic peptides were present.

Discussion

Copy number of expressed α-type gliadin genes and genetic localization

Genes coding α-gliadins present in the bread wheat cultivar Pegaso and the deletion lines were isolated and characterized. The analysis of α-gliadin transcripts leads to the isolation of 49 unique sequences. This number is congruent with that previously reported for other bread wheat cultivars [16, 43].

Although estimated copy numbers of α-type gliadins are consistently higher (from 100 up to 150) in genomic DNA [13], it was estimated that half of the α-type-gliadin genes are pseudogenes and thus not expressed [15]. In agreement with this estimation, Noma and colleagues [43] found that 40 out of 90 genes isolated from the bread wheat cultivar Chinese Spring (ca. 44%) had a premature stop codon.

Here, all the 49 α-type-gliadin genes were associated to one of the three *Gli-2* loci. The *Gli-B2* locus contains a lower number of genes (12) compared to the *Gli-A2* (18) and *Gli-D2* (19) loci. This result agrees with that observed by Salentijn et al. [44] that found a low amount of *Gli-B2* sequences in two landraces and in the bread wheat cultivar Lavett. However, the same authors analyzed several tetraploid and hexaploid wheat varieties and demonstrated that the copy number of expressed α-type gliadins associated to each *Gli-2* locus is extremely variable, either among the wheat species or within the same species among different varieties. For example,

Table 3 Number of deduced proteins and immunodominant/toxic epitopes detected in the bread wheat cv Pegaso and Gli-2 deletion lines

Genotypes	n° proteins	Immunodominant epitopes					Toxic epitopes						
		Glia-α	Glia-α2	Glia-α9	Glia-α20	33-mer	α-Glia (31–43)	α-Glia (31–49)	α-Glia (31–55)	α-Glia (44–55)	α-Glia (51–70)	α-Glia (56–75)	α-Glia (206–217)
Pegaso	49	24	13	30	25	4	22	14	9	27	16	5	18
Gli-A2	31	24	13	18	10	4	22	14	9	13	10	5	0
Gli-D2	30	11	0	13	16	0	6	5	0	15	7	0	18
Gli-A2/GliD2	12	11	0	1	1	0	6	5	0	5	1	0	0

differently from the cv Pegaso, Noma and colleagues [43] found a similar number of intact genes coding α-type gliadins in the three *Gli-2* loci of the bread wheat cv Chinese Spring; in particular, the copy number was 16, 16 and 18 from the *Gli-A2*, *Gli-B2* and *Gli -D2* loci, respectively.

Expression of gliadin fractions in bread wheat cv Pegaso and derived Gli-2 deletion lines

Our data highlighted a differential contribution of the three *Gli-2* loci in the production of α-type gliadin transcripts. In particular, the levels of transcript produced from the *Gli-D2* locus is larger than those deriving from the *Gli-A2* and *Gli-B2* loci. Kawaura and colleagues [45] carried out a study of expression patterns of α-type gliadin genes by using a bioinformatic approach and found that genes from the D genome were preferentially expressed compared to those deriving from the A or B genomes. Similarly, Salentijn et al. [44] investigated the expression of α-type-gliadins in some tetraploid and hexaploid wheats, demonstrating that large differences exist in the expression levels of the α-type gliadins associated at the three *Gli-2* loci. Noteworthy the Gli-A2/Gli-D2 deletion line contains a low level of α-type gliadins (20%), but a drastic increase of the other gliadin fractions was observed in A-PAGE. This observation was confirmed at molecular level by Real-Time PCR and lead us to hypothesize the existence of compensatory mechanisms. A similar mechanism of regulation was never observed among the gliadin fraction, but it has previously reported between glutenins and gliadins [46, 47].

Number and position of the cysteine residues

The number and position of cysteine residues was conserved among the different α-type gliadins; In fact, six cysteine residues were present in most of the α-gliadins, except for six sequences which showed the presence of an additional cysteine residue in the NR2 domain (five associated to the *Gli-D2* locus and one to the *Gli-B2*). The presence of an extra cysteine residue in α-gliadins was also reported by different authors [35, 43]. Interestingly, a search on NCBI revealed that this extra cysteine was also present in the same position in different *T. aestivum* ssp. (*vulgare*, *compactum*, *spelta* and *macha*) and several accessions of *Ae tauschii*, thereby supporting the hypothesis that the point mutation responsible for the appearance of the odd cysteine residue occurred in the wild progenitor of the D genome to hexaploid wheats.

Presence of toxic and immunogenic epitopes

It is scientifically proved that celiac disease is triggered from digestion-resistant gluten peptides, among these α-type gliadins are considered the major responsible [19, 26, 48–50]. In fact, as all the other gluten proteins, α-type gliadins, being rich in proline and glutamine, have few trypsin cleavage sites and are partially digested. Some of these peptides are recognized from T cell lines and can activate the adaptive and innate immune response [25]. Van den Broeck and colleagues [28] compared the genetic diversity of gluten proteins for the presence of CD epitopes in 36 modern European wheat varieties and in 50 landraces, finding that the presence of the Gli-α9 (one of the major immunodominant epitope present on α-type gliadins) was higher in modern wheats, as compared to old varieties. In the present work the number of α-type gliadins that have the epitope Gli-α9 is reduced in the lines lacking *Gli-A2* (18) and *Gli-D2* (14) loci compared to the cultivar Pegaso (31); this reduction is drastically evident in the Gli-A2/Gli-D2 deletion line, where the epitope Gli-α9 was detected only in one α-type gliadin (Table 3).

Some α-type gliadin isoforms contain the peptide 33-mer, that correspond to a cluster of immunogenic epitopes. This peptide is considered the main immunodominant toxic epitope and it is formed by six overlapping copies of three DQ2-restricted T-cell epitopes [16, 42].

Here, the 33-mer epitope was detected in four α-gliadin proteins, all encoded by the genes associated with the *Gli-D2* locus; for this reason, the Gli-D2 and Gli-A2/Gli-D2 deletion lines are completely devoid of this peptide.

The investigation of the presence of immunogenic and toxic epitopes among the 49 α-gliadin isoforms isolated from the bread wheat cv Pegaso highlighted that the proteins associated with the *Gli-D2* locus should be considered more highly reactive in respect to T cell toxicity. This result agrees with other works that found a higher amount of immunogenic and toxic epitopes in the α-gliadin proteins specific of the D genome compared to those derived from A and B genomes [17, 27, 51].

Conclusions

Several wheat breeding programs have focused on gluten proteins, in particular on glutenins as these proteins are the main determinants of wheat technological properties. It has been hypothesized that breeding has contributed to the production of more "toxic" varieties, responsible for the increase in celiac disease and gluten sensitivity [28]. Since the gluten exposure is correlated with the incidence of CD, the use of wheat varieties with a reduced amount of toxic and immunogenic peptides seems to be a promising solution. At this regard, the single and double deletion lines here described could, very likely, be less harmful for people with a genetic predisposition for celiac disease, but that have still not developed the pathology, and would represents a good

material for the development of safer wheat varieties. In particular, the Gli-A2/Gli-D2 deletion line contains a drastic reduction of immunogenic (−85%) and toxic (−85%) epitopes compared to the bread wheat cv Pegaso. The use of novel technologies (e. g. genome editing technologies), capable to introduce accurately specific changes into the wheat genome combined with the use of some deletion lines, lacking already several toxic epitopes, could be effective in the challenge of obtaining a complete safe wheat for people suffering of CD.

Additional files

Additional file 1: Figure S1. Comparison of α-type gliadin genes. Alignment of the 49 full length nucleotidic sequences isolated from the bread wheat cv Pegaso were performed by the Clustal Omega multiple sequence alignment program. (DOC 112 kb)

Additional file 2: Figure S2. Amino acidic alignment of the signal peptide and the five domains corresponding to the 49 α-type gliadins isolated. Heptapeptide QRQLPYP in single copy or duplicated is highlighted in yellow; α-Glia (31–43) in gray; α-Glia (31–49) in gray + light blue; α-Glia (31–55) in gray + light blue +green; α-Glia (44–55) in light blue + green; α-Glia (51–70) underlined; α-Glia (56–75) in italic; α-Glia (206–217) in red; Glia-α in fuchsia; Glia-α2 in orange character; Glia-α9 in the black box; Glia-α20 in the blue box; the 33-mer peptide in the red box; the odd cysteine in bold and red character. (PDF 108 kb)

Additional file 3: Table S1. Accession number, molecular weight (MW) and length of the domains in the α-type gliadins isolated from bread wheat cv Pegaso. S=Signal peptide, R1 = Repetitive N-terminal domain; QR1 and QR2 = Polyglutamine domain 1 and 2; NR1 = Unique domain 1; NR2 = C-terminal unique domain. (DOCX 19 kb)

Additional file 4: Table S2. Presence or absence of seven toxic epitopes (**α-Glia (31–43):** PGQQQPFPPQQPY; **α-Glia (31–49):** PGQQQPFPPQQPYPQPQPF; **α-Glia (31–55):** PGQQQPFPPQQPYPQP QPFPSQQPY; **α-Glia (44–55):** PQPQPFPSQQPY; **α-Glia (51–70):** SQQPYLQLQPFPQPQLPY; **α-Glia (56–75):** LQLQPFPQPQLPYPQPQLPY; **α-Glia (206–217):** LGQGSFRPSQQN) in the 49 deduced amino acidic sequences isolated from the bread wheat cv Pegaso. (DOCX 17 kb)

Additional file 5: Table S3. Presence or absence of five immunogenic epitopes (**Glia-α:** QGSFQPSQQ; **Glia-α2:** PQPQLPYPQPQLPY; **Glia-α9:** PFPQPQLPY; **Glia-α20:** PFRPQQPYPQ; **33-mer:** LQLQPFPQPQLPYPQPQLPYPQPQLPYPQPQPF) in the 49 deduced amino acidic sequences isolated from the bread wheat cv Pegaso. (DOCX 15 kb)

Abbreviations

A-PAGE: Acid-Polyacrylamide gel electrophoresis; CD: Celiac disease; DPA: Days Post Anthesis; EMA: Endomysial antibodies; EMBL-EBI: European Bioinformatics Institute; HMW-GS: High molecular weight glutenin subunits; IPTG: Isopropyl-β-D-1-thiogalactopyranoside; LB: Luria-Bertani; LMW-GS: Low molecular weight glutenin subunits; NCBI: National Center for Biotechnology Information; NR1: Unique domain 1; NR2: C-terminal unique domain; ORF: Open Reading Frame; P: Signal peptide; QR1: Polyglutamine domain 1; QR2: Polyglutamine domain 2; R1: N-terminal domain; tTG: Tissue transglutaminase

Acknowledgements

Not Applicable.

Funding

The research was financially supported by Barilla G. e R. F.lli S.p.A. The funding contributed to the design of the study, interpretation of data and writing the manuscript.

Availability of data and materials

The datasets analysed during the current study are available from the corresponding author on request.

Authors' contributions

FC carried out the isolation of α-gliadins and the molecular analysis; BM and GC produced the deletion lines and 2D gel; FS and DL interpreted the data and drafted the manuscript; RR managed the plant grown in field and made substantial contributions to the acquisition of data; MS, RL and SM participated to the interpretation of data and revised the manuscript critically; DL conceived and coordinated the work. All authors have read and approved the final version of this manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Author details

[1]Department of Agriculture and Forestry Sciences, University of Tuscia, 01100 Viterbo, Italy. [2]Institute of Biosciences and Bioresources, CNR, 70126 Bari, Italy. [3]Barilla G. e R. F.lli Sp.A, 43122 Parma, Italy.

Received: 10 July 2017 **Accepted:** 8 December 2017
Published online: 19 December 2017

References

1. Shewry PR. Wheat. J Exp Bot. 2009;60:1537–53.
2. Shewry PR, Tatham AS. The prolamin storage proteins of cereal seeds: structure and evolution. Biochem J. 1990;267:1–12.
3. Kasarda DD. Glutenin structure in relation to wheat quality. In: Pomeranz Y, editor. Wheat is unique. American Association of Cereal Chemists, St Paul, Minnesota; 1989. p. 277–302.
4. Shewry PR, Halford NG, Lafiandra D. Genetics of wheat gluten proteins. Adv Genet. 2003;49:111–84.
5. Woychik JH, Boundy JA, Dimler RJ. Starch gel electrophoresis of wheat gluten proteins with concentrated urea. Arch Biochem Biophys. 1961;94:477–82.
6. Kasarda DD, Adalsteins AE, Laird NF. Gamma-Gliadins with alpha-type structure coded on chromosome 6B of the wheat (*Triticum aestivum* L.) cultivar 'Chinese Spring'. In: Lasztity R, Bekes F, editors. Proceedings of the 3rd International Workshop on Gluten Proteins, Budapest, Hungary; 1987.
7. Lafiandra D, Kasarda DD, Morris R. Chromosomal assignment of genes coding for the wheat gliadin protein components of the cultivars 'Cheyenne' and 'Chinese spring' by two-dimensional (two pH) electrophoresis. Theor Appl Genet. 1984;68:531–9.
8. Metakovsky EV, Branlard GP, Graybosch RA, Wrigley C, Békés F, Bushuk W. Gliadins of common wheat: polymorphism and genetics. In: Wrigley C, Bekes F, Bushuk W, editors. Gliadin and glutenin: the unique balance of wheat quality. AACC International, St Paul, MN; 2006. p. 35–84.
9. Masci S, Lafiandra D, Porceddu E, Lew EJL, Tao HP, Kasarda DD. D-Glutenin subunits: N-terminal sequences and evidence for the presence of cysteine. Cereal Chem. 1993;70:581–5.
10. Masci S, Rovelli L, Kasarda DD, Vensel WH, Lafiandra D. Characterisation and chromosomal localisation of C-type low-molecular-weight glutenin subunits in the bread wheat cultivar Chinese Spring. Theor Appl Genet. 2002;104:422–8.
11. Vensel WH, Tanaka CK, Altenbach SB. Protein composition of wheat gluten polymer fractions determined by quantitative two-dimensional gel electrophoresis and tandem mass spectrometry. Proteome Sci. 2014;12:1.
12. Harberd NP, Bartels D, Thompson RD. Analysis of the gliadin multigene loci in bread wheat using nullisomic-tetrasomic lines. Mol Gen Genet. 1985;198:234–42.
13. Anderson OD, Litts JC, Greene FC. The α-gliadin gene family. I. Characterization of ten new wheat α-gliadin genomic clones, evidence for limited sequence conservation of flanking DNA, and southern analysis of the gene family. Theor Appl Genet. 1997;95:50–8.

14. YQ G, Crossman C, Kong XY, Luo MC, You FM, Coleman-Derr D, Dubcovsky J, Anderson OD. Genomic organization of the complex alpha-gliadin gene loci in wheat. Theor Appl Genet. 2004;109:648–57.
15. Anderson OD, Greene FC. The α-gliadin gene family. II. DNA and protein sequence variation, subfamily structure, and origins of pseudogenes. Theor Appl Genet. 1997;95:59–65.
16. Ozuna CV, Iehisa J, Giménez MJ, Alvarez JB, Sousa C, Barro F. Diversification of the celiac disease α-gliadin complex in wheat: a 33-mer peptide with six overlapping epitopes, evolved following polyploidization. Plant J. 2015;82:794–805.
17. Van Herpen TW, Goryunova SV, van der Schoot J, et al. Alpha-gliadin genes from the A, B, and D genomes of wheat contain different sets of celiac disease epitopes. BMC Genomics. 2006;7:1.
18. Ribeiro M, Rodriguez-Quijanod M, Nunesc FM, Carrillod JM, Branlard G, Igrejaa G. New insights into wheat toxicity: Breeding did not seem to contribute to a prevalence of potential celiac diseases immunostimulatory epitopes. Food Chem. 2016;213:8–18.
19. Mitea C, Salentijn EM, Van Veelen P, et al. A universal approach to eliminate antigenic properties of alpha-gliadin peptides in celiac disease. PLoS One. 2010;5:e15637.
20. Vaccino P, Becker HA, Brandolini A, Salamini F, Kilian BA. Catalogue of *Triticum monococcum* genes encoding toxic and immunogenic peptides for celiac disease patients. Mol Genet Genomics. 2009;281:289–300.
21. Louka AS, Sollid LMHLA. In coeliac disease unravelling the complex genetics of a complex disorder. Tissue Antigens. 2003;61:105–17.
22. Mazzarella G, Maglio M, Paparo F, et al. An immunodominant DQ8 restricted gliadin peptide activates small intestinal immune response in in vitro cultured mucosa from HLA-DQ8 positive but not HLA-DQ8 negative coeliac patients. Gut. 2003;52:57–62.
23. Margaritte-Jeannin P, Babron MC, Bourgey M, et al. HLA-DQ relative risks for coeliac disease in European populations: a study of the European genetics cluster on coeliac disease. Tissue Antigens. 2004;63:562–7.
24. Sapone A, Bai JC, Ciacci C, et al. Spectrum of gluten-related disorders: consensus on new nomenclature and classification. BMC Med. 2012;10:13.
25. Ciccocioppo R, Di Sabatino A, Corazza GR. The immune recognition of gluten in coeliac disease. Clin Exp Immunol. 2005;140:408–16.
26. Molberg Ø, Uhlen AK, Jensen T, Flæte NS, Fleckenstein B, Arentz-Hansen H, Raki M, Lundin KE, Sollid LM. Mapping of gluten T-cell epitopes in the bread wheat ancestors: implications for celiac disease. Gastroenterology. 2005;128:393–401.
27. Spaenij-Dekking L, Kooy–Winkelaar Y, van Veelen P, Drijfhout JW, Jonker H, van Soest L, Smulders MJ, Bosch D, Gilissen LJ, Koning F. Natural variation in toxicity of wheat: potential for selection of nontoxic varieties for celiac disease patients. Gastroenterology. 2005; 129:797–806.
28. Van den Broeck HC, de Jong HC, Salentijn EM, Dekking L, Bosch D, Hamer RJ, Gilissen LJ, van der Meer IM, Smulders MJ. Presence of celiac disease epitopes in modern and old hexaploid wheat varieties: wheat breeding may have contributed to increased prevalence of celiac disease. Theor Appl Genet. 2010;121:1527–39.
29. Lafiandra D, Colaprico G, Kasarda DD, Porceddu E. Null alleles for gliadin blocks in bread and durum wheat cultivars. Theor Appl Genet. 1987;74:610–6.
30. Redaelli R, Metakovsky EV, Davidov SD, Pogna NE. Two-dimensional mapping of gliadins using biotypes and null mutants of common wheat cultivar Saratovskaya 29. Hereditas. 1994;121:131–7.
31. Lafiandra D, D'Ovidio R, Tanzarella OA, Ciaffi M, Margiotta B. Identification, molecular characterization and use of mutants for wheat storage proteins in quality studies. In: Bushuk W, Tkachuk R, editors. AACC International, St. Paul, MN; 1990. p. 537–44.
32. Branlard G, Dardevet M, Nardjis A, Gilberto L. Allelic diversity of HMW and LMW glutenin subunits and omega gliadins in French bread wheat (*Triticum aestivum* L). Genet Res Crop Evol. 2003;50:669–79.
33. Pflüger LA, D'Ovidio R, Margiotta B, Pena R, Mujeeb-Kazi A, Lafiandra D. Characterisation of high-and low-molecular weight glutenin subunits associated to the D genome of *Aegilops tauschii* in a collection of synthetic hexaploid wheats. Theor Appl Genet. 2001;103:1293–301.
34. Lafiandra D, Kasarda DD. One-and two-dimensional (two-pH) polyacrylamide gel electrophoresis in a single gel: separation of wheat proteins. Cereal Chem. 1985;62:314–9.
35. Xie Z, Wang C, Wang K, et al. Molecular characterization of the celiac disease epitope domains in α-gliadin genes in *Aegilops tauschii* and hexaploid wheats (*Triticum aestivum* L.). Theor Appl Genet. 2010;121:1239–51.
36. Li Y, Xin R, Zhang D, Li S. Molecular characterization of α-gliadin genes from common wheat cultivar Zhengmai 004 and their role in quality and celiac disease. Crop J. 2014;2:10–21.
37. Piston F, Dorado G, Martín A, Barro F. Cloning of nine γ-gliadin mRNAs (cDNA's) from wheat and the molecular characterization of comparative transcript levels of γ-gliadin subclasses. J Cereal Sci. 2006;43:120–8.
38. Altenbach SB, Kothari KM. Omega gliadin genes expressed in *Triticum aestivum* cv. Butte 86: effects of post-anthesis fertilizer on transcript accumulation during grain development. J Cereal Sci. 2007;46:169–77.
39. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Mol Biol Evol. 2016; msw054.
40. Sestili F, Janni M, Doherty A, Botticella E, D'Ovidio R, Masci S, Jones HD, Lafiandra D. Increasing the amylose content of durum wheat through silencing of the SBEIIa genes. BMC Plant Biol. 2010;10:1.
41. Marcussen T, Sandve SR, Heier L, et al. Ancient hybridizations among the ancestral genomes of bread wheat. Science. 2014;345:1250092.
42. Shan L, Molberg Ø, Parrot I, Hausch F, Filiz F, Gray GM, Sollid LM, Khosla C. Structural basis for gluten intolerance in celiac sprue. Science. 2002;297:2275–9.
43. Noma S, Kawaura K, Hayakawa K, Abe C, Tsuge N, Ogihara Y. Comprehensive molecular characterization of the α/β-gliadin multigene family in hexaploid wheat. Mol Genet Genomics. 2016;291:65–77.
44. Salentijn EM, Goryunova SV, Bas N, van der Meer IM, van den Broeck HC, Bastien T, Gilissen LJ, Smulders MJ. Tetraploid and hexaploid wheat varieties reveal large differences in expression of alpha-gliadins from homoeologous *Gli-2* loci. BMC Genomics. 2009;10:1.
45. Kawaura K, Mochida K, Ogihara Y. Expression profile of two storage-protein gene families in hexaploid wheat revealed by large-scale analysis of expressed sequence tags. Plant Physiol. 2005;139:1870–80.
46. Scossa F, Laudencia-Chingcuanco D, Anderson OD, Vensel WH, Lafiandra D, D'Ovidio R, Masci S. Comparative proteomic and transcriptional profiling of a bread wheat cultivar and its derived transgenic line overexpressing a low molecular weight glutenin subunit gene in the endosperm. Proteomics. 2008;8:2948–66.
47. Rakszegi M, Bekes F, Lang L, Tamas L, Shewry PR, Bedő Z. Technological quality of transgenic wheat expressing an increased amount of a HMW glutenin subunit. J Cereal Sci. 2005;42:15–23.
48. Anderson RP, Degano P, Godkin AJ, Jewell DP, Hill AV. Vivo antigen challenge in celiac disease identifies a single transglutaminase-modified peptide as the dominant A-gliadin T-cell epitope. Nat Med. 2000;6:337–42.
49. Arentz-Hansen H, Mcadam SN, Molberg Ø, Fleckenstein B, Lundin KE, Jørgensen TJ, Jung G, Roepstorff P, Sollid LM. Celiac lesion T cells recognize epitopes that cluster in regions of gliadins rich in proline residues. Gastroenterology. 2002;123:803–809.
50. Vader W, Stepniak D, Kooy Y, Mearin L, Thompson A, van Rood JJ, Spaenij L, Koning F. The HLA-DQ2 gene dose effect in celiac disease is directly related to the magnitude and breadth of gluten-specific T cell responses. Proc Nat Acad Sci. 2003;100:12390–5.
51. Wang DW, Da LJ JW, Zhao Y, et al. Genome-wide analysis of complex wheat gliadins, the dominant carriers of celiac disease epitopes. Sci Rep. 2017;7

CHAPTER 3:

PRODUCTION AND CHARACTERIZATION OF BREAD WHEAT LINES IN WHICH GENES CODING FOR ALPHA-AMYLASE/TRYPSIN INHIBITORS (ATI) HAVE BEEN SILENCED BY RNAi

ABSTRACT

Although wheat is used worldwide as a staple food, it can give rise to different adverse reactions, some of which have not been deeply characterized. They are caused mainly by wheat proteins, both gluten and non-gluten proteins. The possibility to develop new wheat genotypes accumulating a lower amount of proteins potentially or effectively involved in such pathologies not only offers the possibility to use them as a basis for the creation of wheat varieties with a lower impact on adverse reaction, but also to test if these proteins are actually implicated in those pathologies for which the triggering factor has not been established yet. In this study we report the characterization of lines derived from the bread wheat cultivar Bobwhite, in which three genes coding for α-amylase and trypsin inhibitors have been silenced. These proteins are involved in baker's asthma and likely in Non-Celiac Wheat Sensitivity. We have obtained different transgenic lines showing an effective decrease of the target genes, that do not show differences in terms of yield, but that have unintended effects on the accumulation of high molecular weight glutenin subunits, that play a crucial role for the technological performances.

3.1 INTRODUCTION

Wheat (*Triticum* spp.) is one the "big three" cereal crops, together with maize and rice, cultivated worldwide thanks to its adaptability, nutritional value and versatility. This latter is mostly due to the unique viscoelastic properties of its dough, capable to give rise to a wide range of products, such as bread, pasta, noodles, biscuits, some of which are characteristic of specific geographical areas.

Such properties derive from wheat grain proteins, classified by Osborne (1924) into four major groups based on solvent solubility: albumins (water), globulins (dilute salt solution), prolamins including gliadins (alcohol/water mixture), and finally glutenins, including glutenins (dilute acid or alkali). At present, gliadins and glutenins are both considered prolamins, because they are soluble in alcohol/water mixtures once glutenins are present in the reduced form.

Gliadins and glutenins make up the gluten, defined as the viscoelastic mass obtained after full flour hydration and removal of water soluble components, composed by soluble carbohydrates, mostly starch, and non-prolamin proteins, namely albumins and globulins. Among gluten proteins, glutenins play the major role, and in particular, their size and amount are the major determinants of dough technological quality (Gupta et al, 1993). Glutenins in fact, are composed of polymeric proteins made up of high (HMW-GS) and low (LMW-GS) molecular weight glutenin subunits.

Compared to the gliadins and glutenins, few studies have been carried out on non-prolamins so far. This is probably because their role in flour quality is not as well defined as that of gluten proteins. Nevertheless, the albumins and globulins constitute around 15-20% of total flour protein (Pence et al, 1954). They are a mixture of structural, metabolic and storage proteins (Goesaert et al, 2006).

The albumins and globulins are mostly located in the seed coat, the aleurone cells and the germ; they are relatively scarce in the starchy endosperm (Šramková et al., 2009).

Their amino acid compositions are relatively well balanced because of their higher lysine content as compared to the prolamin fraction.

Predominant albumins and globulins such as alpha-amylase/trypsin inhibitors (ATI), serpins and purothionins seem to possess multiple functions; indeed, they serve as nutrient reserves for the germinating embryo and as inhibitors of insects and fungal pathogens before germination (Dupont and Altenbach, 2003).

All wheat proteins can cause different adverse reactions, some of them are better characterized, such as Celiac Disease (CD), Wheat-Dependent Exercise-Induced Anaphylaxis (WDEIA), or Baker's Asthma (BA) (Elli et al, 2015). Differently, the role of wheat components in Non-Celiac Wheat/Gluten Sensitivity (NCWS or NCGS) or Irritable Bowel Syndrome (IBS) is still not clear. In particular, this can be deduced by the use of the two names, NCWS or NCGS to describe a pathology, that includes both gastrointestinal and non-gastrointestinal symptoms caused by wheat ingestion, but that excludes CD and wheat allergies. Because specific serological markers are not present so far, actually this is mostly a self-reported condition, whose diagnosis is based on double-blind placebo-controlled wheat challenge (reviewed in Mansueto et al, 2019). This situation makes even more difficult to establish the cause that initially was identified in gluten, mostly for analogies with CD, but that at present indicates ATI or fermentable oligosaccharides disaccharides monosaccharides and polyols (FODMAPs) as possible triggering factors, reason why it is currently preferred to use the name NCWS rather than NCGS. Since the prevalence worldwide is in the range 0.6-13%, it is important to identify at least the real culprit of such pathology.

Wheat ATI are among the putative triggering factors of NCWS and are unquestionably involved in BA, the most common occupational respiratory disease in Western Countries, affecting about 10% of flour workers (Salcedo et al, 2011) Moreover , this class of wheat proteins seems involved in some wheat related food-allergies, and, to a minor extent, with WDEIA (Zapatero et al, 2003)

ATI proteins belong to the so-called CM protein fraction of wheat, because they are soluble in chloroform and methanol solutions (for a review see Sanchez-Monge et al, 1989)

Three classes of ATI are typically described, that correspond to monomeric, dimeric and tetrameric forms, with different specificities against various heterologous α–amylases. In particular, the 12

kDa monomeric inhibitors, also known as 0.28 proteins, are encoded by genes on the short arms of the group 6 chromosomes; the 24 kDa homodimeric inhibitors, also known as the 0.19 and 0.53 proteins, are encoded by genes on the short arms of the group 3 chromosomes; the third group is constituted of the 60 kDa heterotetrameric inhibitors: this latter composed of one copy of either CM1 or CM2, encoded by genes on chromosomes 7D or 7B, plus one copy of either CM16 or CM17, encoded by genes on chromosomes 4B or 4D, plus two copies of CM3, also encoded on chromosomes 4B or 4D (Garcia-Olmedo et al, 1987).

The availability of genotypes with reduced amounts of ATI proteins might allow either to develop novel wheat cultivars with a lower triggering potential and to understand better the role of these components in such wheat-related pathologies. Until the advent of CRISPR-Cas9 silencing technology, RNA interference (*RNAi*) was the most efficient tool to silence multigene families and homoeologous genes in polyploids, such is the case of *ATI* genes in hexaploid wheat. In this study, we report the development and the characterization of wheat lines derived from the bread wheat cultivar Bobwhite, in which specific ATI genes have been silenced by RNAi.

3.2 MATERIALS AND METHODS

3.2.1 Plant Materials

Genes choice, RNAi constructs preparation, wheat transformation by the biolistic procedure and the selection of the transgenic lines were previously performed in our lab (in prep.). Briefly, three ATI genes (*CM3*, *CM16* and *0.28)* were targeted by three RNAi cassettes. The bread wheat cultivar used for transformation was Bobwhite. Wheat calli were co-bombarded with the three RNAi plasmids with the biolistic procedure. Homozygous transgenic lines showing silencing of the three genes have been advanced to the T_4 generation , thus T_4 plants were used for the following analyses, in comparison with the control untransformed line (WT). All the lines were cultivated in growing chambers in the same conditions. Not all the five lines obtained were used in all analyses, thus the identification codes of those used used are indicated in the following paragraphs relative to each analysis performed.

3.2.2 Quantitative Analyses of ATI Transcripts by qRT-PCR

qRT-PCR was performed on total RNA extracted from the caryopses at 10 Days Post Anthesis (DPA), 24 DPA and 30 DPA, using the Spectrum Plant Total RNA kit (Sigma-Aldrich, St. Louis,USA) and following manufacturer's instructions. First strand complementary DNA was

synthesized using the QuantiTect Reverse Transcription Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions from one μg of total RNA per each sample. The experiment was carried out with three specific primer pairs for *0.28, CM3* and *CM16* genes and with one specific primer pair for *HMW-GS Dx5*. The actin gene was used as housekeeping gene. The analyses were performed in the CFX 96 Real-TimePCR Detection System device (Bio-Rad Hercules, CA, USA). The reactions were prepared in a final volume of 15 μL, consisting in 7.5 μL SsoAdvUniver SYBR GRN SMX (Bio-Rad Hercules, CA, USA), 0.5 μM of each primer and1 μL of cDNA, following a three step PCR protocol: 95 °C for 30 s and 39 cycles at 95 °C for 10 s, 60 °C for 25 s and 72°C for 15 s. Three replicates were carried out for each reaction and the maximum difference accepted between Ct values of technical replicates was 0.5. The relative quantification normalized to the housekeeping gene was performed using the $2^{-\Delta\Delta Ct}$ method.

3.2.3 SDS-PAGE

The electrophoretic control of lines was performed on the albumin and globulin fraction, containing the CM proteins, including ATI, in all lines obtained, as well as on HMW-GS.

Albumin and globulin extraction was performed on pools of mature grains, that were crushed and mixed with extraction buffer containing 0.05M phosphate buffer/0.1M NaCl, pH 7.8 (3 g: 80 ml). After centrifugation at 8,000 g for 15 min at 4°C, the supernatant was precipitated with acetone; the pellet, corresponding to the A/G proteins, was then rinsed two times. Protein concentrations of the A/G proteins were determined with the 'BioRad Protein assay' kit (Bradford, 1976). The A/G fraction was solubilized in Tris-HCl buffer, pH 6.8, containing 2% SDS, 10% glycerol, 0.1% 2-mercaptoethanol and bromophenol blue. SDS-PAGE was performed using stacking gels (T=3.75%; C=2.67%) to concentrate the proteins and running gels (T=15%; C=0.9%) to separate the proteins based on their molecular weight. SDS-PAGE was performed with the Mini-PROTEAN Tetra Vertical Electrophoresis Cell (Bio-Rad) and the gels were 1.0 mm thick. For each sample, 15 μg of protein extract were loaded into the gel that was run at a constant voltage of 200V until the bromophenol blue dye front disappeared.

Glutenins subunits were extracted from 10 mg of flour of transgenic lines and Bobwhite control plants in a 1:10 ratio w/v in 70mM Tris-HCl, pH 6.8, 2% SDS, 10% glycerol, 0.02% γ-pironine and 1% DTT, stirring for 1 h at room temperature. After centrifugation at 13,000 g extracts for 10 min, they were quantified with the 'BioRad Protein assay' kit and 10 μg of proteins were used for SDS-PAGE analysis through SE 600 Hoefer (FisherScientific) device. The main gel was T = 11 and C = 1.28, while the stacking gel was T = 3.75 and C = 2.67. The electrophoresis was performed at 40 mA per gel at 10°C and stopped after 30 min the tracking dye reached the bottom of the gel.

In both cases, gels were stained according to Neuhoff et al (1988).

3.2.4 Immunoblotting

SDS-PAGE gels were incubated in transfer buffer (25mM Tris, 160mM glycine, 0.1% SDS and 20% ethanol) for 15 minutes. The blotting was performed in a Trans-Blot Semi-Dry Transfer Cell (Bio-Rad) using a nitrocellulose membrane (0,2 µm, Bio-Rad) and the transfer was achieved at 300 mA for 30 min. The nitrocellulose sheet was washed in PBS and blocked overnight at room temperature in PBS-4% milk. After three washes with PBS-0.05% Tween 20, nitrocellulose membrane was incubated for 1 h with a polyclonal antibody against α-amylase inhibitors on A/G fraction, properly developed in rabbit at the INRA, Nantes (France), diluted at 1:8,000 in PBS-2% milk. The membrane was washed three times with washing buffer and incubated for 1 h with the secondary antibody, a HRP conjugated anti-rabbit IgG diluted at 1:100,000 in PBS-2% milk. After further washing, the Luminata Crescendo Western HRP Substrate (Millipore, Billerica, MA, USA) was applied and the membrane was put immediately into darkness. The chemiluminescent signal was detected using a ChemiDoc XRS+ imaging system (Bio-Rad).

3.2.5 Two Dimensional Gel Electrophoresis

Two transgenic lines (24-2 and 10-10a) were used in comparison with the WT line. Albumin and globulins were extracted from 1 g flour of each line according to Vensel et al. (2005), but amounts were scaled up to perform large 2D gels (18 cm). Briefly flour (150 mg) was suspended in 600 µl of cold KCl buffer (50 mM Tris–HCl, 100 mM KCl, 5 mM EDTA, pH 7.8). The suspension was incubated on ice for 5 min with intermittent mixing and centrifuged (14,500 g, 15 min, 4 °C). The KCl-soluble fraction was collected and 5 volumes (v/v) of cold 0.1 M ammonium acetate in methanol were added at room temperature. Following incubation overnight at -20 °C, the methanol-insoluble fraction was pelleted by centrifugation as above. The pellet (containing metabolic proteins) was rinsed with cold acetone, dried down and stored at -20°C until further use. Protein amounts were assessed by 2-D quant Kit Assay (GeHealthcare).

