# Characterization of the variability and the expression of α-gliadin genes from spelt [*Triticum aestivum* ssp. *spelta* (L.) Thell.] in the prospect of addressing the celiac disease issue

Benjamin Dubois

Septembre 2018

*Thèse présentée en vue de l'obtention du titre de docteur en sciences agronomiques et ingénierie biologique*

| | |
|---|---|
| **Promoteurs :** | Pierre Bertin (UCL) |
| | Dominique Mingeot (CRA-W) |
| **Membres du jury :** | Bruno Henry de Frahan (UCL) |
| | Xavier Draye (UCL) |
| | Yvan Larondelle (UCL) |
| | Jean-François Hausman (LIST) |
| | Yordan Muhovski (CRA-W) |

Earth and Life Institute

Faculté des bioingénieurs

Université catholique de Louvain

# ACKNOWLEDGMENTS

Je tiens à remercier en premier lieu mes deux promoteurs Pierre Bertin et Dominique Mingeot. L'encadrement par des personnes aussi compétentes et complémentaires fut extrêmement précieux pour mener à bien ce travail. Pierre, merci pour ta disponibilité et tes conseils toujours pertinents. Merci d'avoir partagé toutes tes connaissances et compétences scientifiques. Le recul que tu as pu prendre par rapport à nos discussions a souvent permis de trouver les solutions les plus adaptées. Dominique, merci aussi pour ton encadrement sans faille tout au long de ma thèse. Ton professionnalisme, ton ouverture et ta bienveillance, tout cela combiné au fait que nous travaillions dans le même bureau, a permis de construire une vraie relation d'échanges. A tous les deux, merci enfin pour les nombreuses heures que vous avez passées à corriger les différents documents que je vous ai soumis.

J'aimerais ensuite remercier les membres qui ont constitué mon jury de thèse : Bruno Henry de Frahan, Xavier Draye, Yvan Larondelle, Jean-François Hausman et Yordan Muhovski. Votre regard attentif et vos remarques pertinentes durant la défense privée furent d'une grande aide pour l'amélioration de ce manuscrit.

Merci aussi aux équipes de recherche qui m'ont entouré durant cette formidable aventure, que ce soit à l'UCL ou au CRA-W. Grâce aux moments conviviaux que nous avons passés ensemble, aux discussions que nous avons pu avoir ou encore au soutien et aux conseils dont j'ai bénéficié au laboratoire et en dehors, vous avez tous contribué d'une façon ou d'une autre à l'accomplissement de cette thèse.

Enfin, j'aimerais terminer en remerciant famille et amis pour leur présence toujours bienveillante à mes côtés. Entre frère, grands-parents, amis de Charleroi ou de Gembloux, difficile de citer tout le monde mais j'aimerais remercier en particulier mes parents pour m'avoir appris la curiosité et la persévérance. Merci pour votre soutien sans faille. Merci aussi et surtout à Gaëlle, ma compagne devenue ma

femme, présente depuis le premier jour de cette grande aventure et qui aura toujours été là pour m'écouter, me conseiller, me supporter et m'encourager.

# TABLE OF CONTENTS

# LIST OF FIGURES

## BIBLIOGRAPHIC REVIEW



## CHAPTER 1



## CHAPTER 2

## CHAPTER 3

# LIST OF TABLES

## BIBLIOGRAPHIC REVIEW



## CHAPTER 1



## CHAPTER 2

# ACRONYMS

| | |
|---|---|
| ANOVA | Analysis Of Variance |
| CD | Celiac Disease |
| cDNA | Complementary DNA |
| CRISPR | Clustered Regularly Inter-spaced Short Palindromic Repeats |
| DPA | Days Post-Anthesis |
| ELISA | Enzyme-linked Immunosorbent Assay |
| gDNA | Genomic DNA |
| GFD | Gluten-free Diet |
| GMO | Genetically-modified Organism |
| HLA | Human Leukocyte Antigen |
| HMW-GS | High Molecular Weight Glutenin Subunit |
| HPLC | High Performance Liquid Chromatography |
| LC-MS/MS | Liquid Chromatography tandem Mass Spectrometry |
| LMW-GS | Low Molecular Weight Glutenin Subunit |
| mRNA | Messenger RNA |
| NCGS | Non-celiac Gluten Sensitivity |
| ORF | Open Reading Frame |
| PCR | Polymerase Chain Reaction |
| PSC | Premature Stop Codon |
| qPCR | Quantitative PCR |

| | |
|---|---|
| RNAi | RNA Interference |
| RT-PCR | Reverse Transcription Polymerase Chain Reaction |
| SNP | Single Nucleotide Polymorphism |
| SSR | Simple Sequence Repeat |
| tTG | Tissue Transglutaminase |
| WA | Wheat Allergy |

# ABSTRACT

Gluten proteins are found in the seeds of cereals like bread wheat (*Triticum aestivum* L. ssp. *aestivum*) and spelt [*Triticum aestivum* ssp. *spelta* (L.) Thell.], and they play a major role in food processing thanks to the viscoelasticity properties they confer to the dough. They are, however, the source of gluten-related disorders when they are consumed by susceptible individuals. Among them, celiac disease (CD) has the highest clinical relevance and its pathophysiology is well understood, which led to focus this research on this disease only. CD is a pathology affecting genetically predisposed subjects where dietary gluten proteins trigger an inflammatory reaction in the upper small intestine. Among these proteins, the α-gliadins constitute the most immunogenic protein fraction as they display the highest number of peptides (epitopes) recognized by the immune system of CD patients. One of the strategies currently investigated to face the CD issue is to make use of the high genetic variability found in bread wheat and related species and sub-species with the aim of obtaining varieties consumable by CD patients. Among them, spelt belongs to the same species than bread wheat but it has been less studied. It displays interesting nutritional and agronomical properties and spelt germplasm collections hold a high genetic diversity. The objective of this research was thus to evaluate the potential of spelt to respond to the problematic of celiac disease by studying its genetic diversity in CD-related expressed epitope sequences. A cloning and sequencing strategy enabled to study the epitope composition of expressed α-gliadin sequences from a few contrasted spelt accessions selected through a genetic diversity analysis. This revealed a high variability in the epitope amino acid composition and it was linked to the genome from which they were expressed (A, B or D). No clear separation between spelt and bread wheat sequences was observed but different frequencies of some amino acid substitutions were still pointed out. TaqMan probes able to discriminate immunogenic from non-immunogenic epitope variants have then been developed and their specificity has been demonstrated. These probes were subsequently applied to a wide set of spelt accessions in order to study the variability in their epitope expression levels as a function of both genetic and environmental factors. Results showed a very high variability in the α-gliadin epitope composition and expression. Significant differences in the epitope expression were highlighted according to the accession provenance and habit (winter vs spring accessions), whereas spelt breeding did not seem to have influenced this expression. In addition, environmental factors, studied by harvesting samples during four consecutive years and by applying different N fertilization strategies, did not have a major impact on the epitope expression.

In conclusion, this thesis provides important information about the expression of CD-related α-gliadin epitopes in spelt and about its potential to develop varieties consumable by CD patients. Perspectives about the different breeding strategies available to obtain such celiac-safe variety are developed.[1]

[1] This abstract is an extended version of the one displayed on the back cover of this thesis.

# GENERAL INTRODUCTION

Wheat is one of the most cultivated crops all over the world and it significantly contributes to the human diet thanks to the high proportion of starch and proteins in the grain. It is also an important source of dietary fibers, B vitamins (B1, B2, B3, B6 and B9) and mineral micronutrients (especially Fe, Zn and Se) (Shewry and Tatham, 2016). The use of wheat in a high number of food products is explained by the technological properties of the gluten proteins. They constitute the majority of the seed storage proteins and are divided into two main groups: gliadins and glutenins. When water is added to the flour, a continuous network is formed by these two proteins classes, which provides a unique combination of viscosity and elasticity properties to the dough (Shewry et al., 2003). This enables, among other things, to produce leavened foods by entrapping carbon dioxide into the dough during fermentation.

These seducing properties of gluten proteins are, however, counterbalanced by the immune reaction they trigger when they are consumed by some individuals (Scherf et al., 2016). Several pathologies are associated with the gluten consumption including celiac disease (CD), which has the strongest clinical relevance. In this disorder, gluten peptides trigger an inflammatory reaction in the upper small intestine since some of their sequences, called epitopes, are recognized by the immune system of CD patients. Among gluten proteins, the α-gliadins have been shown to display the most immunogenic and the highest number of CD-related epitopes (Arentz-Hansen et al., 2002; Ciccocioppo et al., 2005; Vader et al., 2003).

People affected by gluten-related disorders must thus follow a gluten-free diet (GFD) in order to avoid consuming such immunogenic peptides. Gluten-free foods are, however, expensive, generally of lower tasting and structural quality than gluten-containing ones and they may have a socially isolating effect (Lebwohl et al., 2015). In addition, gluten is found in hidden sources like sauces, drug fillers and processed meats. In consequence, it would be relevant to investigate other approaches enabling to avoid the exposition of CD patients to theses epitopes. In

this framework, the development of varieties with a lowered amount of immunogenic gluten peptides, making them consumable by subjects suffering from pathologies related to the gluten consumption, could be a promising strategy.

Among wheat species, spelt [*Triticum aestivum* ssp. *spelta* (L.) Thell.] belongs to the same species as bread wheat (*Triticum aestivum* L. ssp. *aestivum*). It displays interesting nutritional and agronomical properties, including its rusticity, and Belgium is one of the leading countries in spelt breeding, together with Germany and Switzerland. It has been reported that the spelt germplasm displays a high genetic diversity (Bertin et al., 2004) and this diversity has not been investigated yet for its CD-related immunogenic content. In addition, high variations in the grain protein amount has been pointed out in spelt (Gomez-Becerra et al., 2010) and it has even been shown that some spelt accessions displayed higher protein contents than bread wheat (Bonafaccia et al., 2000). In consequence, all this diversity existing in spelt germplasm could be exploited to find out accessions with low immunogenic potential. Moreover, the close genetic relationships between spelt and bread wheat would make easier the transfer of such desirable trait from one sub-species to the other with the aim of answering to expectations of CD patients.

**References**

Arentz-Hansen, H., Mcadam, S. N., Molberg, Ø., Fleckenstein, B., Lundin, K. E. A., Jørgensen, T. J. D., et al. (2002). Celiac lesion T cells recognize epitopes that cluster in regions of gliadins rich in proline residues. Gastroenterology 123, 803–809. doi:10.1053/gast.2002.35381.

Bertin, P., Grégoire, D., Massart, S., and de Froidmont, D. (2004). High level of genetic diversity among spelt germplasm revealed by microsatellite markers. Genome 47, 1043–1052. doi:10.1139/g04-065.

Bonafaccia, G., Galli, V., Francisci, R., Mair, V., Skrabanja, V., and Kreft, I. (2000). Characteristics of spelt wheat products and nutritional value of spelt wheat-based bread. Food Chemistry 68, 437–441. doi:10.1016/S0308-8146(99)00215-0.

Ciccocioppo, R., Di Sabatino, A., and Corazza, G. R. (2005). The immune recognition of gluten in coeliac disease. Clinical and Experimental Immunology 140, 408–416. doi:10.1111/j.1365-2249.2005.02783.x.

Gomez-Becerra, H. F., Erdem, H., Yazici, A., Tutus, Y., Torun, B., Ozturk, L., et al. (2010). Grain concentrations of protein and mineral nutrients in a large collection of spelt wheat grown under different environments. Journal of Cereal Science 52, 342–349. doi:10.1016/j.jcs.2010.05.003.

Lebwohl, B., Ludvigsson, J. F., and Green, P. H. R. (2015). Celiac disease and non-celiac gluten sensitivity. BMJ, h4347. doi:10.1136/bmj.h4347.

Scherf, K. A., Koehler, P., and Wieser, H. (2016). Gluten and wheat sensitivities – An overview. Journal of Cereal Science 67, 2–11. doi:10.1016/j.jcs.2015.07.008.

Shewry, P. R., and Tatham, A. S. (2016). Improving wheat to remove coeliac epitopes but retain functionality. Journal of Cereal Science 67, 12–21. doi:10.1016/j.jcs.2015.06.005.

Shewry, P. R., Halford, N. G., and Lafiandra, D. (2003). Genetics of wheat gluten proteins. Advances in Genetics 49, 111–84.

Vader, L. W., Stepniak, D. T., Bunnik, E. M., Kooy, Y. M. C., Haan, W. D., Drijfhout, J. W., et al. (2003). Characterization of Cereal Toxicity for Celiac Disease Patients Based on Protein Homology in Grains. Gastroenterology 125, 1105-1113.

# BIBLIOGRAPHIC REVIEW

This chapter aims at introducing each concept approached in the following chapters so that the reader has comprehensive information to fully understand the results obtained in this work. This bibliographic review starts by describing the importance of cereals in the human nutrition, the characteristics of spelt [*Triticum aestivum* ssp. *spelta* (L.) Thell.] and the usefulness of gluten proteins to elaborate bakery products. Despite the interesting properties of gluten proteins, they can still trigger, in susceptible individuals, several disorders that are detailed in the second section. The gluten immunogenic sequences involved in CD are then presented in the third section and their genomic location is detailed according to the complex structure of the wheat genome. Given that environmental conditions can have an impact on the amount of immunogenic sequences found in gluten proteins, their influence on the expression of gluten sequences is approached in the fourth section. The fifth one is dedicated to the medical solutions currently investigated with aim of facing the CD issue. Since these medical approaches have not provided definitive solutions to face CD yet, the sixth section details a promising alternative methodology consisting in breeding strategies to reduce the accession immunogenicity. The seventh section investigates with which main research topics the agri-food industry is dealing by analyzing patented inventions. Finally, given that trying to face the CD issue requires to study the immunogenicity of accessions and food stuffs, tools enabling to do so are presented in the last section of this bibliographic review.

# 1. Cereals and gluten

## 1.1. The importance of cereals and gluten in human nutrition

Cereals are the most widely cultivated crops over the world as they represented 32% of the total harvested crop production in 2016 (FAO, 2016). Among cereals, maize, wheat and rice account for about 90% of the production with respectively 1060, 749 and 740m tons harvested in 2016. Bread wheat (*Triticum aestivum* L. ssp. *aestivum*) represents about 95% of the wheat grown worldwide, while the remaining 5% are made of related species and sub-species like spelt [*Triticum aestivum* ssp. *spelta* (L.) Thell.], durum wheat [*Triticum turgidum* ssp. *durum* (Desf.) Husnot], emmer [*Triticum turgidum* ssp. *diccocum* (Schrank ex Schübler) Thell.] or einkorn (*Triticum monococcum* L. ssp. *monococcum*) (Shewry and Hey, 2015). Wheat displays the widest geographical distribution and is grown and consumed in both industrialized and developing countries (Shewry and Tatham, 2016). It is a staple food and it contributes to the food safety in developing countries thanks to its high nutritional values in proteins, starch, dietary fibers, B vitamins and micronutrients. In addition to their nutritional importance, proteins from wheat grains also display technological properties enabling the flour to be processed into various products such as bread, biscuits, cakes, noodles and pasta. These properties are provided by the gluten complex, which is made of seed storage proteins and represents up to 80% of the total wheat grain proteins (Seilmeier et al., 1991). Gluten is the result of the washing of wheat dough in water or dilute salt solution, leaving a cohesive mass made of about 80% of proteins and 20% of mainly remaining trapped starch granules. Related gluten proteins are also found in barley (*Hordeum vulgare* L.), rye (*Secale cereale* L.) and oat (*Avena sativa* L.) seeds (hordeins, secalins and avenins, respectively), but they do not display the same properties and they thus need to be blended with bread wheat flour to make baking products of acceptable quality (Shewry and Tatham, 2016). When mixing flour and water to make dough, gluten proteins are brought together to form a complex network providing viscoelastic

properties important for the dough to entrap carbon dioxide released during leavening.

### 1.2. Spelt [*Triticum aestivum* ssp. *spelta* (L.) Thell.]

Spelt is hexaploid (AABBDD genome) and belongs to the same species as bread wheat (*Triticum aestivum* L. ssp. *aestivum*). It however displays some typical morphological features like a narrow and lax spike with a brittle rachis, adherent glumes (*i.e*. hulled grains) and non-spherical seeds (Campbell, 1997). Its spikes break up into spikelets during threshing, and further processing is needed to free the grain.

Spelt is seen as an old crop as it has been cultivated since 5000 BC and was one of the major feed and food crops in ancient Europe. During the 20$^{th}$ century, the spelt cultivation drastically dropped since bread wheat almost completely replaced it thanks to its better bread-making properties, higher yield and lower processing cost. In the last decades, however, spelt has gained renewed interest because of the increasing demand for unconventional food and low-input agriculture (An et al., 2005). The hulled seeds of spelt indeed offers a physical protection to the grains and it has thus lower pesticides requirements, making it more suitable for organic farming (Kozub et al., 2014). In addition, it displays good adaptability to poor soils, to harsh climatic conditions, and resistance to biotic and abiotic stresses (An et al., 2005; Sobczyk et al., 2017; Xie et al., 2015).

Spelt is cultivated in Europe, Asia, Australia, North Africa, USA and Canada. In USA, the production is mainly located in Northern Ohio, where less fertile soils and low rainfall levels are found, and approximately 8,000 ha were dedicated to its culture in 1995 (Campbell, 1997; Cubadda and Marconi, 2002). Spelt was planted on 2,000 ha in Canada in 2001 and most of the production was intended for the organic market (Abdel-Aal and Hucl, 2005). The main spelt cultivation areas in Europe are located in Germany, Belgium, Austria and Switzerland. Spelt covered 23,000 ha in Germany in 2008 (Federal Statistical Office Germany, 2008). In Belgium, approximately 10,000 ha are cultivated each year but high variations in the

produced amounts can be observed from one year to another (Statbel, 2018). In 2014 for instance, a strong external demand especially from Germany increased the mean price of spelt and pushed Belgian farmers to grow spelt on 20,000 ha in 2015. This led to an overproduction and the spelt culture area dropped to 9,000 ha in 2016. This is due to the narrowness of the spelt market where small variations can strongly affect prices and thus lead to aberrant production from one year to another (Escarnot, 2016). The surface dedicated to spelt cultivation is increasing in Switzerland where it ranged from 7,300 ha in 2000 to 15,700 ha in 2014 (Office fédéral de la statistique Suisse, 2017). This phenomenon is also observed in Austria, reaching 10,000 ha in 2014 (Statistics Austria, 2018), where most of the production is organic (83% during the 2006-2016 period). Spain also displays small culture surfaces of spelt in the mountainous areas of Asturias, as well as Italy, Slovenia and Czech Republic.

Spelt has undergone less breeding efforts than bread wheat and a significant lower amount of spelt cultivars have been developed, compared to bread wheat. The US National Plant Germplasm is one of the biggest germplasm collections worldwide and provides a good illustration since 54,900 accessions are reported for bread wheat but only 1,345 for spelt, regardless of their breeding status. These spelt accessions, however, display an interesting variety in terms of provenance country since they were reported from 32 countries (Figure 1). In addition, it has been demonstrated several times that collections of spelt display a high genetic diversity based on their high- and low-molecular-weight (HMW and LMW) glutenin pattern (An et al., 2005), their gliadin pattern (Caballero et al., 2004) or their SSR (Simple Sequence Repeat) allelic pattern (Bertin et al., 2004).

### 1.3. Gluten proteins

Wheat gluten proteins consist of gliadins and glutenins which confer viscosity and elasticity to the dough, respectively (Shewry et al., 2003). Gliadins are monomeric proteins with molecular weights (MWs) around 28 – 55 kDa, and were initially divided into four groups according to their electrophoretic mobility: α-, β-, γ- and ω-gliadins. The subsequent comparisons of α- and β-gliadins highlighted similar

sequences and led to the definition of three groups: α/β- (or simply α-), γ- and ω-gliadins. Glutenins are composed of aggregated proteins linked by interchain disulfide bonds with MWs ranging from 500 to 10,000 kDa. The reduction of these disulfide bonds leads to two subunit fractions separated by sodium dodecylsulphate polyacrylamide gel electrophoresis (SDS-PAGE): the high-molecular-weight (HMW) glutenin subunits (MW: 67 – 88 kDa) and the low-molecular-weight (LMW) glutenin subunits (MW: 32 – 35 kDa). Native glutenins consist of a backbone of HMW glutenin subunits ramified with LMW glutenin subunits. Gliadins are aggregated to glutenins thanks to hydrogen, ionic and hydrophobic bonds (Wieser, 2007). All these interactions between gluten fractions provide to the dough outstanding technological properties.

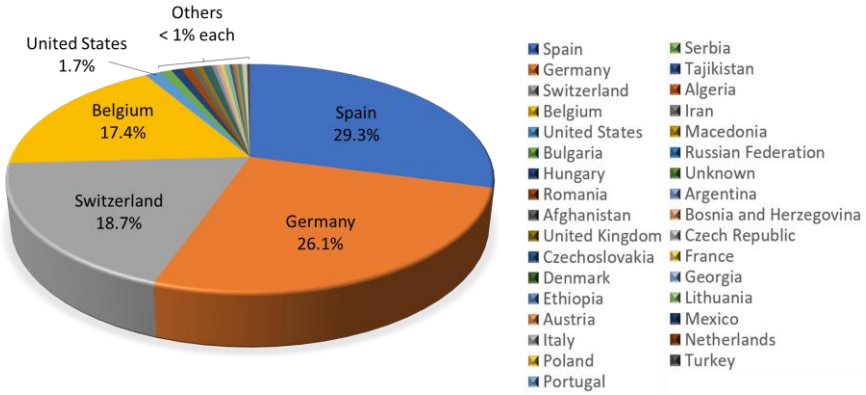



Figure 1. Geographical provenances of the spelt accessions reported in the US National Plant Germplasm in May 2018 (https://www.ars-grin.gov/).

## 2. Gluten proteins-related disorders

Beside its bread-making qualities, gluten has attracted great attention in the last decades because of the raising number of patients suffering from an intolerance to this protein fraction. The prevalence of these disorders is not always easy to assess because of the "gluten-free fashion": some healthy people indeed stop consuming gluten-containing products because of the common preconception that gluten is bad for their health. They then assess they feel better, concluding that they are affected by a gluten-related pathology, while the combination of the placebo and the fashion effect might greatly affect the judgment in such self-diagnosis. The gluten-free market has by the way displayed a substantial growth during the last years, raising from $2.5 billion in 2010 (Sapone et al., 2012) to $11.6 billion in 2015 (Reilly, 2016) in the United States (+ 364%). The increase of the disease prevalence and awareness is not correlated to the disproportionate raise in the gluten-free food industry (Reilly, 2016) and it has been shown that the majority of the gluten-free product consumers have never been medically diagnosed (Topper, 2014).

However, besides this confusion over healthy people consuming gluten-free products, three main pathologies are associated to the consumption of wheat and/or gluten and are detailed hereafter.

### 2.1. Wheat allergy

Wheat allergy (WA) is defined as a reaction to wheat proteins – not only gluten – triggered by immunoglobulin (Ig)-E antibodies. It affects children and adults, with prevalence ranging from 0.5% to 3% (Zuidmeer et al., 2008). It can take different forms according to the underlying immunogenic mechanisms, but two main types are generally recognized (Cianferoni, 2016). One the one hand, the baker's asthma is one of the most prevalent allergies in many countries and is due to the inhalation of wheat and cereal flours and dusts (Tatham and Shewry, 2008). Several wheat protein types have been shown to react with IgE from patients affected by baker's asthma, like α-amylase inhibitors, peroxidases and non-specific lipid transfer

proteins (Elli et al., 2015; Inomata, 2009; Tatham and Shewry, 2008). On the other hand, the dietary allergy to wheat is less widespread than baker's asthma and includes wheat-dependent exercise-induced anaphylaxis (WDEIA), atopic dermatis, urticaria and anaphylaxis (Sapone et al., 2012). The implicated proteins are not clearly defined but it has been shown that, in a population of 28 patients, 60% displayed IgE to α/β-gliadins and LMW glutenin subunits, 55% to γ-gliadins, 48% to ω-gliadins and 26% to HMW glutenin subunits (Battais et al., 2003). The diagnosis of wheat allergy generally involves skin prick tests and *in vitro* IgE assays but, given the cross-reactivity with grass pollens, an oral food challenge is necessary in many cases (Sapone et al., 2012).

### 2.2. Non-celiac gluten sensitivity

Patients affected by non-celiac gluten sensitivity (NCGS) display gastrointestinal and extra-intestinal symptoms which arise a short time after gluten consumption, in the absence of celiac-specific antibodies, intestinal villous atrophy and any allergy-related processes (Catassi et al., 2013). The prevalence of NCGS is hard to estimate firstly because of the self-reporting of the disease. A survey in the United Kingdom showed that 13% of the studied population self-reported to be affected by NCGS whereas only 1% of the subjects underwent a medical diagnosis (Aziz et al., 2014). Moreover, the pathophysiology is unclear and clear serologic or histopathologic criteria are not available to diagnose NCGS, in contrast with celiac disease and wheat allergy. It can be hypothesized that efforts made to find solutions to face CD could also benefit to patients affected by NCGS but the fuzziness surrounding the NCGS determinants and pathophysiology does not enable to claim it in advance. An overlap with other disorders has also been pointed out. Among them, the irritable bowel syndrome (IBS) is the major one since many symptoms are common to both pathologies and it has been shown that 28% of a selected population suffering from IBS was affected by NCGS (Biesiekierski et al., 2011). The fact that the state of health of patients affected by NCGS and IBS improves after reduction of FODMAPs (fermentable, oligo-, di-, monosaccharides and polyols) consumption indicates that some components of wheat other than

gluten proteins may have a role in these pathologies (Biesiekierski et al., 2013; Shepherd et al., 2014).

## 2.3. Celiac disease

### 2.3.1. Prevalence of the disease

Celiac disease (CD) is the most common food intolerance of western populations as its prevalence is about 1% in these countries (Dubé et al., 2005). This prevalence has increased fourfold to fivefold over the past 50 years (Lebwohl et al., 2015) and may greatly fluctuate from one region/population to another since it is dependent on several factors including the underlying genetics of the population, the gluten exposure and possibly the infant feeding patterns, other environmental risk factors, the awareness of the disease among patients and doctors, and the method and frequency of testing. It has indeed been reported that studies based only on serology conclude to higher prevalence levels than screenings requiring confirmation through small intestinal biopsy (Lebwohl et al., 2015). In addition, the majority of celiac patients are still undiagnosed (Rubio-Tapia et al., 2012).

### 2.3.2. Pathophysiology

CD is an autoimmune disorder involving both the innate and the adaptive immunity which affects genetically predisposed individuals. This genetic predisposition comes from two highly polymorphic genes named DQA1 and DQB1, encoding a transmembrane heterodimer expressed on the surface of antigen-presenting cells (APC) (Sollid et al., 1989). About 95% of CD patients display the DQ2 haplotype, which can take two forms: the HLA-DQ2.5 variant (DQA1*0501/DQB1*0201) or less frequently the HLA-DQ2.2 variant (DQA1*0201/DQB1*0202), which displays the same peptide-binding properties than the HLA-DQ2.5 variant. However, the HLA-DQ2.2 must be expressed together with the HLA-DQ2.5 to predispose to CD. The remaining 5% of patients mostly have the DQ8 haplotype (DQA1*0301/DQB1*0302, termed the HLA-DQ8 variant), which thus also predispose to CD. These HLA-DQ2 and -DQ8 variants are expressed on the cellular surface of APC to bind gluten peptides and to present them to a specific population

of T-cells (Escudero-Hernandez et al., 2016). Importantly, gluten peptides must display negatively charged residues to allow a perfect suit into the binding pocket of HLA-DQ molecules. This feature is provided by the tissue transglutaminase (tTG) enzyme which is secreted in several tissues and can be induced under injured or inflammatory conditions (Kim, 2006). This enzyme confers a negative charge to gluten peptides by deaminating glutamine residues into glutamic acid, enabling their appropriate recognition by HLA-DQ molecules. The subsequent presentation to T-cells triggers an inflammatory cascade by secreting cytokines which activate cytotoxic T lymphocytes, macrophages, stromal cells and B lymphocytes, leading to tissue damage and antibody production (Ciccocioppo et al., 2005).

However, it must be mentioned that the presence of these HLA genes is a condition necessary but not sufficient to develop the disease since they account for only 40% of the CD genetic inheritability (Bevan et al., 1999). This suggests that other genes must be determinant for the CD susceptibility. Two genome-wide association studies and an immunochip approach identified a total of 39 loci outside the HLA loci, which were hypothesized to control gene expression (Farh et al., 2015; Trynka et al., 2011; Van Heel et al., 2007). Epigenetic factors like non-coding RNA and the DNA methylation have also been postulated to influence the expression of genes involved in CD. In addition, since the concordance is only 85% between monozygotic twins (Nistico et al., 2006), several environmental factors have been proposed to have a role in the pathology development, like the infant feeding patterns, the perinatal exposure, infections and even smoking and drugs (Lebwohl et al., 2015).

### 2.3.3. Diagnosis and symptoms

The inflammatory reaction triggered by the recognition of gluten peptides by T-cells mainly affects the mucosa of the upper small intestine. It is characterized by a villous atrophy, which can range from partial damage to a complete absence of villi, an increased infiltration of lymphocytes in the epithelium and crypt hyperplasia (Lebwohl et al., 2015). The severity of the mucosal lesions is generally assessed through the Marsh-Oberhuber classification, which takes into account the ratio of

villous height to crypt depth and the number of intraepithelial lymphocytes. The diagnosis of the CD pathogenesis is based on serological tests where the presence of specific antibodies is analyzed. Since the proportion of seronegative CD patients accounts for up to 10% of all diagnosed cases, the diagnosis is generally confirmed by a histological analysis of the small intestine mucosa. In case of doubtful results, a genotyping of HLA-DQ genes can also be carried out. The final diagnosis is assessed by the positive serological, histological and clinical response to a gluten-free diet (Scherf et al., 2016).

Intra- and extra-intestinal symptoms are associated with CD. Intestinal symptoms usually consist in abdominal pain, diarrhea and vomiting, but non-classical symptoms may also be observed like gastroesophageal reflux and constipation. Most extra-intestinal manifestations come from the malabsorption of nutrients and occur as vitamin and mineral deficiencies such as anemia, decreased bone mineral density, bone pain and fractures. In consequence, growth retardation phenomenon may be observed in children. Other extra-intestinal symptoms like depression, migraine, anxiety and epilepsy can also be displayed by some patients (Scherf et al., 2016).

## 3. Wheat genomes and sequences involved in CD

### 3.1. Structure and origin of wheat genome

Bread wheat and spelt are hexaploid wheats (2n = 6X = 42) since their genome is made of three homeologous genomes (A, B and D), each one containing seven pairs of chromosomes (1A to 7A, 1B to 7B and 1D to 7D). The term "homeologous" refers to genes, chromosomes or genomes that originated by speciation and that were brought back together in the same genome by allopolyploidization (Glover et al., 2016). For example, the homeologous group 1 includes the 1A, 1B and 1D chromosomes. In the case of bread wheat and spelt, different polyploidization events occurred (Figure 2). The first hybridization is estimated to have occurred 360,000 to 500,000 years ago (Dvorak and Akhunov, 2005; Huang et al., 2002) between two diploid progenitors: an ancestor of *Triticum urartu* Tumanian ex Gandilyan (2n = 2X = 14, AA genome) and an unknown species (BB genome) related to *Aegilops speltoides* Tausch (2n = 2X = 14, SS genome) to give rise to the tetraploid emmer wheat *Triticum turgidum* ssp. *dicoccum* (Schrank ex Schübler) Thell. (2n = 4X = 28, AABB genome). The second cross occurred 8,000 – 10,000 years ago between a free-threshing tetraploid wheat (which could be *Triticum turgidum* ssp. *turgidum* or *Triticum turgidum* ssp. *durum*) and a diploid ancestor of *Aegilops tauschii* Cosson (DD genome) to form a hexaploid wheat (2n = 6X = 42, AABBDD genome) (Salamini et al., 2002). While bread wheat emerged through this evolutional pathway, spelt is thought to have two distinct geographical and hybridization origins. On the one hand, Asian spelt might have emerged in Iran through the hybridization between emmer and the ancestor of *Aegilops tauschii*, whereas, on the other hand, European spelt seems to be the result of a cross between emmer and bread wheat (Blatter et al., 2004; Dvorak et al., 2012; Kozub et al., 2014).

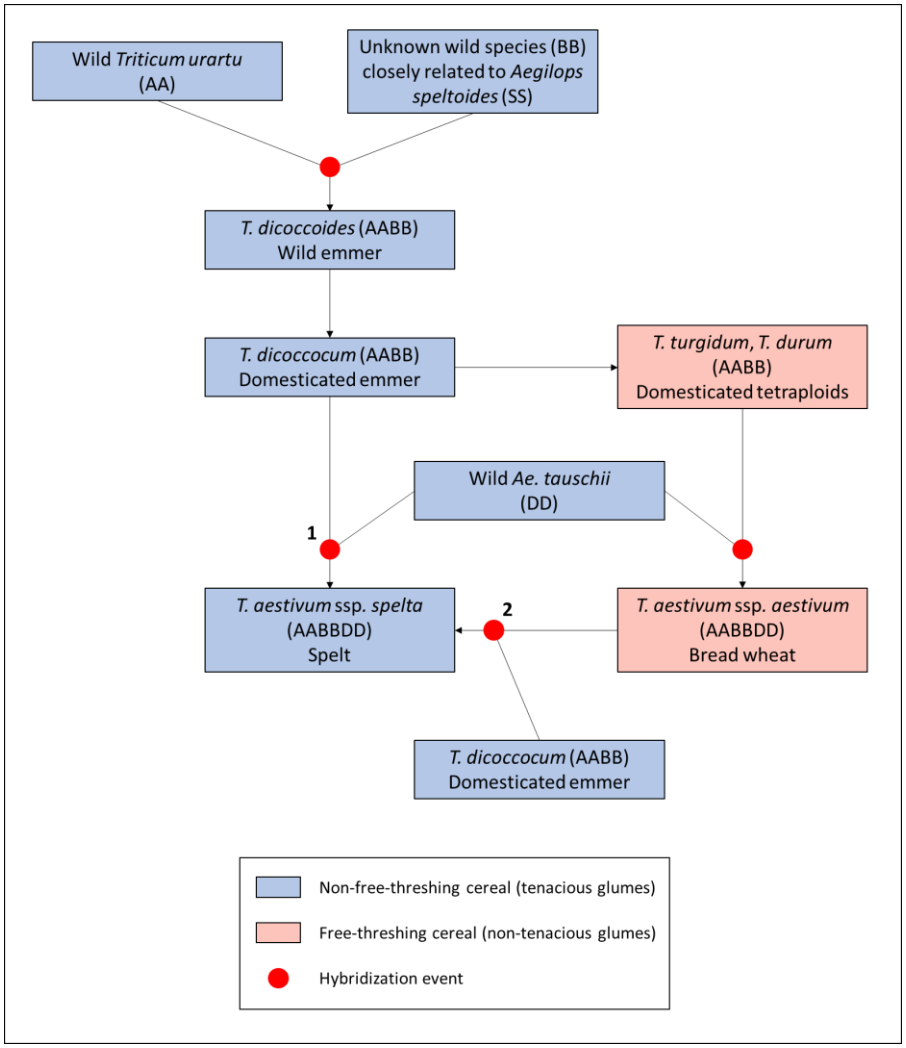



Figure 2. Proposed evolution model of bread wheat and spelt. Spelt has two evolutionary pathways since it is thought to have emerged through two different hybridization events in distinct geographical sites: one in Iran (1) and one in Europe (2).

During recent years, efforts have been undertaken in order to sequence the genomes of bread wheat and related diploid species. While the genomes of *Triticum urartu* and *Aegilops tauschii* have been completely or almost completely assembled (Jia et al., 2013; Ling et al., 2013), the *Aegilops speltoides* genome has not been sequenced yet. Concerning the bread wheat genome, it has been almost completely sequenced. However, its polyploid complexity and its large size have

been major barriers to the bread wheat genome analysis and redundant and multigenic chromosomic regions remain unassembled (Brenchley et al., 2012; International Wheat Genome Sequencing Consortium, 2014).

### 3.2. Gluten protein-encoding sequences involved in CD

The α-gliadin genes represent a complex multigenic family: the haploid genome of wheats may include a number of α-gliadin gene copies ranging from 25 (Harberd et al., 1985) to 100 (Okita et al., 1985) or even up to 150 (Anderson et al., 1997) according to the variety and the resolution of the technique used. However, it has been demonstrated that at least half of the α-gliadin genes from bread wheat displayed a premature stop codon (PSC) and are thus pseudogenes (Anderson and Greene, 1997; Ozuna et al., 2015). Alpha-gliadin genes are located in complex clusters in the Gli-2 loci, on the short arms of the group 6 chromosomes (Gli-A2, Gli-B2 and Gli-D2 on the chromosomes 6A, 6B and 6D, respectively).

The HMW glutenin subunits are encoded by the Glu-1 loci located on the long arms of the group 1 chromosomes (Glu-A1, Glu-B1 and Glu-D1 on the chromosomes 1A, 1B and 1D, respectively), each one including two genes encoding for two types of subunit. LMW glutenin subunits are encoded by three major loci located on the short arms of the group 1 chromosomes (Glu-A3, Glu-B3, Glu-D3 on the chromosomes 1A, 1B and 1D, respectively). These arms also display the Gli-1 loci, including clusters of tightly linked genes encoding γ- and ω-gliadins (Gli-A1, Gli-B1 and Gli-D1). In addition, some minor loci encoding gliadins and LMW glutenin subunits have also been mapped elsewhere on chromosomes 1 and 6 (Shewry et al., 2003; Shewry and Tatham, 2016).

Among gluten proteins, the α-gliadins constitute the most important class as they represent 15-30% of wheat seed proteins (Gu et al., 2004). Moreover, they display the highest immunogenicity since they trigger the strongest T-cell activation (Arentz-Hansen et al., 2002; Camarca et al., 2009; Ciccocioppo et al., 2005; Vader et al., 2003). Given their high content in proline and glutamine residues, α-gliadins are pretty resistant to the gastrointestinal digestion with pepsin and trypsin, and

several intact peptides can thus follow the gastrointestinal tract until the upper small intestine where they stimulate the immune system of CD patients.

The typical structure of an α-gliadin sequence includes an N-terminal signal peptide, a repetitive domain, two regions rich in glutamine residues separated by a first unique domain, and finally a second unique domain (Figure 3).

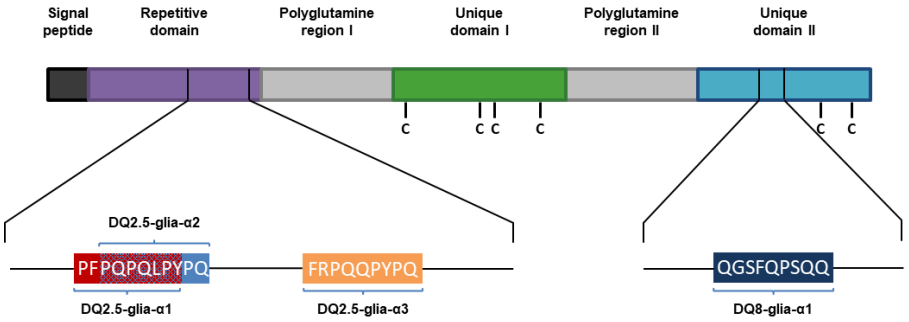



Figure 3. Typical structure of an α-gliadin sequence. The four major T-cell stimulatory epitopes related to CD and their corresponding amino acid sequences are represented below the polypeptide model. 'C' indicates the classical location of the cysteine residues

The unique domains I and II display six cysteine residues together, which play a role in the dough quality since they form intramolecular disulfide bonds stabilizing the compact globular protein fold. Four peptides of nine amino acids displayed by the α-gliadin sequences are particularly recognized by T-cells of CD patients and are thus defined as epitopes: two are major epitopes (the overlapping DQ2.5-glia-α1 and DQ2.5-glia-α2 epitopes) and the other two are minor epitopes (the DQ2.5-glia-α3 and DQ8-glia-α1 epitopes). In the α-gliadin sequences expressed from the D genome, the DQ2.5-glia-α2 epitope can be displayed in one, two or three copies and leads, when three copies are present, to the 33-mer which is the most immunogenic fragment of α-gliadins (Molberg et al., 2005). When this duplication or triplication occurs, the DQ2.5-glia-α1 epitope takes thus two forms of equal immunogenicity: PFPQPQLPY and PYPQPQLPY. In addition, allelic variations in these epitope sequences have been reported (see Table 1 for an example for the DQ2.5-glia-α3 epitope) and they were shown to be specific to the genome they are expressed from (A, B or D). It has been shown by analyzing their recognition by

patient-derived α-gliadin-specific T-cell clones that these allelic variants display lower or suppressed immunogenic properties (Mitea et al., 2010). A high diversity in the composition and the expression of the epitope sequences has been highlighted in diploid species representative of the A genome (Ma et al., 2007; Zhang et al., 2015), the B genome (Van Herpen et al., 2006) and the D genome (Qi et al., 2012), in tetraploid wheats (Salentijn et al., 2009; Salentijn et al., 2013; Zhang et al., 2014) and in bread wheat (Kaur et al., 2016; Li et al., 2014; Noma et al., 2015; Xie et al., 2010). The diversity of α-gliadins expressed by spelt accessions has, however, not been extensively studied in the literature.

Table 1. Composition of the canonical form and the allelic variants of the DQ2.5-glia-α3 epitope

| | Position of the residue in the epitope peptide | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| Canonical epitope | TTT<br>F | CGA<br>R | CCA<br>P | CAA<br>Q | CAA<br>Q | CCA<br>P | TAT<br>Y | CCA<br>P | CA(A/G)<br>Q |
| Allelic variants | TTT<br>F | C**C**A<br>**P** | CCA<br>P | CAA<br>Q | CAA<br>Q | CCA<br>P | TAT<br>Y | CCA<br>P | CAA<br>Q |
| | TTT<br>F | CGA<br>R | C**A**A<br>**Q** | CAA<br>Q | CAA<br>Q | CCA<br>P | TAT<br>Y | CCA<br>P | CAA<br>Q |
| | TTT<br>F | CGA<br>R | CCA<br>P | CAA<br>Q | CAA<br>Q | **T**CA<br>**S** | TAT<br>Y | CCA<br>P | CAA<br>Q |
| | TTT<br>F | CGA<br>R | CCA<br>P | CAA<br>Q | **A**AA<br>**K** | CCA<br>P | TAT<br>Y | CCA<br>P | CAA<br>Q |
| | TTT<br>F | C**A**A<br>**Q** | CCA<br>P | CAA<br>Q | CAA<br>Q | CCA<br>P | TAT<br>Y | CCA<br>P | CAA<br>Q |

For each residue position in the epitope peptide, the codon composition is given together with the amino acid in which it is translated. Nucleotide and consequent amino acid substitutions leading to a reduction or a suppression of the epitope immunogenicity are highlighted in red.

# 4. Environmental influence

The environmental conditions can have a crucial influence on genes expression levels. In the framework of gluten protein expression profiling and beside the genetic variability, two main factors can be identified as possibly influencing their amount and relative proportions in the grain: the meteorological conditions and the N fertilization strategy.

## 4.1. Meteorological conditions

Spelt and bread wheat are divided into two groups according to their habit: winter accessions are generally sown in October-November and harvested in July-August of the following year whereas spring accessions are sown in March and harvested a few days or weeks after winter ones. Interestingly, Maghirang et al. (2006) studied the grain composition of 100 winter and 100 spring bread wheat accessions and they showed that the flour of spring accessions displayed better bread-making properties than winter ones. In addition, another work analyzed the same amount of winter and spring bread wheat accessions to study the size distribution of starch granules since it is correlated to bread-making properties (Park et al., 2009). It was shown that spring accessions displayed larger size and proportion of B-granules than winter ones, and it was hypothesized that the hot and dry growing conditions during the grain filling may be responsible for this difference. Johansson and Svensson (1998) studied the effects of weather parameters on the grain protein composition in two winter and two spring Swedish bread wheat cultivars grown during the period 1975-1996. Interestingly, they showed that the climatic conditions influenced the bread volume and the grain protein concentration and quality. The temperature was the most influencing weather parameter in spring accessions, especially during the grain filling period, whereas the influence of weather conditions on the quality parameters of winter accessions was more complex. All these differences highlighted between winter and spring accessions are thus putatively due to different climatic conditions, but they might also rely on

genetic factors and be linked to differences in the gene pools used in breeding programs for spring vs winter accessions.

Another way to analyze the meteorological impact could consist in the comparison of gene expression levels for identical accessions from one harvest year to another. Noma et al. (2015) analyzed the expression of α-gliadin sequences from bread wheat with samples harvested in 2007, 2008 and 2012 and they noticed that the expression levels and the expression peak differed for the three years. Moreover, they showed that the expression of most α-gliadins was correlated with the sunshine duration. A Swedish research group studied bread wheat cultivars harvested from 1990 to 1996 and they showed significant variations in the grain protein concentration and the gluten strength according to the harvest year (Johansson and Svensson, 1999). In addition to these two parameters, they also showed that the harvest year significantly influenced the amount and the size distribution of most mono- and polymeric proteins when they analyzed another set of bread wheat accessions collected in 1991 and from 1994 to 2000 (Johansson et al., 2003). In another geographical region, Hungarian researchers harvested bread wheat cultivars in 2011, 2012 and 2014 and they highlighted variations in the crude protein content and the relative amount of gliadin subgroups according to the harvest year (Hajas et al., 2017). Some of these accessions were also studied for their 33-mer content by an LC-MS/MS method (Schalk et al., 2017). Significant variations were pointed out according to the harvest year and authors even postulated that the environmental factor might have a greater influence on the 33-mer content than the genetic background of the accessions.

### 4.2. N fertilization

Among environmental factors, the N fertilization strategy is another component that might influence the grain protein content and the relative abundances of the different seed storage proteins (Altenbach et al., 2011; Wieser and Seilmeier, 1998). It has been demonstrated that increasing fertilizer rates led to increased protein concentrations and decreased gluten strength, which resulted in higher bread volumes (Johansson and Svensson, 1999; Johansson et al., 2003). Other

studies highlighted an accumulation of gliadins and total prolamin content when increasing the N fertilizing rates (Matre et al., 2006; Triboï et al., 2003). In addition, a splitting strategy can also be used and consist in the application of N fertilizer at different development stages of the plant. This splitting has indeed been pointed out as an effective way to improve wheat protein quality (Garrido-Lestache et al., 2004) and it was even postulated that N splitting is a more effective way to improve wheat quality than the increase of N rates (Xue et al., 2016).

Wieser and Seilmeier (1998), however, noticed that the degree of the N fertilization effects on the grain protein concentration were strongly dependent on the variety and it has been reported that the application of increasing amounts of N fertilizer might, in some cases, have no impact (Garcia-Molina and Barro, 2017), or even a negative effect (Daniel and Triboi, 2000) on the flour α-gliadin content.

# 5. Medical solutions to face CD

Currently, the only solution for CD patients is to adhere to a strict lifelong gluten-free diet (GFD). Such a diet enables the normalization of symptoms in most patients after a mean period of four weeks (Murray et al., 2004). This is a gradual process with an estimated time of 3.8 years to achieve a normal villous height (Rubio-Tapia et al., 2010). However, this adherence to a GFD is not always carefully observed and this may increase the risk of persistent villous atrophy (Lebwohl et al., 2014). Several factors can explain this incomplete compliance to a GFD, such as the expensive prices of substitutes compared to their gluten-containing counterparts, or the isolating aspect of GFD in social events. In addition, a high number of hidden sources of gluten can hamper the strict avoidance of gluten since it can be found in sauces, drug fillers, processed meats as well as on shared food equipment. All these trapped sources of gluten add anxiety to a pathology already hard to handle and most of patients are in demand of new therapeutic approaches to face the disease (Tennyson et al., 2013). Among them, two main strategies are currently being investigated: the first one consists in approaches enabling CD patients to consume existing bread wheat varieties. The second one, detailed in the following chapter, focuses on the search of non-immunogenic alternatives by selecting other non-immunogenic cereals or by breeding wheats.

## 5.1. Approaches to consume existing wheat varieties

### 5.1.1. Assisted digestion of gluten

Given that the principal toxic components of CD are the proline- and glutamine-rich peptides which are resistant to digestion by gastric, pancreatic and intestinal enzymes, one tested strategy consists in using enzymes capable of breaking the bond between proline and glutamine residues like the propyl endopeptidases (PEPs) produced by several microorganisms. It would thus enable to break down epitope peptides and to annihilate their immunogenicity. Such enzymes can work either in the intestinal lumen – delivered in acid-resistant capsule forms – or as a

pretreatment by detoxifying gluten with bacteria-derived peptidases during food processing (Khosla et al., 2005; Di Cagno et al., 2002). The strategy involving endopeptidases has already reached the phase 2 of clinical trials and no serious adverse events were reported. However, while such enzyme therapy could protect patients from hidden sources of gluten, authors underline that it will not be efficient in the digestion of normal dietary gluten intake (Makharia, 2014).

#### 5.1.2. Prevention of passage of immunogenic peptides between epithelial cells

An increase in the intestinal paracellular permeability has been hypothesized as one of the first occurring mechanisms in CD and the zonulin protein, which regulates the epithelial permeability, has been found to be involved. In consequence, molecules like larazotide acetate are being studied to antagonize zonulin via its receptor blockade (Paterson et al., 2007). The safety profile of larazotide acetate was analyzed in a phase 1 trial and it was shown to be comparable to that of placebo. Phase 2 clinical trials were then carried out but they did not highlight any statistically significant reduction of the intestinal permeability.

#### 5.1.3. HLA inhibition

Another idea is to use gliadin antagonist peptides to block the binding grooves of HLA molecules in order to suppress the antigen presentation to T-cells. This could be carried out using gluten peptide analogs which act as tight-binding HLA-DQ2 ligands (Siegel et al., 2007). This strategy is, however, very challenging since it requires to get access to the binding grooves of the HLA molecules and to avoid the rapid degradation of this peptide agents during their delivery. In addition, the interference with the immunosurveillance functions of HLA molecules by blocking them in tissues other than intestine is a critical point of consideration.

### 5.1.4. Inhibition of tTG

The tTG enzyme, catalyzing the conversion of epitope glutamine residues into glutamic acid, plays a major role in the pathogenesis since it enables the appropriate recognition of gluten peptide by HLA-DQ molecules and their presentation to T-cells. Several strategies involving tTG inhibitors are thus under study to block the further cascade of T-cell stimulation (Marrano et al., 2001). The effectiveness of this approach has only been evaluated on simplified biological systems until now.

### 5.1.5. Induction of tolerance to gluten

The possibility of modulating the immune reaction by inducing tolerance to ingested gluten is currently investigated through the development of desensitizing "vaccines" (Brown et al., 2011). The aim is to induce a tolerogenic response to gluten proteins in CD patients but the efficiency of such a strategy has to be confirmed. Moreover, such peptide vaccines can be designed to include only a finite number of immunogenic peptides. The efficacy and long-term safety of such vaccine has also been questioned as gluten-related gastrointestinal side effects have been reported (Brown et al., 2011).

