Gene editing towards hypoimmunogenic gluten proteins in wheat

Aurélie A. Jouanin

Thesis committee

Promotor
Prof. Dr R.G.F. Visser
Professor of Plant Breeding
Wageningen University & Research

Co-promotors
Dr M.J.M. Smulders
Senior Scientist Plant Breeding
Wageningen University & Research

Dr F.J. Leigh
Senior Research Scientist
National Institute of Agricultural Botany, Cambridge, UK

Other members
Prof. Dr H.D. Jones, Aberystwyth University, Aberystwyth, UK
Prof. Dr M.H. Zwietering, Wageningen University & Research
Dr P.L. Weegels, Sonneveld Group B.V., Papendrecht, The Netherlands
Dr R.A. de Maagd, Wageningen University & Research

This research was conducted under the auspices of the Graduate School Experimental Plant Sciences.

Gene editing towards hypoimmunogenic gluten proteins in wheat

Aurélie A. Jouanin

Thesis

submitted in fulfilment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus,
Prof. Dr A.P.J. Mol,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Monday 28 January 2019
at 4 p.m. in the Aula.

Aurélie A. Jouanin
Gene editing towards hypoimmunogenic gluten proteins in wheat

PhD thesis, Wageningen University, Wageningen, The Netherlands (2019)
With references, with summary in English

DOI: 10.18174/464538
ISBN: 978-94-6343-542-0

Summary

Summary

Bread wheat (*Triticum aestivum*) is an allohexaploid plant, with a genome size five times bigger than the human genome. It is worldwide the third staple crop, after maize and rice, grown in many parts of the world. Wheat grains contain gluten, a polymer of glutenin and gliadin proteins, which build a network that gives wheat dough its unique properties and quality. They are encoded by a total of five large gene families. Gluten proteins harbour immunogenic epitopes which can trigger Coeliac Disease (CD), an auto-immune reaction, to which 1-2% of the world population are sensitive, as described in the *General Introduction*. A gluten-free diet, excluding wheat, barley and rye, is the only remedy. This diet is difficult to adhere to, partly because wheat gluten is added to many processed products for their viscoelastic properties. In addition, gluten-free products typically require the inclusion of numerous additives, resulting in products that are often less healthy than gluten-based equivalents, and more expensive. There is therefore a need to develop alternative methods in order to provide Coeliac patients with healthier food products, yet safe for them to consume.

Various strategies can be deployed towards the development of food products containing gluten that would be hypoimmunogenic for Coeliac patients. *Chapter 1* gives an overview of the food processing and plant breeding strategies to create coeliac-safe and healthy wheat products. We found that several candidate food processing techniques exist that aim to remove or modify gliadins or all gluten, to produce low-gluten food products from wheat. In parallel, wheat breeding strategies are starting to be developed, to remove immunogenic epitopes from the gluten proteins in new varieties, or to remove complete gluten gene families altogether, whilst maintaining the food-processing properties. This chapter highlights the fact that all the existing solutions have limitations. Hence, solutions decreasing gluten immunogenicity from food processing and plant breeding should be combined in order to have the safest wheat product possible for CD patients.

Among the existing plant breeding strategies to reduce wheat gluten immunogenicity, gene editing, a targeted mutagenesis approach, is a very encouraging new approach. It makes it possible to precisely remove gliadin genes or modifying the CD epitopes. It constitutes the core of the present PhD

thesis. In *Chapter 2*, I describe the generation of lines of hexaploid bread wheat cultivar Fielder stably transformed with CRISPR/Cas9 constructs containing up to six multiplex sgRNA targeting α-gliadin genes, γ-gliadin genes or both simultaneously. These sgRNAs guide the Cas9 protein to several positions within these genes to induce double-strand breaks in the DNA. Plants regenerated may have mutations at these positions due to incorrect repair, which may knockout gene expression. Indeed, the resulting grains displayed altered gliadin protein profiles on Acid-PAGE, including absence of certain gliadin proteins, and in some of these lines the changes were similar to what was observed in Paragon mutant lines – obtained using γ-irradiation, which is a random mutagenesis approach.

The gliadin protein mutations observed in the wheat grains of γ-irradiated and gene-edited mutant wheat lines were then quantitatively studied at the genomics level. *Chapter 3* describes the development of a duplex droplet digital PCR (ddPCR) assay to simultaneously assess the copy number of a reference gene and that of the α-gliadin genes. The results enable calculating the copy number of the gliadins, and quantification of gene editing-induced deletions of α-gliadin genes. The duplex ddPCR was validated using Paragon γ-irradiated lines which showed a reduction of over 1/3 of the α-gliadin gene copies. The comparison of two different duplex ddPCR assays made it feasible to distinguish small indels from larger deletions with a single Fielder-CRISPR line containing small indels (1-50bp) in up to 10 α-gliadin genes and large deletions (>300bp) in 20 α-gliadin genes out of the 87 α-gliadin genes present in Fielder.

The gliadin gene mutations induced by γ-irradiation and gene editing were also studied qualitatively. *Chapter 4* presents GlutEnSeq as a qualitative method to reduce the complexity of the wheat genome and enables analysing genomic variation of gluten genes in hexaploid wheat by sequencing. This chapter describes the development of an in-solution gluten gene exome capture for enrichment and sequencing, and the analysis of the resulting sequencing data, against the bread wheat reference genome. Mutations and deletions of complete genes are visible as differences in the gluten coverage profiles between mutant plants and the original cultivar. Paragon γ-irradiated mutant lines were homozygous for deletions of complete gluten-containing loci on chromosome 1 or 6. Fielder-CRISPR mutant lines also showed loss, either in homozygous or heterozygous form, of all α-gliadin genes at locus *Gli-2* on chromosome 6A or γ-gliadin genes at the *Gli-1* locus on

1B. GlutEnSeq thus enables zooming in on where deletions have taken place. For characterizing the mutations in the DNA within gliadin genes, further data analysis will be necessary.

Aside from being a cutting-edge technology, the regulation of gene editing as GM (Genetic Modification) in Europe is currently a hot topic. *Chapter 5* shows the inconsistency of the European regulation, by displaying the similarities of mutations in gliadin genes that are obtained using random γ-irradiation mutagenesis and those obtained by targeted mutagenesis using gene editing, the former being exempted from GM regulation while the latter being subjected to GM regulation. This chapter covers the topics of risks and safety, public acceptance, distinguishability and labelling, and potential politico-economic issues, related to the regulation of gene-edited plants as GM product in Europe, in contrast to other countries, following the ruling of the European Court of Justice in July 2018. It also advises the European Commission to review its position on the matter and to regulate gene editing based on scientific evidence regarding the generated products, and on the innovation principle as part of responsible research innovation initiatives.

Finally, a *General Discussion* elaborates on the necessity of developing hypoimmunogenic wheat using plant breeding. The relative level of success that is obtained with CRISPR/Cas9 gene editing for developing wheat lines with lower gliadin immunogenic epitopes is discussed. Some recently developed CRISPR approaches are presented that may result in faster development of wheat with hypoimmunogenic gluten, or simplify screening during the process. The relevance of high-throughput screening methods and their adaptation is discussed. The benefits and potential risks related to gene-edited wheat with hypoimmunogenic gluten are described. The requirement for producing and processing these varieties are touched upon. Lastly, the development of methods to test for “safe-gluten” food products made of hypoimmunogenic gluten is addressed, since current “gluten-free” tests will not be relevant in this case. To conclude, we indicate that although the first step towards the development of wheat varieties with hypoimmunogenic gluten was promising, few more years will be necessary before commercially releasing wheat varieties with gluten safe for all CD patients.

S

Key words: *Wheat, Allohexaploid, Gene family, Gluten, Gliadin, Epitope, Coeliac disease, Food processing, Plant breeding, Mutation breeding, Deletion, Mutation, Gene editing, New Plant Breeding Technique, CRISPR/Cas9, Acid-PAGE, ddPCR, CNV, GlutEnSeq, GM Regulation, Policy.*

Table of content

	Summary	7
	General Introduction	15
Chapter 1.	Food processing and breeding strategies for coeliac-safe and healthy wheat products	37
Chapter 2.	Outlook for coeliac disease patients: Towards hexaploid wheat with hypo-immunogenic gluten by gene editing of α- and γ-gliadin gene families	63
Chapter 3.	Optimization of droplet digital PCR for determining genetic variation in the large α-gliadin gene family in hexaploid wheat	95
Chapter 4.	Development of GlutEnSeq system to enrich for and sequence gluten gene families in hexaploid bread wheat lines with induced mutations in gluten genes	127
Chapter 5.	Development of wheat with hypoimmunogenic gluten obstructed by the gene editing policy in Europe	151
	General Discussion	165
	References	187
	Acknowledgments	202
	About the author	208
	Education statement	212

General Introduction

Wheat and gluten

From diploid grass to polyploid commercial wheat

Triticum aestivum, commonly known as bread wheat, is an allohexaploid autogamous monocot plant domesticated more than 10,000 years ago in the Fertile Crescent (Eckardt, 2010). Initially, *Triticum* derived from a diploid grass ancestor (2n=14) of which the genome evolved separately into various diploid genomes (A, S/B and D genome Gramineae). Around 0.5 million years ago (Huang *et al.*, 2002), a natural allopolyploidisation occurred between *Triticum urartu* (AA) and *Aegilops speltoides* (BB or SS), resulting in tetraploid emmer wheat, *T. turgidum* spp. *dicoccoides* (AABB, 4n=28). During cultivation of emmer, *T. turgidum* (AABB) underwent a spontaneous allopolyploidisation with *Aegilops tauschii* (DD) resulting in the hexaploid wheat species, *Triticum aestivum* (AABBDD, 6n =42), 10,000 years ago (Gupta *et al.*, 2008) (Fig. 1). Due to their polyploidy, *T. turgidum* and *T. aestivum* are more vigorous and provide bigger grains. Since its domestication, lots of selection has been done on wheat. The focus has been on improving wheat grain yield and on quality for food products.

Wheat gluten specificity

Cereal kernels and derived flours or extracts form the basis of a large range of staple foods (bread, pasta, semolina, etc.) and other products (alcoholic beverages, food additives) worldwide, using the protein and/or starch from the grains. Wheat endosperm contains water-soluble proteins (globulins) and water-insoluble proteins (prolamins or glutelins). Gluten is a polymer of two major protein groups, the glutelins and the prolamins that are grain storage proteins.

For food processing, polymeric wheat glutelins, composed of High and Low Molecular Weight Glutenin Subunits (HMW-GS and LMH-GS), form a strong Glutenin Macro Polymer (GMP) network that provides elastic properties, which determine dough quality. These glutenins are combined with monomeric wheat prolamins called α-, γ- and ω-gliadins that bring viscosity, which has a lesser impact on dough quality. The type of HMW-GS glutenins and the ratio of glutenins to gliadins determines dough viscoelastic superiority required for bread making or other products (Weegels *et al.*, 1996).

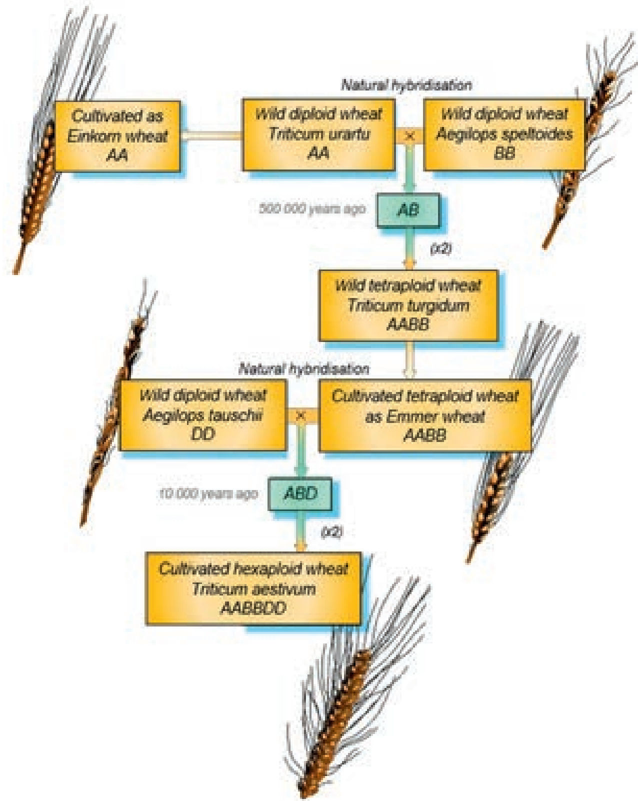


Figure 1. ***The origins of hexaploid bread wheat.***
Bread wheat is allohexaploid and originates from two events of natural hybridisations. The first natural hybridisation occurred between two wild diploid *Gramineae* species, *Triticum urartu* (AA) and *Aegilops speltoides* (BB), up to 0.5 million years ago. The resulting tetraploid *Triticum turgidum* (AABB) was cultivated by farmers when a second natural hybridisation occurred with the wild diploid species *Aegilops tauschii* (DD) about 8-10,000 years ago, resulting in the hexaploid *Triticum aestivum* (AABBDD), cultivated as bread wheat. This Figure has been adapted from http://www.cerealsdb.uk.net/cerealgenomics/WheatBP/Documents/DOC_Evolution.php.

Hexaploid *T. aestivum* (bread wheat) generates better bread dough quality than tetraploid *T. Turgidum spp. durum* (durum or pasta wheat), due to specific types of HMW-GS glutenins that improve bread baking properties and which are located on the D genome in *T. aestivum* (Payne *et al.*, 1984). However, this D genome also codes for gliadin proteins with peptides (epitopes) that are highly immunogenic (van Herpen *et al.*, 2006; Salentijn *et al.*, 2012) for some human individuals.

Gluten from other cereals

Barley (*Hordeum vulgare*), rye (*Secale cereale*) and oat (*Avena sativa*) grains contain gluten as well. However, their gluten do not contain polymeric glutelins and therefore lack the strong GMP network essential for high dough quality. This factor strongly limits their bread-baking potential. Gluten of these cereals mainly consists of monomeric prolamins called hordeins, secalins and avenins, respectively. Hordeins and secalins also contain immunogenic epitopes while avenins have been shown to contain non-immunogenic epitopes, making them safer for consumption (Londono *et al*., 2013; Smulders *et al*., 2018), consistent with the results on long-term cohort studies of CD patients regularly eating oats, amongst others in Nordic countries (Aaltonen *et al*., 2017).

Gliadin immunogenicity and Coeliac Disease

Despite its prevalence in most diets, the ingestion of wheat gluten turns out to be unsafe for one to two percent of the world population (Mäki *et al*., 2003; Rewers, 2005; Catassi *et al*., 2014; Vriezinga *et al.,* 2015), due to the presence of immunogenic gluten epitopes that can trigger an enterogastric autoimmune reaction called Coeliac disease (CD) (Fig. 2). The autoimmune response and associated CD symptoms occur only in the presence of four elements:

1- Ingestion of gluten containing certain immunogenic epitopes (determined by wheat plant genetics)

2- Migration of large gluten fragments in the lamina propria (determined by human intestine condition)

3- Presence of specific HLA-DQ receptors (DQ2, DQ8) subunits on antigen-presenting cells (APC) (determined by human genetics)

4- Presence of specific T-cells receptors subunits (determined by human immune history or sensitisation)

Gluten ingestion and immunogenic epitopes

CD is triggered by the ingestion of gluten from wheat, barley or rye. Gluten proteins, especially gliadins, are glutamine- (Q) and proline- (P) rich, which make them recalcitrant to hydrolysis by gastro-intestinal luminal and brush-border proteases and peptidases (Qiao *et al.*, 2009). This impedes their complete digestion into small three amino-acids peptides (Hausch *et al.*, 2002), necessary for active transport into the bloodstream as nutrients. The 9-mer to 33-mer gliadin epitopes are too large and should be directed to the rectum for excretions.

Gliadin epitopes migration into lamina propria and innate immune response

In instances of lower permeability of the small intestine (e.g. due to infections or due to the presence of zonulin potentially stimulated by gluten itself), large peptides may pass the epithelium, enter the intestinal lamina propria (the sub-epithelial area containing blood vessels) and trigger an innate immune response (Fig. 2). The enterocytes release interleukins IL-15 which stimulate intra-epithelial lymphocytes to migrate in-between enterocytes. This triggers the release of tissue TransGlutaminase (tTG), an enzyme synthesised by the damaged enterocytes. These tTGs will alter the gluten peptides by deaminating glutamines (Q) into negatively charged glutamic acid (E), thus creating more efficiently binding epitopes.

Genetic predisposition

Humans have, in their immune system, Antigen Presenting Cells (APC). These cells harbour on their surface receptors called Human Leucocyte Antigen (HLA), whose role is to bind antigens and present them to T cells. They bind 9-amino acid peptides. HLA-DQ proteins are a combination of various β- and α- subunits forming dimers, creating receptors. The nature of the β (2 or 8) and α (2 or 5) subunits are human-allele dependent. Humans carrying the gene combination DQ 2.5 (β chain gene allele 2, α chain gene allele 5), will have APC receptors HLA-DQ2.5, which can bind gliadin epitopes and present them as antigens to the T cells. A person being heterozygous for both β (2 or 8) and α (2 or 5) allele can develop APC with HLA-DQ2.5, HLA-DQ2.2, HLA-DQ8.5 and HLA-DQ8.2, recognising most epitopes. These persons are genetically predisposed to CD, but do not systematically develop it. In fact, only around 5% of all HLA-DQ2.5 homozygote people develop CD, while heterozygote people have an even lower chance (Koning, personal communication).

HLA-DQ binding gliadin epitopes

Depending on their sequences, the 9-mer gliadin epitopes may interact with the groove of these HLA receptors. Negative charges are necessary, which are not common in gliadin proteins. However, tTG deamination of glutamines (Q) into negatively charged glutamic acid (E) at positions P4 and P6 of specific 9-mer gliadin epitope motifs ensures strong binding to the groove of the HLA-DQ2.5 receptor subunits (Molberg *et al*., 1998; van de Wal *et al*., 1998). In addition, the identity of the amino acids at P1, P7 and P9 influences the binding potential of the epitope (Vartdal *et al*., 1996). A strong gliadin epitope-HLA-DQ complex increases the chance of the APC to successfully present the gliadin epitope as an antigen to the T Cell (Fig. 2). Consequently, a single amino acid change can make the difference between strong binding and no binding (Koning, 2005).