Proteins were dissolved in 800 µl of strip rehydration buffer containing 7 M Urea, 2 M thiourea, 2% (w/v) CHAPS, 2% (v/v) Triton X100, 1.2% (v/v) Destreak reagent (GE Healthcare), and 0.5% (v/v) IPG buffer pH 3-10. IEF linear IPG strips (18 cm, GE Healthcare) pH 3–10 were used as first dimension. Strips were rehydrated overnight at 20°C with 200 µL (about 300 µg) of dissolved proteins and 150 µl of rehydration buffer was added. The same extract volume (200 µl) was loaded. Focusing was performed at 20°C for 80 kVh (200 V 8 h, 300 V 3 h, up to 3500 V in 30 min, then 3500 V 3.5 h, 8000 V 4 h) using IPGphorTM Isoelectric Focusing System (GeHealthcare). The gel

strips were subsequently equilibrated for 15 min in 0.1 M Tris-HCl pH 6.8 containing 8 M Urea, 30% (w/v) glycerol, 1% (w/v) SDS, 1% (w/v) DTT 2.5% iodoacetamide, and bromophenol blue as tracking dye. For the second dimension, the strips were transferred onto 18 x 20 cm polyacrylamide gels (15%T, 1.28%C) (Protean II X-Cell, BioRad) and run at 40 mA per gel for 3-4 h at 11°C until the dye front left the gel. A total of 27 gels were run (3 genotypes x 3 biological replicas x 3 biological replicates). Gel were stained according to Neuhoff et al (1988).

Gels were digitalized by using a scanner to acquire images with the same conditions for each gel (300 dots per inch and 16-bit grayscale pixel depth), according to Scossa et al. (2008). The protein spots that had *p* values (obtained by ANOVA) and *q* values lower than 0.05, and fold change equal to or greater than 1.2, were considered as significant differentially expressed using the software SameSpots Progenesis (version 4.6.1.218, Nonlinear Dynamics, UK). Differential spots were picked from those gels in which the spot color was more intense, (where the protein concentration was supposed to be higher) and transferred into 1.5 ml tubes. In case of low-abundance protein spots, they were picked from replicate gels and pooled. The excision of the spots was carried out by using the instrument EXQuestTM Spot Cutter by Bio-Rad®.

3.2.6 Indirect ELISA

The enzyme-linked immunosorbent assay (ELISA) was performed for quantifying ATI in the A/G fraction of the WT and transgenic lines. For each genotype, three replicates of each of three biological replicates (A/G fraction) were tested, except for the two transgenic lines 24-1-1 and 24-2-1, whose samples were tested in duplicate. The wells on microtiter plate (ELISA plate 82.1581.100, Sarstedt) were coated with 5 μg/ml in 100mM carbonate (pH 9.6) of antigen (A/G fraction) for 1h at room temperature. After three washes with PBS-0, 05% Tween 20, the plate was blocked with PBS-4% milk for 1h at 37°C. The plate was washed three times with PBS-0,05% Tween 20 and then incubated for 1h at 37°C with a serial dilution of primary antibody from 1:2,000 to 1:64,000 in PBS-2% milk; the same anti-ATI antibody used in Immunoblotting was used as primary antibody to complex with the ATI present in the A/G fraction of each line. After three washes with PBS-0, 05% Tween 20, the plate was incubated for 1h at 37°C with the secondary antibody (anti-rabbit IgG labelled with the enzyme HRP). After three additional washes with PBS-0, 05% Tween 20, the colorimetric substrate for HRP OPD (o-Phenylenediamine) in 0,05M citrate buffer plus H_2O_2 were introduced to have a yellow-orange product detectable at 490nm. The OPD reaction was stopped after 30 min with H_2SO_4 4N.

3.2.7 Parameters Related To Yield

Thousand Kernel Weight (TKW) was used as a yield indicator, along with the coleoptile length measured at 5 and 7 days post sowing (DPS), on transgenic and control lines.

3.2.8 Calculation of The Percentage of Unextractable Polymeric Proteins (%UPP)

The percentage of unextractable polymeric proteins (%UPP) was determined according to Gupta et al. (1993). Soluble glutenin polymers were extracted from 10 mg of flour by using 50% ACN 0.05% TFA in a 1:10 (w/v) ratio. The samples were left under stirring for 30 min at room temperature and subsequently centrifuged for 15 min at 10,000 g. The supernatant was filtered through 0.45 μm filters (Ultrafree-MC, centrifugal filters, PVDF, Millipore Corp., Billerica, MA) by centrifugation at 16,900 g for 10 min. The insoluble polymers were extracted by adding to the pellet from the previous extraction 50% ACN 0.05% TFA, in a 1:10 ratio. The extraction of insoluble polymers was carried out by sonication with a sonicator (Vibra cell, Sonics Materials, INC), with a 5 mm probe with the following settings: tune 50, output control 25 for 20 sec. The supernatant was filtered through PVDF 0.45 μm filters by centrifugation at 16,900 g for 10 min. For the SEC analysis of soluble and insoluble polymers, HPLC System Gold 126 NM (BeckmanCoulter) was used with management and acquisition software 32 Karat 7.0, equipped with UV detector Spectra Series UV150 (Thermo Separation Products) set at 214 nm and chromatographic 30 cm column molecular exclusion BioSep-SEC-S 4000 (Phenomenex) with an internal diameter of 4.6 mm. The samples were injected using 100 μl loop connected to a six-way valve, elution occurred under isocratic conditions using 50% ACN 0.05% TFA eluent, at a flow rate of 1 mL/min for a duration of the stroke chromatography of 10 min. The column was kept at the temperature of 30 ° C.

3.2.9 Bioinformatic Comparison

The NCBI (National Center for Biotechnology Information https://www.ncbi.nlm.nih.gov.) data bank was used to compare *ATI 0.28* (accession number AJ223492.1), *ATI CM3* and *ATI CM16* (accession numbers AY436554.1 and X17573.1, respectively), with the following *HMW-GS* genes: *1Dx5* (accession number DQ907161.1), *1Dy10* (accession number X12929.2), *1Ax2** (accession number M22208.2), *1Bx7* (accession number BK006773.1) e *1By9* (accession number X61026.1). Alignment was performed with the software ClustalOmega (https://www.ebi.ac.uk/Tools/msa/clustalo/r).

3.3 RESULTS

3.3.1 Gene Expression Analysis

Relative gene expression analyses were performed on 10-10 and 10-24 lines, as well as on Bobwhite WT plants to investigate the reduction of transcripts of the genes encoding CM3, CM16 and 0.28 subunits in the two RNAi lines during kernel development stages and to evaluate potential pleiotropic effects on *HMW-GS* genes. This latter aspect was added because HMW-GS result downregulated (more details in the paragraph "*HMW-GS decrease in silenced lines*").

Data obtained from RNA extracted at 10 DPA show an immediate reduction in the first development stages of all transcripts investigated. The decrease of the *CM3*, *CM16* and *0.28* messengers was more evident in the line 10-24, but it was also significant in the line 10-10a. In fact, it was found that reduction ranged from 80% to almost 100% in 10-24 line, whereas the reduction ranged from 60% to 90% in 10-10 line; a similar result was obtained from the analysis carried out with RNA extracted at 20 DPA. The almost total silencing of ATI genes was shown at 30 DPA for both RNAi lines. The expression analysis of HMW-GS 1Dx5 revealed a significant reduction of *1Dx5* transcripts, that became even more drastic at 30 DPA (Fig. 3.1).

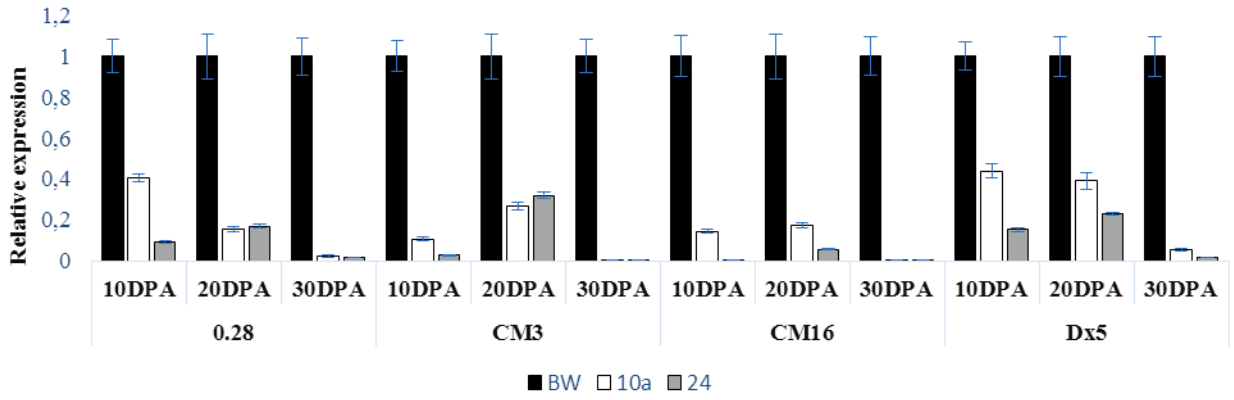

Fig. 3.1. Expression analysis of CM3, CM16, 0.28 ATI subunits and HMW-GS 1Dx5. *Quantitative RT-PCR at different kernel development stages (10, 20 and 30 DPA) in bread wheat RNAi lines 10-10a and 10-24 and Bobwhite control plants*

3.3.2 Effect of Gene Silencing on ATI Content

In order to assess whether the RNAi silencing was successful, we performed indirect ELISA and immunoblotting experiments using a polyclonal antibody against α-amylase inhibitors on A/G

fraction. We included in the experiments all five transgenic lines obtained and Bobwhite control plants.

The ELISA analysis revealed that the amount of ATI reacting with the primary antibody was lower in all the five transgenic lines compared to Bobwhite control plants (Fig. 3.2).

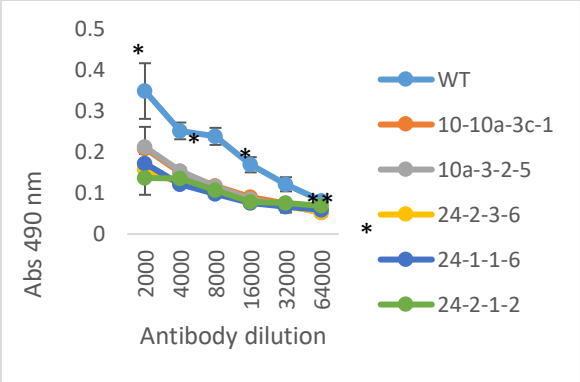

***Fig. 3.2. ELISA**. Test performed with monoclonal antibody against ATI present in the A/G fraction of Bobwhite control plants and the five transgenic lines obtained. Asterisks above the graphic correspond to ranking of Tukey test at 0.95 confidence and P<0.05 level of significance.*

The levels of antigen were significantly different between transgenic lines and Bobwhite control plants in all primary antibody dilutions, except for the lower ratio (1:64,000), for which no difference was observed.

These results are in agreement with immunoblotting experiments by using the same polyclonal antibody. All transgenic lines lacked the hybridization band corresponding to 14 kDa that was present in Bobwhite control plants. On the contrary, both transgenic lines and WT control plants showed a hybridization band of 17 KDa (Fig. 3.3).

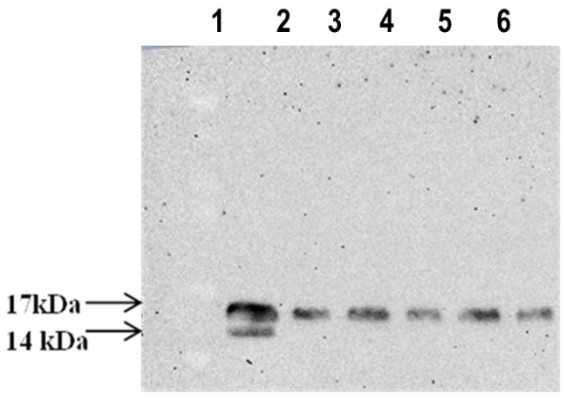

***Fig. 3.3. Immunoblotting experiment with monoclonal antibody against ATI on A/G fraction**. Samples: 1) Bobwhite (WT control); 2) 10-10; 3) 10-5a; 4) 10-24; 5) 10-23; 10-2. Molecular masses of the two bands are indicated on the left.*

3.3.3 Differential Proteomic Analysis of albumin/globulin Fraction

This analysis was performed both for identifying the specific proteins that have been silenced, and to test the presence of possible pleiotropic effects on proteins belonging to the same functional class. For this analysis, we used two transgenic lines (10-10 and 10-24) in comparison with the WT control genotype.

Differential proteomic expression comparison revealed several differentially expressed spots, some of which are either absent in the RNAi silenced, and likely correspond to the target, whereas others show a differential volume, indicating a different regulation (Fig. 3.4).

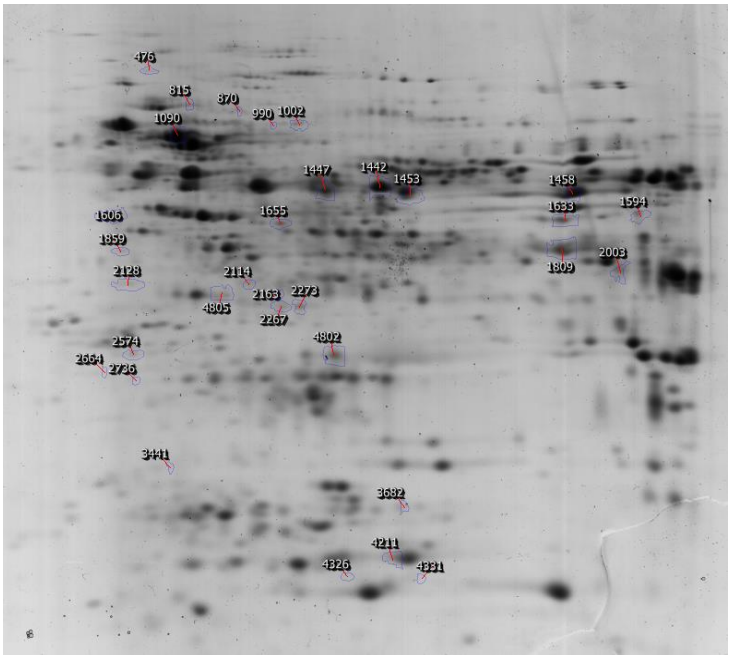

Fig. 3.4: 2D gel *reporting all the protein spots that are differentially accumulated in the two transgenic lines.*

The following Tables (1 and 2) report ANOVA, q value, power, and fold change for all the protein spots that are differentially accumulated in the comparisons between the WT line and each of the transgenic lines

Table 3.1**: Comparison between the WT line and the 10-24 line** *(M stands for missing spot, + or – mean that the protein is either over or down regulated in the transgenic line)*

Spot number	Anova	q value	Power	Fold change	24-2 vs WT
4326	4,235e-005	0.347	0.999	2,6	
1594	1,173e-004	0.0347	0.997	2,1	M
2003	1,504e-004	0.0347	0.99	1,5	+

1002	3,822e-004	0.045	0.987	1,7	-
2273	3,914e-004	0.045	0.987	1,7	-
4211	7,205e-004	0.0592	0.974	2,7	M
1606	8,660e-004	0.06	0.969	1,4	-
2267	0,001	0.0621	0.959	2,0	M
476	0,001	0.0621	0.957	2,7	M
2128	0,001	0.0621	0.951	1,7	M
815	0,001	0.0621	0.951	1,7	-
1458	0,002	0.072	0.933	1,9	+
1442	0,002	0.073	0.928	1,4	+
1655	0,003	0.0742	0.917	1,5	-
4802	0,003	0.0749	0.91	1,5	+
2163	0,003	0.0768	0.906	1,4	M
1447	0,004	0.0861	0.888	1,6	+
1809	0,006	0.107	0.852	1,8	+
1453	0,008	0.127	0.807	1,5	+
990	0,009	0.127	0.797	1,7	M
4805	0,009	0.127	0.794	1,3	-
1090	0,012	0.141	0.76	1,5	+
2574	0,013	0.15	0.743	1,4	-
1859	0,020	0.194	0.676	1,5	-
870	0,029	0.222	0.618	1,9	-
4331	0,044	0.267	0.536	1,7	-

***Table 3.2**: **Comparison between the WT line and the 10-10 line** (M stands for missing spot, + or – mean that the protein is either over or down regulated in the transgenic line)*

Spot number	**Anova**	**q value**	**Power**	**Fold change**	**10-10 vs WT**
4326	4,902e-005	0.0242	0.999	2,4	M
1809	8,455e-005	0.0242	0.998	2,1	+
2163	1,004e-004	0.0242	0.998	1,6	-
4805	0,002	0.109	0.945	1,3	-
2267	0,002	0.109	0.939	2,0	-
815	0,002	0.109	0.936	1,7	-
1458	0,003	0.109	0.91	1,8	+
1606	0,003	0.115	0.9	1,4	-
2273	0,004	0.137	0.878	1,5	-
4331	0,005	0.138	0.869	2,0	-
1859	0,005	0.138	0.865	1,6	-
2003	0,005	0.138	0.856	1,3	+
2574	0,006	0.138	0.851	1,5	-
1002	0,006	0.138	0.85	1,5	-
1090	0,017	0.232	0.708	1,5	+
1447	0,019	0.236	0.684	1,4	+
476	0,021	0.239	6.72	1,9	-
2128	0,038	0.301	0.566	1,4	-
1594	0,041	0.311	0.55	1,4	-

In order to identify them, all these proteins spots will be submitted to MS analyses.

3.3.4 TKW and Coleoptile Length

TKW and Coleoptile Length are correlated with wheat yield and have been measured to test if the silencing of ATI may influence such important aspect.

In both cases, no significant difference was detected between the WT and the transgenic lines (Fig. 3.5).

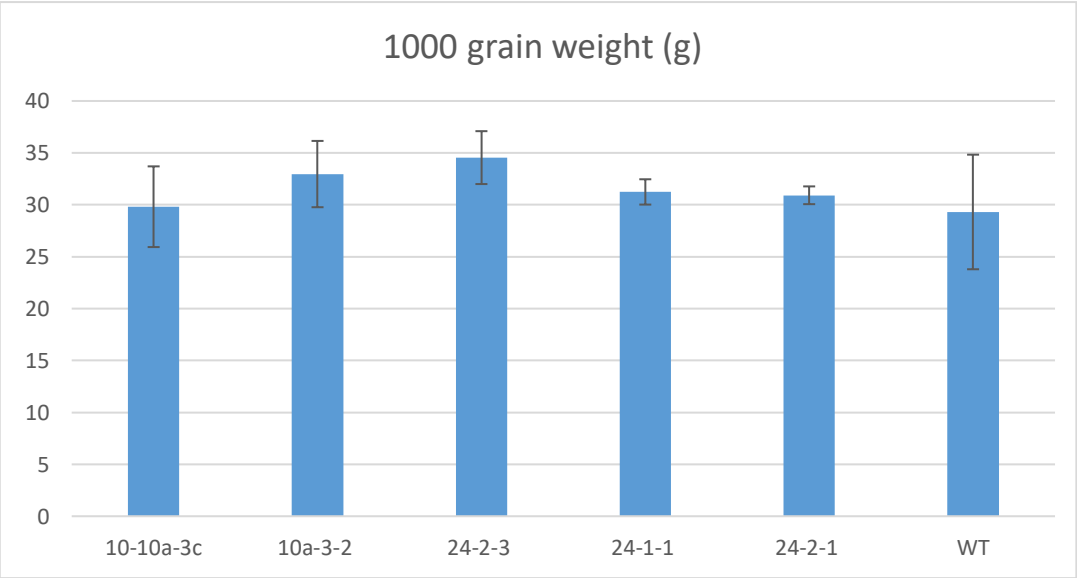

Fig. 3.5: Comparison of the 1000 grain weight between the WT and the transgenic lines.*p<0,05 and* ***p<0,01.*

The coleoptile lengths of the WT and transgenic lines were compared at two different time-points: 5 and 7 DPS (Figs. 3.6 and 3.7). In this case, the measurements were taken from only three RNAi transgenic lines due to a contamination of the 24-2-1 and 24-1-1 lines. As shown in Figs. 3.6 and 3.7, no significant different was observed in any of the two time-points.

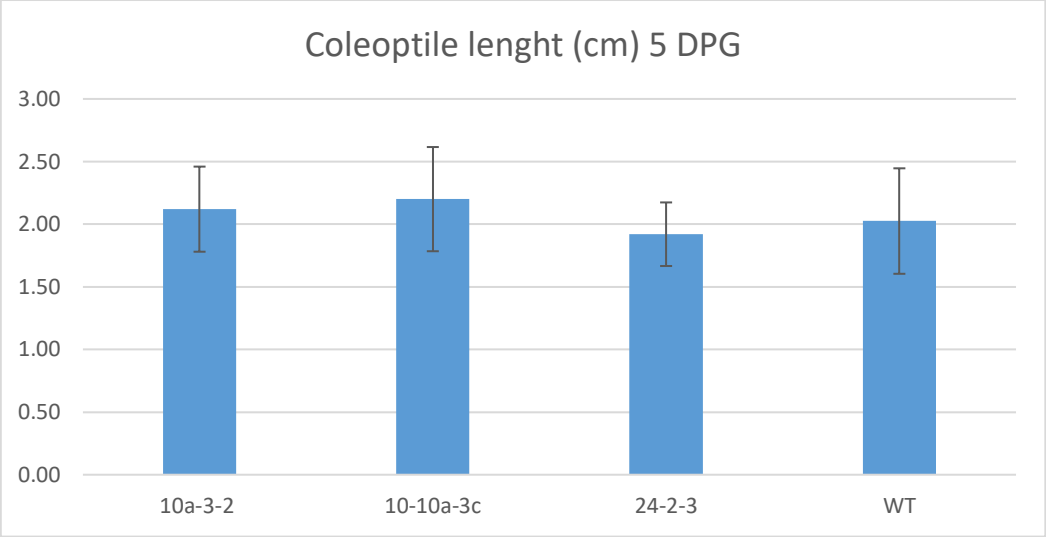

Fig. 3.6: Comparison of the coleoptile length (cm) 5 DPA between the WT and the transgenic lines.*p<0,05 and **p<0,01.*

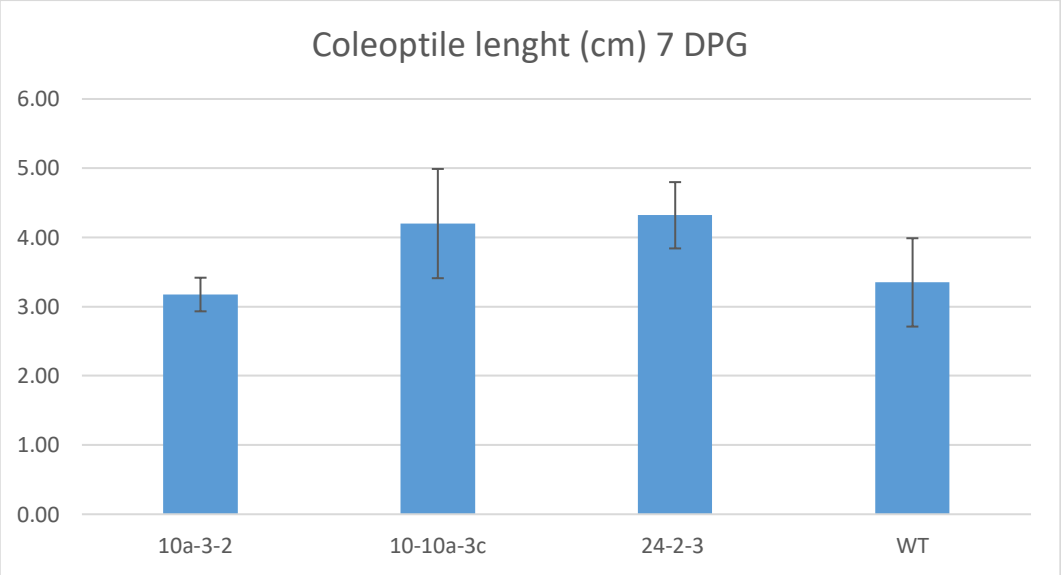

Fig. 3.7: Comparison of the coleoptile length (cm) 7 DPA between the WT and the transgenic lines.*p<0,05 and **p<0,01.*

The absence of significant differences in both the tested parameters showed that yield was not affected by the silencing of ATI proteins.

3.3.5 Prediction of Gluten Quality of RNAi Silenced Plants

%UPP is an index related to gluten strength. Since higher amounts of flour are needed to perform alveographic tests, we used this micro- test to predict gluten quality of RNAi silenced plants.

Fig. 3.8 reports the %UPP values found for the WT line and three transgenic lines. ANOVA test showed that there is a dramatic decrease of the %UPP value in all silenced lines, and that the latter show no significant difference among them.

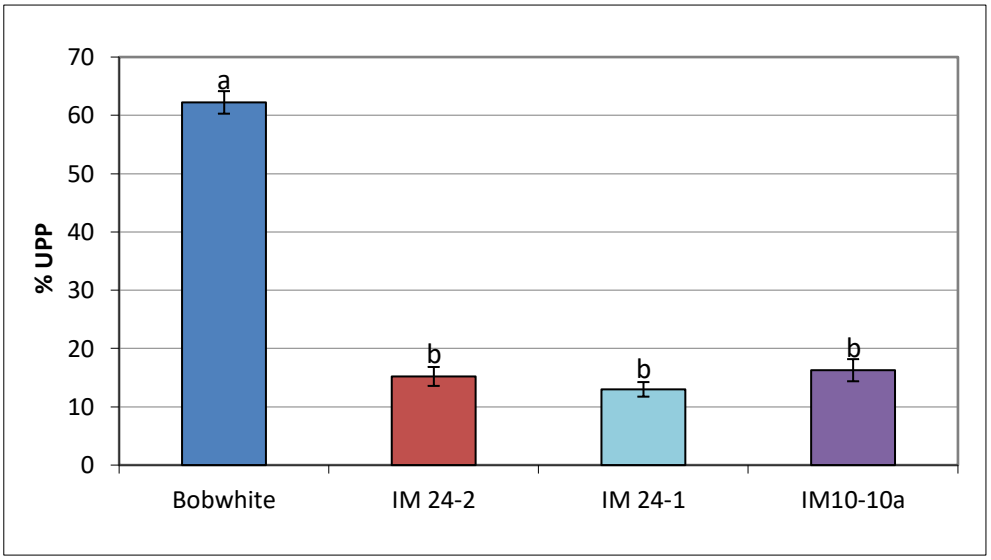

Fig. 3.8: Determination of % of unextractable polymeric proteins (%UPP) in control plants and transgenic lines*. Values represent the average of three biological and three technical replicates. Letters above the histograms correspond to ranking of Tukey test at 0.95 confidence and P<0.01 level of significance.*

3.3.6 HMW-GS Decrease in Silenced Lines

In order to understand the reason of such decrease in %UPP, we have analysed the electrophoretic pattern of HMW-GS, whose allelic variation and amount is related to bread wheat technological quality.

The SDS-PAGE pattern of HMW-GS is reported in Fig. 3.9.

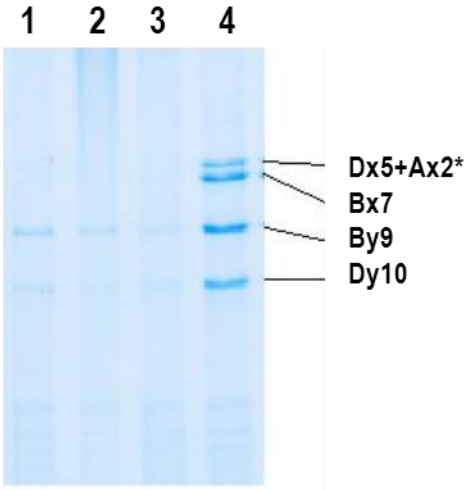

Fig. 3.9: SDS-PAGE. *HMW-GS extracted from three transgenic lines (1-3) and the control line (4)*

It is evident that HMW-GS have been an off-target of RNAi silencing and this explains the decrease in %UPP. To understand if untargeted silencing of HMW-GS was due to possible homologies among ATI and HMW-GS genes, we made a bioinformatics comparison among the three target ATI genes and the genes coding for the HMW-GS genes present in cultivar Bobwhite, that are 1Ax2*, 1Dx5, 1Bx7, 1By9 and 1Dy10.

Such comparison allowed to identify the presence of a conserved region in ATI 0.28 and HMW-GS 1By9 and 1Dy10 (TGCTGCCAGCAGCT), that might explain the off-target silencing, at least of these two subunits. Moreover, the alignment between ATI 0.28 and HMW.GS Dy10 showed a further homology of some nucleotides separated from the conserved region by a guanine pair (Fig. 3.10)

```
127 TACAAGGTGAGCGCACTCACGGGCTGCCGGGCA-ATGGTGAAGCTCCAGT   175
    | |.|||.||||.|.||...||..|||| |||| .|.|||.|.|..||||
104 t-ccaggagagctcgcttgaggcatgcc-ggcaggttgtggaccaacagt   151

176 GTGTGGGCAGTCAGGTGCC--------CGAGGCTGTCCTAAGAGAT----   213
       |||.|.|||.|.||||        ||.|||  |||    ||||
152 ---tggccggtcggctgccatggagcacggggc--tcc----agatgcga   192

214 TGCTGCCAGCAGCTGGCCGACAT--------CAACAACGAATGGTGCAGG   255
    ||||||||||||||  ||||.||        |||         ||||.|.
193 tgctgccagcagct--ccgagatgttagcgccaa---------gtgccgc   231

256 TGCGGGGACCTCAGCAGCATGTTGCGTAGTGTTTATCAG-GAGCTCGGCG   304
    |.||..|.|.|||||  ||.||.|| .||....|||.|| .|.||    |
232 tccgtcgccgtcagc--caagtcgc-aagacaatatgagcaaact----g   274

305 TGCGTG-------AGGGGAAGGAGGTGCT----CCCAGGTTGCCGGAAGG   343
    || |||       ||||  .|||  |.||    |||.|||
275 tg-gtgccgcccaaggg--cgga--tccttctaccctggt----------   309
```

Fig. 3.10. Nucleotide comparison between ATI 0.28 (uppercase) and HMW-GS 1Dy10 (lowercase). *In green the conserved sequence is indicated, whereas the red square includes also the further partially conserved region.*

The conserved sequence has been submitted to a further analysis with PlantGRN (http://plantgrn.noble.org), containing information on plant transcriptional regulation. By using the psRNATarget and ppsRNAi tools, the presence of siRNA and miRNA is suggested. psRNATarget highlights 16 miRNA sequences with sequence homologies with *ATI 0.28* and HMW-GS *1Dy10*. By using the ppsRNAit tool, 18 possible siRNA have been identified (data not shown).

3.4 CONCLUSIONS

Although RNAi is a transgenic procedure, it can be used as a proof of concept before developing new genotypes with non-transgenic techniques. Regardless the procedure used, in fact, any new genetic combination can potentially give rise to unintended effects. Silencing of specific genes can be pursued by exploiting natural variation present in genetic resources, or by generating pseudogenes using classical mutagenesis or advanced biotechnologies, but all the classical methods require time-consuming procedures that, in case undesirable unintended effects are produced, make the work performed useless.

The possibility to develop new wheat genotypes accumulating a lower amount of proteins potentially or effectively involved in such pathologies, not only offers the possibility to use them as a basis for the creation of wheat varieties with a lower impact on adverse reactions, but also to test if these proteins are actually implicated in those pathologies for which the triggering factor has not been established yet.

On these premises, we have developed bread wheat lines in which three genes coding for ATI proteins, involved in baker's asthma and likely in Non-Celiac Wheat Sensitivity, have been silenced by RNAi interference, and characterized them both in terms of protein and gene expression, and of parameters related to technological performances.

We have obtained different transgenic lines showing an effective decrease of the target genes, and of the corresponding protein products, that do not show differences in terms of yield, but that show differences in non-target HMW-GS accumulation, involved in technological performances. There are indications that this latter aspect, rather than being a pleiotropic effect due to silencing of the ATI genes, is likely due to partial homology between ATI and HMW-GS genes. If this is the case, it could be circumvented by crossing the transgenic lines with the WT genotype.

REFERENCES

Bradford, M. M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Analytical Biochemistry, 72(1–2), 248–254. https://doi.org/10.1016/0003-2697(76)90527-3

Dupont, F. M., & Altenbach, S. B. (2003). Molecular and biochemical impacts of environmental factors on wheat grain development and protein synthesis. Journal of Cereal Science, 38(2), 133–146. https://doi.org/10.1016/S0733-5210(03)00030-4

Elli, L.; Branchi, F.; Tomba, C.; Villalta, D.; Norsa, L.; Ferretti, F.; Roncoroni, L.; Bardella, M.T. Diagnosis of gluten related disorders: Celiac disease, wheat allergy and non-celiac gluten sensitivity. World J. Gastroenterol. 2015, 21, 7110–7119.

Garcia-Olmedo, F.; Salcedo, G.; Sanchez-Monge, R.; Gómez, L.; Royo, J.; Carbonero, P. Plant proteinaceous, inhibitors of proteinases and 3-amylases. In Oxford Surveys of Plant Molecular and Cell Biology; Miflin, B. Ed; Oxford University Press: Oxford, UK, 1987; Volume 4, pp. 275–334.

Goesaert, H., Gebruers, K., Courtin, C.M., Brijs, K., Delcour, J.A., Enzymes in breadmaking. In “Bakery products: science and technology”, Hui Y.H. Ed. 2006, pp. 337-364.

Gupta, R.B., Khan, K., and MacRitchie, F. (1993) Biochemical basis of flour properties in bread wheats. I. Effects of variation in the quantity and size distribution of polymeric protein. J. Cereal Sci. 18, 3-41.

Neuhoff V., Arold N., Taube D., Ehrhardt W., Improved staining of proteins in polyacrylamide gels including isoelectric focusing gels with clear background at nanogram sensitivity using Coomassie Brilliant Blue G-250 and R-250, Electrophoresis 9(6) (1988) 255-262)

Mansueto P., Soresi M., Iacobucci R., La Blasca F., Romano G., D’Alcamo A., Carroccio A., 2019. Non-celiac wheat sensitivity: a search for the pathogenesis of a self-reported condition. Italian Journal of Medicine 13:15-23

Pence, J.W., Weinstein, N.E., Mecham, D., The albumin and globulin contents of wheat flour and their relationship to protein quality. Cereal Chem. 1954, 31, 303–311.

Salcedo, G.; Quirce, S.; Diaz-Perales, A. Wheat allergens associated with Baker’s asthma. J. Investig. Allergol. Clin. Immunol. 2011, 21, 81–92.

Sanchez-Monge, R.; Gomez, L.; Garcia-Olmedo, F.; Salcedo, G. New dimeric inhibitor of heterologous □-amylases encoded by a duplicated gene in the short arm of chromosome 3B of wheat (Triticum aestivum L.). Eur. J. Biochem. 1989, 183, 37–40

Scossa F., Laudencia-Chingcuanco D., Anderson O.D., Vensel W.H., Lafiandra D., D'Ovidio R., Masci S., Comparative proteomic and transcriptional profiling of a bread wheat cultivar and its derived transgenic line overexpressing a low molecular weight glutenin subunit gene in the endosperm, Proteomics 8(14) (2008) 2948-2966).

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

Zapatero, L.; Martínez, M.I.; Alonso, E.; Salcedo, G.; Sánchez-Monge, R.; Barber, D.; Lombardero, M. Oral wheat flour anaphylaxis related to wheat α-amylase inhibitor subunits CM3 and CM16. Allergy 2003, 58, 956)

Vensel, W.H., Tanaka, C.K., Cai, N., Wong, J.H., Buchanan, B.B., Hurkman, W.J. 2005. Developmental changes in the metabolic protein profiles of wheat endosperm. Proteomics 5, 1594–1611.