### 5.1.6. Gluten-sequestering polymers

Another solution currently investigated consists in blocking the gluten toxicity thanks to polymeric binders. An illustration of this technique is the poly (hydroxyethyl methacrylate-co-styrene sulfonate) [P(HEMA-co-SS)] which has been shown to complex with α-gliadins in quite a selective way, counteracting thus their toxicity (Liang et al., 2009). However, it is unknown whether similar degrees of peptide inhibition obtained *in vitro* can be reached *in vivo*, since the mechanisms of action of the polymeric binder may be more complex in this latter case. Such sequestering polymer could be used as a supportive therapy in case of inadvertent gluten exposure but not for a complete treatment of CD.

### 5.1.7. Anti-inflammatory drugs

Given that a high number of inflammatory molecules take part in the immunogenic process, another approach is to hinder some of them with anti-inflammatory drugs. Among them, glucocorticoids having an inhibitory effect on both T- and B-cell proliferation have been analyzed (Wahl et al., 1975) but they are generally used in the case of celiac crisis, gliadin shock or refractory CD in order to enhance a better clinical response together with a GFD, rather than for uncomplicated patients. It has however been shown that, despite a reduction of the inflammation, glucocorticoids can also decrease the epithelial cell regeneration (Shalimar et al., 2012). Antibodies raised against cytokines and chemokines like Interferon-γ (Przemioslo et al., 1995), Interleukin-15 (Yokoyama et al., 2009) and Interleukin-10 (Rashtak and Murray, 2012) have also been developed whereas the possibility of blocking the T cell-expressed integrin, enabling the migration of lymphocytes on the site of inflammation, is investigated through the development of blockers (Di Sabatino et al., 2009) and anti-integrin antibody (Ghosh et al., 2003). While some of these drugs have been tested in pilot studies, their pharmacology efficacy has not always been met and some side effects have been pointed out (Mulder et al., 2001).

# 6. Breeding strategies to reduce the accession immunogenicity

## 6.1. Existing celiac-safe species

Some other cereals and pseudocereals than wheat have been reported as safe for CD patients, like rice, maize, sorghum, millet, quinoa, amaranth and buckwheat, and can thus be used as an alternative to wheat in a GFD (Rosell et al., 2014). Even if some of them were shown to increase the nutritional profile (Alvarez-Jubete et al., 2010), the major disadvantage is that their seeds lack gluten storage proteins. In consequence, it must be replaced by mimicking compounds such as starches, gums and hydrocolloids (Gallagher et al., 2004) but the final products may be expensive and of lower tasting and structural quality than the gluten-containing ones.

## 6.2. Plant breeding

Another idea, suggested by several authors (Makharia, 2014; Rashtak and Murray, 2012), is to use related species of wheat since high variations in the immunogenicity of various diploid, tetraploid or hexaploid wheat species have been pointed out. Diploid species representative of A and B genomes as well as tetraploid wheat species (AB genome) were shown to display lower immunogenic properties than hexaploid wheat (Suligoj et al., 2013). This is explained by the absence of D genome which encodes for the most immunogenic CD-related sequences (Salentijn et al., 2009; Van Herpen et al., 2006). However, since bread-making properties are mainly provided by the D genome (Makharia et al., 2014), it is very likely that focusing on diploid or tetraploid species lacking this genome will not provide flours of acceptable technological properties.

In contrast, bread wheat and spelt are both hexaploid (AABBDD genome) and they belong to the same species. It enables thus to cross them easily, providing fertile hybrids, which facilitates the transfer of desirable traits between each other. Bread wheat has been extensively bred to increase features like yield and bread-making

properties, and it has been hypothesized to be associated with an increased immunogenic content. The analysis of ancient and modern wheat varieties, however, did not highlight any clear breeding effect on the immunogenic potential (Ribeiro et al., 2016; Van den Broeck et al., 2010). Even if these intense breeding activities have led to a high number of new varieties, the focus on some features like the bread-making properties and the frequent use of the same parental genotypes led to a relative erosion of the genetic base of the bread wheat utilized diversity (Caballero et al., 2004). In contrast, spelt has been less involved in such breeding programs. It is suspected to have two distinct phylogenetic origins (Nesbitt, 2001; Salamini et al., 2002) and it is grown in a wide range of environmental conditions. Spelt germplasm collections thus display a high genetic diversity which has been poorly studied until now and never characterized for its CD-related immunogenic content. Illustrations of this variability were highlighted for features such as the bread-making qualities and the content in proteins, lipids, micronutrients and fibers (An et al., 2004; Escarnot et al., 2010; Gomez-Becerra et al., 2010; Ruibal-Mendieta et al., 2002). Detailed plant breeding strategies are addressed hereafter.

#### 6.2.1. Traditional breeding

The improvement through plant breeding relies on the identification of genetic variation in interesting traits and on the following transfer of this variation into competitive lines, according to their quality, yield and agronomic performances (Shewry et al., 2016). As already mentioned above, the general “wheat” term includes diploid, tetraploid and hexaploid species, including spelt. Transfers of genes and traits can be achieved through crosses between wild diploid and tetraploid species and their cultivated counterparts, just like for spelt and bread wheat accessions. These transfers can also be carried out between wheat species of different levels of ploidy and it is even possible to resynthetize hexaploid wheats by crossing a tetraploid wheat (AABB genome) and the diploid *Aegilops tauschii* (DD genome) (Ogbonnaya et al., 2013).

In the case of CD, it is important to analyze the distribution of the epitopes among gluten protein classes and their contribution to the immunogenicity according to criteria like the genome from which they are expressed (A, B or D), the breeding status (wild, landrace or cultivar) and the species. Interestingly, naturally occurring variations in the epitope sequences have been reported. While canonical (i. e. intact) epitopes display a full immunogenicity, their allelic variants have been shown to be less immunogenic or even non-immunogenic at all (Mitea et al., 2010). The frequency of these epitope variants is strongly correlated to the genome from which α-gliadins are expressed: B genome sequences are the less immunogenic ones, with almost no canonical epitope expressed. The immunogenicity of A genome α-gliadins can be then qualified as mean, while the immunogenicity of the D genome triggers the strongest immune reaction. The sequences expressed from the D genome can, indeed, display each of the four main CD-related epitopes in the canonical form, as well as the very immunogenic 33-mer peptide (Molberg et al., 2005; Van Herpen et al., 2006). With the aim of lowering the immunogenic content of wheat accessions, avoiding D genome α-gliadin sequences could thus constitute an interesting path. This can be achieved by working with lines displaying a deletion of the chromosomic region where the locus encoding for D genome α-gliadins is located (Van den Broeck et al., 2009). The drawback of this strategy is the loss of dough functionality, mainly determined by D genome sequences, although it could partially be compensated by the addition of related proteins like the avenins from celiac-safe accessions of oat (Van den Broeck et al., 2011). However, the multigenic characteristic of many sequences involved in CD makes traditional breeding strategies challenging. In addition, differences in gene expression levels have been reported according to the genome they belong to (Kawaura et al., 2005) or to the accession (Salentijn et al., 2009), and they must be taken into account as well.

### 6.2.2. Molecular breeding

All the traditional breeding possibilities detailed above should thus enable to develop new varieties with a lower immunogenic content but, because of the multigenic properties of sequences involved in CD, probably not accessions without

any canonical epitope at all. In consequence, since even low levels of canonical epitope are still problematic for CD patients, alternative complementary strategies should be considered when aiming at developing celiac-safe varieties with good technological properties.

Non-targeted mutagenesis is one of the ways to generate novel genetic diversity by the application of chemical agents or radiation to induce mutations. A recently-developed technology named TILLING (Targeting Induced Local Lesions in genomes) enables to generate and identify such mutations. The process includes the seed treatment with a chemical agent like ethyl methane sulphonate (EMS), a polymerase-chain reaction (PCR) and a denaturing high performance liquid chromatography (DHPLC) separation to screen for mutations via the detection of nucleotide mismatches (McCallum et al., 2000). However, the complex genomic structure of the loci containing gluten protein-encoding genes make the use of TILLING to lower the immunogenic potential very challenging since these mutations should occur in every α-gliadin gene copy.

Transgenesis is maybe a more appropriate way to achieve this since it enables to downregulate an entire group of gluten proteins. It can be carried out by using biolistics, *Agrobacterium*-mediated systems or protoplast transformation to introduce specific genes to silence the expression of a target gene. However, this introduction, called transformation, is challenging since the efficiency of transformation and of further cell regeneration is generally very low in bread wheat. One of the major applications of this technology is known as RNA interference (RNAi) since a double-stranded RNA including a sequence complementary to the targeted mRNA provokes its degradation via a two-step post-transcriptional mechanism (Watanabe, 2011). This technology has already enabled the down-regulation of α-gliadins (Becker et al., 2006), ω-gliadins (Altenbach and Allen, 2011), γ-gliadins (Gil-Humanes et al., 2008), all gliadins and LMW glutenin subunits (Gil-Humanes et al., 2010; Wen et al., 2012) in bread wheat as well as C-hordeins in barley (Lange et al., 2007). As a consequence of these down-regulations, a compensatory increase in the expression of other protein

classes was noticed, leading to weaker doughs in some cases and to stronger ones in other cases.

Another promising approach currently investigated in molecular breeding is the development of genome editing technologies. They rely on highly specific nucleases which enable to induce mutations to genomes without leaving any footprint in terms of the presence of foreign DNA. These nucleases are site-directed since they are developed to break DNA at a specific location. Plants own mechanisms may then repair these breaks, but this generally leads to a high frequency of errors, resulting in mutations (Shewry and Tatham, 2016). Three main types of site-directed nucleases are currently being exploited: zing-finger nucleases (ZFNs), transcription activator-like nucleases (TALENs) and clustered regularly interspaced short palindromic repeats nucleases (CRISPR). They are all constituted by a DNA-cutting domain and a DNA-targeting domain. Whereas the DNA recognition domain of ZFNs and TALENs is protein-based, CRISPR nucleases locate the target DNA thanks to a short-guide RNA (gRNA). This recognition domain can thus be engineered to target determined sequences and induce specific mutations in the genome (Jones, 2015). Although transgenesis techniques are required to introduce site-directed nucleases in the genome, the genome editing approach has the advantage that this foreign DNA is then lost due to the segregation and the induced mutations are thus not distinguishable from mutations occurring spontaneously at low frequency. Even though this technology has already been used for example to induce herbicide resistance in canola (Jones, 2015), its application in the framework of CD is promising but it will probably be more complex, given the presence of multiple genes involved in the pathogenesis. It is probable that the application of the CRISPR system will lead to only a certain proportion of mutated CD-related sequences, but subsequent rounds of mutagenizing using specific gRNAs should enable to target the remaining immunogenic sequences.

# 7. Agri-food industry priorities regarding spelt, gluten and CD

Given the importance of CD in terms of the disease incidence and severity, a lot of researches are carried out in the agri-food industry. A good indicator of these efforts lies in patented inventions. A deep analysis of patents linked to the fields of gluten, spelt and celiac disease reveals a high number of duplicated patents. Interestingly, none of them is focused on the breeding of spelt in the framework of CD (Figure 4). After withdrawing duplicated patents, the highest proportion (37%) of the remaining inventions relates to the development of gluten-free products, which confirms the efforts made on existing celiac-safe species approached in the previous chapter. These researches are led on a broad range of plants such as rice, maize, potato, quinoa, teff, Timothy grass, buckwheat, millet, lupin and bean. Some of the inventions provide solution to improve the functional, nutritional and tasting properties of the flour whereas others detail industrial processes.

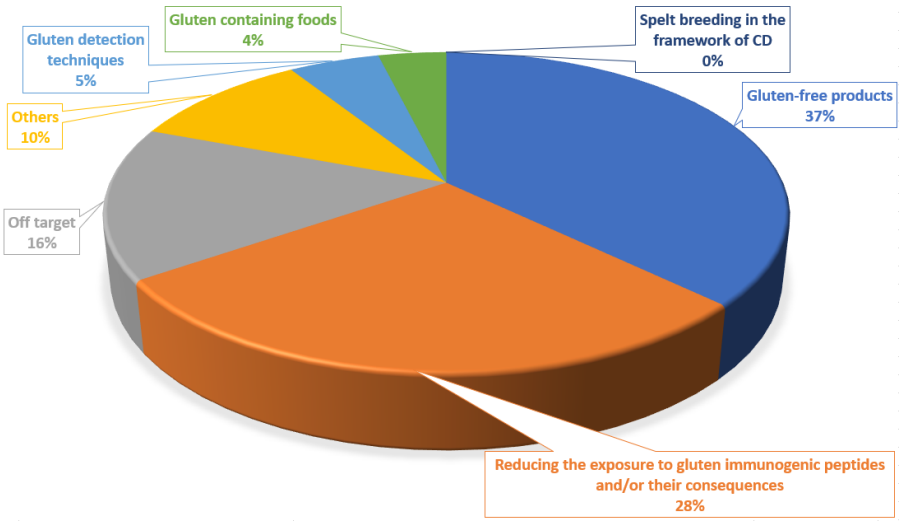



Figure 4. Distribution of patents related to spelt, gluten and celiac disease according to their field of application.

Another field for which the patent analysis gave a high number of occurrences (28%) aims at reducing the exposure to the gluten immunogenic peptides and/or their consequences. They can take the form of different strategies approached in the chapter 5 like gluten-degrading peptidases, immunomodulatory molecules and transglutaminase inhibitors. In addition, the breeding of new varieties displaying mutations in gluten immunogenic peptides is also investigated in some patents.

The development of gluten detection techniques is the subject of another set of inventions (5%). They consist in the identification of novel immunogenic peptides for diagnostic purpose and the development of antibodies to detect them, or in an analytical kit enabling to detect different gluten protein classes.

Surprisingly, the search for inventions linked to gluten, spelt and CD also leads to patents (4%) dealing with gluten-containing foods, without caring about their immunogenic content. Such products consist for example in ready-made meals, or foods where the low content in a component is compensated by an increase in the gluten amount.

Finally, some reported inventions (10%) still belong to the framework of this work but are less clearly related to the research topic, such as the identification of a glutenin allelic variant providing better technological properties to the dough, or new fungal enzymes with different applications, including in the bakery industry. In addition, some patents (16%) are out of the scope of this research and consist, for instance, in the development of an adhesive cream for fixing a dental prothesis or the use of a *Triticum monococcum* extract as active substance in a cosmetic and/or pharmaceutical composition.

In conclusion, the reported patents belong to several fields of application. While some inventions are closely related to the topic of this research, others are not linked to it at all. The majority of the identified patents is dedicated to the development of gluten-free products and to the reduction of the CD patient exposure to gluten immunogenic peptides. Remarkably, none of the reported patents relates to the breeding of spelt in the framework of CD.

# 8. Gluten detection techniques

The breeding and the selection of new varieties for a specific trait require knowledge about the existing variability for this trait. It also requires a tool enabling to track this trait and to guide the breeding process. In the case of CD, complicating features like the multigenic characteristics of sequences involved in this pathology, the high proportion of pseudogenes and the occurrence of epitope allelic variants must be kept in mind.

## 8.1. Methods targeting protein sequences

### 8.1.1. ELISA

The enzyme-linked immunosorbent assay (ELISA) is a technique developed to detect and quantify a specific peptide sequence, named antigen, thanks to antibodies. ELISA kits and different other antibody-based methods targeting gluten epitopes have been developed. These methods can be useful to detect globally gluten proteins in a sample and to quantify the gluten concentration. They are, however, not appropriate to measure the flour immunogenicity since the antibodies used also detect non-immunogenic sequences (Van den Broeck et al., 2015).

ELISA kits have been developed to detect gluten contaminations in food samples and most of them are based on the Skerritt (Skerritt and Hill, 1990), R5 (Valdes et al., 2003), G12 (Moron et al., 2008) and α20 (Mitea et al., 2008) antibodies. The Skerritt antibody was raised against ω-gliadins. It recognizes the sequences PQPQPFPQE and PQQPPFPEE, and reacts with HMW glutenin subunits as well. The R5 antibody is the only one to be approved by the Codex Alimentarius to detect the presence of gluten in food products and to label them “gluten-free” when they contain less than 20 ppm gluten. The R5 antibody was produced against ω-secalins but it recognizes gliadins, secalins and hordeins since it does not specifically binds to only one motif: it detects the peptides QQPFP and the related sequences QQQFP, LQPFP and QLPFP. However, some of these motives do not stimulate the immunity and only one of them (LQPFP) is found in the α-gliadin immunogenic

epitopes (Figure 3). The G12 and A1 antibodies were raised against the strongly immunogenic 33-mer peptide, detecting the sequences QPQLPY and QLPYPQP, respectively. It has however been shown that these last two antibodies also recognize non-immunogenic epitope variants to a lesser extent (Moron et al., 2008). The α20 antibody was recently generated against the sequence RPQQPYP found in a CD-related α-gliadin epitope and it recognizes gliadins, secalins and hordeins. In addition to this broad range of sequences and protein classes recognized by these antibodies, another problem to face when aiming at accurately measuring the immunogenic potential of an accession by ELISA is the short length of the sequences recognized by antibodies. They can be as short as five amino acids whereas the CD-related sequences recognized by T-cell are at least nine amino acids long (Stepniak et al., 2005). This can lead to inaccurate results since an amino acid substitution or deletion suppressing the epitope immunogenicity can be located outside the sequence detected by the antibody and thus lead to an overestimation of the flour immunogenicity.

As previously mentioned, alternative analytical techniques using the same kind of antibodies than ELISA have also been developed but they thus display the same limitations and are not meant to specifically quantify the epitope immunogenicity.

#### 8.1.2. Aptamers

Recently, an original technique has been developed to measure the 33-mer concentration in a sample using single-stranded nucleic acid ligands instead of antibodies (Amaya-Gonzalez et al., 2014). These DNA ligands, called aptamers, were selected through an *in vitro* selection process named systematic evolution of ligands by exponential enrichment (SELEX). The selected aptamer is then used in competitive electrochemical assays where gluten peptide from analyzed samples compete with a fixed amount of 33-mer peptides immobilized on particle beads for binding to the aptamer.

It was demonstrated that the selected aptamer is six times more sensitive than the reference ELISA test and it does not show any cross-reactivity with non-

immunogenic proteins such as those found in maize, soya and rice. This aptamer, however, does not focus on individual epitopes and, despite its high sensitivity, the absence of cross-reactivity with allelic variants of the 33-mer has not been demonstrated.

### 8.1.3. Proteomics methods

Modern proteomics approaches combine a 2D separation, often via chromatography, with mass spectrometry (MS) using soft ionization like matrix-assisted laser desorption/ionization (MALDI) followed by a detection step such as time-of-flight (TOF). The MALDI-TOF technique was the first one to be used for quantification of gluten proteins in food samples without any immunological procedure (Camafeita et al., 1998). However, some weaknesses like insufficient sensitivity and inaccuracies at high mass make the MALDI-TOF strategy applicable only for semi-quantitative analyzes (Ferranti et al., 2007).

A liquid chromatography (LC) separation with electrospray ionization (ESI) and tandem mass spectrometric detection (MS/MS) may overcome these limitations (Scherf and Poms, 2016). Sealey-Voyksner et al. (2010) developed a LC-MS/MS method to quantitate trace levels of six immunogenic gluten marker peptides, whereas Schalk et al. (2018) achieved the quantification of marker peptides in gluten reference proteins in order to calculate gluten concentrations thanks to conversion factors. Van den Broeck et al. (2015) also used nine immunogenic peptides as marker peptides to quantify their concentration in the flour of two bread wheat and one tetraploid accessions. In addition to these peptides, they tested some of their non-immunogenic allelic variants, displaying amino acids substitutions, and they showed that they could be discriminated from canonical peptides. Alternatively, a stable isotope dilution assay (SIDA) has also been used to develop isotopically labelled internal standard enabling to quantitate the amount of 33-mer in bread wheat and spelt flours (Schalk et al., 2017).

The LC-MS/MS approach is thus a promising tool to investigate the immunogenic potential held in cereal flours given its high selectivity, sensitivity and its ability to

detect several immunogenic peptides in the same run. However, it also displays some weaknesses that must be taken into account. Since it requires expertise and expensive equipment only available in specialized laboratories, LC-MS/MS has not been routinely used for the analysis of gluten peptides and it is thus recommended as complementary to other techniques in case of contradictory results (Scherf and Poms, 2016). In addition, the peptide identification following MS requires consultation of databases which are not well curated and generally contain unreviewed amino acid sequences in the case of gluten peptides.

### 8.2. Methods targeting DNA sequences

As already mentioned, epitope variants were shown to display reduced or suppressed immunogenicity compared to the canonical epitope and most of these variants differ from the canonical form by only one nucleotide. Such nucleotide variation is called single nucleotide polymorphism (SNP). However, no SNP detection technique applied to gluten in the framework of CD has been developed until now. With the aim of developing such technique, several strategies are available (for a review on this topic, see Shen et al., 2015). Most of them generally includes a PCR amplification, enabling to increase the number of molecules to analyze and to introduce specificity in some cases (Kim and Misra, 2007).

These strategies rely on one or several of the following properties of the DNA and of the enzymes enabling to handle it: (i) the specificity of the DNA polymerase when adding complementary nucleotides in the PCR reaction, which is used among others in every sequencing technique (*e.g*. pyrosequencing). (ii) The DNA enzymatic cleavage: the specificity of the detection comes from the ability of some enzymes, called restriction enzymes, to recognize a specific DNA motif and to break it. If the targeted SNP is located inside the DNA sequence recognized by the restriction enzyme, such technique can be applied to discriminate allelic variants since the enzyme indeed cleaves only one of the alleles. (iii) The complementary hybridization of two DNA strands: the difference in stability of double-stranded DNA displayed by perfectly matched and mismatched target/probe pairs can be used to discriminate allelic variants by different ways. The use of endonucleases

cleaving mismatched DNA, like the CEL I enzyme, is one solution (Oleykowski et al., 1998). Alternatively, some probes are able to discriminate allelic variants differing by one SNP. The effectiveness of the allelic discrimination depends on the probe length and composition, the localization of the SNP in the probe and the hybridization conditions (Malkki and Petersdorf, 2012). Hybridization systems exist either as individual probe or as microarrays where a high number of different probes are synthetized and immobilized on a solid surface. It is also possible to make use of the complementary hybridization to discriminate allelic variants by using PCR specific primers where the sequence of one of them displays the targeted SNP.

Among all these strategies to genotype SNPs, one technique stands out from the others since it should be particularly adapted to the framework of CD and to the multigenic character of α-gliadins. TaqMan probes relies on the hybridization strategy detailed above and it is especially useful to genotype SNPs with high accuracy. It offers the possibility of multiplexing through the use of several probes labelled with different fluorophores, and it is a relatively cheap approach compared to other existing techniques. In addition, the major advantage of TaqMan probes is that they enable to achieve an accurate quantification of allelic transcripts and to cover several close SNPs with the same probe. This property is particularly useful when aiming at discriminating allelic variants of α-gliadin CD-related epitopes since the SNP location differs from one epitope allelic variant to another (Table 1).

TaqMan probes are classically 18-20 bp long and are labelled with a fluorescent dye at the 5' end and a quencher at the 3' end. In this intact configuration, the proximity of the fluorescent dye and the quencher avoid any fluorescence to be emitted. If the target sequence, perfectly complementary to the TaqMan probe, is present in the analyzed sample, the probe will hybridize to the target sequence during a quantitative PCR (qPCR) reaction and the DNA polymerase will then degrade the hybridized TaqMan probe during the second strand synthesis. This probe degradation moves the fluorescent dye away from the quencher, which causes the emission of a fluorescent signal (Figure 5a). The moment, measured as a

number of PCR cycles, when the fluorescent signal crosses an arbitrarily placed threshold enables to measure the initial amount of target sequences in the sample. Indeed, the more present is the target sequence in the sample, the sooner the fluorescent signal crosses the threshold (Figure 5b).

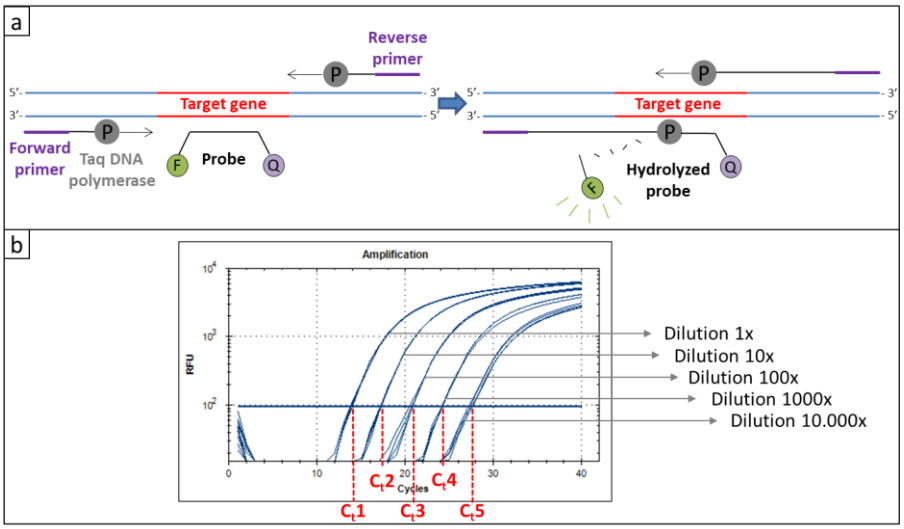



Figure 5. Illustration of TaqMan probe assay to quantify the expression level of a target gene. (a) The TaqMan probe is designed to hybridize to a target sequence; thanks to their high specificity, TaqMan probes are able to discriminate allelic variants differing by only one SNP. The Taq DNA polymerase degrades the hybridized probe during the second strand synthesis leading to the emission of a fluorescence signal which is further measured. (b) The number of PCR cycles after which the fluorescence signal crosses an arbitrarily placed threshold, called cycle threshold ($C_t$), enables to determine the initial amount of the targeted sequence in the sample.

## References

Abdel-Aal, E. and Hucl, P. (2005). Spelt: a specialty wheat for emerging food uses, In: Abdel-Aal, E. and Wood, P. Specialty grains for food and feed. American Association of Cereal Chemists, St-Paul, Minnesota, pp 109-141.

Altenbach, S. B., and Allen, P. V. (2011). Transformation of the US bread wheat 'Butte 86' and silencing of omega-5 gliadin genes. GM Crops 2, 66–73. doi:10.4161/gmcr.2.1.15884.

Altenbach, S. B., Tanaka, C. K., Hurkman, W. J., Whitehand, L. C., Vensel, W. H., and Dupont, F. M. (2011). Differential effects of a post-anthesis fertilizer regimen on the wheat flour proteome determined by quantitative 2-DE. Proteome Science 9, 46. doi:10.1186/1477-5956-9-46.

Alvarez-Jubete, L., Arendt, E. K., and Gallagher, E. (2010). Nutritive value of pseudocereals and their increasing use as functional gluten-free ingredients. Trends in Food Science & Technology 21, 106–113. doi:10.1016/j.tifs.2009.10.014.

Amaya-González, S., de-los-Santos-Álvarez, N., Miranda-Ordieres, A. J., and Lobo-Castañón, M. J. (2014). Aptamer Binding to Celiac Disease-Triggering Hydrophobic Proteins: A Sensitive Gluten Detection Approach. Analytical Chemistry 86, 2733–2739. doi:10.1021/ac404151n.

An, X., Li, Q., Yan, Y., Xiao, Y., Hsam, S. L. K., and Zeller, F. J. (2006). Genetic diversity of European spelt wheat (Triticum aestivum ssp. spelta L. em. Thell.) revealed by glutenin subunit variations at the Glu-1 and Glu-3 loci. Euphytica 146, 193–201. doi:10.1007/s10681-005-9002-6.

Anderson, O. D., and Greene, F. C. The -gliadin gene family. II. DNA and protein sequence variation, subfamily structure, and origins of pseudogenes. Theoretical and Applied Genetics 95, 59-65.

Anderson, O. D., Litts, J. C., and Greene, F. C. The -gliadin gene family. I. Characterization of ten new wheat -gliadin genomic clones, evidence for limited

sequence conservation of flanking DNA, and Southern analysis of the gene family. Theoretical and Applied Genetics 95, 50-58.

Arentz–Hansen, H., Mcadam, S. N., Molberg, Ø., Fleckenstein, B., Lundin, K. E. A., Jørgensen, T. J. D., et al. (2002). Celiac lesion T cells recognize epitopes that cluster in regions of gliadins rich in proline residues. Gastroenterology 123, 803–809. doi:10.1053/gast.2002.35381.

Aziz, I., Lewis, N. R., Hadjivassiliou, M., Winfield, S. N., Rugg, N., Kelsall, A., et al. (2014). A UK study assessing the population prevalence of self-reported gluten sensitivity and referral characteristics to secondary care: European Journal of Gastroenterology & Hepatology 26, 33–39. doi:10.1097/01.meg.0000435546.87251.f7.

Battais, F., Pineau, F., Popineau, Y., Aparicio, C., Kanny, G., Guerin, L., et al. (2003). Food allergy to wheat: identification of immunogloglin E and immunoglobulin G-binding proteins with sequential extracts and purified proteins from wheat flour. Clinical Experimental Allergy 33, 962–970. doi:10.1046/j.1365-2222.2003.01592.x.

Becker, D., Folck, A., Knies, P., Lorz, H., Wieser, H., 2006. Silencing the a-gliadins in hexaploid bread wheat. In: Lookhart, G. L., Ng, P. K. W. (Eds.), Gluten Proteins. A.A.C.C. International, St. Paul, M.N, pp. 86-89.

Bevan, S., Popat, S., Marossy, A. M., Houlston, R. S., Braegger, C. P., Busch, A., et al. Contribution of the MHC region to the familial risk of coeliac disease. Journal of Medical Genetics 36, 687-690.

Biesiekierski, J. R., Newnham, E. D., Irving, P. M., Barrett, J. S., Haines, M., Doecke, J. D., et al. (2011). Gluten Causes Gastrointestinal Symptoms in Subjects Without Celiac Disease: A Double-Blind Randomized Placebo-Controlled Trial. The American Journal of Gastroenterology 106, 508–514. doi:10.1038/ajg.2010.487.

Biesiekierski, J. R., Peters, S. L., Newnham, E. D., Rosella, O., Muir, J. G., and Gibson, P. R. (2013). No Effects of Gluten in Patients With Self-Reported Non-Celiac Gluten

Sensitivity After Dietary Reduction of Fermentable, Poorly Absorbed, Short-Chain Carbohydrates. Gastroenterology 145, 320-328. doi:10.1053/j.gastro.2013.04.051.

Blatter, R. H. E., Jacomet, S., and Schlumbaum, A. (2004). About the origin of European spelt (Triticum spelta L.): allelic differentiation of the HMW Glutenin B1-1 and A1-2 subunit genes. Theoretical and Applied Genetics 108, 360-367. doi:10.1007/s00122-003-1441-7.

Brenchley, R., Spannagl, M., Pfeifer, M., Barker, G. L. A., D'Amore, R., Allen, A. M. et al. (2012). Analysis of the bread wheat genome using whole-genome shotgun sequencing. Nature 491, 705-710. doi:10.1038/nature11650.

Brown, G. J., Daveson, J., Marjason, J. K., Ffrench, R. A., Smith, D., Sullivan, M., et al. (2011). A Phase I Study to Determine Safety, Tolerability and Bioactivity of Nexvax2® in HLA DQ2+ Volunteers With Celiac Disease Following a Long-Term, Strict Gluten-Free Diet. Gastroenterology 140, S-437-S-438. doi:10.1016/S0016-5085(11)61794-0.

Caballero, L., Martín, L. M., and Alvarez, J. B. (2004). Variation and genetic diversity for gliadins in Spanish spelt wheat accessions. Genetic Resources and Crop Evolution 51, 679–686. doi:10.1023/B:GRES.0000034575.97497.6e.

Camafeita, E., Solís, J., Alfonso, P., López, J. A., Sorell, L., and Méndez, E. (1998). Selective identification by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry of different types of gluten in foods made with cereal mixtures. Journal of Chromatography A 823, 299–306. doi:10.1016/S0021-9673(98)00621-9.

Camarca, A., Anderson, R. P., Mamone, G., Fierro, O., Facchiano, A., Costantini, S., et al. (2009). Intestinal T Cell Responses to Gluten Peptides Are Largely Heterogeneous: Implications for a Peptide-Based Therapy in Celiac Disease. The Journal of Immunology 182, 4158–4166. doi:10.4049/jimmunol.0803181.

Campbell, K.G. (1997). Spelt: Agronomy, genetics and breeding. Plant Breeding Reviews 15, 187-213.

Catassi, C., Bai, J., Bonaz, B., Bouma, G., Calabrò, A., Carroccio, A., et al. (2013). Non-Celiac Gluten Sensitivity: The New Frontier of Gluten Related Disorders. Nutrients 5, 3839–3853. doi:10.3390/nu5103839.

Cianferoni, A. (2016). Wheat allergy: diagnosis and management. Journal of Asthma and Allergy 9, 13-25. doi:10.2147/JAA.S81550.

Ciccocioppo, R., Di Sabatino, A., and Corazza, G. R. (2005). The immune recognition of gluten in coeliac disease. Clinical and Experimental Immunology 140, 408–416. doi:10.1111/j.1365-2249.2005.02783.x.

Comino, I., de Lourdes Moreno, M., Real, A., Rodríguez-Herrera, A., Barro, F., and Sousa, C. (2013). The Gluten-Free Diet: Testing Alternative Cereals Tolerated by Celiac Patients. Nutrients 5, 4250–4268. doi:10.3390/nu5104250.

Cubadda, R. and Marconi, E. (2002). Spelt wheat, In: Belton, P. S. and Taylor, J. R. N. (Eds.) Pseudocereals and less common cereals, Springer, New York, pp 153-176.

Daniel, C., and Triboi, E. (2000). Effects of Temperature and Nitrogen Nutrition on the Grain Composition of Winter Wheat: Effects on Gliadin Content and Composition. Journal of Cereal Science 32, 45–56. doi:10.1006/jcrs.2000.0313.

Di Cagno, R., De Angelis, M., Lavermicocca, P., De Vincenzi, M., Giovannini, C., Faccia, M., et al. (2002). Proteolysis by Sourdough Lactic Acid Bacteria: Effects on Wheat Flour Protein Fractions and Gliadin Peptides Involved in Human Cereal Intolerance. Applied and Environmental Microbiology 68, 623–633. doi:10.1128/AEM.68.2.623-633.2002.

Di Sabatino, A., Rovedatti, L., Rosado, M. M., Carsetti, R., Corazza, G. R., and MacDonald, T. T. (2009). Increased expression of mucosal addressin cell adhesion molecule 1 in the duodenum of patients with active celiac disease is associated with depletion of integrin α4β7-positive T cells in blood. Human Pathology 40, 699–704. doi:10.1016/j.humpath.2008.10.014.

Dubé, C., Rostom, A., Sy, R., Cranney, A., Saloojee, N., Garritty, C., et al. (2005). The prevalence of celiac disease in average-risk and at-risk Western European populations: A systematic review. Gastroenterology 128, S57–S67. doi:10.1053/j.gastro.2005.02.014.

Dvorak, J. and Akhunov, E. D. (2005). Tempos of Gene Locus Deletions and Duplications and Their Relationship to Recombination Rate During Diploid and Polyploid Evolution in the *Aegilops-Triticum* Alliance. Genetics 171, 323–332. doi:10.1534/genetics.105.041632.

Dvorak, J., Deal, K. R., Luo, M.-C., You, F. M., von Borstel, K., and Dehghani, H. (2012). The Origin of Spelt and Free-Threshing Hexaploid Wheat. Journal of Heredity 103, 426–441. doi:10.1093/jhered/esr152.

Elli, L., Branchi, F., Tomba, C., Villalta, D., Norsa, L., Ferretti, F., et al. (2015). Diagnosis of gluten related disorders: Celiac disease, wheat allergy and non-celiac gluten sensitivity. World Journal of Gastroenterology 21, 7110–7119. doi:10.3748/wjg.v21.i23.7110.

Escarnot, E. (2016) “L’épeautre panifiable : plusieurs débouchés pour valoriser la production », In: Duvivier, P. and De Coster, S. (Eds.) La Lettre Paysanne (Novembre 2016), Fugea, Namur, pp 29-30.

Escudero-Hernández, C., Peña, A. S., and Bernardo, D. (2016). Immunogenetic Pathogenesis of Celiac Disease and Non-celiac Gluten Sensitivity. Current Gastroenterology Reports 18, 36. doi:10.1007/s11894-016-0512-2.

Farh, K. K.-H., Marson, A., Zhu, J., Kleinewietfeld, M., Housley, W. J., Beik, S., et al. (2014). Genetic and epigenetic fine mapping of causal autoimmune disease variants. Nature 518, 337–343. doi:10.1038/nature13835.

Federal Statistical Office Germany: https://www.destatis.de/EN/Homepage.html

Ferranti, P., Mamone, G., Picariello, G., and Addeo, F. (2007). Mass spectrometry analysis of gliadins in celiac disease. Journal of Mass Spectrometry 42, 1531–1548. doi:10.1002/jms.1361.

Gallagher, E., Gormley, T., Arendt, E. (2004). Recent advances in the formulation of gluten-free cereal-based products. Trends in Food Science & Technology 15, 143–152. doi:10.1016/j.tifs.2003.09.012.

García-Molina, M. D., and Barro, F. (2017). Characterization of Changes in Gluten Proteins in Low-Gliadin Transgenic Wheat Lines in Response to Application of Different Nitrogen Regimes. Frontiers in Plant Science 8, 257. doi:10.3389/fpls.2017.00257.

Garrido-Lestache, E., López-Bellido, R. J., and López-Bellido, L. (2004). Effect of N rate, timing and splitting and N type on bread-making quality in hard red spring wheat under rainfed Mediterranean conditions. Field Crops Research 85, 213–236. doi:10.1016/S0378-4290(03)00167-9.

Ghosh, S., Goldin, E., Gordon, F.H., Malchow, H.A., Rask-Madsen, J., Rutgeerts, P. et al. (2003). Natalizumab for active Crohn's Disease. New England Journal of Medicine 348, 24-32.

Gil-Humanes, J., Pistón, F., Hernando, A., Alvarez, J. B., Shewry, P. R., and Barro, F. (2008). Silencing of γ-gliadins by RNA interference (RNAi) in bread wheat. Journal of Cereal Science 48, 565–568. doi:10.1016/j.jcs.2008.03.005.

Gil-Humanes, J., Piston, F., Tollefsen, S., Sollid, L. M., and Barro, F. (2010). Effective shutdown in the expression of celiac disease-related wheat gliadin T-cell epitopes by RNA interference. Proceedings of the National Academy of Sciences 107, 17023–17028. doi:10.1073/pnas.1007773107.

Glover, N. M., Redestig, H. and Dessimoz, C. (2016). Homeologs: What are they and how do we infer them? Trends in Plant Science 21, 609-621. doi: 10.1016/j.tplants.2016.02.005.

Gu, Y., Crossman, C., Kong, X., Luo, M., You, F., Coleman-Derr, D., et al. (2004). Genomic organization of the complex α-gliadin gene loci in wheat. Theoretical and Applied Genetics 109, 648-657. doi:10.1007/s00122-004-1672-2.

Hajas, L., Scherf, K. A., Bugyi, Z., Török, K., Schall, E., Köhler, P., et al. (2017). ELISA response and gliadin composition of different wheat cultivars grown in multiple harvest years. Acta Alimentaria 46, 187–195. doi:10.1556/066.2016.0019.

Harberd, N. P., Bartels, D., and Thompson, R. D. (1985). Analysis of the gliadin multigene loci in bread wheat using nullisomic-tetrasomic lines. MGG Molecular & General Genetics 198, 234–242. doi:10.1007/BF00383001.

Huang, S., Sirikhachornkit, A., Su, X., Faris, J., Gill, B., Haselkorn, R., et al. (2002). Genes encoding plastid acetyl-CoA carboxylase and 3-phosphoglycerate kinase of the Triticum/Aegilops complex and the evolutionary history of polyploid wheat. Proceedings of the National Academy of Sciences 99, 8133–8138. doi:10.1073/pnas.072223799.

Inomata, N. (2009). Wheat allergy. Current Opinion in Allergy and Clinical Immunology 9, 238–243. doi:10.1097/ACI.0b013e32832aa5bc.

International Wheat Genome Sequencing Consortium (2014). A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. Science 345. doi: 10.1126/science.1251788.

Jia, J., Zhao, S., Kong, X., Li, Y., Zhao, G., He, W. et al. (2013). *Aegilops tauschii* draft genome sequence reveals repertoire for wheat adaptation. Nature 496, 91-95. doi: 10.1038/nature12028.

Johansson, E., Prieto-Linde, M. L., Svensson, G., and Jnsson, J. . (2003). Influences of cultivar, cultivation year and fertilizer rate on amount of protein groups and amount and size distribution of mono- and polymeric proteins in wheat. The Journal of Agricultural Science 140, 275–284. doi:10.1017/S0021859603003162.

Johansson, E., and Svensson, G. (1998). Variation in bread-making quality: effects of weather parameters on protein concentration and quality in some Swedish wheat cultivars grown during the period 1975-1996. Journal of the Science of Food and Agriculture 78, 109-118. doi:10.1002/(SICI)1097-0010(199809)78:1<109::AID-JSFA92>3.0.CO;2-0.

Johansson, E., and Svensson, G. (1999). Influences of yearly weather variation and fertilizer rate on bread-making quality in Swedish grown wheats containing HMW glutenin subunits 2+12 or 5+10 cultivated during the period 1990–96. The Journal of Agricultural Science 132, 13–22. doi:10.1017/S0021859698006029.

Jones, H. D. (2015). Regulatory uncertainty over genome editing. Nature Plants 1, 14011. doi:10.1038/nplants.2014.11.

Kaur, A., Bains, N. S., Sood, A., Yadav, B., Sharma, P., Kaur, S., et al. (2017). Molecular characterization of α-gliadin gene sequences in Indian wheat cultivars vis-à-vis celiac disease eliciting epitopes. Journal of Plant Biochemistry and Biotechnology 26, 106–112. doi:10.1007/s13562-016-0367-5.

Khosla, C., Gray, G. M., and Sollid, L. M. (2005). Putative Efficacy and Dosage of Prolyl Endopeptidase for Digesting and Detoxifying Gliadin Peptides. Gastroenterology 129, 1362–1363. doi:10.1053/j.gastro.2005.08.044.

Kim, S.Y. (2006). Transglutaminase 2 in inflammation. Frontiers in Bioscience 11, 3026–3035.

Kim, S., and Misra, A. (2007). SNP Genotyping: Technologies and Biomedical Applications. Annual Review of Biomedical Engineering 9, 289–320. doi:10.1146/annurev.bioeng.9.060906.152037.

Kozub, N. A., Boguslavskii, R. L., Sozinov, I. A., Tverdokhleb, Y. V., Xynias, I. N., Blume, Y. B., et al. (2014). Alleles at storage protein loci in Triticum spelta L. accessions and their occurrence in related wheats. Cytology and Genetics 48, 33–41. doi:10.3103/S0095452714010046.

Lange, M., Vincze, E., Wieser, H., Schjoerring, J. K., and Holm, P. B. (2007). Suppression of C-Hordein Synthesis in Barley by Antisense Constructs Results in a More Balanced Amino Acid Composition. Journal of Agricultural and Food Chemistry 55, 6074–6081. doi:10.1021/jf0709505.

Lebwohl, B., Ludvigsson, J. F., and Green, P. H. R. (2015). Celiac disease and non-celiac gluten sensitivity. BMJ 351, h4347. doi:10.1136/bmj.h4347.

Lebwohl, B., Murray, J. A., Rubio-Tapia, A., Green, P. H. R., and Ludvigsson, J. F. (2014). Predictors of persistent villous atrophy in coeliac disease: a population-based study. Alimentary Pharmacology & Therapeutics 39, 488–495. doi:10.1111/apt.12621.

Li, Y., Xin, R., Zhang, D., and Li, S. (2014). Molecular characterization of α-gliadin genes from common wheat cultivar Zhengmai 004 and their role in quality and celiac disease. The Crop Journal 2, 10–21. doi:10.1016/j.cj.2013.11.003.

Liang, L., Pinier, M., Leroux, J.-C., and Subirade, M. (2009). Interaction of α-gliadin with poly(HEMA- co -SS): Structural characterization and biological implication. Biopolymers 91, 169–178. doi:10.1002/bip.21109.

Ling, H.-Q., Zhao, S., Liu, D., Wang, J., Sun, H., Zhang, C. et al. (2013). Draft genome of the wheat A-genome progenitor *Triticum urartu*. Nature 496, 87-90. doi: 10.1038/nature11997.

Ma, Z.-C., Wei, Y.-M., Yan, Z.-H., and Zheng, Y.-L. (2007). Characterization of α-gliadin genes from diploid wheats and the comparative analysis with those from polyploid wheats. Russian Journal of Genetics 43, 1286–1293. doi:10.1134/S1022795407110117.

Maghirang, E. B., Lookhart, G. L., Bean, S. R., Pierce, R. O., Xie, F., Caley, M. S., et al. (2006). Comparison of Quality Characteristics and Breadmaking Functionality of Hard Red Winter and Hard Red Spring Wheat. Cereal Chemistry Journal 83, 520–528. doi:10.1094/CC-83-0520.

Makharia, G. K. (2014). Current and Emerging Therapy for Celiac Disease. Frontiers in Medicine 1, 6. doi:10.3389/fmed.2014.00006.

Malkki, M. and Petersdorf, E. W. (2012). Genotyping of Single Nucleotide Polymorphisms by 5' Nuclease Allelic Discrimination. Methods in Molecular Biology 882, 173–182. doi:10.1007/978-1-61779-842-9_10.

Marietta, E. V., and Murray, J. A. (2012). Testing the safety of alternative wheat species and cultivars for consumption by celiac patients. The American Journal of Clinical Nutrition 96, 1247–1248. doi:10.3945/ajcn.112.051425.

Marrano, C., de Macédo, P., Gagnon, P., Lapierre, D., Gravel, C., and Keillor, J. W. (2001). Synthesis and evaluation of novel dipeptide-bound 1,2,4-Thiadiazoles as irreversible inhibitors of guinea pig liver transglutaminase. Bioorganic & Medicinal Chemistry 9, 3231–3241. doi:10.1016/S0968-0896(01)00228-0.

Martre, P., Jamieson, P. D., Semenov, M. A., Zyskowski, R. F., Porter, J. R., and Triboi, E. (2006). Modelling protein content and composition in relation to crop nitrogen dynamics for wheat. European Journal of Agronomy 25, 138–154. doi:10.1016/j.eja.2006.04.007.

McCallum, C. M., Comai, L., Greene, E. A., and Henikoff, S. (2000). Targeting Induced Local Lesions IN Genomes (TILLING) for Plant Functional Genomics. Plant Physiology 123, 439–442. doi:10.1104/pp.123.2.439.

Mitea, C., Kooy-Winkelaar, Y., van Veelen, P., de Ru, A., Drijfhout, J. W., Koning, F., et al. (2008). Fine specificity of monoclonal antibodies against celiac disease-inducing peptides in the gluteome. The American Journal of Clinical Nutrition 88, 1057–1066. doi:10.1093/ajcn/88.4.1057.

Mitea, C., Salentijn, E. M. J., van Veelen, P., Goryunova, S. V., van der Meer, I. M., van den Broeck, H. C., et al. (2010). A Universal Approach to Eliminate Antigenic Properties of Alpha-Gliadin Peptides in Celiac Disease. PLoS ONE 5, e15637. doi:10.1371/journal.pone.0015637.

Molberg, Ø., Uhlen, A. K., Jensen, T., Flæte, N. S., Fleckenstein, B., Arentz–Hansen, H., et al. (2005). Mapping of gluten T-cell epitopes in the bread wheat ancestors: Implications for celiac disease. Gastroenterology 128, 393–401. doi:10.1053/j.gastro.2004.11.003.

Morón, B., Bethune, M. T., Comino, I., Manyani, H., Ferragud, M., López, M. C., et al. (2008). Toward the Assessment of Food Toxicity for Celiac Patients: Characterization of Monoclonal Antibodies to a Main Immunogenic Gluten Peptide. PLoS ONE 3, e2294. doi:10.1371/journal.pone.0002294.

Mulder, C.J.J., Wahab, P.J., Meijer, J.W.R., Metselaar, E. (2001). A pilot study of recombinant human interleukin-10 in adults with refractory coeliac disease. European Journal of Gastroenterology & Hepatology 13, 1183-1188.

Murray, J. A., Watson, T., Clearman, B., and Mitros, F. (2004). Effect of a gluten-free diet on gastrointestinal symptoms in celiac disease. The American Journal of Clinical Nutrition 79, 669–673. doi:10.1093/ajcn/79.4.669.

Nesbitt, M. (2001). Wheat evolution: integrating archaeological and biological evidence. In: Caligari, P.D.S., Brandham, P.E. (Eds.) Wheat taxonomy: the legacy of John Percival, Linnean Society, London, pp 37-59.

Nistico, L. (2006). Concordance, disease progression, and heritability of coeliac disease in Italian twins. Gut 55, 803–808. doi:10.1136/gut.2005.083964.

Noma, S., Kawaura, K., Hayakawa, K., Abe, C., Tsuge, N., and Ogihara, Y. (2016). Comprehensive molecular characterization of the α/β-gliadin multigene family in hexaploid wheat. Molecular Genetics and Genomics 291, 65–77. doi:10.1007/s00438-015-1086-7.

Office fédéral de la statistique Suisse : https://www.bfs.admin.ch/bfs/fr/home.html

Ogbonnaya, F. C. and Abdalla, O. (2013). Synthetic hexaploids: Harnessing Species of the Primary Gene Pool for Wheat Improvement. In: Goldman, I.L., Michler, C.H., Ortiz, R. (Eds.), Plant Breeding Reviews, Wiley-Blackwell, New Jersey, pp. 35-122.

Okita, T.W., Cheesbrough, V., and Reeves, C. D. (1985). Evolution and Heterogeneity of the α-/β-type and γ-type Gliadin DNA Sequences. The Journal of Biological Chemistry 260, 8203-8213.

Oleykowski, C.A., Bronson Mullins, C.R., Godwin, A.K., Yeung, A.T. (1998). Mutation detection using novel plant endonuclease. Nucleic Acids Research 26, 4597-4602.

Ozuna, C. V., Iehisa, J. C. M., Giménez, M. J., Alvarez, J. B., Sousa, C., and Barro, F. (2015). Diversification of the celiac disease α-gliadin complex in wheat: a 33-mer peptide with six overlapping epitopes, evolved following polyploidization. The Plant Journal 82, 794–805. doi:10.1111/tpj.12851.

Park, S.-H., Wilson, J. D., and Seabourn, B. W. (2009). Starch granule size distribution of hard red winter and hard red spring wheat: Its effects on mixing and breadmaking quality. Journal of Cereal Science 49, 98–105. doi:10.1016/j.jcs.2008.07.011.

Paterson, B. M., Lammers, K. M., Arrieta, M. C., Fasano, A., and Meddings, J. B. (2007). The safety, tolerance, pharmacokinetic and pharmacodynamic effects of single doses of AT-1001 in coeliac disease subjects: a proof of concept study. Alimentary Pharmacology & Therapeutics 26, 757–766. doi:10.1111/j.1365-2036.2007.03413.x.

