



Allelic variation of gliadin-encoding genes in a collection of tetraploid wheat genotypes

Domenica Nigro¹ · Linda Monaci³ · Rosa Pilolli³ · Agata Gadaleta^{2,3}

Received: 3 December 2019 / Accepted: 24 July 2020 / Published online: 9 August 2020
© The Author(s) 2020

Abstract

Wheat is one of the main crops bred worldwide. Durum wheat, specifically, is a key element of the Mediterranean diet, representing an élite crop grown in Italy. Durum wheat nutritional and technological values are largely due to the grain protein content (GPC), a complex genetic trait strongly affected by environmental factors and management practices. In the last decades, several breeding programs have been focused on improving GPC by both traditional and innovative approaches. Among seed storage proteins, prolamins, including both gliadins and glutenins, represent the major component. These two classes of proteins are indeed responsible of gluten formation and confer the extensibility and elasticity to the dough. Besides being of crucial importance for both technological properties and rheological characteristics, prolamins, and especially gliadins, have been found to be major triggers for human health, as involved in a number of wheat consumption-related conditions, such as the celiac disease, non-celiac gluten sensitivity, defined as the onset of a variety of manifestations related to wheat, rye and barley ingestion, and wheat allergies, both due to wheat ingestion or inhalation (of flour or pollen). The identification of loci responsible for the gliadin expression, and particularly of polymorphism in the aforementioned genes, which could result in a lower immunogenic/toxic potential, could be of great importance in breeding programs. For this purpose, we screened a collection of tetraploid wheat genotypes for allelic variants of annotated gliadin genes in the durum wheat genome, in order to identify genetic resources available to breeders to improve wheat nutritional and technological properties. Phylogenetic analysis among different species of *Triticum* genus and an in silico expression data analysis may also be useful in the exploitation of the complex scenario of gliadin–glutenin interaction and gluten role in the adverse reactions due to wheat consumption.

Keywords Durum wheat · Gliadin genes · Genetic diversity · Allergens · SNPs

Communicated by International Grain Council, <https://www.igc.int/en/default.aspx>.

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s42976-020-00061-w>) contains supplementary material, which is available to authorized users.

✉ Domenica Nigro
domenica.nigro@uniba.it

¹ Department of Soil, Plant & Food Sciences, University of Bari Aldo Moro, Via G. Amendola, 165/a, 70126 Bari, Italy

² Department of Agricultural & Environmental Science, University of Bari Aldo Moro, Via G. Amendola, 165/a, 70126 Bari, Italy

³ Institute of Sciences of Food Production, National Research Council of Italy (ISPA-CNR), via Giovanni Amendola 122/O, 70126 Bari, Italy

Introduction

Wheat is one of the most widely grown crop worldwide, providing about 20% of the human daily protein intake. The majority of cultivated wheats belong either to the hexaploid *Triticum aestivum* L. (genomes AABBDD, bread wheat) or to the tetraploid *T. turgidum* L. var. *durum* (genomes AABB, durum wheat) species, the latter being an élite crop in drylands, such as the Mediterranean basin.

Grain protein content is one of the most important agronomic traits in wheat, as strictly related to both its nutritional and end-use properties. Breeding programs have been focusing on its improvement for decades, and a number of studies have been trying to better define its genetic control (Gadaleta et al. 2011; Nigro et al. 2014a, 2016, 2019).

Prolamins, which include both gliadins and glutenins, account for up to 80% of the total grain proteins. These

two classes of proteins, which interaction implies the gluten formation, determine the technological quality of flour and semolina, conferring extensibility and elasticity to the dough and allowing it to be processed in a variety of products (Shewry et al. 2002; Shewry 2009).

Despite their importance in contributing to specific technological properties, wheat proteins, and especially gliadins, represent a critical trigger for human health. A number of conditions have been related to wheat consumption, such as the celiac disease (CD), an autoimmune disorder which impacts the 5% of human population, but also wheat allergies (WA), and non-celiac gluten sensitivity (NCGS) (Sollid et al. 2012; Valenti et al. 2017). Recent investigations have been carried out on a set of tetraploid wheats by both genetic and proteomic/peptidomic approaches in order to identify wheat genotypes with reduced gluten content and low immunogenic/toxic potential (Pilolli et al. 2019a, b). Moreover, a first attempt to identify allelic variants for allergenic epitopes in a collection of durum wheat genotypes was reported by Nigro et al. (2014b).

As reported by Huo et al. (2018a), genomic localization, as well as gene structure, have been well characterized for both glutenins and gliadins genes, which are mainly located in three genomic regions, *Glu-3*, *Gli-1* and *Gli-2* (D'Ovidio and Masci 2004; Gu et al. 2004; Dong et al. 2016). Gliadins are subdivided into α , γ , ω and δ -gliadins, with the α -gliadins accounting for 15–30% of the total seed storage proteins in the wheat grain (Altenbach et al. 2011). Several studies focused on sequencing gliadin loci in bread wheat, wild emmer and progenitors *Triticum urartu* and *Aegilops tauschii* (Huo et al. 2017, 2018b, 2019; Zhang et al. 2015), suggested that gliadin gene families organization might actually be less complex than reported in the early studies (Sabelli and Shewry 1991; Anderson et al. 1997).

The recently released bread (IWGS 2018) and durum wheat genomes (Maccaferri et al. 2019) might be greatly helpful tools in unveiling the complexity of the gluten protein gene families, especially gliadins, as well as their allelic variations, which can strongly affect wheat functional properties and its immunogenic potential and have a huge impact on the development of healthier wheat foods.