Specific T cell synthesis against gliadin epitopes: sensitisation

Helper T cells harbour antigen receptors on their surface. If these receptors bind the peptide being presented to them as antigens by the HLA receptor on the APC it is a confirmation for the immune system that the peptide held is an antigen. These Helper T cells will then trigger an immune response to get rid of these antigens (Fig. 2). This binding event depends on the combination of the subunits constituting the Helper T cell receptor and their affinity with the gliadin epitope (Petersen *et al*., 2014).

Adaptive and autoimmune response mechanisms

For individuals carrying Helper T cells with receptors that bind the gliadin epitope presented as an antigen by the APC's HLA-DQ receptors, a chain of adaptive immune responses starts in the small intestine (Petersen *et al*., 2014; Fasano, 2009), initiating CD. Helper T cells will secrete cytokines that activate B cells. B cells will produce specific antibodies against gliadin epitopes but also, following "the guilty by association" rule, against tTG deaminating gliadin epitopes. In addition, Helper T cells will stimulate the development of Killer T cells directed against enterocytes - which are synthesising the tTG- and will release pro-inflammatory cytokines and chemokines causing atrophy of microvilli (reviewed in Fasano, 2009) (Fig. 2). The targeting of the individual's tTG, enterocytes and microvilli by its own immune system constitutes the starting point of the autoimmune reaction.

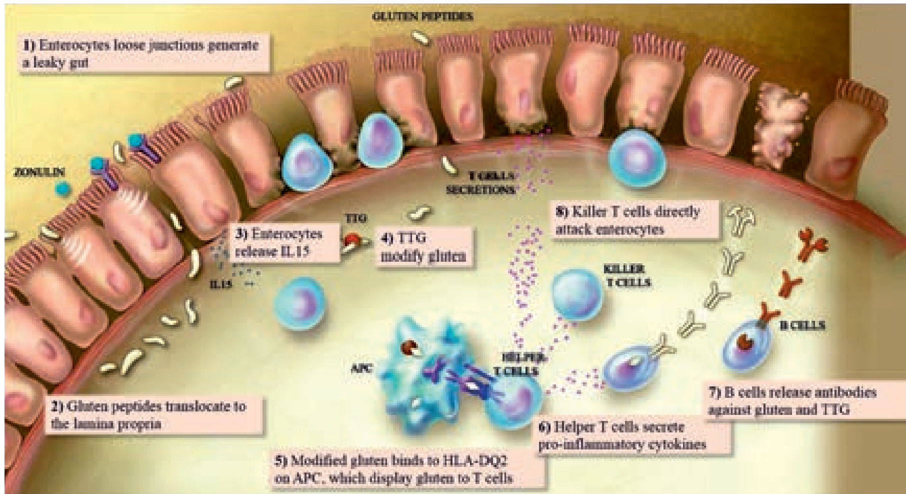


Figure 2. Coeliac Disease mechanism: from gluten protein ingestion to autoimmune reaction.
1) Individual with a leaky gut, where epithelial cells' tight junctions are disrupted. 2) Undigested gluten peptides enter the lamina propria and trigger a first immune response. 3) The enterocytes release IL-15 that stimulates intra-epithelial cells to localise in-between enterocytes, damaging them. 4) Damaged enterocytes release tissue transglutaminase (tTG) that deaminates the gluten peptides. These deaminated gluten peptides may bind to specific HLA-DQ2 or -DQ8 receptors potentially present on the surface of antigen presenting cells (APC). 5) If it is the case, APC display the gluten peptide recognised as an antigen to the Helper T cells. 6) If the Helper T cells carry receptors that also bind the gluten epitope, they secrete pro-inflammatory cytokines. 7) The cytokines also trigger the differentiation of B cells releasing antibodies against the gluten peptides but also against the tTG that deaminated the gluten peptides. 8) In addition, following the "guilty by association principle", these cytokines trigger the differentiation of Killer T cells, which directly attack the enterocytes that generated the tTG. This constitutes the starting point of the CD autoimmune reaction. This Figure has been adapted from Fasano *et al.* (2009).

Coeliac Disease symptoms

The first symptoms of CD are abdominal pain related to the inflammation caused by autoimmune reaction, the enterocyte degradations and the microvilli atrophy. This is gradually degrading the epithelium, causing a loss of its "protective barrier" function. The atrophy of microvilli brush (a border membrane of epithelial cells) decreases the nutrients absorption surface between the small intestine and the blood stream, leading to malnutrition and all sorts of indirectly related symptoms such as fatigue, diarrhoea, weight loss.

The anti-tTG antibodies will travel in the body via the bloodstream and degrade tTG everywhere they are present. This is going do have a general negative impact on the body functioning since tTG enzymes are involved in many processes such as wound healing and more (Upchurch *et al.*, 1991). Other major CD symptoms are headache, dermatitis herpetiformis and eventually death in extreme cases (Zipser *et al.*, 2003; Reunala, 1998).

Gluten-free diet to prevent Coeliac Disease

Up until now, no medical treatment is known. The only remedy for CD is a free-gluten (GF) diet, which allows the small intestine to recover, abolishing the symptoms. However, in daily life, a gluten-free diet is difficult to follow considering the presence of wheat, barley or rye derivatives in a large range of processed food products, often as a “hidden ingredient”, not labelled. In addition, the use of flours from other crops as substitute in GF products means that they lack the nutritional value of wheat and the viscoelastic properties of wheat gluten, which is necessary for good bread dough quality. To compensate for the lack of nutrients and baking quality and to improve taste and texture of these gluten-free products, many additives are used (Caponio *et al*., 2008; Capriles and Arêas, 2014; Belz, 2016; Horstmann *et al*., 2016). GF product are therefore less healthy when consumed on a regular basis. Alternative durable solutions are therefore needed to prevent CD.

Therefore, breeding for wheat varieties that are free of immunogenic epitopes or contain a reduced amount of them, is a potential solution for healthier products safe to consume by CD patients (Jouanin *et al*., 2018; Ribeiro *et al*., 2018).

Complex gliadin gene families and epitope diversity

Gliadin proteins consist of three large gene families: α/β-, γ-, and ω-gliadins (Shewry *et al*., 2003). Gliadin genes do not contain introns but do include high numbers of pseudogenes (90% in case of α-gliadins (van Herpen *et al*., 2006)). The number of gluten genes present in each family is variable between different wheat varieties. For each gliadin family, genes are clustered together as repetitive sequences in characterised loci. In hexaploid wheat, a total of 60-150 α-gliadin gene copies is estimated (Anderson *et al.,* 1997; Ozuna *et al.,* 2015), divided over the six *Gli-2* loci, each being located on the short arm of one of the six chromosomes 6 (pair of chromosome 6AS, 6BS and 6DS) (Fig. 3). Similarly, the short arm of chromosomes 1 (1AS, 1BS and 1DS) contains the *Gli-1* loci with around 39 γ-gliadins (Sabelli and Shewry, 1991; Salentijn *et al.,* 2012) and the *Gli-3* locus, with approximately 16 ω-gliadins copies (Sabelli and Shewry, 1991), in the hexaploid genome. The *Gli-1* and *Gli-3* loci are close to each other and overlap. They also overlap with *Glu-3* loci that contain 35 copies of LMW-glutenins, which carry few immunogenic epitopes

(Sabelli and Shewry, 1991). The 20 HMW-glutenin genes, however, are present at *Glu-1* loci on the long arm of chromosome 1 (AL, 1BL and 1DL) and do not carry known immunologic epitopes (Forde, 1985).

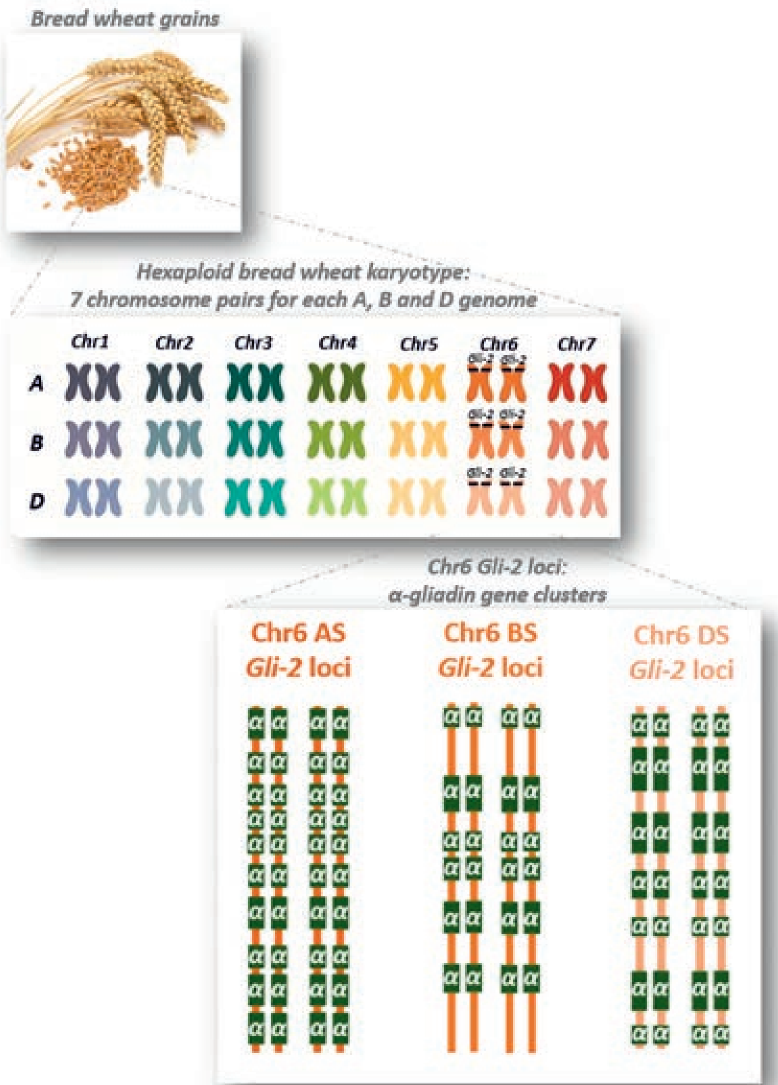


Figure 3. The complex genetics of bread wheat represented by the situation of α-gliadins at the Gli-2 loci on the homoeologous chromosomes 6.

Bread wheat (2n=6x=42) is hexaploid containing three genomes A, B and D, each carrying seven chromosome pairs. On the short arms of the chromosomes 6 are located the *Gli-2* loci, where α-gliadin genes are clustered. The α-gliadin genes present in one *Gli-2* locus differ in length and sequences. *The Gli-2* loci from genome A, B and D contain different numbers of α-gliadin genes and some α-gliadin gene sequences are genome-specific. Similar situations occur for the four other gluten gene families.

Each gene of a gene family is composed of relatively conserved motifs, repeat motifs, and variable motifs. The epitopes reside in the repetitive regions. Not all gene copies from a gene family are similar: Some carry specific variations, deletions or repetition of some motifs within the epitope sequences. Therefore, in one gliadin gene family of one common wheat plant, gene copies encode for gliadin proteins carrying different numbers of epitopes, with some being more immunogenic than others. Not all genes are expressed. The large gene families also contain pseudogenes, estimated up to 90% for α–gliadins (van Herpen *et al.*, 2006), which have early stop codons. A feature of gliadin genes is the absence of introns, enabling protein epitope prediction directly from genomic DNA sequences rather than from messenger RNA. Interestingly, it has been shown that reduction or increase of expression in one family usually leads to compensation by other families (Galili *et al.*, 1986; Pistón *et al.*, 2011). α-gliadins are the most immunogenic types of gliadin, followed by γ-gliadins and ω-gliadins. In this thesis, I focussed on reducing α-gliadins and γ-gliadins as proof of concept.

Gliadin epitopes, immunogenicity and homoeologous genomes

α-gliadins

α-gliadin proteins are up to 320 amino acids long. One can distinguish six domains: signal peptide, repetitive domain, polyQ repeat domain 1, unique domain 1, polyQ repeat domain 2 and unique domain 2 (Fig. 4). The innate epitope and four other overlapping DQ2.5 CD epitopes can be present in the repetitive region while the DQ8.x epitope can be found in the unique domain 2. Several transcriptomics, proteomics and immunology studies revealed differences in gliadins encoded on short arm of different homoeologous chromosomes 6.

The first one to be encountered in the protein, p31-49 innate epitope (PGQQQPFPPQQPYPQPQPF), located at the beginning of the repetitive domain, triggers mainly an innate immune response. It is present in genes from chromosome 6DS, although variants are found in 6AS and 6BS.

Downstream, also within the repetitive domain, another epitope domain is located where one to six overlapping epitopes can be identified. The common α-gliadin epitopes recognised by HLA-DQ2.5 (Sollid *et al.*, 2012) (Fig. 4) are as following:

- DQ2.5-glia-α1a: PFPQP**Q**LPY, on chromosome 6DS (variants on 6AS)

- DQ2.5-glia-α1b: PYPQP**Q**LPY, on chromosome 6DS

- **DQ2.5-glia-α2:** PQP**Q**LPYPQ, on chromosome 6DS (variants on 6AS), is the most common immunogenic α-gliadin epitope

- DQ2.5-glia-α3: FRP**Q**QPYPQ, on chromosome 6AS and 6DS.

(**Bold** and underlined glutamine (Q) can be deaminated into a glutamic acid (E) by tTG):

Some epitopes can be found as overlapping combinations (Fig.4), in proteolysis-resistant peptides, increasing their immunogenicity:

- DQ2.5-glia-α19mer: LQLQPFPQP**Q**LPYPQP**Q**PF,
  2 overlapping epitopes, on chromosome 6DS (or variants on 6AS and 6BS),

- DQ2.5-glia-α26mer: LQLQPFPQP**Q**LPYPQP**Q**LPYPQPQPF,
  4 overlapping epitopes, on chromosome 6DS,

- DQ2.5-glia-α33mer: LQLQPFPQP**Q**LPYPQP**Q**LPYPQP**Q**LPYPQPQPF,
  6 overlapping epitopes, on chromosome 6DS (Molberg *et al.*, 2004).

Within the unique domain 2, another immunogenic epitope is found that is recognised by HLA-DQ8.x: DQ8-glia-α1: **Q**GSFQPSQ**Q** (Sollid *et al.*, 2012), present in genome D.

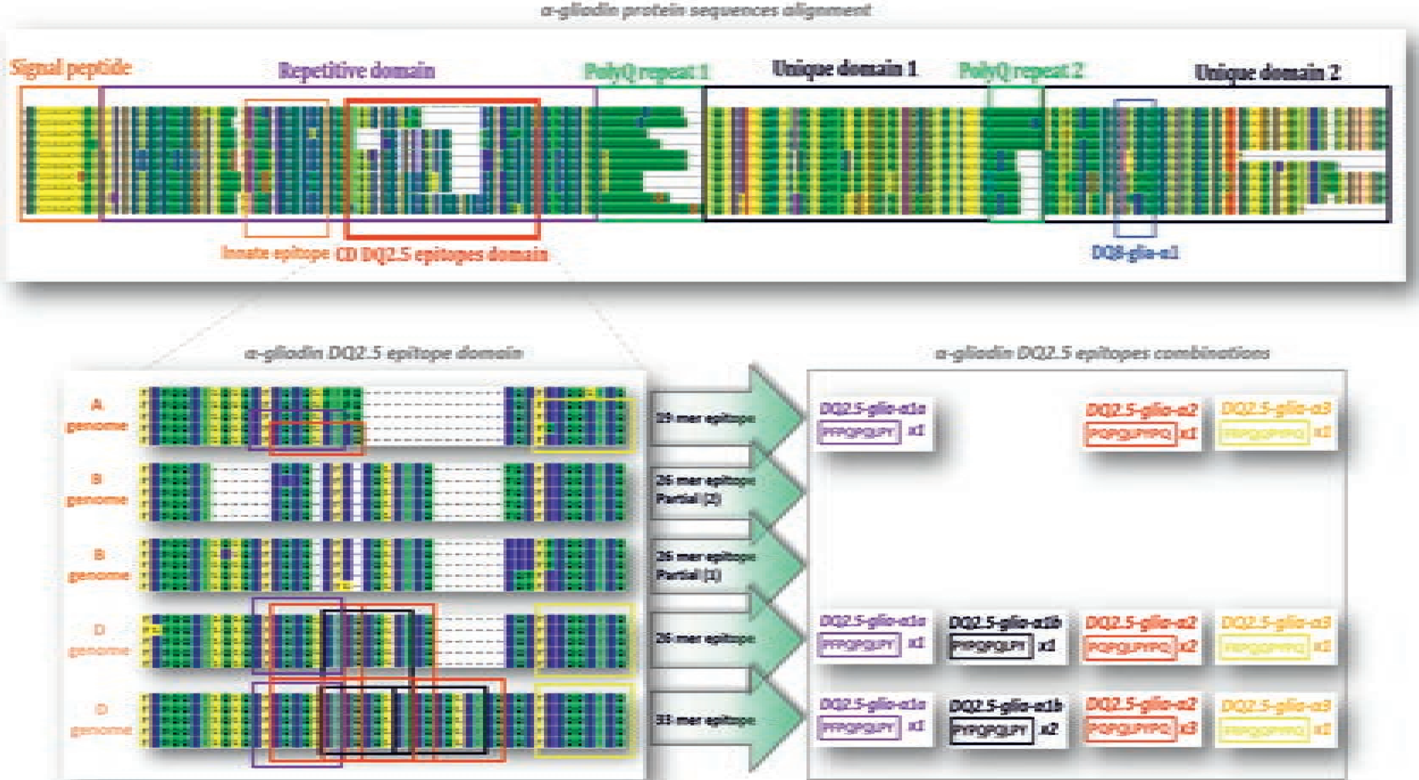


Figure 4. Alignment of α-gliadin protein sequences depicting the different protein domains and the location of CD epitopes.
The alignment displays the diversity of α-gliadin protein sequences. The different protein domains are represented. A zoom-in is made on the region of the DQ2.5 epitopes, where sequences with different epitope combinations can be found. The different sequences were sorted based on the homoeologous genome they originate from, revealing the absence of CD immunogenic epitopes in α-gliadins from genome B in contrast to the high number of immunogenic epitopes found in α-gliadins from genome D.

Typically, the expression level and immunogenicity rate of α-gliadins can be ordered as D > A > B (van Herpen *et al.*, 2006). The differences in immunogenicity are explained by the sequence variation of genes coming from different homoeologous chromosomes, coding for different amino acids at positions P1, P7 and P9 and by different possibilities of Q deamination at P4 and P6, influenced by the flanking amino acids within the epitope (Brusco *et al.*, 1999; Vartdal *et al.*, 1996). These parameters influence the binding affinity of the epitopes to HLA-DQ and the T cells.