Przemioslo, R. T., Lundin, K. E. A., Sollid, L. M., Nelufer, J., and Ciclitira, P. J. Histological changes in small bowel mucosa induced by gliadin sensitive T lymphocytes can be blocked by anti-interferon γ-antibody. Gut 36, 874-879.

Qi, P.-F., Wei, Y.-M., Chen, G.-Y., Jiang, Q.-T., Liu, Y.-X., Li, W., et al. (2012). Development of chromosome 6D-specific markers for α-gliadin genes and their use in assessing dynamic changes at the Gli-2 loci. Molecular Breeding 29, 199–208. doi:10.1007/s11032-010-9539-5.

Rashtak, S., and Murray, J. A. (2012). Review article: coeliac disease, new approaches to therapy. Alimentary Pharmacology & Therapeutics 35, 768–781. doi:10.1111/j.1365-2036.2012.05013.x

Reilly, N. R. (2016). The Gluten-Free Diet: Recognizing Fact, Fiction, and Fad. The Journal of Pediatrics 175, 206–210. doi:10.1016/j.jpeds.2016.04.014.

Ribeiro, M., Rodriguez-Quijano, M., Nunes, F. M., Carrillo, J. M., Branlard, G., and Igrejas, G. (2016). New insights into wheat toxicity: Breeding did not seem to contribute to a prevalence of potential celiac disease's immunostimulatory epitopes. Food Chemistry 213, 8–18. doi:10.1016/j.foodchem.2016.06.043.

Rosell, C. M., Barro, F., Sousa, C., and Mena, M. C. (2014). Cereals for developing gluten-free products and analytical tools for gluten detection. Journal of Cereal Science 59, 354–364. doi:10.1016/j.jcs.2013.10.001.

Rubio-Tapia, A., Rahim, M. W., See, J. A., Lahr, B. D., Wu, T.-T., and Murray, J. A. (2010). Mucosal Recovery and Mortality in Adults With Celiac Disease After Treatment With a Gluten-Free Diet. The American Journal of Gastroenterology 105, 1412–1420. doi:10.1038/ajg.2010.10.

Rubio-Tapia A, Ludvigsson JF, Brantner TL, et al. (2012) The prevalence of celiac disease in the United States. The American Journal of Gastroenterology 107, 1538–1544. doi: doi:10.1038/ajg.2012.219

Salamini, F., Özkan, H., Brandolini, A., Schäfer-Pregl, R., and Martin, W. (2002). Genetics and geography of wild cereal domestication in the near east. Nature Reviews Genetics 3, 429–441. doi:10.1038/nrg817.

Salentijn, E. M., Esselink, D. G., Goryunova, S. V., van der Meer, I. M., Gilissen, L. J., and Smulders, M. J. (2013). Quantitative and qualitative differences in celiac disease epitopes among durum wheat varieties identified through deep RNA-amplicon sequencing. BMC Genomics 14, 905. doi:10.1186/1471-2164-14-905.

Salentijn, E. M., Goryunova, S. V., Bas, N., van der Meer, I. M., van den Broeck, H. C., Bastien, T., et al. (2009). Tetraploid and hexaploid wheat varieties reveal large differences in expression of alpha-gliadins from homoeologous Gli-2 loci. BMC Genomics 10, 48. doi:10.1186/1471-2164-10-48.

Sapone, A., Bai, J. C., Ciacci, C., Dolinsek, J., Green, P. H., Hadjivassiliou, M., et al. (2012). Spectrum of gluten-related disorders: consensus on new nomenclature and classification. BMC Medicine 10. doi:10.1186/1741-7015-10-13.

Schalk, K., Koehler, P., and Scherf, K. A. (2018). Targeted liquid chromatography tandem mass spectrometry to quantitate wheat gluten using well-defined reference proteins. PLoS One 13, e0192804. doi:10.1371/journal.pone.0192804.

Schalk, K., Lang, C., Wieser, H., Koehler, P., and Scherf, K. A. (2017). Quantitation of the immunodominant 33-mer peptide from α-gliadin in wheat flours by liquid chromatography tandem mass spectrometry. Scientific Reports 7, 45092. doi:10.1038/srep45092.

Scherf, K. A., Koehler, P., and Wieser, H. (2016). Gluten and wheat sensitivities – An overview. Journal of Cereal Science 67, 2-11. doi: 10.1016/j.jcs.2015.07.008

Scherf, K. A., and Poms, R. E. (2016). Recent developments in analytical methods for tracing gluten. Journal of Cereal Science 67, 112–122. doi:10.1016/j.jcs.2015.08.006.

Sealey-Voyksner, J. A., Khosla, C., Voyksner, R. D., and Jorgenson, J. W. (2010). Novel aspects of quantitation of immunogenic wheat gluten peptides by liquid chromatography–mass spectrometry/mass spectrometry. Journal of Chromatography A 1217, 4167–4183. doi:10.1016/j.chroma.2010.01.067.

Seilmeier, W., Belitz, H.-D., and Wieser, H. (1991). Separation and quantitative determination of high-molecular-weight subunits of glutenin from different wheat varieties and genetic variants of the variety Sicco. Zeitschrift für Lebensmittel-Untersuchung und-Forschung 192, 124–129. doi:10.1007/BF01202625.

Shalimar, Das, P., Sreenivas, V., Gupta, S.D., Panda, S.K., Makharia, G.K. (2012). Effect of addition of short course of prednisolone to gluten-free diet on mucosal epithelial cell regeneration and apoptosis in celiac disease: a pilot randomized controlled trial. Digestive Diseases and Sciences 57, 3116-3125. doi:10.1007/s10620-012-2294-1

Shen, W., Tian, Y., Ran, T., Gao, Z. (2015). Genotyping and quantification techniques for single-nucleotide polymorphisms. Trends in Analytical Chemistry 69, 1-13. doi: 10.1016/j.trac.2015.03.008.

Shepherd, S. J., Halmos, E., and Glance, S. (2014). The role of FODMAPs in irritable bowel syndrome: Current Opinion in Clinical Nutrition and Metabolic Care 17, 605–609. doi:10.1097/MCO.0000000000000116.

Shewry, P. R., and Tatham, A. S. (2016). Improving wheat to remove coeliac epitopes but retain functionality. Journal of Cereal Science 67, 12–21. doi:10.1016/j.jcs.2015.06.005.

Shewry, P. R., Halford, N. G., and Lafiandra, D. (2003). Genetics of wheat gluten proteins. Advances in Genetics 49, 111–84.

Siegel, M., Xia, J., and Khosla, C. (2007). Structure-based design of α-amido aldehyde containing gluten peptide analogues as modulators of HLA-DQ2 and transglutaminase 2. Bioorganic & Medicinal Chemistry 15, 6253–6261. doi:10.1016/j.bmc.2007.06.020.

Skerritt, J. H. and Hill, A. S. (1990). Monoclonal Antibody Sandwich Enzyme Immunoassays for Determination of Gluten in Foods. Journal of Agricultural and Food Chemistry 38, 1771-1778.

Sobczyk, A., Pycia, K., Jaworska, G., and Kaszuba, J. (2017). Comparison of fermentation strength of the flours obtained from the grain of old varieties and modern breeding lines of spelt (Triticum aestivum ssp. spelta ). Journal of Food Processing and Preservation 41, e13293. doi:10.1111/jfpp.13293.

Sollid, L. M. (1989). Evidence for a primary association of celiac disease to a particular HLA-DQ alpha/beta heterodimer. Journal of Experimental Medicine 169, 345-350. doi:10.1084/jem.169.1.345.

Statbel: https://statbel.fgov.be/fr/themes/agriculture-peche/exploitations-agricoles-et-horticoles#figures

Statistics Austria: https://www.statistik.at/web_en/statistics/index.html

Stepniak, D., Vader, L. W., Kooy, Y., van Veelen, P. A., Moustakas, A., Papandreou, N. A., et al. (2005). T-cell recognition of HLA-DQ2-bound gluten peptides can be influenced by an N-terminal proline at p-1. Immunogenetics 57, 8–15. doi:10.1007/s00251-005-0780-8.

Suligoj, T., Gregorini, A., Colomba, M., Ellis, H. J., and Ciclitira, P. J. (2013). Evaluation of the safety of ancient strains of wheat in coeliac disease reveals heterogeneous small intestinal T cell responses suggestive of coeliac toxicity. Clinical Nutrition 32, 1043–1049. doi:10.1016/j.clnu.2013.02.003.

Tatham, A. S., and Shewry, P. R. (2008). Allergy to wheat and related cereals. Clinical & Experimental Allergy 38, 1712-1726. doi:10.1111/j.1365-2222.2008.03101.x.

Tennyson, C. A., Simpson, S., Lebwohl, B., Lewis, S., and Green, P. H. R. (2013). Interest in medical therapy for celiac disease. Therapeutic Advances in Gastroenterology 6, 358-364. doi:10.1177/1756283X13492580.

Topper A. Non-celiacs Drive Gluten-Free Market Growth. Mintel Group Ltd. Web. http://www.mintel.com/blog/food-market-news/gluten-freeconsumption-trends. Accessed April 30, 2018.

Triboi, E. (2003). Environmentally-induced changes in protein composition in developing grains of wheat are related to changes in total protein content. Journal of Experimental Botany 54, 1731–1742. doi:10.1093/jxb/erg183.

Trynka, G., Hunt, K. A., Bockett, N. A., et al. (2011). Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease. Nature Genetics 43, 1193–1201. doi:10.1038/ng.998.

Vader, L. W., Stepniak, D. T., Bunnik, E. M., Kooy, Y. M. C., Haan, W. D., Drijfhout, J. W., et al. (2003). Characterization of Cereal Toxicity for Celiac Disease Patients Based on Protein Homology in Grains. Gastroenterology 125, 1105-1113.

Valdes, I., García, E., Llorente, M., and Méndez, E. (2003). Innovative approach to low-level gluten determination in foods using a novel sandwich enzyme-linked immunosorbent assay protocol. European Journal of Gastroenterology & Hepatology 15, 465–747. doi:10.1097/01.meg.0000059119.41030.df.

Van den Broeck, H. C., Cordewener, J. H. G., Nessen, M. A., America, A. H. P., and van der Meer, I. M. (2015). Label free targeted detection and quantification of celiac disease immunogenic epitopes by mass spectrometry. Journal of Chromatography A 1391, 60–71. doi:10.1016/j.chroma.2015.02.070.

Van den Broeck, H. C., de Jong, H. C., Salentijn, E. M. J., Dekking, L., Bosch, D., Hamer, R. J., et al. (2010). Presence of celiac disease epitopes in modern and old hexaploid wheat varieties: wheat breeding may have contributed to increased prevalence of celiac disease. Theoretical and Applied Genetics 121, 1527–1539. doi:10.1007/s00122-010-1408-4.

Van den Broeck, H. C., Gilissen, L. J. W. J., Smulders, M. J. M., van der Meer, I. M., and Hamer, R. J. (2011). Dough quality of bread wheat lacking α-gliadins with celiac disease epitopes and addition of celiac-safe avenins to improve dough quality. Journal of Cereal Science 53, 206–216. doi:10.1016/j.jcs.2010.12.004.

Van den Broeck, H. C., van Herpen, T. W., Schuit, C., Salentijn, E. M., Dekking, L., Bosch, D., et al. (2009). Removing celiac disease-related gluten proteins from bread wheat while retaining technological properties: a study with Chinese Spring deletion lines. BMC Plant Biology 9, 41. doi:10.1186/1471-2229-9-41.

Van Heel, D. A., Franke, L., Hunt, K. A., Gwilliam, R., Zhernakova, A., Inouye, M., et al. (2007). A genome-wide association study for celiac disease identifies risk variants in the region harboring IL2 and IL21. Nature Genetics 39, 827–829. doi:10.1038/ng2058.

Van Herpen, T. W. J. M., Goryunova, S. V., van der Schoot, J., Mitreva, M., Salentijn, E., Vorst, O. et al. (2006). Alpha-gliadin genes from the A, B and D genomes of

wheat contain different sets of celiac disease epitopes. BMC Genomics. 7:1. doi:10.1186/1471-2164-7-1

Wahl, S. M., Altman, L. C., Rosenstreich, D. L., and Immunol, J. (2018). Inhibition of in Vitro Lymphokine Synthesis by Glucocorticosteroids. The Journal of Immunology 115, 476-481.

Watanabe, Y. (2011). Overview of plant RNAi. In: Kodama, H., Komamine, A. (Eds.), Methods in Molecular Biology, Humana Press, Totowa, NJ, pp. 1-11.

Wen, S., Wen, N., Pang, J., Langen, G., Brew-Appiah, R. A. T., Mejias, J. H., et al. (2012). Structural genes of wheat and barley 5-methylcytosine DNA glycosylases and their potential applications for human health. Proceedings of the National Academy of Sciences 109, 20543–20548. doi:10.1073/pnas.1217927109.

Wieser, H. (2007). Chemistry of gluten proteins. Food Microbiology 24, 115–119. doi:10.1016/j.fm.2006.07.004.

Wieser, H., and Seilmeier, W. (1998). The influence of nitrogen fertilisation on quantities and proportions of different protein types in wheat flour. Journal of the Science of Food and Agriculture 76, 49–55. doi:10.1002/(SICI)1097-0010(199801)76:1<49::AID-JSFA950>3.0.CO;2-2.

Xie, Q., Mayes, S., and Sparkes, D. L. (2015). Spelt as a Genetic Resource for Yield Component Improvement in Bread Wheat. Crop Science 55, 2753. doi:10.2135/cropsci2014.12.0842.

Xie, Z., Wang, C., Wang, K., Wang, S., Li, X., Zhang, Z., et al. (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, 1239–1251. doi:10.1007/s00122-010-1384-8.

Xue, C., auf'm Erley, G. S., Rossmann, A., Schuster, R., Koehler, P., and Mühling, K.-H. (2016). Split Nitrogen Application Improves Wheat Baking Quality by Influencing

Protein Composition Rather Than Concentration. Frontiers in Plant Science 7. doi:10.3389/fpls.2016.00738.

Yokoyama, S., Watanabe, N., Sato, N., Perera, P.-Y., Filkoski, L., Tanaka, T., et al. (2009). Antibody-mediated blockade of IL-15 reverses the autoimmune intestinal damage in transgenic mice that overexpress IL-15 in enterocytes. Proceedings of the National Academy of Sciences 106, 15849–15854. doi:10.1073/pnas.0908834106.

Zhang, D. L., Gao, A. L., Li, Y. G., Su, Y. R., He, S. B., and Li, S. P. (2014). Molecular Characterization and Phylogenetic Analysis of Novel α-gliadin Genes from *Triticum dicoccoides* Journal of Agricultural Science and Technology 16, 203-215.

Zhang, Y., Luo, G., Liu, D., Wang, D., Yang, W., Sun, J., et al. (2015). Genome-, Transcriptome- and Proteome-Wide Analyses of the Gliadin Gene Families in *Triticum urartu*. PLOS ONE 10, e0131559. doi:10.1371/journal.pone.0131559.

Zuidmeer, L., Goldhahn, K., Rona, R. J., Gislason, D., Madsen, C., Summers, C., et al. (2008). The prevalence of plant food allergies: A systematic review. Journal of Allergy and Clinical Immunology 121, 1210-1218.e4. doi:10.1016/j.jaci.2008.02.019.

# AIMS OF THE RESEARCH

The purpose of this work was to evaluate the potential of spelt [*Triticum aestivum* ssp. *spelta* (L.) Thell.] to respond to the issue of celiac disease by studying its genetic diversity in CD-related expressed epitope sequences.

To achieve this main objective, distinct specific objectives have been defined:

1. Investigate the diversity of α-gliadin expressed genes from spelt compared with bread wheat and diploid species in the Triticeae tribe.

2. Develop and validate a tool to measure the immunogenic content for CD patients held in expressed α-gliadin sequences.

3. Study a large panel of spelt accessions in order to determine the variability in their CD-related immunogenic content.

4. Study the epitope expression stability according to environmental factors.

The first objective was pursued by analyzing 11 contrasted spelt accessions selected through a genetic diversity study based on simple sequence repeat (SSR) markers. α-gliadin sequences expressed by these accessions were cloned and sequenced to study their allelic variation in CD-related epitopes. The subsequent comparison of spelt α-gliadins sequences to those from bread wheat and related diploid species was carried out to look for potential spelt specificities.

The second objective was achieved by developing epitope-specific TaqMan probes that only hybridize to the canonical form of the four main α-gliadin immunogenic epitopes. Reference genes displaying a stable expression were also selected and TaqMan probes targeting them were developed to normalize qPCR data. These epitope- and reference gene-targeting probes were then validated by applying them to cDNA samples from the 11 contrasted spelt accessions and from three diploid species representative of the ancestral genomes of spelt and bread wheat.

The developed TaqMan probes enabled to pursue the third objective by applying them to a wide set of spelt accessions in order to study their epitope expression profile. The putative correlation between this expression and qPCR results from gDNA samples was then analyzed to investigate whether epitope-specific TaqMan probes could directly be used on gDNA.

Finally, the impact of the environment and crop practices on the epitope expression was analyzed by testing the epitope expression stability over consecutive years and by comparing different N fertilization treatments.

# RESULTS

## Chapter 1

Molecular diversity of α-gliadin expressed genes in genetically contrasted spelt (*Triticum aestivum* ssp. *spelta*) accessions and comparison with bread wheat (*T. aestivum* ssp. *aestivum*) and related diploid *Triticum* and *Aegilops* species

This first chapter starts with a genetic diversity analysis based on SSR data in order to select relevant plant materials. Subsequent investigations were carried out with these selected accessions to reach the first objective, by investigating their diversity of α-gliadin expressed genes compared with bread wheat and related diploid species. This chapter provides important information about the allelic variation displayed by the expressed sequences encoding the four main CD-related epitopes found in α-gliadin sequences in spelt. These variations are linked to the genome from which α-gliadin sequences are expressed (A, B or D) and thus provide interesting information about the genome immunogenicity. The phylogeny and potential specificities of spelt α-gliadins compared to those from bread wheat and related diploid *Triticum* and *Aegilops* species is also investigated in this chapter through the comparison of amino acid sequences and a phylogenetic analysis.

This chapter was published as: Dubois, B., Bertin, P., Mingeot, D. (2016). Molecular diversity of α-gliadin expressed genes in genetically contrasted spelt (*Triticum aestivum* ssp. *spelta*) accessions and comparison with bread wheat (*T. aestivum* ssp. *aestivum*) and related diploid *Triticum* and *Aegilops* species. Molecular breeding 36, 152. doi: 10.1007/s11032-016-0569-5

## Abstract

The gluten proteins of cereals such as bread wheat (*Triticum aestivum* ssp. *aestivum*) and spelt (*T. aestivum* ssp. *spelta*) are responsible for celiac disease (CD). The α-gliadins constitute the most immunogenic class of gluten proteins as they include four main T-cell stimulatory epitopes that affect CD patients. Spelt has been less studied than bread wheat and could constitute a source of valuable diversity. The objective of this work was to study the genetic diversity of spelt α-gliadin transcripts and to compare it with those of bread wheat. Genotyping data from 85 spelt accessions obtained with 19 simple sequence repeat (SSR) markers were used to select 11 contrasted accessions, from which 446 full open reading frame α-gliadin genes were cloned and sequenced, which revealed a high allelic diversity. High variations among the accessions were highlighted, in terms of the proportion of α-gliadin sequences from each of the three genomes (A, B and D), and their composition in the four T-cell stimulatory epitopes. An accession from Tajikistan stood out, having a particularly high proportion of α-gliadins from the B genome and a low immunogenic content. Even if no clear separation between spelt and bread wheat sequences was shown, spelt α-gliadins displayed specific features concerning e.g. the frequencies of some amino acid substitutions. Given this observation and the variations in toxicity revealed in the spelt accessions in this study, the high genetic diversity held in spelt germplasm collections could be a valuable resource in the development of safer varieties for CD patients.

## Keywords

spelt, α-gliadin, celiac disease, gluten, genetic diversity

## Introduction

Gluten is the result of denaturation of endosperm storage proteins called prolamins during the dough kneading. These prolamins are water-insoluble proteins found in the seeds of bread wheat (*Triticum aestivum* L. ssp. *aestivum*), spelt [*Triticum aestivum* ssp. *spelta* (L.) Thell.] and durum wheat [*Triticum turgidum* ssp. *durum* (Desf.) Husnot]. Equivalent proteins are also found in other cereals, such as barley and rye. Prolamins are composed of monomeric gliadins and polymeric glutenins (Shewry and Halford, 2002), which can be divided into high-molecular-weight glutenin subunits (HMW-GS) and low-molecular-weight glutenin subunits (LMW-GS). Conventionally, gliadins are classified into α/β-, γ- and ω-gliadins according to their electrophoretic mobility. Gluten proteins mainly determine the functional properties of wheat flour; glutenins are thought to determine dough elasticity, whereas gliadins could determine its viscosity (Shewry et al. 2003).

The ingestion of gluten peptides can lead to three main types of pathologic reactions: allergic (wheat allergy), autoimmune (celiac disease; CD) and possibly immune-mediated non-celiac gluten sensitivity (NCGS), an unclear grouping of patients whose overall state of health improves when gluten is withdrawn from their diet (Mooney et al. 2013). CD is an uncontrolled inflammatory response to partially digested gluten peptides and is triggered by a T-cell activation in the gastrointestinal mucosa. It results in the flattening of intestinal villi and a reduction in its absorptive capacity, leading to such clinical symptoms as diarrhea, bowel pain, fatigue, weight loss, anemia, osteoporosis, headache and growth retardation (Koning et al. 2005; Marsh, 1992; Rashtak and Murray, 2012; Sapone et al. 2012). CD affects genetically predisposed individuals, with a prevalence of about 1% in the human population (Rewers, 2005).

Among the gluten proteins, α-gliadins are the most immunogenic fraction with the strongest T-cell activation (Arentz-Hansen et al. 2002; Camarca et al. 2009; Ciccocioppo et al. 2005; Vader et al. 2003). They display four major T-cell

stimulatory epitopes: the overlapping DQ2.5-glia-α1 and -α2 epitopes (P{F/Y}PQPQLPY and PQPQLPYPQ, respectively), the DQ2.5-glia-α3 epitope (FRPQQPYPQ) and the DQ8-glia-α1 epitope (QGSFQPSQQ) epitopes. Because the DQ2.5-glia-α2 epitope can be duplicated once or twice, the canonical form of the overlapping DQ2.5-glia-α1 epitope shows two variants: DQ2.5-glia-α1a (PFPQPQLPY) and DQ2.5-glia-α1b (PYPQPQLPY). These four epitopes can be displayed in their canonical form (shown above in brackets), as well as with substituted or deleted amino acid residues. Mitea et al. (2010) showed that these mutations reduce or suppress the antigenic properties of the epitope variants. When two duplications of the DQ2.5-glia-α2 epitope occur, this leads to the full 33-mer fragment, displaying three copies of DQ2.5-glia-α2, which is the most immunogenic fragment of α-gliadin sequences (Molberg et al. 2005; Shan et al. 2002). In addition, α-gliadins display the p31-43 peptide, which induces the innate immune response and enhances the T-cell adaptive response (Gianfrani et al. 2005; Maiuri et al. 2003; Stepniak and Koning, 2006).

The α-gliadins constitute the most important class of gliadins as they represent 15-30% of the bread wheat seed proteins (Gu et al. 2004). Encoded by a multigene family, they possess a very high allelic variability and are located at the Gli-2 loci (Gli-A2, Gli-B2 and Gli-D2) on the short arm of the homeologous chromosomes 6A, 6B and 6D, respectively. The haploid genome includes a number of α-gliadin gene copies ranging from 25-35 (Harberd et al. 1985) to 100 (Okita et al. 1985) or even up to 150 (Anderson et al. 1997), depending on the variety.

The development of new cereal varieties that lack immunogenic gluten peptides, but still display good baking properties, constitute one of the new CD therapeutic approaches currently being considered (Rashtak and Murray, 2012). It would therefore be relevant to make use of the high variability existing in bread wheat and its related taxa. Among them, spelt could be particularly interesting because of the high genetic diversity held in spelt germplasm collections (An et al. 2005; Bertin et al. 2004; Caballero et al. 2004). In addition, spelt has been less subject to selection pressure than bread wheat. Selection programs, which have focused,

inter alia, on the improvement of bread wheat baking qualities, have contributed to a decrease in genetic diversity, especially at the level of α-gliadins and their toxic epitope content (Van den Broeck et al. 2010). Spelt was one of the most important cereals in Europe at the beginning of the 20th century, but bread wheat almost completely replaced it because of its better baking qualities, higher yields and lower processing costs (Koening et al. 2015). Spelt, however, has several interesting features, such as high vitamin content and nutrition values, robustness, adaptability to soil and climatic conditions, resistance to diseases and nitrogen use efficiency (Caballero et al. 2004; Campbell, 1997; Kema, 1992). For more than a decade, the popularity of spelt products has been increasing thanks to their pleasant taste and healthy food reputation (Koening et al. 2015; Kozub et al. 2014). Interest in spelt as a crop for organic farming has also increased because of its lower pesticide requirements compared with bread wheat (Kohajdova and Karovicova, 2008; Kozub et al. 2014).

Spelt and bread wheat are both allohexaploids (2n = 6x = 42; AABBDD genome), but they seem to have emerged from distinct hybridization events (Dvorak et al. 2012). The origin of spelt is not yet fully understood, but it seems to have emerged in two different places, one in Iran and one in Europe. Iranian spelt might have emerged, like bread wheat, through hybridization between cultivated emmer [*Triticum turgidum* ssp. *dicoccum* (Schrank ex Schübler) Thell., AABB genome] and *Aegilops tauschii* Cosson (DD genome) whereas European spelt could be the result of a cross between cultivated emmer and hexaploid bread wheat (Blatter et al. 2004; Dvorak et al. 2012; Kozub et al. 2014; Salamini et al. 2002).

The objective of this study was to investigate the diversity of α-gliadin expressed genes from spelt compared with bread wheat and diploid species in the Triticeae tribe based on their α-gliadin amino acid sequence composition. The work involved (i) cloning and sequencing full-ORF α-gliadins from genetically contrasted spelt accessions, (ii) studying the allelic variation in the T-cell stimulatory epitopes, (iii) evaluating the toxicity of spelt accessions by analyzing their canonical epitope composition and (iv) comparing spelt sequences to α-gliadins from bread wheat

and related diploid Triticum and Aegilops species in order to find potential spelt specificities.

## Materials and methods

***Genetic diversity analysis and selection of accessions***

A working collection of 84 spelt accessions, from 23 countries and 4 continents, has been maintained at the Walloon Agricultural Research Center (CRA-W, Belgium, see Online Resource 1). An Iranian accession (CGN06533, Iran77d), thought to originate from ancestors that differ from those of other spelts (Dvorak et al. 2012), was added to this working collection. The microsatellite data from 19 simple sequence repeat (SSR) markers (Bertin et al., 2004) used on the 85 accessions were subjected to the model-based clustering method implemented with Structure software (v2.3.4; Pritchard et al. 2000) in order to infer the optimal number of groups best describing the population structure. The number of clusters (K) was tested from 1 to 20 with 10 iterations per K, each iteration consisting of 100,000 burn-in steps, followed by 100,000 Markov Chain Monte Carlo (MCMC) repetitions. The admixture ancestry model was chosen and the allele frequencies were assumed to be independent. The ΔK statistics developed by Evanno et al. (2005) was calculated using STRUCTURE HARVESTER software (Earl and vonHoldt 2012). CLUMPP software (Jakobsson and Rosenberg 2007) was used to obtain the mean individual Qmatrix. The log probability [LnP(D)] values calculated with Structure software and the ΔK statistics were used to determine the number of clusters that best described the collection structure. One accession was selected in each cluster (10 in total) for all the experiments described here. Based on the results reported by Dvorak et al. (2012) and on the membership coefficients obtained with CLUMPP software (see below), the Iran77d accession was added to this selection.

***Plant materials***

The 11 selected accessions (Table 1) were kindly provided by the United States Department of Agriculture (USDA, Washington, USA), the Vavilov Institute of Plant

Table 1. Total number of pseudogenes and full open reading frames (ORFs) of α-gliadin sequences obtained from 11 contrasted spelt accessions, distribution of the full-ORF sequences and enumeration of the canonical forms of the four T-cell stimulatory epitopes among each genome and each accession.

| Name* | Accession number | Total number of sequences | Pseudogenes | Full ORF | A | B | D | DQ2.5-glia-α1 | | | | DQ2.5-glia-α2 | | | | DQ2.5-glia-α3 | | | | DQ8-glia-α1 | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | | | A | B | D | TOT | A | B | D | TOT | A | B | D | TOT | A | B | D | TOT |
| BEL08 | PI348315 | 42 | 2 | 40 | 11 | 15 | 14 | 9 | 0 | 16 | 25 | 0 | 0 | 16 | 16 | 9 | 0 | 14 | 23 | 0 | 0 | 3 | 3 |
| DK01 | PI361811 | 42 | 0 | 42 | 13 | 19 | 10 | 13 | 0 | 13 | 26 | 0 | 0 | 13 | 13 | 10 | 0 | 7 | 17 | 0 | 0 | 5 | 5 |
| SPA03 | PI348572 | 41 | 1 | 40 | 26 | 8 | 6 | 25 | 0 | 12 | 37 | 0 | 0 | 12 | 12 | 24 | 0 | 6 | 30 | 0 | 0 | 4 | 4 |
| BUL04 | PI295063 | 45 | 2 | 43 | 17 | 15 | 11 | 15 | 0 | 17 | 32 | 0 | 0 | 17 | 17 | 15 | 0 | 9 | 24 | 4 | 0 | 4 | 8 |
| GER11 | PI348114 | 41 | 2 | 39 | 13 | 17 | 9 | 13 | 0 | 11 | 24 | 0 | 0 | 11 | 11 | 12 | 0 | 7 | 19 | 1 | 0 | 3 | 4 |
| GER12 | PI348120 | 36 | 1 | 35 | 18 | 9 | 8 | 16 | 0 | 10 | 26 | 0 | 0 | 10 | 10 | 16 | 0 | 8 | 24 | 1 | 4 | 2 | 7 |
| TAD06 | K52437 | 47 | 1 | 46 | 14 | 27 | 5 | 8 | 0 | 7 | 15 | 0 | 0 | 7 | 7 | 14 | 0 | 5 | 19 | 0 | 0 | 1 | 1 |
| SWI23 | PI347939 | 45 | 4 | 41 | 15 | 17 | 9 | 15 | 0 | 12 | 27 | 0 | 0 | 12 | 12 | 15 | 0 | 7 | 22 | 0 | 2 | 2 | 4 |
| US06 | PI355595 | 38 | 4 | 34 | 21 | 7 | 6 | 14 | 0 | 11 | 25 | 0 | 0 | 11 | 11 | 21 | 0 | 6 | 27 | 0 | 1 | 5 | 6 |
| Iran77d | CGN06533 | 45 | 1 | 44 | 22 | 7 | 15 | 12 | 0 | 22 | 34 | 0 | 0 | 22 | 22 | 22 | 0 | 14 | 36 | 0 | 0 | 8 | 8 |
| IRA03 | CGN12270 | 42 | 0 | 42 | 15 | 11 | 16 | 10 | 0 | 24 | 34 | 0 | 0 | 26 | 26 | 15 | 0 | 10 | 25 | 0 | 0 | 11 | 11 |
| | | **464** | **18** | **446** | **185** | **152** | **109** | **150** | **0** | **155** | **305** | **0** | **0** | **157** | **157** | **173** | **0** | **93** | **266** | **6** | **7** | **48** | **61** |

The 446 full-ORF α-gliadin expressed sequences were checked for the presence of the four T-cell stimulatory epitopes in their canonical forms: P{F/Y}PQPQLPY (DQ2.5-glia-α1), PQPQLPYPQ (DQ2.5-glia-α2), FRPQQPYPQ (DQ2.5-glia-α3) and QGSFQPSQQ (DQ8-glia-α1).

*: Names used for the same accessions in Bertin et al. (2004) except for Iran77d, which was named as in Dvorak et al. (2012).

Genetic Resources (VIR, Saint-Petersburg, Russia) and the Center for Genetic Resources (CGN, Wageningen, The Netherlands). Among these accessions, eight were landraces (BEL08, SPA03, GER11, GER12, TAD06, SWI23, Iran77d and IRA03), two had an uncertain improvement status (DK01 and BUL04) and the final one was breeding material (US06). They were grown in 2014 in field conditions in Belgium and all the immature grains from a self-pollinated ear were harvested 20 days post-anthesis, immediately frozen in liquid nitrogen and stored at -80°C.

***mRNA extraction and RT-PCR***

For each accession, total RNA was extracted from 100 mg seeds using the NucleoSpin® RNA Plant kit (Macherey-Nagel, Germany). The RNA quality was evaluated by a 1% agarose gel electrophoresis and the RNA was quantified by spectrometry. First strand cDNA was synthetized from 250 ng RNA with oligo(dT)18 primer using the RevertAid H Minus First Strand cDNA Synthesis Kit (Thermo Scientific) in a total volume of 20 µl.

***Cloning and sequencing***

The α-gliadin coding sequences were amplified by the specific primers GliFS: 5'-ATGAAGACCTTTCTCATC-3' and GliRS: 5'-GTTRGTACCGAAGATGCC-3'. The reverse primer is degenerated to work with an α-gliadin panel as wide as possible. Despite this precaution, it should however be noted that this could still lead to the amplification of a subset of expressed α-gliadins and thus to an underestimation of the amount of expressed variants. Moreover, the magnitude of this underestimation could vary from one genome to another. The amplification was carried out in 20 µl reaction volume containing 1.25 U Pfu DNA Polymerase (Thermo Scientific), 1 µl cDNA, 2 µl 10x Pfu reaction buffer (with 20 mM MgSO4), 0.2 mM dNTP, 0.25 µM of each primer and nuclease-free water to reach 20 µl. The polymerase chain reaction (PCR) was performed as described by Mitea et al. (2010).

The PCR products were run on a 1% agarose gel, purified with the GeneJET Gel Extraction Kit (Thermo Scientific) and cloned in a pJET 1.2/blunt cloning vector

using the CloneJET PCR Cloning Kit (Thermo Scientific). Chemical competent cells of the E. coli DH5α strain were then transformed and colonies grown after an overnight incubation at 37°C were checked by colony PCR. Subsequently, the PCR products of about 50 positive clones for each spelt accession were sequenced using the Sanger technique (Beckman Coulter Genomics, United Kingdom).

***Sequence sorting and genome assignment***

Given the multigenic character of α-gliadins, there was a risk to obtain chimeric products during the PCR amplification. To avoid it, sequences showing a putative combination of variants were discarded. After withdrawing sequences of poor quality (4), other than α-gliadins (7) or thought to be chimeric (96), the α-gliadin sequences (deposited in GenBank with accession numbers KX173847 through KX174292) were translated into amino acid sequences using BioEdit v7.1.11 (Hall, 1999). The identification of nucleic and amino acid sequences present in more than one copy, as well as the elaboration of clusters grouping identical sequences, were carried out with SeqTools v8.4.042 (http://www.seqtools.dk/). This software was also used to search for homologies with all α-gliadin proteins in GenBank via a BLASTP analysis (date of analysis: 23 October 2015).

The amino acid sequences of the four major T-cell stimulatory epitopes (DQ2.5-glia-α1, DQ2.5-glia-α2, DQ2.5-glia-α3 and DQ8-glia-α1) were investigated in order to assign each sequence to a genome, following Van Herpen et al. (2006). For each accession, this attribution was further confirmed by a phylogenetic analysis. The spelt amino acid sequences were first aligned using ClustalW in MEGA6 (Tamura et al. 2013), together with 67 GenBank sequences: 31 from diploid species (15 from *Triticum urartu* Tumanian ex Gandilyan, five from *Aegilops speltoides* Tausch and 11 from *Aegilops tauschii*) and 36 from *T. aestivum* ssp. *aestivum*, previously assigned to one of the three genomes (13 from chromosome 6A, 12 from 6B and 11 from 6D). These sequences are reported in Online Resource 2a. Neighbor-joining trees were then constructed in MEGA6 based on a distance method (Poisson substitution model), with 1,000 bootstrap replications, uniform rates among sites, homogeneous pattern among lineages and pairwise deletion of gaps and missing

data. In order to compare spelt and bread wheat α-gliadin sequences and to see if spelt characteristics could be pointed out, an overall phylogenetic analysis was conducted in the same way and included all the sequences from the 11 spelt accessions, as well as 210 GenBank α-gliadin sequences from bread wheat and diploid species (see Online Resource 2a and b). The bread wheat sequences included in this analysis came from 31 distinct varieties and corresponded to all bread wheat α-gliadins reported in GenBank after withdrawing pseudogenes and sequences of poor quality.

## Results

***Genetic diversity analysis***

In order to select contrasted spelt accessions, the model-based clustering method in Structure software was applied to the SSR data of an international collection of 85 spelt accessions. The LnP(D) value calculated by Structure was the highest for K = 10. The ΔK statistics (Evanno et al. 2005) showed two clear peaks at K = 2 and 10. The peak at K = 10 being the highest, 10 was therefore assumed as the number of groups that best described the structure of the spelt collection (Figure 1).

Among the 85 accessions, 68 (80%) had a membership coefficient (Q value) higher than 0.9 and the mean Q was equal to 0.91. The composition of some clusters was strongly linked to the geographical provenance of the accessions, e.g. the orange cluster (Figure 1) that grouped all the Spanish accessions, almost exclusively, and the green cluster that included a high proportion of accessions from Eastern Europe. Nine accessions were clearly admixed (Q < 0.7), including Iran77d. One accession was selected in each of the 10 groups and Iran77d, previously assumed to originate from ancestors that differs from those of the other spelts (Dvorak et al. 2012), was added to this selection.

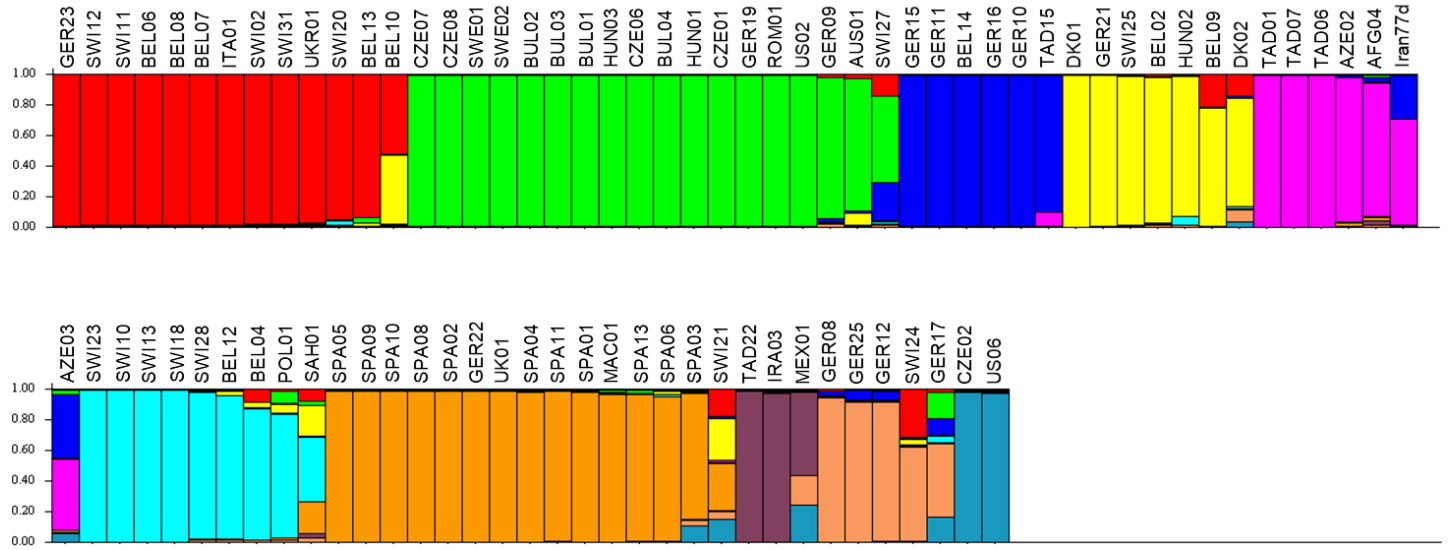



Figure 1. Structure inference in a collection of 85 spelt accessions on the basis of SSR results using Structure software (v2.3.4), clustering for K=10. Microsatellite data from 19 SSR markers (Bertin et al. 2004) were used to perform this analysis. Each color represents one group and each vertical strip corresponds to one accession. The strips are divided into fragments representing the membership proportion to each group

***Cloning and molecular characterization of spelt α-gliadin genes***

In total, 464 α-gliadin expressed sequences were obtained from the 11 spelt accessions, ranging from 36 to 47 sequences per accession (Table 1). Among these, 446 showed a full open reading frame (full ORF). The 18 remaining sequences (3.9%) had at least one premature stop codon (PSC) and were designated as pseudogenes, in line with previous publications (Noma et al. 2015; Van Herpen et al. 2006; Ozuna et al. 2015).

A clustering was carried out on the 446 full-ORF genes to group the identical sequences. This resulted in 260 and 226 different nucleic and amino acid sequences respectively. The 226 amino acid sequences were analyzed using BLASTP for their homology with all α-gliadins from Triticeae species reported in GenBank. Among them, only 26 showed 100% homology with other α-gliadins from Triticeae species.

Most α-gliadins have a typical structure (Online Resource 3), starting with an N-terminal signal peptide, followed by a repetitive domain where three types of CD toxic epitopes are located (DQ2.5-glia-α1, -α2 and -α3), two polyglutamine regions separated by a first unique domain and, finally, a second unique domain at the C-terminal side containing a fourth CD epitope (DQ8-glia-α1). With regard to the sequences obtained in this study, all the 446 α-gliadins displayed this classical structure even when one sequence showed a deletion of almost the entire unique domain I (63 amino acids) and an insertion of eight amino acids at the same location (accession number KX173965). Although showing the typical features, 26 sequences displayed an insertion or deletion in at least one domain.

The typical structure of an α-gliadin also contains six cysteine residues: four in the unique domain I and two in the unique domain II (Online Resource 3). Among the 446 sequences, 427 displayed these cysteines. Seventeen of the 19 remaining α-gliadins had a seventh extra cysteine at the beginning of the second unique domain. A loss of cysteine residues was also observed in two sequences: sequence KX173978 showed five cysteines as the result of a C to Y substitution, whereas

sequence KX173965 had only two cysteines after the deletion of almost the entire unique domain I.

***Genome assignment and T-cell stimulatory epitope diversity***

In order to assign each sequence to its corresponding genome of origin, we used as reference genome-specific amino acid motifs identified by Van Herpen et al. (2006) in wheat diploid species (highlighted in yellow in Figure 2a) in and around the four T-cell stimulatory epitopes DQ2.5-glia-α1, -α2, -α3 and DQ8-glia-α1. All these motifs were found in the spelt sequences in this study (highlighted in yellow in Figure 2b). Some other genome-specific motifs in the same regions, however, were detected in spelt sequences and are highlighted in orange in Figure 2b.

- *DQ2.5-glia-α1 and -α2*

The alignment by genomes of spelt sequences allowed three specific motifs to be detected that had not been reported by Van Herpen et al. (2006): (i) a B-genome specific substitution was found at a high frequency (62.9% of the sequences from the Gli-B2 locus) where the proline amino acid located just after the DQ2.5-glia-α2 epitope was replaced by a threonine (Figure 2b, amino acid at position 112); (ii) a substitution of the glutamine at the last position (Figure 2b, p111) of the DQ2.5-glia-α2 epitope by a histidine (PQPQLPYSH) occurred in some sequences (14.6%) of the A genome, but never in the B and D genomes; and (iii) a D-genome specific substitution of a glutamine by a histidine residue at p113 (Figure 2b) in the second copy of the DQ2.5-glia-α2 epitope (resulting in PQPHLPYPQ) was observed in 11 of the 30 sequences displaying one duplication of this epitope.

- *DQ2.5-glia-α3*

Sequences expressed from the Gli-A2 and Gli-D2 loci are generally not distinguishable from each other because both display the canonical form of the epitope, but we found a substitution of the proline residue at p134 (Figure 2b) by a serine in 13 (11.9%) sequences from the D genome (FRPQQSYPQ).

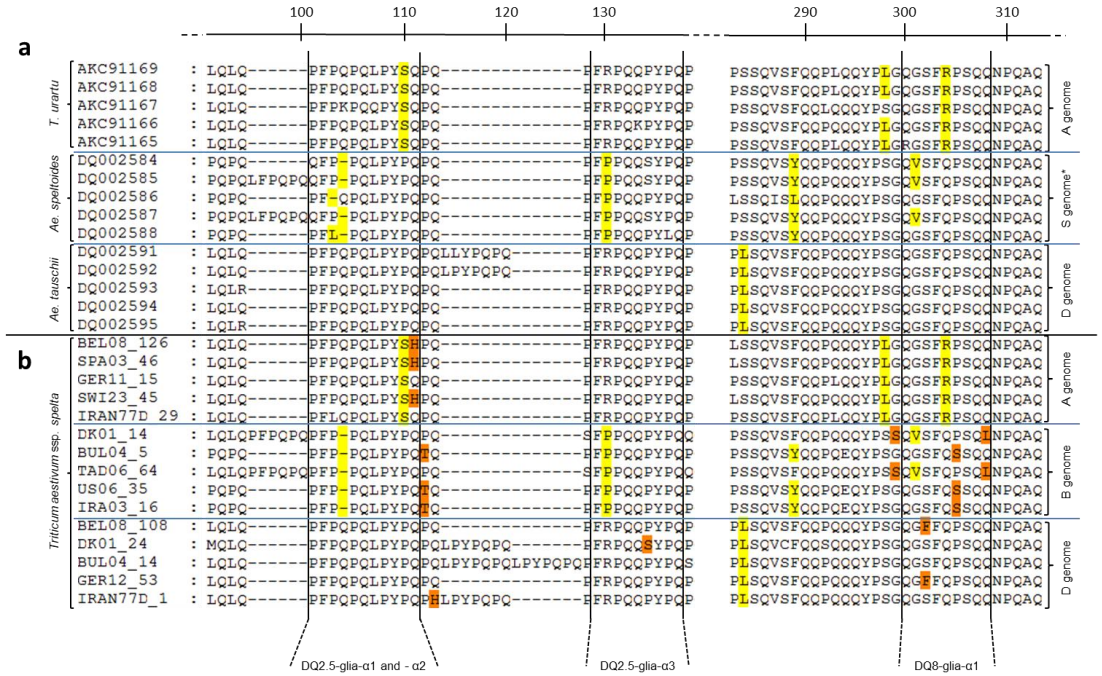



Figure 2. Localization of genome-specific motifs displayed in α-gliadins from (a) diploid species and from (b) spelt. a: For each of the A-, B- and D-genome ancestral species, five α-gliadin sequences selected from GenBank were aligned. b: For each of the three genomes, five representative α-gliadin sequences from spelt obtained in this study were aligned. Yellow residues correspond to genome-specific motifs already identified by Van Herpen et al. (2006) in and around the four major T-cell stimulatory epitopes. Orange residues are new genome-specific motifs discovered in spelt α-gliadins from this study. *: The B genome is hypothesized to be an altered S genome (Von Buren, 2001); *Ae. speltoides* is therefore taken as the closest representative of the B genome

- *DQ8-glia-α1*

We identified two types into which spelt α-gliadins from the B genome could be divided. The first type was characterized by a tyrosine or sometimes a leucine instead of a phenylalanine residue 11 positions before the epitope (p289 in the Figure 2), as described by Van Herpen et al. (2006). Remarkably, this mutation was associated with a proline-to-serine substitution at p305 (QGSFQSSQQ) in 89.4% of the spelt sequences of this type. The second type was characterized by the glycine-to-valine substitution at p301, as reported by Van Herpen (2006). In spelt, this substitution was associated with the replacement of the glutamine residue by a leucine at p308 (QVSFQPSQL) and with a glutamine-to-serine substitution one position before the epitope (p299). Among all the B genome spelt sequences, 73.7% corresponded to the first type ("YG" type) and 25% to the second type ("FV" type). Only two α-gliadin sequences from the American accession (US06) could not be classified according to these types because they displayed both a tyrosine residue at p289 and a valine at p301. We also identified in the spelt sequences a high proportion (52.3%) of α-gliadins from Gli-D2 showing a serine-to-phenylalanine substitution at p302 (QGFFQPSQQ).

For each accession, a neighbor-joining tree resulting from the alignment of the amino acid sequences with α-gliadins of known genomes of origin was constructed (data not shown). Three major groups, corresponding to the A, B and D genomes, were clearly displayed in each phylogenetic tree, confirming the genome assignment of the sequences according to the motifs described by Van Herpen (2006).

***Genomic distribution***

The A, B and D genomes were not equally represented among the 446 full-ORF spelt α-gliadin sequences, with 185 (42%), 152 (34%) and 109 (24%) being counted for the A, B and D genomes, respectively (Table 1). The sequence frequencies in each genome were quite different from one accession to another (Figure 3a). The Spanish (SPA03) and American (US06) accessions, for example, showed the highest

proportion of A genome sequences (more than 60%), the Tajik accession (TAD06) the highest proportion of B genome sequences (about 60%) and the Belgian (BEL08) and both Iranian accessions (Iran77d and IRA03) the highest proportion of sequences from the D genome (35-40%).