In this perspective, we analyzed a collection of 240 tetraploid wheat genotypes for SNPs in gliadin-encoding genes in order to find natural allelic variants. The protein sequences of the identified gliadin genes in durum wheat were then included in a phylogenetic analysis along with their paralogues of bread wheat, wild emmer, *T. urartu* and *A. tauschii*. Also, we evaluated the gene expression in orthologues bread wheat Chinese Spring cv gliadins by analyzing in silico data of the considered genes. The identification of wheat genotypes with lower amounts of toxic epitopes, as well as the

less expressed gliadin genes, could be valuable information to be used in new breeding programs and in the development of new detoxification strategies.

Materials and methods

SNP discovery in gliadins genes in tetraploid wheat

Bread wheat gliadin gene sequences (Juhász et al. 2018; Huo et al. 2018a, b) were used as queries and launched against the publicly available *T. turgidum* L. var. *durum* genome browser (<https://d-gbrowse.interomics.eu/>), and only regions annotated as coding genes were considered for further analysis. The retrieved gene sequences were blasted against the available dataset of SNP marker sequences (Akhunov et al. 2009; Wang et al. 2014), as previously reported in other studies (Nigro et al. 2017a, b, 2019), and the ones aligned with at least 90% identity were considered as markers within the coding sequences of the gliadin genes. Additionally, SNPs markers were investigated and confirmed to locate within gene sequences by searching the variants table section reported for each gene in Ensembl Plants database (<http://plants.ensembl.org/>). The identified SNP markers were then screened in a collection of 240 accessions of tetraploid wheat genotypes (*Triticum turgidum* L., $2n=4x=28$; AABB genome) (Laidò et al. 2013). The collection was recently phenotyped for grain protein content (GPC) and genotyped for SNPs with the wheat 90 K Infinium iSelect array containing 81,587 gene-associated SNP markers, as reported by Nigro et al. (2017a, b, 2019).

Phylogenetic analysis of gliadin genes

Gliadins' aminoacidic sequences retrieved for bread wheat, durum wheat, wild emmer (*Triticum turgidum* ssp. *dicoccoides*), *A. tauschii* and *T. urartu* genomes (<http://plants.ensembl.org/>), were aligned by using the ClustalW method via Mega7 software. Phylogenetic analysis was carried out on a total of 111 sequences using the maximum likelihood method. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using a JTT model, and then selecting the topology with superior log likelihood value. The tree was drawn to scale, with branch lengths measured in the number of substitutions per site. The tree was generated with Mega7 (Kumar et al. 2016) (<http://www.ebi.ac.uk/Tools/phylogeny/>) and modified with the FigTree program (<http://tree.bio.ed.ac.uk/software/figtree/>).

Alpha-gliadins genes in silico expression analysis

Bread wheat gliadin genes identified by Juhász et al. (2018) and Huo et al. (2018a, b) were investigated for their in silico expression, in order to identify an eventual differential expression among all of them.

Each gene was used as a query in the Expression Atlas website (<https://www.ebi.ac.uk/gxa/home>). All experiment were screened, and specifically: four experiments reporting data of RNA-seq by organism part (Pfeifer et al. 2014); three RNA-seq experiments based on different developmental

stage (Gillies et al. 2012; Pingault et al. 2015) and one RNA-seq experiment on different grain developmental stages (Li et al. 2013).

Results

SNPs identification in gliadin-encoding genes

A total of 36 gliadin genes were retrieved from the durum wheat genome browser (<https://d-gbrowse.interomics.eu/>),

Table 1 Gliadin genes IDs retrieved by sequence search in the durum wheat genome browser

Gene ID	Gliadin type	Chr.	Start (bp)	End (bp)	Confidence	Strand
<i>TRITD1Av1G001620</i>	Omega-gliadin	1A	3751615	3832103	LC	–
<i>TRITD1Av1G001890</i>	Omega-gliadin	1A	4352939	4501926	LC	+
<i>TRITD1Av1G002070</i>	Gamma-gliadin	1A	4613285	4614202	HC	+
<i>TRITD1Av1G002120</i>	Gamma-gliadin	1A	4687840	4688874	HC	–
<i>TRITD1Av1G002200</i>	Gamma-gliadin	1A	4871918	4880492	HC	+
<i>TRITD1Av1G002230</i>	Gamma-gliadin	1A	4891405	4892007	HC	+
<i>TRITD1Bv1G001870</i>	Gamma-gliadin	1B	4313476	4519819	HC	+
<i>TRITD1Bv1G001890</i>	Gamma-gliadin	1B	4342064	4342690	LC	+
<i>TRITD1Bv1G001950</i>	Gamma-gliadin	1B	4543310	4543921	HC	+
<i>TRITD1Bv1G001990</i>	Omega-gliadin	1B	4673054	4726968	LC	–
<i>TRITD6Av1G009270</i>	Alpha-gliadin	6A	21959989	21960759	LC	+
<i>TRITD6Av1G009340</i>	Alpha-gliadin	6A	22234850	22235716	HC	+
<i>TRITD6Av1G009360</i>	Alpha-gliadin	6A	22490899	22499261	HC	+
<i>TRITD6Av1G009400</i>	Alpha-gliadin	6A	22582712	22793179	HC	–
<i>TRITD6Av1G009480</i>	Alpha-gliadin	6A	22881581	22948743	LC	–
<i>TRITD6Av1G009500</i>	Alpha-gliadin	6A	22987983	23197938	LC	–
<i>TRITD6Av1G009530</i>	Alpha-gliadin	6A	23061524	23455413	LC	–
<i>TRITD6Av1G009640</i>	Alpha-gliadin	6A	23510267	23510497	LC	–
<i>TRITD6Av1G009660</i>	Alpha-gliadin	6A	23634658	24077830	HC	+
<i>TRITD6Av1G009670</i>	Alpha-gliadin	6A	23689789	23717916	HC	+
<i>TRITD6Av1G009830</i>	Alpha-gliadin	6A	24117681	24118538	HC	+
<i>TRITD6Av1G009860</i>	Alpha-gliadin	6A	24180866	24223897	HC	+
<i>TRITD6Av1G009880</i>	Alpha-gliadin	6A	24341990	24342853	HC	+
<i>TRITD6Bv1G014610</i>	Alpha-gliadin	6B	40750267	40922395	HC	+
<i>TRITD6Bv1G014670</i>	Alpha-gliadin	6B	41175187	41324732	HC	+
<i>TRITD6Bv1G014690</i>	Alpha-gliadin	6B	41452648	41453531	LC	+
<i>TRITD6Bv1G020780</i>	Alpha-gliadin	6B	58542581	58547368	HC	–
<i>TRITD0Uv1G032980</i>	Gamma-gliadin	Un	78424750	78425607	HC	–
<i>TRITD0Uv1G033010</i>	Gliadin-like avenin	Un	78490494	78491336	HC	–
<i>TRITD0Uv1G033050</i>	Gamma-gliadin	Un	78502738	78503256	HC	–
<i>TRITD0Uv1G071610</i>	Gamma-gliadin	Un	186394373	186394879	HC	+
<i>TRITD0Uv1G078340</i>	Alpha-gliadin	Un	212199227	212200107	HC	+
<i>TRITD0Uv1G108590</i>	Alpha-gliadin	Un	330545238	330546146	HC	+
<i>TRITD0Uv1G114050</i>	Alpha-gliadin	Un	346692754	346693617	HC	+
<i>TRITD0Uv1G127200</i>	Alpha-gliadin	Un	389805324	389806011	LC	–
<i>TRITD0Uv1G130960</i>	Alpha-gliadin	Un	405992233	405993108	HC	–