CHAPTER 4:

APPLICATION OF CRISPR-Cas9 TECHNOLOGY TO IMPROVE NUTRITIONAL AND HEALTH VALUE IN DURUM WHEAT

ABSTRACT

Wheat and its derived foods are spread all over the world, representing one of the main food sources. However, in recent years there has been an increase of the incidence of wheat related disorders in the world population. Structural and metabolic proteins, like α-amylase/trypsin inhibitors (ATI) are involved in the onset of wheat allergies (baker's asthma) and likely non-coeliac wheat sensitivity (NCWS). The ATI are encoded by a multigene family dispersed over several chromosomes and they are composed of different subunits. Notably, WTAI-CM3 and WTAI-CM16 subunits are more involved in the onset of bakers' asthma and NCWS.

Here, the editing of WTAI-CM16 and WTAI-CM3 subunits was obtained through CRISPR-Cas9 multiplexing strategy in the Italian durum wheat cultivar Svevo with a marker free approach. The regeneration of plants without selection agents allowed to obtain some T0 homozygous mutant plants without the presence of CRISPR vectors, demonstrating the capability of CRISPR technology to produce safer wheat lines in a reduced time compared to conventional breeding programs.

4.1 INTRODUCTION

Wheat is the major staple food of human diet and is the most cultivated crop all over the world. The nutritional value and the rheological properties of wheat flour dough are due to proteins and starch.
However, in recent years there has been an increase of the incidence of wheat related disorders in the world population. Among the wheat-related pathologies, Coeliac Disease (CD) is the best known, but IgE and non-IgE mediated allergies, wheat-dependent exercise-induced anaphylaxis (WDEIA), non-coeliac wheat sensitivity (NCWS) and hypersensitivity to wheat represent a broad spectrum of disorders resulting from flour ingestion, inhalation or contact.
These adverse reactions to wheat are mainly due to the protein fraction of wheat kernel; gluten proteins are responsible for triggering CD and WDEIA (Tatham and Shewry 2008), whereas structural and metabolic proteins, like α-amylase/trypsin inhibitors (ATI), lipid transfer proteins (LTP), thioredoxins and puroindolines, are involved in the onset of wheat allergies and likely NCWS (Wheichel et al 2006; Battias et al 2008; Henggeller et al 2017, Tatham and Shewry 2008, Mansueto et al 2019).
The ATI form a protein family and represent a substantial fraction of the albumins and globulins of wheat endosperm (2-4 % of total proteins) (Altenbach et al 2011; Henggeler et al 2017). They are inhibitors of heterologous α-amylase and trypsin, thus they play an important protective role against insect, mite and mammalian α-amylase. Their synthesis takes place during seed development as well as main storage proteins, and they are rapidly degraded upon germination and used by the plants as nutritional reserve. They are low molecular weight proteins with molecular weights in the 12-16,000 Da range (about 150 amino acid residues). All ATI contain 10 cysteine residues and form 4-5 interchain disulphide bonds with a common fold: 4-5 α -helices and a short antiparallel β-sheet. Amino acid sequence identity between members of the family ranges from around 30% to 95% (Carbonero and Garcia-Olmedo 1999).
The ATI family includes different subunits and their aggregation gives rise to monomeric (WMAI about 12 kDa), dimeric (WDAI about 24 kDa) and tetrameric forms (WTAI about 60 kDa). The tetrameric inhibitors consist of proteins subunits soluble in chloroform/methanol mixtures (CM proteins) and each WTAI contains one copy of CM1 or CM2 subunits, one copy of CM16 or CM17 subunits and two copies of CM3 subunit (Altenbach et al 2011).
The ATI are encoded by a multigene family dispersed over several chromosomes. They are highly conserved and it was established that there was no intron in the α-amylase inhibitor genes (Wang et al 2008).WMAI are encoded by genes on 6B and 6D chromosomes, WDAI by genes on 3B and 3D and WTAI are encoded by genes on chromosome group 7 and chromosome group 4; in particular the CM3 and CM16 subunits are encoded by genes located on 4B and 4D chromosomes.

In the A genome of wheat (bread wheat and durum wheat) there are some sequences corresponding to the three types of subunits of the tetrameric inhibitors, but these genes are silent or expressed at a low level. No other proteins of this family have been found to be encoded by genes from the A genome and no activity against heterologous α-amylase has been found in extracts from diploid wheat species with the AA genome (Garcia-Maroto et al 1990; Garcia-Maroto et al 1991).

Nowadays, ATI have been considered major allergens in the onset of bakers' asthma and food allergies. They are also claimed to cause non-Coeliac Wheat Sensitivity (NCWS) in predisposed patients along with fermentable oligosaccharides and disaccharides, monosaccharides and polyols (FODMAPs) (Carbonero, Garcia-Olmedo 1999; Thatam & Shewry 2008; Mansueto et al 2019). ATI subunits are the most IgE-reactive proteins of wheat kernel, notably WTAI-CM3 and WTAI-CM16, and the glycosylated form of WTAI-CM16 exhibits the strongest binding to IgE from patients with bakers' asthma.

Wheat breeding can allow the development of varieties safer for both consumers and employees of the food industry. Genome editing tools have rapidly evolved in last years and in particular, CRISPR-Cas9 technology has been reported in many plant species, including durum and bread wheat.

Gene editing remains challenging in transformation-recalcitrant species, with relatively few experiments reported in the major crop species. The progress made in vector assembly, delivery systems and Cas proteins (Xie et al 2015; Zhang et al 2016) has allowed the editing of wheat varieties for multiple purposes, including the production of low immunogenic wheat varieties for patients with CD (Sanchez-Leon et al 2018).

In the present work, the multiplex editing of WTAI-CM16 and WTAI-CM3 subunits of the tetrameric form was obtained through CRISPR-Cas9 knockouts in the Italian durum wheat cultivar Svevo with a marker free approach, in order to obtain edited GM-free plants since the first generation.

4.2 MATERIALS AND METHODS

4.2.1 Plant Materials

The durum wheat cultivar Svevo, used for biolistic transformation, was grown in growing chamber under controlled conditions. After 4 weeks vernalization at 4-5 °C, growing conditions were 18-20 °C day and 14-15 °C night temperatures under a 16-h photoperiod. The same conditions were subsequently used for mutant plants growth.

4.2.2 gRNAs design

The guide RNA targets based on the coding sequence of CM3 and CM16 genes were designed using Deskgen Cloud (www.deskgen.com). The target genes were selected using the Ensembl names Traes_4BL_CE02BBD12 (CM3 gene) and Traes_4BL_B12EAE358 (CM 16 gene) on the reference genome IWGSC RefSeq v1.0 (International Wheat Genome Sequencing Consortium). Four guides were chosen for CM3 gene and three guides for CM16 in order to disrupt the gene sequence (Tab. 4.1). The ON target and OFF target activity predicted by the platform were taken into account along with the location of the OFF target hits across the genome, considering those without target specificity in coding sequences. The secondary structure of each gRNAs was also considered and predicted by RNAfold web server (rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi).

The nucleotide sequences covering around 200 bp upstream the start codon and 100 bp downstream of the stop codon were used to BLAST the durum wheat sequences of IDWGSC, Durum Wheat (cv. Svevo) RefSeq Rel. 1.0 (http://d-data.interomics.eu) in order to analyse their homology, and to assess the presence of other genes with similar coding sequence. Furthermore, two specific primer pairs were used to amplify the CM3 (CM3_104-F and CM3_1009-R) and CM16 genes (CM16_58-F and CM16_821-R) of the wild type plants and the PCR amplicons used for sequencing. All the sequence shared 100 % of homology in their coding sequence.

***Table 4.1. Single guides RNA designed using Deskgen platform**. DNA strand, on and off target activity, distance to start codon, off target hits in coding (COD) and non-coding (Non COD) regions with mismatches (ms) are reported.*

									OFF TARGET HITS		
ID	**SEQ**	**PAM**	**STR**	**%GC**	**ON**	**OFF**	**bp to ATG**		**COD (ms)**	**Non COD(ms)**	**Chr. locus**
CM3_g1	CAGACTTGGCTAGAATACCA	TGG	1	45	63	50	-2		0	1(0);9(3)	7B; 7A; 2B
CM3_g2-	CACCCCTGGGACGCAGCTGC	CGG	-1	75	42	48	82		0	1(0);31(3)	6B; 3A; 7A
CM3_g3	TTTGTTGTAACACATAGTCG	CGG	-1	35	67	49	135		0	4(3);1(1)	2B; 7B
CM3_g5+	TTCTCAGCAGTGCCGGTGCG	AGG	1	65	53	47	287		0	27(3);1(2);1(0)	3B; 7B; 2A; 4A
CM16_g1	GTCCAACTGCGTTCTCCTCT	TGG	1	55	42	99	28		0	1(3)	3A
CM16_g4	CATGTCGCATCGAAACGCCC	GGG	1	60	60	97	172		0	4(3);1(2)	2B; 5B;
CM16_g6	AAGTCTCGTCCGGATCAGAG	CGG	1	55	76	59	288		0	15(3);1(1)	1A; 3B

4.2.3 Plasmid vector assembly

The selected guides were synthesised in a tRNA/gRNA architecture to produce simultaneously multiple gRNAs, as reported by Xie at al (2015). The polycstronic tRNA/gRNA (PTG) containing all the guides was constructed based on the Golden Gate (GG) assembly principle, starting from the pGTR plasmid (Addgene), and cloned into pRRes209RRM_482 vector (provided by Rothamsted Research) under the control of the rice U3 small necleolar RNA promoter (U3 snoRNA promoter). The pRRes226RRM.486_Cas9 vector (Rothamsted Research) contains a wheat optimized Cas9 under the control of a maize ubiquitin promoter (Zm Ubim3), two Nuclear Localization Signals (NLS) and nopaline synthase terminator (NosT) (Fig. 4.1).

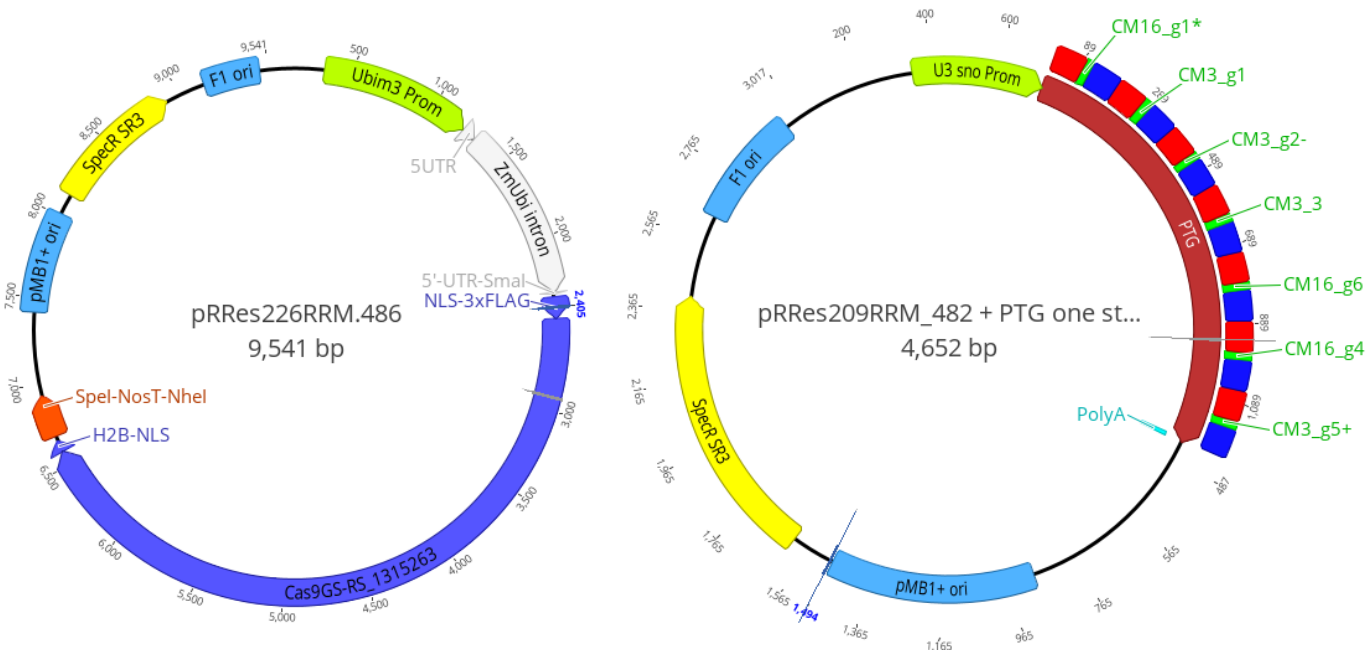

***Figure 4.1. CRISPR CONSTRUTS.** Vectors used for biolistic transformation of wheat embryos.*

4.2.3.1 Design primers for Golden Gate assembly

The GG assembly requires distinct 4-bp overhangs to ligate all DNA parts containing the conserved tRNA and gRNA scaffold sequences and the target-specific spacer. 4-bp long sequences were selected within each spacer as BsaI overhangs in order to split the gRNAs into two overlapping DNA parts; each half of the spacer was synthesised within oligo primers with a *BsaI* site attached to 5' ends (Tab. 4.2). All DNA parts should have different 4-bp overhangs among them and different to the sequences used as adaptors for cloning into pRRes209RRM_482 in order to perform a one direction reaction. Each forward primer anneal to the 5' end of gRNA scaffold and each reverse primer anneal to the 3' end of the pre tRNA of the pGTR, which contains a gRNA/tRNA fused fragments.

***Table 4.2. Primers used for GG assembly**. Highlighted in green the BSAI sites, in red tRNA sequence and in blue gRNA scaffold sequence*

PRIMER ID	5' – 3' sequence
BsaI_tRNA-F	TAGGTCTCAGGCAACAAAGCACCAGTGG
CM16_g1*R	CGGGTCTCACGCAGTTGGACTGCACCAGCCGGGAA
CM16_g1*F	CGGGTCTCCTGCGTTCTCCTCTGTTTTAGAGCTAGAA
CM3_g1R	TAGGTCTCATAGCCAAGTCTGTGCACCAGCCGGG
CM3_g1F	CGGGTCTCCGCTAGAATACCAGTTTTAGAGCTAGAA
CM3_g2-R	TAGGTCTCATCCCAGGGGTGTGCACCAGCCGGG
CM3_g2-F	CGGGTCTCCGGGACGCAGCTGCGTTTTAGAGCTAGAA
CM3_g3R	TAGGTCTCAGTTACAACAAATGCACCAGCCGGG
CM3_g3F	CGGGTCTCCTAACACATAGTCGGTTTTAGAGCTAGAA
CM16_g6R	CGGGTCTCAACGAGACTTTGCACCAGCCGGGAA
CM16_g6F	CGGGTCTCCTCGTCCGGATCAGAGGTTTTAGAGCTAGAA
CM16_g4R	TAGGTCTCAATGCGACATGTGCACCAGCCGGG
CM16_g4F	CGGGTCTCCGCATCGAAACGCCCGTTTTAGAGCTAGAA
CM3_g5+R	TAGGTCTCAACTGCTGAGAATGCACCAGCCGGGAA
CM3_g5+F	CGGGTCTCCCAGTGCCGGTGCGGTTTTAGAGCTAGAA
BsaI_gRNA-R	TAGGTCTCCAAACAAAAAAAAAAGCACCGACTCG

4.2.3.2 PCR to amplify DNA parts

DNA parts used for PTG building were amplified performing 50 µl PCR reactions per each primer pair (reported in tab. 4.2). The PCR reactions were performed with 12.5 µl of 2X GoTaq® G2 Hot Start Colorless Master Mix (Promega), 0.5 µM of each primer, 20 ng of template DNA (pGTR) and nuclease-free water up to 25 µl, in the following conditions: 95 °C for 2 min, followed by 35 cycles at 95 °C for 1 min, 61 °C for 30 s, 72 °C for 30 sec and a final extension at 72 °C for 5 min.

PCR amplicons were checked on 2% agarose gel electrophoresis and purified using NucleoSpin Gel and PCR Clean-up kit (Machery-Nagel, Duren, Germany), recovering PCR amplicons from the gel. Finally, purified DNA was eluted in 30 µl of *Buffer NE* containing 5mM Tris/HCl, pH 8.5 and quantified with Nanodrop (Thermo Scientific™).

4.2.3.3 Golden Gate assembly

The GG assembly reaction to ligate individual DNA parts and to build the PTG was performed in a final volume of 20 µl containing 10 µl of 2x T7 DNA ligase buffer, 2 µl of Bovine Serum Albumin (1mg/ml), 0.5 µl of BsaI-HF (20U/µl, NEB), 0.5 µl of T7 DNA ligase (3000 U/µl, NEB), 50 ng of DNA parts and 15 ng of pRRes209RRM_482-PTG. The GG reaction was incubated at 37 C, 5 min and 20 °C, 10 min for 50 cycles with a final step at 20 °C, 1h. The *BsaI* sites added to 5' ends of BsaI_tRNA-F and BsaI_gRNA-R primers generates 4-bp overhangs that allow to ligate the PTG into pRRes209RRM_482 in one step and one tube reaction.

The GG product (pRRes209RRM_482-PTG) was transformed to *E.coli* 10-beta competent cells (C3019H, NEB), according to NEB protocol. Plates were prepared with YT medium plus spectinomycin (75mg/ml) at final concentration of 75 ng/µl. 100 µl cell culture were plated onto plate and incubated over night at 37 C. Four colonies were inoculated in 5ml of YT medium plus 5µl of spectinomycin (75mg/ml) and incubated over night at 37 C with 230-240rpm shaking. Plasmids were extracted with the Wizard SV Miniprep DNA purification System (Promega) and quantified with Nanodrop. The extracted plasmids and pRREs_209RRM_482 (used as control) were double digested with *MluI*-HF and PstI for 2 h at 37 C, with following 10 µl reaction: 1 µl of 10x NEBuffer 3.1, 0.35 µl of *MluI*-HF (20U/µl), 0.35 µl of *PstI* (20U/µl) and 250 ng of plasmid DNA. Finally, the construct was confirmed by Sanger sequencing using primer.

4.2.4 Biolistic transformation of wheat embryos and tissue culture

Plasmid vectors pRRes226RRM.486_Cas9 and pRRes209RRM_482-PTG were used to co-bombard embryos isolated from durum wheat cultivar Svevo 10-15 days post anthesis. The embryos axis were removed to prevent precocious germination and the scutella were placed within 9-cm Petri dish containing the dissection medium, orientating with the uncut scutellum surface directed upwards. The isolated embryos were pre-cultured in the dark at 22 °C for 2 days prior to bombardment. The plasmid vectors were coated onto 0.6 µm gold particles and used for biolistic bombardment in a 1:1 molar ratio, performed with a PDS1000/He particle bombardment system

(Bio-Rad) according to Rasco-Gaunt et al (2001). After bombardment, the scutella were stored for 24 h at 22 °C and, after that, spread across the medium. After 3-5 weeks of culture for the callus induction phase, calli showing somatic embryos were transferred to regeneration medium in 9-cm Petri dishes and incubated at 22°C in the light for 4 weeks. These were transferred to rooting medium and cultured for 6-8 weeks. Regenerating shoots were transferred to regeneration medium without hormones in GA-7 Magenta vessels and then potted into soil once the root system has been set up and grown on to the flowering stage in controlled environment rooms. The entire tissue culture process was carried out without using selective agents, since no selectable markers have been used for the transformation. The plantlets were sampled for T_0 mutant screening during the Magenta phase.

Table 4.3. Composition of media used during tissue culture.

Medium	Component
Dissection	Murashige & Skoog Salt mixture 4.3 g/l Maltose 40 g/l Thiamine-HCl (25 mg/500 ml) 10 ml/l; L-asparagine 0.15 g/l 0.25% v/v phytagel 2,4 D (0.5 mg/ml) 2 ml/500 ml
Regeneration	Murashige & Skoog Salt mixture 4.3 g/l Maltose 40 g/l Thiamine-HCl (25 mg/500 ml) 10 ml/l L-asparagine 0.15 g/l 0.25% v/v phytagel 2,4 D (0.5 mg/ml) 0.2 ml/500 ml
Rooting	Murashige & Skoog Salt mixture 2.15 g/l Maltose 20 g/l Thiamine-HCl (25 mg/500 ml) 5 ml/l L-asparagine 0.075 g/l 0.25% v/v phytagel

4.2.5 DNA extraction

DNA was extracted from leaves of T_0 plants using NucleoSpin® Plant II (Macherey-Nagel) according to manufacturer's instruction and used for genotyping of mutant plants and detection of transgene.

4.2.6 Large mutations screening

The presence of large deletions in the transformed plants was assessed by PCR. Two specific primer pairs (Tab. 4.4) encompassing the potential mutations induced by CRISPR/Cas9 were used to amplify 437 bp of *CM3* gene and 451 bp of *CM1*6 gene from DNA samples of plantlets, with the following reaction: 10 µl of 2X GoTaq® G2 Hot Start Colorless Master Mix (Promega), 0.5 µM of each primer 20 ng of template DNA and nuclease-free water up to 20 µl, on following conditions: 95 °C for 2 min, followed by 35 cycles at 95 °C for 1 min, 63 °C for 30 s, 72 °C for 45 sec and a final extension at 72 °C for 5 min.

PCR amplicons were checked on 2% agarose gel electrophoresis and amplicons with different mobility as compared to control were sequenced through Sanger Sequencing.

4.2.7 High Resolution Melting Analysis

The HRM analysis was carried out on CM3 and CM16 gene fragments amplified on genomic DNA of T_0 plantlets in order to identify mutant plants carrying short insertion/deletions (IN/DELs) or single nucleotide polymorphisms (SNPs).

Amplicons suitable for HRM analysis were produced by a nested PCR strategy to obtain two short fragments per gene, using primer pairs (Tab. 4.4). The first PCR was carried out to amplify 437 bp fragment from CM3 gene and 451 bp fragment from CM16 gene. The reactions were performed in a 10 µl volume using 5 µl of 2X GoTaq® G2 Hot Start Colorless Master Mix (Promega), 10 ng of genomic DNA, 0.5 µM of each primer, on following conditions: 95 °C for 2 min, followed by 35 cycles at 95 °C for 1 min, 63 °C for 30 s, 72 °C for 45 sec and a final extension at 72 °C for 5 min.

The PCR reactions were diluted 60 fold and 1 µl was used as template for the nested PCR to produce short fragment for analysis. The reactions were performed with 5 µl of 2X GoTaq® G2 Hot Start Colorless Master Mix (Promega), 1 µl of LC Green Plus, 0.5 µM primers (Tab. 4.4) and nuclease-free water up to 10 µl. The reactions were overlaid with 10 µl of mineral oil (Sigma-Aldrich M5904) and the PCR program used was: 95 °C for 2 min, followed by 35 cycles at 95 °C for 1 min, 63 °C for 30 s, 72 °C for 45 sec and a final extension at 72 °C for 5 min. After the final extension step, PCR amplicons were denatured at 95°C for 30 s and reannealed at 25°C for 1 min.

Following nested PCR, plates were transferred to a Lightscanner (Idaho Technology Inc., USA) and subjected to HRM analysis according to Tan et al. (2016). Samples with relative fluorescence differences ($|\Delta F|$) > 0.05 respect to the reference (Hofinger et al. 2009) were considered as putative mutants thus the entire amplicons of the gene were re-amplified from each individual DNA for sanger sequencing.

4.2.8 Tracking of Indels by decomposition analysis (TIDE)

Mutants plants identified by HRM analysis and showing overlapping peaks after Sanger sequences, were re-amplified with primer pair reported in tab 4.4 and submitted to TIDE web tool in order to assess the efficiency of gene editing resulting from NHEJ-mediated repair of the target loci (Brinkman et al. 2014). The PCR conditions for re-amplification of longer fragments of CM3 and CM16 genes were: 95 °C for 2 min, followed by 35 cycles at 95 °C for 1 min, 61.5 °C for 30 s, 72 °C for 1 min and 15 sec and a final extension at 72 °C for 5 min

Table 4.4. Primers used to amplify CM3 and CM16 and for HRM analysis.

	PRIMER ID	5' - 3' SEQUENCE	Fv/Rv	Amplicon size (bp)	Application
***CM3* GENE**	**CM3_104_F**	AGACAGCTTGCATTCCCCTC	Fv	906	**TIDE**
	CM3_1009_R	CGCTTTGGCCTTTGACAGAC	Rv		
	CM3_106_F	ATATGCGAGCATGTGTGGCG	Fv	437	
	CM3_542_R	TCACCAACATTGCCGGACCT	Rv		
	CM3_123_F	GCGCATCCAGCATAAACTCA	Fv	185	Nested PCR
	CM3_307_R	AGTGTGGCAGAAGATTGGTC	Rv		
	CM3_288_F	GACCAATCTTCTGCCACACT	Fv	213	Nested PCR
	CM3_500_R	ACCGGCAACGCTATGAAGTA	Rv		
***CM16* GENE**	**CM16_58_F**	AGCTTGCATTCCCCTCCAAA	Fv		
	CM16_821_R	AGCACCACACATGATGAGCG	Rv	764	**TIDE**
	CM16_128_F	CGTGTAGCATACACCCACCA	Fv	451	
	CM16_578_R	TCGTCAAGTTGCAGTACCCC	Rv		
	CM16_169_F	ATACCAACGAACTGGGCCTG	Fv	127	Nested PCR
	CM16_296_R	TGGGGTGCAATCTTCATTGC	Rv		
	CM16_331_F	CCGTGACTATGTGGAACAAC	Fv	211	**Nested PCR**
	CM16_542_R	TCACGAAGTCCATCTGCACC	Rv		

4.2.9 Detection of transgene

The genomic DNA extracted from plantlets was used as template for PCR to detect the presence of CRISPR constructs. The reactions were performed with four primer pairs (Tab. 4.5) and in a final volume of 20 µl as follow: 10 µl of 2X GoTaq® G2 Hot Start Colorless Master Mix (Promega), 0.5 µM of each primer 20 ng of template DNA and nuclease-free water up to 20 µl, on following conditions: 95 °C for 2 min, followed by 35 cycles at 95 °C for 1 min, 61 °C for 30 s, 72 °C for 1 min and a final extension at 72 °C for 5 min. PCR amplicons were checked on 1.5% agarose gel to confirm the presence of transgene.

Table 4.5. Primers used for the detection of CRISPR constructs.

	PRIMER ID	5' - 3' SEQUENCE	Fv/Rv
pRRes226RRM.486_Cas9	**p486_5394_F**	CAAGGTCCGCGAGATCAACA	Fv
	p486_3529_F	TTCGACCAGTCCAAGAACGG	Fv
	p486_4059_R	CTTCACCTTGGTCAGCTCGT	Rv
	p486_3210_R	GGAGAGAGCGATGAGGTTGC	Rv
	p486_6228_F	CAACGAGCAAAAGCAGCTGT	Fv
	p486_6672_R	CATGCGGCCGCTAAGTTAAC	Rv
pRRes209RRM.482-PTG	**p482_U3_F**	GCTTAATGCGCCGCTACAGG	Fv
	p482_U3_2F	TGGTAGCTCTTGATCCGGCAAA	Fv
	p482_U3_R	TGCCACGGATCATCTGCACA	Rv
	p482_U3_2R	TTGCGCTGCCATTCTCCAAATT	Rv

4.2.10 Analysis of off-target mutation

The off-targets were predicted by DESKGEN Cloud per each gRNA targeting CM3 or CM16 gene. Furthermore, the seed sequences (12 nt upstream of the PAM) of each gRNA were blasted against RefSeq Rel. 1.0 (http://d-data.interomics.eu) and searched among coding sequences, tolerating up to two mismatches.

4.3 RESULTS

The target genes (*CM3* and *CM16*) are present in durum wheat as a single copy on the 4B chromosome. Since the DESKEGEN CLOUD works on bread wheat reference genome, the coding sequences with partial 5' UTR and 3 'UTR of Traes_4BL_CE02BBD12 (*CM3* gene) and Traes_4BL_B12EAE358 (*CM16* gene) used to run the software, were compared to gnl|Td_Svvo|whe_Td_AB_Svevo_4B (*CM3* gene) and to gnl|Td_Svvo|whe_Td_AB_Svevo_4B (*CM16 g*ene) retrieved from IDWGSC, Durum Wheat (cv. Svevo) RefSeq Rel. 1.0 (http://d-data.interomics.eu). The alignment among the coding sequences showed 100% of identity between genes of the two species and it was further confirmed by sequencing the *CM3* and *CM16* genes amplified from Svevo wild type plants.

To disrupt the entire coding sequence of target genes and thus to block the transcription, 4 gRNAs targeting the *CM3* coding sequence and 3 gRNAs targeting *CM16* coding sequence were synthesized in the PTG architecture, as reported by Xie et al (2015) and cloned into expression vector under the control of single U3p (Fig. 4.2). The modified GG assembly, which was improved by replacing the *FokI* sites at the 5'prime and 3' ends of the primers needed for the first and last DNA parts amplification, allowed to obtain the PTG in only one step and one tube reaction; the *MluI* and *PstI* double digestion and subsequent sequencing of construct confirmed that all DNA parts have been ligated in the desired order resulting in a single synthetic gene containing the spacers and gRNA scaffold linked to tRNA sequences at both 3' and 5' ends.

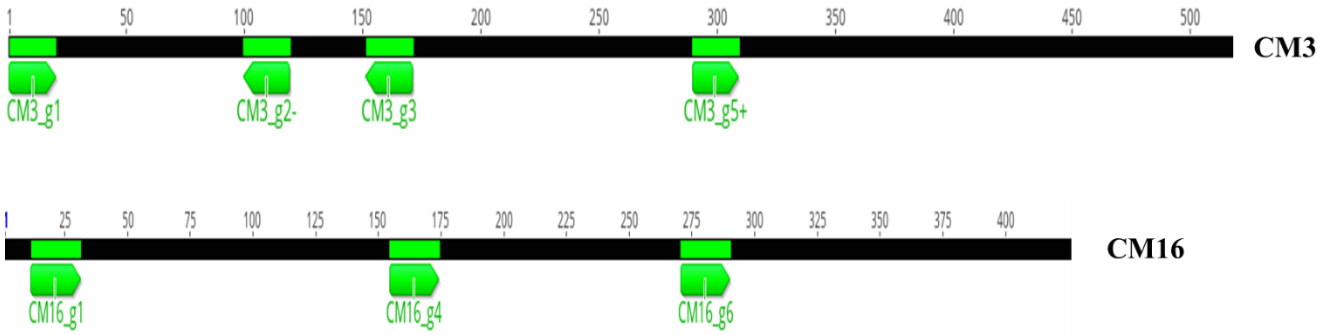

Figure 4.2. Distribution of gRNAs along gene sequences.

The CRISPR/Cas9 vectors pRRes226RRM.486_Cas9 and pRRes209RRM_482-PTG, checked by Sanger sequencing, were used in three rounds of biolistic bombardment to transform embryos isolated from Svevo caryopsis in 1:1 molar ratio.

The tissue culture of transformed calli was carried out using media without selection agents since no vectors containing selectable markers were used for the biolistic transformation. A large number of T_0 plants were obtained in a reduced time as compared to standard procedures of regeneration using selection agents (9-10 weeks instead of 12-15 weeks), although this caused a difficult mutation screening step due to the high number of samples to analyse.

In total, 495 embryos were bombarded and 245 plants were regenerated, 23 of which carried different types of editing events involving IN/DEL and SNP that cause mutations in *CM3* and *CM16* coding sequences (tab. 4.6).

Table 4.6: Summary table. *CRISPR-Cas9 editing events.*

Bombardment ID	**transformed embryos**	**No. regenerated plants**	**No. Edited plants**	**No. editing events (CM3)**	**No. editing events (CM16)**
B3725	190	97	8	8	7
B3726	210	99	7	7	2
B3681	95	49	4	4	4
TOT	495	245	19	19	13

Since the number and positions of gRNAs caused large mutations, some editing events were clearly visible by gel electrophoresis of *CM3* and *CM16* genes amplified by PCR (Fig. 4.3). On the other hand, the HRM analysis allowed to find small IN/DELs and SNPs; the HRM analysis was conducted on two fragments per gene (*CM3* and *CM16*) amplified from genomic DNA of T_0 plantlets (Fig. 4.3, Tab. 4.7). The nucleotide sequence of *CM3* and *CM16* genes of plants identified as potential mutants were genotyped through Sanger sequencing (Fig.4.3).

The *CM3* gene sequence of T_0 plants B3681-16 and B3681-49 showed two identical deletions respect to the wild type reference sequence; 101bp deletion encompassing the first two gRNAs (CM3_g2- and CM3_g3) and 10 bp deletion encompassing the gRNA CM3_g5+. The *CM16* gene sequence of B3681-16 plant showed overlapping peaks resulted from the Sanger sequencing, thus it was submitted to TIDE analysis which estimated a -10 bp deletion frequency of 11 % with p value <0.001 and $R^2 = 0.97$, at the level of the gRNA CM3_g5+. The *CM16* gene sequence of B3681-49 plant showed a large deletion (143 bp), starting 3 bp upstream of the PAM sequence of the guide

CM16_g1 and ending 3 bp upstream of the PAM sequence of guide CM16_g4, and a large insertion of 108 bp starting 3 bp upstream of the PAM sequence of guide CM16_g6.

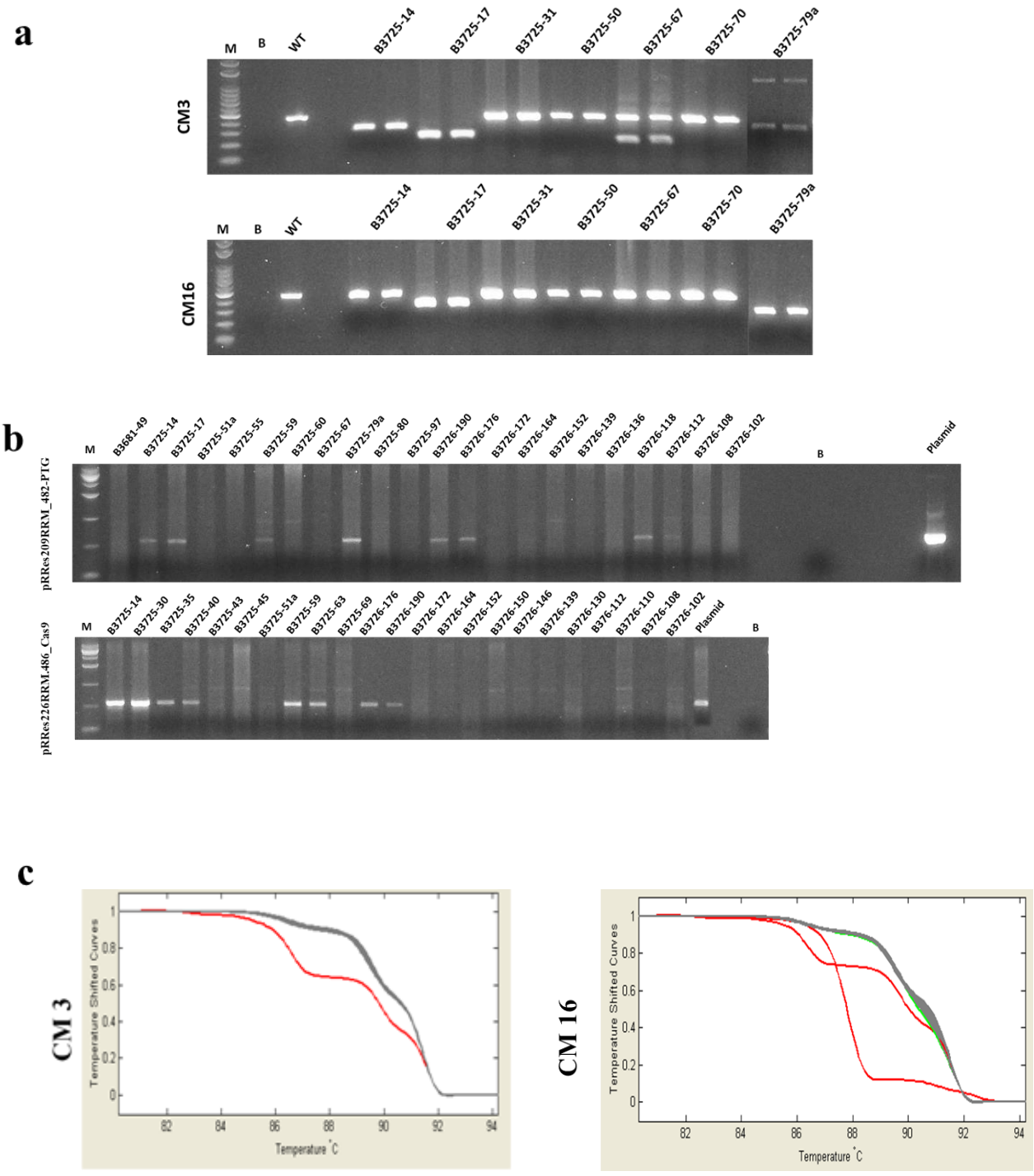

Figure 4.3. Molecular analysis of CRISPR mutants. ***(a)*** *The outcome of PCR assays analysing CM3 and CM16 mutants is shown.* ***(b)*** *CRISPR construct detection.* ***(c)*** *HRM-analysis*

The T_0 plant B3725-14 showed both *CM3* and *CM16* gene sequences mutated; three types of deletions were detected along *CM3* nucleotide sequence of 4, 1 and 116 bp at the level of CM3_g1 and CM3_g5+ guides, whereas the CM16 gene sequence had one SNP in position 270 from the ATG (C -> G) close to CM16_g6 guide. Similarly, the T_0 plant B3725-17 showed three deletions along the *CM3* nucleotide sequence; the first and the second one (-2 bp and -1 bp) caused by CM3_g1 guide before the transcriptional start codon and the third (205 bp) occurred between the CM3_g2- and CM3_g5+ guides. The electrophoretic mobility of CM16 PCR amplicon from plant B3725-17 was clearly different respect to the control and the Sanger sequencing did not give clear results due to the presence of overlapping peaks, not useful to detect precise IN/DELs (probably due to different type of deletions present as biallelic mutations).