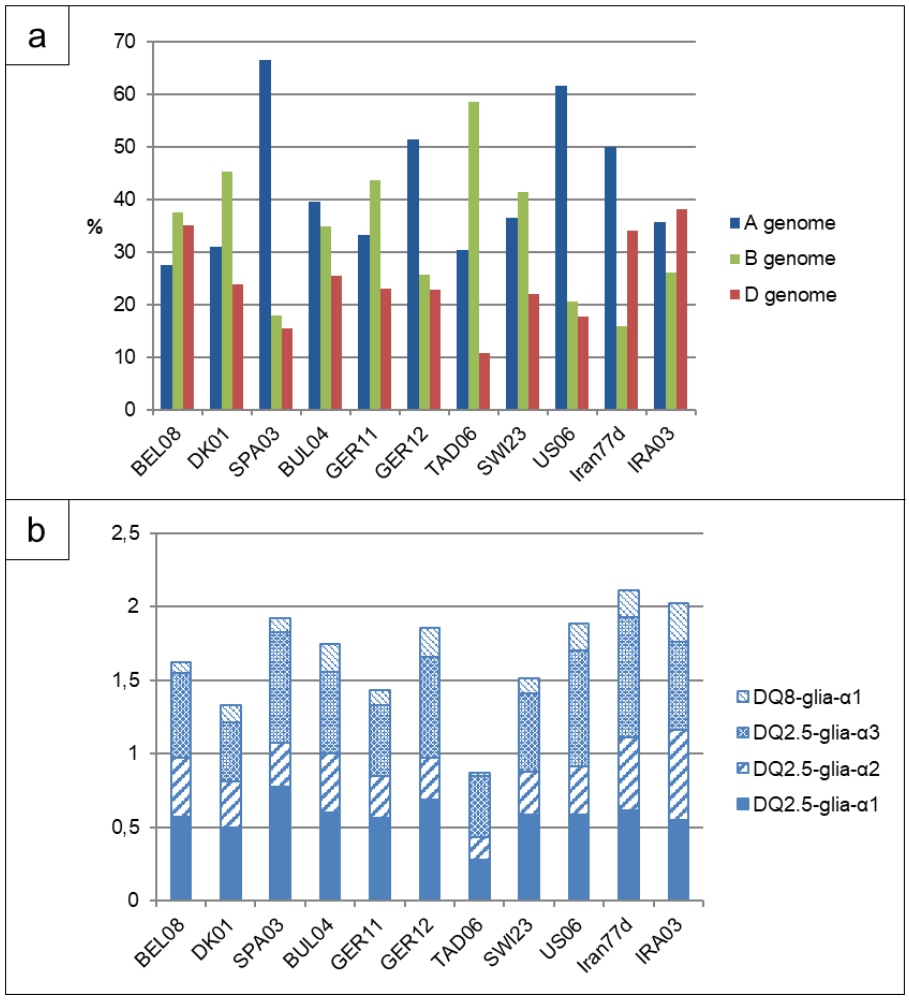



Figure 3. Analysis of α-gliadin transcripts from 11 contrasted spelt accessions: proportion of sequences from the three genomes (a) and average number of canonical epitopes per sequence (b). a: The frequencies were calculated by reporting the number of sequences from each genome to the total number of cloned α-gliadins for each accession. b: Contribution of the four T-cell stimulatory epitopes to the average number of canonical epitopes per sequence

Genome-specific variations in the number of glutamine residues in the two polyglutamine regions (PQI and PQII) have been reported several times in *Triticum* and *Aegilops* species (Li et al. 2012; Li et al. 2013; Van Herpen et al. 2006; Xie et al. 2010). In spelt, significant variations in the length of PQI and PQII regions were observed (Online Resource 4). Overall, the PQI region had a higher average number of glutamine residues than the PQII region and displayed a significantly larger average number of glutamine residues in the α-gliadins from the A genome than from the B and D genomes. In contrast, the PQII region had a significantly lower number of Q-residues in the A genome sequences than the B and D genome sequences. The standard deviation of the mean number of Q-residues in the PQII of the B genome sequences was noticeably high.

***Phylogenetic analysis***

A phylogenetic analysis was performed involving the 226 different spelt amino acid sequences in this study, 31 α-gliadins from diploid species representing the A, B and D genomes (empty triangles in Figure 4) and 179 α-gliadins from bread wheat (empty circles). Three main groups were clearly discernible, corresponding to the three genomes, A, B and D.

The A genome cluster was the largest. Even if no clear separation into spelt, bread wheat and *T. urartu* α-gliadins was seen, some sub-clusters were identified, including exclusively or predominantly sequences from spelt (sub-clusters c, e, g, h, j, k, l), bread wheat (b, f, i) or *T. urartu* (a and d).

In the B genome cluster, a grouping coherent with the YG/FV classification was displayed. The sub-cluster at the top of the figure, predominantly composed of spelt sequences, included only YG-type α-gliadins whereas the middle sub-cluster displayed almost only FV-type spelt and bread wheat α-gliadins. The last sub-cluster at the bottom of the B genome cluster included some YG-type spelt α-gliadins but also bread wheat and *Ae. speltoides* α-gliadins that do not match with this YG/FV classification.

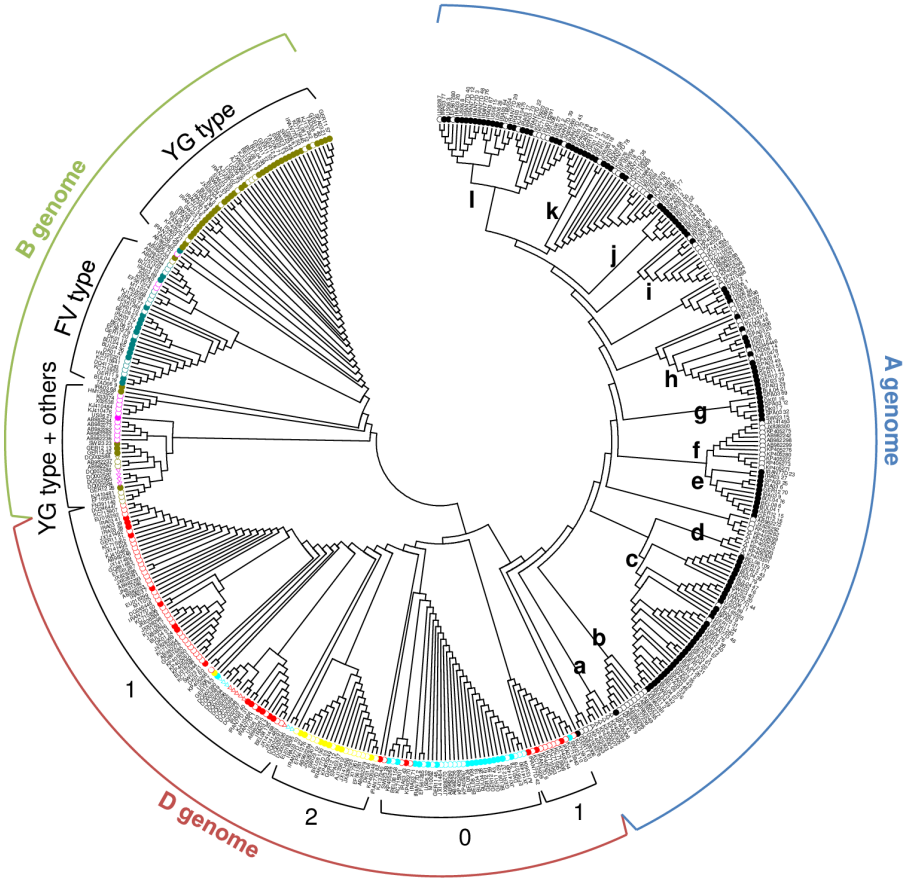



Figure 4. Neighbor-joining tree of 226 spelt α-gliadin amino acid sequences from this study and 210 previously published sequences from diploid species and bread wheat. The neighbor-joining tree is presented in a circular disposition where only the topology is displayed for the sake of clarity. Sequences from diploid species and bread wheat were retrieved from GenBank, and spelt α-gliadins were determined in this study. Alpha-gliadin sequences from diploid species, arising from *Triticum urartu*, *Aegilops speltoides* and *Ae. tauschii*, were labelled by empty triangles, those from bread wheat by empty circles and spelt α-gliadins were labelled by filled circles. Sequences from the A genome were colored black. In the B genome, YG-type and FV-type sequences were colored olive and blue-grey, respectively. Alpha-gliadins matching with neither YG- nor FV-type were colored pink. Spelt sequences from the D genome were marked with a turquoise, red or yellow label when they displayed 0, 1 or 2 duplications of the DQ2.5-glia-α2 epitope, respectively. A full resolution version of this figure is available on the website of the *Molecular Breeding* journal: https://doi.org/10.1007/s11032-016-0569-5

In the D genome cluster, sub-groups corresponding to the number of duplication of the DQ2.5-glia-α2 epitope were observed. Both external sub-clusters on the left and the right parts of the cluster displayed mainly α-gliadins with one duplication. Between them, one sub-cluster included α-gliadins with two duplications, i.e. the full 33-mer sequence, whereas the last one contained mainly α-gliadins without duplication. Among this D genome cluster, spelt and bread wheat α-gliadins were rather homogeneously distributed while *Ae. tauschii* sequences seemed to cluster together.

***Canonical epitope screening and inventory of epitope variants***

Each of the 446 α-gliadin sequences was manually checked for the presence of the four T-cell stimulatory epitopes DQ2.5-glia-α1, -α2, -α3 and DQ8-glia-α1 in their canonical forms: P{F/Y}PQPQLPY, PQPQLPYPQ, FRPQQPYPQ and QGSFQPSQQ, respectively (Table 1). DQ2.5-glia-α1 and -α3 were the most frequent epitopes, followed by DQ2.5-glia-α2 and finally DQ8-glia-α1. This last-mentioned epitope was the only one that was present in its canonical form in each of the three genomes, whereas the three other intact epitopes were systematically absent from the B genome. The DQ2.5-glia-α2 epitope was always mutated in the A and B genome sequences. The D genome was therefore the only one to display the canonical DQ2.5-glia-α2 epitope, but it was still found in relatively high amounts because of its duplication or triplication in some sequences. These duplications and triplications were displayed in 27.5% and 13.8% of the D genome sequences, respectively.

The mean number of canonical epitopes per sequence varied greatly, depending on the accession (Figure 3b). The spelt from Tajikistan TAD06 was clearly distinct from the others, displaying a mean number of canonical epitopes (0.91) significantly smaller than the average of the 10 remaining ones (1.87). The highest values were observed for the Spanish (SPA03) and two Iranian (Iran77d and IRA03) spelt accessions in relation to the large proportion of sequences from the A and D genome, respectively, found in these accessions.

Variants of the DQ2.5-glia-α1, -α2, -α3 and DQ8-glia-α1 epitopes were also searched in the 446 full-ORF sequences, according to their genome of origin (Online Resource 5). The DQ2.5-glia-α1, -α2 and -α3 epitopes displayed 11, 13 and 8 variants, respectively, and the canonical form was always predominant for all of them. Only mutated variants, however, were observed in sequences from the B genome and the mutations consisted either of a residue deletion (DQ2.5-glia-α1 and -α2) or a residue substitution (DQ2.5-glia-α3). The DQ8-glia-α1 epitope showed the lowest diversity, with seven variants that were always the result of a substitution. Its canonical form was preferentially encountered in the D genome α-gliadins, but remarkably it was not the most frequent form as two other variants, almost exclusively found in the A or B genome, appeared in the first and second positions, respectively.

## Discussion

### *Clustering in the spelt collection*

The main objective of this study was to investigate the α-gliadin diversity in spelt through the cloning and sequencing of expressed sequences from contrasted spelt accessions. To this end, we started by studying the genetic diversity of a spelt collection. The analysis with Structure software based on 19 SSR markers led to a clustering of 85 accessions in 10 groups broadly coherent with the spelt geographic provenance. This result was consistent with the findings reported by Bertin et al. (2004), where an unweighted pair-group method with arithmetic averaging (UPGMA)-based dendrogram was generated through the calculation of genetic distances (1 – proportion of shared alleles). Based on the high mean membership coefficient (mean Q = 0.91) and the low number of admixed accessions, we assumed that choosing one accession in each of the 10 clusters would provide a panel that was representative of the spelt diversity. Given that Iran77d does not clearly belong to any of the 10 clusters (Q<0.7) and that it might originate from ancestors that differ from those of the other spelts (Dvorak et al. 2012), it was added to the 10 selected accessions. This panel enabled us to study the diversity of

expressed α-gliadin genes of genetically contrasted spelt accessions, to investigate their allelic variations at the level of four major T-cell stimulatory epitopes, to evaluate the toxicity through their canonical epitope composition and to compare these spelt sequences to α-gliadins from bread wheat and related diploid *Triticum* and *Aegilops* species in order to find potential spelt specificities.

***Molecular characterization of α-gliadin expressed sequences***

The results obtained in this study revealed a high allelic variation among the α-gliadin expressed sequences cloned from the 11 contrasted accessions. We successfully cloned and sequenced 464 α-gliadin expressed sequences corresponding to 226 different complete amino acid sequences. Among these, only 26 displayed 100% homology with other α-gliadins from Triticeae species already reported in GenBank. We therefore provided 200 new α-gliadin sequences from spelt; only 44 spelt α-gliadins had previously been reported. Among these 200 new sequences, a rather high number are unique sequences which could mean that only a subset of all expressed α-gliadins has been amplified. This high diversity is consistent with the multigenic character of the α-gliadin family given that several authors have showed that the duplication of α-gliadin genes led to a great gene copy number with high allelic variation (Anderson et al. 1997; Okita et al. 1985).

Among the 464 sequences, 446 displayed a full ORF, whereas only 18 (3.9%) displayed at least one PSC, and were therefore considered to be pseudogenes. Working on genomic DNA, Anderson and Greene (1997) showed that about half of the α-gliadin genes from bread wheat displayed at least one PSC. More recently, Ozuna et al. (2015) found pseudogene proportions of 39, 76 and 63% in the genomes of diploid, tetraploid and hexaploid wheat species, respectively. Polyploidization might have contributed to this increase in PSC occurrence because the genetic redundancy created by polyploidy can change the dynamics of coding sequence evolution and lead to the accumulation of PSC in duplicated genes (Akhunov et al. 2013; Mighell et al. 2000). The appearance of a PSC usually results from a C-to-T substitution (18 of the 19 substitutions in this study). As much as 20% of the total DNA residues can be methylated in plants and a cytidine methylation at

the 5-position can lead to an incorrect replication as a thymidine, which favors the C-to-T transition (Anderson and Greene, 1997; Gojobori et al. 1982). In this study, cloning the α-gliadin sequences from the transcriptome (cDNA) enabled us to avoid cloning most pseudogenes. The proportion of pseudogenes still observed in the transcript α-gliadins could be explained by the existence of non-functional sequences due to mutations in the protein-coding part while the control elements are maintained, enabling the transcription of the pseudogene (Mighell et al. 2000). The low proportion of sequences displaying PSC (3.9%) could result from the nonsense-mediated mRNA decay (NMD), which is one of several post-transcriptional mechanisms controlling the quality of mRNA function (Maquat, 2004). NMD, also known as mRNA surveillance, eliminates mRNAs displaying PSC in order to prevent the production of potentially deleterious truncated proteins (Maquat, 2004; Mitrovich and Anderson, 2005).

Although most of the full-ORF spelt sequences in this study displayed the classical α-gliadin structure, 19 sequences had an extra cysteine and two sequences had only two and five cysteines. This could have a direct impact on dough quality because six cysteine residues lead to the formation of three intramolecular disulfide bonds which stabilize the compact globular protein fold, whereas additional cysteines enable intermolecular bonds to be created (Khatkar et al. 2002; Shewry and Tatham, 1997). Anderson et al. (2001) and Kasarda (1989) postulated that an odd number of cysteine residues leads to a free cysteine that could participate in gluten polymers.

***T-cell stimulatory epitope diversity***

The expression of the four T-cell stimulatory epitopes DQ2.5-glia-α1, -α2, -α3 and DQ8-glia-α1 in the 11 contrasted accessions showed high diversity in terms of quality (canonical or mutated epitope) and quantity. Phylogenetic analyses with α-gliadins from diploid species enabled the spelt α-gliadins to be grouped into three clusters corresponding to the three genomes, as for bread wheat (Van Herpen et al. 2006). We showed that the genome of origin (A, B or D) greatly influenced the α-gliadin immunogenicity in spelt, as reported by Mitea et al. (2010) for bread wheat.

The 33-mer, involving six overlapping copies of DQ2.5-glia-α1 and DQ2.5-glia-α2, is the most immunogenic fragment of α-gliadin sequences (Molberg et al. 2005; Shan et al. 2002). Alpha-gliadins with these epitope duplications are D-genome specific and the 33-mer fragment is generally found at a low frequency in bread wheat α-gliadins (Molberg et al. 2005; Ozuna et al. 2015). In this study, we noticed the same trend, with only 13.8% of the D-genome spelt α-gliadins containing the full 33-mer fragment. In addition, α-gliadin sequences from the D genome were the only ones to display the four canonical epitopes, as already shown in bread wheat and diploid species (Van Herpen et al. 2006). In bread wheat, α-gliadins expressed from the Gli-A2 locus display two canonical epitopes (DQ2.5-glia-α1 and -α3) and a small proportion of them also contain a canonical DQ8-glia-α1 (Li et al. 2012). Mitea et al. (2010) found that there were no canonical HLA-DQ2.5 T-cell epitopes in B genome α-gliadins from bread wheat and that their mutated substitutes did not display any T-cell stimulatory capacity. Given that these features were reflected in the spelt α-gliadin sequences in our study, a high proportion of expressed α-gliadins from the B genome combined with few sequences from the D genome would be desirable in order to develop new spelt varieties with a reduced CD-immunogenic content. Moreover, Mitea et al. (2010) showed that almost all substitutions occurring in the four major T-cell stimulatory epitopes reduced or suppressed their toxicity, regardless of the genome. They synthetized and tested the toxicity of 14- to 17-mer epitope peptides including two variants highlighted in the figure 2b: the substitution of the proline residue by a threonine one position after the DQ2.5-glia-α2 (P-PQLPYPQT) and the substitution of the serine at the third position of the DQ8-glia-α1 epitope by a phenylalanine (QGFFQPSQQ). Mitea and her colleagues showed that both variants displayed a 1000 times reduced T-cell stimulation compared to the canonical epitope. Such epitope mutations are thus interesting with the aim of lowering the α-gliadin CD-immunogenic content.

Furthermore, we used the mean number of canonical epitopes per α-gliadin sequence as an indicator of the immunogenic content of the 11 spelt accessions and there were great variations, with mean values ranging from 0.87 to 2.11, depending on the accession. Among them, despite a hypothesized difference in

phylogenetic origin, the Iran77d accession did not stand out from the others. The Tajik accession TAD06, in relation to the highest proportion of B-genome α-gliadin transcripts, displayed a low mean number of canonical epitopes. Interestingly, several studies have suggested the existence of two independent origins for spelt (one in Europe and one in Asia) and some genetic differences between them have been reported (Blatter et al. 2004; Dvorak et al. 2012; Jaaska et al. 1978). Thus, the reduced immunogenic content of TAD06 and its geographic origin provide an interesting route for exploring the genetic diversity and the α-gliadin composition of spelt accessions originating from Asia.

***Comparison between spelt and bread wheat α-gliadin sequences***

The phylogenetic analysis based on the spelt α-gliadin sequences in this study and others from bread wheat and diploid species in the Triticeae tribe did not show a clear separation between them. In the A and B genome, sub-clusters showing preferential grouping were however pointed out. Moreover, significant differences in length in the two polyglutamine regions (PQI and PQII) were observed. These regions play an important role in dough properties because large numbers of glutamine side chains can increase the visco-elasticity properties of dough via intermolecular interactions, given that they are both good hydrogen bond donors and acceptors (Masci et al. 2000). Spelt α-gliadins from the A genome had a PQI region with a significantly higher number of glutamine residues than the PQI regions in the B and D genome sequences. We did not, however, observe spelt α-gliadin sequences from a particular genome with significantly larger PQII regions. This does not accord with the findings of previous studies (Li et al. 2012; Li et al. 2013; Van Herpen et al. 2006; Xie et al. 2010) on Triticeae species other than spelt, showing that the B genome sequences had a significantly higher number of Q-residues in the PQII region. The main reason for such a difference not being observed in these spelt α-gliadins lies in the occurrence of two sub-groups in the B genome sequences, the first one with a low number (6 or 7) and the second one with a large number (from 13 to 25) of Q-residues in the PQII region. Interestingly, we also revealed two sub-groups, YG and FV, in the B genome sequences based on

the amino acid patterns located before and in the DQ8-glia-α1 epitope. For 144 out of the 150 B genome sequences (96%), the YG-type classification was systematically associated with a short PQII region, whereas the FV-type was associated with a long PQII region. In addition, the absence of a clear YG-FV dichotomy in *Ae. speltoides* and bread wheat α-gliadins make them distinguishable from spelt α-gliadins.

A previous study also reported that several bread wheat varieties had fewer α-gliadins expressed from the B genome than from the A and D genomes (Kawaura et al. 2005). Spelt accessions in our study did not display this expression pattern, with the mean proportions of spelt α-gliadin transcripts from A, B and D genomes being 42, 34 and 24%, respectively, and the B-genome α-gliadin proportion even reaching 59% in the TAD06 accession. This higher frequency of expressed α-gliadins from the B genome compared with bread wheat suggests that it would be worthwhile paying more attention to spelt in efforts to develop safer varieties for CD patients.

## References

Akhunov ED, Sehgal S, Liang H, Wang S, Akhunova AR, Kaur G, Li W, Forrest KL, See D, Simkova H, Ma Y, Hayden MJ, Luo M, Faris JD, Dolezel J and Gill BS (2013) Comparative analysis of syntenic genes in grass genomes reveals accelerated rates of gene structure and coding sequence evolution in polyploid wheat. Plant Physiol 161:252-265. doi: 10.1104/pp.112.205161

An X, Li Q, Yan Y, Xiao Y, Hsam SLK and Zeller FJ (2005) Genetic diversity of European spelt wheat (Triticum aestivum ssp. spelta L. em. Thell.) revealed by glutenin subunit variations at the Glu-1 and Glu-3 loci. Euphytica 146:193-201. doi: 10.1007/s10681-005-9002-6

Anderson OD, Hsia CC and Torres V (2001) The wheat γ-gliadin genes: characterization of ten new sequences and further understanding of γ-gliadin gene family structure. Theor Appl Genet 103:323-330. doi: 10.1007/s00122-001-0551-3

Anderson OD and Greene FC (1997) The a-gliadin gene family. II. DNA and protein sequence variation, subfamily structure, and origins of pseudogenes. Theor Appl Genet 95:59-65. doi: 10.1007/s001220050532

Anderson OD, Litts JC and Greene FC (1997) The a-gliadin gene family. I. Characterization of ten new wheat a-gliadin genomic clones, evidence for limited sequence conservation of flanking DNA, and Southern analysis of the gene family. Theor Appl Genet 95:50-58. doi: 10.1007/s001220050531

Arentz-Hansen EH, McAdam SN, Molberg O, Kristiansen C and Sollid LM (2000) Production of a panel of recombinant gliadins for the characterization of T cell reactivity in coeliac disease. Gut 46:46-51. doi: 10.1136/gut.46.1.46

Arentz-Hansen H, McAdam SN, Molberg O, Fleckenstein B, Lundin KEA, Jorgensen TJD, Jung G, Roepstorff P and Sollid LM (2002) Celiac lesion T cells recognize epitopes that cluster in regions of gliadins rich in proline residues. Gastroenterology 123:803-809. doi: 10.1053/gast.2002.35381

Bertin P, Grégoire D, Massart S and de Froidmont D (2004) High level of genetic diversity among spelt germplasm revealed by microsatellite markers. Genome 47:1043-1052. doi: 10.1139/G04-065

Blatter RHE, Jacomet S and Schlumbaum A (2004) About the origin of European spelt (Triticum spelta L.): allelic differentiation of the HMW Glutenin B1-1 and A1-2 subunit genes. Theor Appl Genet 108:360-367. doi: 10.1007/s00122-003-1441-7

Caballero L, Martin LM and Alvarez JB (2004) Variation and genetic diversity for gliadins in Spanish spelt wheat accessions. Genet Resour Crop Ev 51:679-686. doi: 10.1023/B:GRES.0000034575.97497.6e

Camarca A, Anderson RP, Mamone G, Flerro O, Facchiano A, Costantini S, Zanzi D, Sidney J, Auricchio S, Sette A, Troncone R and Gianfrani C (2009) Intestinal T cell responses to gluten peptides are largely heterogeneous: implications for a peptide-based therapy in celiac disease. J Immunol 182:4158-4166. doi: 10.4049/jimmunol.0803181

Campbell KG (1997) Spelt: Agronomy, genetics, and breeding. In: Janick J (ed) Plant Breeding Reviews, vol. 15. Wiley, New York, pp 187-213

Ciccocioppo R, Di Sabatino A and Corazza GR (2005) The immune recognition of gluten in coeliac disease. Clin Exp Immunol 140:408-416. doi: 10.1111/j.1365-2249.2005.02783.x

Dvorak J, Deal KR, Luo M-C, You FM, von Borstel K and Dehgani H (2012) The origin of spelt and free-thrishing hexaploid wheat. J Hered 103:426-441. doi: 10.1093/jhered/esr152

Earl DA and vonHoldt BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conserv Genet Resour 4:359-361. doi: 10.1007/s12686-011-9548-7

Evanno G, Regnaut S and Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14:2611-2620. doi: 10.1111/j.1365-294X.2005.02553.x

Garcia-Maroto F, Marana C, Garcia-Olmedo F and Carbonero P (1990) Nucleotide sequence of a cDNA encoding an α/β-type gliadin from hexaploid wheat (Triticum aestivum). Plant Mol Biol 14:867-868. doi: 10.1007/BF00016521

Gianfrani C, Auricchio S and Troncone R (2005) Adaptative and innate immune responses in celiac disease. Immunol Lett 99:141-145. doi: 10.1016/j.imlet.2005.02.017

Gojobori T, Li W-H and Graur D (1982) Patterns of nucleotide substitution in pseudogenes and functional genes. J Mol Evol 18:360-369. doi: 10.1007/BF01733904

Gu YQ, Crossman C, Kong X, Luo M, You FM, Coleman-Derr D, Dubcovsky J and Anderson OD (2004) Genomic organization of the complex α-gliadin gene loci in wheat. Theor Appl Genet 109:648-657. doi: 10.1007/s00122-004-1672-2

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucl Acids Symp Ser 41:95-98. doi: 10.1021/bk-1999-0734.ch008

Harberd NP, Bartels D and Thompson RD (1985) Analysis of the gliadin multigene loci in bread wheat using nullisomic-tetrasomic lines. Mol Gen Genet 198:234-242. doi: 10.1007/BF00383001

Jaaska V (1978) NADP-dependent aromatic alcohol dehydrogenase in polyploid wheats and their diploid relatives. On the origin and phylogeny of polyploid wheats. Theor Appl Genet 53:209-217. doi: 10.1007/BF00277370

Jakobsson M and Rosenberg NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality in analysis

of population structure. Bioinformatics 23:1801-1806. doi: 10.1093/bioinformatics/btm233

Kasarda DD, Okita TW, Bernardin JE, Baecker PA, Nimmo CC, Lew EJ-L, Dietler MD and Greene FC (1984) Nucleic acid (cDNA) and amino acid sequences of α-type gliadins from wheat (Triticum aestivum). Proc Natl Acad Sci USA 81:4712-4716.

Kasarda DD (1989) Glutenin structure in relation to wheat quality. In: Wheat is unique. Am Assoc Cereal Chem 277-302.

Kaur A, Bains NS, Sood A, Yadav B, Sharma P, Kaur S, Garg M, Midha V and Chhuneja P (2016) Molecular characterization of a-gliadin gene sequences in Indian wheat cultivars vis-à-vis celiac disease eliciting epitopes. J Plant Biochem Biot. doi: 10.1007/s13562-016-0367-5

Kawaura K, Mochida K and Ogihara Y (2005) Expression profile of two storage-protein gene families in hexaploid wheat revealed by large-scale analysis of expressed sequence tags. Plant Physiol 139:1870-1880. doi: 10.1104/pp.105.070722

Kema GHJ (1992) Resistance in spelt wheat to yellow rust. III. Phylogenetical considerations. Euphytica 63:225-231.

Khatkar BS, Fido RJ, Tatham AS and Schofield JD (2002) Functional properties of wheat gliadins. II. Effects on dynamic rheological properties of wheat gluten. J Cereal Sci 35:307-313. doi: 10.1006/jcrs.2001.0430

Koening A, Konitzer K, Wieser H and Koehler P (2015) Classification of spelt cultivars based on differences in storage protein compositions from wheat. Food Chem 168:176-182. doi: 10.1016/j.foodchem.2014.07.040

Kohajdova Z and Karovicova J (2008) Nutritional value and baking applications of spelt wheat. Acta Sci Pol, Technol Aliment, 7:5-14.

Koning F, Schuppan D, Cerf-Bensussan N and Sollid LM (2005) Pathomechanisms in celiac disease. Best Pract Res Cl Ga 19:373-387. doi: 10.1016/j.bpg.2005.02.003

Kozub NA, Boguslavskii RL, Sozinov IA, Tverdokhleb YV, Xynias IN, Blume YB and Sozinov AA (2014) Alleles at storage protein loci in Triticum spelta L. accessions and their occurrence in related wheats. Cytol Genet 48:33-41. doi: 10.3103/S0095452714010046

Li J, Wang S, Li S, Ge P, Li X, Ma W, Zeller FJ, Hsam SLK and Yan Y (2012) Variations and classification of toxic epitopes related to celiac disease among α-gliadin genes from four Aegilops genomes. Genome 55:513-521. doi: 10.1139/g2012-038

Li J, Wang S-L, Cao M, Lv D-W, Subburaj S, Li X-H, Zeller FJ, Hsam SLK and Yan Y-M (2013) Cloning, expression, and evolutionary analysis of α-gliadin genes from Triticum and Aegilops genomes. J Appl Genet 54:157-167. doi: 10.1007/s13353-013-0139-z

Li Y, Xin R, Zhang D and Li S (2014) Molecular characterization of α-gliadin genes from common wheat cultivar Zhengmai 004 and their role in quality and celiac disease. Crop J 2:10-21. doi: 10.1016/j.cj.2013.11.003

Maiuri L, Ciacci C, Ricciardelli I, Vacca L, Raia V, Auricchio S, Picard J, Osman M, Quaratino S and Londei M (2003) Association between innate response to gliadin and activation of pathogenic T cells in coeliac disease. Lancet 362:30-37. doi: 10.1016/S0140-6736(03)13803-2

Maquat LE (2004) Nonsense-mediated mRNA decay: splicing, translation and mRNP dynamics. Nat Rev Mol Cell Bio 5:89-99. doi: 10.1038/nrm1310

Marsh MN (1992) Mucosal pathology in gluten sensitivity. In: Marsh MN (ed) Coeliac disease. Blackwell, Oxford, pp136-191.

Masci S, D'Ovidio R and Lafiandra D (2000) A 1B-coded low-molecular-weight glutenin subunit associated with quality in durum wheats shows strong similarity to a subunit present in some bread wheat cultivars. Theor Appl Genet 100:396-400. doi: 10.1007/s001220050052

Mighell AJ, Smith RN, Robinson PA and Markham AF (2000) Vertebrate pseudogenes. FEBS Lett 468:109-114. doi: 10.1016/S0014-5793(00)01199-6

Mitea C, Salentijn EMJ, van Veelen P, Goryunova SV, van der Meer IM, et al. (2010) A universal approach to eliminate antigenic properties of alpha-gliadin peptides in celiac disease. PLoS ONE 5: e15637. doi: 10.1371/journal.pone.0015637

Mitrovich QM and Anderson P (2005) mRNA surveillance of expressed pseudogenes in C. elegans. Curr Biol 15:963-967. doi: 10.1016/j.cub.2005.04.055

Molberg O, Uhlen AK, Jensen T, Flaete NS, Fleckenstein B, Arentz-Hansen H, Raki M, Lundin KEA and Sollid LM (2005) Mapping of gluten T-cell epitopes in the bread wheat ancestors: implications for celiac disease. Gastroenterology 128:393-401. doi: 10.1053/j.gastro.2004.11.003

Mooney PD, Aziz I and Sanders DS (2013) Non-celiac gluten sensitivity: clinical relevance and recommendations for future research. Neurogastroent Motil 25:864-871. doi: 10.1111/nmo.12216

Noma S, Kawaura K, Hayakawa K, Abe C, Tsuge N and Ogihara Y (2015) Comprehensive molecular characterization of the α/β-gliadin multigene family in hexaploid wheat. Mol Genet Genomics 291:65-77. doi: 10.1007/s00438-015-1086-7

Okita TW, Cheesbrough V and Reeves CD (1985) Evolution and heterogeneity of the α-/β-type and γ-type gliadin DNA sequences. J Biol Chem 260:8203-8213.

Ozuna CV, Iehisa JCM, Gimenez MJ, Alvarez JB, Sousa C and Barro F (2015) Diversification of the celiac disease α-gliadin complex in wheat: a 33-mer peptide with six overlapping epitopes, evolved following polyploidization. Plant J 82:794-805. doi: 10.1111/tpj.12851

Pritchard JK, Stephens M and Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945-959.

Rashtak S and Murray JA (2012) Review article: coeliac disease, new approaches to therapy. Aliment Pharm Ther 35:768-781. doi: 10.1111/j.1365-2036.2012.05013.x

Rewers MJ (2005) Epidemiology of celiac disease: what are the prevalence, incidence, and progression of celiac disease? Gastroenterology 128:S47-S51.

Salamini F, Ozkan H, Brandolini A, Schafer-Pregl R and Martin W (2002) Genetics and geography of wild cereal domestication in the Near East. Nat Rev Genet 3:429-441. doi: 10.1038/nrg817

Sander I, Rozynek P, Rihs H-P, van Kampen V, Chew FT, Lee WS, Kotschy-Lang N, Merget R, Bruning T and Raulf-Heimsoth M (2011) Multiple wheat flour allergens and cross-reactive carbohydrate determinants bind IgE in baker's asthma. Allergy 66:1208-1215. doi: 10.1111/j.1398-9995.2011.02636.x

Sapone A, Bai JC, Ciacci C, Dolinsek J, Green PHR, Hadjivassiliou M, Kaukinen K, Rostami K, Sanders DS, Schumann M, Ullrich R, Villalta D, Volta U, Catassi C and Fasano A (2012) Spectrum of gluten-related disorders: consensus on new nomenclature and classification. BMC Med 10:13. doi: 10.1186/1741-7015-10-13

Shan L, Molberg O, Parrot I, Hausch F, Filiz F, Gray GM, Sollid LM and Khosla C (2002) Structural basis for gluten intolerance in celiac sprue. Science 297:2275-2279. doi: 10.1126/science.1074129

Shewry PR and Halford GH (2002) Cereal seed storage proteins: structures, properties and role in grain utilization. J Exp Bot 53:947-958. doi: 10.1093/jexbot/53.370.947

Shewry PR, Halford NG and Lafiandra D (2003) Genetics of wheat gluten proteins. Adv Genet 49:111-184.

Shewry PR and Tatham AS (1997) Disulphide bonds in wheat gluten proteins. J Cereal Sci 25:207-227. doi: 10.1006/jcrs.1996.0100

Stepniak D and Koning F (2006) Celiac disease-sandwiched between innate and adaptive immunity. Hum Immunol 67:460-468. doi: 10.1016/j.humimm.2006.03.011

Sumner-Smith M, Rafalski JA, Sugiyama T, Stoll M and Söll D (1985) Conservation and variability of wheat α/β-gliadin genes. Nucleic Acids Res 13:3905-3916. doi: 10.1093/nar/13.11.3905

Tamura K, Nei M, Peterson D, Filipski A and Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol 30:2725-2729. doi: 10.1093/molbev/mst197

Vader LW, Stepniak DT, Bunnik EM, Kooy YMC, De Haan W, Wouter Drijfhout J, Van Veelen PA and Koning F (2003) Characterization of cereal toxicity for celiac disease patients based on protein homology in grains. Gastroenterology 125:1105-1113. doi: 10.1053/S0016-5085(03)01204-6

Van den Broeck HC, de Jong HC, Salentijn EMJ, Dekking L, Bosch D, Hamer RJ, Gilissen LJWJ, van der Meer IM and Smulders MJM (2010) 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 121:1527-1539. doi: 10.1007/s00122-010-1408-4

Van Herpen TWJM, Goryunova SV, van der Schoot J, Mitreva M, Salentijn E, Vorst O, Schenk MF, van Veelen PA, Koning F, van Soest LJM, Vosman B, Bosch D, Hamer RJ, Gilissen LJWJ and Smulders MJM (2006) Alpha-gliadin genes from the A, B and D genomes of wheat contain different sets of celiac disease epitopes. BMC Genomics 7:1. doi: 10.1186/1471-2164-7-1

Von Buren M (2001) Polymorphisms in two homeologous γ-gliadin genes and the evolution of cultivated wheat. Genet Resour Crop Ev 48:205-220. doi: 10.1023/A:1011213228222

Xie Z, Wang C, Wang K, Wang S, Li X, Zhang Z, Ma W and Yan Y (2010) Molecular characterization of the celiac disease epitope domains in α-gliadin genes in Aegilops tauschii and hexaploid wheats (Triticum aestivum L.). Theor Appl Genet 121:1239-1251. doi: 10.1007/s00122-010-1384-8

## Supplementary materials

***Online resource 1.*** Passport data of the 85 spelt accessions submitted to a genetic assignment method analysis with Structure software.

| **Provenance area** | **Provenance country** | **Name** [1] | **Accession number** | **Accession Name** | **Germplasm** [2] |
|---|---|---|---|---|---|
| Africa | Western Sahara | SAH01 | CGN08284 | Sahara I | CGN (Wageningen, The Netherlands) |
| America | Mexico | MEX01 | PI520066 | 26867-302y-300m-oy | USDA (Washington, USA) |
| | USA | US02 | PI168681 | White spring | USDA (Washington, USA) |
| | | US06 | PI355595 | 69Z5,73 | USDA (Washington, USA) |
| Eastern Europe | Bulgaria | BUL01 | PI295059 | Deutchland | USDA (Washington, USA) |
| | | BUL02 | PI295061 | Ungarn | USDA (Washington, USA) |
| | | BUL03 | PI295062 | Italien | USDA (Washington, USA) |
| | | BUL04 | PI295063 | Ungarn | USDA (Washington, USA) |
| | Czech Republic | CZE01 | RICP 01c0100920 | (svetla) | CRI (Prague, Czech Republic) |
| | | CZE02 | RICP 01c0100921 | (tmava) | CRI (Prague, Czech Republic) |
| | | CZE06 | RICP 01c0200982 | spalda bila jarni | CRI (Prague, Czech Republic) |
| | | CZE07 | RICP 01c0200983 | | CRI (Prague, Czech Republic) |
| | | CZE08 | RICP 01c0201257 | (ruzyne) | CRI (Prague, Czech Republic) |
| | Hungary | HUN01 | PI272574 | I-1-3540 | USDA (Washington, USA) |
| | | HUN02 | PI272579 | I-1-3530 | USDA (Washington, USA) |
| | | HUN03 | PI290514 | White spelt | USDA (Washington, USA) |
| | Macedonia | MAC01 | PI378469 | 1744 | USDA (Washington, USA) |

***Online resource 1.*** *(Continued)*

| **Provenance area** | **Provenance country** | **Name [1]** | **Accession number** | **Accession Name** | **Germplasm [2]** |
|---|---|---|---|---|---|
| Eastern Europe | Poland | POL01 | PI192717 | Dankowska graniatka | USDA (Washington, USA) |
| | Romania | ROM01 | PI306554 | 2947 | USDA (Washington, USA) |
| | Ukraine | UKR01 | K 19372 | | VIR (Saint-Petersburg, Russia) |
| Northwestern Europe | Austria | AUS01 | | Ebners rotkorn | |
| | Belgium | BEL02 | | Franckenkorn | CRA-W (Gembloux, Belgium) |
| | | BEL04 | | Redouté | CRA-W (Gembloux, Belgium) |
| | | BEL06 | PI348303 | 69Z6,472 | USDA (Washington, USA) |
| | | BEL07 | PI348312 | 69Z6,482 | USDA (Washington, USA) |
| | | BEL08 | PI348315 | 69Z6,485 | USDA (Washington, USA) |
| | | BEL09 | PI348329 | 69Z6,499 | USDA (Washington, USA) |
| | | BEL10 | PI348334 | 69Z6,505 | USDA (Washington, USA) |
| | | BEL12 | PI348366 | 69Z6,538 | USDA (Washington, USA) |
| | | BEL13 | PI348409 | 69Z6,582 | USDA (Washington, USA) |
| | | BEL14 | PI348417 | 69Z6,590 | USDA (Washington, USA) |
| | Denmark | DK01 | PI361811 | DN-2267 | USDA (Washington, USA) |
| | | DK02 | PI361812 | DN-2268 | USDA (Washington, USA) |
| | Germany | GER08 | PI348033 | 69Z6,191 | USDA (Washington, USA) |
| | | GER09 | PI348040 | 69Z6,198 | USDA (Washington, USA) |
| | | GER10 | PI348056 | 69Z6,215 | USDA (Washington, USA) |

***Online resource 1.*** *(Continued)*

| Provenance area | Provenance country | Name [1] | Accession number | Accession Name | Germplasm [2] |
|---|---|---|---|---|---|
| Northwestern Europe | Germany | GER11 | PI348114 | 69Z6,275 | USDA (Washington, USA) |
| | | GER12 | PI348120 | 69Z6,282 | USDA (Washington, USA) |
| | | GER15 | PI348159 | 69Z6,322 | USDA (Washington, USA) |
| | | GER16 | PI348168 | 69Z6,332 | USDA (Washington, USA) |
| | | GER17 | PI348171 | 69Z6,335 | USDA (Washington, USA) |
| | | GER19 | PI355552 | T2 | USDA (Washington, USA) |
| | | GER21 | PI355677 | BP 2 | USDA (Washington, USA) |
| | | GER22 | TRI3445 | Blauer Samtiger | IPK (Gatersleben, Germany) |
| | | GER23 | TRI1259 | Rottweiler Frûhkorn St. I | IPK (Gatersleben, Germany) |
| | | GER25 | TRI 303 | Kipperhaus Weisser Spelz | IPK (Gatersleben, Germany) |
| | Sweden | SWE01 | NGB 4495 | Speltvete FR Gotland | NGB (Alnarp, Sweden) |
| | | SWE02 | CGN08300 | | CGN (Wageningen, The Netherlands) |
| | Switzerland | SWI02 | | Balmegg | |
| | | SWI10 | RAC TS2113 | SCHNOTTWILER WEISSKORN 35 | Agroscope Changins (Nyon, Switzerland) |
| | | SWI11 | RAC TS2145 | Thuerig rotkorn Th4 | Agroscope Changins (Nyon, Switzerland) |
| | | SWI12 | RAC TS2144 | Hueslen rotkorn | Agroscope Changins (Nyon, Switzerland) |
| | | SWI13 | RAC TS2117 | Neuegger weisskorn Ngg42 | Agroscope Changins (Nyon, Switzerland) |
| | | SWI18 | RAC TS2116 | Willisauer weisskorn Wil17 | Agroscope Changins (Nyon, Switzerland) |
| | | SWI20 | PI347904 | 69Z6,57 | USDA (Washington, USA) |
| | | SWI21 | PI347913 | 69Z6,66 | USDA (Washington, USA) |

***Online resource 1.*** *(Continued)*

| **Provenance area** | **Provenance country** | **Name [1]** | **Accession number** | **Accession Name** | **Germplasm [2]** |
|---|---|---|---|---|---|
| Northwestern Europe | Switzerland | SWI23 | PI347939 | 69Z6,93 | USDA (Washington, USA) |
| | | SWI24 | PI347950 | 69Z6,105 | USDA (Washington, USA) |
| | | SWI25 | PI348004 | 69Z6,161 | USDA (Washington, USA) |
| | | SWI27 | PI355560 | SK1B | USDA (Washington, USA) |
| | | SWI28 | PI355573 | SK3F | USDA (Washington, USA) |
| | | SWI31 | PI355653 | Rottweiller Fruhkorn 4 | USDA (Washington, USA) |
| | United Kingdom | UK01 | PI355704 | 69Z5.194 | USDA (Washington, USA) |
| Southwestern Europe | Italy | ITA01 | PI355642 | 69Z5,122 | USDA (Washington, USA) |
| | Spain | SPA01 | PI348526 | 69Z6,704 | USDA (Washington, USA) |
| | | SPA02 | PI348537 | 69Z6,715 | USDA (Washington, USA) |
| | | SPA03 | PI348572 | 69Z6,752 | USDA (Washington, USA) |
| | | SPA04 | PI348580 | 69Z6,760 | USDA (Washington, USA) |
| | | SPA05 | PI348588 | 69Z6,768 | USDA (Washington, USA) |
| | | SPA06 | PI348651 | 69Z6,834 | USDA (Washington, USA) |
| | | SPA08 | PI348688 | 69Z6,872 | USDA (Washington, USA) |
| | | SPA09 | PI348716 | 69Z6,900 | USDA (Washington, USA) |
| | | SPA10 | PI348723 | 69Z6,908 | USDA (Washington, USA) |
| | | SPA11 | PI348766 | 69Z6,953 | USDA (Washington, USA) |
| | | SPA13 | PI348463 | 69Z6,638 | USDA (Washington, USA) |

***Online resource 1.*** *(Continued)*

| Provenance area | Provenance country | Name [1] | Accession number | Accession Name | Germplasm [2] |
|---|---|---|---|---|---|
| Middle East | Afghanistan | AFG04 | PI367202 | 625 | USDA (Washington, USA) |
| | Tajikistan | TAD01 | K 52443 | | VIR (Saint-Petersburg, Russia) |
| | | TAD06 | K 52437 | | VIR (Saint-Petersburg, Russia) |
| | | TAD07 | K 52442 | | VIR (Saint-Petersburg, Russia) |
| | | TAD15 | K 52463 | | VIR (Saint-Petersburg, Russia) |
| | | TAD22 | K 56568 | | VIR (Saint-Petersburg, Russia) |
| Near East | Azerbaijan | AZE02 | K 45364 | | VIR (Saint-Petersburg, Russia) |
| | | AZE03 | K 45368 | | VIR (Saint-Petersburg, Russia) |
| | Iran | IRA03 | CGN12270 | Iran 416A | CGN (Wageningen, The Netherlands) |
| | | Iran77d | CGN06533 | Iran 77d | USDA (Washington, USA) |

[1] : Names used for the same accessions in Bertin et al. (2004) except for Iran77d, which was named as in Dvorak et al. (2012).

[2] : USDA = United States Department of Agriculture; VIR = Vavilov Institute of Plant Genetic Resources; CRI = Crop Research Institute; IPK = Institute of Plant Genetics and Crop Plant Research; CGN = Center for Genetic Resources; CRA-W = Centre wallon de Recherches agronomiques; NGB = Nordic Gene Bank.

***Online resource 2.*** (a) List of α-gliadin sequences from ancestral diploid species and from bread wheat used to confirm the genome assignments. (b) List of additional α-gliadin sequences from bread wheat used to conduct an overall phylogenetic analysis together with α-gliadin sequences from bread wheat and diploid species included in the Online resource 2a and with spelt α-gliadins isolated in this work.

**a**

| Genome | Species | | Accession number | Reference |
|---|---|---|---|---|
| A | *T. urartu* | | KP405293, KP405294, KP405295, KP405296, KP405297, KP405298, KP405299, KP405301, KP405302, KP405303, KP405304, KP405305, KP405306, KP405307, KP405308 | Ozuna et al. 2015 |
| | *T. aestivum* ssp. *aestivum* | var. Cheyenne | U08287 | Blechl and Anderson 1994* |
| | | var. Cheyenne | U50984, U51306, U51307 | Anderson et al. 1997 |
| | | var. Chinese Spring | AB982242, AB982245, AB982249, AB982255, AB982272, AB982277, AB982278, AB982281, AB982288 | Noma et al. 2015 |
| S/B | *Ae. speltoides* | | DQ002584, DQ002585, DQ002586, DQ002587, DQ002588 | Van Herpen et al. 2006 |
| | *T. aestivum* ssp. *aestivum* | var. Cheyenne | K02068 | Kasarda et al. 1984 |
| | | var. Cheyenne | M11073 | Okita et al. 1985 |
| | | var. Cheyenne | U51303 | Anderson et al. 1997 |
| | | | X02540 | Sumner-Smith et al. 1985 |
| | | var. Chinese Spring | AB982234, AB982236, AB982237, AB982241, AB982267, AB982273, AB982285, AB982286 | Noma et al. 2015 |
| D | *Ae. tauschii* | | DQ002589, DQ002590, DQ002591, DQ002592, DQ002593, DQ002594, DQ002595, DQ002596, DQ002597, DQ002598, DQ002599 | Van Herpen et al. 2006 |
| | *T. aestivum* ssp. *aestivum* | var. Cheyenne | M11075 | Okita et al. 1985 |
| | | var. Chinese Spring | X17361 | Garcia-Maroto et al. 1990 |

***Online resource 2.*** *(Continued)*

| Genome | Species | | Accession number | Reference |
|---|---|---|---|---|
| D | *T. aestivum* ssp. *aestivum* | var. Chinese Spring | AB982248, AB982253, AB982256, AB982260, AB982265, AB982276, AB982279, AB982284, AB982293 | Noma et al. 2015 |
| **b** | | | | |
| A, B, D | *T. aestivum* ssp. *aestivum* var. Mjoelner | | AJ133602, AJ133604, AJ133605, AJ133612 | Arentz-Hansen et al. 2000 |
| | *T. aestivum* ssp. *aestivum* var. Jimai20 | | JN831386 | Li and Li 2011* |
| | *T. aestivum* ssp. *aestivum* var. Zhengmai9023 | | KC715927, KC715929, KC715930, KC715932, KC715940, KC715941, KC715944, KC715945, KC715950, KC715951, KC715952, KC715953 | Li et al. 2013* |
| | *T. aestivum* ssp. *aestivum* var. Zhengmai9023 | | JX828270 | Li et al. 2012* |
| | *T. aestivum* ssp. *aestivum* var. C273 | | KJ410475 | Kaur et al.2016 |
| | *T. aestivum* ssp. *aestivum* var. GW273 | | KJ410476 | Kaur et al.2016 |
| | *T. aestivum* ssp. *aestivum* var. HI617 | | KJ410477 | Kaur et al.2016 |
| | *T. aestivum* ssp. *aestivum* var. Hyb65 | | KJ410478 | Kaur et al.2016 |
| | *T. aestivum* ssp. *aestivum* var. Lok1 | | KJ410481 | Kaur et al.2016 |
| | *T. aestivum* ssp. *aestivum* var. NP4 | | KJ410483 | Kaur et al.2016 |
| | *T. aestivum* ssp. *aestivum* var. NP824 | | KJ410484 | Kaur et al.2016 |
| | *T. aestivum* ssp. *aestivum* var. VL404 | | KJ410486 | Kaur et al.2016 |
| | *T. aestivum* ssp. *aestivum* var. WG377 | | KJ410487 | Kaur et al.2016 |
| | *T. aestivum* ssp. *aestivum* var. WL711 | | KJ410488 | Kaur et al.2016 |
| | *T. aestivum* ssp. *aestivum* var. Zhengmai004 | | KC715889, KC715890, KC715892, KC715896, KC715897, KC715901, KC715906, KC715907, KC715908, KC715910, KC715913, KC715917, KC715920, HM120221, HM120222 | Li et al. 2014 |
| | *T. aestivum* ssp. *aestivum* var. Yamhill | | K03074, K03075, K03076 | Sumner-Smith et al. 1985 |
| | *T. aestivum* ssp. *aestivum* var. Wyuna | | FN391140 | Sander et al. 2011 |

***Online resource 2.*** *(Continued)*

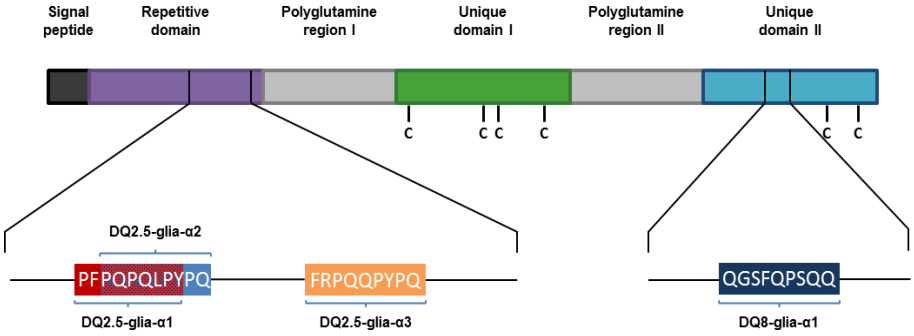
| Genome | Species | Accession number | Reference |
|---|---|---|---|
| A, B, D | *T. aestivum* ssp. *aestivum* var. Gaocheng 8901 | EF561274, EF561275, EF561276, EF561278, EF561279, EF561280 | Xie et al. 2010 |
| | *T. aestivum* ssp. *aestivum* var. Zhongyou 9507 | EF561281, EF561283, EF561285, EF561287, EF561288 | Xie et al. 2010 |
| | *T. aestivum* ssp. *aestivum* var. AS1643 | DQ166376 | Chen et al. 2005* |
| | *T. aestivum* ssp. *aestivum* var. Jinan177 | EU018291, EU018292, EU018295, EU018296, EU018297, EU018298, EU018299 | Chen et al. 2007* |
| | *T. aestivum* ssp. *aestivum* var. Shaan253 | GQ891683, GQ891685 | Li et al. 2009* |
| | *T. aestivum* ssp. *aestivum* var. Chinese Spring | JX141483, JX141484, JX141485, JX141486, JX141488, JX141489, JX141490, JX141492, JX141493, JX141494, JX141495 | Li et al. 2012* |
| | *T. aestivum* ssp. *aestivum* var. Yumai50 | JX828260 | Li et al. 2012* |
| | *T. aestivum* ssp. *aestivum* var. Zhengfeng5 | JX828284, JX828286, JX828291, JX828292, JX828293, JX828296, JX828298, JX828299, JX828300, JX828304, JX828306, JX828307, JX828310 | Li et al. 2012* |
| | *T. aestivum* ssp. *aestivum* var. Zhengmai366 | JX828318 | Li et al. 2012* |
| | *T. aestivum* ssp. *aestivum* | DQ417343 | Liu et al. 2006* |
| | *T. aestivum* ssp. *aestivum* var. Chuannong 16 | DQ246446, DQ246447, DQ246448 | Liu et al. 2005* |
| | *T. aestivum* ssp. *aestivum* var. Chinese Spring | KJ137236, KJ137239, KJ137240, KJ137242, KJ137243 | Liu et al. 2014* |
| | *T. aestivum* ssp. *aestivum* var. Chinese Spring | AB982235, AB982282, AB982291, AB982292, AB982294, AB982296, AB982298, AB982299 | Noma et al. 2015 |
| | *T. aestivum* ssp. *aestivum* var. Cheyenne | M10092 | Okita et al. 1985 |
| | *T. aestivum* ssp. *aestivum* var. Conil | KP405248, KP405249, KP405252, KP405254, KP405255, KP405260, KP405261, KP405262, KP405263, KP405266, KP405267, KP405268, KP405269, KP405270, KP405271, KP405272, KP405273, KP405275, | Ozuna et al. 2015 |

***Online resource 2.*** *(Continued)*

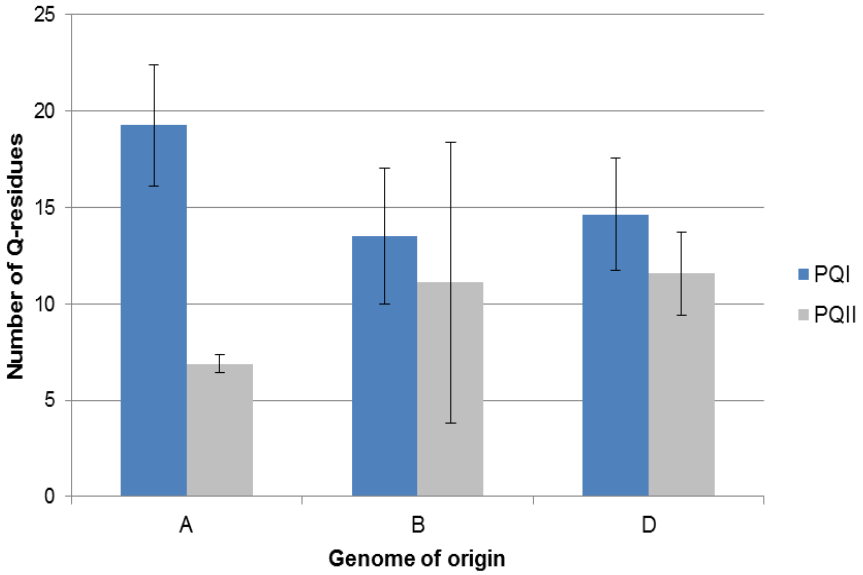
| Genome | Species | Accession number | Reference |
|---|---|---|---|
| A, B, D | *T. aestivum* ssp. *aestivum* var. Conil | KP405276, KP405277, KP405278, KP405279, KP405280, KP405281, KP405282 | Ozuna et al. 2015 |
| | *T. aestivum* ssp. *aestivum* | X02538, X02539 | Sumner-Smith et al. 1985 |
| | *T. aestivum* ssp. *aestivum* var. 96-212-2 | EF165553, EF165556 | Tang et al. 2006* |
| | *T. aestivum* ssp. *aestivum* var. 96-132-1 | EF165554 | Tang et al. 2006* |
| | *T. aestivum* ssp. *aestivum* var. 96-137-3 | EF165555 | Tang et al. 2006* |

*: Direct submission

***Online resource 3.*** Typical structure of an α-gliadin. The four major T-cell stimulatory epitopes and their corresponding amino acid sequences are represented below the polypeptide model. 'C' indicates the classical location of the cysteine residues.