The gene span position is reported, along with their physical localization, confidence level and coding strand

25 of which were annotated as high confidence genes. In Table 1, gene IDs, functional annotation, physical map position, level of confidence as well as the strand of each gene are reported. As showed, 19 sequences belonged to the A genome, 8 to the B genome and the remaining ones were unmapped. Also, only genes within alpha- and gamma-gliadins groups were annotated as high confidence, while all the omega gliadins genes were annotated as low confidence ones. Alpha-gliadin genes were the most represented ones. Out of 25 identified 25 genes, 13 mapped on 6a chromosome, 8 of which as high confidence genes, while 4 were located on homoeologues 6B, one of which as low confidence gene. Nine additionally alpha-gliadin genes were identified and unmapped, even though only 4 of them were HC genes. All the 36 gene sequences were used as queries in a BLASTn with the whole set of wheat SNPs of the 90 K iSelect Illumina chip, allowing the identification of 30 SNPs markers within gene sequences. Specifically, 13 SNPs were identified in the genomic sequence of gliadin genes

on the short arm of homoeologous chromosome group 6, while the remaining SNPs are located on the short arm of homoeologous chromosome group 1. Sixteen of them were monomorphic among the 240 genotypes, one resulted to be failed, while the other thirteen showed different levels of polymorphism (Table 2). The minor allele frequency (MAF) ranged from 1.25% for IWB11263 to 32.92% for IWB65373, both located in gamma-gliadin genes. The 30 SNPs markers sequences aligned with a total of 59 genomic regions on 1S and 6S homoeologous chromosomes groups on durum wheat genome, probably due to the gene evolution process and the presence of a high number of genes in the *Gli* loci, as well as the high number of transcript for each gene. In Table S1, all the aligned regions and information retrieved from Svevo browser (<https://d-gbrowse.interomics.eu/>) are reported. Sixteen of the identified 59 regions corresponded to coding regions, while the remaining were located in introns. Interestingly, SNP marker IWB63758 aligned with *TRITD6Bv1G014610* gene in four different positions, three

Table 2 SNP markers identified by blasting the gliadin gene sequences to the durum wheat genome, along with their eventual polymorphism in a collection of 240 tetraploid wheat genotypes and the minimum allele frequency (MAF %) reported for the polymorphic ones

Gene family	SNP name	SNP ID	SNP	MAF %
Gamma-gliadin	GENE-3593_37	IWB33486	C	–
Gamma-gliadin	RAC875_c41240_73	IWB57694	T	–
Gamma-gliadin	Tdurum_contig42344_2003	IWB71090	C	–
Gamma-gliadin	tplb0046h22_935	IWB74870	A	–
Gamma-gliadin	BS00082565_51	IWB11262	T/C	3.75
Gamma-gliadin	BS00082566_51	IWB11263	T/C	1.25
Gamma-gliadin	BS00082567_51	IWB11264	A/G	11.25
Gamma-gliadin	Excalibur_c42979_666	IWB26397	A/G	2.5
Gamma-gliadin	TA001286-0611-w	IWB65373	T/C	32.92
Gamma-gliadin	BS00105383_51	IWB12241	G	–
Gamma-gliadin	Excalibur_c58300_540	IWB27797	G	–
Gamma-gliadin	Excalibur_c58300_564	IWB27798	T	–
Gamma-gliadin	Excalibur_c85221_345	IWB29135	C	–
Gamma-gliadin	Kukri_c19709_407	IWB42185	C	–
Gamma-gliadin	RFL_Contig1457_465	IWB63643	T	–
Gamma-gliadin	Tdurum_contig28703_113	IWB69596	C	–
Gamma-gliadin	Tdurum_contig28703_293	IWB69597	T/G	25
Alpha-gliadin	BS00066207_51	IWB9743	G	–
Alpha-gliadin	Kukri_c78263_217	IWB47752	C	–
Alpha-gliadin	Kukri_c78263_296	IWB47753	C	–
Alpha-gliadin	Tdurum_contig4885_588	IWB71933	Failed	–
Alpha-gliadin	Tdurum_contig4885_1307	IWB71927	A/G	7.92
Alpha-gliadin	Tdurum_contig4885_1536	IWB71928	T/C	22.08
Alpha-gliadin	Tdurum_contig4885_1870	IWB71930	A/G	22.5
Alpha-gliadin	Tdurum_contig4885_843	IWB71934	A/G	10
Alpha-gliadin	RFL_Contig1877_207	IWB63758	A/C	31.67
Alpha-gliadin	CAP11_c1455_128	IWB12688	C	–
Alpha-gliadin	IACX557	IWB35996	T/C	22.5
Alpha-gliadin	Kukri_c4630_113	IWB45640	A	–
Alpha-gliadin	wsnp_BF482960B-Ta_1_4	IWB75679	T/G	12.5