The plants B3725-31 and B3725-59 showed the same type of mutations for both *CM3* and *CM16* genes: -1 bp deletion along the *CM3* gene sequence 2 bp upstream of the PAM sequence of CM3_g1 guide and one SNP in position 270 (C -> G) along the *CM16* gene sequence.

The T_0 plants B3725-46, B3725-51a, B3726-118, B3726-136 and B3726-176 also contained the same mutation already described in B3725-31 and B3725-59 in the *CM3* gene sequence. The TIDE analysis of CM3 PCR amplicon of plant B3725-51a estimated the frequency of -46 bp deletion was of 7 % with p value <0.001 (p value=3.2^{e-05}) and R^2 = 0.94, whereas for the plant B3726-118 the frequency of -4 bp deletion was 5.2% with p value = 7.8^{e-10} and R^2 = 0.91.

The Sanger sequencing of *CM16* gene obtained from B3725-51a plant showed 4 SNPs and 6 IN/DELs of 1 bp respect to the wild type sequence. An insertion of + 2 bp was estimated with a frequency of 13.8% with p value = $2.9e^{-27}$ R^2 = 0.93 by TIDE along the *CM16* gene sequence of B3726-118. TIDE has also found a -24 bp deletion in *CM3* gene of B3726-176 plant with a frequency of 3.9 % (p value = 0.00035, R^2 = 0.91).

The electrophoresis of CM3 PCR amplicon of plant B3725-67 showed two bands with different mobility. Each band was sequenced: the higher band nucleotide sequence reported the same -1 bp deletion of B3725-31 plant, whereas the lower band carried two deletions of 86 bp between CM3_g1 and CM3_g2- guides and 153 bp between CM3_g3 and CM3_g5+ guides. The *CM16* gene sequence of plant B3725-67 showed one SNP in position 270 (C -> G).

Large IN/DELs were also detected along *CM3* and *CM16* gene sequences of the B3725-79a T_0 plant: the *CM3* gene sequence could not be reconstructed from Sanger sequencing results and showed a particular electrophoretic mobility, whereas the *CM16* gene sequence showed a deletion of -163 bp starting 11 bp upstream of the PAM sequence of CM3_g1 guide. This is probably due to the presence of high percentage/rates of biallelic mutations.

The *CM3* gene sequence of B3681-14 plant showed two SNPs in position 276 (G -> T) and 289 (G-> A) from ATG start codon, whereas the *CM16* gene sequence resulting from Sanger sequencing showed overlapping peaks that did not allow the reconstruction of the entire gene sequence.

Figure 4.4. Genotyping of CRISPR mutants. (a) *Sanger sequencing of some representative CM3 mutants.* ***(b)*** *Sanger sequencing of some representative CM16 mutants*

Some samples showing relative fluorescence differences [(|ΔF|) > 0.05] respect to the reference after the HRM analysis of *CM3* and *CM16* genes, that resulted non-mutated on the basis of Sanger sequencing but with overlapping peaks, were re-amplified and submitted to TIDE. The TIDE analysis of *CM3* PCR amplicon of plant B3681-17 estimated the frequency of -3 bp deletion was of 17.5 % with p value <0.001 (p value=1.8^{e-27}) and R^2 = 0.92 and the frequency of -9 bp deletion of CM16 PCR amplicon was 12.4 % with p value <0.001 (p value=1^{e-20}) and R^2 = 0.94. In the plant B3726-176, deletions of – 24 bp and – 12 bp in *CM3* and *CM16* genes, respectively, were estimated by TIDE with a frequency of 3.9 %, p value = 8.4 e^{-07} and R^2 = 0.90 for *CM3* gene and with a frequency of 7.4 %, p value = 1.7 e^{-12} and R^2 = 0.96 for *CM16* gene. The *CM3* PCR amplicons of B3726-179, B3726-164 and B3726-190 plants showed deletions of -46 bp, with a frequency of 2 % (p value = 3.2 e^{-08} and R^2 = 0.90), 6.2 % (p value = 9.4 e^{-15} and R^2 = 0.90) and 13 % (p value = 1.5 e^{-87} and R^2 = 0.90), respectively. The TIDE analysis of B3726-164 also found an insertion of + 6 bp in *CM3* gene with a frequency of 1.1% (p value = 0.0002 and R^2 = 0.97).

To assess the presence of CRISPR vectors in the T_0 plants, PCR were performed with primer pairs designed across ubiquitin promoter region and gRNAs cassette for p482 plasmid and across Cas9 sequence and spectinomycin gene for p486 plasmid, reported in tab 4.7. The CRISPR vectors were absent in some T_0 edited plants and the gRNAs vector was found more frequently respect to Cas9 vector (Fig. 4.3, Tab. 4.7). The plants showing the presence of both plasmid vectors were also selected and will be analysed in the T_1 generation (Tab. 4.7).

None off-targets were predicted and detected by *in silico* analysis in coding region along the bread wheat reference genome. The *in silico* search for off-target sites to the whole genome of durum wheat were expanded by searching for perfect matches of the seed sequence (12 nt) plus PAM in the reference genome of durum wheat.

Table 4.7. ***Summary table of edited plants.*** *Plants showing different type of editing in CM3 and CM16 genes. gRNAs and Cas9 indicates plants containing plasmid vectors.*

EXP.ID	PLANT ID	EDITED GENES	CM3	TYPE	CM16	TYPE2	gRNAs	Cas9
B3681	B3681-17	2	DEL	TIDE (-3)	DEL	TIDE (-10)	√	√
	B3681-16	1	DEL	101;10	-			
	B3681-14	1	SNP	276G->T;289G->A	-			√
	B3681-49	2	DEL	101;10	DEL/INS	143;108		
B3725	B3725-14	2	DEL	4;1;116	1SNP	270C->G	√	√
	B3725-17	2	DEL	2;1;205	DEL	-	√	
	B3725-31	2	DEL	1	1SNP	270C->G		
	B3725-46	1	DEL	1	-			
	B3725-51A	2	DEL	1	DEL/SNP	-		
	B3725-59	2	DEL	1	SNP	270C->G	√	√
	B3725-67	2		86;153	SNP	270C->G		
	B3725-79A	2	DEL		DEL	163	√	
B3726	B3726-118	2	DEL	1	INS	TIDE (+2)	√	
	B3726-136	1	DEL	1	-			
	B3726-176	2	DEL	TIDE(-24)	DEL	TIDE (-12)	√	√
	B3726-190	1	DEL	TIDE (-46)	-		√	√
	B3726-164	1	DEL	TIDE (-46)	-			
	B3726-179	1	DEL	TIDE (-46)	-			

4.4 CONCLUSION

In the last five years, the CRISPR-Cas9 has become the predominant gene editing tool for crop breeding. Since the perceived increase of wheat related disorders in the last 50 years, wheat breeding focuses also on healthy aspects and the CRISPR-Cas9 technology can led to the development of safer wheat varieties in a reduced time compared to conventional breeding programs (Sanchez-Leon et al 2018). Furthermore, CRISPR has received much more attention in crop breeding because of its main features: the high efficiency and accuracy of gene editing, the multiplexing capability (Baltes and Voytas 2015; Xie et al 2015), the minimized chance of off-target mutations and the possibility to obtain transgene-free mutant plants in the first generation (Zhang et al 2016).

Here, to obtain T_0 mutant plants for *CM3* and *CM16* genes in the durum wheat cultivar *Svevo*, we adopted the multiplex gene editing strategy reported for the first time by Xie et al (2015). In total, 7 gRNAs targeting the two genes of interest were assembled under the control of a single RNA polymerase III promoter and processed inside the cell, exploiting the tRNA processing system of the plants. No substantial differences were observed in the efficiency of each gRNAs. The use of multiple guides targeting each gene allowed to obtain plants with large deletions, such as B3681-16, B3681-49, B3725-14, B3725-17, B3725- 67 and B3725-79a plants; the plants B3681-49, B3725-17 B3725-79a showed deletions in both *CM3* and *CM16* genes.

The majority of mutant plants showed homozygous mutations for CM3 gene (B3681-14, B3681-16, B3681-49, B3725-14, B3725-17, B3725-31, B3725-46, B3725-51a, B3725-59, B3725-67, B3725-79a, B3726-118 and B3726-136) and for CM16 gene (B3681-49, B3725-14, B3725-17, B3725-31, B3725-51a, B3725-59, B3725-67 and B3725-79a), whereas some plants carried putative biallelic mutations in one or both target genes, probably due to a lower expression of gRNAs or Cas9 cassettes.

The regeneration of plants without selection genes allowed to obtain some T_0 mutant plants without the presence of CRISPR vectors. Interestingly, B3681-49 and B3725- 67 plants containing large deletions in both *CM3* and *CM16* genes did not show the presence of CRISPR vectors. Homozygous transgene-free wheat lines edited for the *CM3* and *CM16* gene were obtained in the T_0 generation, also demonstrating the capability of CRISPR technology to produce improved wheat lines in a reduced time compared to conventional breeding and with a substantial equivalence to its parental line. Moreover, we selected also plants containing both plasmid vectors in order to evaluate the transgenerational activity of Cas9 and gRNAs in the following generations, as reported by Wang et al. (2018).

None off-target effect were predicted by DESKGEN Cloud in the coding region along the A and B genome of Chinese Spring reference sequence and no hits were found when blasting the core sequence of each gRNAs against the reference sequence of durum wheat cv. Svevo, taking into account the coding regions.

Since we transformed an elite durum wheat variety, the edited plants could be grown as safer wheat line for predisposed individuals to baker's asthma, food allergies and NCWS. This material could be also used for plant breeding programmes for the introgression of hypoallergenic traits into other elite wheat varieties.

REFERENCES

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*(1), 242.

Baltes, N. J., & Voytas, D. F. (2015). Enabling plant synthetic biology through genome engineering. *Trends in biotechnology*, *33*(2), 120-131.

Battais, F., Richard, C., Jacquenet, S., Denery-Papini, S., & Moneret-Vautrin, D. A. (2008). Wheat grain allergies: an update on wheat allergens. *European annals of allergy and clinical immunology*, *40*(3), 67.

Carbonero Zalduegui, P., & García Olmedo, F. (1999). A multigene family of trypsin/a-amylase inhibitors from cereals.

García-Maroto, F., Maraña, C., Mena, M., García-Olmedo, F., & Carbonero, P. (1990). Cloning of cDNA and chromosomal location of genes encoding the three types of subunits of the wheat tetrameric inhibitor of insect α-amylase. *Plant Molecular Biology*, *14*(5), 845-853.

García-Maroto, F., Carbonero, P., & García-Olmedo, F. (1991). Site-directed mutagenesis and expression inEscherichia coli of WMAI-1, a wheat monomeric inhibitor of insect α-amylase. *Plant molecular biology*, *17*(5), 1005-1011.

Henggeler, J. C., Veríssimo, M., & Ramos, F. (2017). Non-coeliac gluten sensitivity: a review of the literature. *Trends in Food Science & Technology*, *66*, 84-92.

Mansueto, P., Soresi, M., Iacobucci, R., La Blasca, F., Romano, G., D'Alcamo, A., & Carroccio, A. (2019). Non-celiac wheat sensitivity: a search for the pathogenesis of a self-reported condition. *Italian Journal of Medicine*, *13*(1), 15-23.

Sánchez-León, S., Gil-Humanes, J., Ozuna, C. V., Giménez, M. J., Sousa, C., Voytas, D. F., & Barro, F. (2018). Low-gluten, nontransgenic wheat engineered with CRISPR/Cas9. *Plant biotechnology journal*, *16*(4), 902-910.

Tatham, A. S., & Shewry, P. R. (2008). Allergens to wheat and related cereals. *Clinical & Experimental Allergy*, *38*(11), 1712-1726.

Wang, W., Pan, Q., He, F., Akhunova, A., Chao, S., Trick, H., & Akhunov, E. (2018). Transgenerational CRISPR-Cas9 activity facilitates multiplex gene editing in allopolyploid wheat. *The CRISPR journal*, *1*(1), 65-74.

Weichel, M., Glaser, A. G., Ballmer-Weber, B. K., Schmid-Grendelmeier, P., & Crameri, R. (2006). Wheat and maize thioredoxins: a novel cross-reactive cereal allergen family related to baker's asthma. *Journal of allergy and clinical immunology*, *117*(3), 676-681.

Xie, K., Minkenberg, B., & Yang, Y. (2015). Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system. *Proceedings of the National Academy of Sciences*, *112*(11), 3570-3575.

Zhang, Y., Liang, Z., Zong, Y., Wang, Y., Liu, J., Chen, K., ... & Gao, C. (2016). Efficient and transgene-free genome editing in wheat through transient expression of CRISPR/Cas9 DNA or RNA. *Nature communications*, *7*, 12617.

CHAPTER 5:

INVESTIGATION OF *Gli-D1* and *Glu-D3* LOCI OF BREAD WHEAT AND ITS D GENOME DONOR *Aegilops tauschii.*

ABSTRACT

Bread wheat (T. aestivum ssp. aestivum) is a hexaploid crop arisen from hybridization between T. turgidum ssp. dicoccum and Ae. tauschii. The genetic relationships between the D genome of Ae. tauschii and T. aestivum are one of the fundamental aspects to better understand the evolution of modern bread wheat. The gluten proteins, in particular gliadins and glutenins encoded by the genes located at *Glu-D3* and *Gli-D1* on chromosome 1D, are widely studied because of their essential role in influencing bread-making proprieties of wheat dough but also allow to investigate the phylogenetic relationship existing between *Ae. tauschii* subspecies and *T.aestivum* ssp. *aestivum*.
Here, two translocation lines Sv 1BL.1RS 1AL.1DScs and Sv 1BL.1RS 1AL.1DScnn lack the genes located at Gli-1, Glu-3 and minor loci on the short arm of 1A and 1B chromosomes and contain 1DS arm from cultivar Chinese Spring or cultivar Cheyenne were used to investigate the genetic relationships between genes encoding ω-, γ- gliadins and LMW-GS from *Aegilops tauschii* spp. strangulata (G1276) and from CS-type and CNN-type lines.

5.1 INTRODUCTION

The prolamins are the largest class of wheat proteins, accounting about 80% of endosperm protein content (Altenbach et al 2011). They consists of glutenins and gliadins, which confer viscoelastic proprieties to gluten enabling flour to be processed into a variety of foods (Payne et al. 1984, Shwery et al 1994).

Glutenin proteins are divided into high-molecular-weight glutenin subunits (HMW-GS) and low-molecular-weight glutenin subunits (LMW-GS), based on their molecular weight in SDS-PAGE (Payne 1987). They are widely studied because of their essential role in influencing bread-making proprieties of wheat dough but also represent a route to investigate the origin of wheat.

Multigene families, organised in complex loci on the chromosomes of groups 1 and 6, encode for glutenin and gliadin subunits. Despite two genes per haploid genome encode for the x- and y-type HMW-GS at *Glu-1* loci on the long arms of homeologous group 1chromosomes, the number of genes encoding gliadins and LMW-GS is much higher: 100-150 genes per haploid genome coding for α-gliadins are present, and 30-40 genes per haploid genome encode for LMW-GS, γ- and ω-gliadins (Jackson et al. 1983; Cassidy et al. 1998, Sabelli and Shewry 1991).

5.1.2 Proteins encoded at the *Glu-3* loci

LMW-GS represent about 60% of the total glutenin fraction and they contribute to viscoelastic proprieties of dough (D'Ovidio and Masci 2004). Based on electrophoretic mobility in SDS-PAGE, three groups of LMW-GS were identified, referred to as the B, C and D groups (Payne and Corfield 1979).

The D-group contains modified ω-gliadins that have acquired a cysteine residue (Masci et al., 1993, 1999) and their molecular weight (MW) is greater than other groups in SDS-PAGE. The C group consists mostly of modified α- and γ-gliadins with an odd number of cysteine residues (D'Ovidio et al. 1995; Scheets and Hedgcoth 1988; Masci et al. 2002; Camerlengo et al. 2017, Anderson et al. 1997). The B group includes mostly the typical LMW-GS. The structure of a typical LMW-GS consists of a short N terminal domain (N-term) that may contain the first cysteine residue, a repetitive domain (R-domain) rich in glutamine codons that may contain the first cysteine residue, and a C-terminal (C-term) domain further subdivided into three subdomains: (i) C-term I containing five cysteine residues, (i) C-term II containing a cysteine residue and glutamine repeated motifs,

and (i) C-term III containing a cysteine residue and a conserved sequence (D'Ovidio and Masci 2004).

Functionally, based on the number of cysteine residues and on the structure of the B, C and D groups, LMW-GS may be distinguished in two functional groups (Kasarda 1989). The B-type LMW-GS (i) act as chain extenders of gluten polymer due to the first and the seventh cysteine residues involved in inter-molecular disulphide bonds, the C and D-type LMW-GS (ii) act as chain terminators probably because they have only one cysteine available to form an intermolecular disulphide bond (D'Ovidio and Masci 2004).

The position of cysteine residues along LMW-GS sequence is mostly conserved, and three types of distribution can be recognized: (i) the first cysteine residue is present in the N-terminal domain and seventh in the C-terminal domain, (ii) the first cysteine residue present at the beginning of the R-domain and the seventh in the C-terminal domain and (iii) all eight cysteine are present in the C-terminal domain (D'Ovidio and Masci 2004).

On the basis of the first N-terminal amino acid residue, which could corresponds to isoleucine, methionine or serine; they are classified in LMW-i, LMW-m and LMW-s types, respectively. Whereas the LMW-m type subunits represent the most abundant at gene sequence level, the LMW-s type subunits are the predominant proteins determined by N-terminal protein sequencing (Johal et al. 2004). Ikeda et al. (2002) classified LMW-GS in six types (based on the first and seventh cysteine residue position) and twelve groups (based on N-term and C-term sequences). As reported in table 5.1, the LMW-m consists of six types of N-terminal sequence, the LMW-s have an asparagine residue (N) in position 3 of the N-term and the LMW-i types start directly with the repetitive domain after the signal peptide probably due to the lack of the N-terminal region in their gene sequences.

Table 5.1: N-terminal amino acid sequence and chromosome location of three LMW-types.

LMW type	N-Terminal sequence	Chromosome
LMW-m	METSRV	1A, 1D
	METRCIP	1A, 1D
	METSHIP	1A, 1D
	METSCIP	1A, 1D
	METSCIS	1A, 1D
LMW-s	MENSHIP	1B, 1D
LMW-i	ISQQQ	1A
	IENSHIP	1A

Since the N-terminal of LMW-s type is identical to N-terminal of some LMW-m type, Egidi et al. (2014) demonstrated that a maturation process likely conducted by an asparaginyl endopeptidase give rise to the LMW-s type due to the presence of the N residue instead of threonine (MEN – MET). Thus, considering the nucleotide sequences of typical LMW-GS, LMW-s and LMW-m types can be distinguished on the basis of a SNP.

LMW-GS are encoded by the genes located at complex *Glu-3* loci on the short arm of homoeologous chromosomes 1A, 1B and 1D (Jackson et al. 1983; Payne et al. 1985; Singh and Shepherd 1988; Gupta and Shepherd 1990). The *Glu-3* loci contain a variable number of genes, ranging from 10-15 (Harberd et al. 1985) to 35-40 (Cassidy et al. 1998; Sabelli and Shewry 1991) in bread wheat (Dong et al., 2010).

As reported in the Catalogue of Gene Symbols for Wheat (McIntosh et al. 2013), there are 7, 15 and 14 different allelic variants for *Glu-A3*, *Glu-B3* and *Glu-D3* loci in bread wheat, respectively (Ibba et al 2017c). Generally, more than 15 LMW-GS genes are present in individual wheat varieties; the *Glu-A3* locus consists of 4–6 genes, the *Glu*-B3 locus contains 3–5 genes and eight genes are located at the *Glu-D3* locus (Zhang et al. 2013; Ibba et al. 2017a).

The LMW-s type seems to be encoded by genes located at *Glu-B3* and *Glu-D3* loci (with MENSHIP N-terminal sequence), whereas LMW-i type are encoded by genes at *Glu-A3* locus (with ISQQQQ N-terminal sequence). The most of LMW-m type subunits (METSHIP, METSRV, METSCIS, METSCIP AND METRCIP) are encoded at the *Glu-D3* locus (Ikeda et al 2006).

The LMW-GS allele composition of *Glu-3* loci is highly variable among wheat cultivars and many laboratories developed different methods to identify *Glu-3* alleles and to assign each one to a specific locus (*Glu-A3/B3/D3*) (Liu et al 2010; Zhang et al 2011; Ibba et al 2018).

The *Glu-D3* seems to be the most complex locus showing a lower polymorphism compared to *Glu-A3* and *Glu-B3* loci, resulting in higher number of genes. Ibba et al. (2017c) analysed 30 different varieties and most of them exhibited eight amplicons corresponding to six expressed and two unexpressed genes at *Glu-D3* loci. In each variety, seven m-type and one s-type genes were identified. According to previous studies, the genes of *Glu-3* loci should be considered as individual genes rather than a unique genic block. In fact, recombination at *Glu-3* loci was observed and different haplotypes have been characterized at each locus (Dong et al. 2010; Zhang et al. 2013).

5.1.3 Proteins encoded at the *Gli-1* loci

Gliadins are the most abundant proteins in wheat endosperm and are divided into α/β-gliadins, γ-gliadins and ω-gliadins based on their electrophoretic mobility in A-PAGE. Since their similar structure, especially in proline- and glutamine-rich repeat motifs, they probably arose from a

common ancestor gene along with LMW-GS (Payne 1987; Gao et al. 2007). LMW-GS, α/β-gliadins and γ-gliadins are classified as S-rich protein due to their cysteine content and show a very similar structure consisting of N-terminal domain, a repeated domain variable in length and a C-term domain. They differ mainly in length and patterns of repeat motifs and in the number and position of cysteine residues.

Differently, ω-gliadins do not contain cysteine and methionine residue, thus they are known as S-poor proteins and they are not able to form inter- or intramolecular bonds. Their structure consists of a long repeat motif following the signal peptide and a short C-term domain. Some modified ω-gliadins with a free cysteine available for inter- or intramolecular bond were found and labelled as D-type LMW-GS; similarly, α/β-gliadins and γ-gliadins with an odd number of cysteine residues were labelled as C-type LMW-GS (Masci et al. 1993; Masci et al. 1999, D'Ovidio and Masci 2004). The γ- and ω- gliadins are encoded by multigene families located on the short arm of the homeologous group 1 chromosomes at the *Gli-A1*, *Gli-B1* and *Gli-D1* loci (Payne 1987; Shewry et al. 2003). Several minor dispersed loci encoding ω-gliadins have been identified and mapped on short arms of the homeologous chromosomes of group 1 arising from unequal crossing-over, translocation or gene duplication. Major and minor gliadins loci are tightly linked or very close each other and to *Glu-3* loci.

Based on the consensus sequence of the N-terminal domain, ω-gliadins may be classified as ARE, ARQ, KEL and SRL types (Kasarda et al 1983). *Gli-1* loci on the 1A and 1D chromosomes encode for the ARE, ARQ and KEL type, whereas *Gli-B1* encodes for the SRL-type. The KEL type arises from the ARQ type likely due to the cleavage of an asparaginyl endopeptidase. Similarly to LMW-s post-translational processing. According to their electrophoretic mobility in A-page, proteins encoded by the genes of 1A and 1D chromosomes are known as ω-1 and ω-2 gliadins and the SRL-type associated at the *Gli-B1*locus is known as ω-5 (DuPont et al. 2004; Kasarda et al. 1983).

The γ-gliadin family is the less investigated, although it has been reported that its gene expression is ten-fold greater than other gliadin families (Piston et al 2006). The typical γ-gliadins contain eight cysteine residues involved in intramolecular bonds with a large mobility range in A-PAGE, due to their variable content in charged amino acids.

Based on A-PAGE patterns of gliadins extracted from several bread wheat cultivars, Masci et al (1991) found essentially two main types of protein components associated with the *Gli-D1* locus, referred to as Chinese Spring-type (CS-type) or Cheyenne-type (CNN-type) because of their similarity with patterns exhibited from these two cultivars (Fig. 5.1).

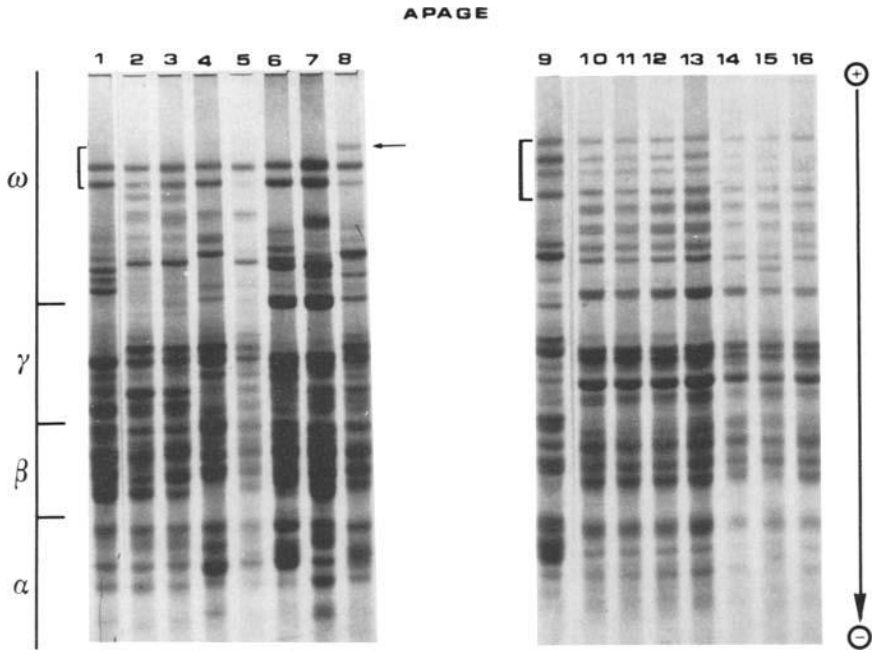

Figure 5.1:APAGE of some Chinese Spring-type (lanes 1-8) and Cheyenne-type (lanes 9-16) *genotypes. 1, Chinese Spring; 2, Marquis; 3, Hope; 4, Lutescens; 5, Timstein; 6, Wichita; 7, Ciano; 8, Capelle Desprez; 9, Cheyenne; 10, Thatcher; 11, Katepwa; 12, Neepawa; 13, Benito; 14, Solar; 15, Conway; 16, Canthatch. Brackets show the 1D-encoded omega gliadins, while the arrow in lane 8 indicates the 1D-controlled extra omega band possessed by Capelle Desprez.*

5.1.4 The contribution of the D-genome donor

The D genome donor of hexaploid common wheat is the wild annual diploid wheat (2n = 14, DD) *Aegilops tauschii*. The southern coast of the Caspian Sea and Azerbaijan are the center of distribution of *Ae. tauschii,* which spreads in central Eurasia towards China and westward to central Syria (Van Slageren 1994; Matsuoka et al. 2008). Hammer (1980) classified *Ae. tauschii* into two morphologically subspecies, spp. *tauschii* and ssp. *strangulata*. The subspecies *tauschii* is found in central Eurasia from Iran to Turkey and from Afghanistan to western China; it is also predominant in all areas of Iran except for the center of distribution of ssp *strangulata* in Golestan and Mazandaran provinces.

The subspecies *tauschii* exhibits elongated, cylindrical spikelets, whereas the subspecies *strangulata* exhibits more quadrate spikelets (Fig.5.3). *Ae. tauschii* is a self-pollinated species, but continuous cross-pollination events occurred between the two subspecies giving rise to intermediate forms (Kihara et al 1965; Dudnikov 1998).

Many studies conducted on several accession of *Ae. tauschii* revealed some discrepancies between morphological categories and genetic relationships (Dvorak et al.1998; Mizuno et al. 2010; Sohail et al. 2010), probably due to the existence of intermediate forms.

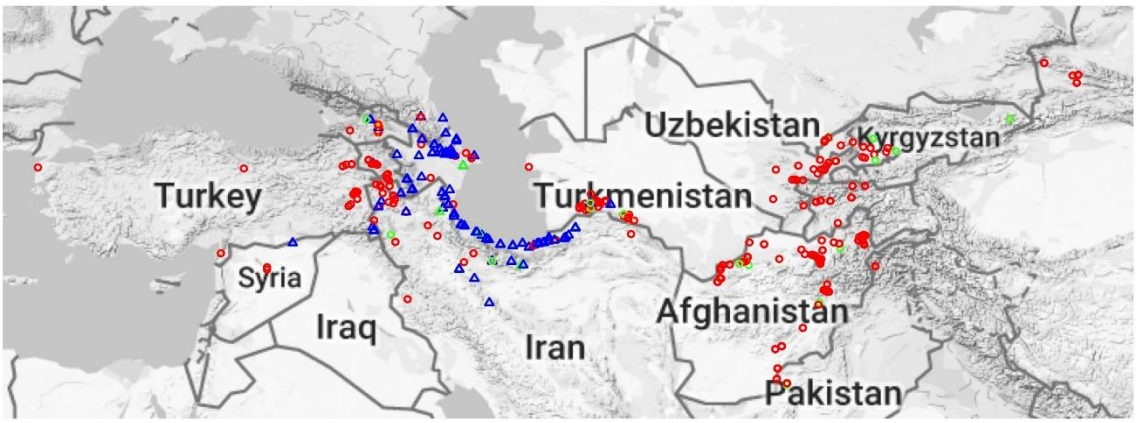

Figure 5.2: .Geographical distribution of Aegilops tauschii accessions. *Red circles represent Lineage 1 (L1), blue triangles Lineage 2 (L2), and gold plus sign (+) are putative hybrids. Green circles and triangles represent MiniCore accessions, and their shapes represent their lineage.*

However, significantly genetic diversity was found between ssp. *strangulata* and ssp. *tauschii* populations and between ssp. *strangulata* populations and intermediate forms (Aghaei et al. 2008; Shen et al. 2018). It is almost certain that limited hybridization events between a small number of *Ae. tauschii* ssp. *strangulata* accessions and the tetraploid wheat *Triticum turgidum* ssp. *dicoccum* gave rise to common hexaploid wheat (Masci et al. 1991; Cox 1997; Caldwell et al. .2004).

Figure 5.3: Spike morphology of D genome Aegilops species. *Ae. tauschii ssp. strangulata (a), Ae. tauschii ssp. tauschii (b), Ae. crassa (c), Ae. cylindrica (d), Ae. juvenalis (e) and Ae. vavilovii (f) and Ae. ventricosa (g)*

The loss of genetic variation within the D genome of bread wheat is evident compared with that estimated for *Ae*. *tauschii* (Arora et al 2017; Aghaei et al. 2008; Wang et al 2012; Shen et al 2018), especially in *Glu-D3*, *Gli-D1* and associated minor loci. Therefore, *Ae. tauschii* genepool may represent an interestingly source of genetic diversity to improve several traits, such as biotic and abiotic stress resistance, yield, quality and adaptability in bread wheat (Miranda et al 2006; Watanabe et al 2006). To this aim, the International Center for Maize and Wheat Improvement (CIMMYT) over the last 25 years released more than 1200 synthetic hexaploid wheat lines, obtained crossing elite tetraploid durum wheat with different *Ae. tauschii* accession (Trethowan and Mujeeb-Kazi 2008).

On these premises, the investigation of genes encoding gliadins and glutenins located on 1D chromosome of modern bread wheats and of the wheat D genome donor will give a broader knowledge necessary to improve essential traits in wheat.

5.2 MATERIALS AND METHODS

5.2.1 Plant Materials

Two translocation lines derived from durum wheat cultivar Svevo and one accession (G1276) of *Aegilops tauschii* spp. *strangulata* have been used in this work to investigate the ω-, γ-gliadins and LMW-GS encoded by the D genome. The two lines *Sv* 1BL.1RS 1AL.1DS$_{cs}$ and *Sv* 1BL.1RS 1AL.1DS$_{cnn}$ carries 1RS arm from rye replacing 1BS and 1DS arm from cultivar Chinese Spring or cultivar Cheyenne, respectively, in place of 1AS thanks to a second translocation. Eventually, these two lines lack the genes located at *Gli-1*, *Glu-3* and minor loci on the short arm of 1A and 1B chromosomes and only contain genes located at *Gli-D1* and *Glu-D3* loci.

These genotypes were analysed by polyacrylamide gel electrophoresis along with deletion lines derived from cv Pegaso lacking short arms of one or two group 1 chromosomes (Pegaso 1D-, Pegaso 1B-, Pegaso 1A-, Pegaso 1A-/1D-, Pegaso 1A-/1B-, Pegaso 1B-/1D, Pegaso 1B-/1D-) and four cultivars used as control (Pegaso, Svevo, Chinese Spring, Cheyenne)

5.2.2 Electrophoretic Separation Of Proteins

5.2.2.3 Protein extraction

The extraction of gliadin and glutenin fractions was carried out from the same kernel according to Pflüger et al., 2001. After grinding the seeds, the gliadins were extracted in ratio 1:10 flour weight to solvent volume with 1.5 M dimethyl formamide (DMF) from 20 mg of flour. The samples were centrifugated at 14,000 g for 10 minutes and the supernatant containing gliadins was transferred to a new tube. The pellet was washed twice with 500 µl of 50% 1-propan-1-ol and after centrifugation the supernatant was discarded. The pellet was then re-suspended in 200 µl of 50 % 1-propanol containing 1% of dithiothreitol (DTT) for glutenins extraction. After incubation for 45 min at 60°C and centrifugation at 14,000 g for 10 min, about 200 µl of supernatant containing the glutenin fraction was recovered. The alkylation of glutenin subunits was obtained by adding 2.8 µl of 4-vinylpiridine followed by incubation for 15 min at 60°C. Glutenins were precipitated with 1 ml of cold acetone overnight and then centrifuged. The dried pellet was re-suspended in *Laemmli Buffer* (Tris 0.125 M pH 6.8, SDS 2%, 20% glycerol) with 1% DTT and Pyronin y.