***Online resource 4.*** Average number of glutamine residues in the first (PQI) and second (PQII) polyglutamine regions in 226 α-gliadin transcripts from 11 spelt accessions, according to the genome of origin (A, B and D). The PQI region overall had a higher average number of glutamine residues (16.8 +/- 4.1) than the PQII region (9.0 +/- 4.4). Alpha-gliadins from the A genome displayed a significantly higher average number of glutamine residues in the PQI region (19.3 +/- 3.1) than B (13.5 +/- 3.5) and D (14.6 +/- 2.9) genome sequences. For the PQII region, A genome sequences had a significantly lower average number of glutamine residues (6.9 +/- 0.5) than B (11.1 +/- 7.3) and D (11.6 +/- 2.2) genome α-gliadins.



***Online resource 5.*** Variants of the DQ2.5-glia-α1, -α2, -α3 and DQ8-glia-α1 epitopes according to their genome of origin.

| Epitope | Amino acid sequence | A genome | B genome | D genome | Total |
|---|---|---|---|---|---|
| DQ2.5-glia-α1 | PFPQPQLPY | 150 | 0 | 104 | 254 |
| | PFP-PQLPY | 0 | 138 | 0 | 138 |
| | P**Y**PQPQLPY | 0 | 0 | 51 | 51 |
| | PF**S**QPQLPY | 16 | 0 | 0 | 16 |
| | PF**L**QPQLPY | 16 | 0 | 0 | 16 |
| | P**Y**PQP**H**LPY | 0 | 0 | 11 | 11 |
| | PF**L**-PQLPY | 0 | 10 | 0 | 10 |
| | **Q**FP-PQ**Q**PY | 0 | 3 | 0 | 3 |
| | PFPQPQL**S**Y | 1 | 0 | 0 | 1 |
| | P**S**P-PQLPY | 0 | 1 | 0 | 1 |
| | P**Y**P**R**PQLPY | 0 | 0 | 1 | 1 |
| | Total | 183 | 152 | 167 | 502 |
| DQ2.5-glia-α2 | PQPQLPYPQ | 0 | 0 | 157 | 157 |
| | P-PQLPYPQ | 0 | 139 | 0 | 139 |
| | PQPQLPY**S**Q | 123 | 0 | 0 | 123 |
| | PQPQLPY**SH** | 27 | 0 | 0 | 27 |
| | **S**QPQLPY**S**Q | 16 | 0 | 0 | 16 |
| | **L**QPQLPY**S**Q | 16 | 0 | 0 | 16 |
| | PQP**H**LPYPQ | 0 | 0 | 11 | 11 |
| | **L**-PQLPYPQ | 0 | 8 | 0 | 8 |
| | P-PQ**Q**PYPQ | 0 | 3 | 0 | 3 |
| | **L**-PQLPYP**R** | 0 | 2 | 0 | 2 |
| | PQPQLPYP**R** | 0 | 0 | 1 | 1 |
| | P**R**PQLPYPQ | 0 | 0 | 1 | 1 |
| | PQPQL**S**Y**S**Q | 1 | 0 | 0 | 1 |
| | Total | 183 | 152 | 170 | 505 |
| DQ2.5-glia-α3 | FRPQQPYPQ | 173 | 0 | 93 | 266 |
| | F**P**PQQPYPQ | 0 | 149 | 0 | 149 |
| | FR**Q**QQPYPQ | 60 | 0 | 0 | 60 |
| | FRPQQ**S**YPQ | 0 | 0 | 13 | 13 |
| | FRPQ**K**PYPQ | 4 | 0 | 0 | 4 |
| | **SP**PQQPYPQ | 0 | 2 | 0 | 2 |
| | F**Q**PQQPYPQ | 1 | 0 | 0 | 1 |

***Online resource 5.*** *(Continued)*

| Epitope | Amino acid sequence | A genome | B genome | D genome | Total |
|---|---|---|---|---|---|
| DQ2.5-glia-α3 | F**P**PQQ**S**YPQ | 0 | 1 | 0 | 1 |
| | Total | 238 | 152 | 106 | 496 |
| DQ8-glia-α1 | QGSFRPSQQ | 176 | 2 | 2 | 180 |
| | QGSFQ**S**SQQ | 0 | 102 | 1 | 103 |
| | QGSFQPSQQ | 6 | 7 | 48 | 61 |
| | QG**F**FQPSQQ | 0 | 1 | 57 | 58 |
| | Q**V**SFQPSQ**L** | 0 | 40 | 0 | 40 |
| | QGSF**R**PSQ**L** | 1 | 0 | 0 | 1 |
| | QGSFQPSQ**L** | 0 | 0 | 1 | 1 |
| | Total | 183 | 152 | 109 | 444 |

Amino acid residues differing from the canonical epitope are highlighted in bold and deleted residues are represented by a dash.

# CHAPTER 2

Development of TaqMan probes targeting the four major celiac disease epitopes found in α-gliadin sequences of spelt (*Triticum aestivum* ssp. *spelta*) and bread wheat (*Triticum aestivum* ssp. *aestivum*)

In the previous chapter, a large number of α-gliadin expressed sequences from spelt accessions were cloned and sequenced, highlighting the high genetic diversity existing for this trait. This enabled to study the variations in their composition in the four main T-cell stimulatory epitopes involved in CD. The genome assignment of the α-gliadins enabled to highlight genome-specific allelic variants and the count of canonical epitopes provided an estimation of the accession immunogenicity expressed as the mean number of canonical epitopes per sequence. Such estimation, however, did not take into account the expression levels of α-gliadin sequences. Hence, this second chapter is dedicated to respond to the second specific objective, *i.e.* the development and the validation of TaqMan probes molecular markers enabling to quantify the expression of the four main α-gliadin CD-related epitopes. In order to normalize qPCR results, TaqMan probes were also developed to measure the expression of stable reference genes. These designed epitope- and reference gene-targeting TaqMan probes were then used to obtain a more accurate information about the immunogenicity of 11 contrasted spelt accessions studied in chapter 1 and three related diploid species.

This chapter was published as: Dubois, B., Bertin, P., Muhovski, Y., Escarnot, E., Mingeot, D. (2017). Development of TaqMan probes targeting the four major celiac disease epitopes found in α-gliadin sequences of spelt (*Triticum aestivum* ssp. *spelta*) and bread wheat (*Triticum aestivum* ssp. *aestivum*). Plant Methods, 13, 72. doi: 10.1186/s13007-017-0222-2

## Abstract

Celiac disease (CD) is caused by specific sequences of gluten proteins found in cereals such as bread wheat (*Triticum aestivum* ssp. *aestivum*) and spelt (*T. aestivum* ssp. *spelta*). Among them, the α-gliadins display the highest immunogenicity, with four T-cell stimulatory epitopes. The toxicity of each epitope sequence can be reduced or even suppressed according to the allelic form of each sequence. One way to address the CD problem would be to make use of this allelic variability in breeding programs to develop safe varieties, but tools to track the presence of toxic epitopes are required. The objective of this study was to develop a tool to accurately detect and quantify the immunogenic content of expressed α-gliadins of spelt and bread wheat. Four TaqMan probes that only hybridize to the canonical – i.e. toxic – form of each of the four epitopes were developed and their specificity was demonstrated. Six TaqMan probes targeting stable reference genes were also developed and constitute a tool to normalize qPCR data. The probes were used to measure the epitope expression levels of 11 contrasted spelt accessions and three ancestral diploid accessions of bread wheat and spelt. A high expression variability was highlighted among epitopes and among accessions, especially in Asian spelts, which showed lower epitope expression levels than the other spelts. Some discrepancies were identified between the canonical epitope expression level and the global amount of expressed α-gliadins, which makes the designed TaqMan probes a useful tool to quantify the immunogenic potential independently of the global amount of expressed α-gliadins. In conclusion, the results obtained in this study provide useful tools to study the immunogenic potential of expressed α-gliadin sequences from Triticeae accessions such as spelt and bread wheat. The application of the designed probes to contrasted spelt accessions revealed a high variability and interesting low canonical epitope expression levels in the Asian spelt accessions studied.

## Keywords

spelt, α-gliadin, celiac disease, TaqMan probe, reference genes, gluten

## Introduction

Gluten is a group of water-insoluble proteins which display useful properties for dough formation. These proteins are found in the seed of cereals such as bread wheat (*Triticum aestivum* L. ssp. *aestivum*), spelt [*Triticum aestivum* ssp. *spelta* (L.) Thell.], barley (*Hordeum vulgare* L.) and rye (*Secale cereale* L.). The gluten complex is composed of monomeric gliadins and polymeric glutenins which confer visco-elasticity properties to the dough [1, 2].

The consumption of gluten-containing foods can lead to several disorders in a fraction of the population, the best known being mediated by the adaptive immune system: celiac disease (CD) and wheat allergy (WA). They are both mediated by a T-cell activation in the gastrointestinal mucosa. WA involves immunoglobulins (Ig)-E that link to repeat sequences in wheat proteins whereas CD is an autoimmune disorder with a genetic predisposition of the patient, mediated by Ig-A and Ig-G antibodies. Beside these two disorders, a novel pathologic entity – called non-celiac gluten sensitivity (NCGS) – has gained importance and involves neither autoimmune nor allergic mechanisms. People affected by this illness display symptoms similar to CD but with usually normal small intestinal histology. These patients become healthier when gluten or wheat is withdrawn from the diet [3, 4, 5, 6]. CD affects about 1% of the human population and is thus one of the most important food sensitivities worldwide [6, 7]. Some gluten peptides are partially resistant to proteolysis and are presented by HLA-DQ2 or -DQ8 molecules on antigen-presenting cells (APC) to T-cells which are hence activated. The subsequent production of inflammatory cytokines leads to small intestinal injuries such as reduction of villus height and a decrease of its absorptive capacity, which may cause severe malnutrition [8, 9]. Intra- and extraintestinal symptoms are encountered like diarrhea, bowel pain, fatigue, weight loss, anemia, osteoporosis, headaches and growth retardation [5, 6, 10, 11].

Different strategies are currently being investigated to address the CD problem, such as gluten detoxification, modulation of mucosal permeability, antigen

presentation blockade, raising monoclonal antibodies against inflammatory cytokines, inhibition of T-cell recruitment, or oral tolerance induction [10]. Among them, the development of new cereal varieties lacking immunogenic gluten peptides but still displaying good baking properties is a promising approach given the high variability existing in bread wheat and its related taxa [10, 12]. Spelt is one of these taxa and is a member of the *Triticum aestivum* species, just like bread wheat. It is particularly interesting because of the high genetic diversity held in spelt germplasm collections [13, 14, 15]. Moreover, genetic variations in their immunogenic potential, their bread-making qualities and their content in proteins, lipids, micronutrients and fibers have been pointed out [13, 16, 17, 18, 19].

Gliadins are usually classified into the α/β-, γ- and ω-types. The α-gliadins are the most studied ones since they trigger the strongest T-cell activation [8, 20, 21, 22]. The Gli-2 loci, which encode α-gliadins, are located on the short arm of the three homeologous chromosomes from group 6 and they include a number of α-gliadin gene copies that might reach 150 per haploid genome in some accessions [23]. However, a high proportion of these copies are pseudogenes: Ozuna et al. [24] showed that 39, 76 and 63% of α-gliadin genes from diploid, tetraploid and hexaploid wheat species respectively displayed a premature stop codon. The α-gliadin strong immunogenicity is mainly due to four T-cell stimulatory epitopes. Two are major epitopes (the overlapping DQ2.5-glia-α1 and -α2 epitopes) and the other two are minor epitopes (the DQ2.5-glia-α3 and DQ8-glia-α1 epitopes). The DQ2.5-glia-α2 epitope can be displayed in one, two or three copies and leads, when three copies are present, to the most immunogenic fragment of α-gliadin sequences known as the 33-mer fragment [25, 26]. Each of these four epitopes can be displayed in its canonical form, which is toxic, but some amino acid substitutions or deletions can reduce or suppress antigenic properties, according to the allelic variant [27]. Interestingly, the epitope variants highlighted in bread wheat α-gliadin sequences by Mitea et al. [27] were similar to those found in spelt accessions [16]. The multigenic character of the α-gliadin family results in a very high allelic variability and leads, among other things, to high variations in the immunogenic content from one α-gliadin sequence to another. The exploitation of this variability

to develop safe spelt or wheat cultivars will probably require the combination of classical breeding and molecular technologies [12], and tools to infer the presence of sequences coding for toxic epitopes are needed.

ELISA test kits have been developed to measure the amount of gluten in food samples and to detect gluten contamination. Generally speaking, they are based on different antibodies [28, 29, 30, 31, 32, 33] which have been raised against gluten proteins. However, although ELISA kits provide information on toxicity, the restricted specificity of the antibodies limits the accuracy of ELISA tests regarding the immunogenic potential of varieties. Indeed, each antibody is raised against only a short stretch of one of the toxic epitopes, which may lead to the simultaneous detection of non-toxic epitope variants together with the canonical – toxic – epitope form, because the mutation is located outside the antibody hybridization site. It has been shown that G12 and A1 antibodies display the highest affinity for canonical epitopes included in the 33-mer fragment, but they also recognize allelic variants of these epitopes to a lesser extent [30]. Also, no ELISA kit can be used to study the four major α-gliadin T-cell stimulatory epitopes at the same time. Furthermore, antibodies used in ELISA test kits generally recognize more than one site. For example, the R5 antibody links to the epitope QQPFP and the related sequences QQQFP, LQPFP and QLPFP [33], but only one of these (LQPFP) is found in the composition of the four major canonical T-cell stimulatory epitopes of α-gliadin sequences. Lastly, such antibodies generally recognize different types of gluten proteins, such as the Skerritt antibody that detects ω-gliadins and HMW glutenin subunits. In addition to ELISA kits, some other techniques have recently been developed to detect and quantify different CD-immunogenic peptides, such as aptamer receptors and liquid chromatography combined with mass spectrometry [34, 35, 36].

The multigenic nature of α-gliadins and the high proportion of pseudogenes would require markers able to detect and quantify toxic sequences in expressed genes. The TaqMan probe system is one of the most commonly used to perform a quantitative detection of amplicons [37], and is particularly adapted to the study of

the epitope expression profile since it makes it possible to highlight single-nucleotide polymorphisms (SNPs) and thus to discriminate allelic variants [38, 39].

The objective of this study was to develop and validate a tool to measure the immunogenic content for CD patients held in expressed α-gliadins. This objective was pursued by (i) developing epitope-specific TaqMan probes that only hybridize to the canonical form of the four epitopes, (ii) selecting stable reference genes and developing specific TaqMan probes to normalize qPCR data and (iii) validating the designed probes through their application to cDNA samples from 11 contrasted spelt accessions and three diploid species representative of the ancestral genomes of spelt and bread wheat.

# Materials and methods

***Plant materials***

Eleven contrasted spelt accessions selected in a previous genetic diversity study [16] as a representative panel of spelt diversity were used (BEL08, DK01, SPA03, BUL04, GER11, GER12, TAD06, SWI23, US06, Iran77d and IRA03). Three diploid accessions representative of the three ancestral genomes of spelt and bread wheat were added to this selection (Table 1): LB01 (*Triticum urartu* Tumanian ex Gandilyan, A genome), TR08 (*Aegilops speltoides* Tausch, S genome which is suspected to be the ancestor of the B genome) and TR10 (*Aegilops tauschii* Cosson, D genome). The accessions were kindly provided by the United States Department of Agriculture (USDA, Washington, USA), the Vavilov Institute of Plant Genetic Resources (VIR, Saint Petersburg, Russia) and the Center for Genetic Resources (CGN, Wageningen, The Netherlands). They were grown in 2013-2014 in Gembloux (Belgium) in field conditions. To ensure the self-pollination of the ears, they were enclosed with cellophane bags. All the immature grains were harvested 20 days post-anthesis (DPA), immediately frozen in liquid nitrogen and stored at -80°C.

Table 1. Genetic material used to analyze the α-gliadin canonical epitope expression with TaqMan probes

| Species | Genome | Name | Accession name | Accession number | Germplasm [1] | Provenance country |
|---|---|---|---|---|---|---|
| *Triticum aestivum* ssp. *spelta* | ABD | BEL08 | 69Z6,485 | PI348315 | USDA (Washington, USA) | Belgium |
| | | DK01 | DN-2267 | PI361811 | USDA (Washington, USA) | Denmark |
| | | SPA03 | 69Z6,752 | PI348572 | USDA (Washington, USA) | Spain |
| | | BUL04 | Ungarn | PI295063 | USDA (Washington, USA) | Bulgaria |
| | | GER11 | 69Z6,275 | PI348114 | USDA (Washington, USA) | Germany |
| | | GER12 | 69Z6,282 | PI348120 | USDA (Washington, USA) | Germany |
| | | TAD06 | - | K 52437 | VIR (Saint-Petersburg, Russia) | Tajikistan |
| | | SWI23 | 69Z6,93 | PI347939 | USDA (Washington, USA) | Switzerland |
| | | US06 | 69Z5,73 | PI355595 | USDA (Washington, USA) | USA |
| | | Iran77d | Iran 77d | CGN 06533 | CGN (Wageningen, The Netherlands) | Iran |
| | | IRA03 | Iran 416A | CGN12270 | CGN (Wageningen, The Netherlands) | Iran |
| *Triticum urartu* | A | LB01 | G3178 | PI 428293 | USDA (Washington, USA) | Lebanon |
| *Aegilops speltoides* | S* | TR08 | 2733 | PI 170204 | USDA (Washington, USA) | Turkey |
| *Aegilops tauschii* | D | TR10 | 79TK057-322-1 | PI 486267 | USDA (Washington, USA) | Turkey |

The 11 spelt accessions are contrasted accessions previously selected after a genetic diversity analysis [14]. *Triticum urartu*, *Aegilops speltoides* and *Aegilops tauschii* are diploid species representative of the three ancestral genomes (A, B and D, respectively) of spelt and bread wheat. They were selected to validate the specificity of the designed TaqMan probes.

[1] USDA = United States Department of Agriculture; VIR = Vavilov Institute of Plant Genetic Resources; CGN = Center for Genetic Resources.

* The B genome is hypothesized to be an altered S genome; Ae. sepltoides is therefore taken as the closest representative of the B genome.

***RNA extraction, cDNA synthesis and plasmid extraction***

Total RNA was extracted for each accession from 100 mg of ground seeds using the NucleoSpin® RNA Plant kit (Macherey-Nagel, Germany) and quantified by spectrometry. First-strand cDNA was synthesized with oligo(dT)18 primer from 250 ng RNA using the RevertAid H Minus First Strand cDNA Synthesis Kit (Thermo Scientific) in a volume of 20 µl. It was then quantified by spectrometry.

The plasmids of previously cloned α-gliadin sequences [16] were extracted with the GeneJet Plasmid Miniprep Kit (Thermo Scientific) after the overnight incubation of bacterial cultures at 37°C. The GeneBank accession numbers of the sequences are reported in Additional file 1.

***Design of epitope-targeting primers and probes***

For each epitope, primers were developed (Table 2) so that they hybridize in conserved regions on both sides of the epitope. The design was carried out by aligning previously cloned spelt α-gliadin sequences [16] with the Vector NTI software (v6.0).

The probe development was carried out in accordance with the following important probe characteristics: (i) the amplicon should have a length of between 75 and 150 bp; (ii) no guanine residue should be located at the 5' end of the probe as it quenches the fluorescent dye; (iii) runs of identical nucleotides should be avoided, especially four or more consecutive guanine residues; (iv) the probe should have a Tm of 65-67 °C and a minimum length of 13 nucleotides; (v) the SNP position in the probe should be as central as possible to maximize the detection specificity [40]. TaqMan probes focusing on the four major T-cell stimulatory epitopes DQ2.5-glia-α1, DQ2.5-glia-α2, DQ2.5-glia-α3 and DQ8-glia-α1 were designed so that they only hybridized to the canonical form of each epitope (Table 2). This was done by aligning spelt α-gliadin sequences cloned in a previous study [16] with the Vector NTI software. The probes were designed to cover the entire regions where epitope mutations are found in spelt and bread wheat sequences.

Table 2. List of the primers and TaqMan probes designed in this study

| Target epitope/gene | Oligonucleotide | Sequence |
|---|---|---|
| ***Epitopes*** | | |
| DQ2.5-glia-α1 | Forward primer | 5'-GCAACCATTTCCATCACAACWAC-3' |
| | Reverse primer | 5'-GTGSTTGCGAATACTGTGGTTG-3' |
| | Probe | 5'-FAM-WTCCRCAGCCGCAACTACCA-TAMRA-3' |
| DQ2.5-glia-α2 | Forward primer | 5'-GCAACCATTTCCATCACAACWAC-3' |
| | Reverse primer | 5'-GTGSTTGCGAATACTGTGGTTG-3' |
| | Probe | 5'-FAM-AGCCGCAACTACCATATCCGC-TAMRA-3' |
| DQ2.5-glia-α3 | Forward primer | 5'-GCAACCATTTCCATCACAACWAC-3' |
| | Reverse primer | 5'-GTGSTTGCGAATACTGTGGTTG-3' |
| | Probe | 5'-FAM-TTCGACCACAACAACCATATCCAC-TAMRA-3' |
| DQ8-glia-α1 | Forward primer | 5'-CCACAATGTYGTTCATGCTATTATTCTGC-3' |
| | Reverse primer | 5'-CAGAGCCCTGGGCCTGTGG-3' |
| | Probe | 5'-FAM-AGGGCTCCTTCCAGCCAT-TAMRA-3' |
| ***Reference** genes* | | |
| ADP-ribosylation factor (ARF) | Forward primer | 5'-GCTCTCCAACAACATTGCCAAC-3' |
| | Reverse primer | 5'-GCTTCTGCCTGTCACATACGC-3' |
| | Probe | 5'-FAM-CAAGAAACAAACGTGCTGGATGTC-TAMRA-3' |
| Similar to RNase L inhibitor-like protein (RLI) | Forward primer | 5'-CGATTCAGAGCAGCGTATTGTTGC-3' |
| | Reverse primer | 5'-AGTTGGTCGGGTCTCTTCTAAATGTAATG-3' |
| | Probe | 5'-FAM-CTTAGCGGACAAGGTTATTGTTTATGAGG-TAMRA-3' |
| Vacuolar ATP synthase 16 kDa proteolipid sub. (VAS) | Forward primer | 5'-GCTGGAGTGCTCGGTATCTACGG-3' |
| | Reverse primer | 5'-TGCGAAGATGAGGATGAGGATCA-3' |
| | Probe | 5'-FAM-ATCGGCATTGTTGGTGATGCT-TAMRA-3' |
| Ubiquinol-cytochrome C reduct. iron-sulfur sub. (UCC) | Forward primer | 5'-CCTGCCCCGTACAACCTTGAG-3' |
| | Reverse primer | 5'-TCACCGTTGCGATAGTCCTGAAAC-3' |
| | Probe | 5'-FAM-ACAGGAGTGAATTCCTGTTGCGC-TAMRA-3' |
| GABA-receptor-associated protein (GABA) | Forward primer | 5'-TTACGAGGAGAACAAGGACGAGGA-3' |
| | Reverse primer | 5'-CAGGAGGCATTCAGAGCGATTG-3' |
| | Probe | 5'-FAM-CACCTTCGGATTGCTCTAGATGGC-TAMRA-3' |
| Protein of unknown function [DUF52 family] (DUF52) | Forward primer | 5'-TGGTGCCATTCACAAATCAATCG-3' |
| | Reverse primer | 5'-GCGAACAAACCCGACCTTAATCTTC-3' |
| | Probe | 5'-FAM-CATGGAGATCATAGAGACTGGTGACC-TAMRA-3' |

Table 2 *(Continued)*

| Target epitope/gene | Oligonucleotide | Sequence |
|---|---|---|
| Cell division control protein (CDC) | Forward primer | 5'-CAAATACGCCATCAGGGAGAACATC-3' |
| | Reverse primer | 5'-CGCTGCCGAAACCACGAGAC-3' |
| Protein transport protein Sec23A (SEC23) | Forward primer | 5'-AGCAATTCGCACAATTATTACAAGCTC-3' |
| | Reverse primer | 5'-GATGCTCACAGAAGACCTGGAAGC-3' |
| Superoxide dismutase [Cu-Zn] (SOD) | Forward primer | 5'-CCTTACTGGACCAAATTCAATTGTTGG-3' |
| | Reverse primer | 5'-GGTGCACACTAACAAGTGATCAAAGATC-3' |
| S-adenosylmethionine decarboxylase (SAD) | Forward primer | 5'-GGCTGGACAAGAAGAAGGCCTCT-3' |
| | Reverse primer | 5'-ATGGATGGTGGAGACGGCAGAT-3' |

The primers and probes were designed to measure the expression levels of the four α-gliadin major T-cell stimulatory epitopes in their canonical form and the expression levels of reference genes. Each TaqMan probe was labeled with the fluorogenic dye FAM (fluorescein) at its 5' end and with the quencher TAMRA (tetramethylrhodamine) at its 3' end.

### *Quantitative PCR*

Amplifications with TaqMan probes were carried out using 10 µl of Takyon™ No ROX Probe 2x Mastermix dTTP Blue (Eurogentec, Belgium), 300 nM of each primer and 100 nM of TaqMan probe labeled with the FAM fluorophore and the TAMRA quencher (Eurofins Genomics, Germany), 400 ng of cDNA and nuclease-free water (Thermo Scientific) for a total volume of 20 µl. For amplifications using the SYBR® dye, each reaction was carried out using 10 µl of Takyon™ No ROX SYBR® 2x MasterMix dTTP Blue (Eurogentec, Belgium), 300 nM of each primer, 400 ng of cDNA and nuclease-free water for a total volume of 20 µl. Samples were loaded in a Hard-Shell® 96-Well PCR skirted white plate and sealed with a Microseal® 'B' PCR Plate Sealing Film (Bio-Rad). PCR amplifications were performed by the C1000 Touch™ Thermal Cycler coupled to the CFX96™ Real-Time detection system (Bio-Rad). The following thermal cycling protocol was used: initial denaturation at 95 °C for 3 min followed by 40 cycles at 95 °C for 10 s and 69 °C for 1 min.

***Validation of primer efficiency and TaqMan probe specificity***

Each set of primers was validated for its efficiency by carrying out qPCR amplifications with successive 10-fold cDNA dilutions. For each qPCR reaction, the Ct (threshold cycle) value, defined as the fractional qPCR cycle number at which the fluorescent signal crosses an arbitrarily placed threshold, was measured for quantification. A calibration curve plotting the logarithm of the cDNA concentration on the x-axis and the corresponding Ct values on the y-axis was drawn. The efficiency was then calculated based on the curve slope: Efficiency (%) = $[(10^{-1/slope})-1] * 100$. The primers were selected so that their efficiency was comprised between 96 and 100%.

The four probes were checked for their specificity by performing qPCR amplifications with previously cloned α-gliadin sequences [16] displaying either the canonical form or one of the allelic variants of the four epitopes (see Additional file 1). The probe specificity was validated when a fluorescent signal was measured only for the α-gliadin clone showing the canonical epitope.

***Design of primers and probes targeting reference genes***

With the aim of accurately normalizing qPCR data, Paolacci et al. [41] evaluated the expression stability of 32 reference genes in different wheat tissues: roots, shoots, stems, flag leaves, spikes, single floral organs and caryopses. On this basis, ten of the most stable genes were selected for the present study, and primers and TaqMan probes were designed to quantify their expression. Primers and probes were developed based on wheat contig sequences from the TIGR wheat Gene index database (ftp://occams.dfci.harvard.edu/pub/bio/tgi/data/Triticum_aestivum/, Table 2).

For each reference gene, TIGR contigs mentioned in Paolacci et al. [41] were aligned with Vector NTI and primers were designed in conserved regions to amplify 150-200 bp-long amplicons. The amplifications were checked in qPCR experiments with SYBR® dye on 400 ng of cDNA from 11 spelt accessions and three diploid species representative of the A, B and D genomes to ensure that the reference

genes are present in the three genomes. Each cDNA sample was loaded in triplicate. The specificity of the primers targeting reference genes was checked by running the qPCR products on a 2% agarose gel.

TaqMan probes were then designed for the reference genes that showed an amplification, and each set of primers was validated for its efficiency (same procedure as detailed above). These probes were tested on the 14 accessions. The geNorm software was used following the procedure detailed by Vandesompele et al. [42] to calculate the M and V values. The M value can be used to classify the reference genes according to their expression stability by the stepwise exclusion in the calculation process of the least stable gene. The V value helps to determine the appropriate number of reference genes to use by analyzing the expression stability of the two most stable genes and by the stepwise addition of the next most stable reference gene to the analysis.

***Epitope expression profiling with designed TaqMan probes***

The four TaqMan probes designed to focus on the canonical epitopes and the four TaqMan probes targeting the four most stable reference genes were applied to the cDNA of the 11 spelt and three diploid accessions. Each accession was tested in two biological replications and each sample was loaded in triplicate. Measured $C_t$ values were reported to $C_t$ values of a calibrator. For the epitope-targeting probes, the calibrator was a previously cloned α-gliadin sequence (GER12_63, accession number KX174081) displaying one copy of each of the four canonical epitopes. This enabled the comparison of expression values from one epitope to another. For each reference gene, the calibrator consisted of the sample with the lowest $C_t$ value, according to the geNorm user manual instructions. To allow comparison from one accession to another, the epitope expression levels were normalized to the expression of the reference genes, thanks to the calculation of normalization factors. The calculation of the epitope relative quantities was carried out using the $2^{\Delta Ct}$ method, where $\Delta C_t$ corresponds to $C_{t,\ calibrator} - C_{t,\ sample}$, and dividing by the normalization factor as recommended by Vandesompele et al. [42].

***Alpha-gliadin expression profiling***

The global α-gliadin expression level was evaluated in each accession by SYBR® qPCR amplifications using primers hybridizing on both sides of the DQ2.5-glia-α1, -α2 and -α3 epitopes (Table 2). Melting curve analyses followed qPCR amplifications to check that only one amplicon was amplified in each sample: the temperature was increased from 65 to 95°C, with increments of 0.5°C for 5 seconds (Additional file 2). Primers designed in TaqMan experiments to target reference genes were also used in qPCR amplification with SYBR® dye for normalization. The sample with the lowest $C_t$ value was chosen as the calibrator for the $2^{\Delta Ct}$ calculation.

# Results

***Development of TaqMan probes specific to the four canonical α-gliadin epitopes***

Four TaqMan probes and primer pairs were designed to focus on the canonical form of the four major α-gliadin T-cell stimulatory epitopes recognized by the immune system of CD patients: DQ2.5-glia-α1, DQ2.5-glia-α2, DQ2.5-glia-α3 and DQ8-glia-α1 epitopes (Table 2). These epitopes are found in α-gliadin sequences from both spelt and bread wheat, as previously stated. In order to avoid any overestimation of the immunogenic content, the probes could not hybridize to any known allelic variant of these epitopes. The specificity of each probe was confirmed by qPCR analyses: a fluorescent signal was clearly observable with the α-gliadin clones displaying the canonical epitope, while it was insignificant or absent with clones containing allelic variants (Figure 1). The specificity of the probes was further confirmed on the cDNA of the three diploid species representative of the ancestral genomes (A, B and D) of spelt and bread wheat (see below). The primer efficiency was comprised between 96.4 and 99.5%, according to the primer pairs (see Additional file 3).

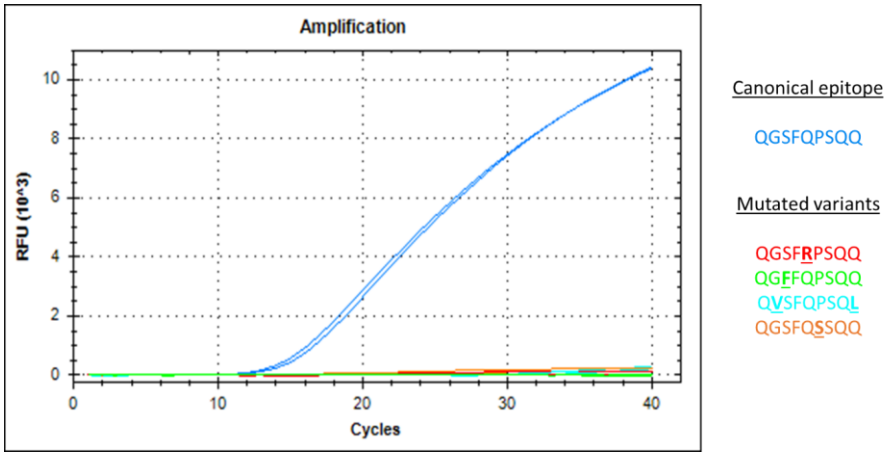



Figure 1. Illustration of the specificity of the TaqMan probe designed to target the canonical DQ8-glia-α1 epitope. The high fluorescent signal observed with the α-gliadin clone displaying the canonical epitope (QGSFQPSQQ) combined with the insignificant or absent fluorescence with the clones containing the allelic variants validated the probe specificity. RFU: Relative fluorescent units.

***Selection of stable reference genes and development of their specific primers and TaqMan probes***

In order to normalize qPCR data, Paolacci et al. [41] evaluated the expression stability of 32 reference genes in different wheat samples. Based on their results, ten of the most stable genes were selected for the present study. Sets of primers and probes were developed to analyze their expression stability in immature seeds harvested 20 DPA from spelt and diploid species representative of the A, B and D genomes (Table 2). Among them, four reference genes (CDC, SOD, SAD and Sec23) were discarded as they showed under- or no expression in one of the three diploid species. Among the six remaining reference genes tested, the analysis with the geNorm software showed that ARF and DUF52 were the two most stable according to their average expression stability (M value, figure 2a), followed by RLI, VAS, GABA and UCC.

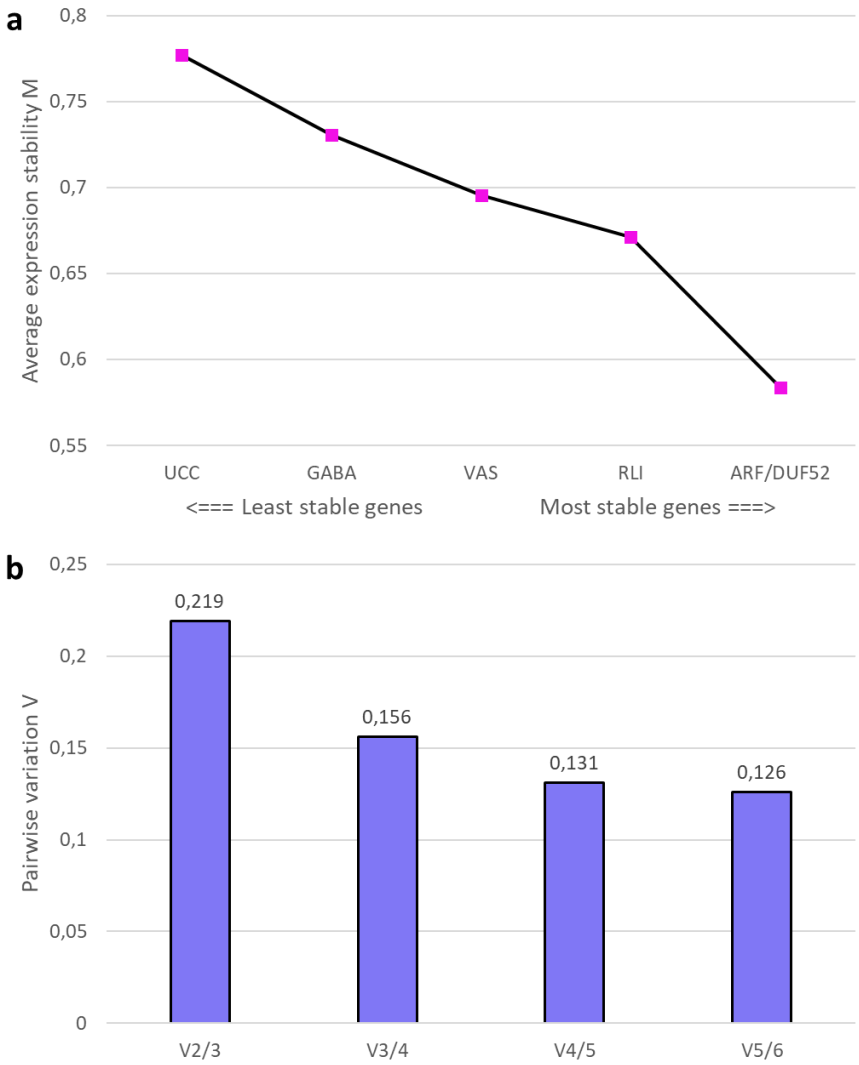



Figure 2. geNorm output charts of M and V values for the six tested reference genes. (a) Average expression stability M of the remaining reference genes after stepwise exclusion of the least stable gene. (b) Pairwise variation V between two sequential normalization factors containing an increased number of reference genes, calculated for the determination of the optimal number of reference genes for normalization.

The evolution of the pairwise variation (V) between two sequential normalization factors ($NF_n$ and $NF_{n+1}$) showed decreasing values when an additional reference gene was used for calculation (Figure 2b). The combination of the four most stable

reference genes resulted in a V value of 0.156 between $NF_3$ and $NF_4$, and was chosen as the best compromise between an optimal cut-off (0.15) below which the inclusion of an additional reference gene is not required, and technical adequacy. The reference genes ARF, DUF52, RLI and VAS were thus selected to normalize qPCR data. The primers targeting these reference genes showed good efficiency (comprised between 97 and 100%, see Additional file 3) and specificity (Additional file 4).

***Epitope expression profile***

The qPCR experiments were carried out on cDNA samples of 11 spelt and three diploid accessions to measure $C_t$ values (Additional file 5). These values were then used to calculate the relative quantity of the four canonical T-cell stimulatory epitopes compared to reference genes (Figure 3). From one accession to another, high variations in the expression levels were displayed but the same trend was observed for each epitope. The analysis of variance showed, from one accession to another, statistically significant differences in the expression levels of the DQ2.5-glia-α1 (ANOVA $F_{13}$ = 4.3829 ; P = 4.874e-3), DQ2.5-glia-α2 (ANOVA $F_{13}$ = 7.9597 ; P = 2.202e-4), DQ2.5-glia-α3 (ANOVA $F_{13}$ = 8.4628 ; P = 1.558e-4) and DQ8-glia-α1 (ANOVA $F_{13}$ = 12.971 ; P = 2.251e-5) epitopes. The spelt accessions SPA03 and US06 showed the highest expression values for the DQ2.5-glia-α1, DQ2.5-glia-α2 and DQ2.5-glia-α3 epitopes, and US06 displayed the highest expression for the DQ8-glia-α1 epitope. Conversely, the Tajik TAD06 and Iranian Iran77d accessions almost always showed the lowest expression levels among the spelt accessions. The diploid accessions displayed more contrasting results. The TR08 accession showed no fluorescence for the DQ2.5-glia-α1, DQ2.5-glia-α2 and DQ2.5-glia-α3 epitopes. In the same way, the LB01 accession showed no expression for the DQ2.5-glia-α2 epitope and almost none for the DQ8-glia-α1 epitope.

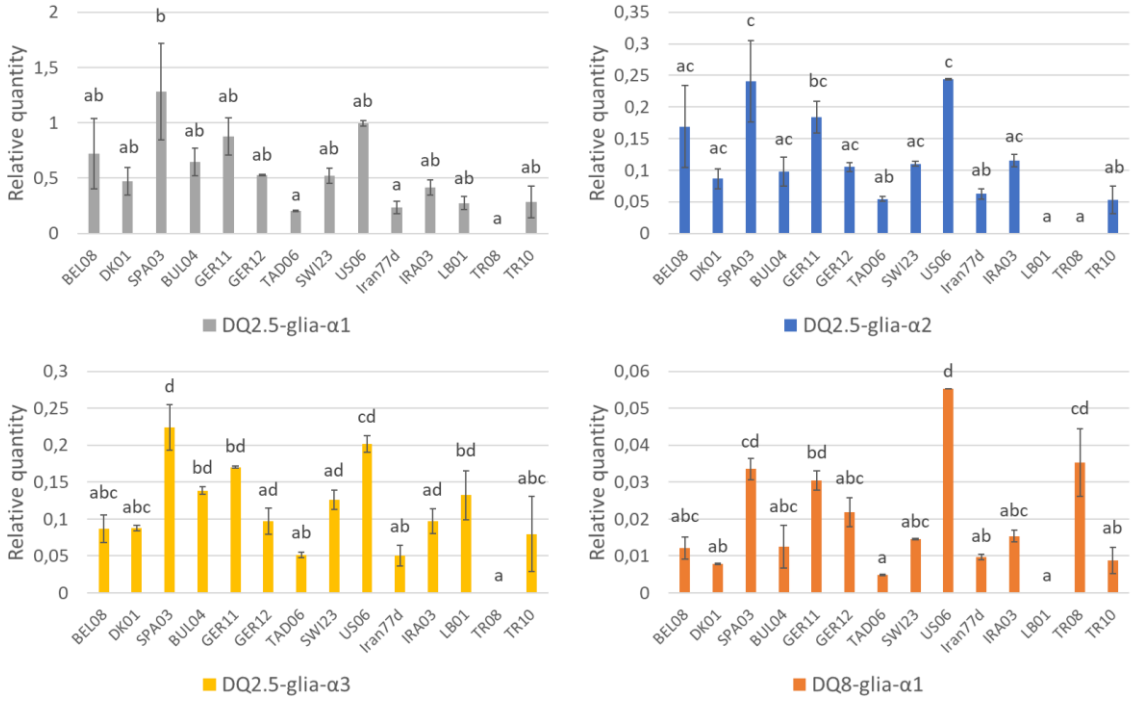



Figure 3. Relative abundance of the four canonical α-gliadin epitopes among contrasted spelt accessions and diploid species. Developed TaqMan probes were used for the relative quantification of the expression of the four major α-gliadin T-cell stimulatory epitopes in their canonical form among 11 contrasted spelt accessions and the three diploid accessions representative of the ancestral genomes of spelt and bread wheat: *Triticum urartu* (LB01, A genome), *Aegilops speltoides* (TR08, B genome) and *Aegilops tauschii* (TR10, D genome). The relative quantities were calculated by dividing the $2^{\Delta Ct}$ values by a normalization factor obtained through the expression analysis of four stable reference genes. Data are presented with standard error of the mean and significant differences detected by Tukey's multiple comparison test are shown by different letters.

The relative quantity of each canonical epitope was compared thanks to the calibration of qPCR values to a unique α-gliadin clone displaying the four canonical epitopes in one copy (Figure 4a). Unsurprisingly, the SPA03 and US06 accessions showed the highest global expression values and the TAD06 and Iran77d accessions the lowest ones. The *Ae. speltoides* accession TR08 showed the highest expression for the canonical DQ8-glia-α1 epitope, but the low relative expression of this epitope compared to the others combined with the absence of expressed canonical DQ2.5-glia-α1, -α2 and -α3 epitopes in this accession led to a very low cumulative expression level for this accession. Intermediate comparable expression levels were measured for the DQ2.5-glia-α2 and DQ2.5-glia-α3 epitopes, whereas the DQ2.5-glia-α1 epitope clearly showed the highest expression compared to the three other epitopes.

***Alpha-gliadin expression analysis***

With the aim of investigating to what extent the measured relative quantities of canonical epitopes are related to the global α-gliadin expression level, an analysis was carried out using the SYBR® dye to detect α-gliadin amplicons, whatever the epitope variants (Figure 4b). Again, the analysis of variance revealed statistically significant differences in the α-gliadin expression levels from one accession to another (ANOVA $F_{13}$ = 5.1367 ; P = 2.256e-3). The diploid species *Ae. speltoides* (TR08) and *Ae. tauschii* (TR10) showed the lowest α-gliadin expression levels whereas a much higher value was reached for the remaining diploid species, *T. urartu* (LB01). In spelt accessions, the American US06 and the Iranian Iran77d accessions were identified as having the highest and the lowest expression levels, respectively. Interestingly, the amount of expressed canonical epitopes did not always correspond to the global expression level of α-gliadin sequences.