of which corresponding to exon regions. SNP IWB63758-6B.1 was found to align to 16 different alternative mRNAs of gene *TRITD6Bv1G014610* (*TRITD6Bv1G014610.5–TRITD6Bv1G014610.21*), which shows quite a complex structure, as 25 different splicing form were identified. SNP IWB63758-6B.3 was found to align to 2 other different alternative mRNAs of gene *TRITD6Bv1G014610* (*TRITD6Bv1G014610.24* and *TRITD6Bv1G014610.25*). The single nucleotide polymorphism determined a glutamine (Q)–lysine (K) aminoacid change at position 46 of the protein sequence.

SNP IWB63758-6B.4 aligned to gene *TRITD6Bv1G020780*, which also shows 9 different transcripts, and also in this gene, a non-synonymous glutamine–lysine aminoacid change was observed due to the SNP. Three SNP markers located in gamma-gliadin genes blasted with *TRITD1Av1G002230* identifying three different point mutations. Specifically, while IWB11262 corresponded to a synonymous substitution, two others, IWB11263 and IWB11264, caused two non-synonymous changes, a serine (S)–proline (P) and isoleucine (I)–valine (V), respectively.

Phylogenetic analysis

Orthologues gliadin gene sequences were retrieved for bread wheat (42 sequences), wild emmer (20), *A. tauschii* (21) and *T. urartu* (14) genomes from the Ensembl Plants database (<http://plants.ensembl.org/>).

These identified protein sequence, along with the 23 HC previously identified in durum wheat (Table 1), were checked for sequence structure and similarity. A total of 111 gene sequences located on the short arms of 1 and 6 chromosome groups were retained to build a phylogenetic tree comprising the five considered species, which were firstly aligned by using the ClustalW method via Mega7 software. Nine sequences were discarded, probably subjected to some evolutionary polymorphism such as TE insertion or deletion, and multiple alternative splicing events resulting in a number of different transcripts for the same gene. The evolutionary history was inferred by using the maximum likelihood method based on the JTT matrix-based model. The tree with the highest log likelihood (−26,965.5799) is shown in Fig. 1. The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. The analysis involved 111 aminoacid sequences. There were a total of 609 positions in the final dataset.

Two main clades were generated: the first one, grouped 19 sequences belonging to the short arm of chromosome group 6, annotated as alpha-gliadins. Therefore, their protein sequence was quite different from the other groups of gliadins reported in the tree. The same situation was observed for the second clade, which showed several differentiation nodes. As expected, all gliadins sequences belonging to 1S

chromosome groups, corresponding to gamma-gliadin genes grouped together, as well as the highest number of 6S genes (29) which shared the highest sequence similarity. The phylogenetic result reflected the complexity of gliadin gene family, showing the high number of speciation and duplication events the gene family underwent.

Differential expression of gliadin genes

Despite being the main components of seed storage proteins, alpha-gliadins are also among the main triggers for celiac disease and other wheat-related pathologies. As no similar experiments have been yet carried out on durum wheat, we exploited the close relation and synteny with bread wheat in order to detect eventual differential expression among known alpha-gliadins gene in seeds and in different experimental condition. To this purpose, an in silico analysis was carried out on 21 alpha-gliadin gene sequences belonging either to A and B genomes (reported in Juhász et al. 2018; Huo et al. 2018a, b) through by the Expression Atlas website (<https://www.ebi.ac.uk/gxa/home>). Expression level were reported in TPM (transcripts per million), a normalized measure of the proportion of transcripts in an RNA sample. TPM were calculated from the number of reads that mapped to each particular gene sequence taking into account the gene length (more reads might be produced from longer genes) and the sequencing depth (more reads might be produced from the sample that has been sequenced to a greater depth). All the experiment carried out on *Triticum aestivum* and available in the Atlas were considered (see Materials and method section). The first experiment (Pfeifer et al. 2014) considered the transcriptional profile in different tissues 10, 20 and 30 days post-anthesis (Fig. 2). Results showed that gene expression in highly variable among genes and cell tissues, but basically the alpha-gliadin genes are more expressed in the endosperm during all the seed development. Indeed, 10 days post-anthesis, expression was detectable only in the endosperm, with *TraesCS6A02G049400*, *TraesCS6A02G049100* and *TraesCS6B02G065856* showing the highest expression levels (Fig. 2a). At 20 and 30 days post-anthesis, the maximum gene expression was detected in endosperm, with lower expression was detected in the aleurone layer (Fig. 2b, c). This was not unexpected, considering that alpha-gliadins are among the most abundant seed storage proteins.