5.2.2.4 Acid Polyacrylamide Gel Electrophoresis (A-PAGE)

The gliadin subunits were separated according to Lookhart et al. (1982) by *Acid Polyacrylamide Gel Electrophoresis* (A-PAGE) in aluminium lactate buffer, pH 3.1. Polyacrylamide gels had a T value of 8% (total acrylamide concentration) and a C value (cross-linker concentration) of 1.25% and contained 40 mM aluminium lactate, 260 mM lactic acid, 0.002 g of ascorbic acid. After cooling to 1-4 C, 3% of Hydrogen peroxide and 0.0032 mg of ferrous sulphate eptahydrate (FeSO4 7H2O) were quickly added to the solution for polymerization. The solution was poured in 14 cm X 16 cm X 1.5 mm. Methyl violet was added to samples as tracking dye and 10 µl of sample were loaded on the gel. The run was performed at 20°C with inverted polarity at constant current (25mA per gel), the upper and lower chambers of electrophoresis apparatus were filled with buffer contains 40 mM aluminium lactate and 260 mM lactic acid. The run was stopped 30 mins after the dye reach the bottom of the gel (about 3.5 hours) and the gels were stained overnight in Commassie R-250 with 5% ethanol and 12% trichloacetic acid.

5.2.2.5 Sodium Dodecyl Sulphate–Polyacrylamide Gel electrophoresis (SDS-PAGE)

The glutenin subunits were separate in *Sodium Dodecyl Sulphate–Polyacrylamide Gel electrophoresis* (SDS-PAGE). The main gel used for glutenins subunits resolution had a T value of 12% and C value of 2% and contained 0.38 M Tris-HCl, pH 8.8 and 0.1% SDS. The solution was poured in 14 cm X 16 cm X 1 mm and the gels were polymerised using 7.5 mM ammonium persulfate and 0.6% (v/v) TEMED. The stacking gel (T=3.75 and C= 2.67), needed to pack the proteins in the comb before starting run, was poured after the main gel polymerization. The run was performed at constant current (35 mA per gel in a running buffer containing 0.2 M glycine, 25 mM Tris at pH 8.8 and 0.1% SDS) at 15°C. The run was stopped 1.5 hours after the dye reached the bottom of the gels. The gels were stained overnight in Coomassie G-250 with 10% H_3PO_4 and 20% methanol.

5.2.3 DNA Extraction

Genomic DNA was extracted from leaves of *Ae. tauschii* spp. *strangulata*, *Sv* 1BL.1RS 1AL.1DS$_{cs}$ and *Sv* 1BL.1RS 1AL.1DS$_{cnn}$ lines according to de Kochko et al. (1990). The samples were grinded and well-homogenized in presence of liquid nitrogen to obtain a fine powder. 200 µl of TE buffer (100 mM Tris, 50 mM EDTA and 500 mM NaCl) were added to the samples along with 40 µl of 20% SDS. After mixed thoroughly, the samples were incubated at 65°C for 10 mins enabling cellular lysis. 66.75 µl of 5M Sodium acetate were added to precipitate cellular debris and SDS-protein complex. After centrifugation at 14,000 g for 10 mins the supernatant was transferred to a

new tube for DNA precipitation in 2-Propanol. The pellet obtained through centrifugation at 14,000 g for 10 mins was washed twice with 70% ethanol. The dried pellet was re–suspended in 30 µl of 10 mM Tris, 1 mM EDTA, pH 8 solution and used for *Polymerase Chain Reaction* (PCR).

5.2.4 A-Tailing Cloning Strategy For *LMW-GS*, *γ-* And *α/β* Gliadin Genes

5.2.4.1 PCR to amplify LMW-GS, γ-gliadin and α/β gliadin genes

In order to amplify genes encoding LMW-GS, γ-gliadins and α/β gliadins from *Ae. tauschii* spp. *strangulata*, *Sv* 1BL.1RS 1AL.1DS$_{cs}$ and *Sv* 1BL.1RS 1AL.1DS$_{cnn}$ lines, PCR were performed with different primers pairs reported in table 5.2 The reactions were conducted in a final volume of 40 µl containing 20 µl of GoTaq Green Master Mix 2x (Promega), 0.5 µM of each primer and 1 µl of genomic DNA. The PCR conditions were: an initial denaturation at 95°C for 5 mins, followed by 35 cycles at 95°C for 1 min, 61°C for 30 s, 72°C for 1 min and a final extension at 72°C for 5 mins. Touch-down PCR method was used to amplify γ-gliadins, on the following conditions: 95°C for 5 mins, followed by 14 cycles at 95°C for 1 min, 68°C for 30 s with an increment decreasing of 0.5°C per cycle, 72°C for 1 min and a final extension at 72°C for 5 min.

PCR amplicons were checked on 1% agarose gel electrophoresis and purified using NucleoSpin Gel and PCR Clean-up kit (Machery-Nagel, Duren, Germany), recovering PCR amplicons from the gel. Finally, purified DNA was eluted in 30 µl of Buffer NE containing 5 mM Tris/HCl, pH 8.5.

Table 5.2: Primers used for gliadin genes amplification.

	PRIMER ID	5'→3' SEQUENCE	DIRECTION
γ-gliadins	G-gli_F	ATGAAGACCTTTACTCATC	Fv
	G-gli_R	TCATTGGCCACCAATGCC	Rv
	G-11_R	GGACAWAGACRTTGCACATG	Rv
LMW-GS	GLU31_P4	ATGGAGACTAGCCACATCCCT	Fv
	GLU3A3_P10F	ATGGAGACTAGCTGCATCCC	Fv
	GLU_18	CATCACAAGCACAAGCATCAA	Fv
	HB1133_R	TCAGTAGGCACCAACTCCGGTG	Rv
	LMWGS1,2F	ATGAAGACCTTCCTCRTCTTTG	Fv
	Glu8	CCACCATGAAGACCTTCCTC	Fv
	Glu1 R	GACACTTTATTTGTCACCGCT	Rv
	Glu-D3_r	TCCTTATCAGTAGRCACCAAC	Rv

5.2.4.2 Transformation of *E.coli* competent cells

Transformation of competent cells was carried out according to Sambrook et al (1989). Firstly, the purified PCR products were ligated into pGEM-T Easy vectors with T4 DNA ligase. The pGEM-T Easy vectors are linearized vectors with a single 3´-terminal thymidine at both ends (T-overhangs) that improves the efficiency of ligation of PCR products generated by A-tailing polymerases (GoTaq master mix, Promega). The molar ratio of PCR product: vector was optimized to 3:1, calculating the appropriate amount of insert needed for ligation reaction on the following equation:

$$\frac{ng\ of\ vector\ X\ Kb\ size\ of\ insert}{Kb\ size\ of\ vector}\ X\ insert{:}\ vector\ molar\ ratio = ng\ of\ insert$$

The ligation reaction was performed with reagents supplied with pGEM®-T and pGEM®-T Easy Vector Systems (Promega) and prepared as reported in table.

Table 5.3:. Ligation reaction components.

Reaction component	Amount
2x Rapid Ligation Buffer, T4 DNA Ligase	5 µl
pGEM - T Easy Vector (50ng)	1 µl
PCR product	Xµl
T4 DNA Ligase (3 Weiss units/µl)	
nuclease-free water to a final volume of	10 µl

The ligation reactions (2 µl) were used to transform 50 µl of JM109 High Efficiency Competent Cells (Promega) following producer recommendation. Transformed cells were plated onto LB/ampicillin/IPTG/X-gal plates for blue/white screening and incubate over-night at 37°C.

5.2.4.3 Plasmid DNA extraction and sequencing

After over-night incubation at 37°C, white colonies were inoculate in LB plus ampicillin (100 µg/ml) and the plasmids were extracted the day after with the NucleoSpin® Plasmid kit (Macherey-Nagel, Duren,Germany). Digestion of recombinant plasmids were performed with EcoRI enzyme

(10 Weiss units/µl, Thermo Fisher Scientific) with an incubation of 2h at 37°C in a volume of 20 µl per reaction containing 2 µl of 10X Buffer EcoRI, 250 ng of plasmid DNA and 0.5 µl of EcoRI. Digested plasmids were checked on agarose gel (0.8%) and only the plasmids containing an insert were sequenced with the universal primers T7 and SP6 by Eurofins Genomics (Ebersberg, Germany).

5.2.5 Gateway Cloning for ω-Gliadin Genes

5.2.5.1 PCR to amplify ω-gliadin genes

Genes encoding ω-gliadins of *Ae. tauschii* spp. *strangulata*, *Sv* 1BL.1RS 1AL.1DS$_{cs}$ and *Sv* 1BL.1RS 1AL.1DS$_{cnn}$ lines were cloned with Gateway® Cloning Technology (ThermoFisher Scientific) in order to avoid unstable recombinant clones due to vector/host combination and CAA repetition along gene sequences.

Two primer pairs used by Chen et al. (2010) and Hassani et al. (2007) were modified by adding attachment site (*att*B1-*att*B2) sequences to 5'end of each primer in order to insert PCR fragments into Gateway™ pDONR™221 Vector through BP recombination reaction in anti-sense direction (Tab. 5.4). The PCR reaction were performed in a volume of 25 µl containing 12.5 µl of 2X Platinum™ SuperFi™ PCR Master Mix, 0.5 µM of each primer 1 µl of template DNA and 5 µl of 5X SuperFi™ GC Enhancer, on following conditions: 95°C for 2 min, followed by 35 cycles at 95°C for 1 min, 61°C for 30 s, 72°C for 1 min and a final extension at 72°C for 5 min.

Table 5.4: Primers used for Gateway Cloning.

PRIMER ID	5'-3' SEQUENCE	REF
attB1	GGGGACAAGTTTGTACAAAAAAGCAGGCTNNN	
attB2	GGGGACCACTTTGTACAAGAAAGCTGGGTNNN	
attB2-omegaF	GGGGACCACTTTGTACAAGAAAGCTGGGTATGAAGACCTTCCTCATCTTTG	Chen et al 2010.
attB2-G3F5	GGGGACCACTTTGTACAAGAAAGCTGGGTTCAAAGGCAAGCAAGAAGTAGTA	Hassani et al. 2007
attB1-omegaR	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATTGGCCACCGATGCTTGT	Chen et al 2010.
attB1-wF20bR	GGGGACAAGTTTGTACAAAAAAGCAGGCTTCATTGGCCACCGATGCTTGTAAGACTA	Hassani et al. 2007

5.2.5.2 Gateway® Recombination reactions and Transformation of competent cells

After PEG (polyethylene glycol) purification of *att*B-PCR products from agarose gel following producer recommendations, the Gateway® BP Clonase® II was used in recombination reaction to facilitate transfer of *att*B-PCR products into Gateway™ pDONR™221 Vector (*att*P-containg donor vector). The reactions were performed in 10 µl with 50 fmol of *att*B-PCR products, 1 µl (150 ng) of Gateway™ pDONR™221 Vector, 8 µl of TE buffer, pH 8.0 and 2 µl of Gateway® BP Clonase® II. After incubation at 25°C for 2 h, 1 µl of Proteinase K was added to each reaction and incubated for 1 min at 37°C.

One Shot™ TOP10 Chemically Competent *E. coli* cells were transformed with 1 µl of each BP recombination reactions following the manufacturer's instructions. Transformed cells were plated onto LB/Kanamycin plates (50µg/mL) and ccdB gene located between the two attP sites on Gateway™ pDONR™221 Vector allowed negative selection of clones. The plates were incubated over-night at 37°C.

5.2.5.3 Analyses of positive clones and sequencing

PCR colony was used to analyse colonies obtained from each transformation using 10 µl of GoTaq Green Master Mix 2x (Promega), 0.5 µM universal M13 Reverse primer and 0.5 µM of reverse primer (attB2 omega R or attB1 W F20 BR) in a volume of 20 µl (PCR mix). Colonies were picked and re-suspended in PCR mix and amplified on the following conditions: 95°C for 10 min, followed by 35 cycles at 95°C for 1 min, 50°C for 15 s, 72°C for 1 min and a final extension at 72°C for 5 min.

Positive colonies were inoculated in LB medium containing 50 µg/mL Kanamycin and, the day after, the plasmids were extracted with the NucleoSpin® Plasmid kit (Macherey-Nagel, Duren,Germany).

Plasmids were sequenced with the universal primers M13 Forward (-20) and M13 Reverse using Power Read Upgrade service of Eurofins Genomics (Ebersberg, Germany) for special cycling conditions or chemistry to achieve best sequencing results.

5.2.6 Retrieving Gene Sequences From Databases

Some sequences of genes encoding LMW-GS, ω-and γ-gliadins were also obtained from International Wheat Genome Sequencing Consortium (IWGSC) RefSeq v1.0 (cultivar Chinese

Spring) and from National Center for Biotechnology Information (NCBI) database. These sequence were used in pairwise identity analyses of genes and deducted proteins and in the construction of phylogenetic trees.

5.3 RESULTS

5.3.1 Analyses of Protein Electrophoretic Patterns

The gliadin fraction extracted from *Aegilops tauschii* spp. *strangulata* (G1276), *Sv* 1BL.1RS 1AL.1DS$_{cs}$, *Sv* 1BL.1RS 1AL.1DS$_{cnn}$ flours was analysed by A-PAGE comparing them with other deletion and translocation lines. The electrophoretic pattern of ω-gliadins from *Aegilops tauschii* spp. *strangulata* (G1276) is quite similar to the ω-gliadins pattern of *Sv* 1BL.1RS 1AL.1DS$_{cnn}$, which only contains gliadins encoded by *Gli-D1* locus of Cheyenne, *Gli-2* loci (on 6A and 6B) of Svevo and secalins encoded on the short arm of 1R chromosome of rye. The comparison among protein electrophoretic patterns of G1276 and the translocations lines along with standard cultivars Chinese Springs and Cheyenne allow to assign many proteins bands to *Gli-1* loci (Figure 5.4).

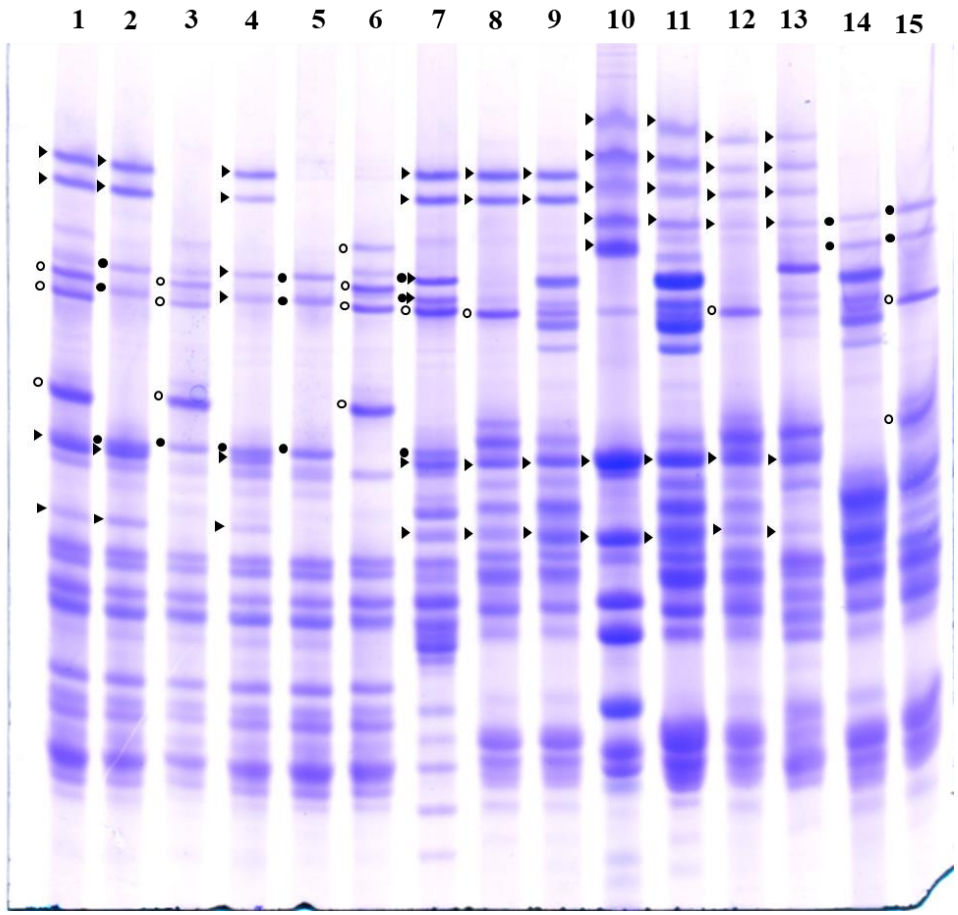

Figure 5.4: A-PAGE of some durum and bread wheat genotypes. *1 = Pegaso 1A-2 = Pegaso 1B-3 = Pegaso 1D-4 = Pegaso 1A-/1B-5 = Pegaso 1B-/1D-6 = Pegaso 1A-/1D-7 = CS8 = Sv 1AL.1DS$_{CS}$9 = Sv 1BL.1RS 1AL.1DS$_{CS}$10 = G127611 = Sv 1BL.1RS 1AL.1DS$_{CNN}$12 = Sv 1AL.1DS$_{CNN}$13 = CNN14 = Sv 1BL.1RS 15 = Svevo.* ▶ *= Gli-D1,* ○ *= Gli-B1,* ● *= Gli-A1.*

The LMW-GS protein profile of *Aegilops tauschii* spp. *strangulata* (G1276), *Sv* 1BL.1RS 1AL.1DScs, *Sv* 1BL.1RS 1AL.1DScnn lines were analysed by SDS-PAGE. Some protein bands have the same electrophoretic mobility in each line, with particular regard to those corresponding to proteins encoded at *Glu-D3* of *Ae. tauschii* spp. *strangulata*. Notably, the LMW-GS pattern of *Ae. tauschii* is more similar to that of *Sv* 1BL.1RS 1AL.1DS$_{CNN}$ rather than that of *Sv* 1BL.1RS 1AL.1DS$_{CS}$. In particular, a protein component with fast mobility and associated to *Glu-D3* locus of *Ae. tauschii* is present in the *Sv* 1BL.1RS 1AL.1DScnn lines only. As reported for A-PAGE, the electrophoretic patterns of the analysed wheat genotypes allow to assign several proteins bands to *Gli-1* loci.

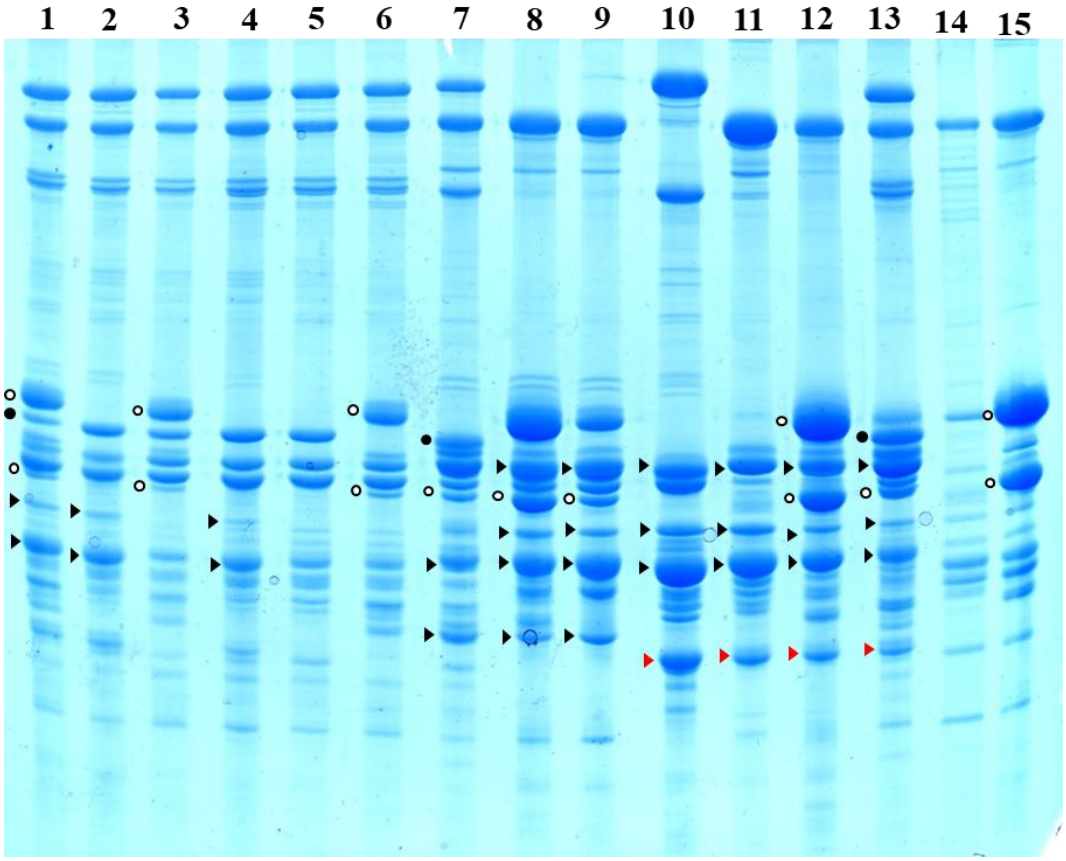

Figure 5.5: SDS-PAGE of some durum and bread wheat genotypes*.1 = Pegaso 1A-, 2 = Pegaso 1B-, 3 = Pegaso 1D-, 4 = Pegaso 1A-/1B-, 5 = Pegaso 1B-/1D-, 6 = Pegaso 1A-/1D-, 7 = CS, 8 = Sv 1AL.1DS$_{CS}$, 9 = Sv 1BL.1RS 1AL.1DSCS, 10 = G1276, 11 = Sv 1BL.1RS 1AL.1DS$_{CNN}$, 12 = Sv 1AL.1DS$_{CNN}$, 13 = CNN, 14 = Sv 1BL.1RS, 15 = Svevo.* ▶ *= Gli-D1,* ○ *= Gli-B1,* ● *= Gli-A1.*

5.3.2 Isolation of Genes Encoding ω-gliadins

The ω-gliadin genes are difficult to clone probably due to their large repetitive domain. Hsia and Anderson (2001) published for the first time a full-length ω-gliadin gene sequence (wF20b –AREL) isolated from wheat genomic lambda bacteriophage (λ) libraries and they also reported one

pseudogene (ωG3 - AREL). In later years, different authors tried to isolate and sequence ω-gliadins from wheat and wild relatives, but most results included pseudogenes, fast ω- gliadins, or gliadin genes modified by the host (Dupont et al 2004; Chen et al 2010; Du et al 2015).

Initially, PCR products obtained using primers without *att*B1-*att*B2 recombination sites were cloned into pGEM-T easy vector and transformed into JM109 competent cells. The resulting clones contained partial ω-gliadin sequences with various size, indicating the ω-gliadins genes were unstable in bacterial host or the sequencing system failed because of long repetition in the repetitive domain may form secondary structure. To overcome this problem, The Gateway® Cloning Technology (ThermoFisher Scientific) was used for directional cloning and sequencing of ω-gliadin PCR fragments.

Since the large size of amplicon fragments (>1200bp) and the potential secondary structures, clones resulting in expected product sizes from colony PCR were used for sequencing with special cycling conditions.

In total, 14 full-length ORFs were isolated from genomic DNA extracted from *Aegilops tauschii* spp. *strangulata* (G1276), *Sv* 1BL.1RS 1AL.1DScs, *Sv* 1BL.1RS 1AL.1DScnn lines (5, 2, 7, sequences respectively). The coding sequences ranged from 1038 to 1182 nucleotides (nt) and the corresponding deduced amino acid sequences ranged in size from 345 to 394 amino acids with a molecular weight (MW) mean of 44.438 KDa. Among 14 isolated ORFs, 11 resulted unique sequences; 4 unique genes were isolated from *Aegilops tauschii*, 1 from *Sv* 1BL.1RS 1AL.1DScs and 5 from *Sv* 1BL.1RS 1AL.1DScnn (Tab 5.5).

***Table 5.5: ω-gliadin genes isolated from** Aegilops tauschii strangulata (G1276) Sv 1BL.1RS 1AL.1DScs Sv 1BL.1RS 1AL.1DScnn lines.*

GENOTYPE	Sequence ID	Class	Type	N-TERM	MW (KDa)	Lenght	N-TERM domain		REP domain		C-TERM domain	
							Lenght	Cysteine num.	Lenght	Cysteine num.	Lenght	Cysteine num.
Ae.tauschii ssp. *Strangulata*	Ae.tauschii_AREL1	ω-glia	ARE	ARELNPS	38.019	326	0	0	314	0	12	0
	Ae.tauschii_AREL2	ω-glia	ARE	ARELNPS	42.744	366	0	0	354	0	12	0
	Ae.tauschii_AREL3	ω-glia	ARE	ARELNPS	43.739	375	0	0	362	0	12	0
	Ae.tauschii_ARQ	ω-glia	KEL	ARQLNPS	41.803	359	0	0	346	0	12	0
Sv 1BL.1RS 1AL.1DS$_{CS}$	SvCS_AREL2	ω-glia	ARE	ARELLNPS	42.744	367	0	0	354	0	12	0
Sv 1BL.1RS 1AL.1DS$_{CNN}$	SvCNN_AREL1	ω-glia	ARE	ARELLNPS	43.739	375	0	0	362	0	12	0
	SvCNN_AREL4	ω-glia	ARE	ARELLNPS	42.744	365	0	0	354	0	12	0
	SvCNN_AREL5	ω-glia	ARE	ARELLNPS	43.772	375	0	0	362	0	12	0
	SvCNN_ARQ1	ω-glia	KEL	ARQLNPS	41.805	358	0	0	346	0	12	0

Three genes isolated from *Sv* 1BL.1RS 1AL.1DScnn had two SNPs in position 399 and 465 resulting from a synonymous substitution (CCA > CCG, CAG >CAA) that led to the translation of the same protein (SvCNN_AREL1, SvCNN_AREL2, SvCNN_AREL3).

All the deduced amino acid sequences had the typical structure of ω-gliadins: a signal peptide followed by a long repetitive domain (R) and a short C-TERM domain. The signal peptide (19 aa) and the C-TERM domain (12 aa) were highly conserved among sequences whereas the R domain was mostly variable in length showing few amino acid changes/replacements. The R domain ranged from 314 to 361 amino acid residues with highly proline and glutamine content. Their length depends on the presence and number of typical repetitions, such as QQPQQPFP or QPQQPLP.

They lacked of cysteine residues and only 3 deduced proteins had one methionine residue in R domain (G1276_ARQ, SvCNN_ARQ1, SvCNN_AREL5).

According to their classification based on N-terminal sequence, 3 AREL type and 1 ARQ type were isolated from G1276, 3 AREL type and 1 ARQ type were found in *Sv* 1BL.1RS 1AL.1DScnn and only 1 AREL type was found in *Sv* 1BL.1RS 1AL.1DScs. None SRL type were detected according to 1B chromosomal location of their encoding genes. This observation confirms that the ARQ type ω-gliadins were processed to KEL type likely by the cleavage of an asparaginyl endopeptidase (Egidi et al 2014).

After removing redundant amino acid sequences, the alignment of deduced proteins, using Geneious Prime software, revealed a pairwise identity (ID) among sequences of 92.7%. In particular, no proteins sharing 100% of ID with the ARQ type ω-gliadin isolate from *Ae.tauschii* were found in the other genotypes (*Sv* 1BL.1RS 1AL.1DScs and *Sv* 1BL.1RS 1AL.1DScs), whereas the deduced protein *Ae.tauschii* _AREL2 showed 100% of ID with SvCNN_AREL4 and SvCS_AREL2, furthermore *Ae.tauschii* _AREL1 had 100% of ID with SvCNN_AREL1.

The alignments of deduced proteins performed in pairs showed that pairwise ID between sequences from G1276 and Sv 1BL.1RS 1AL.1DScnn was higher than pairwise ID between sequences from G1276 and Sv 1BL.1RS 1AL.1DScs (Fig.5.6). However, the highest pairwise ID was found between sequences from Sv 1BL.1RS 1AL.1DScnn and Sv 1BL.1RS 1AL.1DScs. The same trend was observed comparing the AREL type sequences (Fig. 5.7).

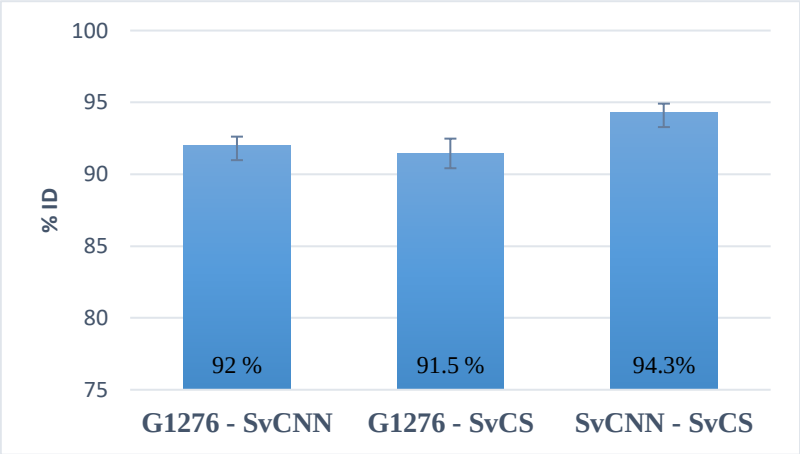

Figure 5.6: Pairwise identity of ω-gliadin amino acid sequences deduced from genes isolated from three genotypes. *SvCNN* = Sv 1BL.1RS 1AL.1DScnn; SvCS = Sv 1BL.1RS 1AL.1DScs; G1276 = *Aegilops tauschii* spp. *strangulata*

The ARQ type was found in *Ae.tauschii* and Sv 1BL.1RS 1AL.1DS$_{CNN,}$ with 99.4% of pairwise ID. They only had two different amino acid residues in position 88 Thr→ Ile and 185 Pro→Ser caused by two nonsynonymous substitutions at positions 263 and 553 of their nucleotide sequences (C→T) (Fig. 5.7).

```
AE.tauschii_AREL_1  ARELNPSNKELQSPQQSFSYQQQPFPQQPYPQQPYPSQQPYPSQQPFPTPQQQFPEQSQQ
AE.tauschii_AREL_2  ARELNPSNKELQSPQQSFSHQQQPFPQQPYPQQPYPSQQPYPSQQPFPTPQQQFPQQSQQ
AE.tauschii_AREL_3  ARELNPSNKELQSPQQSFSYQQQPFPQQPYPQQPYPSQQPYPSQQPFPTPQQQFPEQSQQ
AE.tauschii_ARQ     ARQLNPSNKELQSPQQSFSHQQQPFPQQPYPQQPYPSQQPYPSQQPFPTPQPQFPQQSQQ
Sv_CNN_AREL_1       ARELNPSNKELQSPQQSFSYQQQPFPQQPYPQQPYPSQQPYPSQQPFPTPQQQFPEQSQQ
Sv_CNN_AREL_4       ARELNPSNKELQSPQQSFSHQQQPFPQQPYPQQPYPSQQPYPSQQPFPTPQQQFPQQSQQ
Sv_CNN_AREL_5       ARELNPSNKELQSPQQSFSYQQQPFPQQPYPQQPYPSQQPYPSQQPFPTPQQQFPEQSQQ
Sv_CNN_ARQ_1        ARQLNPSNKELQSPQQSFSHQQQPFPQQPYPQQPYPSQQPYPSQQPFPTPQPQFPQQSQQ
Sv_CS_AREL_2        ARELNPSNKELQSPQQSFSHQQQPFPQQPYPQQPYPSQQPYPSQQPFPTPQQQFPQQSQQ

AE.tauschii_AREL_1  PFTQPQQPTPIQPQQPFPQQPQQPQQPFPQPQQPFPWQPQQPFPQTQQSFPLQPQQPFPQ
AE.tauschii_AREL_2  PFTQPQQPTPLQPQQPFPQQPQQPQQPFPQPQQPFPWQPQQPFPQTQQSFPLQPQQPFPQ
AE.tauschii_AREL_3  PFTQPQQPTPIQPQQPFPQQPQQPQQPFPQPQQPFPWQPQQPFPQTQQSFPLQPQQPFPQ
AE.Tauschii_ARQ     PFPQPQQPTPLQPQQPFPQQPQQPQQPFPQPQQPFPWQPQQPFPQTQQSFPLQPQQPFPQ
Sv_CNN_AREL_1       PFTQPQQPTPIQPQQPFPQQPQQPQQPFPQPQQPFPWQPQQPFPQTQQSFPLQPQQPFPQ
Sv_CNN_AREL_4       PFTQPQQPTPLQPQQPFPQQPQQPQQPFPQPQQPFPWQPQQPFPQTQQSFPLQPQQPFPQ
Sv_CNN_AREL_5       PFPQPQQPIPLQPQQPFPQQPQQPQQPFPQPQQPFPWQPQQPFPQTQQSFPLQPQQPFPQ
Sv_CNN_ARQ_1        PFPQPQQPIPLQPQQPFPQQPQQPQQPFPQPQQPFPWQPQQPFPQTQQSFPLQPQQPFPQ
Sv_CS_AREL_2        PFTQPQQPTPLQPQQPFPQQPQQPQQPFPQPQQPFPWQPQQPFPQTQQSFPLQPQQPFPQ

AE.tauschii_AREL_1  QPQQPFPQPQLPFPQQSEQIIPQQLQQPFPLQPQQPFPQQPQQPFPQPQQPIPVQPQQSF
AE.tauschii_AREL_2  QPQQPFPQPQLPFPQQSEQIIPQQPQQPFPLQPQQPFPQQPQQPFPQPQQPIPVQPQQSF
AE.tauschii_AREL_3  QPQQPFPQPQLPFPQQSEQIIPQQLQQPFPLQPQQPFPQQPQQPFPQPQQPIPVQPQQSF
AE.tauschii_ARQ     QPQQPFSQPQLQFPQQPEQIIPQQPQQPFLLESQQPFPQQPQQPFPQPQQLIPMQPQQPF
Sv_CNN_AREL_1       QPQQPFPQPQLPFPQQSEQIIPQQLQQPFPLQPQQPFPQQPQQPFPQPQQPIPVQPQQSF
Sv_CNN_AREL_4       QPQQPFPQPQLPFPQQSEQIIPQQPQQPFPLQPQQPFPQQPQQPFPQPQQPIPVQPQQSF
Sv_CNN_AREL_5       QPQQPFSQPQLQFPQQPEQIIPQQPQQPFLLESQQPFPQQPQQPFSQPQQLIPMQPQQPF
Sv_CNN_ARQ_1        QPQQPFSQPQLQFPQQPEQIIPQQPQQPFLLESQQPFPQQPQQPFSQPQQLIPMQPQQPF
Sv_CS_AREL_2        QPQQPFPQPQLPFPQQSEQIIPQQPQQPFPLQPQQPFPQQPQQPFPQPQQPIPVQPQQSF
```

```
AE.tauschii_AREL_1  PQQSQQSQQPFAQPQQLFPELQQPIPQQPQQPFPLQPQQPFPQQPQQPFPQQPQQSFP--
AE.tauschii_AREL_2  PQQSQQSQQPFAQPQQLFPELQQPIPQQPQQPFPLQPQQPFPQQPQQPFPQQPQQSFPQQ
AE.tauschii_AREL_3  PQQSQQSQQPFAQPQQLFPELQQPIPQQPQQPFPLQPQQPFPQQPQQPFPQQPQQSFPQQ
AE.tauschii_ARQ     PQQSQQSQQPFPGPQQLFPELQQPIPQQPQQPFPLQPQQPFPQQSQQPFPQQPQQPFPLQ
Sv_CNN_AREL_1       PQQSQQSQQPFAQPQQLFPELQQPIPQQPQQPFPLQPQQPFPQQPQQPFPQQPQQSFPQQ
Sv_CNN_AREL_4       PQQSQQSQQPFAQPQQLFPELQQPIPQQPQQPFPLQPQQPFPQQPQQPFPQQPQQSFPQQ
Sv_CNN_AREL_5       PQQSQQSQQPFPGPQQLFPELQQPIPQQPQQPFPLQPQQPFPQQPQQPFPQQPQQSFPQQ
Sv_CNN_ARQ_1        PQQSQQSQQPFPGPQQLFPELQQPIPQQPQQPFPLQPQQPFPQQSQQPFPQQPQQPFPLQ
Sv_CS_AREL_2        PQQSQQSQQPFAQPQQLFPELQQPIPQQPQQPFPLQPQQPFPQQPQQPFPQQPQQSFPQQ

AE.tauschii_AREL_1  ---------------------------------------------QQPQQSQQSFPQPQ
AE.tauschii_AREL_2  PQQPFPQQP--------QQPFPQQPQQPFPQQPQQPFPLRPQQPFPQQPQQSQQSFPQPQ
AE.tauschii_AREL_3  PQQPFPQQPQQPFPLQLQQPFPQQPQQPFPQQPQQPFPLRPQQPFPQQPQQSQQSFPQPQ
AE.tauschii_ARQ     PQQPFPQQP--------QQPFPQQPQQPFPLQPQQPFPLRPQQPFSQQPQQSQQSFPQPQ
Sv_CNN_AREL_1       PQQPFPQQPQQPFPLQLQQPFPQQPQQPFPQQPQQPFPLRPQQPFPQQPQQSQQSFPQPQ
Sv_CNN_AREL_4       PQQPFPQQP--------QQPFPQQPQQPFPQQPQQPFPLRPQQPFPQQPQQSQQSFPQPQ
Sv_CNN_AREL_5       PQQPFPQQPQQPFPLQLQQPFPQQPQQPFPQQPQQPFPLRPQQPFPQQPQQSQQSFPQPQ
Sv_CNN_ARQ_1        PQQPFPQQP--------QQPFPQQPQQPFPLQPQQPFPLRPQQPFSQQPQQSQQSFPQPQ
Sv_CS_AREL_2        PQQPFPQQP--------QQPFPQQPQQPFPQQPQQPFPLRPQQPFPQQPQQSQQSFPQPQ

AE.tauschii_AREL_1  PQQPQQPSILQPQQPLPQQPQQPFQQPQQQLSQQPEQTISQQPQQPFPQQPHQPQQPYPQ
AE.tauschii_AREL_2  PQQPQQPSILQPQQPLPQQPQQPFQQPQQQLSQQPEQTISQQPQQPFPQQPHQPQQPYPQ
AE.tauschii_AREL_3  PQQPQQPSILQPQQPLPQQPQQPFQQPQQQLSQQPEQTISQQPQQPFPQQPHQPQQPYPQ
AE.tauschii_ARQ     PQQPQQPSIL--------QPQQPFLQPQQQLSQQLEQTISQQPQQPFPQQPHQPQQPYPQ
Sv_CNN_AREL_1       PQQPQQPSILQPQQPLPQQPQQPFQQPQQQLSQQPEQTISQQPQQPFPQQPHQPQQPYPQ
Sv_CNN_AREL_4       PQQPQQPSILQPQQPLPQQPQQPFQQPQQQLSQQPEQTISQQPQQPFPQQPHQPQQPYPQ
Sv_CNN_AREL_5       PQQPQQPSILQPQQPLPQQPQQPFQQPQQQLSQQPEQTISQQPQQPFPQQPHQPQQPYPQ
Sv_CNN_ARQ_1        PQQPQQPSIL--------QPQQPFLQPQQQLSQQLEQTISQQPQQPFPQQPHQPQQPYPQ
Sv_CS_AREL_2        PQQPQQPSILQPQQPLPQQPQQPFQQPQQQLSQQPEQTISQQPQQPFPQQPHQPQQPYPQ

AE.tauschii_AREL_1   QQPYGSSLTSIGGQ
AE.tauschiiAREL_2    QQPYGSSLTSIGGQ
AE.tauschiiAREL_3    QQPYGSSLTSIGGQ
AE.tauschii_ARQ      QQPYGSSLTSIGGQ
Sv_CNN_AREL_1        QQPYGSSLTSIGGQ
Sv_CNN_AREL_4        QQPYGSSLTSIGGQ
Sv_CNN_AREL_5        QQPYGSSLTSIGGQ
Sv_CNN_ARQ_1         QQPYGSSLTSIGGQ
Sv_CS_AREL_2         QQPYGSSLTSIGGQ
```

Figure 5.7: Amino acid alignment of ω-gliadin deduced proteins of three genotypes.