- The α-gliadin genes from genome D usually code for the highest number, most complex combinations and the most immunogenic epitopes, with Q's at positions P4 and P6 that can easily be deaminated by tTG (Salentijn *et al.*, 2013).

- The α-gliadin genes from the B genomes code for a variant of the 26-mer DQ2.5-glia that carry deletions of Lysine (L) at P11 and Glutamine (Q) at P15, that affect the four overlapping DQ2.5-glia-α1a, DQ2.5-glia-α2, DQ2.5-glia-α1b, and 2nd DQ2.5-glia-α2 epitopes at P7, P5, P9, P4 and P2 respectively. This renders those epitopes much less or non-immunogenic (Mitae *et al.*, 2010), both individually and combined.

- The α-gliadin genes from genome A usually code for a variant of the 19-mer DQ2.5-glia that carry substitution of Proline (P) by Serine (S) at P12 that affect the two overlapping DQ2.5-glia-α1a, DQ2.5-glia-α2 at P8 and P6 respectively, rendering them less immunogenic (Mitae *et al.*, 2010) both individually and combined.

γ-gliadins

γ-gliadin proteins are up to 370 amino acids long. One can distinguish five domains: signal peptide, repetitive domain, unique domain 1, Q repeat domain and unique domain 2. The DQ2.5-glia-γ4d and six other overlapping DQ2.5 epitopes are present in the repetitive region while the DQ2.5-glia-γ2 epitope can be found in the unique domain 2. Several studies showed differences in gliadins related to homoeologous genomes.

The first one, DQ2.5-glia-γ4d (PQP**Q**QPFCQ), located at the beginning of the repetitive domain is absent from chromosome 1AS.

Downstream, also within the repetitive domain, another epitope domain is located where one to six overlapping epitopes can be identified. The common γ-gliadin epitopes recognised by HLA-DQ2.5 are (Sollid *et al.*, 2012):

- **DQ2.5-glia-γ1:** PQQSFP**Q**QQ, found on 1AS, 1BS and 1DS, is the most immunogenic γ-gliadin epitope

- DQ2.5-glia-γ3: QQP**Q**QPYPQ, probably present only on chromosome 1BS

- DQ2.5-glia-γ4a: SQP**Q**Q**Q**FPQ, only present in a few pseudogenes from chromosome 1DS

- DQ2.5-glia-γ4b: PQP**Q**Q**Q**FPQ, probably present only on chromosome 1BS

- DQ2.5-glia-γ4c: QQP**Q**QPFPQ, found on 1AS, 1BS and 1DS, most common epitope

- DQ2.5-glia-γ5: QQPFP**Q**QPQ, on chromosome 1BS and 1DS, one of the most common epitopes, rarely found on chromosome 1AS

Some epitopes can be found as overlapping combinations, proteolysis-resistant, increasing their immunogenicity:

- DQ2.5-glia-γ17mer: **Q**QP**Q**QPFP**Q**QP**Q**QPFPQ, 3 overlapping epitopes, on chromosome 1DS

- DQ2.5-glia-γ26mer: FLQP**Q**QPFP**Q**QP**Q**QPYP**Q**QP**Q**QPFPQ, 3 overlapping epitopes, on chromosome 1DS

Within the unique domain 2, another immunogenic epitope, DQ2.5-glia-γ2: IQP**Q**QPAQL, is found except in γ-gliadins on chromosome 6AS.

Typically, the expression level and immunogenicity rate of γ-gliadins can be ordered as D > A > B (Salentijn *et al*., 2012), as for α-gliadins (van Herpen *et al.,* 2006). Salentijn *et al.* (2012) showed that specific DQ2.5-glia-γ1 epitope variants in which Q2 and Q9 are deamidated stimulate the T cell clone while variants in which Q8 instead of Q9 is deamidated or the F5 is replaced by S5 do not display such activity. Thus, natural variants of the DQ2-γ-I epitope exist that influence the T-cell stimulatory capacity. They also showed the effect of flanking amino acids on glutamine deamination.

Breeding for hypoimmunogenic gluten in wheat

A challenge too complex for conventional breeding

An alternative to gluten-free food for CD patients is to breed for wheat varieties that are free of immunogenic epitopes or contain a reduced amount of them, is an attractive solution for healthier products safe to consume by CD patients (Jouanin *et al*., 2018; Ribeiro *et al*., 2018). Breeding wheat without immunogenic epitopes (Gilissen *et al*., 2014) would be a definitive solution for CD patients (Shewry and Tatham, 2016). As a first step, developing "hypoimmunogenic gluten" wheat varieties that retain baking quality is already very challenging. Firstly, gluten proteins are encoded by five gene families containing many immunogenic epitopes. Within these families, α-gliadins on chromosomes 6 trigger CD strongly, followed by γ-gliadins, ω-gliadins and, to a lesser extent, LMW glutenins on chromosomes 1. Secondly, bread wheat is allohexaploid, with three sets of chromosomes referred to as genome A, B and D. Each of these genomes contains all gluten gene families. As a result, a single bread wheat variety has a combination of gliadins and glutenins, some without any CD epitopes, others with one or more immunogenic epitopes (van Herpen *et al*., 2006; Tye-Din *et al*., 2010; Salentijn *et al*., 2013). No cultivated wheat or wild relative has been identified that contains only CD safe gluten epitopes (van den Broeck *et al*., 2010a; 2010b). Consequently, gluten hypoimmunogenic wheat cannot be produced by conventional breeding that aim at identifying and introgressing naturally occurring CD-safe gliadin gene variants into a variety.

Hurdles faced by Genetically Modified plants and mutation breeding

RNAi targeting all three gliadin families successfully decreased 97% of gliadin expression in bread wheat grain and its gluten extract did not stimulate CD patient T cells while dough quality was barely affected (Gil-Humanes *et al.*, 2010; 2014). Similarly, Becker *et al.* (2012) decreased the expression of 20 α-gliadins although increasing expression of other storage proteins. Wen *et al.* (2012) reduced the expression of DEMETER, preventing DNA-methylation changes and thus repressing glutenin and gliadin gene expression in the endosperm. However, RNAi construct has to remain in the wheat genome to silence the gliadin genes. RNAi wheat lines are therefore hold back by GM regulation, despite the considerable improvement they represent for CD patients.

Another approach towards wheat with hypoimmunogenic gluten is mutation breeding based on random mutagenesis. This method which is exempted from GM regulation, has recently been applied to develop "ultra-low gluten" barley (Tanner *et al.*, 2016), and is being used to produce gluten-free beer in Germany. Developing wheat with hypoimmunogenic gluten using a similar approach is theoretically possible although it constitutes a greater challenge (Ribeiro *et al.*, 2018). Wheat γ-irradiated lines have to be identified that miss large region on short arm of chromosomes 1 or 6, having lost gliadin or LMW-glutenin genes in one of the three homoeologous genome. Plants with these 12 different events have to be self to become homozygous for the deletion and then inter-crossed to obtain hypoimmunogenic gluten while maintaining bread wheat unique baking quality conferred by HMW-glutenin loci. However, a potential consequence of supressing all gliadin loci is an effect on bread dough viscosity. Moreover, this process is highly time consuming and lots of plants need to be screened since the mutations are random. In addition to the deletion of gliadin loci, many other mutations occur, which can have an effect on the plant.

Project aim: use of gene editing to decrease α- and γ-gliadin immunogenicity in bread wheat

An alternative approach to generate wheat varieties with non-immunogenic gluten is the precise editing of gluten genes, which can be achieved with a targeted mutagenesis approach, such as using the CRISPR/Cas9 system.

The CRISPR/Cas9 system

The CRISPR/Cas9 system was originally described as a bacterial defence mechanism against viral infections (Barrangou *et al.*, 2007) and has since been customised to induce mutations and deletions at specific locations within a target gene in animals and plants (Doudna and Charpentier, 2014). CRISPR/Cas9 is relatively fast and easy to customise to target different genes of interest and mutate them at specific sites. It is composed of a single guide RNA (sgRNA) that contains a sub-sequence of 20-mer which may be customised to be complementary to the genome site to target and mutate. The sgRNA guides a Cas9 endonuclease to the target region within the genome and binds to it. Then, the Cas9 generates a double-strand break in the DNA, three bases upstream of a specific protospacer adjacent motif (PAM) (Fig. 5). This triggers the cell's DNA repair mechanism, which in case of a double-stranded break in plants is mostly done by non-homologous end joining (NHEJ). This can result in nucleotide insertion/deletions (indels) or other mutations (Belhaj *et al.*, 2013), and offspring in which such mistakes have been made can be selected.

CRISPR/Cas9 constructs can be integrated in plant genomes to generate the desired double-strand breaks, and the construct can be out-crossed in subsequent generations while the mutations of interests are retained. This is a considerable advantage compared to GM techniques such as RNAi, in which the transgene must remain present for the desired effect. The CRISPR/Cas9 complex can even, in certain cases, be delivered into plant protoplast cells as riboprotein complex (Woo *et al.*, 2015), which generates the targeted double-strand break and is subsequently naturally degraded. The mutations that are generated would remain in the regenerated plant and its progenies. In addition, CRISPR/Cas9, being a targeted mutagenesis method, generates less off-target mutations compared to random mutagenesis, or even no off-type mutations at all (Feng *et al.,* 2014). The targeted approach also decreases the minimal number of mutant plants to generate and screen before identifying a desired mutation, which is time and cost efficient. Furthermore, when dealing with multiple genes, CRISPR/Cas9 enables simultaneous mutations either by multiplexing sgRNA with different targets or by targeting conserved regions (Xing *et al.*, 2014). The mutations that are generated may be homozygous for all copies within a unique generation.

This PhD project

The concept of the project is to use CRISPR/Cas9 to generate mutations within α- and γ-gliadins in order to modify or remove the immunogenic epitopes, to completely remove the genes, or to prevent their expression (Fig. 5). This approach would enable obtaining wheat plants producing grains containing fewer or no immunogenic gliadin epitopes yet retaining nutritious values and acceptable bread dough quality provided by the untargeted glutenins and newly generated CD-safe gliadins. Although "safe-gluten" wheat is the ultimate goal, we use the term "hypoimmunogenic" gluten and wheat.

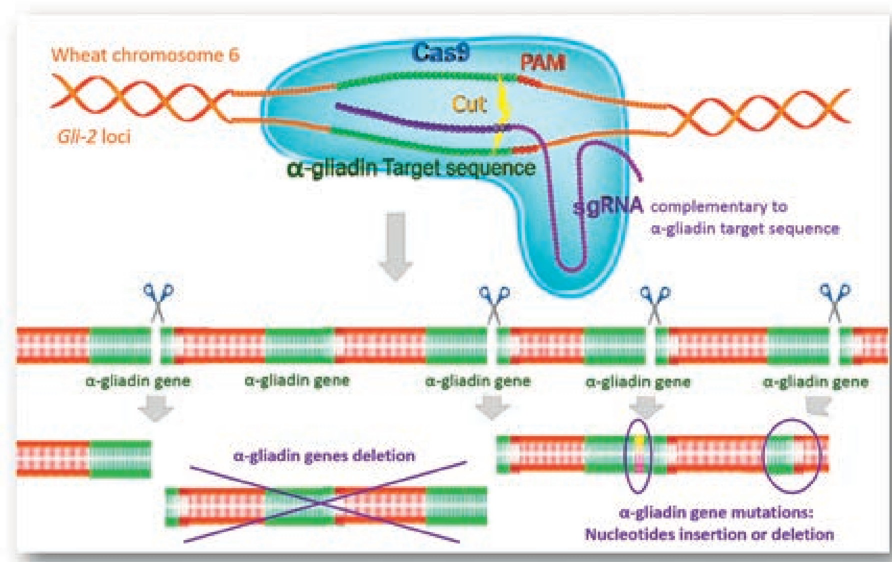


Figure 5. Representation of a CRISPR/Cas9 complex targeting a Gli-2 locus containing α-gliadin genes. CRISPR/Cas9 consists of a sgRNA (in purple) that contain a protospacer sequence designed to be complementary to the DNA sequence of the genes to target - which are here α-gliadins or γ-gliadins – (in green). The sgRNA finds the gliadin sequence targeted within the wheat genome (represented as the orange strand) and bind to it, bringing alongside a nuclease enzyme called Cas9 (in blue). The Ca9 envelops the double-stranded DNA, recognises the protospacer adjacent motive PAM (in red) and generates a DNA double strand break three base pairs upstream of the PAM. The DNA repair mechanism of the plant cell may produce a mistake during the damage repair, generating mutations such as insertion or deletion of few nucleotides. In addition, as α-gliadin genes occur in tandem repeats, simultaneous double-strand breaks in two non-consecutive copies could generate the deletion of a large DNA fragment carrying several α-gliadins gene copies. This figure has been adapted from GenScript, Nanjing, China.

When the project started, CRISPR/Cas9 had successfully been used in polyploid wheat protoplasts to induce mutations in one genes in the homoeologous genomes (Shan *et al.*, 2013) or in diploid mice embryogenic stem cells to target a small

gene family containing 5 genes (Wang *et al.*, 2013a). However, no one had ever reported the use of CRISPR/Cas9 to target large gene families with clustered genes in polyploid crops. At that time, the way of regulating gene-edited plants was not decided in Europe nor elsewhere. Therefore, the three main questions driving the research were:

- Can CRISPR/Cas9 successfully mutate a sufficient number of gliadin gene copies to modify gliadin proteins expressed in hexaploid wheat grain?

- What screening method can be used first to identify the mutants and then to characterise the mutations generated?

- How much can gene-edited regulatory decisions impact research, development and commercialisation of gene-edited wheat with hypoimmunogenic gluten?

This project should therefore be seen as a pilot research project regarding the development of a methodology and the applicability of CRISPR/Cas9 gene editing approaches to create bread wheat containing hypoimmunogenic gluten for CD patients. This proof of concept will not immediately remove all epitopes in the targeted gene families, let alone those in ω-gliadins and LMW-glutenins. Furthermore, once all known epitopes would be removed, the Cas9 construct should be outcrossed, lines should be made homozygous and gluten should be tested for immunogenicity and dough rheology.

The structure of the thesis

In this PhD thesis, *Chapter 1* gives an overview of the food processing and plant breeding strategies to create coeliac-safe and healthy wheat products. This chapter highlight the fact that solutions decreasing gliadin immunogenicity exists in food processing but should be combined with plant breeding strategies in order to have the safest wheat product possible for CD patients.

Among the plant breeding strategies described to reduce gliadin immunogenicity in wheat, the use of CRISPR/Cas9 targeted mutagenesis editing α- and γ- gliadin genes constitutes the theme of *Chapter 2***.** This chapter describes the generation

of hexaploid Fielder wheat plants containing CRISPR/Cas9 constructs with gRNA editing α- and γ- gliadin genes. It presents the Acid-PAGE screening results identifying α- and γ- gliadin Fielder-CRISPR mutant plants. These basic proteomics results were compared to the ones obtained for the mutants of bread wheat cultivar Paragon that were generated using γ-irradiation, a random mutagenesis approach.

These basic proteomics evidences of wheat gliadin mutations needed to be further studied using genomic analysis. Consequently, *Chapter 3* describes the development of a droplet digital PCR assay as a quantitative method to determine gene copy number variation and gene-editing detection of α-gliadin gene family, in hexaploid wheat. In this chapter, the methodology for setting up a ddPCR assay that can estimate gene copy number variations in large gene families, in polyploids, is described. After optimisation, ddPCR was implemented on wild type and γ-irradiated Paragon lines previously revealing gliadins mutations in order to test the possibility of quantifying deleted gene copies. The ddPCR assay was then applied to the Fielder-CRISPR gliadin mutant plants to quantify and distinguish between Cas9 induced small indels and large DNA fragments deletion.

Quantitative data regarding gliadin genomic mutations are not sufficient and mutations generated must be qualitatively characterised. Thus, *Chapter 4* presents GlutEnSeq, gluten gene enrichment and sequencing, as a qualitative method to analyse genomic variation of gluten genes in hexaploid wheat. This chapter describes the design and development of a solution-based gluten exome capture baits assays used to perform gluten gene enrichment followed by next generation sequencing, based on REnSeq. The method was first tested on reference cultivar Chinese Spring, before it was implemented on Paragon and Fielder cultivars and associated gliadin mutant lines in order to get qualitative information on the type of genetic mutations generated.

Gene editing being the core of the project and its new regulation as GM in Europe is currently a hot topic. *Chapter 5*, entitled "Development of wheat with hypoimmunogenic gluten obstructed by the gene editing policy in Europe" covers the topics of safety, public acceptance, distinguishability and labelling, and potential politico-economic issues, related to the regulation of gene-edited as GM product in Europe in contrast to other countries, following the ruling of the

European Court of Justice in July 2018. It also advises the European Commission to review its position on the matter and to regulate gene editing based on the generated products and the innovation principle.

Finally, a *General Discussion* elaborates on the necessity of developing gluten-safer products using plant breeding. The level of success of CRISPR/Cas9 gene editing for developing wheat lines with lower gliadin immunogenic epitopes is discussed. Alternative solutions are presented for faster development of wheat with hypoimmunogenic gluten. The relevance of fast mutant screening methods and their adaptation is discussed. The necessity of developing methods to test food for "safe-gluten" rather than "gluten-free" is also suggested.

Chapter 1

Food processing and breeding strategies for coeliac-safe and healthy wheat products

Aurélie Jouanin[1,2], Luud J.W.J. Gilissen[1], Lesley A. Boyd[2], James Cockram[2], Fiona J. Leigh[2], Emma J. Wallington[2], Hetty C. van den Broeck[1], Ingrid M. van der Meer[1], Jan G. Schaart[1], Richard G.F. Visser[1], Marinus J.M. Smulders[1]

[1]Wageningen University and Research, Wageningen, The Netherlands
[2]NIAB, Cambridge, UK

Published in Food Research International, 110, 11-21 (2018)
https://doi.org/10.1016/j.foodres.2017.04.025

Abstract

A strict gluten-free diet is currently the only treatment for the 1-2% of the world population who suffer from coeliac disease (CD). However, due to the presence of wheat and wheat derivatives in many food products, avoiding gluten consumption is difficult. Gluten-free products, made without wheat, barley or rye, typically require the inclusion of numerous additives, resulting in products that are often less healthy than gluten-based equivalents. Here, we summarise and contrast two broad approaches to decrease wheat gluten immunogenicity for CD patients. The first is based on food processing strategies, which aim to remove gliadins or all gluten from edible products. We find that several of the candidate food processing techniques to produce low gluten-immunogenic products from wheat already exist. The second approach focuses on wheat breeding strategies to remove immunogenic epitopes from the gluten proteins, whilst maintaining their food-processing properties. A combination of breeding strategies, including mutation breeding and possibly genome editing, will be necessary to produce coeliac-safe wheat. Individuals suffering from CD and people genetically susceptible who may develop CD after prolonged gluten consumption would benefit from reduced CD-immunogenic wheat. Although the production of healthy and less CD-toxic wheat varieties and food products will be challenging, increasing global demand may require these issues to be addressed in the near future by food processing and cereal breeding companies.