The second experiment reports RNA-seq data from different phenological phases. Gillies et al. (2012) reported RNA-seq analysis on central endosperm during seed development 6, 9 and 14 days post-anthesis. Interestingly, and in accordance with previous data, gliadin gene expression increased notably 9 days after anthesis and reached their highest values at 14 daa. *TraesCS6A02G049100* gene showed the highest expression (Figure S1a). Also, Pingault et al. (2015)

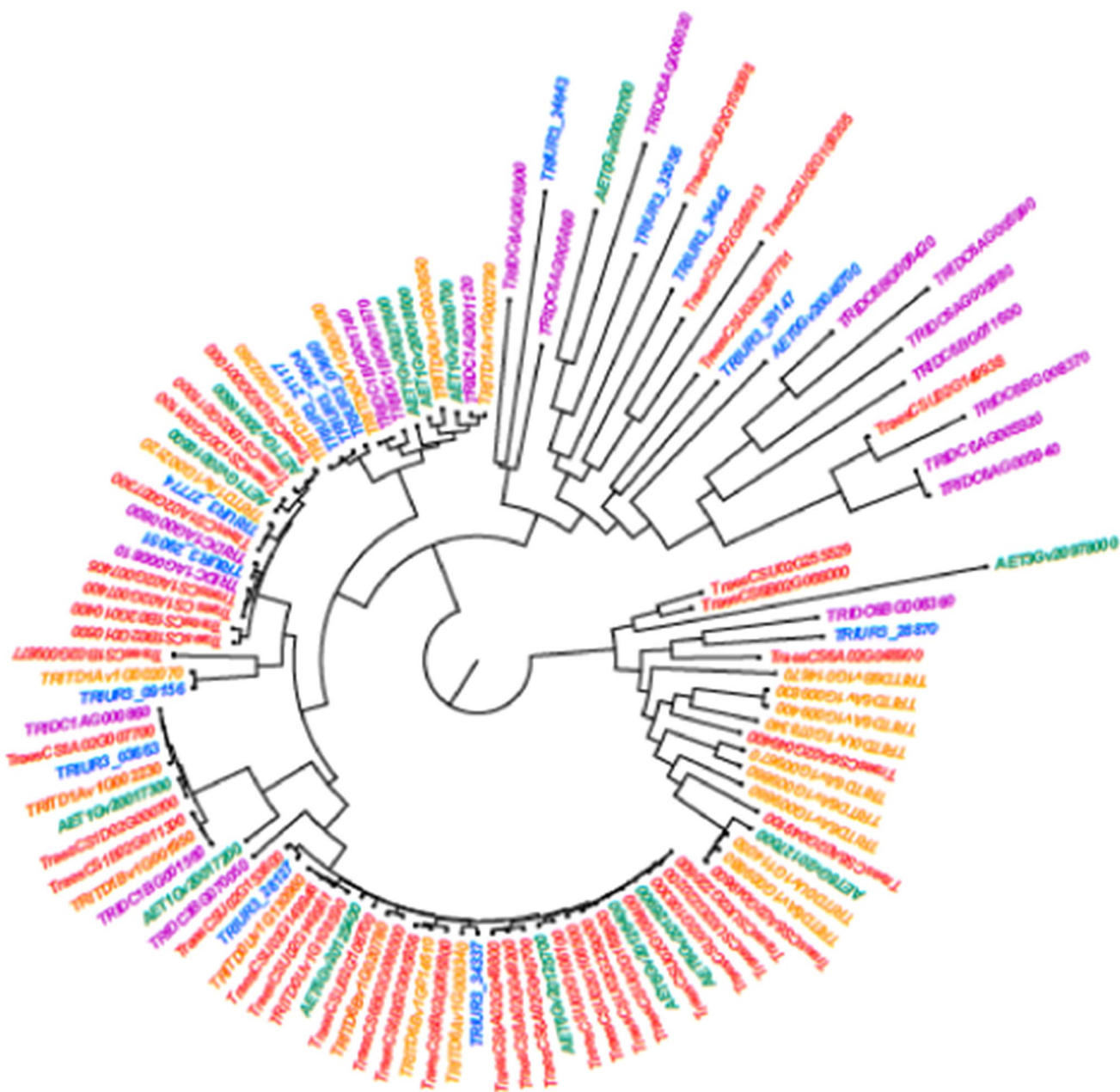


Fig. 1 Phylogenetic analysis on 111 gliadin gene sequences belonging to durum wheat, bread wheat, *A. tauschii*, *T. urartu* and *Triticum dicoccoides*, by using the maximum likelihood method based on the

JTT matrix-based model. Branch lengths are in the same units as those of the evolutionary distances used to infer the phylogenetic tree

analyzed fruit transcription profiling by high throughput sequencing of five different organs of wheat (leaf, shoot, root, spike and grain) at three developmental stages each (Figure S1b). The highest gene expression was observed when plants were between 70% of their fruit formation to final size, and, also in this case, *TraesCS6A02G049100* gene showed the highest expression. Before and after this phenological phase, no expression at all was detected for none of the considered gliadin genes.

The last considered experiment was carried out by Li et al. (2013), who run an RNA-seq experiment on grains at different developmental stages (Figure S1c). The highest gene expression was found in samples collected 12 days after pollination, and particularly, two genes showed very high expression level, both belonging to A genome gliadins, *TraesCS6A02G049100*, which has been shown the highest expression in all experiments, and *TRAESCS6A02G049800* gene. By comparison of the three experiments, two genes,