5.3.3 Isolation of Genes Encoding γ-gliadins

The full length ORFs of **γ**-gliadin genes in individual genotype were amplified using two primer pairs (Tab. 5.2) and cloned into pGEM-T easy. After removing redundant and pseudo genes, 42 full-length ORFs were characterised; 15 unique genes from *Aegilops tauschii* spp. *strangulata* (G1276), 12 unique genes from *Sv* 1BL.1RS 1AL.1DScs, 12 unique genes from *Sv* 1BL.1RS 1AL.1DScnn lines (Tab. 5.6).

The deduced amino acid sequences had a mean MW of 31.560 KDa with a minimum and maximum length of 268 and 280 amino acid residues, respectively. The protein structure corresponded to those of **γ**-gliadins: a signal peptide (19 aa) followed by a short N-term domain of 12 amino acids, a repetitive domain (REP) ranging from 103 to 175 amino acid residues and a N-term domain containing 8 cysteine residues.

By the alignment of all deduced amino acid sequences, the pairwise ID reached 81.3 % with a higher ID between sequences from *Ae. tauschii* and *Sv* 1BL.1RS 1AL.1DScnn lines compared to sequences from *Sv* 1BL.1RS 1AL.1DScs (Fig. 5.8). Noteworthy, the sequence named SvCNN_GAMMA11 is identical to sequence named G1276_GAMMA7.

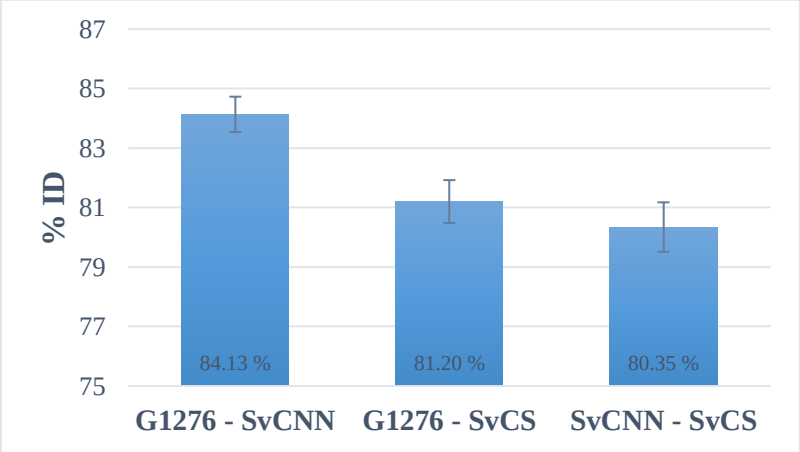

Figure 5.8: Pairwise identity of γ-gliadin amino acid sequences deduced from genes isolated from three genotypes. *SvCNN* = Sv 1BL.1RS 1AL.1DScnn; SvCS = Sv 1BL.1RS 1AL.1DScs; G1276 = *Aegilops tauschii* spp. *strangulata*

Whereas the REP domain is the most variable in length and amino acid composition, the N-term domain is the most conserved with 87.3 % of pairwise ID among all deduced proteins, followed by the C-term domain with 84.4 % of ID. The length of REP domains depends on the presence and number of peptide repetitions PLSQQ, PQQQLVPQ, QM/FHHQP, QPQQTFP. The mean length of REP domains in γ-gliadins extracted from *Ae. tauschii* and *Sv* 1BL.1RS 1AL.1DScnn was similar

(125.5 and 125.2 amino acids, respectively), on the other hand the mean length of REP domains from *Sv* 1BL.1RS 1AL.1DScs were the longest (138.9 amino acid residues).

The same pairwise ID trend of the alignment among all deduced proteins was observed comparing the pairwise ID obtained from the alignment of REP domains performed in genotype pairs (Fig. 5.9).

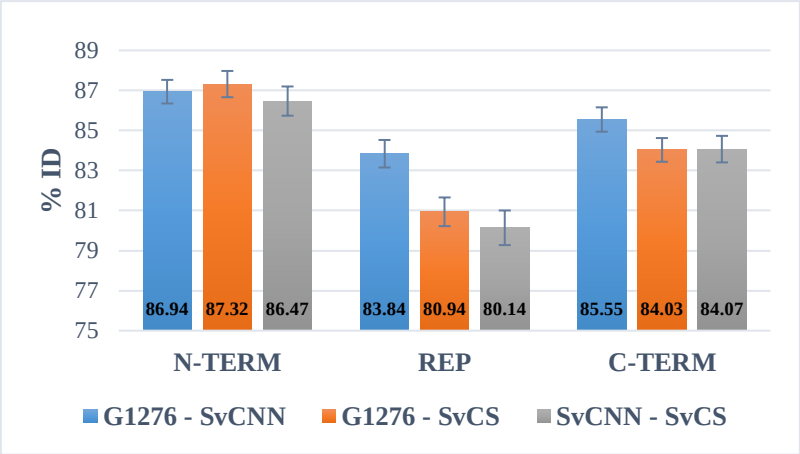

Figure 5.9: Pairwise identity of amino acid domains of γ-gliadin isolated from three genotypes. *SvCNN* = Sv 1BL.1RS 1AL.1DScnn; SvCS = Sv 1BL.1RS 1AL.1DScs; G1276 = *Aegilops tauschii* spp. *strangulata*

The typical 8 cysteine residues of γ-gliadins, located at the C-term domain, were conserved in all deduced amino acid sequences, but 9 sequences (tab. 5.6) showed the presence of an additional cysteine residue in the REP domain at position 26. The additional cysteine residue resulted from a nonsynonymous substitution (TCC→TGC) at position 137 of the nucleotide sequence. The pairwise ID among all sequences with 9 cysteine residues reached 94.7% (with a maximum pairwise ID of 97.3% between sequences from *Ae.tauschi* and Sv 1BL.1RS 1AL.1DScnn) and, notably, sequences obtained from Sv 1BL.1RS 1AL.1DScs contained an extra peptide repetitions QPQQTFP at the beginning of REP domain compare to other deduced amino acid sequences.

***Table 5.6: γ-gliadin genes isolated from** Aegilops tauschii strangulata (G1276)Sv 1BL.1RS 1AL.1DScs Sv 1BL.1RS 1AL.1DScnn lines**.*

GENOTYPE	Sequence ID	Class	N-TERM	MW (KDa)	Length	N-TERM domain		REP domain		C-TERM domain	
						Length	Cysteine num.	Length	Cysteine num.	Length	Cysteine num.
*Ae.tauschi*i ssp. *Strangulata*											
	Ae.tauschii_GAMMA1	γ-glia	NIQ	30.579	268	12	0	121	0	135	8
	Ae.tauschii_GAMMA1A	γ-glia	NMQ	31.66	276	12	0	121	0	143	8
	Ae.tauschii_GAMMA2	γ-glia	NMQ	31.767	276	12	0	129	1	135	8
	Ae.tauschii_GAMMA3	γ-glia	NMQ	31.739	276	12	0	129	1	135	8
	Ae.tauschii_GAMMA3A	γ-glia	NMQ	31.872	279	12	0	121	0	146	8
	Ae.tauschii_GAMMA4	γ-glia	NMQ	32.059	280	12	0	121	0	147	8
	Ae.tauschii_GAMMA4A	γ-glia	NMQ	31.844	279	12	0	121	0	146	8
	Ae.tauschii_GAMMA5	γ-glia	NMQ	31.828	276	12	0	129	1	135	8
	Ae.tauschii_GAMMA6	γ-glia	NIQ	31.722	276	12	0	129	1	135	8
	Ae.tauschii_GAMMA6A	γ-glia	NMQ	31.872	279	12	0	121	0	146	8
	Ae.tauschii_GAMMA7	γ-glia	NIQ	35.198	308	12	0	149	0	147	8
	Ae.tauschii_GAMMA8	γ-glia	NMQ	31.8	276	12	0	129	1	135	8
	Ae.tauschii_GAMMA9	γ-glia	NMQ	31.904	280	12	0	121	0	147	8
	Ae.tauschii_GAMMA10	γ-glia	NMQ	30.765	268	12	0	121	0	135	8
	Ae.tauschii_GAMMA11	γ-glia	NMQ	31.641	276	12	0	121	0	143	8
Sv 1BL.1RS 1AL.1DS$_{CS}$	SvCS_GAMMA1	γ-glia	NMQ	31.567	275	12	0	121	0	143	8
	SvCS_GAMMA1A	γ-glia	NMQ	31.865	279	12	0	121	0	146	8
	SvCS_GAMMA2	γ-glia	NMQ	32.33	283	12	0	136	1	135	8
	SvCS_GAMMA3	γ-glia	NMQ	31.589	276	12	0	121	0	143	8
	SvCS_GAMMA5	γ-glia	NMQ	31.619	276	12	0	121	0	143	8
	SvCS_GAMMA6A	γ-glia	NMQ	32.634	283	12	0	136	1	135	8
	SvCS_GAMMA7	γ-glia	NMQ	39.982	349	12	0	175	0	162	8
	SvCS_GAMMA8	γ-glia	NMQ	30.716	283	12	0	136	1	135	8
	SvCS_GAMMA9	γ-glia	NIQ	33.258	308	12	0	149	0	147	8
	SvCS_GAMMA10	γ-glia	NMQ	39.938	349	12	0	175	0	162	8
	SvCS_GAMMA11	γ-glia	NIQ	33.23	307	12	0	149	0	146	8
	SvCS_GAMMA12	γ-glia	NMQ	29.003	272	12	0	114	0	146	8
Sv 1BL.1RS 1AL.1DS$_{CNN}$	SvCNN_GAMMA1	γ-glia	NMQ	32.052	279	12	0	121	0	146	8
	SvCNN_GAMMA1A	γ-glia	NMQ	32.018	279	12	0	121	0	146	8
	SvCNN_GAMMA2	γ-glia	NMQ	30.732	268	12	0	121	0	135	8
	SvCNN_GAMMA3	γ-glia	NMQ	31.797	276	12	0	129	1	135	8
	SvCNN_GAMMA3A	γ-glia	NMQ	31.835	279	12	0	121	0	146	8
	SvCNN_GAMMA5	γ-glia	NIQ	35.17	308	12	0	149	0	147	8
	SvCNN_GAMMA5A	γ-glia	NMQ	31.946	279	12	0	121	0	146	8
	SvCNN_GAMMA6	γ-glia	NMQ	29.887	279	12	0	121	0	146	8
	SvCNN_GAMMA8	γ-glia	NMQ	29.912	279	12	0	121	0	146	8
	SvCNN_GAMMA9	γ-glia	NMQ	30.062	279	12	0	121	0	146	8
	SvCNN_GAMMA11	γ-glia	NIQ	33.24	308	12	0	149	0	147	8

```
 Ae.tuschii_GAMMA-1   NIQVDPSGQVQWLQQQLVPQLQQPLSQQPQQTFPQP---------------QQTFPHQPQ
Ae.tuschii_GAMMA-1A   NMQVDPSSQVQWPQQQPVPQPHQPFSQQPQQTFPQP---------------QQTFPHQPQ
 Ae.tuschii_GAMMA-2   NMQVDPGYQVQWPQQQPFPQPQQPFCQQPQQTIPQPHQMFHH---Q----PQQTYPHQPQ
 Ae.tuschii_GAMMA-3   NMQVDPGYQVQWPQQQPFPQPQQPFCQQPQQTIPQPHQMFHH---Q----PQQTYPHQPQ
Ae.tuschii_GAMMA-3A   NMQVDPSGQVQWPQQQPFPQPHQPFSQQPQQTFPQP---------------QQTFPHQPQ
 Ae.tuschii_GAMMA-4   NMQVDPSSQVQWPQQQPVPQPHQPFSQQPQQTFPQP---------------QQTFPHQPQ
Ae.tuschii_GAMMA-4A   NMQVDPSGQVQWPQQQPFPQPHQPFSQQPQQTFPQP---------------QQTFPHQPQ
 Ae.tuschii_GAMMA-5   NMQVDPGYQVQWPQQQPFPQPQQPFCQQPQQTIPQPHQMFHH---Q----PQQTYPHQPQ
 Ae.tuschii_GAMMA-6   NIQVDPSGQVQWLQQQPFPQPQQPFCQQPQQTIPQPHQMFHH---Q----PQQTYPHQPQ
Ae.tuschii_GAMMA-6A   NMQVDPSGQVQWPQQQPFPQPHQPFSQQPQQTFPQP---------------QQTFPHQPQ
 Ae.tuschii_GAMMA-7   NIQVDPSGQVQWLQQQLVPQLQQPLSQQPQQTFPQP---------------QQTFPHQPQ
 Ae.tuschii_GAMMA-8   NMQVDPGYQVQWPQQQPFPQPQQPFCQQPQQTIPQPHQMFHH---Q----PQQTYPHQPQ
 Ae.tuschii_GAMMA-9   NMQVDPSGQVQWPQQQPFPQPHQPFSQQPQQTFPQP---------------QQTFPHQPQ
Ae.tuschii_GAMMA-10   NMQVDPSSQVQWPQQQPVPQPHQPFSQQPQQTFPQP---------------QQTFPHQPQ
Ae.tuschii_GAMMA-11   NMQVDPSSQVQWPQQQPVPQPHQPFSQQPQQTFPQP---------------QQTFPHQPQ
     Sv_CNN_GAMMA-1   NMQVDPSSQVQWPQQQPVPQPHQPFSQQPQQTFPQP---------------QQTFPHQPQ
    Sv_CNN_GAMMA-1A   NMQVDPSSQVQWPQQQPVPQPHQPFSQQPQQTFPQP---------------QQTFPHQPQ
     Sv_CNN_GAMMA-2   NMQVDPSGQVPWPQQQPFPQPHQPFSQQPQQTFPQP---------------QQTYPHQPQ
     Sv_CNN_GAMMA-3   NMQVDPGYQVQWPQQQPFPQPQQPFCQQPQQTIPQPHQTFHH---Q----PQQTYPHQPQ
    Sv_CNN_GAMMA-3A   NMQVDPSGQVPWPQQQPFPQPHQPFSQQPQQTFPQP---------------QQTFPHQPQ
     Sv_CNN_GAMMA-5   NIQVDPSGQVQWLQQQLVPQLQQPLSQQPQQTFPQP---------------QQTFPHQPQ
    Sv_CNN_GAMMA-5A   NMQVDPSSQVQWPQQQPVPQPHQPFSQQPQQTFPQP---------------QQTFPHQPQ
     Sv_CNN_GAMMA-6   NMQVDPSGQVPWPQQQPFPQPHQPFSQQPQQTFPQP---------------QQTFPHQPQ
     Sv_CNN_GAMMA-8   NMQVDPSGQVPWPQQQPFPQPHQPFSQQPQQTFPQP---------------QQTFPHQPQ
     Sv_CNN_GAMMA-9   NMQVDPSSQVQWPQQQPVPQPHQPFSQQPQQTFPQP---------------QQTFPHQPQ
    Sv_CNN_GAMMA-11   NIQVDPSGQVQWLQQQLVPQLQQPLSQQPQQTFPQP---------------QQTFPHQPQ
      Sv_CS_GAMMA-1   NMQVDPSSQVQWPQQQPVPQPHQPFSQQPQQTFPQP---------------QQTFPHQPQ
     Sv_CS_GAMMA-1A   NMQVDPSGQVPWPQQQPFPQPHQPFSQQPQQTFPQP---------------QQTFPHQPQ
      Sv_CS_GAMMA-2   NMQVDPSGQVQWPQQQPFPQPQQPFCEQPQRTIPQPHQTFHHQPQQTFPQPEQTYPHQPQ
      Sv_CS_GAMMA-3   NMQVDPSSQVQWPQQQPVPQPHQPFSQQPQQTFPQP---------------QQTFPHQPQ
      Sv_CS_GAMMA-5   NMQVDPSSQVQWPQQQPVPQPHQPFSQQPQQTFPQP---------------QQTFPHQPQ
     Sv_CS_GAMMA-6A   NMQVDPGYQVHWPQQQPFPQPQQPFCQQPQQTIPQPHQTFHHQPQQTFPQPQQTYPHQPQ
      Sv_CS_GAMMA-7   NMQVDPSGQVQWPQQQLVPQPQQPLSQQPQQAFPQP---------------QQTFPHQPQ
      Sv_CS_GAMMA-8   NMQVDPGYQVHWPQQQPFPQPQQPFCQQPQQTIPQPHQTFHHQPQQTFPQPQQTYPHQPQ
      Sv_CS_GAMMA-9   NIQVDPSGQVQWLQQQLVPQLQQPLSQQPQQTFPQP---------------QQTFPHQPQ
     Sv_CS_GAMMA-10   NMQVDPSGQVQWPQQQLVPQPQQPLSQQPQQAFPQP---------------QQTFPHQPQ
     Sv_CS_GAMMA-11   NIQVDPSGQVQWLQQQLVPQLQQPLSQQPQQTFPQP---------------QQTFPHQPQ
     Sv_CS_GAMMA-12   NMQVDPSGQVQWPQQQPFLQPHQPFSQQPQQIFPQP---------------QQTFPHQPQ

 Ae.tuschii_GAMMA-1   QQVPQPQQPQQPFPQPQQTFPQQPQLPFP--Q-----------------------------
Ae.tuschii_GAMMA-1A   QQFPQPRQPQQQFLQPQQPFPQQPQQPYP--Q-----------------------------
 Ae.tuschii_GAMMA-2   QQFPQTQQPQQPFPQPQQTFPQQPQLPFP--Q-----------------------------
 Ae.tuschii_GAMMA-3   QQFPQTQQPQQPFPQPQQTFPQQPQLPFP--Q-----------------------------
Ae.tuschii_GAMMA-3A   QQFSQPQQPQQQFIQPQQPFPQQPQQPYP--Q-----------------------------
 Ae.tuschii_GAMMA-4   QQFPQPQQPQQQFLQPQQPFPQQPQQPYP--Q-----------------------------
Ae.tuschii_GAMMA-4A   QQFSQPQQPQQQFIQPQQPFPQQPQQPYP--Q-----------------------------
 Ae.tuschii_GAMMA-5   QQFPQTQQPQQPFPQPQQTFPQQPQLPFP--Q-----------------------------
 Ae.tuschii_GAMMA-6   QQFPQTQQPQQPFPQPQQTFPQQPQLPFP--Q-----------------------------
Ae.tuschii_GAMMA-6A   QQFSQPQQPQQQFIRPQQPFPQQPQQPYP--Q-----------------------------
 Ae.tuschii_GAMMA-7   QQVPQPQQPQQPFLQPQQPFPQQPQQPFPQTQQPQQPFPQQPQQPFPQTQ-----------
 Ae.tuschii_GAMMA-8   QQFPQTQQPQQPFPQPQQTFPQQPQLPFP--Q-----------------------------
 Ae.tuschii_GAMMA-9   QQFSQPQQPQQQFIQPQQPFPQQPQQPYP--Q-----------------------------
Ae.tuschii_GAMMA-10   QQFPQPQQPQQQFLQPQQPFPQQPQQPYP--Q-----------------------------
Ae.tuschii_GAMMA-11   QQFPQPQQPQQQFLQPQETFPQQPQLPFP--Q-----------------------------
     Sv_CNN_GAMMA-1   QQFPQPQQPQQQFLQPQQPFPQQPQQPYP--Q-----------------------------
    Sv_CNN_GAMMA-1A   QQFPQPQQPQQQFLQPQQPFPQQPQQPYP--Q-----------------------------
     Sv_CNN_GAMMA-2   QQFPQTQQPQQPFPQPQQTFPQQPQLPFP--Q-----------------------------
     Sv_CNN_GAMMA-3   QQFPQTQQPQQPFPQPQQTFPQQPQLPFP--Q-----------------------------
    Sv_CNN_GAMMA-3A   QQFSQPQQPQQQFIQPQQPFPQQPQQTYP--Q-----------------------------
     Sv_CNN_GAMMA-5   QQAPQPQQPQQPFLQPQQPFPQQPQQPFPQTQQPQQPFPQQPQQPFPQTQ-----------
    Sv_CNN_GAMMA-5A   QQFPQPQQPQQQFLQPQQPFPQQPQQPYP--Q-----------------------------
     Sv_CNN_GAMMA-6   QQFSQPQQPQQQFIQPQQPFPQQPQQTYP--Q-----------------------------
     Sv_CNN_GAMMA-8   QQFSQPQQPQQQFIQPQQPFPQQPQQTYP--Q-----------------------------
     Sv_CNN_GAMMA-9   QQFPQPQQPQQQFLQPQQPFPQQPQQPYP--Q-----------------------------
    Sv_CNN_GAMMA-11   QQVPQPQQPQQPFLQPQQPFPQQPQQPFPQTQQPQQPFPQQPQQPFPQTQ-----------
      Sv_CS_GAMMA-1   QQFPQPQQPQQQFLQPQQPFPQQPQQPYP--Q-----------------------------
     Sv_CS_GAMMA-1A   QQFSQPQQPQQQFIQPQQPFPQQPQQTYP--Q-----------------------------
      Sv_CS_GAMMA-2   QQFPQTQQPQQPFPQPQQTFPQQPQLPFP--Q-----------------------------
```

```
       Sv_CS_GAMMA-3   QQFPQPQQPQQQFLQPQQPFPQQPQQPYP--Q----------------------------
       Sv_CS_GAMMA-5   QQFPQPQQPQQQFLQPQQPFPQQPQQPYP--Q----------------------------
      Sv_CS_GAMMA-6A   QQFPQTQQPQQPFPQPQQTFPQQPQLPFP--Q----------------------------
       Sv_CS_GAMMA-7   QQVPQPQQPQQPFLQPQQAFPQQPQQPFPQTQQPQQPFPQQPQQPFPQTQQPQQPFPQQP
       Sv_CS_GAMMA-8   QQFPQTQQPQQPFPQPQQTFPQQPQLPFP--Q----------------------------
       Sv_CS_GAMMA-9   QQVPQPQQPQQPFLQPQQPFPQQPQQPFPQTQQPQQPFPQQPQQPFPQTQ----------
      Sv_CS_GAMMA-10   QQVPQPQQPQQPFLQPQQAFPQQPQQPFPQTQQPQQPFPQQPQQPFPQTQQPQQPFPQQP
      Sv_CS_GAMMA-11   QQVPQPQQPQQPFLQPQQPFPQQPQQPFPQTQQPQQPFPQQPQQPFPQTQ----------
      Sv_CS_GAMMA-12   QQFPQPQQPQQQFLQPRQPFPQQPQQPYP--Q----------------------------

 Ae.tuschii_GAMMA-1   ----------------Q--------PQQPFPQPQQPQQQFPQSQQPQQPFPQPQQQFLQP
Ae.tuschii_GAMMA-1A   ----------------E--------PQQPFPQTQQPQQLFPQSQQPQQQFPQPQQQFPQP
 Ae.tuschii_GAMMA-2   ----------------Q--------PQQPFPQPQQPQQQFPQSQEPQQPFPQPQQQFLQP
 Ae.tuschii_GAMMA-3   ----------------Q--------PQQPFPQPQQPQQQFPQSQQPQQPFPQPQQQFPQP
Ae.tuschii_GAMMA-3A   ----------------Q--------PQQPFPQTQQPQQPFPQSQQPQQPFPRPQQQFPQP
 Ae.tuschii_GAMMA-4   ----------------E--------PQQPFPQTQQPQQLFPQSQQPQQQFPQPQQQFPQP
Ae.tuschii_GAMMA-4A   ----------------Q--------PQQPFPQTQQPQQPFPQSQQPQQPFPQPQQQFPQP
 Ae.tuschii_GAMMA-5   ----------------Q--------PQQPFPQPQQPQQQFPQSQQPQQPFPRPQQQFLQP
 Ae.tuschii_GAMMA-6   ----------------Q--------PQQPFPQPQQPQQQFPQSQQPQQPFPQPQQQFLQP
Ae.tuschii_GAMMA-6A   ----------------Q--------PQQPFPQTQQPQQPFPQSQQPQQPFPQPQQQFPQP
 Ae.tuschii_GAMMA-7   ----------------QPQQPFPQQPQQPFPQTQQPQQPFPQLQQPQQPFPQPQQQLPQP
 Ae.tuschii_GAMMA-8   ----------------Q--------PQQPFPQPQQPQQQFPQSQQPQQPFPQPQQQFLQP
 Ae.tuschii_GAMMA-9   ----------------Q--------PQQPFPQTQQPQQPFPQSQQPQQPFPQPQQQFPQP
Ae.tuschii_GAMMA-10   ----------------E--------PQQPFPQTQQPQQLFPQSQQPQQQFPQPQQQFLQP
Ae.tuschii_GAMMA-11   ----------------Q--------PQQPFPQPQQPQQLFPQSQRPQQQFPQPQQQFPQP
      Sv_CNN_GAMMA-1   ----------------Q--------PQQPFPQTQQPQQLFPQSQQPQQQFSQPQQQFPQP
     Sv_CNN_GAMMA-1A   ----------------Q--------PQQPFPQTQQPQQLFPQSQQPQQQFSQPQQQFPQP
      Sv_CNN_GAMMA-2   ----------------Q--------PQQPFPQPQQPQQQFPQSQQPQQPFPQPQQQFLQP
      Sv_CNN_GAMMA-3   ----------------Q--------PQQPFPQPQQPQQQFPQSQQPQQPFPQPQQQFLQP
     Sv_CNN_GAMMA-3A   ----------------R--------PQQPFPQTQQPQQPFPQSQQPQQPFPQPQQQFPQP
      Sv_CNN_GAMMA-5   ----------------QPQQPFPQQPQQPFPQTQQPQQPFPQLQQPQQPFPQPQQQLPQP
     Sv_CNN_GAMMA-5A   ----------------Q--------PQQPFPQTQQPQQLFPQSQQPQQQFSQPQQQFPQP
      Sv_CNN_GAMMA-6   ----------------R--------PQQPFPQTQQPQQPFPQSQQPQQPFPQPQQQFPQP
      Sv_CNN_GAMMA-8   ----------------R--------PQQPFPQTQQPQQPFPQSQQPQQPFPQPQQQFPQP
      Sv_CNN_GAMMA-9   ----------------Q--------PQQPFPQTQQPQQLFPQSQQPQQQFSQPQQQFPQP
     Sv_CNN_GAMMA-11   ----------------QPQQPFPQQPQQPFPQTQQPQQPFPQLQQPQQPFPQPQQQLPQP
       Sv_CS_GAMMA-1   ----------------Q--------PQQPFPQTQQPQQLFPQSQQPQQQFSQPQQQFPQP
      Sv_CS_GAMMA-1A   ----------------R--------PQQPFPQTQQPQQPFPQSQQPQQPFPQPQQQFPQP
       Sv_CS_GAMMA-2   ----------------Q--------PQQPFPQPQQPQQPFPQSQQPQQPFPQPQQQFPQP
       Sv_CS_GAMMA-3   ----------------Q--------PQQPFPQTQQPQQLFPQSQQPQQQFSQPQQQFPQP
       Sv_CS_GAMMA-5   ----------------Q--------PQQPFPQTQQPQQLFPQSQQPQQQFSQPQQQFPQP
      Sv_CS_GAMMA-6A   ----------------Q--------PQQPFPQPQQPQQQFPQSQQPQQPFPQPQQQFLQP
       Sv_CS_GAMMA-7   QQPFPQQPQQPFPQTQQPQQPFPQQPQQPFPQTQQPQQPFPQFQQPHQPFPQPQQQFPQP
       Sv_CS_GAMMA-8   ----------------Q--------PQQPFPQPQQPQQQFPQSQQPQQPFPQPQQQFLQP
       Sv_CS_GAMMA-9   ----------------QPQQPFPQQPQQPFPQTQQPQQPFPQLQQPQQPFPQPQQQLPQP
      Sv_CS_GAMMA-10   QQPFPQQPQQPFPQTQQPQQPFPQQPQQPFPQTQQPQQPFPQFQQPHQPFPQPQQQFPQP
      Sv_CS_GAMMA-11   ----------------QPQQPFPQQPQQPFPQTQQPQQPFPQLQQPQQPFPQPQQQLPQP
      Sv_CS_GAMMA-12   ----------------Q--------PQQPFPQTQQPQQPFPQSKQPQQPFP-------QP

 Ae.tuschii_GAMMA-1   QQPQQSFPQQQQPLIQLSLQQQMNPCKNFLLQQCNPVSLVSSLISMILPRSDCQVMQQQC
Ae.tuschii_GAMMA-1A   QQPQQSFPQQQPPFIQPSLQQQVNPCKNFLLQQCKPVSLVSSLWSMIWPQSDCQVMRQQC
 Ae.tuschii_GAMMA-2   QQPQQSFPQQQQPLIQLSPQQQMNPCKNFLLQQCNPVSLVSSLISMILPRSDCQVMQQQC
 Ae.tuschii_GAMMA-3   QQPQQSFPQQQPPFIQPSLQQQVNPCKNFLLQQCNPVSLVSSLISMILPRSDCQVMQQQC
Ae.tuschii_GAMMA-3A   QQPQQSFPQQQPSLIQQSLQQQLNPCKNFLLQQCKPVSLVSSLWSMILPRSDCQVMRQQC
 Ae.tuschii_GAMMA-4   QQPQQSFPQQQPPFIQPSLQQQVNPCKNFLLQQCKPVSLVSSLWSMIWPQSDCQVMRQQC
Ae.tuschii_GAMMA-4A   QQPQQSFPQQQPSLIQQSLQQQLNPCKNFLLQQCKPVSLVSSLWSMILPRSDCQVMRQQC
 Ae.tuschii_GAMMA-5   QQPQQSFPQQQQPLIQLSLQQQMNPCKNFLLQQCNPVSLVSSLISMILPRSDCQVMQQQC
 Ae.tuschii_GAMMA-6   QQPQQSFPQQQQPLIQLSLQQQMNPCKNFLLQQCNPVSLVSSLISMILPRSDCQVMQQQC
Ae.tuschii_GAMMA-6A   QQPQQSFPQQQPSLIQQSLQQQLNPCKNFLLQQCKPVSLVSSLWSMILPRSDCQVMRQQC
 Ae.tuschii_GAMMA-7   QQPQQSFPQQQRPFIQPSLQQQLNPCKNILLQQCKPASLVSSLWSIIWPQSDCQVMRQQC
 Ae.tuschii_GAMMA-8   QQPQQSFPQQQQPLIQLSLQQQMNPCKNFLLQQCNPVSLVSSLISMILPRSDCQVMQQQC
 Ae.tuschii_GAMMA-9   QQPQQSFPQQQPSLIQQSLQQQLNPCKNFLLQQCKPVSLVSSLWSMILPRSDCQVMRQQC
Ae.tuschii_GAMMA-10   QQPQQSFPQQQQPLIQLSLQQQMNPCKNFLLQQCNPVSLVSSLISMILPRSDCQVMQQQC
Ae.tuschii_GAMMA-11   QQPQQSFPQQQPPFIQPSLQQQVNPCKNFLLQQCKPVSLVSSLWSMIWPQSDCQVMRQQC
      Sv_CNN_GAMMA-1   QQPQQSFPQQQPPFIQPSLQQQVNPCKNFLLQQCKPVSLVSSFWSMIWPQSDCQVMRQQC
     Sv_CNN_GAMMA-1A   QQPQQSFPQQQPPFIQPSLQQQVNPCKNFLLQQCKPVSLVSSLWSMIWPQSDCQVMRQQC
      Sv_CNN_GAMMA-2   QQPQQSFPQQQQPLIQLSLRQQMNPCKNFLLQQCNPVSLVSSLISMILPRSDCQVMQQQC
      Sv_CNN_GAMMA-3   QQPQQSFPQQQQPLIQLSLQQQMNPCKNFLLQQCNPVSLVSSLISMILPRSDCQVMQQQC
     Sv_CNN_GAMMA-3A   QQPQQSFPQQQPSLIQQSLQQQLNPCKNFLLQQCKPVSLVSSLWSMILPRSDCQVMRQQC
      Sv_CNN_GAMMA-5   QQPQQSFPQQQRPFIQPSLQQQLNPCKNILLQQCKPASLVSSLWSIIWPQSDCQVMRQQC
```

```
     Sv_CNN_GAMMA-5A   QQPQQSFPQQQPPFIQPSLQQQVNPCKNFLLQQCKPVSLVSSLWSMIWPQSDCQVMRQQC
      Sv_CNN_GAMMA-6   QQPQQSFPQQQPSLIQQSLQQQLNPCKNFLLQQCKPVSLVSSLWSMILPRSDCQVMRQQC
      Sv_CNN_GAMMA-8   QQPQQSFPQQQPSLIQQSLQHQLNPCKNFLLQQCKPVSLVSSLWSMILPRSDCQVMRQQC
      Sv_CNN_GAMMA-9   QQPQQSFPQQQPPFIQPSLQQQVSPCKNFLLQQCKPVSLVSSLWSMIWPQSDCQVMRQQC
     Sv_CNN_GAMMA-11   QQPQQSFPQQQRPFIQPSLQQQLNPCKNILLQQCKPASLVSSLWSIIWPQSDCQVMRQQC
       Sv_CS_GAMMA-1   QQPQQSFPQQQPPFIQPSLQQQVNPCKNFLLQQCKPVSLVSSLWSMIWPQSDCQVMRQQC
      Sv_CS_GAMMA-1A   QQPQQSFPQQQPSLIQQSLQQQLNPCKNFLLQQCKPVSLVSSLWSMILPRSDCQVMRQQC
       Sv_CS_GAMMA-2   QQPQQSFPQQQQPAIQSFLQQQMNPCKNFLLQQCNHVSLVSSLVSIILPRSDCQVMQQQC
       Sv_CS_GAMMA-3   QQPQQSFPQQQPPFIQPSLQQQVNPCKNFLLQQCKPVSLVSSLWSMIWPQSDCQVTRQQC
       Sv_CS_GAMMA-5   QQPQQSFPQQQPPFIQPSLQQQVNPCKNFLLQQCKPVSLVSSLWSMIWPQSDCQVMRQQC
      Sv_CS_GAMMA-6A   QQPQQSFPQQQQPLIQLSLQRQMNPCKNFLLQQCNPVSLVSSLISMILPRSDCQVMQQQC
       Sv_CS_GAMMA-7   QQPQQSFLQQQRPFIQPSLQQRLNPCKNILLQQCKPASLVSSLWSIIWPQSDCQVMQQQC
       Sv_CS_GAMMA-8   QQPQQSFPQQQQPLIQLSLQQQMNPCKNFLLQQCNPVSLVSSLISMILPRSDCQVMQQQC
       Sv_CS_GAMMA-9   QQPQQSFPQQQRSFIQPSLQQQLNPCKNILLQQCKPASLVSSLWSIIWPQSDCQVMRQQC
      Sv_CS_GAMMA-10   QQPQQSFLQQQRPFIQPSLQQRLNPCKNILLQQCKPASLVSSLWSIIWPQSDCQVMQQQC
      Sv_CS_GAMMA-11   QQPQQSFPQQQRSFIQPSLQQQLNPCKNILLQQCKPASLVSSLWSIIWPQSDCQVMRQQC
      Sv_CS_GAMMA-12   QQPQQSFPQQQPSLIQQSLQQQLNPCKNFLLQQCKPVSLVSSLWSIILPPSDCQVMRQQC

  Ae.tuschii_GAMMA-1   CQQLAQIPQQLQCAAIHSVVHSIIMQQ--------EQRQGVQIRRPLFQ--------LVQ
 Ae.tuschii_GAMMA-1A   CQQLAQIPQQLQCAAIHTVIHSITMQQ--------EQQQGMHILLPLYQQQQVGQGTLVQ
  Ae.tuschii_GAMMA-2   CQQLAQIPQQLQCAAIHSVVHSIIMQQ--------EQRQGVQIRRPLFQ--------LVQ
  Ae.tuschii_GAMMA-3   CQQLAQIPQQLQCAAIHSVVHSIIMQQ--------EQRQGVQIRRPLFQ--------LVQ
 Ae.tuschii_GAMMA-3A   CQQLAQIPQQLQCAAIHSIVHSIIMQQEQQ-----EQQQGVQILVPLSQQQQVGQGTLVQ
  Ae.tuschii_GAMMA-4   CQQLAQIPQQLQCAAIHSVVHSIIMQQQQQQ----QQQQGMHIFLPLSQQQQVGQGSLVQ
 Ae.tuschii_GAMMA-4A   CQQLAQIPQQLQCAAIHSIVHSIIMQQEQQ-----EQQQGVQILVPLSQQQQVGQGTLVQ
  Ae.tuschii_GAMMA-5   CQQLAQIPQQLQCAAIHSVVHSIIMQQ--------EQRQGVQIRRPLFQ--------LVQ
  Ae.tuschii_GAMMA-6   CQQLAQIPQQLQCAAIHSVVHSIIMQQ--------EQRQGVQIRRPLFQ--------LVQ
 Ae.tuschii_GAMMA-6A   CQQLAQIPQQLQCAAIHSIVHSIIMQQEQQ-----EQQQGVQILVPLSQQQQVGQGTLVQ
  Ae.tuschii_GAMMA-7   CQQLAQIPQQLQCAAIHSVVHSIIMQQQQQQ----QQQQGMHIFLPLSQQQQVGQGSLVQ
  Ae.tuschii_GAMMA-8   CQQLAQIPQQLQCAAIHSVVHSIIMQQ--------EQRQGVQIRRPLFQ--------LVQ
  Ae.tuschii_GAMMA-9   CQQLAQIPQQLQCAAIHSVVHSIIMQQQQQQ----QQQQGMHIFLPLSQQQQVGQGSLVQ
 Ae.tuschii_GAMMA-10   CQQLAQIPQQLQCAAIHSVVHSIIMQQ--------EQRQGVQIRRPLFQ--------LVQ
 Ae.tuschii_GAMMA-11   CQQLAQIPQQLQCAAIHTVIHSIIMQQ--------EQQQGMHILLPLYQQQQVGQGTLVQ
      Sv_CNN_GAMMA-1   CQQLAQIPQQLQCAAIHTIIHSIIMQQEQQ-----EQQQGMHILLPLYQQQQVGQGTLVQ
     Sv_CNN_GAMMA-1A   CQQLAQIPQQLQCAAIHTIIHSIIMQQEQQ-----EQQQGMHILLPLYQQQQVGQGTLVQ
      Sv_CNN_GAMMA-2   CQQLAQIPQQLQCAAIHSVVHSIIMQQ--------EQRQGVQIRRPLFQ--------LVQ
      Sv_CNN_GAMMA-3   CQQLAQIPQQLQCAAIHSVVHSIIMQQ--------EQRQGVQIRRPLFQ--------LVQ
     Sv_CNN_GAMMA-3A   CQQLAQIPQQLQCAAIHSIVHSIITQQEQQ-----EQRQGVQILVPLSQQQQVGQGTLVQ
      Sv_CNN_GAMMA-5   CQQLAQIPQQLQCAAIHSVVHSIIMQQQQQQ----QQQQGMHIFLPLSQQQQVGQGSLVQ
     Sv_CNN_GAMMA-5A   CQQLAQIPQQLQCAAIHTIIHSIIMQQEQQ-----EQQQGMHILLPLYQQQQVGQGTLVQ
      Sv_CNN_GAMMA-6   CQQLAQIPQQLQCAAIHSIVHSIIMQQEQQ-----EQRQGVQILVPLSQQQQVGQGTLVQ
      Sv_CNN_GAMMA-8   CQQLAQIPQQLQCAAIHSIVHSIIMQQEQQ-----EQRQGVQILVPLSQQQQVGQGTLVQ
      Sv_CNN_GAMMA-9   CQQLAQIPQQLQCAAIHTIIHSIIMQQEQQ-----EQQQGMHILLPLYQQQQVGQGTLVQ
     Sv_CNN_GAMMA-11   CQQLAQIPQQLQCAAIHSVVHSIIMQQQQQQ----QQQQGMHIFLPLSQQQQVGQGSLVQ
       Sv_CS_GAMMA-1   CQQLAQIPQQLQCAAIHTVIHSIIMQQ--------EQQQGMHILLPLYQQQQVGQGTLVQ
      Sv_CS_GAMMA-1A   CQQLAQIPQQLQCAAIHSIVHSIIMQQEQQ-----EQRQGVQILVPLSQQQQVGQGTLVQ
       Sv_CS_GAMMA-2   CQQLAQIPQQLQCAAIHSVAHSIIMQQ--------EQQQGVPILRPLFQ--------LAQ
       Sv_CS_GAMMA-3   CQQLAQIPQQLQCAAIHTVIHSIIMQQ--------EQQQGMHILLPLYQQQQVGQGTLVQ
       Sv_CS_GAMMA-5   CQQLAQIPQQLQCAAIHTVIHSIIMQQ--------EQQQGMHILLPLYQQQQVGQGTLVQ
      Sv_CS_GAMMA-6A   CQQLAQIPQQLQCAAIHSVVHSIIMQQ--------EQRQGVQIRRPLFQ--------LVQ
       Sv_CS_GAMMA-7   CQELAQIPQQLQCAAIHSVVHSIIMQQQQQQQQRQQQQQQGMHILLTLSQQQQLGQGTLVQ
       Sv_CS_GAMMA-8   CQQLAQIPQQLQCAAIHSVVHSIIMQQ--------EQRQGVQIRRPLFQ--------LVQ
       Sv_CS_GAMMA-9   CQQLAQIPQRLQCAAIHSVVHSIIMQQQQQQ----QQQQGMHIFLPLSQQQQVGQGSLVQ
      Sv_CS_GAMMA-10   CQELAQIPQQLQCAAIHSVVHSIIMQQQQQQQQQQQQQQQGMHILLTLSQQQQLGQGTLVQ
      Sv_CS_GAMMA-11   CQQLAQIPQQLQCAAIHSVVHSIIMQQQQQQ----QQQQGMHIFLPLSQQQQVGQGSLVQ
      Sv_CS_GAMMA-12   CQQLAQIPQQLQCAAIHSVVHSIIMQQEQQ-----EQLQGVQILVPLSQQQQVGQGILVQ

  Ae.tuschii_GAMMA-1   GQGIIQPQQPAQLEVIRSLVLRTLPTMCNVYVSPDCSTINAPFASIVVGIGGQ-------
 Ae.tuschii_GAMMA-1A   GQGIIQPQQPAQLEAIRSLVLQTLPTMCNVYVPPECSIIKAPFSSIVAGIGGQ-------
  Ae.tuschii_GAMMA-2   GQGIIQPQQPAQLEAIRSLVLQTLPTMCNVYVPPECSIIKAPFSSIVAGIGGQ-------
  Ae.tuschii_GAMMA-3   GQGIIQPQQPAQLEVIRSLVLRTLPTMCNVYVSPDCSTINAPFASIVVGIGGQ-------
 Ae.tuschii_GAMMA-3A   GQGIIQPQQPAQLEVIRSLVLQTLPTICNVYVPPYCSTIRAPFASIVAGIGGQ-------
  Ae.tuschii_GAMMA-4   GQGIIQPQQPAQLEAIRSLVLQTLPSMCNAYVPPECSIMRAPFASIVAGIGGQ-------
 Ae.tuschii_GAMMA-4A   GQGIIQPQQPAQLEVIRSLVLQTLPTICNVYVPPYCSTIRAPFASIVAGIGGQ-------
  Ae.tuschii_GAMMA-5   GQGIIQPQQPAQLEVIRSLVLRTLPTMCNVYVSPDCSTINAPFASIVVGIGGQ-------
  Ae.tuschii_GAMMA-6   GQGIIQPQQPAQLEVIRSLVLRTLPTMCNVYVSPDCSTINAPFASIVVGIGGQ-------
 Ae.tuschii_GAMMA-6A   GQGIIQPQQPAQLEVIRSLVLQTLPTICNVYVPPYCSTIRAPFASIVAGIGGQ-------
  Ae.tuschii_GAMMA-7   GQGIIQPQQPAQLEAIRSLVLQTLPSMCNVYVPPECSIMRAPFASIVAGIGGQ-------
  Ae.tuschii_GAMMA-8   GQGIIQPQQPAQLEVIRSLVLRTLPTMCNVYVSPDCSTINAPFASIVVGIGGQ-------
  Ae.tuschii_GAMMA-9   GQGIIQPQQPAQLEAIRSLVLQTLPSMCNVYVPPECSIMGAPFASIVAGIGGQ-------
```

```
Ae.tuschii_GAMMA-10   GQGIIQPQQPAQLEVIRSLVLRTLPTMCNVYVSPDCSTINAPFASIVVGIGGQ-------
Ae.tuschii_GAMMA-11   GQGIIQPQQPAQLEAIRSLVLQTLPTMCNVYVPPECSIIKAPFSSIVAGIGGQ-------
     Sv_CNN_GAMMA-1   GQGIIQPQQPAQLEAIRSLVLQTLPTMCNVYVPPECSIIKAPFSSVVAGIGGQ-------
    Sv_CNN_GAMMA-1A   GQGIIQPQQPAQLEAIRSLVLQTLPTMCNVYVPPECSIIKAPFSSVVAGIGGQ-------
     Sv_CNN_GAMMA-2   GQGIIQPQQPAQLEVIRSLVLRTLPTMCNVYVSPDCSTINAPFANIVVGIGGQ-------
     Sv_CNN_GAMMA-3   GQGIIQPQQPAQLEVIRSLVLRTLPTMCNVYVSPDCSTINAPFANIVVGIGGQ-------
    Sv_CNN_GAMMA-3A   GQGIIQPQQPAQLEVIRSLVLQTLATMCNVYVPPYCSTIRAPFASIVAGIGGQ-------
     Sv_CNN_GAMMA-5   GQGIIQPQQPAQLEAIRSLVLQTLPSMCNVYVPPECSIMRAPFASIVAGIGGQ-------
    Sv_CNN_GAMMA-5A   GQGIIQPQQPAQLGAIRSLVLQTLPTMCNVYVPPECSIIKAPFSSVVAGIGGQ-------
     Sv_CNN_GAMMA-6   GQGIIQPQQPAQLEVIRSLVLQTLATMCNVFVPPYCSTIRAPFASIVAGIGGQ-------
     Sv_CNN_GAMMA-8   GQGIIQPQQPAQLEVIRSLVLQTLATMCNVYVPPYCSTIRAPFASIVAGIGGQ-------
     Sv_CNN_GAMMA-9   GQGIIQPQQPAQLEAIRSLVLQTLPTMCNVFVPPECSIIKAPFSSVVAGIGGQ-------
    Sv_CNN_GAMMA-11   GQGIIQPQQPAQLEAIRSLVLQTLPSMCNVYVPPECSIMRAPFASIVAGIGGQ-------
      Sv_CS_GAMMA-1   GQGIIQPQQPAQLEAIRSLVLQTLPTMCNAYVSPDCSTINAPFASIVVGIGGQ-------
     Sv_CS_GAMMA-1A   GQGIIQPQQPAQLEVIRSLVLQTLATMCNVYVPPYCSTIRAPFASIVAGIGGQ-------
      Sv_CS_GAMMA-2   GLGIIQPQQPAQLEGIRSLVLKTLPTMCNVYVPPDCSTINVPYANIDAGIGGQ-------
      Sv_CS_GAMMA-3   GQGIIQPQQPAQLEAIRSLVLQTLPTMCNVYVPPECSIIKAPFSSVVAGIGGQ-------
      Sv_CS_GAMMA-5   GQGIIQPQQPAQLEAIRSLVLQTLPTMCNVYVPPECSIIKAPFSSVVAGIGGQ-------
     Sv_CS_GAMMA-6A   GQGIIQPQQPAQLEVIRSLVLRTLPTMCNVYVSPDCSTINAPFASIVVGIGGQ-------
      Sv_CS_GAMMA-7   GQGIIQPQQLAQLEAIRSLVLQTLPTMCNVYVPPECSIIRAPFASIRASSGHHLPALVAG
      Sv_CS_GAMMA-8   GQGIIQPQQPAQLEVIRSLVLRTLPTMCNVYVSPDCSTINAPFASIVVGIGGQ-------
      Sv_CS_GAMMA-9   GQGIIQPQQPAQLEAIRSLVLQTLPSMCNVYVPPECSIMRAPFASIVAGIGGQ-------
     Sv_CS_GAMMA-10   GQGIIQPQQLAQLEAIRSLVLQTPPTMCNVYVPPECSIIRAPFASIRASSGHHLPALVAG
     Sv_CS_GAMMA-11   GQGIIQPQQPAQLEAIRSLVLQTLPSMCNVYV-PECSIMRAPFASIVAGIGGQ-------
     Sv_CS_GAMMA-12   GQGIIQPQQPAQLEVIRSLVLQTLPTMCNVFVPPYCSTIRAPFASIVASIGGQ-------

 Ae.tuschii_GAMMA-1   ----
Ae.tuschii_GAMMA-1A   ----
 Ae.tuschii_GAMMA-2   ----
 Ae.tuschii_GAMMA-3   ----
Ae.tuschii_GAMMA-3A   ----
 Ae.tuschii_GAMMA-4   ----
Ae.tuschii_GAMMA-4A   ----
 Ae.tuschii_GAMMA-5   ----
 Ae.tuschii_GAMMA-6   ----
Ae.tuschii_GAMMA-6A   ----
 Ae.tuschii_GAMMA-7   ----
 Ae.tuschii_GAMMA-8   ----
 Ae.tuschii_GAMMA-9   ----
Ae.tuschii_GAMMA-10   ----
Ae.tuschii_GAMMA-11   ----
     Sv_CNN_GAMMA-1   ----
    Sv_CNN_GAMMA-1A   ----
     Sv_CNN_GAMMA-2   ----
     Sv_CNN_GAMMA-3   ----
    Sv_CNN_GAMMA-3A   ----
     Sv_CNN_GAMMA-5   ----
    Sv_CNN_GAMMA-5A   ----
     Sv_CNN_GAMMA-6   ----
     Sv_CNN_GAMMA-8   ----
     Sv_CNN_GAMMA-9   ----
    Sv_CNN_GAMMA-11   ----
      Sv_CS_GAMMA-1   ----
     Sv_CS_GAMMA-1A   ----
      Sv_CS_GAMMA-2   ----
      Sv_CS_GAMMA-3   ----
      Sv_CS_GAMMA-5   ----
     Sv_CS_GAMMA-6A   ----
      Sv_CS_GAMMA-7   IGGQ
      Sv_CS_GAMMA-8   ----
      Sv_CS_GAMMA-9   ----
     Sv_CS_GAMMA-10   IGGQ
     Sv_CS_GAMMA-11   ----
     Sv_CS_GAMMA-12   ----
```