Key words: *Gluten, Gliadin, T cell Epitope, Coeliac Disease, Plant breeding, Food processing, Mutation breeding, Genome editing.*

Introduction

Wheat forms the basis of a diverse range of staple foods globally, which includes bread and bakery items in addition to other products such as alcoholic beverages and food additives. Two major storage protein families are present in wheat grains: glutenins and gliadins (belonging to glutelins and prolamins, respectively), which are collectively known as gluten. The qualitative and quantitative properties of wheat gluten are important factors for baking quality. Despite its predominance in the human diet, wheat can cause allergies and/or intolerances after consumption in susceptible individuals (Fasano, 2006). Wheat contains immunogenic proteins, but the prevalence of wheat allergy, which is mediated by IgE antibodies, is low (Zuidmeer *et al.,* 2008; Gilissen *et al.,* 2014). However, gluten can also trigger intolerance reactions in 1-2% of the human population, the most common of which is coeliac disease (CD) (Mäki *et al.,* 2003; Rewers 2005; Catassi *et al.,* 2014; Vriezinga *et al.,* 2015).

CD is a chronic inflammation of the small intestine, which leads to a variety of symptoms that can include malnutrition, bowel disorders, skin, bone, nerve, and muscle problems. It is induced by epitopes (peptide sequences of nine amino acids) present in gluten proteins from wheat, barley and rye. These epitopes are bound by human leucocyte antigen sub-units (HLA-DQ2 or HLA-DQ8) located on antigen presenting cells in humans, which are subsequently recognised by T cells, triggering an immune reaction (for reviews see Fasano 2009; Koning 2012). CD develops in genetically predisposed individuals, and is induced by gluten ingestion. A strict gluten-free diet is currently the only treatment to stop the immune reaction and inflammation, allowing the intestine to recover, after which the symptoms gradually disappear.

Avoiding consumption of gluten is difficult to do in practice, due to the presence of wheat derivatives in many supermarket products (Atchison *et al.,* 2010). Therefore there is a need for new strategies to create food products with low gluten immunogenicity. Here, we summarise several strategies to decrease wheat gluten toxicity for CD patients, based on food processing and on wheat breeding (Fig. 1). The food processing strategies aim at removing gliadins or whole gluten from edible products (Gilissen *et al.,* 2016b) whereas the wheat breeding strategies focus on the removal of immunogenic epitopes from the gluten proteins, whilst

potentially maintaining their food-technological properties (Smulders *et al.*, 2015; Shewry and Tatham 2016). If successful, the latter strategy may result in food products considered coeliac-safe, despite containing large amounts of gluten, which would require alternative labelling from the term 'gluten-free'. A comparison of the applicability of these approaches, in terms of technical and economic aspects, is presented. Two groups of people may benefit especially from food products derived from such reduced CD-immunogenic wheat: (1) individuals suffering from CD, whether they are already or not yet diagnosed as CD patient, and (2) genetically predisposed individuals who may develop CD after prolonged wheat or gluten consumption.

Food processing-related strategies to manufacture gluten-free products or to lower the content of CD epitopes

Wheat gluten is a polymer composed of two types of proteins: glutenins and gliadins. The type of HMW subunits and the ratio of glutenins and gliadins determine the viscoelasticity properties required for bread, cakes or other baking products (Weegels *et al.,* 1996; Shewry *et al.,* 2009). The high and low molecular weight glutenin subunits (HMW-glutenins and LMW-glutenins) form a strong glutenin macro polymer (GMP) network, formed by inter-chain disulphide bonds, which determines dough elasticity. The monomeric gliadin proteins (α/β-gliadins, γ-gliadins and ω-gliadins) are linked to the GMP and can be envisaged to 'dilute' the strength of the GMP, thus influencing dough viscosity (Shewry 2009). The glutenins are the less immunogenic gluten proteins for CD. The monomeric gliadin proteins (α/β-gliadins, γ-gliadins and ω-gliadins) are the most immunogenic.

Two types of products can be generated during food production: gluten-free products made without ingredients from wheat, barley or rye (Fig. 1a), or products with low gluten immunogenicity based on the processing of these cereal grains to remove the immunogenic gluten fraction (Fig, 1b) (as reviewed by Kucek *et al.,* 2015). Food products can be labelled as 'gluten-free' as long as they contain less than 20 ppm of gluten (Saturni *et al.,* 2010).

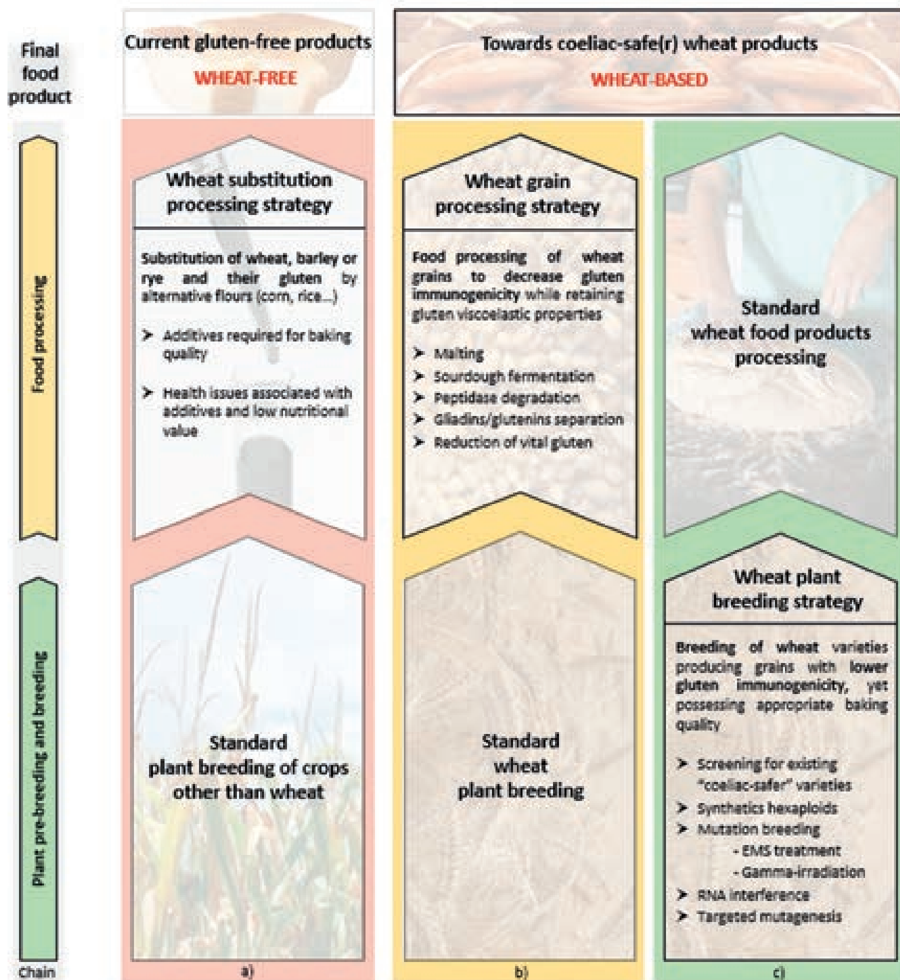


Figure 1. Food supply chains producing gluten-free or coeliac-safe(r) products.
Depicted are simplified food supply chains comprising of two steps, plant breeding and food processing, leading to the final food product. Three variants, a), b) and c), of this food supply chain can be envisaged, depending on the step at which efforts are made to generate gluten-free or coeliac-safe(r) products and whether the process includes wheat ingredients or not. These three variants are presented here in the same order as described in the paper. a) Gluten-free products are generated using substitutions for wheat to avoid the presence of its immunogenic gluten fraction. b) Coeliac-safer of celiac-safe products could be made at the food-processing step by processing wheat grains to decrease their immunogenic gluten fraction before proceeding to processing of the final product. c) Coeliac-safer or coeliac-safe products could be made based on plant breeding of wheat varieties with reduced immunogenic gluten fraction, used in a standard grain processing and final food product processing. Note that b) and c) can complement each other to achieve a further reduction of gluten immunogenicity.

Gluten-free products made without ingredients from wheat, barley and rye

With increasing numbers of patients with CD being diagnosed, the gluten-free market is becoming more important, and gluten-free products are increasingly found on grocery shop shelves. Such products generally do not contain any wheat, barley or rye, but include several additives (Fig. 1a) to recreate the viscoelastic properties of gluten present in these cereal grains (as summarised by Gustafson 2016).

Current composition of gluten-free products

Currently commercial gluten-free bread is usually made from a mixture of flours, including maize, rice, sorghum and oat. Oat can be consumed safely by CD patients without a daily limit (Kaukinen *et al.,* 2013; Hardy *et al.,* 2015; Gilissen *et al.,* 2016a and references therein). Alternatively, flour from pseudo-cereals (such as amaranth, buckwheat and quinoa), legumes and nuts have been proposed as new candidates for gluten-free products.

Issues encountered with these gluten-free flours are their lack of viscoelasticity and their low ability to retain water and CO_2, the properties responsible for dough rising conferred by the gluten network. Starch may be added to increase viscosity and provide some binding abilities to the dough. Starches from potato, corn, rice and cassava (tapioca) are mostly used. Hydrocolloids, also known as gums, are also important additives in gluten-free baking products. They are used as substitute for glutenins to confer elasticity to the dough, as well as gas and moisture retention for higher loaf volumes and longer shelf life. Inulin, xanthan, guar gum, and hydroxypropyl methylcellulose (HPMC) are the most commonly used hydrocolloids in gluten-free bread. Protein concentrates such as soybean, egg and whey proteins are also used for their gelling properties, contributing to dough quality improvement. The addition of lipids, such as sunflower and safflower oil, retards the staling process, due to the gas-retaining properties of polar lipids (Horstmann *et al.,* 2016). Additionally, emulsifiers such as glycerol monostearate and lecithin are commonly used to reduce staling and to increase shelf-life. Sodium stearoyl-2-lactylate is also an emulsifier but acts by stretching the gluten network. High salt content is often used to improve taste, texture and bread stability (Belz, 2016). The addition of dietary fibre through incorporation of oat or rice bran could become a solution to compensate for the low fibre content of gluten-free bread (Gilissen *et al.,* 2016b). Rice bran also appears to improve

rheology and sensory aspects of gluten-free bread, whereas dough acidification with acetic, citric, or lactic acid, or monosodium phosphate can also help improve bread volume.

Health issues associated with current gluten-free products

Gluten-free bread thus contains far more additives than standard wheat bread. As a result, regular consumption of gluten-free bread may in itself cause health issues, e.g. the mineral content can be very low, leading to iron and calcium deficiency, and despite the inclusion of protein additives for dough improvement, it is a relatively poor source of protein. Gluten-free bread is usually low in fibre, which decreases the sensation of satiety; in contrast, it is rich in starch, thereby enhancing the risk of hyperglycaemia and diabetes. Gluten-free bread also contains high levels of salt, which is a risk for cardiovascular and renal disease. In addition to being high in lipids, gluten-free products contain *trans* fatty acids, which may provoke metabolic imbalance (Caponio *et al.,* 2008). Several countries try to reduce or eliminate them from food products. Fructan-type inulin is used in gluten-free bread as a hydrocolloid but is known to be a possible cause of irritable bowel syndrome (IBS) (Capriles and Arêas, 2014). IBS is mainly caused by fermentable oligosaccharides, disaccharides, monosaccharides and polyols (FODMAPs) such as fructan, a fructose polymer. Since fructans are present in the wheat grain, IBS is often associated with wheat as non-coeliac gluten sensitivity or non-coeliac wheat sensitivity (NCGS), and is sometimes wrongly auto-diagnosed as CD. While a low FODMAPs diet will usually decrease IBS symptoms, some gluten-free bread products containing inulin may not be suitable for these patients. In such cases, it is important to be diagnosed with the correct disorder. Nonetheless, based on the frequency of people with clinical bowel complaints that appear to improve under a gluten-free diet, the prevalence of NCGS might be as high as 6% of the population (Gilissen *et al.,* 2014), but more recent evidence suggests that 'true' NCGS is only slightly more than 1% of the population (Catassi, 2015). Although no biomarkers have been identified yet, a state of systemic immune activation in conjunction with a suggested compromised intestinal epithelium could be observed in a subset of individuals with NCGS (Uhde *et al.,* 2016). In addition to individuals diagnosed with CD or IBS, other consumers have actively decided to avoid wheat and gluten based on personal choice. However, such choices are often made following self-diagnosis based on personal health-related preference, rather than for clear medical reasons, despite the dietary concerns described above.

Gluten contamination elimination diet (GCED)

Some coeliac patients have persistent symptoms and villous atrophy despite strict adherence to a gluten-free diet. They are referred to as having 'non-responsive CD'. Some non-responding CD patients may react to minute amounts of gluten, likely present as contaminants in processed gluten-free foods. For example, industrial conversion of wheat starch into other food-grade carbohydrates may not eliminate all gluten (Scherf *et al.,* 2016). Therefore, foods containing these modified carbohydrates may still contain very low levels of stable, and potentially immunogenic, gluten. Furthermore, when chemically deamidated gluten protein, which is even more immunogenic than natural gluten proteins, is used for increased emulsifying applications (Wu *et al.,* 1976) it may be undetectable in the R5 monoclonal antibody assay used for routine gluten screening by commercial companies (Kanerva, 2011). These undetected deamidated gluten fractions can also contribute to contamination of food labelled as 'gluten-free'.

Over 80% of the non-responsive patients showed health improvement under a 'gluten contamination elimination diet' (GCED), that includes only fresh and unprocessed foods. These patients were able to return to a traditional gluten-free diet without return of symptoms (Hollon *et al.,* 2015). A subset of 'non-responsive CD' patients do not show any decrease in symptoms under GCED, and therefore suffer from true 'refractory CD' (RCD).

Development of wheat-based products with low gluten immunogenicity

Due to the low dietary value and health risks associated with non-wheat based gluten-free products, alternative food processing methods for wheat are being investigated (Fig. 1b). The main aim is to use wheat (conserving the rheological properties conferred by the glutenins), whilst removing the most immunogenic gluten fraction (the gliadins). This approach would decrease the need for additives to reach adequate dough quality and dietary values, thus improving the healthiness of gluten-free or 'gluten-safe' products.

Malting

The malting process is not only relevant for brewing beer, but also for the baking sector. Malting represents the first step in seed germination, during which amylase enzymes become active, degrading starch into maltose to provide the source of energy for the embryo to germinate. Malting produces higher amounts of maltose,

which enhances further fermentation. In addition, endogenous proteolytic enzymes in the grain degrade storage proteins (i.e. gluten), releasing amino acids that are used by the growing seedling. Gliadins become degraded through the activity of proteases (Bigiarini *et al.,* 1995). Because of this degradation process, in practice many brands of beer are technically gluten-free or contain only small amounts of gluten.

Malting has been used for many years to add flavour and crust colour to bread, and can even enhance sourdough fermentation when the malting process is not fully stopped. It has been shown that malted and fermented wheat sourdough effectively degrade 95% of the gliadins (Loponen *et al.,* 2007), leading to bread that is low in gliadins. An alternative approach is the use of enzyme extracts from a small stock of germinated seeds (particularly barley seeds). This has recently been suggested to be applicable to the large-scale production of flour with a reduced gliadin content (Stenman *et al.,* 2010; Schwalb *et al.,* 2012).

Sourdough fermentation and dough acidification

Sourdough is the result of the activity of micro-organisms (*Lactobacilli* and yeast) naturally present in flour, becoming active when water is added. *Lactobacillus* ferments the flour/water mixture creating lactic acid, acidifying the mixture. This enhances the activity of propyl endopeptidase (PEP) enzymes produced by other micro-organisms to degrade gluten proteins (Hartmann *et al.,* 2006; Gänzle *et al.,* 2008; Loponen *et al.,* 2009). Sourdough fermentation has been suggested as a way of enabling the manufacture of 'gluten-free' or 'low-in-gluten' (< 100 ppm) wheat-based products through proteolytic gluten degradation (Loponen, 2006). Similarly, chemical acidification of conventional dough has been shown to activate PEP activity, resulting in increased prolamin degradation (Kanerva, 2011), although not as much as in traditional sourdough. However, artificially acidified dough may trigger more intestinal permeability (Di Cagno *et al.,* 2004) and enhance non-enzymatic deamination of gluten peptides, which may render them more immunogenic (Arentz-Hansen *et al.,* 2000).

Sourdough fermentation is a traditional baking process with a long history, particularly in Germany. The prevalence of CD in Germany (measured using a single antibody test) has been found to be remarkably low, 0.4% and 0.2% in women and men, respectively (Kratzer *et al.,* 2013), confirming earlier data showing a 0.1-

0.4% prevalence on the basis of anti-tTG and anti-endomysial antibody (EMA-IgA) tests (Mustalahti *et al.,* 2010). Recent estimations of the celiac prevalence in Germany have increased to 0.8%, especially in children and young adults up to 17 years old (Laass *et al.,* 2015). This may reflect under-diagnosis in earlier studies. However, if sourdough indeed reduces the exposure to intact gluten epitopes, then the recent increase in CD prevalence in young people may reflect a change in the diet of German children towards white and yeast-based bread.

Recently, Greco *et al.* (2011) described a two-step hydrolysis including fungal proteases and *Lactobacillus* endo-peptidase activity, which resulted in a wheat flour-derived product with less than 10 ppm gluten. This product was found not to be harmful to CD patients (Greco *et al.,* 2011). These observations are promising, but further research on the potential of sourdough products in a safe 'gluten-free' diet is needed, examining individual cases. Firstly, a strict definition on the type of sourdough is required, and the process of gluten degradation needs to be analysed in detail. Secondly, it needs to be determined whether large-scale consumption by the general population of well-defined sourdough products can indeed contribute to the reduction of the incidence and symptom severity of CD and NCWS. This will require large-scale epidemiological studies.