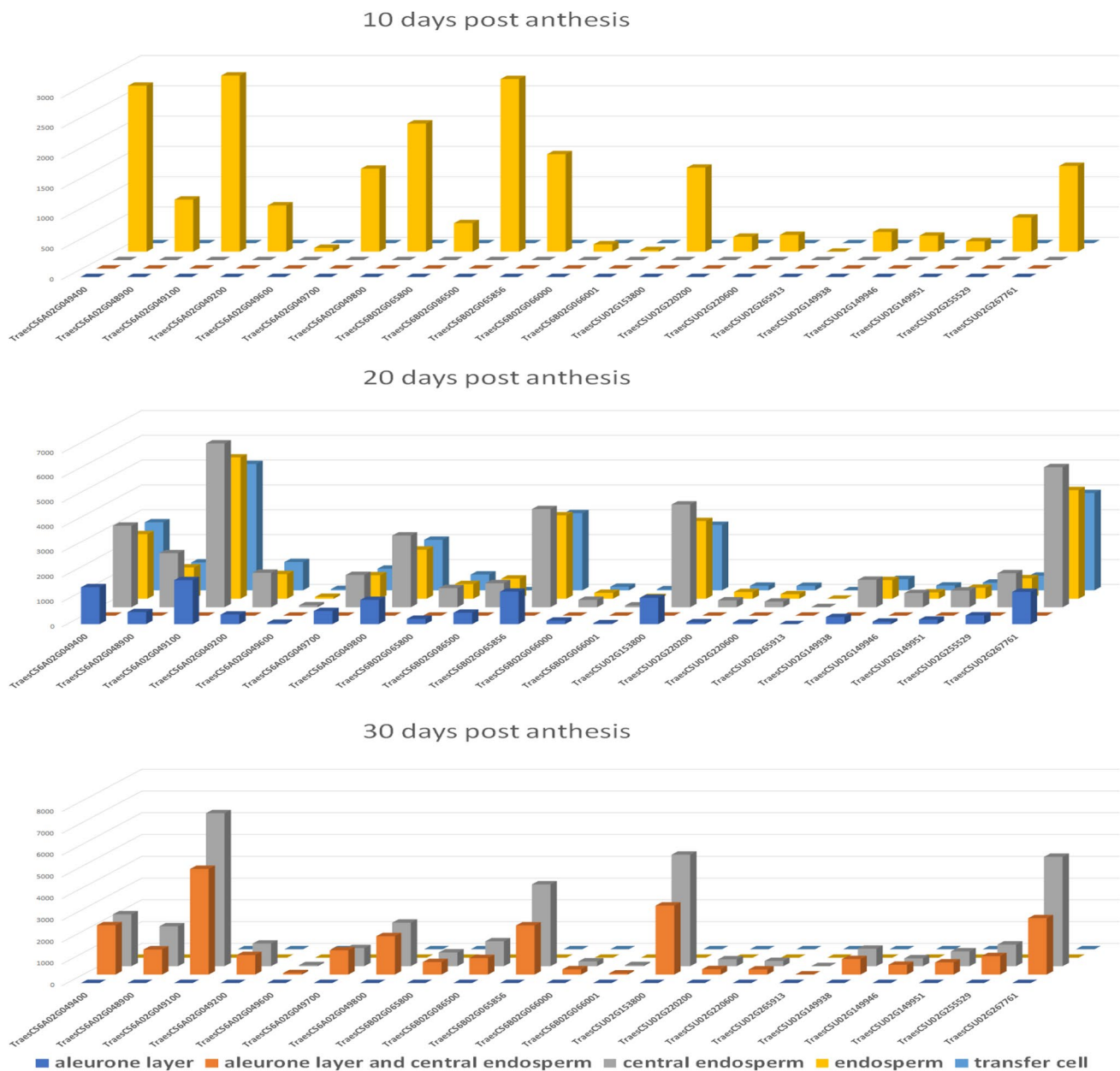


Fig. 2 Transcriptional profile of eight bread wheat gliadin genes in different cell types at: **a** 10, **b** 20 and **c** 30 days post-anthesis (experiment by Pfeifer et al. 2014)

TraesCS6B02G066001 and *TraesCSU02G265913*, were never expressed, suggesting these two genes might have undergone to pseudogenization events.

Discussion

Studies on gliadin genes have been carried out for bread wheat, wild emmer and the two progenitors of A and D genome, *T. urartu* and *A. tauschii*, respectively (Zhang et al. 2015; Huo et al. 2017, 2018b, 2019).

Despite being among the most effective seed storage protein on food processing properties, wheat gliadins are also a main trigger for health-related conditions associated with the consumption of its derived products (Shewry et al. 2002). So far, understanding gliadins structure, evolution and expression pattern, as well as allelic variation within their coding sequence, might be extremely important in order to focus on new breeding programs to develop wheat varieties with a reduced immunogenic potential.

In this study, we exploited the recent genome sequencing data of tetraploid durum wheat (Maccaferri et al.

2019), in order to assess the number of gliadin genes annotated in the genome, as well as SNPs polymorphism in a collection of tetraploid wheat genotypes. SNPs are nucleotide variations in the DNA sequence of individuals in a population and represent the class of molecular markers more frequent in the genome. Indeed, SNPs are widely distributed along the genomes, may vary in frequency and distribution among species and are usually prevalently but not exclusively located in the noncoding regions of the genome (Akhunov et al. 2009; Wang et al. 2014).

This work provides a new insight into the genetic diversity and allelic variation at gliadin genes in durum wheat. Indeed, a collection of 240 tetraploid wheat genotypes was screened for polymorphic SNPs in the gliadin gene sequences retrieved in the durum wheat genome browser, which allowed us to identify several SNPs within their genomic sequence. As expected, most of them were localized in intronic region, but nine of them were found to be located in exon regions. Moreover, four of them were also found to be polymorphic within the screened collection, resulting in one silent mutation and three non-synonymous aminoacidic change. Despite none of them seems to be located in known major CD epitopes, the approach we followed showed that the combination of recent genome sequencing and annotation data with SNP polymorphism analysis it could be a valid tool to quickly screen and identify gliadins allelic variation, as previously reported for other allergenic wheat epitopes (Nigro et al. 2014b). Indeed, more high throughput data needs to be generated and analyzed in order to detect useful information about point mutations for the development of new wheat varieties.

The analysis of in silico expression data run in different tissues and different phenological phases showed how highly variable the gliadin gene expression is. Due to high sequence similarity, it is quite complex to characterize each member of a large gene family, as well as their gene expression. Another important fact to be considered about gliadin gene family is the high pseudogenization rate in wheat. As we reported, some of the identified genes showed a low expression level in all the considered situations, probably due to the instability of their transcript, as also identified by Huo et al. (2018b, 2019) in other related species. A deeper analysis of all the identified durum wheat gene sequences will be required in order to establish the pseudogenization rate also in durum wheat, and identify the ones actually expressed. These information could be useful in further silencing and genome editing experiments to reduce their toxicity.