Figure 5.10: Amino acid alignment of ω-gliadin deduced proteins of three genotypes.

5.3.4 Isolation of Genes Encoding LMW-GS

The genes encoding LMW-GS were isolated from *Aegilops tauschii spp. strangulata* (G1276), Sv 1BL.1RS 1AL.1DScs and Sv 1BL.1RS 1AL.1DScnn lines using different combination of primer pairs (Tab 5.2). Excluding several redundant and pseudo genes, 21 unique genes encoding LMW-GS were found ranging from 894 to 1134 nucleotides. After the removal of the signal peptide, the deduced mature proteins were aligned and showed the typical structure of the LMW-GS with MW mean of 34.582 KDa and a length between 278 and 358 amino acid residues. The N-terminal domain length was the same for all deduced amino acid sequences (12 aa) whereas the length of REP domain and the C term domain ranged from 84 to 160 and from 172 to 187 aa.

Among 21 isolated genes only one was found in Sv 1BL.1RS 1AL.1DScs encoding for LMW-s type with MENSHIP N-terminal amino acid sequence and, the other 20 genes encoded for LMW-m type with different N-terminal sequences (METRCIP, METSCIS, METSRV, METSCIP, METSHIP). Ten genes encoding LMW-m were isolated from *Aegilops tauschii*, six genes were isolated from 1D chromosome of *Cheyenne* genotype and four genes were isolated from 1D chromosome of *Chine Spring* genotype (Tab 5.7).

Table 5.7: LMW-GS genes isolated from *Aegilops tauschii strangulata (G1276) Sv 1BL.1RS 1AL.1DScs Sv 1BL.1RS 1AL.1DScnn lines***.**

GENOTYPE	Sequence ID	Class	Type	N-TERM	MW (KDa)	Length	N-TERM domain		REP domain		C-TERM domain		Total cysteine residues
							Length	Cysteine num.	Length	Cysteine num.	Length	Cysteine num.	
Ae.tauschii ssp. *strangulata*	Ae.tauschii_LMW9	LMW-GS	lmw-m	METRCIP	32.619	285	12	1	101	0	97	7	8
	Ae.tauschii_LMW14	LMW-GS	lmw-m	METRCIP	32.637	285	12	1	101	0	97	7	8
	Ae.tauschii_LMW6	LMW-GS	lmw-m	METSCIS	32.5	286	12	1	87	0	112	7	8
	Ae.tauschii_LMW7	LMW-GS	lmw-m	METSCIS	32.533	286	12	1	87	0	112	6	7
	Ae.tauschii_LMW10	LMW-GS	lmw-m	METSCIS	32.472	286	12	1	87	0	112	7	8
	Ae.tauschii_LMW8	LMW-GS	lmw-m	METSCIS	32.461	286	12	1	87	0	112	7	8
	Ae.tauschii_LMW12	LMW-GS	lmw-m	METSHIP	32.576	285	12	0	101	0	97	7	8
	Ae.tauschii_LMW13	LMW-GS	lmw-m	METSHIP	32.602	285	12	0	101	0	97	7	8
	Ae.tauschii_LMW11	LMW-GS	lmw-m	METSHIP	37.813	331	12	0	133	1	111	7	8
	Ae.tauschii_LMW1	LMW-GS	lmw-m	METSRV	36.949	319	12	0	133	1	112	6	7
Sv 1BL.1RS 1AL.1DS$_{CS}$	SvCS_LMW5	LMW-GS	lmw-s	MENSHIP	38.025	334	12	0	150	1	117	7	8
	SvCS_LMW6	LMW-GS	lmw-m	METRCIP	32.879	287	12	1	103	0	97	7	8
	SvCS_LMW3	LMW-GS	lmw-m	METSCIP	31.773	278	12	1	84	0	107	7	8
	SvCS_LMW2	LMW-GS	lmw-m	METSCIS	32.047	283	12	1	87	0	109	7	8
	SvCS_LMW1	LMW-GS	lmw-m	METSRV	41.049	358	12	0	160	1	111	7	8
Sv 1BL.1RS 1AL.1DS$_{CNN}$	SvCNN_LMW2	LMW-GS	lmw-m	METRCIP	32.841	287	12	1	103	0	97	7	8
	SvCNN_LMW3	LMW-GS	lmw-m	METRCIP	32.879	287	12	1	103	0	97	7	8
	SvCNN_LMW4	LMW-GS	lmw-m	METRCIP	32.907	286	12	1	103	0	97	7	8
	SvCNN_LMW1	LMW-GS	lmw-m	METSHIP	37.881	330	12	0	134	1	109	7	8
	SvCNN_LMW3	LMW-GS	lmw-m	METSHIP	37.845	330	12	0	134	1	109	7	8
	SvCNN_LMW7	LMW-GS	lmw-m	METSHIP	39.712	345	12	0	149	1	109	7	8

```
  Ae.tauschii_METRCIP-9   METRCIPGLERPWQQQPLP-PQQTFP---QQPLFSQQQ------QLFP----QQPSFSQQ
 Ae.tauschii_METRCIP-14   METRCIPGLERPWQQQPLP-PQQTFP---QQPLFSQQQ------QLFP----QQPSFSQQ
  Ae.tauschii_METSCIS-6   METSCISGLERPWQQQPLP-PQQSFS---QQPPFSQQQQQ----QPLP----QQPSFSQQ
  Ae.tauschii_METSCIS-7   METSCISGLERPWQQQPLP-PQQSFS---QQPPFSQQQQR----QPLP----QQPSFSQQ
 Ae.tauschii_METSCIS-10   METSCISGLERPWQQQPLP-PQQSFS---QQPPFSQQQQQ----QPLP----QQPSFSQQ
  Ae.tauschiI_METSCIS-8   METSCISGLERPWQQQPLP-PQQSFS---QQPPFSQQQQQ----QPLP----QQPSFSQQ
 Ae.tauschii_METSHIP-12   METSHIPGLERPWQQQPLP-PQQTFP---QQPLFSQQQ------QLFP----QQPSFSQQ
 Ae.tauschii_METSHIP-13   METSHIPGLERPWQQQPLP-PQQTFP---QQPLFSQQQ------QLFP----QQPSFSQQ
Ae.tauschii_METSHIP- 11   METSHIPGLERPWQQQPLP-PQQQPPCSQQQQPFPQQQQPIIILQQSPFSQQQQPVLPQQ
   Ae.tauschii_METSRV-1   METSRVPGLEQPWQQQPLP-PQQQPPCSQQQQPFPQQQQPIIILQQSPFSQQQQPVLPQQ
       Sv_CNN_METRCIP-2   METRCIPGSERPWQQQPLP-PQQTFP---QQPLFSQQQQQ----QLFP----QQPSFSQQ
       Sv_CNN_METRCIP-3   METRCIPGLERPWQQQPLP-PQQTFP---QQPLFSQQQQQ----QLFP----QQPSFSQQ
       Sv_CNN_METRCIP-4   METRCIPGLERPWQQQPLP-PQQTFP---QQPLFSQQQQQ----QLFP----QQPSFSQQ
       Sv_CNN_METSHIP-1   METSHIPSLEKPLQQQPLPLQQILWY--QQQQPIQQQPQPFP--QQPPCSQQQQPPLSQQ
       Sv_CNN_METSHIP-3   METSHIPSLEKPLQQQPLPLQQILWY--QQQQPIQQQPQPFP--QQPPCSQRQQPPLSQQ
       Sv_CNN_METSHIP-7   METSHIPGLEKPSQQQPLPLQQILWY--HQQQPIQQQPQPFP--QQPPCSQQQQPPLSQQ
        Sv_CS_MENSHIP-5   MENSHIPGLEKPSQQQPLPLQQTLSHQQQQQPVQQQPQPFPQQQPCSQQQQPPLSQQQQP
        Sv_CS_METRCIP-6   METRCIPGLERPWQQQPLP-PQQTFP---QQPLFSQQQQQ----QLFP----QQPSFSQQ
        Sv_CS_METSCIP-3   METSCIPGLERPWQQQPLQ-QKETFP---QQPPSSQQQ------QPFP----QQPPFLQQ
        Sv_CS_METSCIS-2   METSCISGLERPWQQQPLP-PQQSFS---QQPPFSQQQQ-----QPLP----QQPSFSQQ
         Sv_CS_METSRV-1   METSRVPGLEKPWQQQPLP-PQQQPPCSQQQQPFPQQQQPIIILQQSPFSQQQQPVLPQQ

  Ae.tauschii_METRCIP-9   -QPPFWQQQPPFSQQQ-PILPQQ-------P---------PFSQQQQLVLPQQPPFS---
 Ae.tauschii_METRCIP-14   -QPPFWQQQPPFSQQQ-PILPQQ-------P---------PFSQQQQLVLPQQPPFS---
  Ae.tauschii_METSCIS-6   --------QPPFSQQQ-PILSQQ-------P---------PFSQQQQPVLPQQSPFS---
  Ae.tauschii_METSCIS-7   --------QPPFSQQQ-PILSQQ-------P---------PFSQQQQPVLPQQSPFS---
 Ae.tauschii_METSCIS-10   --------QPPFSQQQ-PILSQQ-------P---------PFSQQQQPVLPQQSPFS---
  Ae.tauschiI_METSCIS-8   --------QPPFSQQQ-PILSQQ-------P---------PFSQQQQPVLPQQSPFS---
 Ae.tauschii_METSHIP-12   -QPPFWQQQPPFSQQQ-PILPQQ-------P---------PFSQQQQLVLPQQPPFS---
 Ae.tauschii_METSHIP-13   -QPPFWQQQPPFSQQQ-PILPQQ-------P---------PFSQQQQLVLPQQPPFS---
Ae.tauschii_METSHIP- 11   -QPVIILQQPPFSQQQQPVLPQQPPSSQQQQQQ-QQQQQPPFSQQQQPVLPQQPPFS---
   Ae.tauschii_METSRV-1   -QPVIILQQPPFSQQQQPVLPQQPPFSQQQQQQ-QQQQQPPFSQQQQPVLPQQPPFS---
       Sv_CNN_METRCIP-2   -QPPFWQQQPPFSQQQ-PILPQQP----------------PFSQQQQLVLPQQPPFS---
       Sv_CNN_METRCIP-3   -QPPFWQQQPPFSQQQ-PILPQQP----------------PFSQQQQLVLPQQPPFS---
       Sv_CNN_METRCIP-4   -QPPFWQQQPPFSQQR-PILPQQP----------------PFSQQQQLVLPQQPPFS---
       Sv_CNN_METSHIP-1   QQPPFSQQQPPFSQQQQPILPQQPPFSQQQQ---------QFPQQQQPLLPQQPPFSQQR
       Sv_CNN_METSHIP-3   QQPPFSQQQPPFSQQQQPILPQQPPFSQQQQ---------QFPQQQQPLLPQQPPFSQQQ
       Sv_CNN_METSHIP-7   QQPPFSQQQPPFSQQELPILPQQPPFSQQQQPQ-FSQQQQPFPQQQQPLLLQQPPFSQQR
        Sv_CS_MENSHIP-5   PFSQQQPPFSQQQQPSFSQQQQPPFSQQQPPFSQQQQPVLPQQPSFSQQQLPPF-SQQQP
        Sv_CS_METRCIP-6   -QPPFWQQQPPFSQQQ-PILPQQP----------------PFSQQQQLVLPQQPPFS---
        Sv_CS_METSCIP-3   -QPSFS-QQPLFSQKQQPVLPQQP------A----------FSQQQQTVLPQQPAFS---
        Sv_CS_METSCIS-2   --------QPPFSQQQ-PILSQQP------P----------FSQQQQPVLPQQSPFS---
         Sv_CS_METSRV-1   -QPVIILQQPPFSQQQQPVLPQQPPFSQQQQQQ--QQQQPPFSQQQQPVLPQQPPFSQQQ


  Ae.tauschii_METRCIP-9   ----------QQQQPVLP----------P---------QQSPFPQQQQQHQQLVQQQI
 Ae.tauschii_METRCIP-14   ----------QQQQPVLP----------P---------QQSPFPQQQQQHQQLVQQQI
  Ae.tauschii_METSCIS-6   ----------QQQQLVLP----------P---------------QQQQQ---LVQQQI
  Ae.tauschii_METSCIS-7   ----------QQQQLVLP----------P---------------QQQQQ---LVQQQI
 Ae.tauschii_METSCIS-10   ----------QQQQLVLP----------P---------------QQQQQ---LVQQQI
  Ae.tauschiI_METSCIS-8   ----------QQQQLVLP----------P---------------QQQQQ---LVQQQI
 Ae.tauschii_METSHIP-12   ----------QQQQPVLP----------P---------QQSPFPQQQQQHQQLVQQQI
 Ae.tauschii_METSHIP-13   ----------QQQQPVLP----------P---------QQSPFPQQQQQHQQLVQQQI
Ae.tauschii_METSHIP- 11   -----Q-----QQQPPFSQ--------------QQQPSSQQPPFPQQHQQ---FPQQQI
   Ae.tauschii_METSRV-1   -----Q-----QQQPPFSQ----------Q----QQPSSQQPPFPQQHQQ---FPQQQI
       Sv_CNN_METRCIP-2   ----------QQQQPVLP----------P---------QQSPFPQQQQQHQQLVQQQI
       Sv_CNN_METRCIP-3   ----------QQQQPVLP----------P---------QQSPFPQQQQQHQQLVQQQI
       Sv_CNN_METRCIP-4   ----------QQQQPVLP----------P---------QQSPFPQQQQQHQQLVQQQI
       Sv_CNN_METSHIP-1   PPFSQQ-----QQQPPFSQQ-------------QQQPILLQQPPFSRHQQP--VLPQQQI
       Sv_CNN_METSHIP-3   PPFSQQ-----QQQPPFSQQ-------------QQQPILLQQPPFSQHQQP--VLPQQQI
       Sv_CNN_METSHIP-7   PPFSQ------QQQQPVLPQQPPFSQ-----QQQQQPILPQQPPFSQHQQP--VLPQQQI
        Sv_CS_MENSHIP-5   PFSQQQQPVLPQQPPFS--------------QQQQPILPQ--------------------
        Sv_CS_METRCIP-6   ----------QQQQPVLP----------P---------QQSPFPQQQQQHQQLVQQQI
        Sv_CS_METSCIP-3   ------------QQ-----------------Q---------------HQQ----LLQQQI
        Sv_CS_METSCIS-2   ----------QQQQLVLP------------P---------------QQQQQ--QLVQQQI
         Sv_CS_METSRV-1   QQQQQQQPPFSQQQQPVLPQQPPFSQQQQPPFSQQQQPSSQQPPFPQQHQQ---FPQQQI
```

```
  Ae.tauschii_METRCIP-9   PVVQPSILQQLNPCKLFLQQQCSPVAIPQRLARSQMLQQSSCHVMQQQCCQQLPQIPQQS
 Ae.tauschii_METRCIP-14   PVVQPSILQQLNPCKLFLQQQCSPVAMPQRLARSQMLQQSSCHVMQQQCCQQLPQIPQQS
  Ae.tauschii_METSCIS-6   PIVQPSVLQQLNPCKVFLQQQCSPVAMPQRLARSQMWQQSSCHVMQQQCCQQLQQIPEQS
  Ae.tauschii_METSCIS-7   PIVQPSVLQQLNPCKVFLQQQCSPVAMPQRLARSQMWQQSSCHVMQQQCCQQLQQIPEQS
 Ae.tauschii_METSCIS-10   PIVQPSVLQQLNPCKVFLQQQCSPVAMPQRLARSQMWQQSSCHVMQQQCCQQLQQIPEQS
  Ae.tauschiI_METSCIS-8   PIVQPSVLQQLNPCKVFLQQQCSPVAMPQRLARSQMWQQSSCHVMQQQCCQQLQQIPEQS
 Ae.tauschii_METSHIP-12   PVVQPSILQQLNPCKLFLQQQCSPVAMPQRLARSQMLQQSSCHVMQQQCCQQLPQIPQQS
 Ae.tauschii_METSHIP-13   PVVQPSILQQLNPCKLFLQQQCSPVAMPQRLARSQMLQQSSCHVMQQQCCQQLPQIPQQS
Ae.tauschii_METSHIP- 11   PVVQPSVLQQLNPCKVFLQQQCSHVAMSQRLARSQMWQQSSCHVMQQQCCQQLPQIPEQS
   Ae.tauschii_METSRV-1   PVVQPSVLQQLNPCKVFLQQQCSHVAMSQRLARSQMWQQSSCHVMQQQCCQQLPQIPEQS
        Sv_CNN_METRCIP-2   PVVQPSILQQLNPCKVFLQQQCSPVAMPQRLARSQMLQQSSCHVMQQQCCQQLPQIPQQS
        Sv_CNN_METRCIP-3   PVVQPSILQQLNPCKVFLQQQCSPVAMPQRLARSQMLQQSSCHVMQQQCCQQLPQIPQQS
        Sv_CNN_METRCIP-4   PVVQPSILQQLNPCKVFLQQQCSPVAMPQRLARSQMLQQSSCHVMQQQCCQQLPQIPQQS
        Sv_CNN_METSHIP-1   PSVQPSILQQLNPCKVFLQQQCSPVAMPQSLARSQMLWQSSCHVMQQQCCRQLPQIPEQS
        Sv_CNN_METSHIP-3   PSVQPSILQQLNPCKVFLQQQCSPVAMPQSLARPQMLWQSSCHVMQQQCCRQLPQIPEQS
        Sv_CNN_METSHIP-7   PYVQPSILQQLNPCKVFLQQHCSPVAMPQSLARSQMLWQSSCHVMQQQCCQQLPRIPEQS
         Sv_CS_MENSHIP-5   -QPPFSQQQQQ-PVLPQQQILFVHPSILQQLNPCKVFLQQQCSPVAMPQSLARSQMLQQS
         Sv_CS_METRCIP-6   PVVQPSILQQLNPCKVFLQQQCSPVAMPQRLARSQMLQQSSCHVMQQQCCQQLPQIPQQS
         Sv_CS_METSCIP-3   PIVHPSILQQLNPCKVFLQQQCSPVAMPQHLARSQMWQQSSCNVMQQQCCQQLPRIPEQS
         Sv_CS_METSCIS-2   PIVQPSVLQQLNPCKVFLQQQCSPVAMPQRLARSQMWQQSSCHVMQQQCCQQLQQIPEQS
          Sv_CS_METSRV-1   PVVQPSVLQQLNPCKVFLQQQCSHVAMSQRLARSQMWQQSSCHVMQQQCCQQLPQIPEQS

  Ae.tauschii_METRCIP-9   RYEAIRAIIYSIILQEQQQVQGSIQSQQQQPQQLGQCVSQPQQ--QSQQ-----------
 Ae.tauschii_METRCIP-14   RYEAIRAIIYSIILQEQQQVQGSIQSQQQQPQQLGQCVSQPQQ--QSQQ-----------
  Ae.tauschii_METSCIS-6   RYEAIRAIIYSIILQEQQR--GFVQPQQQQPQQSGQGVSQSQQ--QSQQQLGQCSFQQPQ
  Ae.tauschii_METSCIS-7   RYEAIRAIIYSIILQEQQQ--GFVQPQQQQPQQSGQGVSQSQQ--QSQQQLGQCSFQQPQ
 Ae.tauschii_METSCIS-10   RYEAIRAIIYSIILQEQQQ--GFVQPQQQQPQQSGQGVSQSQQ--QSQQQLGQCSFQQPQ
  Ae.tauschiI_METSCIS-8   RYEAIRAIIYSIILQEQQQ--GFVQPQQQQSQQSGQGVSQSQQ--QSQQQLGQCSFQQPQ
 Ae.tauschii_METSHIP-12   RHEAIRAIIYSIILQEQQQVQGSIQSQQQQPQQLGQCVSQPQQ--QSQQ-----------
 Ae.tauschii_METSHIP-13   RYEAIRAIIYSIILQEQQQVQGSIQSQQQQPQQLGQCVSQPQQ--QSQQ-----------
Ae.tauschii_METSHIP- 11   RSEAIRAIVYSIILQEQQQ--GFVQPQQQQPQQSGQGVSQHQQQSQQQQQLGQCSFQQPQ
   Ae.tauschii_METSRV-1   RSEAIRAIVYSIILQEQQQ--GFVQPQQQQPQQSGQGVSQHQQQSQQQQQLGQCSFQQPQ
        Sv_CNN_METRCIP-2   RYEAIRAIIYSIILQEQQQVQGSIQSQQQQPQQLGQCVSQPQQ--QSQQ-----------
        Sv_CNN_METRCIP-3   RYEAIRAIIYSIILQEQQQVQGSIQSQQQQPQQLGQCVSQPQQ--QSQQ-----------
        Sv_CNN_METRCIP-4   RYEAIRAIIYSIILQEQQQVQGSIQSQQQQPQQLGQCVSQPQQ--QSQQ-----------
        Sv_CNN_METSHIP-1   RYDAIRAIIYSIVLQEQQHGQGLNQPQQQQPQQSVQGVSQPQ---QQQKQLGQCSFQRPQ
        Sv_CNN_METSHIP-3   RYDAIRAIIYSIVLQEQQHGQGLNQPQQQQPQQSVQGVSQPQ---QQQKQLGQCSFQQPQ
        Sv_CNN_METSHIP-7   RYDAIRAIIYSIVLQEQQHGQGFNQPHQQQPQQSVQGVSQPQ---QQQKQLGQCFFQRPQ
         Sv_CS_MENSHIP-5   SCHVMQQQCCQQLLQIPQQSRYEAIRAIIYSIILQEQQQVQGSIQTQQQQPQELGQCVSQ
         Sv_CS_METRCIP-6   RYEAIRAIIYSIILQEQQQVQGSIQSQQQQPQQLGQCVSQPQQ--QSQQ-----------
         Sv_CS_METSCIP-3   RYEAIRAIIFSIILQEQQQ--GFVQPQQQQPQQSVQGVYQPQQ--QSQQQLGQCSFQQPQ
         Sv_CS_METSCIS-2   RYEAIRAIIYSIILQEQQQ--GFVQPQQQQPQQSGQGVSQSQQ--QSQQQLGQCSFQQPQ
          Sv_CS_METSRV-1   RSEAIRAIVYSIILQEQQQ--GFVQPQQQQPQQSGQGVSQHQQQSQQQQQLGQCSFQQPQ

  Ae.tauschii_METRCIP-9   ---QLGQQP-----QQQQLAQGTFLQPHQIAQLEVMTSIALRILPTMCSVNV-PLYRTTT
 Ae.tauschii_METRCIP-14   ---QLGQQP-----QQQQLAQGTFLQPHQIAQLEVMTSIALRILPTMCSVNV-PLYRTTT
  Ae.tauschii_METSCIS-6   Q--QLGQQPQQQQQQQQQQVLQGTFLQPHQIAHLEVVLSIALRTLPTMCSVNV-PLYSATT
  Ae.tauschii_METSCIS-7   Q--QLGQQPQQQQQQQQQQVLQGTFLQPHQIAHLEVVLSIALRTLPTMRSVNV-PLYSATT
 Ae.tauschii_METSCIS-10   Q--QLGQQPQQQQQQQQQQVLQGTFLQPHQIAHLEVVLSIALRTLPTMCSVNV-PLYSATT
  Ae.tauschiI_METSCIS-8   Q--QLGQQPQQQQQQQQQQVLQGTFLQPHQIAHLEVVLSIALRTLPTMCSVNV-PLYSATT
 Ae.tauschii_METSHIP-12   ---QLGQQP-----QQQQLAQGTFLQPHQIAQLEVMTSIALRILPTMCSVNV-PLYRTTT
 Ae.tauschii_METSHIP-13   ---QLGQQP-----QQQQLAQGTFLQPHQIAQLEVMTSIALRILPTMCSVNV-PLYRTTT
Ae.tauschii_METSHIP- 11   QLQQLGQQP-----QQQQIPQGIFLQPHQISQLEVMTSIALRTLPTMCGVNV-PLYSSTT
   Ae.tauschii_METSRV-1   QLQQLGQQPQ-----QQQIPQGIFLQPHQISQLEVMTSIALRTLPTMVRVSTWPLYSSTT
        Sv_CNN_METRCIP-2   ---QLGQQP-----QQQQLAQGTFLQPHQIAQLEVMTSIALRTLPTMCSVNV-PLYRTTT
        Sv_CNN_METRCIP-3   ---QLGQQP-----QQQQLAQGTFLQPHQIAQLEVMTSIALRILPTMCSVNV-PLYRTTT
        Sv_CNN_METRCIP-4   ---QLGQQP-----QQQQLAQGTFLQPHQIAQLEVMTSIALRILPTMCSVNV-PLYRTTT
        Sv_CNN_METSHIP-1   Q-QQLGQWP-----QQQQVPQGTLLQPHQIAQLELMTSIALRTLPMMCSVNV-PVYGTTT
        Sv_CNN_METSHIP-3   Q-QQLGQWP-----QQQQVPQGTLLQPHQIAQLEVMTSIALRTLPTMCSVNV-PVYGTTT
        Sv_CNN_METSHIP-7   Q-QQLGQWP-----QQQQVPQGTLLQPHQIAQLELMTSIALRTLPMMCSVNV-PVYGTTT
         Sv_CS_MENSHIP-5   PQQ--QSQQ--------------QLGQQP-----QQQQLAQGTFLQPHQIAQLEVMTSIA
         Sv_CS_METRCIP-6   ---QLGQQP-----QQQQLAQGTFLQPHQIAQLEVMTSIALRILPTMCSVNV-PLYRTTT
         Sv_CS_METSCIP-3   Q--QLGQQP-----QQQQVQKGTFLQPHQIARLEVMTSIALRTLPTMCSVNV-PLYSSIT
         Sv_CS_METSCIS-2   Q--QLGQQPQQ---QQQQVLQGTFLQPHQIAHLEAVTSIALRTLPTMCSVNV-PLYSATT
          Sv_CS_METSRV-1   QLQQLGQQP-----QQQQIPQGIFLQPHQISQLEVMTSIALRTLPTMCGVNV-PLYSSTT
```

```
  Ae.tauschii_METRCIP-9   SVPFGVGTGVGAY--------------------
 Ae.tauschii_METRCIP-14   SVPFGVGTGVGAY--------------------
  Ae.tauschii_METSCIS-6   SVPFGVGTGVGAY--------------------
  Ae.tauschii_METSCIS-7   SVPFGVGTGVGAY--------------------
 Ae.tauschii_METSCIS-10   SVPFGVGTGVGAY--------------------
  Ae.tauschiI_METSCIS-8   SVPFGVGTGVGAY--------------------
 Ae.tauschii_METSHIP-12   SVPFGVGTGVGAY--------------------
 Ae.tauschii_METSHIP-13   SVPFGVGTGVGAY--------------------
Ae.tauschii_METSHIP- 11   IMPFSIGTGVGAY--------------------
   Ae.tauschii_METSRV-1   IMPFSIGTGVGGY--------------------
       Sv_CNN_METRCIP-2   SVPFGVGTGVGAY--------------------
       Sv_CNN_METRCIP-3   SVPFGVGTGVGAY--------------------
       Sv_CNN_METRCIP-4   SVPFGVGTGVGAY--------------------
       Sv_CNN_METSHIP-1   SVPFGVGTQVGAY--------------------
       Sv_CNN_METSHIP-3   IVPFGVGTRVGAY--------------------
       Sv_CNN_METSHIP-7   SVPFGVGTQVGAY--------------------
        Sv_CS_MENSHIP-5   LRTLPTMCRVNV-PLYRTTTSVPFGVGAGVGAY
        Sv_CS_METRCIP-6   SVPFGVGTGVGAY--------------------
        Sv_CS_METSCIP-3   SAPLGVGSRVGAY--------------------
        Sv_CS_METSCIS-2   SVPFGVGTGVGAY--------------------
         Sv_CS_METSRV-1   IMPFSIGTGVGGY--------------------
```