Degradation of gluten using a peptidase

Walter *et al.* (2014) showed that AN-PEP, an *Aspergillus niger* prolyl endopeptidase, can be used to degrade gluten in wheat bran to below 20 ppm. Such treated wheat bran could have a use in increasing the nutritional value of the gluten-free diet, e.g. by increasing soluble fibre content.

Separation of gliadins from total gluten

Extraction of HMW-glutenins from total gluten and substitution of the highly immunogenic wheat gliadins with, for instance, coeliac-safe oat avenins in order to maintain good baking quality (van den Broeck *et al.,* 2011), is a realistic option in foods aiming at a strongly reduced immunogenic gluten load. Nearly complete separation can be achieved at the lab scale (Bassi *et al.,* 1997; Gilissen *et al.,* 2012; Lagrain *et al.,* 2013). Industrial scale-up to economically and technologically viable levels appears to be more problematic, and would need further research. Alternatively, α-gliadins, the most immunogenic gluten proteins, can be partially removed mechanically using roller-milling to remove the seed sub-aleurone layer,

in which a large fraction of the α-gliadins are deposited (van Herpen *et al.,* 2008; Sutton *et al.,* 2015).

Reduction of the use of vital gluten

Vital (or purified) wheat gluten is a major side-product of wheat starch production. Starch is used to make native and modified starch, glucose syrup, maltodextrins, and liquid and crystalline polyols, with wide applications in the food industry. Wheat starch is also used for the production of alcohol, including bioethanol. Since wheat starch is economically more important than flour for bakery purposes, vital gluten production (as side-product of starch isolation) is large, and its cost relatively low.

Vital gluten is increasingly applied as a protein additive in food products to improve technological characteristics. In bread and bakery products vital gluten is added to maintain high loaf volumes, but most of its use is in the form of a 'hidden' ingredient, often not labelled, in a large number of processed food products. This may have contributed to the increase in total gluten consumption: between 1977 and 2007, gluten intake in the western world has tripled (Kasarda, 2013). Currently, a 'back to basics' philosophy in modern bakeries in The Netherlands has led to 'clean label' strategies, and avoidance (or strong reduction) of bread quality-improving ingredients such as enzymes and vital gluten in artisanal and industrial whole grain breads. There is, however, no indication of a reduction in the use of vital gluten in other food production sectors.

Wheat breeding strategies for products with low gluten immunogenicity

An alternative strategy for the production of gluten-free wheat-based food is to develop wheat varieties with fewer or no immunogenic CD epitopes (Fig. 1c) (Gilissen *et al.,* 2008; Shewry and Tatham 2016). To be able to generate such varieties, one must be aware of the complexity of wheat genome and the structure of gene families coding for gliadin immunogenic CD epitopes.

Wheat breeding background: genomics, gluten gene families and gliadin protein epitopes

The ancestor of all wild and cultivated wheat species was a diploid grass (2n=14), which evolved into several diploid species (e.g. Goncharov, 2011; Goryunova *et al.,* 2012). Natural allopolyploidisation occurred through hybridisation between the diploid species *Triticum urartu* (AA genome) and *Aegilops speltoides* (BB or SS) followed by chromosome doubling, resulting in tetraploid emmer wheat, *T. turgidum* spp. *dicoccoides* (AABB). Much later, tetraploid emmer hybridized with goat grass, *Aegilops tauschii* (DD), followed by chromosome doubling, resulting in hexaploid bread wheat, *Triticum aestivum* (AABBDD) (Gupta *et al.,* 2008). The latter allopolyploidisation took place in agricultural fields early after the start of cultivation of emmer wheat; bread wheat does not exist in nature.

The α-gliadins are encoded by a large gene family around 100 genes, located on the short arms of the group 6 chromosomes (6AS, 6BS and 6DS) (Ozuna *et al.,* 2015). The γ-gliadins, ω-gliadins and LMW-glutenins are encoded by gene families located on the short arms of the group 1 chromosomes (1AS, 1BS and 1DS), while the HMW-glutenins are encoded by gene families located on 1AL, 1BL and 1DL.

The presence of the D genome in hexaploid *T. aestivum* provides additional gluten proteins, which considerably improve its bread baking properties compared to tetraploid wheat, particularly HMW-glutenin subunits (Payne *et al.,* 1984). However, the D genome also encodes for gliadins that have been reported as highly immunogenic.

CD epitopes are 9 amino acids in length, and form the peptide fragment that fits into the T cell HLA-DQ2 and HLA-DQ8 epitope-presenting receptor grooves. The list of well-defined CD epitopes (Sollid *et al.,* 2012) comprises: twenty-four DQ2 epitopes and seven DQ8 epitopes, including six epitopes present in α-gliadins and eleven in γ-gliadins, complemented by two ω-gliadin epitopes, three LMW-glutenin, and two HMW-glutenin epitopes. The HLA-DQ8 epitopes play a minor role in the development of CD, since over 90% of CD patients are HLA-DQ2 positive and react to α-gliadins (Koning, 2012; 2015). The HLA-DQ / gliadin peptide / T cell interactions have been characterised extremely well. The positions that are crucial for immune reactivity of α-gliadins, identified by T cell

tests (e.g., Mitea *et al.,* 2010), correspond exactly with the amino acid residues that form the crucial hydrogen bridges that connect HLA-DQ2 (Petersen *et al.,* 2014) to the T cell receptor, a connection which, if sufficiently stable, triggers the immune reaction in humans. The structure of the HLA-DQ8 / gliadin peptide / T cell interaction has also been characterised in detail (Petersen *et al.,* 2016).

At the genomic level, the occurrence of CD epitopes varies between gliadin genes (each locus comprising multiple gene copies), between homoeologous wheat chromosomes (A, B, D), and among wheat varieties and species (Spaenij-Dekking *et al.,* 2005; Molberg *et al.,* 2005; van Herpen *et al.,* 2006; Salentijn *et al.,* 2009; 2012; van den Broeck *et al.,* 2010a; 2010b). The α-gliadins from the D genome contribute most CD epitopes, while those from the B genome contribute the least (van Herpen *et al.,* 2006; Salentijn *et al.,* 2009). For γ-gliadins, the highest number of CD epitopes has also been found in genes present on the D genome (Salentijn *et al.,* 2009; 2013). While less sequence information is available for ω-gliadins, their involvement in CD immunogenicity has been recently recognized from cross-reactivity of T cells with rye and barley epitopes (Tye-Din *et al.,* 2010).

We will first discuss research approaches aimed at identifying naturally occurring variation within gliadin gene sequences and loci. This genetic variation can be used in conventional breeding and forward genetics approaches, by introgression into new wheat varieties. However, there is limited prospect for success, as multiple gliadin loci occur in the polyploid wheat genome, and glutenins and gliadins are encoded by large gene families, so multiple genes need to be screened. Many of the genes are closely linked in the genome, which limits the frequency of recombination, making selection of appropriate variation through a breeding approach difficult. Therefore, the use of reverse genetics approaches, such as random or directed mutagenesis and transgenic methods (such as gene silencing) or gene editing are considered attractive options to create wheat varieties with reduced gluten immunogenicity. We will discuss such approaches later in this review.

When these approaches lead to wheat varieties with fewer gliadin proteins, this may affect baking quality. The effect should be compensated for, and one approach is the addition of celiac-safe proteins that perform the same function in the GMP (Gilissen *et al.,* 2012, Smulders *et al.,* 2015), as was demonstrated for oat avenins

added to flour of wheat deletion lines lacking certain groups of gliadins (van den Broeck *et al.,* 2009).

Selection of low CD-toxic wheat accessions and varieties

Genebanks around the world contain modern and old varieties of hexaploid and tetraploid wheat varieties, as well as landraces (locally adapted wheat accessions), wild emmer (*Triticum dicoccoides*) and wild diploid species that are related to the ancestors of cultivated wheat (crop wild relatives). These collections are a useful source of wheat germplasm for analysing the relationships between individual gluten genes or proteins, the occurrence of epitope variants, and the resulting CD immunogenicity.

Whilst there may be large variation in the presence of intact CD epitopes amongst individual α -gliadin genes, all wheat varieties or accessions express multiple gliadins (Salentijn *et al.,* 2013), which reduces the variation detected between varieties by immunogenic screening approaches. Indeed, screening with monoclonal antibodies (mAbs) for two CD epitopes revealed limited genetic variation in gliadins immunogenicity among modern wheat varieties (van den Broeck *et al.,* 2010a). Some older hexaploid wheat varieties were identified with somewhat lower mAb response, as well as two lines derived from a heterogeneous tetraploid durum wheat landrace (van den Broeck *et al.,* 2010b). In addition, analysis of nine landraces of farro wheat (*T. turgidum* ssp. *dicoccum*) revealed three landraces that caused negligible proliferation of T-cell lines from CD patients, whereas the other landraces studied showed intermediate to very high responses (Vincentini *et al.,* 2009).

Diploid einkorn wheat (*T. monococcum*, A genome) has good nutritional characteristics and bread making quality. Food products made from the variety 'Monlis' have been shown to be tolerated by CD patients, revealing similar absence of CD toxicity-related symptoms as found for rice (Zanini *et al.,* 2013). New combinations of gluten genes can be made by crossing and selection at the diploid level. Although this is a time-consuming procedure, the increasing incidence of CD in recent decades would justify selection for reduced CD epitopes levels as a novel quality trait for wheat breeding (van den Broeck *et al.,* 2011).

The utility of screening wheat germplasm using epitope-specific mAbs is, however, limited. The specificity of mAbs is not exactly the same as that of the T cell epitopes, as mAbs only recognise peptide sequences of 4-6 amino acids in length, which is substantially shorter than the 9 amino acid T cell epitopes that cause CD (Londono *et al.*, 2013). Therefore, the identification of wheat genotypes with reduced CD toxicity using mAbs should be regarded as a preliminary screen, to be followed by screening of selected lines with more accurate methods.

Salentijn *et al.* (2013) have examined the CD immunogenic potential of tetraploid durum wheat by deep sequencing of α-gliadin transcripts, including the sequences coding for both the N-terminal domain and the repetitive domain, where the CD epitopes occur, from developing grains. They found differences among varieties, but no variety was identified as safe for CD patients. Dubois *et al.* (2016) screened spelt lines by sequencing α-gliadin genes expressed in the developing grains. They saw no clear separation in the occurrence of epitopes between spelt and bread wheat accessions, although some spelt lines expressed more B-genome α-gliadins, which contain fewer epitopes.

A direct approach would be to use quantitative proteomics to identify the gliadins in mature grains, as this would determine the exact amino acid sequence and the quantity of the proteins produced during grain development. However, this is complicated by the diversity of gluten genes in any single wheat variety, and the relative insensitivity of these proteins to the proteolysis step necessary for analysis. A method has recently been developed that enables identification and quantification of specific CD-epitopes in chymotryptic gluten digests (van den Broeck *et al.*, 2015).

Reconstituting hexaploid bread wheat: synthetic hexaploids

Overall, the genetic variation of the D genome in hexaploid bread wheat is much lower than the genetic variation present in the A and B genomes. This is consistent with the theory that the hybridisation that led to hexaploid bread wheat resulted in a strong genetic bottleneck (Reif *et al.*, 2005; Jones *et al.*, 2013). Indeed, several studies have found high levels of genetic diversity among wild *Ae. tauschii* accessions (Wang *et al.*, 2013b; Jones *et al.*, 2013). In order to introduce additional genetic diversity into bread wheat, programmes that recreate the hybridisation process to generate 'synthetic hexaploid wheat' (SHW) have been undertaken at

various institutes around the world, including the International Maize and Wheat Improvement Centre (CIMMYT, Mexico), NIAB (UK), and the Commonwealth Scientific and Industrial Research Organisation (CSIRO, Australia). The process involves hybridising *T. turgidum* spp. *durum* with *Ae. tauschii*, followed by rescue of the triploid embryo, and subsequent chromosome doubling using colchicine to generate hexaploid individuals. Careful selection of diverse *Ae. tauschii* donors can result in SHW accessions that maximise the D genome variation captured, and so increase the potential benefit of SHW in wheat breeding programmes. *Ae. tauschii* accessions which are low in CD-toxic gliadins could for instance be hybridised with two tetraploid durum wheat landraces that have been shown to possess a relatively low mAb response (van den Broeck *et al.,* 2010b), to generate a low gluten immunogenic SHW wheat. These SHWs could then be used in pre-breeding programmes for crossing with commercial varieties, combining screening for low gluten immunogenic progeny with the presence of competitive agronomic characters.

Mutation breeding

Mutation breeding is based on the random induction of genome-wide mutations, followed by selection of plants carrying mutations in the gene(s) of interest; products of mutation breeding can be released on the market without regulation. The type and frequency of mutations obtained in classical mutagenesis programmes depends on the method used (chemical or ionizing radiation), the type and concentration of mutagen, and the duration of the treatment, which are adjusted on a case by case basis. Polyploid crops in general tolerate higher mutation loads than diploid crops (Shu *et al.,* 2011; Oladosu *et al.,* 2016). This is certainly necessary for gliadin mutagenesis in wheat: stacking mutations in multiple gliadin genes into a single plant through crossing is very challenging, due to the number of gliadin genes and their close physical linkage, so this strategy can only be effective if each of the starting lines already contain many mutated gliadin genes.

EMS

Ethyl methane sulfonate (EMS) is a chemical mutation approach that primarily results in transitions of G/C to A/T nucleotides in the DNA. In the context of α-gliadin and γ-gliadin genes, EMS treatment may result in mis-sense mutations within epitopes, which could disrupt binding to the Antigen Presenting Cells,

or non-sense mutations upstream of the epitopes which truncate the protein or prevent its translation. EMS mutations are detected using an approach termed Targeting Local Lesions IN Genomes (TILLING). Detection of induced mutations originally relied on enzyme-mediated cleavage of mismatched nucleotides (McCallum *et al.*, 2000), but is now commonly done by DNA sequencing of either the complete genome or a subset of the genome, either produced with restriction enzymes or using approaches such as exome capture. Acevedo-Garcia *et al.,* (2017) used TILLING to identify missense mutations in each of the three wheat Mildew Locus O (*TaMLO)* homoeologues, which were subsequently stacked to produce plants with enhanced, yet incomplete, resistance towards powdery mildew (*Blumeria graminis f.* sp. *tritici* (Bgt)).

A simple screening system for EMS TILLING populations has been constructed in tetraploid (*T. turgidum* cv. Kronos) and hexaploid (*T. aestivum* cv. Cadenza) wheat. Both TILLING populations have recently been sequenced using exome capture (Krasileva *et al.,* 2017), allowing identification of TILLING mutants in target genes via DNA sequence searches of online databases (www.wheat-tilling.com/). Gluten gene mutants can be identified bioinformatically in these populations, but in practice few mutated gliadin genes have been found. The high copy number and sequence similarity make it difficult to correctly assemble gliadin sequences. As a consequence, assembly of sequence reads from a mutated population runs the risk of combining reads with point mutations from multiple genes into single sequence assemblies. If filtered out, this could lead to an artificially low mutation frequency, but if not filtered, it could result in an over-estimation of the average number of mutations per gene. The use of Pacific Bioscience Circular Consensus Sequencing (PacBio CCS), which produces longer sequence reads, may solve some assembly issues. However, the resulting sequence data would likely still need to be combined with more accurate short read sequencing approaches (e.g. Illumina) to compensate for the relatively high PacBio sequencing error rate.

γ-irradiation

γ-irradiation produces reactive oxygen species (ROS) that cause DNA oxidative damage or DNA single/double strand breaks. Double strand breaks are repaired by the cell's non-homologous end joining (NHEJ) mechanism, which is a relatively error-prone process. Most errors lead to small deletions, although base-pair substitutions, inversions, and small insertions also may be found (Cockram

et al., 2007; Morita *et al.,* 2009). γ-irradiation therefore has the potential to mutate epitopes, remove whole gliadin genes, or even delete multiple gliadin genes.

In barley, γ -irradiated accession Risø 56 (lacking B-hordeins), Risø 1508 (lacking C-hordeins) and R118 (an Ethiopian derived line which lacks D-hordeins) have been intercrossed to produce an Ultra Low Gluten variety (ULG 2.0) with a reduction of hordein (gluten) content below 5 ppm, thus allowing it to be classified as 'gluten-free' (Tanner *et al.*, 2016). Subsequent crossing and selection with commercial varieties resulted in a variety with good malting and brewing characteristics (now marketed as 'Kebari'). Interestingly, in several countries the products made with this variety cannot legally be marketed as gluten-free, solely because it contains barley.

In wheat a strategy to introduce similar resources into commercial breeding programs could be based on screening γ-irradiated hexaploid wheat populations, such as that created in the UK from the commercial cultivar Paragon. We are currently screening these lines for changes in their gluten pattern on acid PAGE protein gels (Cooke 1995), prior to further characterisation of a subset of lines using approaches such as RNA-seq, qPCR, and/or proteomics.

Deletion lines

Wheat deletion lines are available in which whole chromosomes, or parts of chromosomes, have been deleted by crossing monosomic with tetrasomic lines, or by introgression of alien addition or translocation lines into cv. 'Chinese Spring' to induce deletions (Endo and Gill, 1996). A set of these hexaploid wheat cv. 'Chinese Spring' deletion lines has been used to test the effects of individual deletions on the reduction of CD epitopes and on technological properties. A deletion line missing the short arm of chromosome 6D (6DS) in which the 6D α-gliadin locus had been deleted, was found to have strongly decreased mAb responses against Glia-α1 and Glia-α3 epitopes (van den Broeck *et al.,* 2009). The absence of these α-gliadin proteins also led to a significant change in dough mixing properties and dough rheology, as the dough became stiffer and less elastic. In contrast, the deletion of the ω-gliadins, γ-gliadins and LMW-glutenin subunits on chromosome 1DS removed some epitopes, but the technological properties were retained (van den Broeck *et al.,* 2009).

Transgenic approaches to reduce CD toxicity

Until recently, wheat was considered a challenging crop for genetic transformation; particle bombardment was routinely used for delivery of gene constructs, despite its many drawbacks in the quality of plants generated, which included multiple insertions with complex rearrangements. Furthermore, only a limited range of germplasm could be utilised. The development of supervirulent *Agrobacterium tumefaciens* strains (Wu *et al.,* 2008), plus changes to media components and the introduction of new techniques (Wu *et al.,* 2003; Risacher *et al., 2009*) improved efficiency. However, the major breakthrough was made by Ishida *et al.* (2015), using the US soft white wheat, Fielder. This new technology is implemented at NIAB and has been used to both extend the range of UK spring and winter germplasm which can be transformed, and to support a high throughput wheat transformation platform.