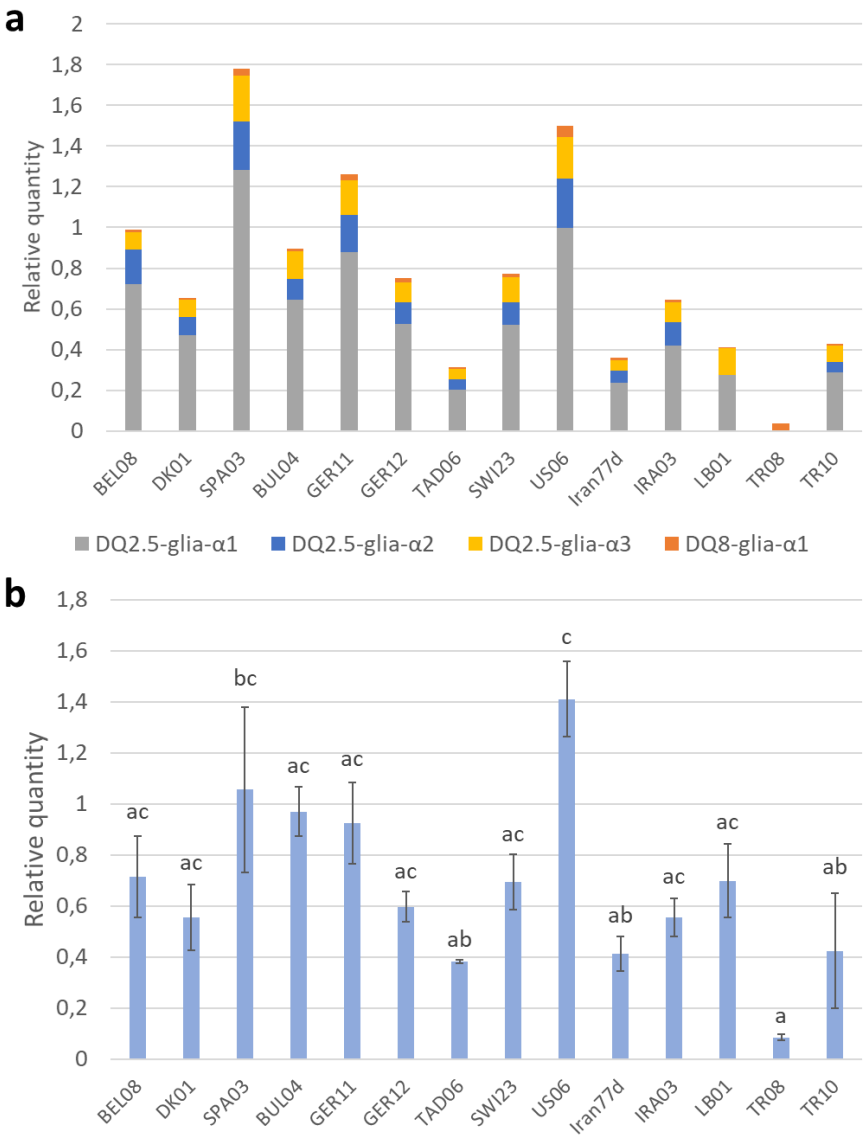



Figure 4. Expression profile of the cumulated canonical epitopes and the global α-gliadins among spelt and diploid accessions. (a) Relative quantification of the expression of the four cumulated canonical α-gliadin epitopes affecting CD patients among the 11 contrasted spelt accessions and the three diploid species *T. urartu*, *Ae. speltoides* and *Ae. tauschii*. The relative quantity of each epitope in the α-gliadin transcripts was calculated by the $2^{\Delta Ct}$ method through the normalization of qPCR values to those obtained for the four most stable reference genes and the calibration to a unique α-gliadin clone displaying the four canonical epitopes in one copy. (b) Relative quantification of the expressed α-gliadin sequences in the same accessions. The sample with the lowest $C_t$ value was chosen as calibrator. Error bars represent the standard error of the mean and the letters displayed above denote significant differences highlighted by Tukey's multiple comparison test.

## Discussion

***Development of epitope-targeting TaqMan probes***

The objective of this study was to develop a tool to detect and accurately quantify the expression of the four α-gliadin toxic epitopes. The developed tool had to discriminate canonical epitopes from their allelic variants because, even if some natural variants can cause a T-cell response in some patients, only the canonical epitopes display a full toxicity [27]. It was therefore decided to develop TaqMan probes given their ability to discriminate among allelic variants differing by only one SNP.

Four TaqMan probes targeting the four α-gliadin canonical epitopes were developed and their specificity was demonstrated with previously cloned α-gliadin sequences. This specificity was confirmed by testing the probes with three diploid accessions (*Triticum urartu*, *Aegilops speltoides* and *Aegilops tauschii*), which are representative of the three ancestral genomes of spelt and bread wheat: the A, B and D genomes respectively. No fluorescent signal was noted for *Ae. speltoides* with the DQ2.5-glia-α1-, DQ2.5-glia-α2- and DQ2.5-glia-α3-targeting probes, and a very low expression level was measured for *T. urartu* with the DQ8-glia-α1-targeting probe (Figure 3). This is consistent with the previously demonstrated absence of canonical DQ2.5-glia-α1, DQ2.5-glia-α2 and DQ2.5-glia-α3 epitopes in B genome α-gliadins [16, 27, 43] and the very low proportion of canonical DQ8-glia-α1 epitope in A genome α-gliadins [44].

The TaqMan probe's specificity was tested with α-gliadin clones, ensuring that the probe does not hybridize to any known allelic variant of the epitopes found in spelt sequences. This specificity makes TaqMan probes a very accurate tool to infer the expression of the different toxic epitopes.

Mitea et al. [27] studied the epitope composition of 3022 expressed α-gliadin sequences originating from 11 bread wheat varieties to analyze the toxicity of each variant of the four epitopes. Almost every allelic variant found in bread wheat

sequences was also identified in spelt sequences [16]. Moreover, the SNPs displayed in the remaining allelic variants in bread wheat sequences are located in the hybridization site of the probes developed in the present study. This makes the TaqMan probes designed in this project useful to study the immunogenic potential of both spelt and bread wheat α-gliadin sequences. Selection of stable reference genes and development of specific TaqMan probes

The quantification of the gene expression level through qPCR requires a normalization step in order to accurately compare samples with one another. Reference genes showing a stable expression are usually used for this purpose. The normalization of qPCR data obtained with TaqMan probes thus required the use of additional TaqMan probes specific to such reference genes. Paolacci et al. [41] studied the expression stability of 32 reference genes using the cDNAs from 24 bread wheat plant samples, including different tissues, developmental stages and temperature stresses. They developed specific primers and used them in qPCR amplifications to assess the expression stability of the reference genes. We selected the most stable reference genes in developing grains on the basis of their results and designed specific primers/TaqMan probes systems to fit with the TaqMan amplification requirements. TaqMan probes specific to six reference genes were developed in this way and the most stable genes as well as the appropriate number of genes to use for normalization were determined. These reliable probes represent a normalization tool that could be useful in other studies where the expression level of sequences other than α-gliadins is analyzed.

***Epitope expression profile in spelt accessions***

The qPCR values measured with the probes focusing on the four canonical epitopes showed very different expression patterns according to the epitope (Figure 4a). The canonical DQ2.5-glia-α1 epitope is by far the most strongly expressed, followed by DQ2.5-glia-α2, DQ2.5-glia-α3 and finally DQ8-glia-α1 to a lesser extent. DQ2.5-glia-α1 and DQ2.5-glia-α2 are overlapping epitopes which are displayed in one, two or three copies in the D genome. Three copies of each epitope lead to the full 33-mer fragment, which is the most immunogenic fragment of α-gliadin sequences [25,

26]. Moreover, the DQ2.5-glia-α1 epitope is encountered in its canonical form in the D but also in the A genome [43]. Combined with its possible duplication or triplication, this explains why the highest expression is observed for this epitope.

The analysis of epitope expression levels in different accessions also highlighted high variations. It has been shown by a sequencing analysis in a previous study [16] that spelt α-gliadins from the A genome were expressed most strongly and could display three out of four epitopes in their canonical form. The same study highlighted that SPA03 and US06 were the two accessions with the highest proportion of A genome α-gliadins. This supports the results obtained in the present study identifying the SPA03 and US06 accessions as those with the highest canonical epitope expression levels. The lowest expression values measured for TAD06 are also consistent with our previous results, in which the greatest proportion of B genome α-gliadins – displaying a high proportion of epitope variants – were found in this Tajik accession.

Unsurprisingly, the amount of expressed canonical epitopes is related to the expression level of α-gliadin transcripts. However, the global amount of expressed α-gliadins does not provide enough information to study the immunogenic content of the accessions, since some opposed trends can be identified between Figures 4a and 4b: the quantification of the global α-gliadin expression level can lead to an under- or an over-estimation of the canonical epitope expression level, as noticed for the SPA03 and BUL04 or LB01 accessions, respectively. Generally speaking, although a correlation might appear for some accessions between the toxic epitope detection and the global α-gliadin expression levels, measures with epitope-targeting probes provide more precise information and demonstrates the usefulness of this TaqMan probes for developing new varieties with reduced immunogenic content without lowering the overall amount of expressed α-gliadins.

Interestingly, the three Asian spelt accessions included in the present study displayed the lowest canonical epitope expression levels, compared to the seven European accessions and the American one – presumably of European origin [14]. Several authors have suggested the existence of two independent origins for spelt

– one in Asia and one in Europe – and genetic differences have been reported [45, 46, 47, 48]. Asian spelt seems to have emerged, like bread wheat, through hybridization between the cultivated free-threshing tetraploid wheat *Triticum turgidum* ssp. *turgidum* (AABB genome) and the wild *Aegilops tauschii* (DD genome), whereas European spelt may be the result of a secondary hybridization between bread wheat and cultivated non-free-threshing tetraploid emmer [*Triticum turgidum* ssp. *diccocum* (Schrank ex Schübler) Thell., AABB genome]. The lower canonical epitope expression levels measured in Asian spelt accessions could be linked to their putative distinct origin from European spelts. This opens up an interesting route to explore the α-gliadin immunogenic content of Asian spelt accessions with the aim of developing safer varieties for CD patients.

***Comparison between the developed TaqMan probe method and other existing techniques***

The developed probes involve the extraction of mRNA from immature seeds which requires a careful handling. However, this methodology enables to quantify the expression of the four major T-cell stimulatory epitopes related to CD independently on the same sample. This tool has been developed to carry out genetic studies of the toxic potential found in α-gliadin transcripts from spelt or bread wheat accessions, with the aim of applying them in breeding programs. It could be useful to track modifications at the sequence or at the expression level in molecular breeding. Conversely, ELISA kits focus only on one or two CD-related epitopes but are not completely specific to canonical epitopes. They are thus mostly intended to measure the global amount of gluten in food samples rather than to quantify the toxic content held in these samples. Beside ELISA, the methodology based on aptamers – binding the immunodominant 33-mer peptide – developed by Amaya-Gonzalez et al. [36] enables to quantify the gluten content through an electrochemical competitive enzyme-linked assay on magnetic particles. The selected aptamer is six times more sensitive than the reference ELISA test and it does not show any cross-reactivity with non-toxic proteins such as those found in maize, soya and rice. Nevertheless, these aptamers do not focus on

individual epitopes and, despite their high sensitivity, the absence of cross-reactivity with allelic variants of the 33-mer has not been demonstrated.

In the recent years, another strategy coupling liquid chromatography (LC) to mass spectrometry (MS) has been developed [34, 35]. Like TaqMan probes developed in this work, the LC-MS technique enables to discriminate canonical epitopes from their allelic variants.

ELISA tests, aptamer assays and LC-MS are intended to detect gluten at the protein level, whereas TaqMan probes focus on α-gliadin transcripts. Given that the component responsible for triggering the disease is the protein, the toxic content of a flour cannot be irrevocably stated from analyses carried out on mRNA samples since slight variations can still occur between the mRNA and the protein stages. However, Van den Broeck et al. [35] showed that their quantification results, based on LC-MS method, were correlated with those of Salentijn et al. [49], who developed an RNA sequencing pipeline to analyze which α-gliadin sequences are being expressed.

The identification or the development of spelt or bread wheat accessions with a reduced immunogenic content is not sufficient for celiac disease patients since the low amount of toxic epitopes will still trigger the disease. However, the developed TaqMan probes could be used in conventional and/or molecular breeding programs (genome editing,…) aiming at developing celiac-safe varieties by tracking the immunogenic potential at the genetic level. This longer-term application of the designed probes could thus find an interest for the development of safe varieties for CD patients.

In addition to the application of the developed probes on α-gliadin transcripts, they could be used as a first approximation on the genomic DNA (gDNA) of spelt and bread wheat accessions if the results obtained from gDNA samples are globally correlated with those obtained from cDNA samples.

## Conclusion

In this study, we have developed a tool to study the expression level of the four major α-gliadin T-cell stimulatory epitopes involved in celiac disease. This was achieved by designing, first, TaqMan probes that only hybridize to the canonical form of each epitope, and, second, TaqMan probes targeting stable reference genes. This work thus provides reliable tools to study the expression of the four α-gliadin canonical epitopes as well as for normalization purposes in studies where the expression level of sequences other than α-gliadins is analyzed. The application of the designed probes to contrasted spelt accessions revealed high variations in the expression levels of the canonical epitopes. In particular, 11 spelt accessions were analyzed and the three Asian spelts showed a low expression of these epitopes compared to the eight European ones. Even if more accessions are needed to draw definitive conclusions, these differing results between European and Asian spelts seem to be a way worth exploring, since it could open up an interesting route in the development of cereal varieties which are safe for CD patients.

## References


1. Shewry PR, Halford NG, Lafiandra D. Genetics of wheat gluten proteins. Adv Genet. 2003;49:111-184.

2. Shewry PR, Halford NG. Cereal seed storage proteins: structures, properties and role in grain utilization. J Exp Bot. 2002;53:947-958.

3. Fasano A, Sapone A, Zevallos V, Schuppan D. Nonceliac gluten and wheat sensitivity. Gastroenterology. 2015;148:1195-1204.

4. Mooney PD, Aziz I, Sanders DS. Non-celiac gluten sensitivity: clinical relevance and recommendations for future research. Neurogastroent Motil. 2013;25:864-871.

5. Sapone A, Bai JC, Ciacci C, Dolinsek J, Green PHR, Hadjivassiliou M, Kaukinen K, Rostami K, Sanders DS, Schumann M, Ullrich R, Villalta D, Volta U, Catassi C, Fasano A. Spectrum of gluten-related disorders: consensus on new nomenclature and classification. BMC Med. 2012;10:13.

6. Scherf KA, Koehler P, Wieser H. Gluten and wheat sensitivities – An overview. J Cereal Sci. 2016;67:2-11.

7. Rewers MJ. Epidemiology of celiac disease: what are the prevalence, incidence, and progression of celiac disease? Gastroenterology. 2005;128:S47-S51.

8. Ciccocioppo R, Di Sabatino A, Corazza GR. The immune recognition of gluten in celiac disease. Clin Exp Immunol. 2005;140:408-416.

9. Stoven S, Murray JA, Marietta E. Celiac Disease: Advances in Treatment via Gluten Modification. Clin Gastroenterol H. 2012;10:859-862.

10. Rashtak S, Murray JA. Review article: coeliac disease, new approaches to therapy. Aliment Pharm Ther. 2012;35:768-781.

11. Tonutti E, Bizzaro N. Diagnosis and classification of celiac disease and gluten sensitivity. Autoimmun Rev. 2014;13:472-476.

12. Shewry PR, Tatham AS. Improving wheat to remove coeliac epitopes but retain functionality. J Cereal Sci. 2016;67:12-21.

13. An X, Li Q, Yan Y, Xiao Y, Hsam SLK, Zeller FJ. Genetic diversity of European spelt wheat (Triticum aestivum ssp. spelta L. em. Thell.) revealed by glutenin subunit variations at the Glu-1 and Glu-3 loci. Euphytica. 2005;146:193-201.

14. Bertin P, Grégoire D, Massart S, de Froidmont D. High level of genetic diversity among spelt germplasm revealed by microsatellite markers. Genome. 2004;47:1043-1052.

15. Caballero L, Martin LM and Alvarez JB. Variation and genetic diversity for gliadins in Spanish spelt wheat accessions. Genet Resour Crop Ev. 2004;51:679-686.

16. Dubois B, Bertin P, Mingeot D. Molecular diversity of α-gliadin expressed genes in genetically contrasted spelt (Triticum aestivum ssp. spelta) accessions and comparison with bread wheat (T. aestivum ssp. aestivum) and related diploid Triticum and Aegilops species. Mol Breed. 2016;36-152.

17. Gomez-Becerra HF, Erdem H, Yazici A, Tutus Y, Torun B, Ozturk L, Cakmak I. Grain concentrations of protein and mineral nutrients in a large collection of spelt wheat grown under different environments. J Cereal Sci. 2010;52:342-349.

18. Ruibal-Mendieta NL, Delacroix DL, Meurens M. A comparative analysis of free, bound and total lipid content on spelt and winter wheat wholemeal. J Cereal Sci. 2002;35:337-342.

19. Escarnot E, Agneessens R, Wathelet B, Paquot M. Quantitative and qualitative study of spelt and wheat fibres in varying milling fractions. Food Chem. 2010;122:857-863.

20. Arentz-Hansen H, Mcadam SN, Molberg O, Fleckenstein B, Lundin KEA, Jorgensen TJD, Jung G, Roepstorff P, Sollid LM. Celiac lesion T cells recognize epitopes that cluster in regions of gliadins rich in prolamine residues. Gastroenterology. 2002;123:803-809.

21. Camarca A, Anderson RP, Mamone G, Fierro O, Facchiano A, Constantini S, Zanzi D, Sidney J, Auricchio S, Sette A, Troncone R, Gianfrani C. Intestinal T cell responses to gluten peptides are largely heterogeneous: implications for a peptide-based therapy in celiac disease. J Immunol. 2009;182:4158-4166.

22. Vader LW, Stepniak DT, Bunnik EM, Kooy YMC, De Haan W, Drijfhout JW, Van Veelen PA, Koning F. Characterization of cereal toxicity for celiac patients based on protein homology in grains. Gastroenterology. 2003;125:1105-1113.

23. 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;95:50-58.

24. Ozuna CV, Iehisa JCM, Gimenez MJ, Alvarez JB, Sausa 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.

25. Molberg O, Uhlen AK, Jensen T, Flaete NS, Fleckenstein B, Arentz-Hansen H, Raki M, Lundin KEA, Sollid LM. Mapping of gluten T-cell epitopes in bread whaet ancestors: implications for celiac disease. Gastroenterology. 2005;128:393-401.

26. Shan L, Molberg O, Parrot I, Hausch F, Filiz F, Gray GM, Sollid LM, Khosla C. Structural basis for gluten intolerance in celiac sprue. Science. 2002;297:2275-2279.

27. Mitea C, Salentijn EMJ, van Veelen P, Goryunova SV, van der Meer IM, et al. A universal approach to eliminate antigenic properties of alpha-gliadin peptides in celiac disease. PLoS ONE. 2010;5:e15637.

28. Skerritt JH, Hill AS. Monoclonal antibody sandwich enzyme immunoassays for determination of gluten in foods. J Agr Food Chem. 1990;38:1771-1778.

29. Valdes I, Garcia E, Llorente M, Mendez E. Innovative approach to low-level gluten determination in foods using a novel sandwich enzyme-linked immunosorbent assay protocol. Eur J Gastroen Hepat. 2003;15:465-474.

30. Moron B, Bethune MT, Comino I, Manyani H, Ferragud M, Lopez C, Cebolla A, Khosla C, Sousa C. Toward the assessment of food toxicity for celiac patients: characterization of monoclonal antibodies to a main immunogenic gluten peptide. PLoS One. 2008;3:e2294.

31. Mitea C, Kooy-Winkelaar Y, van Veelen P, de Ru A, Drijfhout JW, Koning F, Dekking L. Fine specificity of monoclonal antibodies against celiac disease-inducing peptides in the gluteome. Am J Clin Nutr. 2008;88:1057-1066.

32. Scherf KA, Poms RE. Recent developments in analytical methods for tracing gluten. J Cereal Sci. 2016;67:112-122.

33. Kahlenberg F, Sanchez D, Lachmann I, Tuckova L, Tlaskalova H, Méndez E, Mothes T. Monoclonal antibody R5 for detection of putatively coeliac-toxic gliadin peptides. Eur Food Res Technol. 2006;222:78-82.

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

35. Van den Broeck HC, Cordewener JHG, Nessen MA, America AHP, Van der Meer I. Label free targeted detection and quantification of celiac disease immunogenic epitopes by mass spectrometry. J Chromatogr A. 2015;1391:60-71.

36. Amaya-Gonzalez S, de-los-Santos-Alvarez N, Miranda-Ordieres AJ, Lobo-Castanon MJ. Aptamer Binding to Celiac Disease-Triggering Hydrophobic Proteins: A Sensitive Gluten detection Approach. Anal Chem. 2014;86:2733-2739.

37. Pfaffl MW. Quantification strategies in real-time PCR. In: Bustin SA, editor. A-Z of quantitative PCR. La Jolla: International University Line; 2004. p. 87-112.

38. Kim S, Misra A. SNP genotyping: technologies and biomedical applications. Annu Rev Biomed Eng. 2007;9:289-320.

39. Vignal A, Milan D, SanCristobal M, Eggen A. A review on SNP and other types of molecular markers and their use in animal genetics. Genet Sel Evol. 2002;34 :275-305.

40. Malkki M, Petersdorf W. Genotyping of single nucleotide polymorphisms by 5' nuclease allelic discrimination. Method Mol Biol. 2012;882:173-182.

41. Paolacci AR, Tanzarella OA, Porceddu E, Ciaffi M. Identification and validation of reference genes for quantitative RT-PCR normalization in wheat. BMC Mol Biol. 2009;10:11.

42. Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol. 2002;3:research0034.1.

43. Van Herpen TWJM, Goryunova SV, van der Schoot J, Mitreva M, Salentijn E, Vorst O, Schenk MF, van Veelen PA, Koning F, van Soest LJM, Vosman B, Bosch D, Hamer RJ, Gilissen LJWJ, Smulders MJM. Alpha-gliadin genes from the A, B and D genomes of wheat contain different sets of celiac disease epitopes. BMC Genomics. 2006;7:1.

44. Li J, Wang S, Li S, Ge P, Li X, Ma W, Zeller FJ, Hsam SLK, Yan Y. Variations and classification of toxic epitopes related to celiac disease among α-gliadin genes from four Aegilops genomes. Genome. 2012;55:513-521.

45. Blatter RHE, Jacomet S, Schlumbaum A. About the origin of European spelt (Tritcum spelta L.): allelic differentiation of the HMW Glutenin B1-1 and A1-2 subunit genes. Theor Appl Genet. 2004;108:360-367.

46. Dvorak J, Deal KR, Luo M-C, You FM, von Borstel K, Dehgani H. The origin of spelt and free-threshing hexaploid wheat. J Hered. 2012;103:426-441.

47. Nesbitt M. Wheat evolution: integrating archaeological and biological evidence. In: Caligari PDS, Brandham PE, editors. Wheat taxonomy: the legacy of John Percival. London: Linnean Society; 2001. p. 37-59.

48. Zohary D, Hopf M, Weiss E. Domestication of plants in the Old World. 4th ed. Oxford: Oxford University Press ; 2012.

49. Salentijn EMJ, Esselink DG, Goryunova SV, van der Meer I, Gilissen LJWJ, Smulders MJM. Quantitative and qualitative differences in celiac disease epitopes among durum wheat varieties identified through deep RNA-amplicon sequencing. BMC Genomics. 2013;14:905.

## Supplementary materials

***Additional file 1:*** Composition of the canonical form and the allelic variants of the four α-gliadin T-cell stimulatory epitopes used to optimize each TaqMan probe's specificity.

| Epitopes | Variants | Composition [1] | | GeneBank accession numbers [2] |
|---|---|---|---|---|
| DQ2.5-glia-α1 | Canonical epitope | NA | CCA T{T/A}T CC{G/A} CAA CTA CCA TAT | KX173976, KX174272, KX174055 |
| | | AA | P{F/Y}PQPQLPY | |
| | Mutated epitopes | NA | CCA TTT CCG --- CC**A** CAA CTA CCA TAT | KX173988 |
| | | AA | PFP-PQLPY | |
| | | NA | CCA TTT **T**CG CAG CCG CAA CTA CCA TAT | KX173847 |
| | | AA | PF**S**QPQLPY | |
| | | NA | CCA TTT C**T**G CAG CCG CAA CTA CCA TAT | KX174174 |
| | | AA | PF**L**QPQLPY | |
| | | NA | CCA TTT CCG CAG CCG CAA CTA **T**CA TAT | KX174120 |
| | | AA | PFPQPQL**S**Y | |
| | | NA | CCA TAT CCG C**G**G CCG CAA CTA CCA TAT | KX174253 |
| | | AA | PYP**R**PQLPY | |
| | | NA | CCA TAT CCG CAG CCG CA**T** CTA CCA TAT | KX174252 |
| | | AA | PYPQP**H**LPY | |
| DQ2.5-glia-α2 | Canonical epitope | NA | CC{A/G} CAG CCG CAA CTA CCA TAT CCG CAG | KX173976, KX174272, KX174055 |
| | | AA | PQPQLPYPQ | |
| | Mutated epitopes | NA | CCG --- CC**A** CAA CTA CCA TAT CCG CAG | KX173988 |
| | | AA | P-PQLPYPQ | |
| | | NA | CCG CAG CCG CAA CTA CCA TAT **T**C**A** CAG | KX174127 |
| | | AA | PQPQLPY**S**Q | |
| | | NA | CCG C**G**G CCG CAA CTA CCA TAT CCG CAG | KX174253 |
| | | AA | P**R**PQLPYPQ | |
| | | NA | CCG CAG CCG CA**T** CTA CCA TAT CCG CAG | KX174252 |
| | | AA | PQP**H**LPYPQ | |

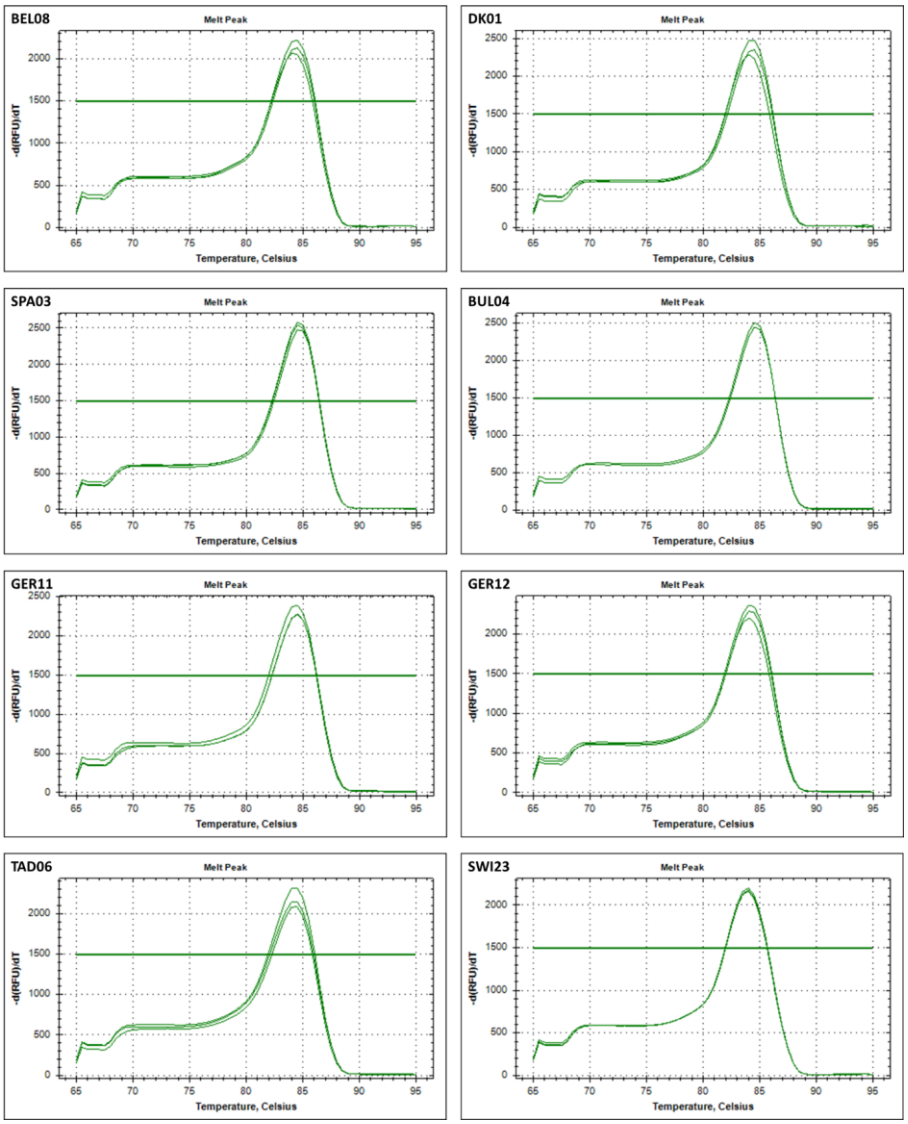
***Additional file 1.*** *(Continued)*

| Epitopes | Variants | Composition [1] | | GeneBank accession numbers [2] |
|---|---|---|---|---|
| DQ2.5-glia-α3 | Canonical epitope | NA | TTT CGA CCA CAA CAA CCA TAT CCA CA{A/G} | KX174127 |
| | | AA | FRPQQPYPQ | |
| | Mutated epitopes | NA | TTT C**C**A CCA CAA CAA CCA TAT CCA CAA | KX173988 |
| | | AA | F**P**PQQPYPQ | |
| | | NA | TTT CGA C**A**A CAA CAA CCA TAT CCA CAA | KX173847 |
| | | AA | FR**Q**QQPYPQ | |
| | | NA | TTT CGA CCA CAA CAA **T**CA TAT CCA CAA | KX174255 |
| | | AA | FRPQQ**S**YPQ | |
| | | NA | TTT CGA CCA CAA **A**AA CCA TAT CCA CAA | KX173906 |
| | | AA | FRPQ**K**PYPQ | |
| | | NA | TTT C**A**A CCA CAA CAA CCA TAT CCA CAA | KX174012 |
| | | AA | F**Q**PQQPYPQ | |
| DQ8-glia-α1 | Canonical epitope | NA | CAG GGC TCC TTC CAG CCA TCT CAG CAA | KX174252 |
| | | AA | QGSFQPSQQ | |
| | Mutated epitopes | NA | CAG GGC TCC TTC C**G**G CCA TCT CAG CAA | KX173906 |
| | | AA | QGSF**R**PSQQ | |
| | | NA | CAG GGC TCC TTC CAG **T**CA TCT CAG CAA | KX173989 |
| | | AA | QGSFQ**S**SQQ | |
| | | NA | CAG GGC T**T**C TTC CAG CCA TCT CAG CAA | KX173976 |
| | | AA | QG**F**FQPSQQ | |
| | | NA | CAG G**T**C TCC TTC CAG CCA TCT CAG C**T**A | KX173988 |
| | | | Q**V**SFQPSQ**L** | |

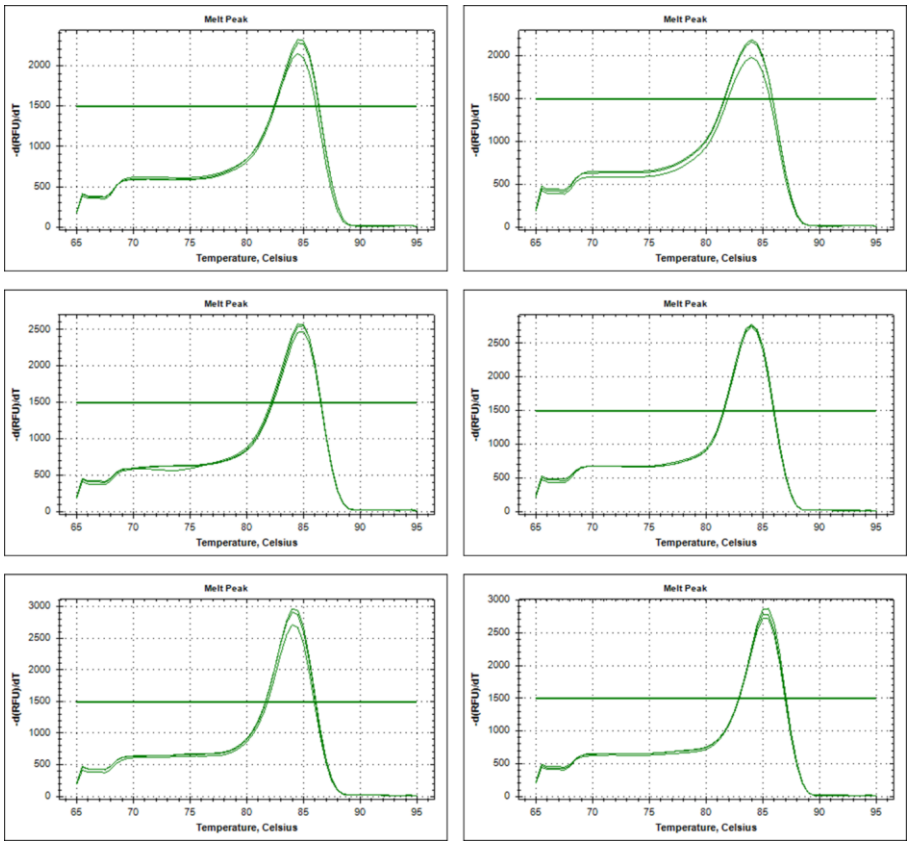
[1] : NA : Nucleic acid composition ; AA : Amino acid composition

[2] : α-gliadin sequences previously cloned in Dubois et al. 2016

***Additional file 2.*** Melting curve analyses carried out after measurement by qPCR of the global amount of expressed α-gliadin sequences in 11 spelt (BEL08, DK01, SPA03, BUL04, GER11, GER12, TAD06, SWI23, US06, Iran77d and IRA03) and three diploid (LB01, TR08 and TR10) accessions representative of the ancestral genomes of spelt and bread wheat.



***Additional file 2.*** *(Continued)*



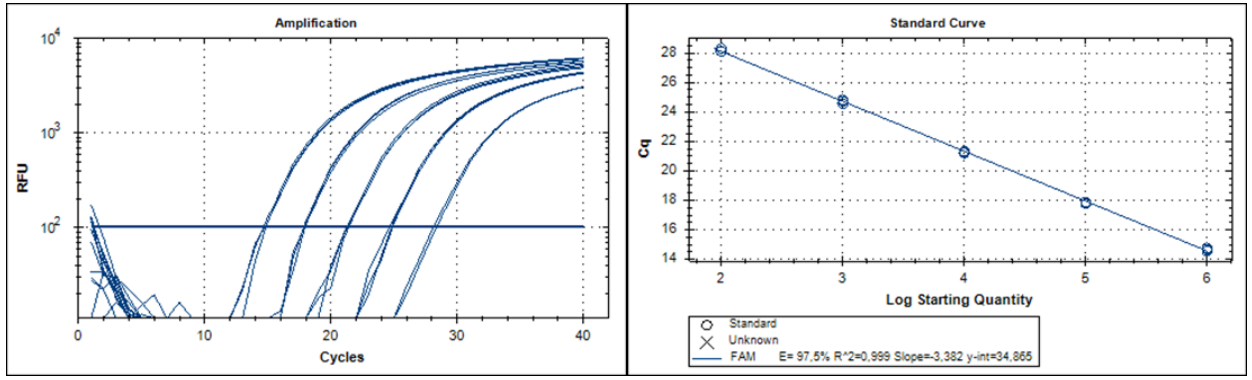
***Additional file 3.*** Primer efficiency calculation for the four T-cell stimulatory epitopes involved in CD and for the four reference genes used to normalize the epitope expression levels.

| | Targeted sequence | Slope | $R^2$ | Efficiency (%) |
|---|---|---|---|---|
| Epitopes | DQ2.5-glia-α1 | -3.382 | 0.999 | 97.5 |
| | DQ2.5-glia-α2 | -3.350 | 0.999 | 98.8 |
| | DQ2.5-glia-α3 | -3.333 | 0.999 | 99.5 |
| | DQ8-glia-α1 | -3.412 | 0.997 | 96.4 |
| Reference genes [1] | ARF | -3.344 | 0.994 | 99.1 |
| | RLI | -3.395 | 0.998 | 97.0 |
| | VAS | -3.322 | 0.998 | 100.0 |
| | DUF52 | -3.347 | 0.994 | 99.0 |

***Additional file 3.*** *(Continued)*

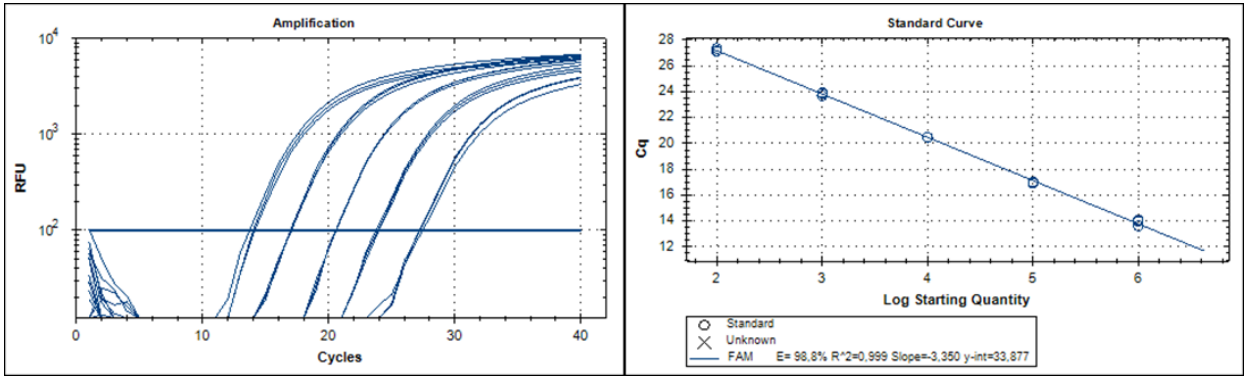
Probes targeting canonical epitopes:

*1. DQ2.5-glia-α1*

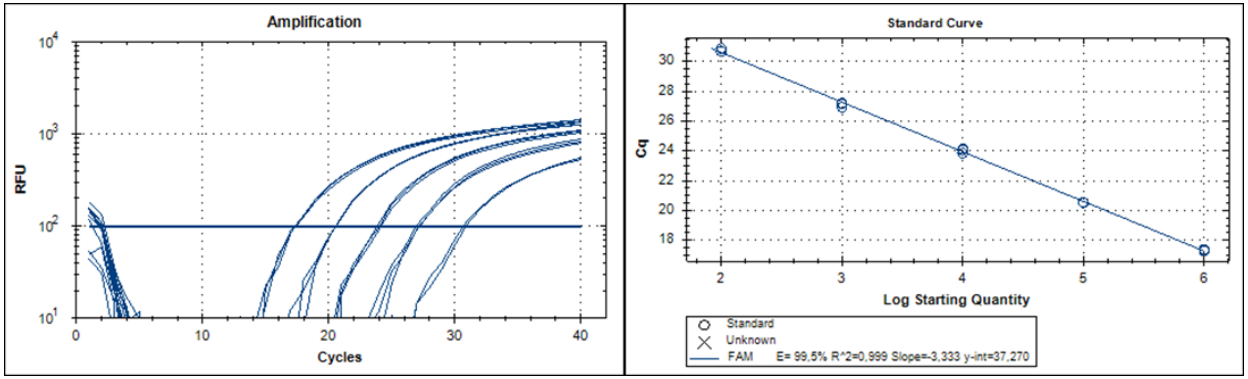


***Additional file 3.*** *(Continued)*

*2. DQ2.5-glia-α2*

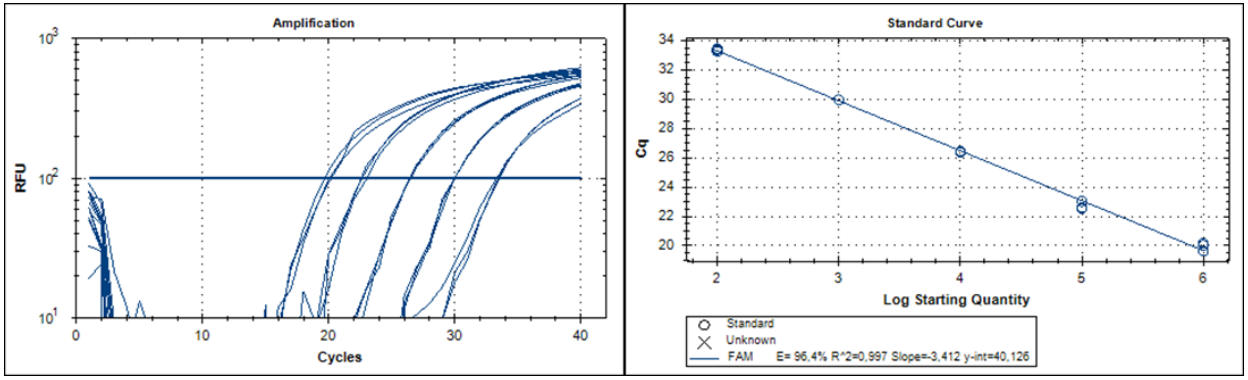



*3. DQ2.5-glia-α3*



***Additional file 3.*** *(Continued)*

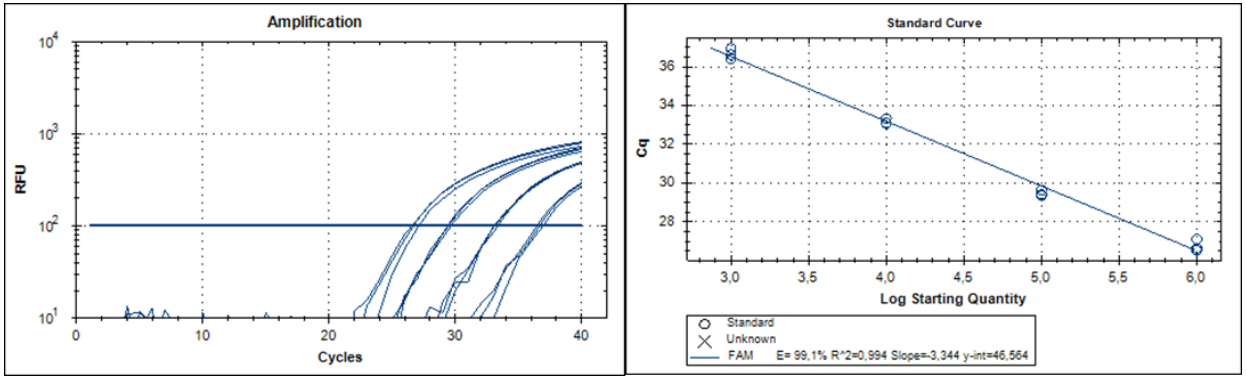
*4. DQ8-glia-α1*



***Additional file 3.*** *(Continued)*

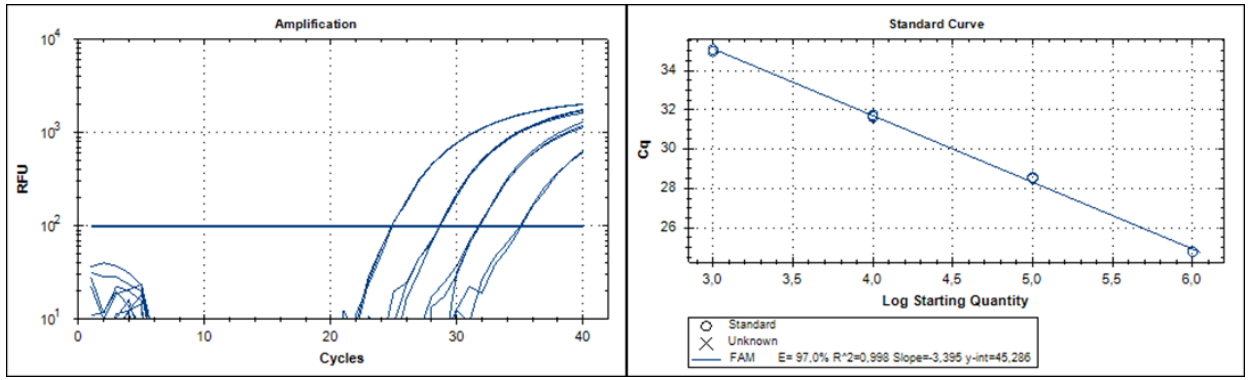
Probes targeting reference genes:

*5. ARF*

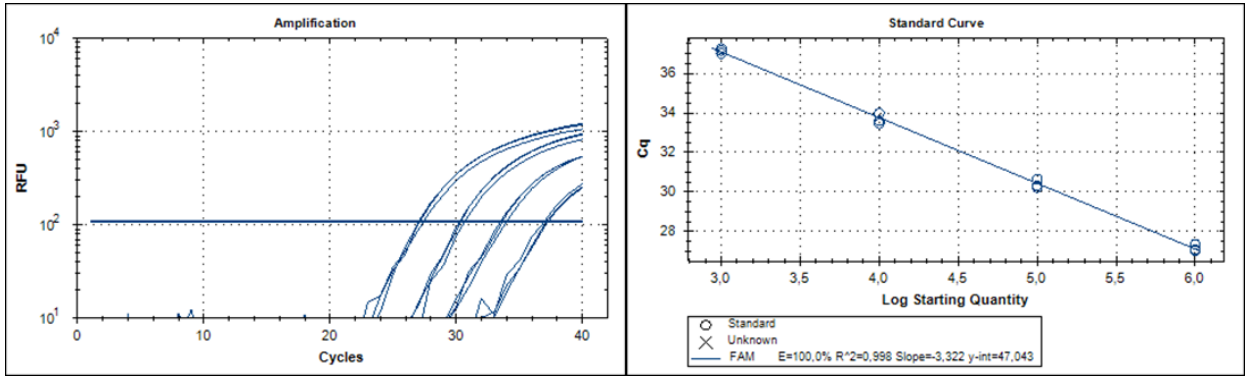


***Additional file 3.*** *(Continued)*

*6. RLI*

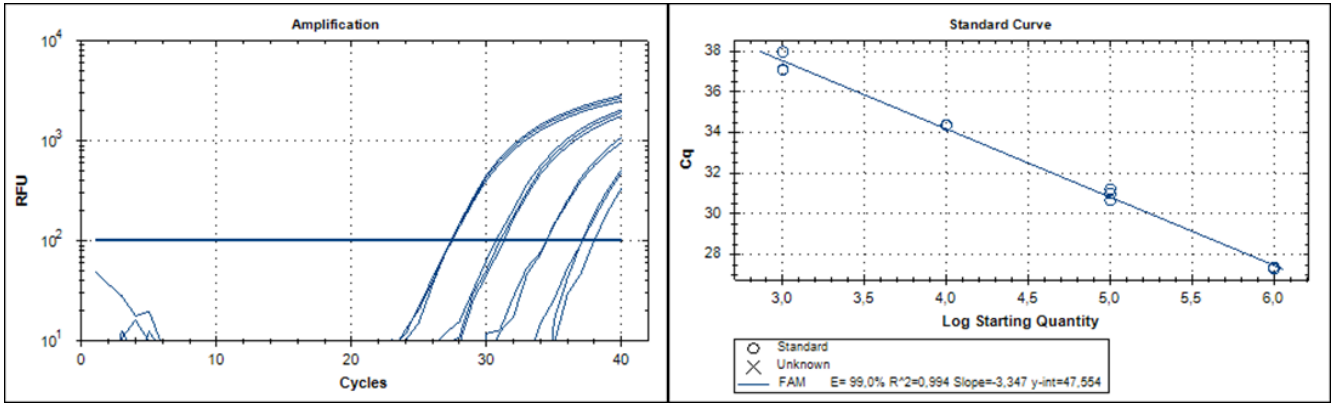



*7. VAS*



***Additional file 3.*** *(Continued)*

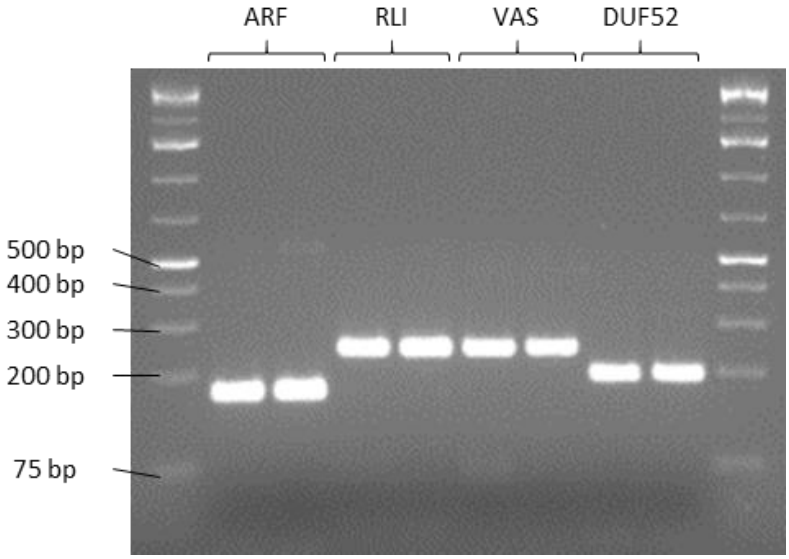
*8. DUF52*




Each set of primers was validated for its efficiency by carrying out qPCR amplifications with successive 10-fold cDNA dilutions. Calibration curves were drawn by plotting the logarithm of the cDNA concentration on the x-axis and the corresponding $C_t$ values on the y-axis. The efficiency was then calculated based on the curve slope: Efficiency (%) = $[(10^{-1/slope})-1] * 100$.

[1] : ARF: ADP-ribosylation factor, RLI: Similar to RNase L inhibitor-like protein, VAS: Vacuolar ATP synthase 16 kDa proteolipid sub., DUF52: Protein of unknown function [DUF52 family].

***Additional file 4.*** Agarose gel electrophoresis of the qPCR products amplified with primers targeting the four most stable reference genes ARF, RLI, VAS and DUF52.




ARF: ADP-ribosylation factor, RLI: Similar to RNase L inhibitor-like protein, VAS: Vacuolar ATP synthase 16 kDa proteolipid sub., DUF52: Protein of unknown function [DUF52 family].

***Additional file 5.*** Average Ct values measured for the four T-cell stimulatory epitopes involved in CD and for the four most stable references genes in 11 spelt accessions and 3 diploid accessions representative of the ancestral genomes of spelt and bread wheat.

| Species | Accession name | TaqMan probes targeting epitopes | | | | TaqMan probes targeting reference genes | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | A20 | A1 | A9 | A2 | ARF | RLI | VAS | DUF52 |
| | BEL08 | 19.62 | 21.90 | 16.59 | 15.15 | 27.64 | 29.92 | 32.03 | 29.36 |
| | DK01 | 18.88 | 21.80 | 16.44 | 15.37 | 26.70 | 29.55 | 31.49 | 28.60 |
| | SPA03 | 18.30 | 20.47 | 15.80 | 14.68 | 28.01 | 29.89 | 31.32 | 30.16 |
| | BUL04 | 21.05 | 24.14 | 18.76 | 17.99 | 29.27 | 32.54 | 34.32 | 31.36 |
| | GER11 | 18.83 | 20.75 | 16.43 | 15.18 | 27.06 | 30.39 | 32.49 | 29.52 |
| *Triticum aestivum* ssp. *spelta* | GER12 | 19.60 | 21.19 | 17.08 | 15.91 | 27.82 | 30.34 | 31.96 | 29.61 |
| | TAD06 | 19.85 | 22.67 | 17.77 | 16.16 | 27.27 | 29.68 | 31.10 | 28.86 |
| | SWI23 | 18.68 | 21.23 | 16.58 | 15.32 | 27.25 | 29.76 | 31.66 | 28.94 |
| | US06 | 19.74 | 18.91 | 17.37 | 15.91 | 29.00 | 31.58 | 33.21 | 30.77 |
| | Iran77d | 21.03 | 22.79 | 18.72 | 17.11 | 28.59 | 30.66 | 32.02 | 30.19 |
| | IRA03 | 19.19 | 21.28 | 17.03 | 15.38 | 27.36 | 30.14 | 31.58 | 29.03 |
| *Triticum urartu* | LB01 | 19.97 | 35.08 | 18.83 | | 27.90 | 32.00 | 33.58 | 29.35 |
| *Aegilops speltoides* | TR08 | | 22.09 | | | 28.07 | 31.41 | 34.68 | 31.83 |
| *Aegilops tauschii* | TR10 | 22.61 | 24.97 | 20.53 | 19.40 | 29.73 | 33.45 | 34.75 | 31.25 |

# CHAPTER 3

Genetic and environmental factors affecting the expression of α-gliadin canonical epitopes involved in celiac disease in a wide collection of spelt (*Triticum aestivum* ssp. *spelta*) cultivars and landraces

The previous chapter was dedicated to the development and the validation of TaqMan probes enabling to quantify the immunogenic content related to CD in α-gliadin sequences. These probes were designed to detect the four main α-gliadin epitopes only in their canonical – *i.e.* immunogenic – form and their specificity was validated using diploid species representative of the A, B and D genomes, as well as α-gliadin clones displaying the canonical form or an allelic variant of each epitope. TaqMan probes targeting potential reference genes were also developed and the expression stability of these genes was evaluated in order to select the most stable ones and to normalize the epitope expression levels. In this next chapter, these probes were used to reach the third and the fourth specific objectives by investigating the CD-related immunogenic potential of spelt according to both genetic and environmental factors. To reach the third objective, the expression of the four main α-gliadin epitopes was first studied in a wide set of spelt accessions representative of different geographical provenances, status and habits. In addition, the relation between the genomic occurrence of sequences encoding for these epitopes and their expression was analyzed. With the aim of reaching the fourth objective, the environmental effect on the epitope expression was studied by quantifying the amount of epitope transcripts in accessions harvested during four consecutive years and by evaluating the influence of different N fertilization strategies.