Acknowledgements Open access funding provided by Università degli Studi di Bari Aldo Moro within the CRUI-CARE Agreement.

Funding The research was funded by the Ministry of Education, Universities and Research (MIUR-Italy), program SIR 2014 within the project titled “S. Wheat Pro.—Proteomic characterization of Selected durum Wheat cultivars for PROduction of low toxicity-food products toward celiac disease patients (RBS114QQ1W)” and by “Intervento cofinanziato dal Fondo di Sviluppo e Coesione 2007–2013—APQ Ricerca Regione Puglia “Programma regionale a sostegno della specializzazione intelligente e della sostenibilità sociale ed ambientale—FutureInResearch.”

Compliance with ethical standards

Conflict of interest The authors declare that no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Akhunov E, Nicolet C, Dvorak J (2009) Single nucleotide polymorphism genotyping in polyploid wheat with the Illumina GoldenGate assay. *Theor Appl Genet* 119(3):507–517. <https://doi.org/10.1007/s00122-009-1059-5>
- Altenbach SB, Tanaka CK, Hurkman WJ, Whitehand LC, Vensel WH, Dupont FM (2011) Differential effects of a post-anthesis fertilizer regimen on the wheat flour proteome determined by quantitative 2-DE. *Proteome Sci* 9:46. <https://doi.org/10.1186/1477-5956-9-46>
- Anderson OD, Litts JC, Greene FC (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. *Theor Appl Genet* 95:50–58
- Dong L, Huo N, Wang Y, Deal K, Wang D, Hu T, Dvorak J, Anderson OD, Luo MC, Gu YQ (2016) Rapid evolutionary dynamics in a 2.8-Mb chromosomal region containing multiple prolamins and resistance gene families in *Aegilops tauschii*. *Plant J* 87:495–506. <https://doi.org/10.1111/tbj.13214>
- D’Ovidio R, Masci S (2004) The low-molecular-weight glutenin subunits of wheat gluten. *J Cereal Sci* 39:321–339. <https://doi.org/10.1016/j.jcs.2003.12.002>
- Gadaleta A, Nigro D, Giancaspro A, Blanco A (2011) The glutamine synthetase (*GS2*) genes in relation to grain protein content of durum wheat. *Funct Integr Genom* 11:665–670. <https://doi.org/10.1007/s10142-011-0235-2>
- Gillies SA, Futardo A, Henry RJ (2012) Gene expression in the developing aleurone and starchy endosperm of wheat. *Plant Biotechnol J* 10(6):668–679. <https://doi.org/10.1111/j.1467-7652.2012.00705.x>
- Gu YQ, Coleman-Derr D, Kong X, Anderson OD (2004) Rapid genome evolution revealed by comparative sequence analysis of orthologous regions from four triticeae genomes. *Plant Physiol* 135:459–470. <https://doi.org/10.1104/pp.103.038083>