Gene silencing using RNA interference

RNA interference (RNAi) is a conserved biological process initially identified as a plant defence mechanism, in which a viral double-stranded RNA (dsRNA) is cleaved, silencing viral gene expression and inhibiting further replication (Ratcliff *et al.,* 1997). It has since been demonstrated that genetic transformation of plants with a construct encoding a dsRNA molecule similar to an endogenous plant single-stranded mRNA, triggers silencing of that specific plant gene (Waterhouse and Helliwell, 2003).

Two research groups have successfully silenced gliadin expression using RNAi. Becker *et al.* (2012) silenced α-gliadins, eliminating 20 different storage proteins from the grain, whereas Gil-Humanes *et al.* (2010) down-regulated gliadins from all groups in bread wheat, with an average reduction of 92.2% in the R5 mAb assay. T cell tests demonstrated a 10-100-fold reduction in DQ2 and DQ8 epitopes from α-gliadins, γ-gliadins and ω-gliadins. Indeed, total gluten extracts of three transgenic wheat lines did not to stimulate T cell clones. Down-regulation of γ-gliadins resulted in an increase of other gluten proteins in the grains (Pistón *et al.,* 2011), but little or no effect was found on dough strength or gluten and starch properties (Pistón *et al.,* 2011; Gil-Humanes *et al.,* 2012; 2014).

RNAi has also been used to suppress the *DEMETER* (*DME*) homoeologues in wheat (Wen *et al.,* 2012). The *DME* gene encodes a 5-methylcytosine DNA

glycosidase that demethylates the promotor regions of gliadins and LMW-glutenins in the wheat endosperm. This demethylation is essential for activation of the genes during endosperm development. *DME* RNAi transgenic plants showed a high degree of *DME* transcript suppression, resulting in >75% reduction in the amount of immunogenic prolamins.

Providing their agronomic properties and yields are acceptable, such RNAi wheat lines may become candidates for the production of wheat-based products for 'gluten-free' or 'low-in-gluten' diets. However, it should be noted that these strategies require a stable integration of the RNAi cassette into the genome, thus resulting in a genetically modified (GM) plant. It is therefore subject to relevant national and international GM regulations, with the accompanying long, expensive, and uncertain procedure for market approval.

Targeted mutagenesis to generate small indels and to delete genes

In contrast to mutation breeding, genome (gene) editing approaches aim to induce mutations only within specific targets. These methods include Meganucleases, Zincfinger nucleases, TALENs and CRISPR-Cas9. The regulatory status of plants and plant products derived from gene editing is currently unclear in the European Union (Sprink *et al.,* 2016b). The current method of choice is CRISPR-Cas9 due to its ease of use (e.g., Schaart *et al.,* 2016). The CRISPR-Cas9 system was originally described as a bacterial defence mechanism against viral infections (Barrangou *et al.,* 2007) and has since been customised to induce mutations and deletions at specific locations within a target gene in animals and plants (Zhou *et al.,* 2014; Schaart *et al.,* 2016). Mutations can be induced by introducing genomic RNA (gRNA) complementary to the gene of interest, plus an expression cassette to express the Cas9 endonuclease protein. The 20-mer gRNA directs the Cas9 nuclease to the target site within the gene of interest, where the nuclease generates a double strand DNA break, three bases upstream of a specific protospacer adjacent motif (PAM). In plants, repair by NHEJ can result in insertion/deletions (indels) and other mutations that can result in beneficial phenotypes (Belhaj *et al.,* 2013; van de Wiel *et al.,* 2017).

CRISPR-Cas9 and other site-directed nucleases have been successfully used to target homeologous genes in diploid and polyploid crops (Brooks *et al.,* 2014; Lawrenson *et al.,* 2015). In potato, the four copies of *Granule-bound starch*

synthase (*GBSS*) have been simultaneously modified to alter starch composition in tetraploid potato (Andersson *et al.,* 2017). In bread wheat, the TALEN system was used to simultaneously mutate the three homoeologs of a Ta*MLO* gene (Wang *et al.,* 2014). Using a Cas9/guide ribonucleoprotein complex in plant callus, protoplasts and immature embryos, homoeologous genes were successfully edited in hexaploid and durum wheat (Zhang *et al.,* 2016; Liang *et al.,* 2017). In this system no foreign DNA is added to the plant cell, so plants developed in this manner might not be regulated as GM, as was recently propositioned by the U.S. Department of Agriculture's (USDA) Animal and Plant Health Inspection Service (APHIS) in its proposed revision of regulations (USDA, 2017).

With these examples, it should also be feasible to modify and/or delete gliadin genes from hexaploid wheat. Such experiments are ongoing at NIAB and elsewhere. In theory, several types of mutations could be generated in the gliadin genes. Small indels, within or upstream of the epitope region, can generate premature stop codons resulting in truncated proteins that lack CD epitopes or within modified epitopes which are no longer immunogenic. Moreover, genes of the α-, γ- and ω- gliadin families are each grouped at single loci on the short arms of the group 6 and 1 chromosomes, respectively. The α-gliadin *Gli-2* loci contains α-gliadin interspersed with retrotransposons (Gu *et al.,* 2004). The γ-gliadin locus also includes other prolamin and non-prolamin genes (Gao *et al.,* 2007). Cas9-mediated double strand breaks generated simultaneously in at least two gliadin genes on the same chromosome may result in a deletion of the intervening genes. Similarly, multiplexing gRNAs to target different sites within one gliadin gene could, together with a high mutation rate, delete only the immunogenic epitopes from the gliadin protein.

Following wheat transformation with the gliadin gene editing construct, T1 seed could be analysed by Acid-PAGE separation of storage proteins, allowing rapid identification of plants with modified gliadin profiles. Lines of interest could then be further analysed using methods such as deep sequencing of the genome, digital drop real-time PCR to determine gliadin copy number, and RNA-seq of developing grains to determine gene expression profiles. In *Chapter 4*, we developed a process termed 'Gluten genes Enrichment and Sequencing' (GlutEnSeq), a protocol based on REnSeq methods (Jupe *et al.,* 2013). This approach uses exome capture to enrich for gluten gene DNA from lines of interest for targeted re-sequencing,

preferably using a long read technology such as PacBio. Proteomics methods such as LC-QTOF-MSMS could be used to generate information on the composition of the gliadins produced in mutant lines, to quantify epitope levels, and to determine the occurrence of amino acid changes in epitope-harbouring peptide fragments.

Prospects for the low gluten immunogenicity strategies

The food processing and wheat breeding strategies described represent two major routes for the development of cereal products with low gluten immunogenicity. Gluten-free products, made without wheat, barley or rye, are now increasingly found on grocery store shelves. However, these products typically require the inclusion of numerous additives, resulting in products that are often less healthy than gluten-based equivalents. To overcome this issue, other strategies have to be deployed.

Malting may be one of the most applicable methods for large-scale preparation of products with low gluten immunogenicity in the food processing industry, and it can even enhance sourdough fermentation when the malting process is not fully stopped. Similarly, sourdough is a traditional method used to increase loaf volume for bread baking which is possible to implement at industrial scale. Sourdough can also be combined with malting for further reduction of gluten immunogenicity. However, in addition to being more time consuming, sourdough has a taste that may not be widely accepted by consumers.

While the separation of gliadins from glutenins, followed by substitution of gliadins for non-immunogenic oat avenins, is possible at a laboratory scale, large-scale implementation could be challenging and would require considerable infrastructure. From another perspective, the labelling and reduction in the use of vital gluten as an additive, used as source of protein to improve food structure, is a simple approach that would improve the healthiness of gluten-sensitive patients. Unlike GM breeding techniques, most of the candidate food processing techniques for making low gluten-immunogenic products are already implemented, and do not arouse ethical and political issues.

From a wheat breeding perspective, very few polyploid wheat accessions show low gliadin immunogenicity, and those that do lack the agronomic characteristics required for commercial varieties. Introgression of the multiple low immunogenic gliadin loci involved into commercial varieties would be a complex undertaking. Since the expected result would only be a low-immunogenic line, not a CD-safe wheat, it is not anticipated to be a likely near-term target in wheat breeding.

Mutation breeding represents an attractive option to create wheat with low gluten immunogenicity. EMS TILLING on members of large gene families is challenging. Again multiple mutations are necessary, but if these occur in multiple plants then a programme of backcrossing followed by recurrent self-pollination would be required to remove background mutations, while retaining as many homozygous gliadin mutations as possible. However, as the gliadins are closely linked on chromosome 1 and 6, the recombination events required to combine all mutations in a single genetic background might be very rare. Induced mutations using γ-irradiation would be a more tractable approach, since multiple gliadin genes may be removed in single deletion events. This approach has been successfully undertaken in the diploid crop barley, where inter-crossing γ-irradiated lines, each missing a different prolamin gene family, generated an Ultra Low Gluten line that was successfully crossed with commercial varieties (Tanner *et al.,* 2016), leading to a commercial variety with less than 20 ppm gluten. In theory, this line could be used for breeding more low-gluten barley varieties, although it would represent a major bottleneck for genetic diversity around these deletions. The variety also comes with a yield penalty. It is not expected that this line will be used by other breeders to generate low-gluten barley in the short term (Tanner, personal communication). However, the case for wheat may be stronger, as CD patients encounter wheat gluten in many food products, but can easily avoid barley by drinking only gluten-free beer.

The use of RNAi in wheat has been shown to be very efficient in decreasing the amount of gliadins expressed in the grain, without greatly altering dough properties. Nonetheless, the RNAi construct must be retained in the plant genome, resulting in a GM plant. Multiple factors restrict the commercial feasibility of using GM crops, including the costs of the regulatory process in those countries that allow GM crops (Laursen, 2016). However, this may not be the case with CRISPR-Cas9 technology, as after having generated inheritable mutations in target genes,

the T-DNA can be crossed out. Whether the resulting products, which do not contain foreign DNA, will be regulated as GM plants, is under debate (Sprink *et al.,* 2016b). In some countries, process-based regulation may define them as GM, whereas other countries with a more product-based approach may consider them to be identical to plants produced by conventional or mutation breeding. The European Food Safety Authority (EFSA) GMO panel (2015) considered the use of site-directed nucleases for inducing small deletions or insertions at a precisely defined location in the genome (*SDN-1*) as a form of mutagenesis. Acceptance by consumers may also be an issue (Ishii and Araki, 2016) but acceptance of products that are tailored to the needs of specific groups of consumers (e.g., hypoallergenic fruits, nuts etc.) is probably much higher than that of products tailored to the needs of producers or the general public (Schenk *et al.,* 2008; 2011; Kronberger *et al.,* 2014).

In the future, low gluten immunogenicity may become one of many traits included in commercial breeding programs. Achieving low gluten-immunogenic wheat, and ultimately food products, will require a great deal of effort and is likely to require changes in industrial grain and food processing processes. Despite these challenges, the importance of wheat (and other gluten-containing immunogenic grains) in the human diet, along with the increasing worldwide incidence of immunogenic diseases such as CD, means that it is imperative that these challenges are now addressed.

Acknowledgements

Writing this study was partially funded by the EFRO project 'Nieuwe detectiemethoden voor coeliakie en coeliakie-inducerende gluten in voeding' (2011-018974), FP7-PEOPLE-2013-ITN, EID Max-imising the potential of CROP researchers (Max-CROP), grant no. 607178, and the Global One Health project. It does not necessarily reflect the views of the European Commission and in no way anticipates the Commission's future policy in this area.

Author contribution statement

MS and LG initiated the development of the manuscript. AJ wrote the first draft. All the other co-author contributed to the review and improvement of the manuscript. AJ, MS and LG edited the final version of the paper. All authors approved the final version.

Chapter 2

Outlook for coeliac disease patients: Towards hexaploid wheat with hypoimmunogenic gluten by gene editing of α- and γ-gliadin gene families

Aurélie Jouanin[1,2], Emma J. Wallington[2], Fiona J. Leigh [2], Lesley A. Boyd [2], Jan G. Schaart[1], Richard G.F. Visser[1], Marinus J.M. Smulders[1],

[1]*Wageningen University and Research, Plant Breeding, Wageningen, The Netherlands*
[2]*The John Bingham Laboratory, NIAB, Huntingdon Road, Cambridge, United Kingdom*

Submitted

Abstract

Wheat grains contain gluten proteins, which harbour immunogenic epitopes that trigger Coeliac disease in 1-2% of the human population. Wheat varieties or accessions containing only safe gluten have not been identified and conventional breeding alone struggles to achieve such a goal, as the epitopes occur in gluten proteins encoded by five multigene families, these genes are partly located in tandem arrays, and bread wheat is allohexaploid. Gluten immunogenicity can be reduced by modification or deletion of epitopes. Mutagenesis technologies provide a route to obtain bread wheat containing gluten proteins with fewer immunogenic epitopes. However, CRISPR/Cas9 has never been used to edit multigene families in a polyploid crop. In this study, we analysed the genetic diversity of over 600 α- and γ-gliadin gene sequences to design six sgRNA sequences on relatively conserved domains that we identified near coeliac disease epitopes. They were combined in four CRISPR/Cas9 constructs to target the α- or γ-gliadins, or both simultaneously, in the hexaploid bread wheat cultivar Fielder. We compared the results with those obtained with random mutagenesis in cultivar Paragon by γ-irradiation. For this, Acid-PAGE was used to identify T1 grains with altered gliadin protein profiles compared to the wild-type endosperm. We first optimised the interpretation of Acid-PAGE gels using Chinese Spring deletion lines. We then analysed the changes generated in 360 Paragon γ-irradiated lines and in 117 Fielder CRISPR/Cas9 lines. Similar gliadin profile alterations, with missing protein bands, could be observed in grains produced by both methods. The results demonstrate the feasibility and efficacy of using CRISPR/Cas9 to simultaneously edit multiple genes in the large α- and γ-gliadin gene families in polyploid bread wheat. Additional methods, generating genomics and proteomics data, will be used to determine the exact nature of the mutations generated with both methods.

Key words: *Gene editing, CRISPR/Cas9, Mutation breeding, γ-irradiation, Wheat, Polyploid, Gene family, Gluten, α-gliadin, γ-gliadin, Epitope, Coeliac disease.*

Introduction

Bread wheat (*Triticum aestivum*) is a staple crop worldwide and an important source of energy, fibre and protein. The largest protein fraction in wheat grains is gluten, a polymer of glutenins and gliadins. Glutenins provide elastic properties, which are essential for bread dough quality, while gliadins provide viscosity with less impact on dough quality (Shewry *et al.*, 2009).

Wheat can cause allergies and/or intolerances after consumption in susceptible individuals (Fasano, 2006; Gilissen *et al.*, 2014). The most common disorder is an autoimmune reaction triggered by gluten immunogenic epitopes, known as coeliac disease (CD), which occurs in 1-2% of the human population (Dicke *et al.*, 1953; Mäki *et al.*, 2003; Rewers, 2005; Catassi and Fasano, 2014; Vriezinga *et al.*, 2015). Currently, the only treatment is a gluten-free (GF) diet, excluding all wheat, barley and rye. On a daily basis, the GF diet is difficult to follow as wheat gluten is added to a large range of food products. In addition GF products are less healthy, with low levels of proteins and nutrients plus high levels of salt and, many additives are needed to simulate the unique rheological properties of wheat gluten (Caponio *et al.*, 2008; Capriles and Arêas, 2014; Belz, 2016; Horstmann *et al.*, 2016). Breeding for wheat varieties that are free of, or have reduced levels of immunogenic epitopes, is therefore a potential solution for healthier products which are safe to consume by CD patients (Jouanin *et al.*, 2018; Ribeiro *et al.*, 2018).

Bread wheat is hexaploid, so each locus is present on the homoeologous chromosome pairs of the three different sub-genomes, A, B and D. Immunogenic epitopes occur in α-, γ- and ω- gliadins and to a lesser extent in low molecular weight (LMW) glutenins while high molecular weight (HMW) glutenins are mostly safe for CD patients and constitute the main gluten family responsible for bread dough quality (Payne *et al.*, 1984; Shewry *et al.*, 2009). Gliadin proteins are encoded by large gene families. They are clustered together as repetitive sequences in characterised loci, do not contain introns but do include high numbers of pseudogenes (90% in case of α-gliadins (Van Herpen *et al.*, 2006)). In hexaploid wheat, the α-gliadin genes are grouped in the *Gli-2* locus on the short arm of each of the six chromosome 6 and number between 60 and 150 copies (Anderson *et al.*, 1997; Ozuna *et al.*, 2015). The short arm of each of the six chromosomes 1 contains

the *Gli-1* locus with around 40 γ-gliadins and the *Gli-3* locus, with approximately 16 ω-gliadins copies (Sabelli and Shewry, 1991), in hexaploid genome. Variants of the immunogenic epitopes that are not recognised by T cell receptors, also exist in hexaploid bread wheat (Mitea *et al*., 2010), but in all genotypes they are found in combination with other, highly immunogenic epitope variants (Van Herpen *et al*., 2006; van den Broeck et *al.,* 2010a; 2010b; Salentijn *et al*., 2013). There is a trend between the level of immunogenicity of gliadin epitopes and the sub-genome by which they are coded (Salentijn *et al*., 2012); gliadins from genome B tend to be less immunogenic while gliadins from genome D trigger the most CD. Given this level of complexity, it is not surprising that conventional breeding alone has not yet succeeded in generating "gluten-safe" wheat, containing only gliadins with non-immunogenic epitopes or even with no gliadin at all (Smulders *et al*., 2015; Sánchez-León *et al*., 2018; Jouanin *et al*., 2018).