This chapter has been submitted as an original research paper to the *BMC Plant Biology* journal.

## Abstract

Celiac disease (CD) is an autoimmune disorder affecting genetically predisposed individuals where dietary gluten proteins trigger an inflammatory reaction in the small intestine. Gluten is found in the seeds of cereals like bread wheat (*Triticum aestivum* ssp. *aestivum*) and spelt (*Triticum aestivum* ssp. *spelta*). The development of new varieties lacking immunogenic peptides is one of the strategies currently investigated to face the CD issue. Among gluten proteins, α-gliadins display the strongest immunogenicity with four main T-cell stimulatory epitopes. The objective of this work was to study the expression of α-gliadin epitopes related to CD in a wide collection of 121 spelt accessions (landraces and varieties, spring and winter accessions) from different provenances and to analyze the correlation between the presence of epitope sequences in gDNA and their expression (cDNA). The effect of environmental factors (harvest year and N fertilization) on the epitope expression was also investigated. TaqMan probes targeting the canonical form of the epitopes were used to evaluate the epitope expression levels. Very highly significant variations in the amount of epitope transcripts were highlighted between accessions and according to the provenances. Spring accessions showed a significantly higher immunogenicity than winter ones and no influence of spelt breeding on the epitope expression levels could be assessed comparing landraces and varieties from Northwestern Europe. No correlation was observed between quantitative PCR results obtained from cDNA and gDNA for 45 accessions tested, stressing the need to use markers focusing on epitope transcripts rather than on genomic sequences. A relative stability of the amount of epitopes expressed by a same accession was revealed across harvest years and fertilization strategies. Indeed, no significant difference was highlighted according to the harvest year for 10 accessions harvested during 4 consecutive years. Comparing seven N fertilization modalities applied to two commercial spelt varieties, a slight effect of the fertilization strategy on the amount of epitope transcripts was revealed for one variety while no significant difference was

revealed for the other. This study provides important information to set up the best strategies to develop new varieties tolerated by CD patients.

## Keywords

Spelt, Wheat, α-gliadin, Celiac disease, TaqMan probe, N fertilization, Epitope, Gluten

## Introduction

Celiac disease (CD) is an immune disorder of the upper small intestine triggered by the ingestion of gluten. The inflammatory response of the immune system to the presence of dietary gluten leads to the flattening of intestinal mucosa, resulting in the malabsorption of nutrients. This pathology affects genetically predisposed individuals. Its occurrence has raised during last decades to reach ~1% of human population but most patients affected by CD are still undiagnosed (Fasano et al., 2003; Rubio-Tapia et al., 2012). Gluten is composed of storage proteins from wheat (*Triticum aestivum* L. ssp. *aestivum*), spelt [*Triticum aestivum* ssp. *spelta* (L.) Thell.], barley (*Hordeum vulgare* L.), rye (*Secale cereale* L.) and oat (*Avena sativa* L.). The daily intake of gluten should be kept under 50 mg for CD patients (Catassi et al., 2007) and they must thus follow a strict life-long gluten-free diet to avoid intra- and extra-intestinal symptoms such as diarrhea, bowel pain, fatigue, weight loss, anemia, osteoporosis, headaches and growth retardation (Sapone et al., 2012; Scherf et al., 2016; Tonutti and Bizzaro, 2014).

Gluten proteins are usually classified according to their alcohol solubility into soluble monomeric gliadins and insoluble polymeric glutenins. Gliadins confer viscosity to the dough and are divided into three structural groups according to their electrophoretic mobility: α/β-, γ- and ω-gliadins. Glutenins are responsible for the dough elasticity and are classified into high- and low-molecular-weight glutenin subunits (HMW- and LMW-GS) (Shewry et al., 2003). Gluten proteins display a high content in proline and glutamine amino acids, which make them partially resistant to the gastrointestinal digestion (Shewry and Tatham, 2016).

The α-gliadins are recognized as the most toxic group of gluten proteins affecting CD patients since they trigger the strongest T-cell activation (Arentz-Hansen et al., 2002; Camarca et al., 2009; Ciccocioppo et al., 2005; Vader et al., 2003). The α-gliadins are encoded by a multigene family. Indeed, the Gli-2 loci, where the α-gliadin genes are located, include a number of α-gliadin gene copies that might reach 150 per haploid genome in some accessions (Anderson et al., 1997).

However, it has been shown that a high proportion of these genes are pseudogenes as they display a premature stop codon in their reading frame (Anderson and Greene, 1997; Ozuna et al., 2015). The α-gliadin proteins display four main T-cell stimulatory epitopes involved in CD: the major DQ2.5-glia-α1 and DQ2.5-glia-α2 (P{F/Y}PQPQLPY and PQPQLPYPQ, respectively) and the minor DQ2.5-glia-α3 and DQ8-glia-α1 epitopes (FRPQQPYPQ and QGSFQPSQQ, respectively). In the α-gliadin from the D genome, the DQ2.5-glia-α2 epitope can be displayed in one, two or three copies and leads, when three copies are present, to the most immunogenic fragment of α-gliadin sequences, known as the 33-mer fragment (Molberg et al., 2005; Shan et al., 2002).

Each of the four main α-gliadin T-cell stimulatory epitopes involved in CD can be displayed in its canonical form recognized by the immune system of CD patients. However, allelic variants of these epitopes, resulting from amino acids substitution or deletion, also exists among the α-gliadin genes. Interestingly, these allelic variants have a reduced or suppressed immunogenicity (Mitea et al., 2010). Several techniques have been developed to measure the toxic potential held in the grain of cereals. The first one is the enzyme-linked immunosorbent assay (ELISA), which mainly exists in the form of test kits to quantitate the amount of gluten in food samples and to detect gluten contamination. These kits are based on different antibodies which have been raised against gluten proteins (Scherf and Poms, 2016). However, the restricted specificity of the antibodies regarding canonical epitopes limits the accuracy of ELISA tests to study the immunogenic potential of varieties. Amaya-Gonzalez et al. (2014) developed another technique based on aptamers to detect the 33-mer fragment. This tool enables to quantify the gluten content through an electrochemical competitive enzyme-linked assay on magnetic particles with a six-times improved sensitivity compared to the ELISA reference test. Nevertheless, these aptamers do not focus on individual epitopes and the absence of cross-reactivity with allelic variants of the 33-mer has not been demonstrated. Two other strategies have recently been developed and enable to discriminate canonical epitopes from their allelic variants: liquid chromatography combined to mass spectrometry (LC-MS) on the one hand (Schalk et al., 2017; Schalk et al.,

2018; Van den Broeck et al., 2015), and TaqMan probes on the other hand (Dubois et al., 2017). The LC-MS technique is designed to study the epitope composition at the protein level, whereas TaqMan probes focus on α-gliadin transcripts. The data obtained by these two techniques should provide comparable results given the close relationship highlighted between the results of Van den Broeck et al. (2015), who developed a LC-MS method, and those of Salentijn et al. (2013) who carried out a high throughput RNA sequencing of α-gliadin sequences coming from the same accessions.

Several possibilities are currently being investigated to address the CD problem, like gluten detoxification, modulation of mucosal permeability, antigen presentation blockade, raising monoclonal antibodies against inflammatory cytokines, inhibition of T-cell recruitment or oral tolerance induction (Rashtak and Murray, 2012). Among them, the development of new cereal varieties that lack immunogenic peptide but still display good baking characteristics is a promising approach (Rashtak and Murray, 2012; Shewry and Tatham, 2016). Indeed, bread wheat and related taxa display a high genetic variability which can be exploited. Among them, spelt is also a member of the *Triticum aestivum* species. It displays interesting features such as its adaptability to poor soils, to harsh climatic conditions and its low-input tolerance. Moreover, spelt germplasm collections hold a high genetic diversity (An et al., 2005; Bertin et al., 2004; Caballero et al., 2004; Dubois et al., 2016). Illustrations of this variability were highlighted for features such as the immunogenic potential, the bread-making qualities and the content in proteins, lipids, micronutrients and fibers (An et al., 2005; Dubois et al., 2016; Escarnot et al., 2010; Gomez-Becerra et al., 2010; Ruibal-Mendieta et al., 2002). Furthermore, tools used to study the CD-related immunogenic potential of bread wheat can also be used for spelt since it displays the same epitope allelic variants (Dubois et al., 2016).

Beyond the search for non- or low-immunogenic accessions, attention should also be paid to the putative influence of cultivation techniques on the amount of immunogenic sequences expressed. Among the factors influencing the expression

of gluten proteins, the nitrogen (N) fertilization strategy has been pointed out as an important component (Garrido-Lestache et al., 2004; Wieser and Seilmeier, 1998; Xue et al., 2016). In particular, several studies showed that the N rates applied influenced the expression of α-gliadins (Altenbach et al., 2011; Daniel and Triboi, 2000; Garcia-Molina and Barro, 2017). The expression of the four main α-gliadin epitopes involved in CD could thus be affected by the fertilization strategy since similar trends between the global expression of α-gliadins and the amount of epitope transcripts have previously been observed (Dubois et al., 2017). In consequence, it would be relevant to investigate this hypothesis and to analyze whether an appropriate fertilization strategy could have a beneficial impact to reduce the amount of expressed epitopes.

In this work, the objective was to study the potential of spelt regarding the immunogenicity related to CD as a function of both genetic and environmental factors. This was carried out by (i) analyzing the expression profile of the four main α-gliadin T-cell stimulatory epitopes in a wide collection of spelt accessions using epitope-specific TaqMan probes, (ii) analyzing the correlation between this expression and qPCR results from gDNA samples, and (iii) investigating whether environment and crop practices could have an impact on this expression by testing the expression stability over consecutive years and comparing different N fertilization treatments.

## Material and Methods

### *Plant materials*

A total of 121 spelt accessions (Additional file 1), including cultivars and landraces from the three main spelt breeding countries (Belgium, Germany and Switzerland) and landraces from different provenances in Europe and Asia were studied for their CD-related content. This panel included ten contrasted spelt accessions previously selected as representative of spelt diversity (Dubois et al., 2016) and previously used to validate the development of epitope-specific TaqMan probes (Dubois et al., 2017). All these accessions were grown in field conditions in Gembloux (Belgium) in

2015 (Additional file 2). The growing conditions did not include any pest control and each plot was 1.3 m in length, 0.4 m in width and consisted in two ear-rows. In order to investigate the interannual variation of the epitope expression, the 10 contrasted spelt accessions mentioned above were grown in the field in Gembloux during four years from 2014 to 2017 (Additional file 2). For epitope expression analyses, the ears were enclosed with cellophane bags to ensure self-pollination and immature grains were harvested 20 days post-anthesis (DPA), immediately frozen in liquid nitrogen and stored at -80°C. For studies focusing on genomic DNA (gDNA), shoots were collected after the immature grain harvest, frozen in liquid nitrogen and stored at -80°C as well.

For the trial focusing on the putative influence of the nitrogen (N) fertilizer amount on the epitope expression, two spelt cultivars widely cultivated in Belgium and Germany, i.e. Cosmos and Zollernspelz respectively, were used. The experiment was designed as a randomized complete block consisting of two 1.5 m*10 m blocks. Seven N modalities were applied and included one control without N application and three increasing N amounts (105, 165 and 225 kg N/ha), fractionated into three application periods. The first and second applications were at tillering and first node stages, respectively, and the last application was carried out either at the last leaf stage (LL) or the post anthesis stage (PA). Immature grains were harvested for each modality 10, 15 and 20 DPA.

***RNA extraction, cDNA synthesis and gDNA extraction***

Total RNA was extracted from 100 mg of ground seeds using the NucleoSpin® RNA Plant kit (Macherey-Nagel, Germany) and quantified by spectrometry. First-strand complementary DNA (cDNA) was synthetized from 200 ng RNA using the RevertAid H Minus First Strand cDNA Synthesis Kit (Thermo Scientific) with oligo(dT)18 primer in a volume of 20 µl. The cDNA was then quantified by spectrometry. The gDNA was extracted from 100 mg of ground spelt shoot following a modified version of the Doyle protocol (Doyle, 1991): the 2/3 volumes of cold isopropanol were replaced by the addition of NaCl 2M and 2 volumes of ethanol. The wash buffer was replaced by ethanol 70% and the protocol was terminated by directly storing

the samples at -20°C after the RNase incubation during 30 minutes at 37°C. The extracted DNA was further quantified by spectrometry.

***Expression of the four α-gliadin CD-related canonical epitopes***

Primers and TaqMan probes previously developed to study the expression of the four α-gliadin canonical epitopes involved in CD (DQ2.5-glia-α1, -α2, -α3 and DQ8-glia-α1) and to normalize their expression to the one of stable reference genes (Dubois et al., 2017) were used (Additional file 3). The three most stable reference genes*, i.e*. ADP-ribosylation factor (ARF), similar to RNase L inhibitor-like protein (RLI) and a protein of unknown function [DUF52 family] (DUF), were selected as the most appropriate number of genes to normalize epitope expression values, taking into account calculation accuracy and technical considerations. The primers targeting reference genes were adapted from those designed by Paolacci et al. (2009).

Quantifications were carried out using 10 µl of Takyon™ No Rox Probe 2x Mastermix dTTP Blue (Eurogentec, Belgium), 300 nM of each primer, 100 nM of TaqMan probe labeled with the FAM fluorophore and the TAMRA quencher (Eurofins Genomics, Germany), 2µg of cDNA and nuclease-free water (Thermo Scientific) for a total volume of 20 µl. Two biological replicates of each sample were loaded in triplicate in a Hard-Shell® 96-Well PCR skirted white plate and sealed with a Microseal® 'B' PCR Plate Sealing Film (Bio-Rad). PCR amplifications were performed by the C1000 Touch™ Thermal Cycler coupled to the CFX96™ Real-Time detection system (Bio-Rad). The following thermal cycling protocol was used: initial denaturation at 95°C for 3 min followed by 40 cycles at 95°C for 10 s and 69°C for 1 min. The epitope expression levels were calculated following the 2ΔCt method detailed in Dubois et al. (2017).

***Quantitative PCR on gDNA samples***

With the aim of investigating whether epitope-specific TaqMan probes could directly be used on gDNA, the correlation between the epitope genomic occurrence (gDNA) and their expression (cDNA) was studied. A sampling among the

accessions was thus carried out to study their gDNA epitope content. Based on epitope expression levels measured on the 121 spelt accessions harvested in 2015, five accessions were selected for each of the nine sub-groups (i.e. Belgian cultivars, German cultivars, Swiss cultivars, Belgian landraces, German landraces, Swiss landraces and landraces from Eastern Europe, Spain and Near and Middle East). Within each sub-group, the five selected accessions consisted in those displaying the highest (2), the lowest (2) and a mean (1) epitope expression level. Four hundred ng gDNA were used to quantify the gDNA sequences coding for the four α-gliadin canonical epitopes involved in CD (DQ2.5-glia-α1, -α2, -α3 and DQ8-glia-α1), using the same primers, TaqMan probes and conditions as above.

Results could not be normalized thanks to reference genes since these genes were selected as stable reference genes based on their expression rather than on the genomic occurrence. In consequence, the equal amount, from one sample to another, of gDNA used as template in qPCR reactions was ensured by strict spectroscopy measurements.

The putative correlation between results obtained from cDNA and gDNA samples was evaluated using the non-parametric Spearman's rank correlation test and the *cor.test* function.

***Data analysis***

The deviation between expression values for biological replicates were expressed as standard deviations. The significance of the observed differences ($P < 0.05$) was assessed through ANOVA tests using the *lm* function (normal linear models) and time was considered as a fixed factor. Multiple comparison tests (Tukey) were performed thanks to the *glht* function. The resulting p-values were adjusted for multiple comparisons by a Bonferroni correction. The putative correlation between results obtained from cDNA and gDNA samples was evaluated using the non-parametric Spearman's rank correlation test and the *cor.test* function. All statistical analyses were performed in R (R Development Core Team). All assumptions were checked graphically.

## Results

***Epitope expression profiling in a wide set of spelt accessions***

The expression levels of the four main canonical epitopes involved in CD was determined on 121 spelt accessions, including cultivars and landraces from Northwestern Europe (Belgium, Germany, Switzerland) as well as landraces from Eastern Europe, Spain and Near and Middle East (Figure 1). A very high variability in the epitope expression levels was revealed with cumulated relative quantities ranging from 0.14 for the Belgian accession W-BE24 to 3.97 for the Spanish accession S-ES06. Spanish accessions globally displayed the highest epitope expression levels (Figure 2A) and the analysis of variance revealed very highly significant differences between accession groups (P = 2.666e-11).

The putative effect of breeding strategies on the amount of CD-related epitope transcripts was investigated by comparing epitope expression levels measured on cultivars and landraces from a same geographical provenance, i.e. Belgium, Germany and Switzerland (Figure 2A and 2B). No statistical difference was found in expression level neither between cultivars and landraces taking all three countries together (P = 0.576, figure 2B), nor between cultivars and landraces taking each country separately (Belgium: P = 0.63, Germany: P = 0.248, Switzerland: P = 0.993) (Figure 2A).

The influence of the habit, i.e. winter-sown or spring-sown accessions, was studied by comparing epitope expression levels from all winter spelt accessions against those of all spring accessions (Figure 2C). This resulted in a very highly significant difference (P = 2.634e-8) between this two groups, the spring accessions displaying

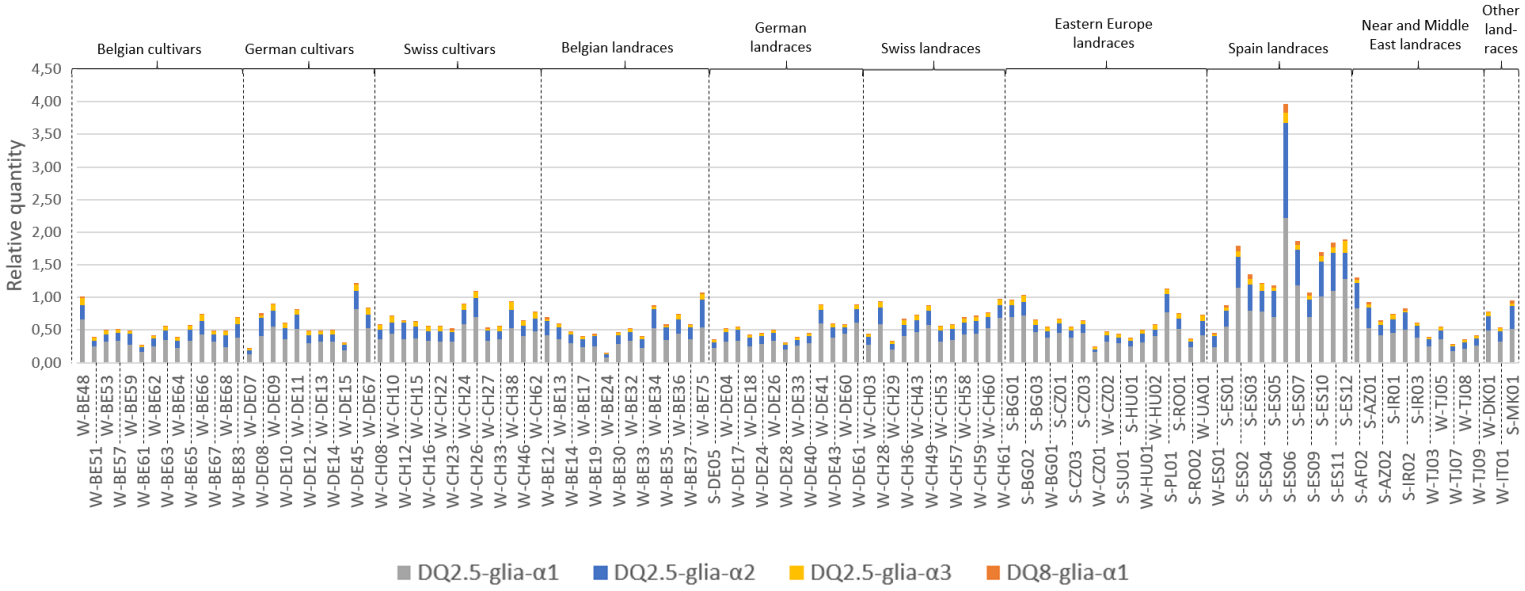



Figure 1. Epitope expression profile of 121 spelt cultivars and landraces from different geographical provenances through the use of specific TaqMan probes targeting the canonical form of the four main α-gliadin epitopes involved in CD. The relative quantities were calculated by dividing the 2ΔCt values by a normalization factor obtained through the expression analysis of stable reference genes.

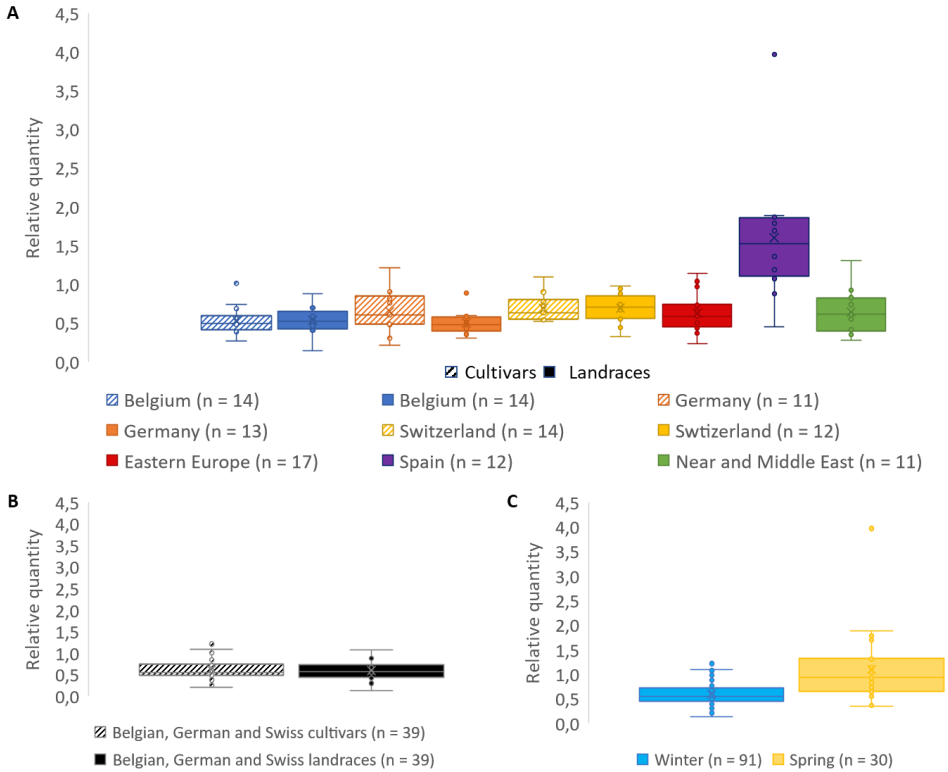



Figure 2. Relative quantities of epitope transcripts in different subsets of spelt cultivars and landraces, divided according to (A) their geographical provenance, (B) their breeding status, and (C) their habit.

higher epitope expression levels than the winter ones. In addition, this trend was also noticed within sub-groups representing geographical provenances, i.e. Eastern Europe, Spain and Near and Middle East.

***Quantitative PCR on gDNA samples and correlation with epitope expression***

Measures of the gDNA epitope content were carried out on a sub-group of accessions representative of the geographical provenance, the breeding status and the epitope expression levels. Epitope relative quantities measured on gDNA samples ranged from 0.11 for the Belgian cultivar W-BE48 to 0.63 for the German landrace W-DE28 (see additional file 4). The comparison of the accession habit showed that winter accessions displayed a significantly higher genomic epitope content (0.37 ± 0.12) than the spring ones (0.21 ± 0.04) (P = 5.286e-4). However, no significant correlation was found between cDNA and gDNA results (r = -0.222; P = 0.14). Moreover, spring accessions displayed higher epitope expression levels but lower genomic epitope occurrence than winter ones.

***Interannual epitope expression analysis***

The interannual variations of the epitope expression levels were investigated with samples harvested in 2014, 2015, 2016 and 2017 for 10 contrasted spelt accessions previously studied (Dubois et al. 2016, Dubois et al. 2017) (Figure 3). A two-way ANOVA revealed a very highly significant effect of the accession, but none for the harvest year. Moreover, a significant interaction between these two factors was pointed out (Additional file 5).

No significant interannual variation was found for eight accessions while slight variations were observed for the W-DE61 and W-TJ08 accessions (P = 1.445e-2 and 8.341e-5, respectively). As shown in the figure 3, the general ranking of the accessions was conserved from one harvest year to another, with the Spanish S-ES04 and the Tajik W-TJ08 accessions showing the highest and the lowest epitope expression levels, respectively. A deeper analysis revealed significant differences in the epitope expression, whatever the harvest year.

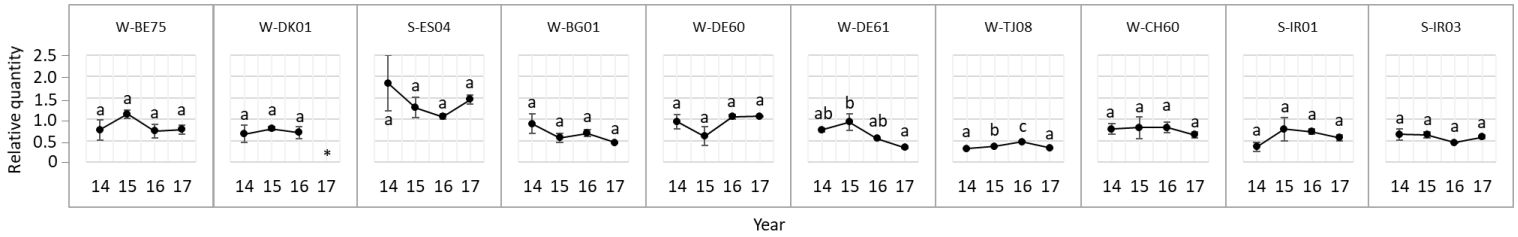



Figure 3. Study of the expression of the four cumulated α-gliadin canonical epitopes involved in CD in 10 contrasted spelt accessions harvested in 2014, 2015, 2016 and 2017. Intra-annual significant differences between accessions are denoted by different letters. *: Missing data.

***Epitope expression according to harvest date and N fertilization***

The epitope expression analysis showed very highly significant differences between samples harvested 10, 15 and 20 DPA for both studied spelt cultivars Cosmos and Zollernspelz (P = 2.2e-16 and P = 3.437e-16, respectively), with increasing epitope expression levels from early to later harvest dates (Figure 4).

For Cosmos, no statistical difference appeared between modalities of N fertilization, neither for samples harvested 10 DPA (P = 0.073), 15 DPA (P = 0.485) or 20 DPA (P = 0.124).

For Zollernspelz at 20 DPA, the treatment modalities 165N-LL and 225N-PA (*i.e.* total amount of N applied of 165 and 225 kg/ha with the last of the three applications carried out at the last leaf and the post anthesis stage, respectively) gave rise to a significantly higher epitope expression compared to the other ones, whereas the modality 105N-PA yielded the lowest one (P = 4.31e-3). On the contrary, no difference between modalities was pointed out for samples collected 10 DPA (P = 0.13) and 15 DPA (P = 0.619).

## Discussion

***CD-related epitope expression in a wide spelt collection***

The main objective of this work was to gain insight the potential of spelt regarding its immunogenicity for CD patients. To reach this objective, a collection of 121 spelt accessions, representative of different geographical provenances, status (cultivar and landrace) and habits (winter and spring) was assembled and their expression profile of CD-related epitopes was evaluated.

The profiling of the epitope expression among the 121 spelt accessions revealed very high variations, with a factor 28.4 between the lowest and the highest epitope expression level measured, which is in line with the high level of spelt genetic diversity previously reported according to microsatellite markers (Bertin et al., 2004), glutenin subunit variation (An et al., 2005; Caballero et al., 2004) and

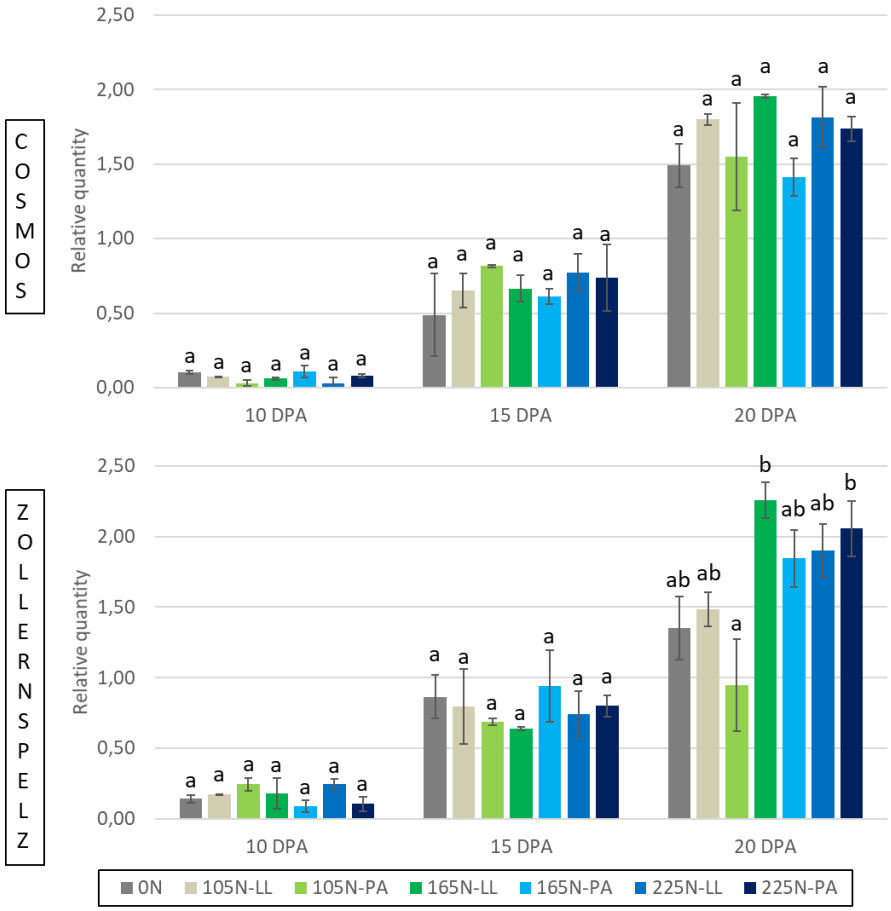



Figure 4. Epitope expression levels measured after different N fertilization strategies for Cosmos and Zollernspelz spelt cultivars. Seven N modalities were tested and included one control without N application (0N) and three increasing N amounts (105, 165 and 225 kg N/ha), fractionated into three application periods. The last application was carried out either at the last leaf stage (LL) or the post anthesis stage (PA). The expression levels were studied on seed samples harvested 10, 15 and 20 DPA. Within each of these harvest modalities, significant differences are denoted by different letters.

variance in agronomic traits (Longin and Würschum, 2014). Highly significant differences between accession provenances were highlighted. The Spanish and Tajik accessions stood out from the others by, respectively, high and low amounts of epitope transcripts measured for each accession of these provenances. This feature can be linked to the work of Bertin et al. (2004). Studying the genetic diversity and structure of the spelt gene pool by using microsatellite markers, they

showed that almost all Spanish and Tajik accessions clustered in two distinct groups and clearly set apart from spelts of other geographical provenance. In addition, these results are also in line with previous studies which focused on only one accession of each geographical provenance (Dubois et al., 2016; 2017).

The main spelt breeding programs in the world have been led in Belgium, Germany and Switzerland, with official research programs starting around 1950. The epitope expression levels of cultivars and landraces from these countries were compared to each other to highlight a putative effect of spelt breeding on CD immunogenicity. The statistical analysis did not show any difference between these two groups. Thus, the selection carried out in spelt breeding programs do not seem to influence the epitope expression. This is in accordance with the results of Kasarda et al. (2013) on wheat, who showed that breeding during the 20$^{th}$ century did not increase the gluten content in the United States. More recently, Schalk et al. (2017), using a liquid chromatography tandem mass spectrometry assay, did not show any higher amount of 33-mer peptide in modern than in old wheat cultivars. In contrast, Ribeiro et al. (2016) showed in ELISA assays that landraces presented a higher reactivity than varieties in bread wheat. However, they used the R5 monoclonal antibody, which is usually used to quantitate the amount of gluten in food samples and not specific to CD-related epitopes, thus revealing both immunogenic and non-immunogenic epitopes altogether.

The analysis of expression data according to the habit, *i.e.* winter or spring spelt, showed that spring accessions displayed on average a higher amount of epitope transcripts than the winter ones, whether provenances were analyzed altogether or separately. These differences in the epitope expression can find two explanations. Firstly, they could be linked to the differences in the gene pools used in breeding programs for spring vs winter accessions. Secondly, they could also be attributed to the differential environmental effect due to the sowing date (October vs March, respectively). The epitope expression levels may indirectly give an idea of the α-gliadin expression, as shown in Dubois et al. (2017), and could thus be linked to the bread-making properties. Interestingly, Maghirang et al. (2006), comparing

spring and winter bread wheat accessions, showed that the grains and flour of spring accessions displayed better bread-making qualities than winter ones. If these higher quality properties go hand in hand with a higher immunogenicity, it thus means that, in the framework of CD, it is more interesting to deal with winter spelt varieties than spring ones, which is already the case in spelt breeding countries like Belgium, Germany and Switzerland, to develop low celiac potential accessions.

***Evaluation of the putative correlation between expressed and genomic epitope sequences***

The aim was to investigate the relation between the genomic occurrence of sequences encoding for immunogenic epitopes and their expression. In addition, this comparison also enabled to analyze whether the epitope-specific TaqMan probes could be used on gDNA, which would be simpler and cheaper than on cDNA, to have an accurate idea of the accession immunogenicity. The qPCR results obtained from gDNA samples of 45 accessions were compared to those from cDNA samples. Measured relative quantities showed a factor 28.4 (from 0.14 to 3.97) and 5.7 (from 0.11 to 0.63) between the lowest and the highest values for cDNA and gDNA samples respectively. This indicates that the use of epitope-specific TaqMan probes on gDNA leads to an underestimation of the variability in epitope content among the accessions. Moreover, the analysis of variance revealed an absence of correlation between these two sets of data. Thus, using epitope-specific TaqMan probes on gDNA does not constitute a valid alternative to the epitope expression study, when aiming at evaluating the immunogenicity of an accession.

The absence of correlation between the two sets of data (cDNA vs gDNA) is striking but could be explained by the multigenic character of the α-gliadin family and the abundance of pseudogenes. Previous studies pointed out the high and variable number of α-gliadin coding sequences held in the haploid genome of bread wheat, ranging from 25 (Harberd et al., 1985) to 150 (Anderson et al., 1997), depending on the variety. Among these coding sequences, it was demonstrated by Anderson and Greene (1997) that about half of them were pseudogenes, due to the presence of

at least one premature stop codon (PSC). More recently, Ozuna et al. (2015) found pseudogenes proportions of 39, 76 and 63 % in the genomes of diploid, tetraploid and hexaploidy wheat species, respectively. When measuring the genomic epitope content with the TaqMan probe/primer systems used in this work, pseudogenes are also detected, unless the PSC is located in the epitope sequence targeted by the TaqMan probe or the primer hybridization site. On the contrary, pseudogenes are almost absent in cDNA samples (Dubois et al., 2016) and are thus not detected when studying the epitope expression profile on the cDNA. This may explain the absence of correlation highlighted in this study between the occurrence of α-gliadin epitope sequences in the spelt genome and their expression levels.

Despite the absence of correlation between qPCR values from gDNA and cDNA samples, spring spelt accessions displayed a significantly lower occurrence of epitope sequences in their genome, while their expression of epitope transcripts was significantly higher than in winter spelts. This could be explained by two hypotheses: either the putative higher amount of pseudogenes in winter spelt accessions and/or the spring environmental conditions that would favor the epitope expression. These observations underline the complexity to lead breeding strategies when dealing with the CD problematics, and demonstrate the necessity to work with molecular markers focusing on epitope transcripts rather than on genomic sequences.

***Influence of harvest year and impact of N fertilization***

It has already been shown that environmental conditions, and thus the harvest year, can influence gluten composition in bread wheat (Noma et al., 2015; Schalk et al., 2017) and that the N fertilization strategy (amount and splitting) can impact the grain protein content and the relative abundances of the different seed storage proteins (Altenbach et al., 2011; Wieser and Seilmeier, 1998; Xue et al., 2016)

To investigate whether the harvest year could have an impact on the immunogenicity level, the expression of the four major α-gliadin canonical epitopes was evaluated in 10 contrasted accessions harvested during four consecutive years.

This revealed a significant effect of the accessions on the epitope expression levels, whereas no impact of the harvest year was pointed out. This contrasts with the results of Schalk et al. (2017) who showed that the harvest year significantly influenced the 33-mer content in four bread wheat cultivars grown at the same location in Hungary and harvested in 2011, 2012 and 2014. The authors even suggested that the environmental factor had a greater influence on the 33-mer content than the genetic background of the four wheat cultivars. However, these results are not really comparable since the years and locations differ; these discrepancies could be partly attributed to the different growing conditions as Hungary's continental climate may more often lead to periods of drought.

Interestingly, the analysis of climatic conditions for each year in Gembloux during the grain filling period revealed some differences, especially between the 2016 and 2017 harvest years (see additional file 2). Whereas the climatic conditions have been by far the wettest during the 2016 grain filling period, climatic data for 2017 display the opposite trend with a higher mean temperature and a low level of rainfalls. Despite these fluctuating meteorological conditions, no significant differences in the epitope expression levels have been pointed out according to the harvest year.

However, the statistical analysis still highlighted an interaction between the accession effect and the harvest year. This means that, even if the harvest year parameter taken alone does not have a significant influence, the epitope expression evolution from one year to another is not the same for each accession, some accession/harvest year combinations leading to significantly different epitope expressions levels than others.

The influence of N fertilization on the amount of epitope transcripts was evaluated on one Belgian and one German spelt cultivars, i.e. Cosmos and Zollernspelz, respectively. Seven N modalities, differing by the N amount and splitting, were applied. This split application of N fertilizer has been pointed out as an effective way to improve wheat protein content (Garrido-Lestache et al., 2004). Xue et al. (2016) showed that N splitting changes the grain protein composition by increasing

the proportions of gliadins and glutenins, leading to an improved baking quality of wheat flour. They even postulated that N splitting is a more effective way to improve wheat quality than the increase in the N applied, offering the potential to reduce the amount of N fertilizer.

The highest epitope expression levels were reached 20 DPA, whatever the accession or the N treatment modality. This is in accordance with the α-gliadin expression peak previously observed around 20 DPA in bread wheat (Li et al. 2013). The analysis of variance revealed a slight effect of the fertilization strategy on the amount of epitope transcripts measured at 20 DPA for Zollernspelz, but not for Cosmos. Wieser and Seilmeier (1998) also noticed that the degree of the N fertilization effects on the quantities and proportions of flour protein groups were strongly dependent on the variety. The absence of difference in epitope expression for Cosmos cultivar can be linked to the work of Garcia-Molina and Barro (2017) who worked with the bread wheat cultivar Bobwhite and showed that the application of increasing amounts of N fertilizer did not modify the α-gliadin content for some fertilizing strategies. For Zollernspelz, the global trend – *i.e.* increasing epitope expression when raising the N fertilization – is in line with the accumulation of gliadins and total prolamin content when increasing N fertilization highlighted in several studies (Matre et al., 2006, Triboï et al., 2003, Xue et al., 2016).

## Conclusion

The development of new varieties lacking immunogenic peptides is one of the strategies currently investigated to face the CD issue. The knowledge of the available diversity is a crucial information for breeding purposes. This work provides useful information about the diversity held in a wide collection of spelt cultivars and landraces from different geographical provenances, in terms of CD-related epitope expression levels. Some accessions were shown to display a low CD-related content but they cannot be safely consumed by celiac patients since even low amounts of immunogenic epitopes still stimulate the immune system. This work, however, provides important knowledges about the available plant material which could be mobilized in breeding programs combining traditional and new promising molecular approaches. A correlation analysis revealed no correlation between the epitope contents measured on cDNA and gDNA samples of identical accessions. This could be linked to the multigenic character of the α-gliadin family and the high number of pseudogenes, making conventional breeding alone probably not efficient enough to develop celiac-safe varieties. However, several promising molecular approaches are currently being investigated in bread wheat or barley, like RNA interference strategies with hairpin RNA constructs (Becker et al., 2012; Gil-Humanes et al., 2010) or antisense constructs (Lange et al., 2007), and the generation of deletion lines lacking loci encoding for α-gliadins (Camerlengo et al., 2017). In addition, new genome editing technologies have been shown to be efficient to significantly reduce the α-gliadin immunogenicity (Sanchez-Leon et al., 2018; Smulders et al., 2017). The epitope expression stability, which was evaluated by harvesting seeds during four consecutive years and by studying the influence of different N fertilization strategies, did not reveal any major fluctuation. In consequence, lowering the CD-related epitope by modifying environmental conditions and/or applying particular N fertilization would probably be inefficient.

## References

Altenbach, S. B., Tanaka, C. K., Hurkman, W. J., Whitehand, L. C., Vensel, W. H., Dupont, F. M. (2011). Differential effects of a post-anthesis fertilizer regimen on the wheat flour proteome determined by quantitative 2-DE. Proteome Sci. 9, 46. doi: 10.1186/1477-5956-9-46

Amaya-Gonzalez, S., De-los-Santos-Alvarez, N., Miranda-Ordieres, A. J., Lobo-Castanon, M. J. (2014). Aptamer binding to celiac disease-triggering hydrophobic proteins: a sensitive gluten detection approach. Anal. Chem. 86, 2733-2739. doi: 10.1021/ac404151n

An, X., Li, Q., Yan, Y., Xiao, Y., Hsam, S. L. K., Zeller, F. J. (2005). Genetic diversity of European spelt wheat (*Triticum aestivum* ssp. *spelta* L. em. Thell.) revealed by glutenin subunit variations at the Glu-1 and Glu-3 loci. Euphytica 146, 193-201. doi: 10.1007/s10681-005-9002-6

Anderson, O. D. and Greene F. C. (1997). The α-gliadin gene family. II. DNA and protein sequence variation, subfamily structure, and origins of pseudogenes. Theor. Appl. Genet. 95, 59-65. doi: 10.1007/s001220050532

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. Theor. Appl. Genet. 95, 50-58. doi: 10.1007/s001220050531

Arentz-Hansen, H., McAdam, S. N., Molberg, O., Fleckenstein, B., Lundin, K. E. A., Jorgensen, T. J. D. et al. (2002). Celiac lesion T cells recognize epitopes that cluster in regions of gliadins rich in proline residue. Gastroenterology 123, 803-809. doi: 10.1053/gast.2002.35381

Becker, D., Wieser, H., Koehler, P., Folck, A., Mühling, K. H., Zörb, C. (2012). Protein composition and techno-functional properties of transgenic wheat with reduced α-gliadin content obtained by RNA interference. J. Appl. Bot. Food Qual. 85, 23-33.

Bertin, P., Grégoire, D., Massart, S., De Froidmont, D. (2004). High level of genetic diversity among spelt germplasm revealed by microsatellite markers. Genome 47, 1043-1052. doi: 10.1139/G04-065

Caballero, L., Martin, L. M., Alvarez, J. B. (2004). Variation and genetic diversity for gliadins in Spanish spelt wheat accessions. Genet. Resour. Crop Ev. 51, 679-686. doi: 10.1023/B:GRES.0000034575.97497.6e

Camarca, A., Anderson, R. P., Mamone, G., Fierro, O., Facchiano, A., Costantini, S. et al. (2009). Intestinal T cell responses to gluten peptides are largely heterogeneous: implications for a peptide-based therapy in celiac disease. J. Immunol. 182, 4158-4166. doi: 10.4049/jimmunol.0803181

Camerlengo, F., Sestili, F., Silvestri, M., Colaprico, G., Margiotta, B., Ruggeri, R. et al. (2017). Production and molecular characterization of bread wheat lines wheat reduced amount of α-type gliadins. BMC Plant Biol. 17, 248. doi: 10.1186/s12870-017-1211-3

Catassi, C., Fabiani, E., Iacono, G., D'Agate, C., Francavilla, R., Biagi, F. et al. (2007). A prospective, double-blind, placebo-controlled trial to establish a safe gluten threshold for patients with celiac disease. Am. J. Clin. Nutr. 85, 160-166. doi: 10.1093/ajcn/85.1.160

Ciccocioppo, A., Di Sabatino, A., Corazza, G. R. (2005). The immune recognition of gluten in celiac disease. Clin. Exp. Immunol. 140, 408-416. doi: 10.1111/j.1365-2249.2005.02783.x

Daniel, C. and Triboi, E. (2000). Effects of temperature and nitrogen nutrition on the grain composition of winter wheats: effects on gliadin content and composition. J. Cereal Sci. 32, 45-56. doi: 10.1006/jcrs.2000.0313

Doyle, J. (1991). "DNA Protocols for Plants", In: Molecular Techniques in Taxonomy, eds. Hewitt, G.M., Johnston, A.W.B. and Young, J.P.W. (Berlin, Heidelberg: Springer), 283-285.

Dubois, B., Bertin, P., Mingeot, D. (2016). Molecular diversity of α-gliadin expressed genes in genetically contrasted spelt (*Triticum aestivum* ssp. *spelta*) accessions and comparison with bread wheat (*T. aestivum* ssp. *aestivum*) and related diploid *Triticum* and *Aegilops* species. Mol. Breeding 36, 152. doi: 10.1007/s11032-016-0569-5

Dubois, B., Bertin, P., Muhovski, Y., Escarnot, E., Mingeot, D. (2017). Development of TaqMan probes targeting the four major celiac disease epitopes found in α-gliadin sequences of spelt (*Triticum aestivum* ssp. *spelta*) and bread wheat (*Triticum aestivum* ssp. *aestivum*). Plant Methods 13, 72. doi: 10.1186/s13007-017-0222-2

Escarnot, E., Agneessens, R., Wathelet, B., Paquot, M. (2010). Quantitative and qualitative study of spelt and wheat fibres in varying milling fractions. Food Chem. 122, 857-863. doi: 10.1016/j.foodchem.2010.02.047

Fasano, A., Berti, I., Geraduzzi, T., Not, T., Colletti, R.B., Drago, S. et al. (2003). Prevalence of celiac disease in at-risk and not-at-risk groups in the United States. Arch. Intern. Med. 163, 286-292. doi: 10.1001/archinte.163.3.286

Garcia-Molina, M. D. and Barro, F. (2017). Characterization of changes in gluten proteins in low-gliadin transgenic wheat lines in response to application of different nitrogen regimes. Front. Plant Sci. 8, 257. doi: 10.3389/fpls.2017.00257

Garrido-Lestache, E., Lopez-Bellido, R. J., Lopez-Bellido, L. (2004). Effect of N rate, timing and splitting and N type on bread-making quality in hard red spring wheat under rainfed Mediterranean conditions. Field Crop. Res. 85, 213-236. doi: 10.1016/S0378-4290(03)00167-9

Gil-Humanes, J., Piston, F., Tollefsen, S., Sollid, L. M., Barro, F. (2010). Effective shutdown in the expression of celiac disease-related wheat gliadin T-cell epitopes by RNA interference. Proc. Natl. Acad. Sci. 107, 17023-17028. doi: 10.1073/pnas.1007773107

Gomez-Becerra, H. F., Erdem, H., Yazici, A., Tutus, Y., Torun, B., Ozturk, L. et al. (2010). Grain concentrations of protein and mineral nutrients in a large collection of spelt wheat grown under different environments. J. Cereal Sci. 52, 342-349. doi: 10.1016/j.jcs.2010.05.003

Harberd, N. P., Bartels, D., Thompson, R. D. (1985). Analysis of the gliadin multigene loci in bread wheat using nullisomic-tetrasomic lines. Mol. Gen. Genet. 198, 234-242. doi: 10.1007/BF00383001

Kasarda, D. D. (2013). Can an increase in celiac disease be attributed to an increase in the gluten content of wheat as a consequence of wheat breeding? J. Agr. Food Chem. 61, 1155-1159. doi: 10.1021/jf305122s

Lange, M., Vincze, E., Wieser, H., Schjoerring, J. K., Holm, P. B. (2007). Suppression of C-hordein synthesis in barley by antisense constructs results in a more balanced amino acid composition. J. Agr. Food Chem. 55, 6074-6081. doi: 10.1021/jf0709505

Li, J., Wang, S.-L., Cao, M., Lv, D.-W., Subburaj, S., Li, X.-H. et al. (2013). Cloning, expression and evolutionary analysis of α-gliadin genes from *Triticum* and *Aegilops* genomes. J. Appl. Genet. 54, 157-167. doi: 10.1007/s13353-013-0139-z

Longin, C. F. H., and Würschum T. (2014). Genetic variability, heritability and correlation among agronomic and disease resistance traits in a diversity panel and elite breeding material of spelt wheat. Plant Breeding 133, 459-464. doi: 10.1111/pbr.12182

Maghirang, E. B., Lookhart, G. L., Bean, S. R., Pierce, R. O., Xie, F., Caley, M. S. et al. (2006). Comparison of quality characteristics and breadmaking functionality of Hard Red winter and Hard Red spring wheat. Cereal Chem. 83, 520-528. doi: 10.1094/CC-83-0520

Matre, P., Jamieson, P. D., Semenov, M. A., Zyskowski, R. F., Porter, J. R., Triboi, E. (2006). Modelling protein content and composition in relation to crop nitrogen dynamics for wheat. Eur. J. Agron. 25, 138-154. doi: 10.1016/j.eja.2006.04.007

Mitea, C., Salentijn, E. M. J., Van Veelen, P., Goryunova, S. V., Van der Meer, I. M., Van den Broeck H. C. et al. (2010). A universal approach to eliminate antigenic properties of alpha-gliadin peptides in celiac disease. PLoS One 5, e15637. doi: 10.1371/journal.pone.0015637

Molberg, O., Uhlen, A. K., Jensen, T., Flaete, N. S., Fleckenstein, B., Arentz-Hansen, H. et al. (2005). Mapping of gluten T-cell epitopes in the bread wheat ancestors: implications for celiac disease. Gastroenterology 128, 393-401. doi: 10.1053/j.gastro.2004.11.003

Noma, S., Kawaura, K., Hayakawa, K., Abe, C., Tsuge, N., Ogihara, Y. (2015). Comprehensive molecular characterization of the α/β-gliadin multigene family in hexaploidy wheat. Mol. Genet. Genomics 291, 65-77. doi: 10.1007/s00438-015-1086-7

Ozuna, C. V., Iehisa, J. C. M., Gimenez, M. J., Alvarez, J. B., Sousa, C., Barro, F. (2015). Diversification of the celiac disease α-gliadin complex in wheat: a 33-mer peptide with six overlapping epitopes, evolved following polyploidization. Plant J. 82, 794-805. doi: 10.1111/tpj.12851

Paolacci, A. R., Tanzarella, A. A., Porceddu, E., Ciaffi, M. (2009). Identification and validation of reference genes for quantitative RT-PCR normalization in wheat. BMC Mol. Biol. 10, 11. doi: 10.1186/1471-2199-10-11

Rashtak, S. and Murray, J. A. (2012). Review article : coeliac disease, new approaches to therapy. Aliment. Pharm. Ther. 35, 768-781. doi: 10.1111/j.1365-2036.2012.05013.x

Ribeiro, M., Rodriguez-Quijano, M., Nunes, F. M., Carrillo, J. M., Branlard, G., Igrejas, G. (2016). New insights into wheat toxicity: breeding did not seem to contribute to a prevalence of potential celiac disease's immunostimulatory epitopes. Food Chem. 213, 8-18. doi: 10.1016/j.foodchem.2016.06.043

Rubio-Tapia, A., Ludvigsson, JF., Brantner, TL., Murray, J. A., Everhart, J. E. (2012). The prevalence of celiac disease in the United States. Am. J. Gastroenterol. 107, 1538-1544. doi: 10.1038/ajg.2012.219

Ruibal-Mendieta, N. L., Delacroix, D. L., Meurens, M. (2002). A comparative analysis of free, bound and total lipid content on spelt and winter wheat wholemeal. J. Cereal Sci. 35, 337-342. doi: 10.1006/jcrs.2001.0434

Salentijn, E. M. J., Esselink, D. G., Goryunova, S. V., Van der Meer, I., Gilissen, L. J. W. J., Smulders, M. J. M. (2013). Quantitative and qualitative differences in celiac disease epitopes among durum wheat varieties identified through deep RNA-amplicon sequencing. BMC Genomics 14, 905. doi: 10.1186/1471-2164-14-905

Sanchez-Leon, S., Gil-Humanes, J., Ozuna, C. V., Gimenez, M. J., Sousa, C., Voytas, D. F. et al. (2018). Low-gluten, non-transgenic wheat engineered with CRISPR/Cas9. Plant Biotechnol. J. 16, 902-910. doi: 10.1111/pbi.12837

Sapone, A., Bai, J. C., Ciacci, C., Dolinsek, J., Green, P. H. R., Hadjvassiliou, M. et al. (2012). Spectrum of gluten-related disorders: consensus on new nomenclature and classification. BMC Med. 10, 13. doi: 10.1186/1741-7015-10-13

Schalk, K., Lang, C., Wieser, H., Koehler, P., Scherf, K. A. (2017). Quantitation of the immunodominant 33-mer peptide from α-gliadin in wheat flours by liquid chromatography tandem mass spectrometry. Sci. Rep. 7, 45092. doi: 10.1038/srep45092

Schalk, K., Koehler, P., Scherf, K. A. (2018). Targeted liquid chromatography tandem mass spectrometry to quantitate wheat gluten using well-defined reference proteins. PLoS One 13, e0192804. doi: 10.1371/journal.pone.0192804

Scherf, K. A., Koehler, P., Wieser, H. (2016). Gluten and wheat sensitivities – An overview. J. Cereal Sci. 67, 2-11. doi: 10.1016/j.jcs.2015.07.008

Scherf, K. A. and Poms, R. E. (2016). Recent developments in analytical methods for tracing gluten. J. Cereal Sci. 67, 112-122. doi: 10.1016/j.jcs.2015.08.006

Shan, L., Molberg, O., Parrot, I., Hausch, F., Filiz, F., Gray, G. M. et al. (2002). Structural basis for gluten intolerance in celiac sprue. Science 297, 2275-2279. doi: 10.1126/science.1074129

Shewry, P. R. and Halford N. G. (2003). “Genetics of wheat gluten proteins”, in Advances in genetics, eds. Hall, J. C., Dunlap, J. C., Friedmann, T. (San Diego, California: Academic Press), 111-184.