- Huo N, Dong L, Zhang S, Wang Y, Zhu T, Mohr T, Altenbach S, Liu Z, Dvorak J, Anderson OD, Luo MC, Wang D, Gu YQ (2017) New insights into structural organization and gene duplication in a 1.75-Mb genomic region harboring the alpha-gliadin gene family in *Aegilops tauschii*, the source of wheat D genome. *Plant J* 92:571–583. <https://doi.org/10.1111/tpj.13675>
- Huo N, Zhang S, Zhu T, Dong L, Wang Y, Mohr T, Hu T, Liu Z, Dvorak J, Luo MC, Wang D, Lee JY, Altenbach S, Gu YQ (2018a) Gene duplication and evolution dynamics in the homeologous regions harboring multiple prolamin and resistance gene families in hexaploid wheat. *Front Plant Sci* 9:673. <https://doi.org/10.3389/fpls.2018.00673>
- Huo N, Zhu T, Altenbach S, Dong L, Wang Y, Mohr T, Liu Z, Dvorak J, Luo MC, Gu YQ (2018b) Dynamic evolution of alpha-gliadin prolamin gene family in homeologous genomes of hexaploid wheat. *Sci Rep* 8:5181. <https://doi.org/10.1038/s41598-018-23570-5>
- Huo N, Zhu T, Zhang S, Mohr T, Luo MC, Lee JY, Gu YQ (2019) Rapid evolution of α -gliadin gene family revealed by analyzing *Gli-2* locus regions of wild emmer wheat. *Funct Integr Genom* 19(6):993–1005. <https://doi.org/10.1007/s10142-019-00686-z>
- International Wheat Genome Sequencing Consortium (IWGSC) (2018) Shifting the limits in wheat research and breeding through a fully annotated and anchored reference genome sequence. *Science* 361:eaar7191
- Juhász A, Belova T, Florides CG, Maulis C, Fischer I, Gell G, Birinyi Z, Ong J, Keeble-Gagnère G, Maharajan A, Ma W, Gibson P, Jia J, Lang D, Mayer KFX, Spannagl M, International Wheat Genome Sequencing Consortium, Tye-Din JA, Appels R, Olsen OA (2018) Genome mapping of seed-borne allergens and immunoresponsive proteins in wheat. *Sci Adv* 4:eaar8602
- Kumar S, Stecher G, Tamura K (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol* 33(7):1870–1874
- Laidò G, Mangini G, Taranto F, Gadaleta A, Blanco A, Cattivelli L, Marone D, Mastrangelo AM, Papa R, De Vita P (2013) Genetic diversity and population structure of tetraploid wheats (*Triticum turgidum* L.) estimated by SSR, DArT and pedigree data. *PLoS ONE* 8(6):e67280. <https://doi.org/10.1371/journal.pone.0067280>
- Li HZ, Gao X, Li XY, Chen QJ, Dong J, Zhao WC (2013) Evaluation of assembly strategies using RNA-seq data associated with grain development of wheat (*Triticum aestivum* L.). *PLoS ONE* 8(12):e83530. <https://doi.org/10.1371/journal.pone.0083530>
- Maccaferri M, Harris NS, Twardziok SO, Pasam RK, Gundlach H, Spannagl M, Ormanbekova D, Lux T, Prade VM, Milner SG, Himmelbach A, Cattivelli L (2019) Durum wheat genome highlights past domestication signatures and future improvement targets. *Nat Genet* 51(5):885. <https://doi.org/10.1038/s41588-019-0381-3>
- Nigro D, Blanco A, Anderson OD, Gadaleta A (2014a) Characterization of ferredoxin-dependent glutamine-oxoglutarate amidotransferase (Fd-GOGAT) genes and their relationship with grain protein content QTL in wheat. *PLoS ONE* 9:e103869. <https://doi.org/10.1371/journal.pone.0103869>
- Nigro D, Giove SL, Fortunato S, Incerti O, Zacheo SA, Blanco A, Gadaleta A (2014b) Allelic variation of wheat flour allergens in a collection of wheat genotypes. *J Chem.* <https://doi.org/10.1155/2014/629379>
- Nigro D, Fortunato S, Giove SL, Paradiso A, Gu YQ, Blanco A, de Pinto MC, Gadaleta A (2016) Glutamine synthetase in durum wheat: genotypic variation and relationship with grain protein content. *Front Plant Sci* 7:971. <https://doi.org/10.3389/fpls.2016.00971>
- Nigro D, Fortunato S, Giove SL, Mangini G, Yacoubi I, Simeone R, Blanco A GA (2017a) Allelic variants of glutamine synthetase and glutamate synthase genes in a collection of durum wheat and association with grain protein content. *Diversity* 9:52. <https://doi.org/10.3390/d9040052>
- Nigro D, Laddomada B, Mita G, Blanco E, Colasuonno P, Simeone R, Gadaleta A, Pasqualone A, Blanco A (2017b) Genome-wide association mapping of phenolic acids in tetraploid wheats. *J Cereal Sci* 75:25–34. <https://doi.org/10.1016/j.jcs.2017.01.022>
- Nigro D, Gadaleta A, Mangini G, Colasuonno P, Marcotuli I, Giancaspro A, Giove SL, Simeone R, Blanco A (2019) Candidate genes and genome-wide association study of grain protein content and protein deviation in durum wheat. *Planta* 249(4):1157–1175. <https://doi.org/10.1007/s00425-018-03075-1>
- Pfeifer M, Kugler KG, Sandve SR, Zhan B, Rudi H, Hvidsten TR, International Wheat Genome Sequencing Consortium (2014) Genome interplay in the grain transcriptome of hexaploid bread wheat. *Science* 345(6194):1250091. <https://doi.org/10.1126/science.1250091>
- Pilolli R, Gadaleta A, Di Stasio L, Lamonaca A, De Angelis E, Nigro D, De Angelis M, Mamone G, Monaci L (2019a) A comprehensive peptidomic approach to characterize the protein profile of selected durum wheat genotypes: implication for coeliac disease and wheat allergy. *Nutrients* 11(10):2321. <https://doi.org/10.3390/nu11102321>
- Pilolli R, Gadaleta A, Mamone G, Nigro D, De Angelis E, Montemurro N, Monaci L (2019b) Scouting for naturally low-toxicity wheat genotypes by a multidisciplinary approach. *Sci Rep* 9(1):1646. <https://doi.org/10.1038/s41598-018-36845-8>
- Pingault L, Choulet F, Alberti A, Glover N, Wincker P, Feuillet C, Paux E (2015) Deep transcriptome sequencing provides new insights into the structural and functional organization of the wheat genome. *Genome Biol* 16(1):29. <https://doi.org/10.1186/s13059-015-0601-9>
- Sabelli P, Shewry PR (1991) Characterization and organization of gene families at the *Gli-1* loci of bread and durum wheats by restriction fragment analysis. *Theor Appl Genet* 83:209–216
- Shewry PR (2009) Wheat. *J Exp Bot* 60:1537–1553. <https://doi.org/10.1093/jxb/erp058>
- Shewry PR, Halford NG, Belton PS, Tatham AS (2002) The structure and properties of gluten: an elastic protein from wheat grain. *Philos Trans R Soc Lond B Biol Sci* 357(1418):133–142
- Sollid LM, Qiao SM, Anderson RP, Gianfrani C, Konig F (2012) Nomenclature and listing of celiac disease relevant gluten T-cell epitopes restricted by HLA-DQ molecules. *Immunogenetics* 64:455–460. <https://doi.org/10.1007/s00251-012-0599-z>
- Valenti S, Corica D, Ricciardi L, Romano C (2017) Gluten-related disorders: certainties, questions and doubts. *Ann Med* 49:569–581
- Wang S, Wong D, Forrest K, Allen A, Chao S, Huang BE, Maccaferri M, Salvi S, Milner SG, Cattivelli L, Mastrangelo AM, Akhunov E (2014) Characterization of polyploid wheat genomic diversity using a high-density 90 000 single nucleotide polymorphism array. *Plant Biotechnol J* 12(6):787–796. <https://doi.org/10.1111/pbi.12183>
- Zhang Y, Luo G, Liu D, Wang D, Yang W, Sun J, Zhang A, Zhan K (2015) Genome-, transcriptome- and proteome-wide analyses of the gliadin gene families in *Triticum urartu*. *PLoS ONE* 10:e0131559. <https://doi.org/10.1371/journal.pone.0131559>