Alternatively, RNAi targeting all three gliadin families has successfully decreased 97% of gliadin expression in bread wheat grain and the gluten extract did not stimulate CD patient T cells while dough quality was barely affected (Gil-Humanes *et al*., 2010; 2014). Similarly, Becker *et al.* (2012) decreased the expression of 20 α-gliadins yet increased expression of other storage proteins. Wen *et al.* (2012) reduced the expression of DEMETER, preventing DNA-methylation changes and thus repressing glutenin and gliadin gene expression in the endosperm. As the transgenic RNAi construct remains in the wheat genome to silence the genes, these plants are subject to the hurdles of GM regulation. Another approach towards wheat with hypoimmunogenic gluten is mutation breeding. This method which is exempted from GM regulation, has recently been applied to develop "ultra-low gluten" barley (Tanner *et al.,* 2016), and is being used to produce gluten-free beer in Germany. Developing wheat with hypoimmunogenic gluten using a similar approach is theoretically possible although it constitutes a greater challenge (Ribeiro *et al*., 2018). Wheat γ-irradiated lines have to be identified that miss large regions on the short arm of chromosomes 1 or 6, yet have lost gliadin or LMW-glutenin genes in one of the three homoeologous genomes. Plants with these 12 different events have to be self-fertilised to become homozygous for the deletion and then inter-crossed to obtain hypoimmunogenic gluten while maintaining bread wheat unique baking quality conferred by HMW-glutenin loci. Gene editing using CRISPR/Cas9 represents an alternative approach that enables modification or deletion of the immunogenic gliadin genes in order to generate non-immunogenic gluten while

retaining the viscosity provided by gliadins required for good bread dough quality. CRISPR/Cas9 has successfully been used in polyploid wheat to induce mutations in all six alleles of single copy genes. Zhang *et al.* (2016) targeted up to four genes in bread wheat, using various gene editing approaches. Gil-Humanes *et al.* (2017) targeted two single-copy genes simultaneously for gene targeting (i.e. with a DNA repair template) without integrating a Cas9 construct in the genome. CRISPR/Cas9 has been used once before to modify gluten, when Sánchez-León *et al.* (2018) targeted multiple α-gliadin genes.

In this study, we have transformed hexaploid bread wheat cultivar Fielder with constructs containing Cas9 and combinations of single guide RNAs (sgRNA) targeting specific sites upstream of, or within, the CD epitopes in the α-gliadin as well as the γ-gliadin gene family. These constructs would potentially create nucleotide changes and small indels in the epitope regions, or deletion of the epitope regions, either in, or out of frame. In the ideal scenario, small in-frame mutations would be generated, producing non-immunogenic gliadins which retained the desired rheological properties. Small out of frame mutations upstream the epitope region, would generate a knock out of the gliadin that should prevent the immunogenicity but may slightly decrease bread dough quality. However, the transcription of the truncated protein could prevent the often reported compensation by other gliadin genes (Pistón *et al.*, 2011; Galili *et al.*, 1986). Since gliadin genes are clustered as repeat sequences, simultaneously targeting non-consecutive genes could generate the deletion of intervening gliadin gene copies. This possibility, similar to but more precise than γ-irradiation events, presents the advantage of fully supressing the epitopes, but often triggers compensation of expression by the remaining gliadin genes. This attempt to use CRISPR/Cas9 to mutate gliadin families in polyploid wheat was a pilot experiment. The aim was to mutate as many gliadin genes as possible, in any manner, to test whether CRISPR/Cas9 could be an appropriate method to simultaneously modify sufficient gene copies in order to change qualitatively and/or quantitatively the gliadin content in wheat grains. An additional aim was to study the frequency of the different mutation types and to identify the most relevant methods to use in order to screen for wheat grain with mutated gliadins.

Due to the high sequence complexity, variability and number of targeted gliadin genes, a simple PCR on wild type plants would already give several bands. A

CAPS assay to reveal mutations by identifying restriction site loss, would not be possible since not all sequences would carry the site. In addition, regular gene cloning and sequencing would not be an efficient pre-screening method, given the uncertainly that any editing occurred in targeted gliadin genes. Indeed, the cloning of the 50 to 100 single different gene copies present in a plant cannot be guaranteed and would require deep sequencing for every single Fielder-CRISPR plant tested. In addition, as many gliadin genes are pseudogenes (van Herpen *et al.*, 2006), the actual influence of potential mutations on the phenotype could not quickly be assessed. Moreover, using this approach on cDNA would have been inappropriate at this stage since mRNA extraction would have to be done on endosperm from immature grain, without damaging the embryo required to grow the potential mutant progeny. Advanced proteomics techniques for gliadin identification are currently under development at WUR. However, being both time consuming and expensive, they would be more appropriately used as a final epitope characterisation method, rather than pre-screening one. Consequently, Acid-PAGE was employed as a first screening method to identify gliadin protein profile changes simultaneously in targeted and untargeted gliadin gene families. This method has the advantage of being cheap and revealing mutations in expressed genes with an impact on the proteome, necessary to develop wheat lines with hypoimmunogenic gluten. However, this method reveals mainly full gene copy deletions, non-sense or off-frame mutations implying absence or truncation of gene translation but not mis-sense or small in frame mutations.

We compared the changes that occurred in our Fielder-CRISPR gene-edited lines with those we found in selected lines from a γ-irradiated mutant collection in variety Paragon (Shaw *et al.*, 2013), and in selected chromosome arm deletion lines in the reference wheat cultivar Chinse Spring (Sears, 1966; Endo, 1988).

Materials and methods

Gliadin sequence alignment

Over 438 α-gliadin gene sequences from 30 wheat accessions and 187 γ-gliadin gene sequences from hexaploid wheat were downloaded from NCBI database in September 2014 and July 2015 respectively, translated into amino acid sequences, manually aligned using Mega_6 and clustered based on their combination of

known CD epitopes (Sollid *et al*., 2012). To deduce the sub-genome of origin of different epitopes, which is linked with the level of immunogenicity, sequences from diploid bread wheat ancestors *Triticum monococcum* and *Triticum urartu* (A genome), *Aegilops speltoides* (for the B genome), *Aegilops tauschii* (D genome) and tetraploid durum wheat *Triticum turgidum* (genome AB) were added that were present in NCBI or generated at WUR-PBR (Schaart, personal communication). To complete the alignments and facilitate sgRNA design, primers F: 5'-ATGAARACMTTTCYCATC-3' (Salentijn *et al*., 2013) and R: 5'-YAGTTRGTACCRAAGATGM-3' were used to clone and sequence seven intact α–gliadin genes from Fielder, the spring bread wheat variety used for transformation. These α-gliadin gene sequences were similar to the one already present in the databases. This incorporation increased the total number of sequences used to 1273 α-gliadins and 442 γ-gliadins.

sgRNA protospacer design

Six sgRNA protospacers were designed on conserved regions that were identified based on the sequence alignments and were present in the sequences of the hexaploid wheat cultivar Fielder. Of these six sgRNAs, three targeted α-gliadins and the other three targeted γ-gliadins (Fig. 2). In each case, one targeted downstream of the signal peptide, the other two targeted in or near epitopes. These sgRNA target the complementary DNA strand and should therefore be reverse-complemented to be found in the Genbank sequences.

sgRNA_α87: 5'-GATTTTGTGGCTGCAATTG-3' targets α-gliadins downstream the signal peptide, P87
sgRNA_α213: 5'-ATGGTTGTTGTGATGGAAA-3' targets α-gliadins upstream the epitope region, P213
sgRNA_α324: 5'-GTTGTGGTCGAAATGGTTG-3' targets α-gliadins downstream the epitope region, P324
sgRNA_γ86: 5'-TTGTTGTGGCCATTGTACT-3' targets γ-gliadins downstream the signal peptide, P86
sgRNA_γ272: 5'-AATGGTTGTTGTGGTTGCTG-3' targets γ-gliadins within the epitope region, P272
sgRNA_γ603: 5'-TGCTGGGGGAATGATTGTTG-3' targets γ-gliadins downstream the epitope region, P603

These sgRNA protospacers where tested *in silico* for the absence of off-targets using BLAST in the Ensembl! plant wheat database, for absence of secondary structures using RNAfold web server and for absence of cross-dimers between multiplexed sgRNAs using ThermoFisher Scientific Primer Analyzer.

Each sgRNA, including a wheat-optimised U6 promoter, gliadin-specific protospacer and sgRNA scaffold sequence (Shan *et al.*, 2013), flanked by multiple unique restriction sites, was individually synthesised by GenScript.

CRISPR/Cas9-sgRNA constructs

Type II-A *Streptococcus pyogenes* 2NLS-Cas9 gene, codon-optimized for rice and wheat expression, was cloned from the pJIT163-2NLS-Cas9 plasmid (Shan *et al.*, 2013). Its ribosome binding site (RBS) "CACC" was mutated into "CCACC", using site directed mutation PCR approach, for better expression in wheat. The optimised 2NLS-Cas9 gene plus CaMV terminator sequence was transferred into binary vector pSC4Act-R1R2-SCV (Perochon *et al.* 2015), containing a rice actin promoter to drive the expression of the Cas9 gene. Each sgRNA, was combined using multiple unique restriction sites and then integrated into the final binary plasmids (Fig. 1) (named pAJ2_ followed by sgRNA names).

Figure 1. CRISPR/Cas9 T-DNA construct 2α _sgRNA.
Construct 2α_sgRNA that contain sgRNA_ α213 and sgRNA_ α324. The four T-DNA constructs generated are similar, only the number and nature of sgRNA integrated are different.

Four T-DNA constructs were produced by combining different sgRNAs, using multiple unique restriction sites:

Construct 1α_sgRNA	=	sgRNA_α87
Construct 2α_sgRNA	=	sgRNA_α213 + sgRNA_α324
Construct 3γ_sgRNA	=	sgRNA_γ86 + sgRNA_γ272 + sgRNA_γ603
Construct 2α3γ_sgRNA	=	sgRNA_α213 + sgRNA_α324 + sgRNA_γ86 + sgRNA_γ272 + sgRNA_γ603 (i.e., it combines Construct 2α_sgRNA and Construct 3γ_sgRNA).

Stable transformation and generation of Fielder-CRISPR plants and derived grains

The CRISPR/Cas9-sgRNA binary vectors were transferred to *A. tumefaciens* and used to transform immature embryos of Fielder, (Segovia *et al.,* 2014). Selection of transformed tissues was based on the presence of the *nptII* gene conferring resistance to the G418 antibiotic. Regenerated T0 plantlets were transferred to soil and tested for T-DNA copy number using an *nptII*-based qPCR assay (Milner *et al.* 2018), for presence of Cas9 gene and all sgRNA using PCR, and whether Cas9 was expressed using RT-PCR. Plantlets positive in all these tests were grown in climate controlled growth chambers, bagged during anthesis, and T1 grains harvested.

The individual T1 grains were first cut transversally, alongside the embryo, so that this could be germinated later. The resulting endosperm section was then cut in half longitudinally. One half was used for Acid-PAGE. Selected T1 embryos were germinated in Petri dishes containing filter paper soaked with water. Seedlings were potted in compost after sufficient root development and transferred to growth chambers until grain set. Ears containing T2 grains were harvested individually.

Other plant materials

Chinese Spring (CS) wild type (WT) and selected CS Kansas deletion lines (Endo, 1988; Endo and Gill, 1996) that lacked parts of chromosome 1 or 6, were obtained from Kansas State University (Table 1). In addition, CS nullisomic-tetrasomic lines (Sears, 1966) were obtained from the Germplasm Resources Unit (GRU JIC, Norwich, UK) (Table 1). The CS lines were used to set up and optimise the Acid-PAGE separation.

The Paragon γ-irradiated population (Shaw *et al*., 2013), obtained from JIC (Norwich, UK) based on hexaploid spring wheat cultivar Paragon, was also screened. Mature grains from a subset of 360 lines self-pollinated for 4 generations (M4) were analysed on Acid-PAGE. The lines screened were numbered from P3-47 to P6-79 and from P10-19 to P10-96 (Table 1).

Table 1. List of Triticum aestivum lines used to set up the screening method and subsequent comparison.

Triticum aestivum	Lines	Description
Cultivar 'Chinese Spring' (CS)	**WT**	**Control**
	1AS-1 1AS-3 1BS-9 1BS-10 1DS-1 1DS-5	**Kansas Deletion lines** **missing part of Chr 1 short arm** that contain *Gli-1* and/or *Gli-3* loci where γ- and/or ω-gliadin genes are located
	6AS-1 6BS-4/5BS-2 6DS-4 6DS-4/1BS-19	**Kansas Deletion lines** **missing part of Chr 6 short arm** that contain *Gli-2* loci where α-gliadin genes are located
	N1AT1D = 1DDB N1AT1B = 1BBD N1DT1A = 1AAB N1DT1B = 1BBA N1BT1D = 1DDA N1BT1A = 1AAD	**Nullisomic-tetrasomic lines** **with one homoeologous pair of Chr 1** that contain *Gli-1* and/or *Gli-3* loci where γ- and/or ω-gliadin genes are located substituted by another homoeologous pair of Chr 1
	N6AT6D = 6DDB N6AT6B = 6BBD N6DT6A = 6AAB N6DT6B = 6BBA N6BT6D = 6DDA	**Nullisomic-tetrasomic lines** **with one homoeologous pair of Chr 6** that contain *Gli-2* loci where α-gliadin genes are located substituted by another homoeologous pair of Chr 6
Cultivar 'Paragon'	**WT**	**Control**
	Line P3-47 to P6-79 and Line P10-19 to P10-96	**360 γ-irradiated lines** M4 generation

Acid-Polyacrylamide Gel Electrophoresis (Acid-PAGE)

For each Fielder-CRISPR plant, a minimum of 8 T1 grains were screened individually, alongside Fielder. The grain samples were loaded in duplicate. As Fielder exhibit some heterogeneity in its storage protein profile on Acid-PAGE, grains from multiple Fielder plants were loaded alongside on the gel. For T2, four grain samples from four different ears were loaded in duplicates. A similar procedure was applied for screening of CS and Paragon, with the exception that only 2 grains were tested as they are homogeneous for the mutations due to several generations of self-pollination.

Each half endosperm sample was crushed into fine powder, and the gliadin fraction extracted overnight at 4°C, in 150 µl of 25% dichloroethanol solution containing 0.05% Pyronin Y. Duplicate 30µl samples were loaded in wide slots on 13% polyacrylamide gels (acrylamide-bis 19:1) and run at 180V, for 4 hours at room temperature (adapted from Cooke, 1987). Gels were stained overnight in 10:1 solution of 15% trichloroacetic acid (TCA): industrial methylated spirits (IMS) containing 10g/L Coomassie Brilliant Blue G250, then destained overnight in water.

Results

Sequence alignment and sgRNA design

Genomic sequences of 1273 α-gliadin and 442 γ-gliadin genes from several species and varieties were translated into protein sequences and manually aligned to infer position of epitope and most conserved regions, and to superimpose this information on the DNA sequence. The alignment was optimised by hand, especially around the repetitive parts of α- and γ-gliadin as presented in Supplementary Figure 1 and 2, respectively. These Supplementary Figures are useful to quickly visualise the species and homoeologous genome to which they are associated and the sub-grouping based on the α- and γ-gliadin sequence pattern and the position or absence of the different overlapping CD canonical epitopes.

In α-gliadins, five protein subgroups exist based on sequence patterns, two associated to genome A, two to genome B, and one that presents a unique amino-acid variant specific for genome A or D. Some CD epitopes were present or absent in specific protein subgroups. As a consequence, some CD epitopes are specific for one homoeologous genome (Suppl. Fig.1 and 2). Pseudogenes were identified having similar sequences to intact genes but harbouring early stop codons at various places (Anderson, 1991).

In γ-gliadins, six protein subgroups exist based on sequence patterns, two associated to genome A, one to genome B and the three others are at least not associated to genome A (Suppl. Fig. 1 and 2). Some CD epitopes were only present in some specific protein subgroups. However, this does not lead to a good correlation between epitopes and a homoeologous genome, most epitopes occur in γ-gliadins from all three genomes. (Suppl. Fig.1 and 2). Six groups of pseudogenes were observed that did not simply contain an early stop codon but also had divergent sequences compared to the 6 groups of full length γ-gliadins. This is consistent with the hypothesis of Goryunova *et al.* (2012) that the various γ-gliadin groups largely predate the evolution of the genomes within the genus Triticum/Aegilops. Consequently, sgRNA were designed focusing on full length genes while considering pseudogenes as much as possible.

Six sgRNAs were designed, three of which target α-gliadins (Fig. 2a) and the other three targeting γ-gliadins (Fig. 2b). In each gene family one target site was placed

upstream in the gene, soon after the signal peptide, with the aim of disrupting the open reading frame and the two others were near or in the innate or DQ2.5 coeliac disease epitope regions in order to modify or remove the epitope region.

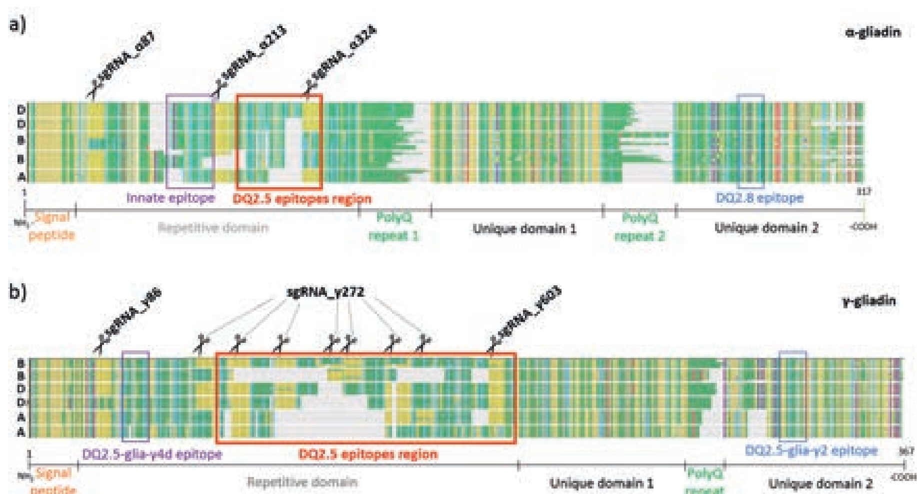


Figure 2. Alignment of α- and γ-gliadin protein sequences with sgRNA position and potential sites of gene editing.

A representation of the protein sequence alignments of α-gliadins (a) and γ-gliadins (b) based on a total of 438 and 187 DNA sequences, respectively. The variation in the sequences form patterns which are grouped here and associated to the genome in which they are mostly found (A, B, D on the left side of each group), based on comparison of hexaploid wheat sequences with sequences from diploid relatives. The different gliadin protein domains are indicated and the position of the CD immunogenic epitopes are boxed. The DQ2.5 epitopes box includes the DQ2.5-α1, - α2 and - α3 epitopes (Supplementary Fig.1 and 2). The sgRNAs targeted motifs are highlighted in yellow and the potential gene editing sites are marked with scissors. SgRNA_γ272 may cut multiple times, depending on the number of repetitions of the most abundant γ-gliadin CD epitope, DQ2.5-glia-γ4c, which it targets. Details on the alignment, sequence patterns and CD epitopes can be found in Supplementary Fig. 1 and 2.