Shewry, P. R. and Tatham A. S. (2016). Improving wheat to remove coeliac epitopes but retain functionality. J. Cereal Sci. 67, 12-21. doi: 10.1016/j.jcs.2015.06.005

Smulders, M. J. M., Jouanin, A., Gilissen, L. J. W. J. (2017). 4.8 Gene editing using CRISPR/Cas9 to modify or remove gliadins from wheat and produce coeliac disease epitope-free wheat. WORKING GROUP on PROLAMIN ANALYSIS and TOXICITY, Proceedings of the 31st Meeting, September 28-30, 2017, Minden, Germany, pp 63-68.

Tonutti, E. and Bizzaro, N. (2014). Diagnosis and classification of celiac disease and gluten sensitivity. Autoimmun. Rev. 13, 472-476. doi: 10.1016/j.autrev.2014.01.043

Triboï, E., Matre, P., Triboï-Blondel, A.-M. (2003). Environmentally-induced changes in protein composition in developing grains of wheat are related to changes in total protein content. J. Exp. Bot. 54, 1731-1742. doi: 10.1093/jxb/erg183

Vader, L. W., Stepniak, D. T., Bunnik, E. M., Kooy, Y. M. C., De Haan, W., Drijfhout, J. W. et al. (2003). Characterization of cereal toxicity for celiac disease patients based on protein homology in grains. Gastroenterology 125, 1105-1113. doi: 10.1053/S0016-5085(03)01204-6

Van den Broeck H. C., Cordewener, J. H. G., Nessen, M. A., America, A. H. P., Van der Meer, I. (2015). Label free targeted detection and quantification of celiac disease immunogenic epitopes by mass spectrometry. J. Chromatogr. A 1391, 60-71. doi: 10.1016/j.chroma.2015.02.070

Wieser, H. and Seilmeier, W. (1998). The influence of nitrogen fertilisation on quantities and proportions of different protein types in wheat flour. J. Sci. Food Agric. 76, 49-55. doi: 10.1002/(SICI)1097-0010(199801)76:1<49::AID-JSFA950>3.0.CO;2-2

Xue, C., Erley, G. S., Rossmann, A., Schuster, R., Koehler, P., Mühling, K.-H. (2016). Split nitrogen application improves wheat baking quality by influencing protein composition rather than concentration. Front. Plant Sci. 7, 738. doi: 10.3389/fpls.2016.00738

## Supplementary materials

***Additional file 1:*** Spelt accessions used to study the α-gliadin canonical epitope expression with TaqMan probes.

| Type | Provenance | Country | Accession code[1] (Previous code[2]) | Year of release | Accession number | Accession name |
|---|---|---|---|---|---|---|
| Cultivars | Belgium | Belgium | W-BE48 | 1977 | TRI 9885 | Renval |
| | | | W-BE51 | 1979 | | Albin |
| | | | W-BE53 | 1979 | TRI 16607 | Rouquin |
| | | | W-BE57 | 1982 | TS2044 | Hercule |
| | | | W-BE59 | 1995 | | Franckenkorn |
| | | | W-BE61 | 1995 | | Redouté |
| | | | W-BE62 | 1998 | | Poeme |
| | | | W-BE63 | 1998 | | Ressac |
| | | | W-BE64 | 1999 | | Spy |
| | | | W-BE65 | 2000 | | Cosmos |
| | | | W-BE66 | 2003 | | Stone |
| | | | W-BE67 | 2008 | | Epanis |
| | | | W-BE68 | 1965 | TRI 9883 | Ardenne |
| | | | W-BE83 | 2015 | | 34927 |
| | Germany | Germany | W-DE07 | 1997 | TRI 29474 | Holstenkorn |
| | | | W-DE08 | 2000 | | Schwabenspelz |
| | | | W-DE09 | 2001 | | Ceralio |
| | | | W-DE10 | 2005 | | Badengold |
| | | | W-DE11 | 2006 | | Zollernspelz |
| | | | W-DE12 | 2010 | | Divimar |
| | | | W-DE13 | 2011 | | Badenkrone |
| | | | W-DE14 | 2011 | | Badenstern |
| | | | W-DE15 | 2012 | | Filderstolz |
| | | | W-DE45 | 1980 | TRI 27760 | Hohenheimer Weisser Kolbendinkel |
| | | | W-DE67 | 1988 | TRI 16604 | Schwabenkorn |
| | Switzerland | Switzerland | W-CH08 | 1978 | TS2045 | Ostro |

***Additional file 1.*** *(Continued)*

| Type | Provenance | Country | Accession code[1] (Previous code[2]) | Year of release | Accession number | Accession name |
|---|---|---|---|---|---|---|
| **Cultivars** | Switzerland | Switzerland | W-CH10 | 1990 | TS2083 | Lueg |
| | | | W-CH12 | 1992 | TS2085 | Hubel |
| | | | W-CH15 | 1995 | TS2247 | Ostar |
| | | | W-CH16 | 2004 | | Alkor |
| | | | W-CH22 | 2002 | TS2236 | Sertel |
| | | | W-CH23 | 2004 | | Sirino |
| | | | W-CH24 | 2005 | | Tauro |
| | | | W-CH26 | 2004 | | Titan |
| | | | W-CH27 | 2005 | | Samir |
| | | | W-CH33 | 1952 | TS2024 | Altgold |
| | | | W-CH38 | 1972 | TRI 4685 | Bregenzer Roter Spelz |
| | | | W-CH46 | 1972 | TS2068 | Lenzburg |
| | | | W-CH62 | 2011 | | Zor |
| **Landraces** | Belgium | Belgium | W-BE12 | | | RL111 |
| | | | W-BE13 | | | RL115 |
| | | | W-BE14 | | | RL118 |
| | | | W-BE17 | | | RL135 |
| | | | W-BE19 | | | RL140 |
| | | | W-BE24 | | | RL160 |
| | | | W-BE30 | | | RL226 |
| | | | W-BE32 | | | RL234 |
| | | | W-BE33 | | | RL238 |
| | | | W-BE34 | | | RL239 |
| | | | W-BE35 | | | RL240 |
| | | | W-BE36 | | | RL244 |
| | | | W-BE37 | | | RL253 |
| | | | W-BE75 (BEL08) | | PI348315 | 69Z6,485 |
| | Germany | Germany | S-DE05 | | TRI 27708 | Weisser Kolbenspelz |
| | | | W-DE04 | | RAC TS0146 | OSCH 31F |

***Additional file 1.*** *(Continued)*

| Type | Provenance | Country | Accession code[1] (Previous code[2]) | Year of release | Accession number | Accession name |
|---|---|---|---|---|---|---|
| **Landraces** | Germany | Germany | W-DE17 | | TS2130 | Zeiners Weisser Schlegeldinkel |
| | | | W-DE18 | | TRI 27838 | Waggershauser Hohenheimer weisser kolbendinkel |
| | | | W-DE24 | | TRI 3802 | Blauer Samtiger Winterspelz |
| | | | W-DE26 | | TRI 1786 | Brauner Winter-Grannendinkel aus Nördlingen |
| | | | W-DE28 | | TRI 251 | Müllers Gaiberger |
| | | | W-DE33 | | TRI 1784 | Schwarzer Behaarter Winter-Kolbendinkel aus Nördlingen |
| | | | W-DE40 | | TRI 16603 | Bauländer |
| | | | W-DE41 | | TS2077 | Blauer Winter Kolbendinkel |
| | | | W-DE43 | | TS2026 | Fuggers Babenhauser Zuchtveesen |
| | | | W-DE60 (GER11) | | PI348114 | 69Z6,275 |
| | | | W-DE61 (GER12) | | PI348120 | 69Z6,282 |
| | Switzerland | Switzerland | W-CH03 | | TRI 17250 | Oberkulmer Rotkorn |
| | | | W-CH28 | | RAC TS2145 | Thuerig rotkorn Th4 |
| | | | W-CH29 | | RAC TS2116 | Willisauer weisskorn Wil17 |
| | | | W-CH36 | | TRI 4684 | Brauner Spelz aus Schefflenz |
| | | | W-CH43 | | RAC TS2144 | Hueslen rotkorn |

***Additional file 1.*** *(Continued)*

| Type | Provenance | Country | Accession code[1] (Previous code[2]) | Year of release | Accession number | Accession name |
|---|---|---|---|---|---|---|
| **Landraces** | Switzerland | Switzerland | W-CH49 | | RAC TS2117 | Neuegger weisskorn Ngg42 |
| | | | W-CH53 | | TS2308 | Salez |
| | | | W-CH57 | | PI347950 | 69Z6,105 |
| | | | W-CH58 | | PI347904 | 69Z6,57 |
| | | | W-CH59 | | PI347913 | 69Z6,66 |
| | | | W-CH60 (SWI23) | | PI347939 | 69Z6,93 |
| | | | W-CH61 | | PI348004 | 69Z6,161 |
| | Nothern Europe | Denmark | W-DK01 (DK01) | | PI361811 | DN-2267 |
| | Eastern Europe | Bulgaria | S-BG01 | | PI295059 | Deutchland |
| | | | S-BG02 | | PI295062 | Italien |
| | | | S-BG03 | | PI295061 | Ungarn |
| | | | W-BG01 (BUL04) | | PI295063 | Ungarn |
| | | Czech Republic | S-CZ01 | | RICP 01c0201257 | (ruzyne) |
| | | | S-CZ02 | | RICP 01c0200982 | spalda bila jarni |
| | | | S-CZ03 | | RICP 01c0200983 | |
| | | | W-CZ01 | | RICP 01c0100920 | (svetla) |
| | | | W-CZ02 | | RICP 01c0100921 | (tmava) |
| | | Former Soviet Union | S-SU01 | | TRI 474 | Spelz aus Tzaribord |
| | | Hungary | S-HU01 | | PI290514 | White spelt |
| | | | W-HU01 | | PI272579 | I-1-3530 |
| | | | W-HU02 | | PI272574 | I-1-3540 |
| | | Poland | S-PL01 | | PI192717 | Dankowska graniatka |
| | | Romania | S-RO01 | | PI306554 | 2947 |
| | | | S-RO02 | | PI 306551? | 2944 |
| | | Ukraine | W-UA01 | | K 19372 | |
| | Southern Europe | Italy | W-IT01 | | PI355642 | 69Z5,122 |
| | | Macedonia | S-MK01 | | PI378469 | 1744 |

***Additional file 1.*** *(Continued)*

| Type | Provenance | Country | Accession code[1] (Previous code[2]) | Year of release | Accession number | Accession name |
|---|---|---|---|---|---|---|
| Landraces | Southern Europe | Spain | W-ES01 | | RAC TS1671 | AST 24G |
| | | | S-ES01 | | PI348463 | 69Z6,638 |
| | | | S-ES02 | | PI348526 | 69Z6,704 |
| | | | S-ES03 | | PI348537 | 69Z6,715 |
| | | | S-ES04 (SPA03) | | PI348572 | 69Z6,752 |
| | | | S-ES05 | | PI348580 | 69Z6,760 |
| | | | S-ES06 | | PI348588 | 69Z6,768 |
| | | | S-ES07 | | PI348651 | 69Z6,834 |
| | | | S-ES09 | | PI348688 | 69Z6,872 |
| | | | S-ES10 | | PI348716 | 69Z6,900 |
| | | | S-ES11 | | PI348723 | 69Z6,908 |
| | | | S-ES12 | | PI348766 | 69Z6,953 |
| | Near and Middle East | Afghanistan | S-AF02 | | PI367202 | 625 |
| | | Azerbaijan | S-AZ01 | | K 45364 | |
| | | | S-AZ02 | | K 45368 | |
| | | Iran | S-IR01 (Iran77d) | | CGN 06533 | Iran 77d |
| | | | S-IR02 | | TRI 28863 | Tri 28863 |
| | | | S-IR03 (IRA03) | | CGN12270 | Iran 416A |
| | | Tajikistan | W-TJ03 | | K 56568 | |
| | | | W-TJ05 | | K 52463 | |
| | | | W-TJ07 | | K 52443 | |
| | | | W-TJ08 (TAD06) | | K 52437 | |
| | | | W-TJ09 | | K 52442 | |

[1]: The first letter of the name denotes the accession habit: W = Winter, S = Spring.

[2]: Accessions names used in Dubois et al. (2016) and Dubois et al. (2017). These accessions were selected, after a genetic diversity analysis, as 10 contrasted spelt accessions representative of the spelt diversity.

***Additional file 2:*** Cultivation conditions in which the spelt accessions studied in this work have been grown.

| | **Plot** | | | | | **Climate** | | |
|---|---|---|---|---|---|---|---|---|
| | **Geographic coordinates** | **Altitude (m)** | **Soil type** | **Previous crop** | **N amount applied (kg/ha)[1]** | **Climate type** | **Mean temperature (°C)[2]** | **Rainfalls (mm)[2]** |
| 2014 | 50.558596N, 4.716483E | 166 | Silty | Sugar beet | 90 | Temperate oceanic | 15.4 | 65.9 |
| 2015 | 50.568452N, 4.741172E | 161 | Silty | Pea | 20 | Temperate oceanic | 15.3 | 47.2 |
| 2016 | 50.560465N, 4.711088E | 167 | Silty | Rapeseed | 65 | Temperate oceanic | 16.4 | 149.3 |
| 2017 | 50.558596N, 4.716483E | 166 | Silty | Rapeseed | 15 | Temperate oceanic | 17.8 | 56.5 |

[1] : The N amount applied was adapted each year according to the remaining N amount in the soil in such a way that an identical amount of N has been available for the crop from one year to another.

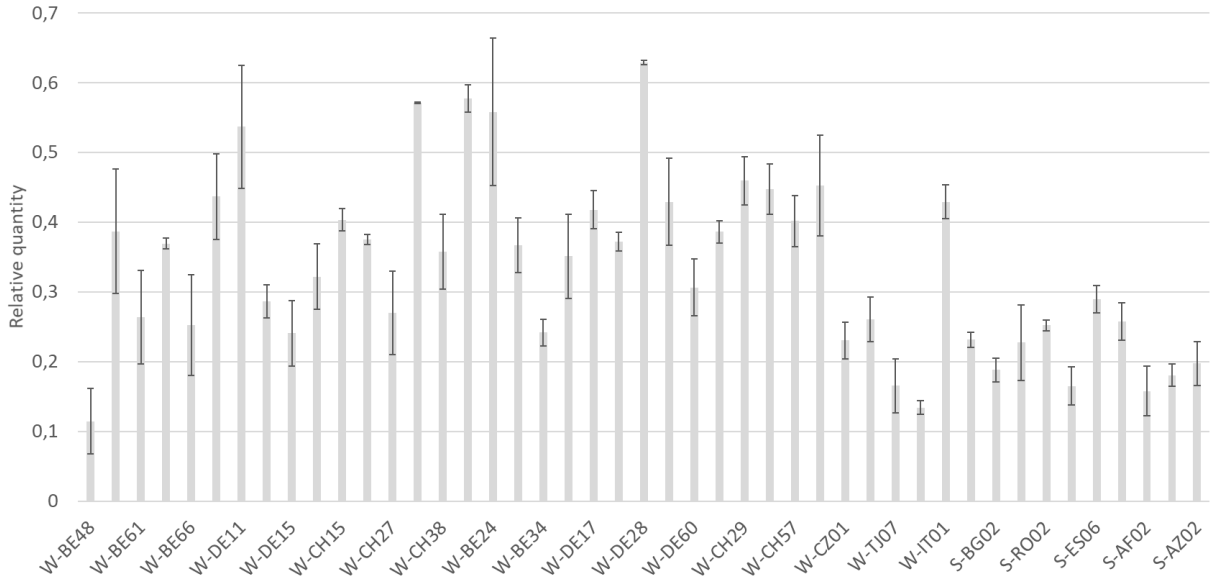
[2] : Data collected in June (grain filling period) by the unit of agrarian systems, territories and information technology from the Walloon agricultural Research Center (Gembloux, Belgium).

***Additional file 3:*** List of the primers and probes used to measure the expression levels of the four α-gliadin major T-cell stimulatory epitopes in their canonical form and the expression levels of reference genes.

| | Target epitope/gene | Oligonucleotide | Sequence |
|---|---|---|---|
| **Epitopes** | DQ2.5-glia-α1 | Forward primer | 5'-GCAACCATTTCCATCACAACWAC-3' |
| | | Reverse primer | 5'-GTGSTTGCGAATACTGTGGTTG-3' |
| | | Probe | 5'-FAM-WTCCRCAGCCGCAACTACCA-TAMRA-3' |
| | DQ2.5-glia-α2 | Forward primer | 5'-GCAACCATTTCCATCACAACWAC-3' |
| | | Reverse primer | 5'-GTGSTTGCGAATACTGTGGTTG-3' |
| | | Probe | 5'-FAM-AGCCGCAACTACCATATCCGC-TAMRA-3' |
| | DQ2.5-glia-α3 | Forward primer | 5'-GCAACCATTTCCATCACAACWAC-3' |
| | | Reverse primer | 5'-GTGSTTGCGAATACTGTGGTTG-3' |
| | | Probe | 5'-FAM-TTCGACCACAACAACCATATCCAC-TAMRA-3' |
| | DQ8-glia-α1 | Forward primer | 5'-CCACAATGTYGTTCATGCTATTATTCTGC-3' |
| | | Reverse primer | 5'-CAGAGCCCTGGGCCTGTGG-3' |
| | | Probe | 5'-FAM-AGGGCTCCTTCCAGCCAT-TAMRA-3' |
| **Reference genes** | ADP-ribosylation factor (ARF) | Forward primer | 5'-GCTCTCCAACAACATTGCCAAC-3' |
| | | Reverse primer | 5'-GCTTCTGCCTGTCACATACGC-3' |
| | | Probe | 5'-FAM-CAAGAAACAAACGTGCTGGATGTC-TAMRA-3' |
| | Similar to RNase L inhibitor-like protein (RLI) | Forward primer | 5'-CGATTCAGAGCAGCGTATTGTTGC-3' |
| | | Reverse primer | 5'-AGTTGGTCGGGTCTCTTCTAAATGTAATG-3' |
| | | Probe | 5'-FAM-CTTAGCGGACAAGGTTATTGTTTATGAGG-TAMRA-3' |
| | Protein of unknown function [DUF52 family] (DUF52) | Forward primer | 5'-TGGTGCCATTCACAAATCAATCG-3' |
| | | Reverse primer | 5'-GCGAACAAACCCGACCTTAATCTTC-3' |
| | | Probe | 5'-FAM-CATGGAGATCATAGAGACTGGTGACC-TAMRA-3' |

Each TaqMan probe is labeled with the fluorogenic dye FAM (fluorescein) at its 5' end and with the quencher TAMRA (tetramethylrhodamine) at its 3' end. The primers targeting reference genes were adapted from those designed by Paolacci et al. (2009).

***Additional file 4:*** Relative quantities of the four cumulated α-gliadin epitopes involved in CD measured on gDNA samples with TaqMan probes targeting only the canonical form of these epitopes.




On the basis of the epitope expression levels measured for 120 spelt accessions (Figure 1), five accessions were selected for each of the nine sub-groups (i.e. Belgian cultivars, German cultivars, Swiss cultivars, Belgian landraces, German landraces, Swiss landraces and landraces from Eastern Europe, Spain and Near and Middle East) to study their gDNA epitope content. These results were then compared to those obtained on cDNA samples to investigate whether a correlation could be assessed between these two sets of data.

***Additional file 5:*** Two-way analysis of variance for the epitope expression analysis of 10 contrasted spelt accessions during 4 consecutive years.

| Source of variation | MS | *d.f.* | *F* | *P* |
|---|---|---|---|---|
| Spelt accessions | 0.618 | 9 | 22.471 | 8.844e-13 |
| Year | 0.051 | 3 | 1.859 | 0.153 |
| Spelt accessions* Year | 0.075 | 26 | 2.718 | 2.322e-3 |

# CONCLUSIONS AND PERSPECTIVES

The results obtained in this research provide important information about the potential of spelt regarding its immunogenic content in the four main α-gliadin T-cell stimulatory epitopes involved in CD. The structure of spelt α-gliadin sequences, their composition in canonical epitope and allelic variants and their expression had not previously been extensively studied in the literature. Moreover, the analysis of patented inventions showed that, in addition to the development of gluten-free products, reducing the patient exposure to gluten immunogenic peptides is one of the major options considered to face the CD issue in the agri-food industry. However, none of them has looked at the spelt potential in this framework. This work thus provides valuable results with the aim of facing CD and investigating the possibility of exploiting the genetic diversity displayed by related (sub-) species of bread wheat to develop non-immunogenic varieties in a long-term perspective.

The following paragraphs will draw conclusions and detail perspectives based on the results obtained in this research, by linking them to each of the objectives mentioned at the beginning of this work.

**Objective 1: Investigate the diversity of α-gliadin expressed genes from spelt compared with bread wheat and diploid species in the Triticeae tribe.**

The analysis by neutral markers of the genetic diversity held in a wide spelt collection revealed a structuration of the accessions in 10 distinct groups, which illustrated the high diversity held in spelt germplasm collections. This structuration enabled to select contrasted accessions, hopefully representative of the spelt diversity. The distribution among these groups was consistent with previous results from the literature showing that spelt accessions from some geographical provenances tend to group together, such as those from Near and Middle East, Eastern Europe and especially Spain. Such geographic structuration can be taken into account when searching interesting traits: if an accession displays such trait, it

can be hypothesized that the probability that other accessions from the same provenance also display this feature should be higher. This postulate has been investigated in the third chapter.

Starting from 11 contrasted spelt accessions, 464 α-gliadin sequences were obtained. One of the consequences of polyploidization, which led to the emergence of spelt, is the increase in the proportion of pseudogenes due to a genetic redundancy (Akhunov et al., 2013) and it has been shown that hexaploid wheats displayed a higher proportion of pseudogenes than diploid ones (Ozuna et al., 2015). In this framework, trying to avoid any overestimation of the immunogenicity by studying expressed α-gliadins rather than their genomic occurrence revealed the relevance of this approach since almost no pseudogene was present in the large set of cloned and sequenced α-gliadin sequences. The molecular characterization of the spelt α-gliadin sequences highlighted interesting features in some sequences like an additional cysteine residue or long polyglutamine domains. As previously discussed, such characteristic may have a favorable impact on the bread-making properties of the dough and could also be targeted in further breeding programs. One of the approach to reduce the CD-related immunogenic content could aim at lowering the expression of D genome sequences since they are the most immunogenic ones. However, they are also the one contributing the most to the bread-making properties and features like additional cysteine residue and long polyglutamine domains could thus be interesting elements to counterbalance the decrease in the expression of D genome sequences. Moreover, newly developed technologies could enable to remove the D-genome sequence immunogenicity while retaining the bread-making qualities and are detailed in the general perspectives section hereafter.

The genomic assignment of spelt α-gliadins led to the same classification of the genome immunogenicity than the one reported for bread wheat sequences in the literature. D-, A- and B-genome α-gliadin sequences had decreasing immunogenic content according to the epitope allelic variants they displayed. In particular, B genome sequences showed almost no celiac immunogenicity and some spelt

accessions had higher proportions of sequences from this genome than those reported for bread wheat in the literature. It could thus be relevant to make use of such spelt accessions with a low immunogenic potential to include them in a breeding strategy.

The phylogenic analysis showed that α-gliadin sequences were grouped according to the genome they are expressed from. The comparison of the amino acid sequences from spelt and bread wheat did not highlight any clear separation between them, but B-genome spelt sequences, however, were divided into two sub-groups according to amino acid motifs that are not found in bread wheat α-gliadins. This observation may support the conclusions of several authors who hypothesized that bread wheat and spelt emerged from different hybridization events.

The analysis of the canonical epitope occurrence in spelt α-gliadins revealed interestingly that the Tajik accession displayed less canonical epitope than the others. Since the genetic diversity analysis showed that Near and Middle East accessions preferentially grouped together, a deeper analysis of the immunogenic potential of other accessions from the same genepool seems promising. This consideration has been investigated in the third chapter where significant differences in the immunogenic content were indeed highlighted according to the geographical provenances.

It must be kept in mind that the results presented in relation to the first objective are based on the mean number of canonical epitopes per sequence, without taking into account the expression level of these sequences. It was essential to consider the expression level as well, in order to have a more complete picture of the immunogenicity of spelt expressed α-gliadins. This consideration was the subject of the second chapter of this research.

**Objective 2: Develop and validate a tool to measure the immunogenic content for CD patients held in expressed α-gliadin sequences.**

This second objective was pursued by choosing the TaqMan probe system as an appropriate tool to quantify the immunogenic content of spelt α-gliadin transcripts thanks to its ability to discriminate SNPs.

A first set of TaqMan probes was developed and validated in order to target only the canonical – *i. e.* immunogenic – form of each of the four main T-cell stimulatory epitopes found in α-gliadin sequences and involved in CD. Such tool may be useful in a breeding program aiming at developing new varieties with a low or no immunogenic content. It could, indeed, be used to track the immunogenic content from one generation to the other and to select the lines displaying the most reduced epitope expression profiles. In addition, the fact that the developed probes are able to discriminate every epitope allelic variant from the canonical form offers interesting perspectives for other researches where the discrimination of several SNPs is a key point. This work indeed demonstrates that TaqMan probes are able, when their design and the qPCR conditions are optimized, to discriminate at the same time different allelic variants whose SNP location differs from one variant to another.

A second set of TaqMan probes has been developed to target reference genes. With the aim of normalizing epitope expression levels, the expression stability of potential reference genes was studied. The ranking of the gene expression stability coming from this analysis provides a useful information for researchers aiming at normalizing the expression values of a gene of interest in similar conditions. Moreover, the genes selected to study their expression stability were chosen because they were among the most stable ones in another study which evaluated the expression stability of 32 potential reference genes in 24 bread wheat plant samples, including different tissues, developmental stages and temperatures stresses. Interestingly, the selected reference genes also displayed a good expression stability in the present work. This underlines the wide application range of these genes for normalization purposes. The primer/TaqMan probes systems

developed can thus be directly used to analyze the expression stability of the gene they target in order to normalize the expression of genes of interest other than α-gliadins.

In the first result chapter, the immunogenic potential was estimated through an average number of canonical epitopes per sequence. In contrast, the immunogenic potential measured in this chapter was expressed as epitope expression levels, which revealed a higher variability in the results. This underlines the higher resolution in the results obtained from the TaqMan probe approach compared to the cloning and sequencing one. This could, however, be partly due to the low number of cloned and sequenced α-gliadins, which might be circumvented using next generation sequencing techniques.

The epitope expression profiling in the diploid species representative of the ancestral genomes of bread wheat and spelt showed that *Aegilops speltoides* α-gliadin sequences displayed only one out of the four epitopes in the canonical form and confirmed the interest of B genome α-gliadins to develop low CD potential accessions.

The epitope expression levels measured with TaqMan probes confirmed the global trend observed when the epitope composition of cloned α-gliadins was analyzed in the previous chapter. In addition, the analysis of the epitope expression levels revealed a great diversity among accessions and very highly significant differences were pointed out. It would thus be relevant to push such epitope expression analysis further by studying it in a wider spelt collection. This investigation has been carried out in the third chapter of this research.

**Objective 3: Study a large panel of spelt accessions in order to determine the variability in their CD-related immunogenic content.**

The application of the developed TaqMan probes to a wide collection of spelt accessions representative of different geographical provenances, breeding status and habits provided interesting information about their immunogenic content. High variations in the epitope expression were highlighted from one accession to

another, which provides an important knowledge to choose spelt accessions as source material for plant breeding programs aiming at developing new varieties with a reduced or suppressed CD-related immunogenic content. Previous genetic diversity analyses showed that Spanish spelt accessions tended to group together and to form a distinct genepool (Bertin et al., 2004; Dubois et al., 2016). Spelt accessions from Spain studied in this work also set apart from the others with particularly high epitope expression levels. In contrast, some accessions showed low or very low amounts of epitope transcripts, such as a German cultivar, a Belgian landrace and, more generally, spelts from Tajikistan. The analysis of the epitope expression according to the breeding status enabled to investigate the impact of breeding led in Belgium, Germany and Switzerland on the immunogenic content. The compared cultivars and landraces did not display significant differences in their epitope expression, leading to the conclusion that breeding strategies did not increase the immunogenicity of spelt from Northwestern Europe, which is in line with other recent observations on species of the Triticeae tribe (Ozuna and Barro, 2018). According to the habit, results showed that spring accessions displayed significantly higher epitope expression levels than winter ones. With aim of developing new varieties with a low immunogenic potential, it is thus preferable to focus on winter spelt accessions.

In the framework of a breeding strategy, it could be convenient to use the developed probes directly on the gDNA of fresh leaves rather than waiting for the development of the grain and harvest it at a precise developmental stage to study its cDNA content. Hence, it has been investigated whether epitope-targeting TaqMan probes could be directly applied to gDNA samples. The results, however, pointed out an absence of correlation between the epitope occurrence in the genome and their expression, which underlines the importance of focusing on the epitope sequences at the transcript level. This means that the marker-assisted selection applied on gDNA would not be efficient in this case. In addition, winter accessions displayed a significantly higher epitope occurrence in their gDNA whereas spring accessions were the one showing the highest epitope expression. This could mean that winter accessions have a higher proportion of α-gliadin

pseudogenes, and/or that the growing conditions undergone by spring accessions are more favorable for the expression of these epitopes.

**Objective 4: Study the epitope expression stability according to environmental factors.**

The environmental influence on the epitope expression has been studied in the last part of the third chapter by harvesting the same accessions during four consecutive years, and by applying different N fertilization strategies. Results showed that the variations in the epitope expression levels were mainly due to the accession factor rather than to the harvest year. This is an interesting observation since a new variety bred to display a low immunogenic content should thus keep this feature from year to year and not be significantly affected by the harvest year. This is illustrated by the Tajik accession which was globally the lowest immunogenic one over the four harvest years.

The impact of the N fertilization on the epitope expression was evaluated by applying three fractionated N doses according to several modalities. Theses modalities differed from each other by the total amount of N applied and by the timing of the last application. Despite slight variations, the assay did not highlight major fluctuations in the epitope expression according to the modalities. Even if this trial was carried out with only two accessions – one Belgian and one German cultivar –, it still seems to indicate that trying to modulate the spelt immunogenic content through the N fertilization would be inefficient. The relative stability of the α-gliadin epitope expression according to the harvest year and to the N fertilization strategy seems to indicate that the heritability for this trait is high, which is an important point in a breeding framework.

## General perspectives

Spelt has previously been shown to display interesting agronomical and nutritional properties. Among other things, it displays high levels of resistance to pathogens, a good adaptability to harsh climatic conditions (Campbell, 1997) and high nutritional values in protein (Bonafaccia et al., 2000; Marconi et al., 1999; Ranhotra et al., 1996), lipid (Ruibal-Mendieta et al., 2002), and micronutrient content (Grela, 1996; Ruibal-Mendieta et al., 2005). A high genetic diversity in spelt germplasm collections has also been reported (An et al., 2005; Bertin et al., 2004; Caballero et al., 2004). In the present research, a wide collection of spelt cultivars and landraces was studied for its CD-related immunogenic content, but it is far from embracing the entire genetic diversity displayed by the spelt sub-species. Hence, it would be interesting to carry on such investigations given the high variability highlighted in the epitope expression profiles. It is, indeed, very likely that other spelt accessions with a low immunogenic potential are yet to be discovered.

The α-gliadin immunogenicity is a quantitative character and its genetic determinism is mostly understood. It is widely known that bread wheat and spelt display a complex genomic structure with three homeologous genomes and complex loci including a high number of gene copies encoding α-gliadins. These loci also display a high proportion of α-gliadin pseudogenes. In addition, it has been shown that α-gliadins expressed from the D genome sequenced so far systematically display immunogenic epitopes.

This research has shown that the epitope expression in spelt is probably not significantly influenced by the environment, which is a primary condition to hypothesize a high heritability of this character. To the best of our knowledge, the segregation of this character has never been studied; such initiative could be relevant in order to sharpen its heritability. Several crosses between very and poorly immunogenic spelt accessions have recently been carried out in this respect. Given the high variability in the epitope expression levels, the known determinism

and the probable high heritability of the character, these features open the door to a selection strategy. Traditional breeding strategies could thus be applied to reduce the amount of immunogenic epitopes but providing celiac-safe varieties without any expressed epitope by traditional techniques only will probably be very challenging. However, it can be hypothesized that the crossing of very poorly immunogenic accessions may be interesting to find transgressive lines in the offspring, thanks to the appropriate combination of non-immunogenic alleles coming from both parents. Another possibility would be the search of wheat or related diploid species lacking immunogenic epitopes, especially in D genome sequences. In this prospect, a wide screening of diploid *Aegilops tauschii* accessions, which provided the D genome to hexaploid wheats, should be carried out in priority, since such celiac-free diploid species has not been identified yet. If these efforts lead to the identification of a diploid accession lacking immunogenic epitopes, it could be used to resynthesize a synthetic hexaploid wheat. However, the existence of such celiac-free *Aegilops tauschii* accessions remains hypothetical.

Importantly, given the complex genetic structure of the α-gliadins, with a high number of genes and pseudogenes, the selection strategy will have to be led at the phenotype level – *i.e*. the amount of expressed immunogenic epitopes – rather than at the genotype level. Such consideration has been confirmed by the absence of correlation between gDNA and cDNA samples in our study. In this prospect, the TaqMan probes developed in this work should constitute a relevant tool.

The traditional breeding for a quantitative character should enable to lower its expression but it must be completely suppressed in the case of celiac disease. Hence, even if the possibility of obtaining celiac-safe varieties through traditional breeding cannot be completely discarded, it is still unlikely nowadays. In consequence, such limitation of traditional breeding strategies has encouraged the development of new molecular breeding approaches which should be considered in the framework of CD.

The down-regulation of gene expression through the RNA interference (RNAi) technology is a promising approach since it offers the possibility of silencing

multigene families as well as more than one group of proteins (Rosell et al., 2014). It has been successfully used to down-regulate the expression of gliadins in bread wheat (up to 97% reduction). This enabled to significantly reduce the immunogenicity of α-, γ-, and ω-gliadins, but efforts to develop a celiac-safe variety through the RNAi interference technology must be carried on since no line showed a complete absence of immunogenicity (Gil-Humanes et al., 2010). The lowering of the gliadin expression had a general weakening effect on the dough. This is probably due to the fact that, even if bread-making properties are mainly determined by glutenins, gliadins are still important since it is the interaction between these two protein classes that confers the unique viscoelastic properties to the dough. Gil-Humanes et al. showed that the silencing of gliadins was compensated by an increase in HMW glutenin subunits, which provided to the dough better tolerance to over-mixing. Moreover, such low-gliadin lines were used to make bread with baking and sensory properties and overall acceptance similar to those of the wild type cultivar (Gil-Humanes et al., 2014).

Another molecular breeding approach consists in editing target genes by inducing modifications of different types: insertions or deletions of several nucleotides, SNPs or large insertions or deletions. Three main technologies have been developed – i.e. zinc-finger nucleases (ZFNs), transcription activator-like nucleases (TALENS) and clustered regularly inter-spaced short palindromic repeats nucleases (CRISPR) – and have already been used to create healthier and more nutritious foods, like improving the oil composition of soybean (Demorest et al., 2016; Haun et al., 2014), developing reduced-acrylamide potatoes (Clasen et al., 2016) or fruits with increased antioxidants (Cermak et al., 2015).

In the framework of CD, the CRISPR technology has recently been applied to bread wheat to generate mutant lines showing strong reduction in the α-gliadin content (Sanchez-Leon et al., 2018). Up to 78% of α-gliadin genes were mutated and γ- and ω-gliadins were also strongly decreased in some lines. Authors showed that the induced mutations reduced the immunoreactivity by 85% and that they were heritable from one generation to the other. One of the drawbacks of this approach

is that it may induce unwanted mutations elsewhere in the genome, but no off-target mutations were detected in this case.

Both CRISPR and RNAi systems are effective to drastically reduce the amount of immunogenic epitopes but one of the major advantages of CRISPR compared to RNAi is that the foreign DNA encoding the CRISPR system can be lost in the following generation thanks to the segregation, while the desired mutation remains in the genome. Such transgene-free lines have been identified in the work of Sanchez-Leon et al. (2018) and are interesting to face regulatory limitations, as detailed further. A second advantage relies in the fact that, in contrast with RNAi, it provides phenotypes that are independent of environmental conditions. Furthermore, the CRISPR technology should enable to cut chromosome fragments containing gliadin genes or even to replace immunogenic fragments with non-immunogenic ones, keeping the gliadin functionality. However, the developed CRISPR and RNAi methodologies have proved to significantly reduce the immunogenicity of mutant lines, but they did not lead to plants completely lacking immunogenic epitopes. Subsequent rounds of genome editing will thus be needed to target the remaining gliadin genes.

In the case of both RNAi and CRISPR, a transgenic step is necessary to introduce either the silencing or the nuclease machinery in the host cell and this stage can be very challenging. This cell transformation as well as its following regeneration are indeed low efficient and time consuming, which represent a significant bottleneck in the implementation of gene silencing or editing. Hence, new strategies are currently being investigated to overcome these limitations, such as the infection of a crop plant with an engineered virus that harbors a genome-editing machinery (Ali et al., 2015; Yin et al., 2015).

Another limitation relies in the fact that only a limited number of genotypes of a given species are amenable to transformation. In addition, this amenability to transformation and regeneration is generally correlated with poor agronomic performance, which implies genome-edited crops to be placed into breeding programs to transfer the mutations into elite varieties (Baltes et al., 2017). This

laborious process makes it expensive (Shewry and Tatham, 2016). Another drawback of these techniques is the regulations about GMO, which greatly fluctuates about what is considered as a genetically modified crop and what is acceptable or not. In the United States for instance, crop varieties generated through genome editing are not considered as genetically modified organisms whereas the situation is more complicated in the European Union (Jones, 2015). One of the alternative currently investigated is the non-transgenic genome engineering where sequence-specific nucleases are delivered as transient DNA – i.e. which is not incorporated in the host genome – (Zhang et al., 2016), mRNA (Stoddard et al., 2016) or protein (Liang et al., 2017; Wolter and Puchta, 2017).

In conclusion, the development of celiac-safe wheat varieties is a realistic target, but it will probably require a combination of traditional and molecular breeding strategies. In this framework, the TaqMan probe systems developed in this research should be useful to track the immunogenic content throughout such complex breeding process. A high genetic diversity in the CD-related immunogenic content has been highlighted in a wide spelt collection and a larger genepool available in gene banks is yet to be characterized. In addition, the availability of traditional breeding techniques combined to the emergence of molecular approaches provide promising prospects to develop celiac-safe varieties.

## References


Akhunov, E.D., Sehgal, S., Liang, H., Wang, S., Akhunova, A.R., Kaur, G. et al. (2013). Comparative analysis of syntenic genes in grass genomes reveals accelerated rates of gene structure and coding sequence evolution in polyploid wheat. Plant Physiology, 161, 252-265. doi: 10.1104/pp.112.205161.

Ali, Z., Abul-faraj, A., Li, L., Ghosh, N., Piatek, M., Mahjoub, A., et al. (2015). Efficient Virus-Mediated Genome Editing in Plants Using the CRISPR/Cas9 System. Molecular Plant 8, 1288–1291. doi:10.1016/j.molp.2015.02.011.

An, X., Li, Q., Yan, Y., Xiao, Y., Hsam, S. L. K., and Zeller, F. J. (2006). Genetic diversity of European spelt wheat (Triticum aestivum ssp. spelta L. em. Thell.) revealed by glutenin subunit variations at the Glu-1 and Glu-3 loci. Euphytica 146, 193–201. doi:10.1007/s10681-005-9002-6.

Baltes, N. J., Gil-Humanes, J., and Voytas, D. F. (2017). “Genome Engineering and Agriculture: Opportunities and Challenges,” in Progress in Molecular Biology and Translational Science (Elsevier), 1–26. doi:10.1016/bs.pmbts.2017.03.011.

Bertin, P., Grégoire, D., Massart, S., and de Froidmont, D. (2004). High level of genetic diversity among spelt germplasm revealed by microsatellite markers. Genome 47, 1043–1052. doi:10.1139/g04-065.

Bonafaccia, G., Galli, V., Francisci, R., Mair, V., Skrabanja, V., and Kreft, I. (2000). Characteristics of spelt wheat products and nutritional value of spelt wheat-based bread. Food Chemistry 68, 437–441. doi:10.1016/S0308-8146(99)00215-0.

Caballero, L., Martín, L. M., and Alvarez, J. B. (2004). Variation and genetic diversity for gliadins in Spanish spelt wheat accessions. Genetic Resources and Crop Evolution 51, 679–686. doi:10.1023/B:GRES.0000034575.97497.6e.

Campbell, K.G. (1997). Spelt: Agronomy, genetics and breeding. Plant Breeding Reviews 15, 187-213.

Čermák, T., Baltes, N. J., Čegan, R., Zhang, Y., and Voytas, D. F. (2015). High-frequency, precise modification of the tomato genome. Genome Biology 16. doi:10.1186/s13059-015-0796-9.

Clasen, B. M., Stoddard, T. J., Luo, S., Demorest, Z. L., Li, J., Cedrone, F., et al. (2016). Improving cold storage and processing traits in potato through targeted gene knockout. Plant Biotechnology Journal 14, 169–176. doi:10.1111/pbi.12370.

Demorest, Z. L., Coffman, A., Baltes, N. J., Stoddard, T. J., Clasen, B. M., Luo, S., et al. (2016). Direct stacking of sequence-specific nuclease-induced mutations to produce high oleic and low linolenic soybean oil. BMC Plant Biology 16. doi:10.1186/s12870-016-0906-1.

Gil-Humanes, J., Pistón, F., Altamirano-Fortoul, R., Real, A., Comino, I., Sousa, C., et al. (2014). Reduced-Gliadin Wheat Bread: An Alternative to the Gluten-Free Diet for Consumers Suffering Gluten-Related Pathologies. PLoS ONE 9, e90898. doi:10.1371/journal.pone.0090898.

Gil-Humanes, J., Piston, F., Tollefsen, S., Sollid, L. M., and Barro, F. (2010). Effective shutdown in the expression of celiac disease-related wheat gliadin T-cell epitopes by RNA interference. Proceedings of the National Academy of Sciences 107, 17023–17028. doi:10.1073/pnas.1007773107.

Grela, E. R. (1996). Nutrient Composition and Content of Antinutritional Factors in Spelt (*Triticum spelta* L) Cultivars. Journal of the Science of Food and Agriculture 71, 399–404. doi:10.1002/(SICI)1097-0010(199607)71:3<399::AID-JSFA609>3.0.CO;2-Q.

Haun, W., Coffman, A., Clasen, B. M., Demorest, Z. L., Lowy, A., Ray, E., et al. (2014). Improved soybean oil quality by targeted mutagenesis of the fatty acid desaturase 2 gene family. Plant Biotechnology Journal 12, 934–940. doi:10.1111/pbi.12201.

Jones, H. D. (2015). Regulatory uncertainty over genome editing. Nature Plants 1, 14011. doi:10.1038/nplants.2014.11.

Liang, Z., Chen, K., Li, T., Zhang, Y., Wang, Y., Zhao, Q., et al. (2017). Efficient DNA-free genome editing of bread wheat using CRISPR/Cas9 ribonucleoprotein complexes. Nature Communications 8, 14261. doi:10.1038/ncomms14261.

Marconi, E., Carcea, M., Graziano, M., and Cubadda, R. (1999). Kernel Properties and Pasta-Making Quality of Five European Spelt Wheat (*Triticum spelta* L.) Cultivars. Cereal Chemistry Journal 76, 25–29. doi:10.1094/CCHEM.1999.76.1.25.

Ozuna, C. V., Iehisa, J. C. M., Gimenez, M. J., Alvarez, J. B., Sousa, C., Barro, F. (2015). Diversification of the celiac disease α-gliadin complex in wheat: a 33-mer peptide with six overlapping epitopes, evolved following polyploidization. The Plant Journal 82, 794-805. doi: 10.1111/tpj.12851

Ranhotra, G. S., Gelroth, J. A., Glaser, B. K., and Stallknecht, G. F. Nutritional Profile of Three Spelt Wheat Cultivars Grown at Five Different Locations. Cereal Chemistry 73, 533-535.

Rosell, C. M., Barro, F., Sousa, C., and Mena, M. C. (2014). Cereals for developing gluten-free products and analytical tools for gluten detection. Journal of Cereal Science 59, 354–364. doi:10.1016/j.jcs.2013.10.001.

Ruibal-Mendieta, N. L., Delacroix, D. L., and Meurens, M. (2002). A Comparative Analysis of Free, Bound and Total Lipid Content on Spelt and Winter Wheat Wholemeal. Journal of Cereal Science 35, 337–342. doi:10.1006/jcrs.2001.0434.

Ruibal-Mendieta, N. L., Delacroix, D. L., Mignolet, E., Pycke, J.-M., Marques, C., Rozenberg, R., et al. (2005). Spelt (*Triticum aestivum* ssp. *spelta*) as a Source of Breadmaking Flours and Bran Naturally Enriched in Oleic Acid and Minerals but Not Phytic Acid. Journal of Agricultural and Food Chemistry 53, 2751–2759. doi:10.1021/jf048506e.

Sánchez-León, S., Gil-Humanes, J., Ozuna, C. V., Giménez, M. J., Sousa, C., Voytas, D. F., et al. (2018). Low-gluten, nontransgenic wheat engineered with CRISPR/Cas9. Plant Biotechnology Journal 16, 902–910. doi:10.1111/pbi.12837.

Shewry, P. R., and Tatham, A. S. (2016). Improving wheat to remove coeliac epitopes but retain functionality. Journal of Cereal Science 67, 12–21. doi:10.1016/j.jcs.2015.06.005.

Stoddard, T. J., Clasen, B. M., Baltes, N. J., Demorest, Z. L., Voytas, D. F., Zhang, F., et al. (2016). Targeted Mutagenesis in Plant Cells through Transformation of Sequence-Specific Nuclease mRNA. PLOS ONE 11, e0154634. doi:10.1371/journal.pone.0154634.

Wolter, F., and Puchta, H. (2017). Knocking out consumer concerns and regulator's rules: efficient use of CRISPR/Cas ribonucleoprotein complexes for genome editing in cereals. Genome Biology 18. doi:10.1186/s13059-017-1179-1.

Yin, K., Han, T., Liu, G., Chen, T., Wang, Y., Yu, A. Y. L., et al. (2015). A geminivirus-based guide RNA delivery system for CRISPR/Cas9 mediated plant genome editing. Scientific Reports 5, 14926. doi:10.1038/srep14926.

Zhang, Y., Liang, Z., Zong, Y., Wang, Y., Liu, J., Chen, K., et al. (2016). Efficient and transgene-free genome editing in wheat through transient expression of CRISPR/Cas9 DNA or RNA. Nature Communications 7, 12617. doi:10.1038/ncomms12617.