Figure 5.11: Amino acid alignment of ω-gliadin deduced proteins of three genotypes.

From the alignment among all deduced amino acid sequences (Fig. 5.11), the pairwise ID was 74.6 % with a minimum pairwise ID of 72.7 % between the alignment of the sequences from Sv 1BL.1RS 1AL.1DScs and Sv 1BL.1RS 1AL.1DScnn lines (Fig 5.12). Otherwise, the pairwise ID between deduced proteins from Sv 1BL.1RS 1AL.1DScs and *Aegilops tauschii* and from Sv 1BL.1RS 1AL.1DScnn and *Aegilops tauschii* was almost equal (Fig.5.12).

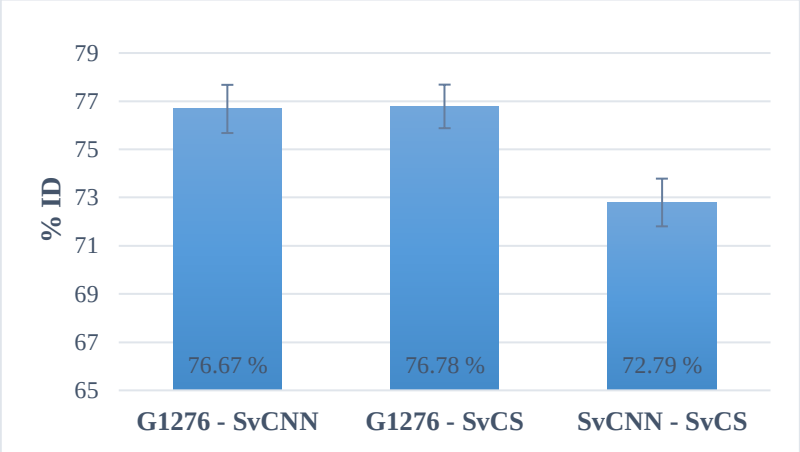

Figure 5.12: Pairwise identity of LMW-GS amino acid sequences deduced from genes isolated from three genotypes. *SvCNN* = Sv 1BL.1RS 1AL.1DScnn; SvCS = Sv 1BL.1RS 1AL.1DScs; G1276 = *Aegilops tauschii* spp. *strangulata*

The LMW-m type subunits with N-terminal sequence METRCIP were the most conserved among the three genotypes with pairwise ID of 98.97 %. At this regards, the deduced proteins Sv_CNN_METRCIP-3 and Sv_CS_METRCIP-6 from were identical.

LMW-GS subunits with the METSHIP N-terminal sequence were isolated from *Aegilops tauschii* and Sv 1BL.1RS 1AL.1DScnn lines whereas subunits with the METSRV and METSCIS N-terminal sequences were identified in Sv 1BL.1RS 1AL.1DScs and *Aegilops tauschii*. Notably, LMW subunits with MENSHIP and METSCIP N-terminal sequence were only isolated in Sv 1BL.1RS 1AL.1DScs and not found in the other genotypes.

The alignment performed with each domain of all deduced proteins highlighted the highest variability of the REP domain compared to other domains among all genotypes; whereas the C-TERM domain was the most conserved. For each domain the pairwise ID of the alignment performed with sequences from in pairs between genotypes Sv 1BL.1RS 1AL.1DScs and Sv 1BL.1RS 1AL.1DScnn was lower than sequence alignment between other genotypes (Fig. 5.13).

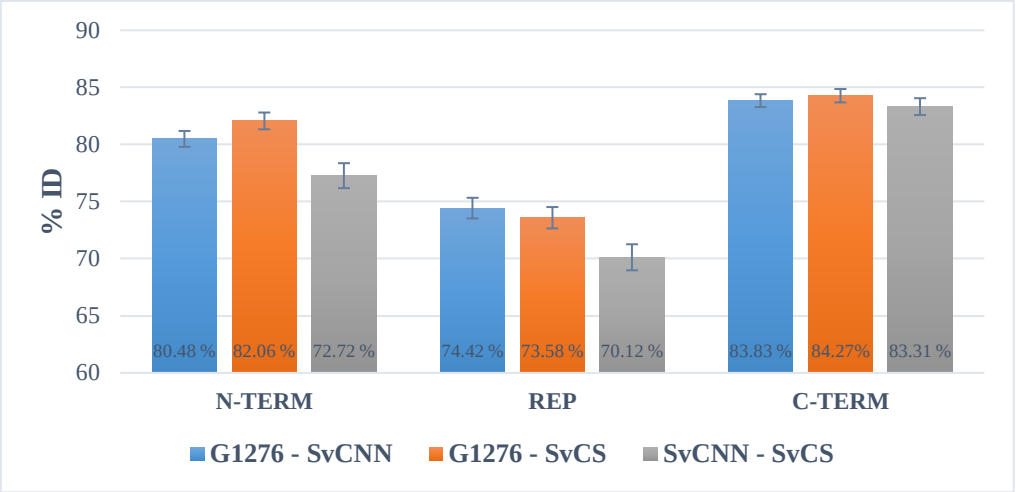

Figure 5.13: Pairwise identity of amino acid domains of LMW-GS isolated from three genotypes. *SvCNN* = Sv 1BL.1RS 1AL.1DScnn; SvCS = Sv 1BL.1RS 1AL.1DScs; G1276 = *Aegilops tauschii* spp. *strangulata*

Regarding the number and positions of cysteine residues, eighteen deduced proteins contained eight cysteine residues, as reported for all the most LMW-GS, of which twelve sequences in N-TERM domain and seven sequences contain the first residue in the C-TERM domain. Furthermore, three sequences contained seven cysteine residues lacking either the first one or the last one.

5.4 CONCLUSION

Several events of allopolyploidization between *T. turgidum* spp *dicoccum* e *Aegilops tauschii* spp. *strangulata* led to hexaploid bread wheat (*T.aestivum*) (Feldman 2001). *Ae. tauschii* is categorized into two subspecies, ssp. *tauschii*, with elongated cylindrical spikes, and ssp. *strangulata*, with quadrate spikes (Hammer 1980, Huo et al 2018). Several hybridizations occurred among haplotypes of the two subspecies, giving rise to intermediate forms and to different haplotypes (Shen et al 2018, Huo et al 2018). *Ae. tauschii* is widely recognized as the D genome donor of the hexaploid wheats (McFadden and Sears 1946) and the subspecies *strangulata* is considered the main candidate in hybridization with *T. turgidum* spp *dicoccum*, in particular the haplotypes founded in southeastern Caspian Iran (such as the G1276 accession) (Shen et al 2018).

The genetic relationships between the D genome of *Ae. tauschii* and *T. aestivum* are one of the fundamental aspects to better understand the evolution of modern bread wheat. The gluten proteins, in particular gliadins and glutenins encoded by the genes located at *Glu-D3* and *Gli-D1* on

chromosome 1D, allow to investigate the phylogenetic relationship existing between *Ae. tauschii* subspecies and *T.aestivum* ssp. *aestivum*.

Masci et al (1991) identified the Chinese Spring-type (CS-type) and the Cheyenne-type (CNN-type) as the two main types of proteins associated with the *Gli-D1* locus, among modern bread wheat genotypes. The CS-type had electrophoretic patterns of ω- and γ-gliadins very similar to cv Chinese Spring, whereas CNN-type had electrophoretic patterns overlapping to cv Cheyenne (Fig. 5.1).

Here, genes encoding ω-, γ- gliadins and LMW-GS were isolated from *Aegilops tauschii* spp. *strangulata* (G1276) and from *Sv* 1BL.1RS 1AL.1DS$_{cs}$ and *Sv* 1BL.1RS 1AL.1DS$_{cnn}$ lines. The two translocation lines *Sv* 1BL.1RS 1AL.1DS$_{cs}$ and *Sv* 1BL.1RS 1AL.1DS$_{cnn}$ lack the genes located at *Gli-1*, *Glu-3* and minor loci on the short arm of 1A and 1B chromosomes and contain 1DS arm from cultivar Chinese Spring or cultivar Cheyenne, respectively. These allowed the investigation of the genetic relationship of *Gli-D1* and *Glu-D3* loci among CS-type, CNN-type and *Ae. tauschii* ssp. *strangulata.*

The electrophoretic patterns of gliadins, analysed in A-PAGE, showed the distinction between the CS-type and CNN-type wheats and highlighted the similarity between the G1276 accession and CNN-type, especially for ω- gliadins. Furthermore, different genotypes and their deletion and substitution lines has made possible to assign specific *Gli-1* loci to many protein bands (Fig. 5.4).

The same trend was observed in the SDS-PAGE electrophoretic patterns of LMW-GS of the analysed wheat genotypes and allow to assign many proteins bands to *Glu-3* loci (Fig. 5.5)

The ARQ and ARE types were isolated from *Aegilops tauschii* spp. *strangulata* and from *Sv* 1BL.1RS 1AL.1DS$_{cnn}$ whereas no ARQ-type were found in *Sv* 1BL.1RS 1AL.1DS$_{cs}$. According to Kasarda et al. (1983) the *Gli-B1* encodes for SRL type, thus no sequences corresponding to SRL type were isolated from *Aegilops tauschii* spp. *strangulata*, *Sv* 1BL.1RS 1AL.1DS$_{cnn}$ and *Sv* 1BL.1RS 1AL.1DS$_{cs}$ lines. The pairwise identity (92.7%) among all deduced protein indicates the high phylogenetic relationships between *Aegilops tauschii* spp. *strangulata* and the D genome of bread wheat. *Sv* 1BL.1RS 1AL.1DS$_{cnn}$ line. Interestingly, only one AREL type was obtained from *Sv* 1BL.1RS 1AL.1DS$_{cs}$, which probably could be processed to KEL type by an asparaginyl endopeptidase, since it contains the NK (positions 8-9) cleavage site (Dupont et al 2004). On the other hand, the same number of genes was isolated from *Aegilops tauschii* spp. *strangulata* and *Sv* 1BL.1RS 1AL.1DS$_{cnn}$; the deduced proteins Sv_CNN_AREL_1 and Sv_CNN_AREL_4 are identical to *Ae. tauschii* _AREL_3 and *Ae. tauschii*_AREL_2, respectively, and the Sv_CNN_ARQ_1 showed only two different amino acid residues compared to *Ae. tauschii*_ARQ. All the deduced proteins lacked cysteine residues, but Sv_CNN_AREL_1, Sv_CNN_AREL_5 and *Ae. tauschii*_ARQ deduced proteins showed a methionine residue in R domain.

As regard to γ-gliadins, the deduced proteins obtained from *Sv* 1BL.1RS 1AL.1DS$_{cnn}$ line showed a pairwise ID with *Aegilops tauschii* deduced proteins higher compared to those obtained from *Sv* 1BL.1RS 1AL.1DS$_{cs}$. Interestingly, the REP domain of deduced proteins from *Sv* 1BL.1RS 1AL.1DS$_{cnn}$ and from *Aegilops tauschii* was similar in length, whereas the REP domain of *Sv* 1BL.1RS 1AL.1DS$_{cs}$ deduced proteins was the longest due to the presence of an extra peptide repetion (QPQQTFP). Proteins with an additional cysteine residue in REP domain were identified in all genotypes, probably belonging to the C group of LMW_GS (D'Ovidio et al. 1995; Masci et al. 2002).

According to Ikeda et al (2006), all the LMW-GS genes isolated from three genotypes encoded for LMW-m type and only one gene encoding for LMW-s type was found in *Sv* 1BL.1RS 1AL.1DS$_{cs}$.

The pairwise identity of *Sv* 1BL.1RS 1AL.1DS$_{cs}$ and *Sv* 1BL.1RS 1AL.1DS$_{cnn}$ LMW-GS deduced proteins with *Aegilops tauschii* LMW-GS deduced proteins was almost equal whereas the ID between *Sv* 1BL.1RS 1AL.1DS$_{cs}$ and *Sv* 1BL.1RS 1AL.1DS$_{cnn}$ deduced proteins was the lowest.

Finally, the comparison among deduced proteins obtained from *Aegilops tauschii, Sv* 1BL.1RS 1AL.1DS$_{cs}$ and *Sv* 1BL.1RS 1AL.1DS$_{cnn}$ lines encoded by the genes located at *Gli-1* and *Glu-3* loci highlighted a closer correlation between CNN-type wheat genotypes and the *Aegilops tauschii* ssp. *strangulata,* confirming the hypothesis that the hybridization took place between ssp. *strangulata* and *T. turgidum* spp *dicoccum.* These data will be validated by mass spectrometry of polyacrylamide proteins bands corresponding to genes located at *Gli-1* and *Glu-3* loci.

These also confirm that allohexaploidization has led to a bottleneck because few *Ae. tauschii* accessions participated in the events of hybridization with *T. turgidum* spp *dicoccum* (Dubcovsky and Dvorak 2007; Dvorak et al. 1998 from Luo et al 2018).

Thus, a better understanding of the genetic resource represented by several *Ae. tauschii* accessions will allow to increase genetic variability in the modern bread wheat, useful for breeding programs.

REFERENCES

Aghaei, M. J., Mozafari, J., Taleei, A. R., Naghavi, M. R., & Omidi, M. (2008). Distribution and diversity of Aegilops tauschii in Iran. *Genetic resources and crop evolution*, *55*(3), 341.

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(1), 242.

Anderson, O. D., & Greene, F. C. (1997). The α-gliadin gene family. II. DNA and protein sequence variation, subfamily structure, and origins of pseudogenes. *Theoretical and applied genetics*, *95*(1-2), 59-65.

Arora, S., Singh, N., Kaur, S., Bains, N. S., Uauy, C., Poland, J., & Chhuneja, P. (2017). Genome-wide association study of grain architecture in wild wheat Aegilops tauschii. *Frontiers in plant science*, *8*, 886.

Caldwell, K. S., Dvorak, J., Lagudah, E. S., Akhunov, E., Luo, M. C., Wolters, P., & Powell, W. (2004). Sequence polymorphism in polyploid wheat and their D-genome diploid ancestor. *Genetics*, *167*(2), 941-947.

Camerlengo, F., Sestili, F., Silvestri, M., Colaprico, G., Margiotta, B., Ruggeri, R., ... & Lafiandra, D. (2017). Production and molecular characterization of bread wheat lines with reduced amount of α-type gliadins. *BMC plant biology*, *17*(1), 248.

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*(6-7), 743-750.

Chen, F., Yang, L., Zhao, F., Min, H., & Xia, G. (2011). Molecular cloning and variation of ω-gliadin genes from a somatic hybrid introgression line II-12 and parents (Triticum aestivum cv. Jinan 177 and Agropyron elongatum). *Journal of genetics*, *90*(1), 137-142.

Cox, T. S. (1997). Deepening the wheat gene pool. *Journal of Crop Production*, *1*(1), 1-25.

Dong, L., Zhang, X., Liu, D., Fan, H., Sun, J., Zhang, Z., ... & Wang, D. (2010). New insights into the organization, recombination, expression and functional mechanism of low molecular weight glutenin subunit genes in bread wheat. *PLoS One*, *5*(10), e13548.

D'Ovidio, R., & Masci, S. (2004). The low-molecular-weight glutenin subunits of wheat gluten. *Journal of cereal science*, *39*(3), 321-339.

D'Ovidio, R., Simeone, M., Masci, S., Porceddu, E., & Kasarda, D. D. (1995). Nucleotide sequence of a gamma-type glutenin gene from a durum wheat: correlation with a gamma-type glutenin subunit from the same biotype. *Cereal chemistry (USA)*.

Du, X., Zhao, J., Ma, X., Yin, H., Wang, H., Li, A., & Kong, L. (2015). Cloning and characterization of novel fast ω-gliadin genes in Triticum monococcum. *Journal of genetics*, *94*(2), 323-327.

Dudnikov, A. J. (1998). Allozyme variation in Transcaucasian populations of Aegilops squarrosa. *Heredity*, *80*(2), 248.

DuPont, F. M., Vensel, W., Encarnacao, T., Chan, R., & Kasarda, D. D. (2004). Similarities of omega gliadins from Triticum urartu to those encoded on chromosome 1A of hexaploid wheat and evidence for their post-translational processing. *Theoretical and applied genetics*, *108*(7), 1299-1308.

Dvorak, J., Luo, M. C., Yang, Z. L., & Zhang, H. B. (1998). The structure of the Aegilops tauschii genepool and the evolution of hexaploid wheat. *Theoretical and Applied Genetics*, *97*(4), 657-670.

Egidi, E., Sestili, F., Janni, M., D'Ovidio, R., Lafiandra, D., Ceriotti, A., ... & Masci, S. (2014). An asparagine residue at the N-terminus affects the maturation process of low molecular weight glutenin subunits of wheat endosperm. *BMC plant biology*, *14*(1), 64.

Feldman, M., 2001 The origin of cultivated wheat, pp. 1–56. In The Wheat Book, edited by A. Bonjean, and W. Angus. Lavoisier Tech & Doc, Paris.

Gao, S., Gu, Y. Q., Wu, J., Coleman-Derr, D., Huo, N., Crossman, C., ... & Kong, X. (2007). Rapid evolution and complex structural organization in genomic regions harboring multiple prolamin genes in the polyploid wheat genome. *Plant molecular biology*, *65*(1-2), 189.

Gupta, R. B., & Shepherd, K. W. (1990). Two-step one-dimensional SDS-PAGE analysis of LMW subunits of glutelin. *Theoretical and Applied Genetics*, *80*(1), 65-74.

Hammer, K. (1980). Vorarbeiten zur monographischen darstellung von wildpflanzensortimenten: Aegilops L. *Die Kulturpflanze*, *28*(1), 33-180.

Harberd, N. P., Bartels, D., & Thompson, R. D. (1985). Analysis of the gliadin multigene loci in bread wheat using nullisomic-tetrasomic lines. *Molecular and General Genetics MGG*, *198*(2), 234-242.

Ibba, M. I., Kiszonas, A. M., & Morris, C. F. (2017)a. Evidence of intralocus recombination at the Glu-3 loci in bread wheat (Triticum aestivum L.). *Theoretical and applied genetics*, *130*(5), 891-902.

Ibba, M. I., Kiszonas, A. M., & Morris, C. F. (2018). Development of haplotype-specific molecular markers for the low-molecular-weight glutenin subunits. *Molecular breeding*, *38*(5), 68.

Ibba, M. I., Kiszonas, A. M., Guzmán, C., & Morris, C. F. (2017)c. Definition of the low molecular weight glutenin subunit gene family members in a set of standard bread wheat (Triticum aestivum L.) varieties. *Journal of cereal science*, *74*, 263-271.

Ikeda, T. M., 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*(2), 327-334.

Ikeda, T. M., Nagamine, T., Fukuoka, H., & Yano, H. (2002). Identification of new low-molecular-weight glutenin subunit genes in wheat. *Theoretical and Applied Genetics*, *104*(4), 680-687.

Jackson, E. A., Holt, L. M., & Payne, P. I. (1983). Characterisation of high molecular weight gliadin and low-molecular-weight glutenin subunits of wheat endosperm by two-dimensional electrophoresis and the chromosomal localisation of their controlling genes. *Theoretical and Applied Genetics*, *66*(1), 29-37.

Johal, J., Gianibelli, M. C., Rahman, S., Morell, M. K., & Gale, K. R. (2004). Characterization of low-molecular-weight glutenin genes in Aegilops tauschii. *Theoretical and Applied Genetics*, *109*(5), 1028-1040.

Kasarda, D. D. (1989). Glutenin structure in relation to wheat quality. *Wheat is unique*, *277*, 302.

Kasarda, D. D., Autran, J. C., Lew, E. J. L., Nimmo, C. C., & Shewry, P. R. (1983). N-terminal amino acid sequences of ω-gliadins and ω-secalins: Implications for the evolution of prolamin genes. *Biochimica et Biophysica Acta (BBA)-Protein Structure and Molecular Enzymology*, *747*(1-2), 138-150.

Kihara, H. (1965). The origin of wheat in the light of comparative genetics. *The Japanese Journal of Genetics*, *40*(1), 45-54.

Liu, L., Ikeda, T. M., Branlard, G., Peña, R. J., Rogers, W. J., Lerner, S. E., ... & Appels, R. (2010). Comparison of low molecular weight glutenin subunits identified by SDS-PAGE, 2-DE, MALDI-TOF-MS and PCR in common wheat. *BMC plant biology*, *10*(1), 124.

Masci, S. M., Porceddu, E., Colaprico, G., & Lafiandra, D. (1991). Comparison of the B and D subunits of glutenin encoded at the Glu-D3 locus in two biotypes of the common wheat cultivar Newton with different technological characteristics. *Journal of Cereal Science*, *14*(1), 35-46.

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*(1), 17-25.

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 (USA)*.

Masci, S., Rovelli, L., Kasarda, D. D., Vensel, W. H., & Lafiandra, D. (2002). Characterisation and chromosomal localisation of C-type low-molecular-weight glutenin subunits in the bread wheat cultivar Chinese Spring. *Theoretical and Applied Genetics*, *104*(2-3), 422-428.

Matsuoka, Y., Takumi, S., & Kawahara, T. (2008). Flowering time diversification and dispersal in central Eurasian wild wheat Aegilops tauschii Coss.: genealogical and ecological framework. *PLoS One*, *3*(9), e3138.

McIntosh, R. A., Yamazaki, Y., Dubcovsky, J., Rogers, J., & Morris, C. (1973, August). A catalogue of gene symbols for wheat. In *Proc. 4th Int. Wheat Genet. Symp* (pp. 893-937).

Miranda, L. M., Murphy, J. P., Marshall, D., & Leath, S. (2006). Pm34: a new powdery mildew resistance gene transferred from Aegilops tauschii Coss. to common wheat (Triticum aestivum L.). *Theoretical and Applied Genetics*, *113*(8), 1497-1504.

Mirzaghaderi, G., & Mason, A. S. (2019). Broadening the bread wheat D genome. Theoretical and Applied Genetics, 1-13.

Mizuno, N., Yamasaki, M., Matsuoka, Y., Kawahara, T., & Takumi, S. (2010). Population structure of wild wheat D-genome progenitor Aegilops tauschii Coss.: implications for intraspecific lineage diversification and evolution of common wheat. *Molecular ecology*, *19*(5), 999-1013.

Payne, P. I. (1987). Genetics of wheat storage proteins and the effect of allelic variation on bread-making quality. *Annual Review of Plant Physiology*, *38*(1), 141-153.

Payne, P. I., & Corfield, K. G. (1979). Subunit composition of wheat glutenin proteins, isolated by gel filtration in a dissociating medium. *Planta*, *145*(1), 83-88.

Payne, P. I., Holt, L. M., Jarvis, M. G., & Jackson, E. A. (1985). Two-dimensional fractionation of the endosperm proteins of bread wheat (Triticum aestivum): biochemical and genetic studies. *Cereal Chem*, *62*(5), 319-326.

Payne, P. I., Jackson, E. A., Holt, L. M., & Law, C. N. (1984). Genetic linkage between endosperm storage protein genes on each of the short arms of chromosomes 1A and 1B in wheat. *Theoretical and Applied Genetics*, *67*(2-3), 235-243.

Piston, F., Dorado, G., Martin, A., & Barro, F. (2006). Cloning of nine γ-gliadin mRNAs (cDNAs) from wheat and the molecular characterization of comparative transcript levels of γ-gliadin subclasses. *Journal of cereal science*, *43*(1), 120-128.

Sabelli, P. A., & Shewry, P. R. (1991). Characterization and organization of gene families at the Gli-1 loci of bread and durum wheats by restriction fragment analysis. *Theoretical and Applied Genetics*, *83*(2), 209-216.

Scheets, K., & Hedgcoth, C. (1988). Nucleotide sequence of a γ gliadin gene: Comparisons with other γ gliadin sequences show the structure of γ gliadin genes and the general primary structure of γ gliadins. *Plant Science*, *57*(2), 141-150.

Shen, L., Luo, G., Song, Y., Song, S., Li, Y., Yang, W., ... & Zhang, A. (2018). Low molecular weight glutenin subunit gene composition at Glu-D3 loci of Aegilops tauschii and common wheat and a further view of wheat evolution. *Theoretical and Applied Genetics*, *131*(12), 2745-2763.

Shewry, P. R. (1994). The prolamin storage proteins of wheat and related cereals. *Prog. Biophys. Mol. Biol.*, *61*, 37-59.

Shewry, P. R., Halford, N. G., & Lafiandra, D. (2003). Genetics of wheat gluten proteins. In *Advances in genetics* (Vol. 49, pp. 111-184). Academic Press.

Singh, N. K., & Shepherd, K. W. (1988). Linkage mapping of genes controlling endosperm storage proteins in wheat. *Theoretical and Applied Genetics*, *75*(4), 628-641.

Sohail, Q., Shehzad, T., Kilian, A., Eltayeb, A. E., Tanaka, H., & Tsujimoto, H. (2012). Development of diversity array technology (DArT) markers for assessment of population structure and diversity in Aegilops tauschii. *Breeding science*, *62*(1), 38-45.

Trethowan, R. M., & Mujeeb-Kazi, A. (2008). Novel germplasm resources for improving environmental stress tolerance of hexaploid wheat. *Crop Science*, *48*(4), 1255-1265.

Van Slageren, MW, (1994). Wild Wheats: A Monograph of Aegilops L. and Amblyopyrum (Jaub. & Spach) Eig (Poaceae). Wageningen: Wageningen Agricultural University. 513 p.

Wang, J., Luo, M. C., Chen, Z., You, F. M., Wei, Y., Zheng, Y., & Dvorak, J. (2013). Aegilops tauschii single nucleotide polymorphisms shed light on the origins of wheat D-genome genetic diversity and pinpoint the geographic origin of hexaploid wheat. *New phytologist*, *198*(3), 925-937.

Watanabe, N., Fujii, Y., Takesada, N., & Martinek, P. (2006). Cytological and microsatellite mapping of the gene for brittle rachis in a Triticum aestivum-Aegilops tauschii introgression line. *Euphytica*, *151*(1), 63-69.

Zhang, X., Liu, D., Yang, W., Liu, K., Sun, J., Guo, X., ... & Zhang, A. (2011). Development of a new marker system for identifying the complex members of the low-molecular-weight glutenin subunit gene family in bread wheat (Triticum aestivum L.). *Theoretical and Applied Genetics*, *122*(8), 1503-1516.

Zhang, X., Liu, D., Zhang, J., Jiang, W., Luo, G., Yang, W., ... & Zhang, A. (2013). Novel insights into the composition, variation, organization, and expression of the low-molecular-weight glutenin subunit gene family in common wheat. *Journal of experimental botany*, *64*(7), 2027-2040.