Our alignment shows that as a result of the existing sequence differences, largely within the DQ2.5 epitope region (Suppl. Fig.1 and 2), one sgRNA cannot theoretically target all the intended sites present in the hexaploid bread wheat genome. For instance, sgRNA_α324 may not target the two groups of genes associated with genome B while sgRNA_α87 may only target some genes related to one of the two α-gliadin groups of the B genome. sgRNA_α213, however, should recognise its target motif in all 5 different α-gliadins DQ2.5 groups, regardless of the genome they are associated to (Fig 2a). For the γ-gliadins, sgRNA_γ86 may not target one of the DQ2.5 groups typical for hexaploid wheat (possibly genome D) and sgRNA_γ603 may not target a group of genes associated to genome A.

sgRNA_γ272 should target most of the γ-gliadin DQ2.5 sequences and enable cutting them between 1 and 6 times depending on the number of repetitions of the targeted motif, corresponding to the most abundant γ-gliadin CD epitope DQ2.5-glia-γ4c, throughout the epitope region (Fig 2b; Suppl. Fig.2).

Multiplexing sgRNAs in clustered gene families can generate many different types of mutation (Fig. 3). Indeed, a simultaneous Cas9 cut within two non-consecutive gliadin genes can delete the intervening genes. In addition, a simultaneous Cas9 cut upstream of and downstream of the epitope region, may enable a minimal deletion of only the potentially immunogenic region. Finally, simple single cuts may create small indels or base substitutions at that site.

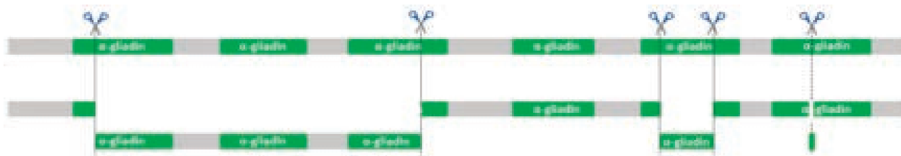


Figure 3. Representation of one α-gliadin Gli-2 locus and different mutation types potentially induced by CRISPR/Cas9.
α-gliadin Gli-2 loci representation with genes clustered and different types of mutations that can be induced by sgRNA. Simultaneous cuts in non-consecutive genes can delete the intervening genes. Similarly, two simultaneous cuts flanking the epitope can delete only this region, whilst simple small indels or base substitutions can also occur.

Generation of CRISPR/Cas9_sgRNA constructs and transgenic Fielder wheat plants

Four different CRISPR/Cas9_sgRNA binary constructs were created containing Cas9 plus different combinations of the sgRNAs (Table 2), transferred to *A. tumefaciens* and used in stable transformation experiments with immature embryos from cultivar Fielder. The number of T0 plants regenerated and expressing the Cas9 (as determined by RT-PCR on leaf samples) are presented in Table 2, together with the number of plantlets harbouring 1or 2 copies of the *nptII* gene, part of the T-DNA containing the CRISPR/Cas9 construct.

Table 2. Summary of T0 Fielder-CRISPR regenerated plants expressing Cas9 and their transgene copy number

	T0 Fielder wheat plants		
	Regenerated	**Expressing Cas9**	**1 or 2 transgene copies**
Construct α1_sgRNA sgRNA_α87	40	38 (95%)	21 (55%)
Construct α2_sgRNA sgRNA_α213 + sgRNA_α324	32	12 (38%)	4 (33%)
Construct γ3_sgRNA sgRNA_γ86 + sgRNA_γ272 + sgRNA_γ603	36	35 (97%)	26 (74%)
Construct α2γ3_sgRNA sgRNA_α213 + sgRNA_α324 + sgRNA_γ86 + sgRNA_γ272 + sgRNA_γ603	42	32* (76%)	12 (38%)
TOTAL	**150**	**117 (78%)**	**63 (54%)**

For each construct, between 32 and 42 plants were regenerated. The * indicates that 34 plants expressed Cas9 but 2 were discarded since they did not contain all 5 sgRNAs. On average, 78% of plants expressed the Cas9, among which 54% had 1 or 2 *npTII* copies.

A total of 150 T0 Fielder-CRISPR plantlets were regenerated, using the four constructs, among which 31 plants did not express the Cas9, had prematurely aborting grains or died. Among the plants carrying the T-DNA, only two plants transformed with the α2γ3_sgRNA construct did not carry all the sgRNAs. Generally, T0 wheat plants generated 4 to 6 ears and 80-250 T1 grains. A subset of grains samples was analysed using Acid-PAGE.

Acid-PAGE

Acid-Polyacrylamide Gel Electrophoresis has been used for decades to differentiate and identify wheat varieties based on their characteristic gliadin protein profile (Cooke, 1987). It is a non-denaturing procedure, separating on a combination of molecular weight and charge. Here, we used it to identify grains with modified gliadin protein profiles from Fielder-CRISPR plants compared with wild-type Fielder, and to determine the type of changes that had occurred. We first optimised the interpretation of the gels using Chinese Spring deletion lines, and also analysed a Paragon γ-irradiated population to be able to compare the type of changes induced by irradiation mutagenesis with those induced by gene editing.

Chinese Spring

Chinese Spring (CS), a model variety for hexaploid bread wheat, plus the CS nullisomic-tetrasomic and CS Kansas deletion lines, identified as either missing and/or having substitutions of gliadin genes on known homoeologous chromosomes arms, were used to set up and optimise the gliadin protein profile

screening method. We renamed the nulli-tetra lines based on the homoeologous genomes they contain, e.g. N1AT1B was renamed 1BBD and lack 1A.

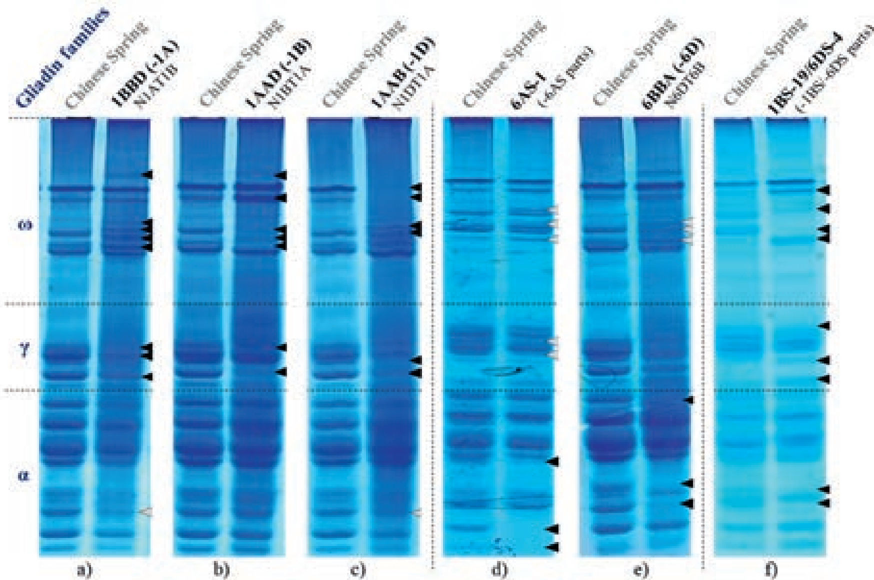


Figure 4. Acid-PAGE of Chinese Spring deletion lines showing altered endosperm gliadin protein profiles. Gliadin extracts from grains of Chinese Spring nulli-tetra lines and deletion lines were run on non-denaturing Acid-PAGE alongside a gliadin extract from Chinese Spring. The lanes displayed next to each other have been run alongside each other on the same gel but each panel represents a different gel. Each sample was always run alongside CS as a control. The black and grey arrows point at the changes in the protein groups from the deleted chromosome arms and in others respectively. CS gliadin profile in absence of a) Chr1-AS, b) Chr1BS, c) Chr1-DS, d) part of Chr6-AS, e) Chr6-DS, f) part of Chr1-BS and Chr6-DS.

These deletion and substitution lines revealed differences in gliadin protein patterns compared with to CS. Various bands were missing or shifted on the Acid-PAGE, depending on which chromosomes (Chr1 γ- and ω-gliadins or Chr6 carrying α-gliadins) and which homoeologous genome (A, B or D) were changed (Fig. 4). As expected, deletion lines and nullisomic-tetrasomic lines supported each other. For instance, in panel e) and f) of Fig. 4, the same two α-gliadin proteins are missing at the bottom of the gel. Band shifts or changes in band intensity were sometimes also seen in a gene family for which the loci had not been changed, such as ω-gliadins in case of a deletion in chromosome 6 (Fig. 4d, 4e). Deletion or substitution of genome 6B did not show any perceptible change, neither in the α-gliadin proteins, nor in the other gliadin families (picture not shown).

In addition to optimising the method, the use of the CS set of lines gave indications on the position of bands from α-, γ- and ω-gliadins which are specifically associated with gliadin proteins originating from A, B or D homoeologous genomes. This will constitute benchmarks to infer, on non-characterised mutant lines, which homoeologous chromosome is the most likely to have been impaired.

Paragon γ-irradiated population

Wheat grains from various Paragon γ-irradiated lines were screened using Acid-PAGE. Differences of gliadin proteins profiles were observed between Paragon and M4 grains in 14 out of 360 γ-irradiated lines tested (3.88%).

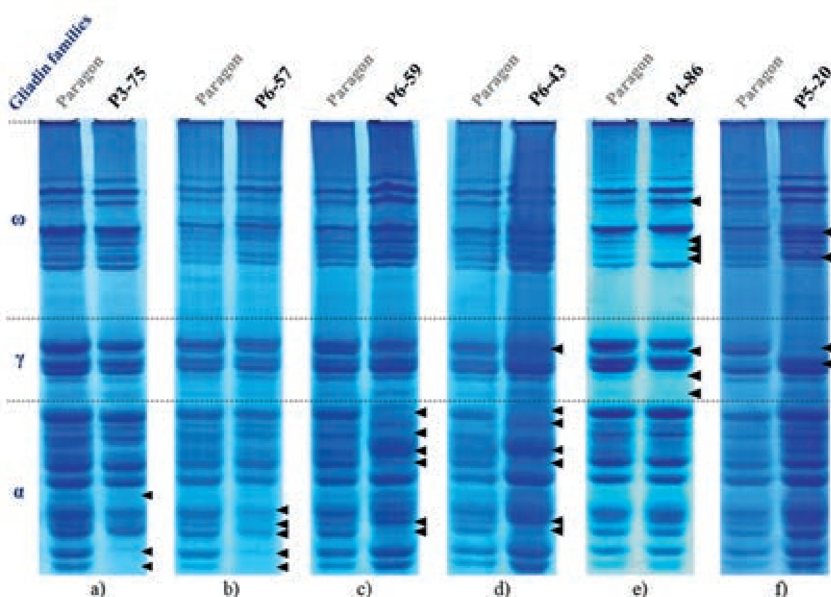


Figure 5. Acid-PAGE of selected Paragon γ-irradiated mutant lines that showed changes in gliadin protein profiles.
Gliadin extracts from grains of the M4 generation of Paragon γ-irradiated mutant population were run on non-denaturing Acid-PAGE alongside a gliadin extract from Paragon (Paragon). Each panel represent a different gel. The lanes displayed next to each other have been run alongside each other. The black arrows point at the changes observed in the irradiated lines. a) and b) α-gliadin bands missing, probably correlated to changes in Chr6-AS, c) and d) α-gliadin bands changes that are different from any change observed in deletion lines and nullisomic-tetrasomic lines, e) γ- and ω-gliadin bands missing, probably due to changes in 1BS, f) γ-gliadin bands missing and ω-gliadin bands shifts and intensity changes, probably due to changes in Chr1-AS.

Three lines showed differences in only the α-gliadins (Fig. 5a, 5b, 5c), one line displayed variations in only the γ-gliadins while four lines presented variations in only the ω-gliadins. One line showed differences in both α- and γ-gliadins (Fig.

5d) while two lines presented changes in both γ-and ω-gliadins (Fig. 5e, 5f). No line screened showed variations in all three gliadin families (Table 3).

By comparing the Paragon mutant gliadin protein profiles with those obtained from CS deletion lines and CS nullisomic-tetrasomic lines, it was possible to infer the homoeologous chromosome location of the gliadin genes altered by γ-irradiation mutagenesis in these lines (Table 3).

Table 3. γ-irradiated Paragon lines and the likely altered chromosome arms underlying the gliadin protein profile variations

	Altered protein profile				Altered protein profile		
	α	γ	ω		α	γ	ω
P3-75	6AS	-	-	**P6-43**	6DS	1AS	-
P4-34	-	1BS	-	**P6-57**	6AS	-	-
P4-84	-	1BS	1BS	**P6-59**	6DS	-	-
P4-86	-	1BS	1BS	**P6-60**	-	-	1AS
P5-20	-	1AS	1AS	**P6-65**	-	-	1AS
P5-53	-	-	1BS	**P6-74**	-	-	1AS
P6-02	-	1AS	1AS	**P10-73**	-	1AS	1AS

Paragon γ-irradiated lines and the gliadin families in which altered protein profiles are seen on Acid-PAGE. The chromosome arms inferred to be altered by the γ-irradiation and to cause the protein profiling changes are indicated in the table. In total, 4 lines showed changes in the α-gliadins, 7 lines had differences in the γ-gliadins and 9 lines displayed variations in the ω-gliadins. P6-57 had large changes in α-gliadins from most likely 6AS.

2

Fielder CRISPR-Cas9 plants

For all of the 117 T0 plants expressing the Cas9 mRNA and carrying the full complement of expected sgRNAs, 8 or more randomly selected T1 grains per plants were screened on Acid-PAGE. Differences in the gliadin protein profile were observed between Fielder wild type and T1 grains harvested from some T0 CRISPR/Cas9 plants (Fig. 6; Table 4.).

Modified profiles were found for each of the 4 CRISPR/Cas9_sgRNA constructs used (Fig. 6). When a modified grain was found, most of the other T1 grains of that transformed plant showed a wild type profile. When several mutated T1 grains were found from one T0 plant, in some cases they contained similar profile modifications (Fig. 6e), in other cases the changes were different. Similarly altered gliadin protein profiles were sometimes seen in grains coming from different T0 individuals carrying the same CRISPR/Cas9_sgRNA constructs or even from

different constructs (Fig. 6a, 6d, at α-gliadin level). The "clear type of protein profile changes" observed were protein bands disappearing or appearing on the gel (Fig. 6a, 6b, 6e). Bands could also be present at a different intensity (Fig. 6c, 6d) or have a shifted position (Fig. 6c, 6e) that was not always very clear and therefore qualified as "potential changes" type. The term "any changes" refer to both clear and potential changes. The changes in protein expression were mostly present in the targeted gliadin families. However, differences in intensity or shifts in size were sometimes seen in bands associated with non-targeted gliadin families (Fig. 6c, 6d, 6e). Based on the deletion of 3 α-gliadin bands obtained with construct α1 in grain α1-14_G1, 21 more grains from the T0 α1-14 plant were screened. However, none of them had an altered gliadin protein profile compared with Fielder.

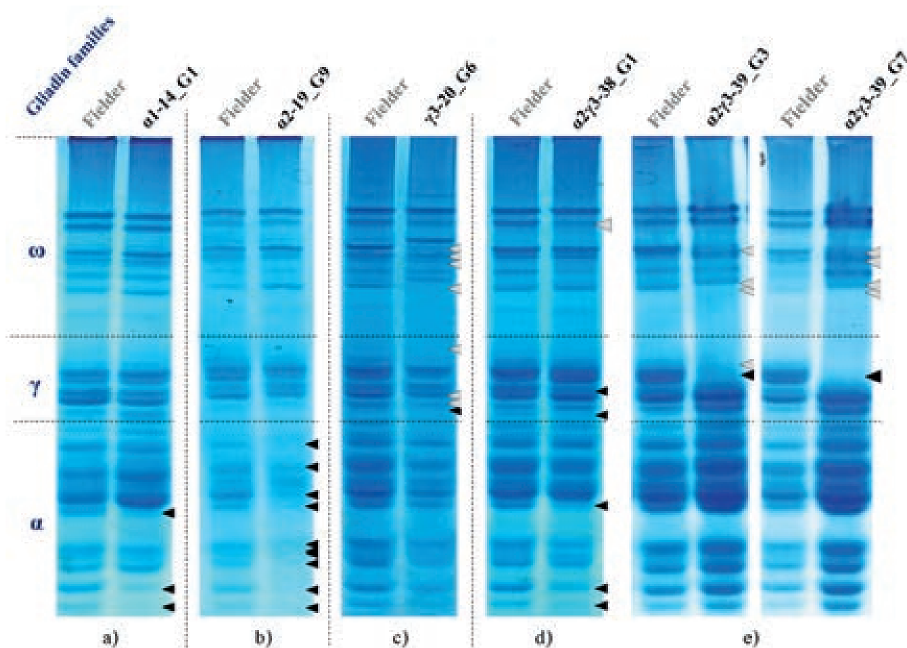


Figure 6. Acid-PAGE of Fielder-CRISPR T1 grains showing altered gliadin protein profiles
Gliadin extracts from Fielder-CRISPR T1 grain from each of the 4 constructs were run on non-denaturing Acid-PAGE alongside the gliadin extract from Fielder wild type. Each panel represent a different gel. The lanes displayed next to each other have been run alongside each other. The start of the sample names refers to the constructs with the sgRNAs they include (α1, α2, γ3 or α2γ3), followed by the T0 plant line and grain number. The black and grey arrows point respectively at the changes intended or unintended by the construct present in the plant that set the grains. a) α-gliadin bands missing likely related to mutations on Chr6-AS, b) α-gliadin bands missing or lower expressed likely related to mutations on Chr6 in all 3 homoeologous genomes, c) ω-gliadin bands shifted up and γ-gliadin band with lower expression level likely related to mutations on Chr1-BS or DS, d) ω- γ- and α-gliadin bands with lower expression likely related to mutations on Chr6-AS, e) ω-gliadin bands shifted up and γ-gliadin bands missing likely to be related to Chr1-AS, similar mutant profile in two different T1 grains from the same T0 plant.

The constructs differed significantly in mutation efficiency for any changes (Chi square, P= 0.015) but not for clear changes (P=0.097) (Table 4). Construct α1_sgRNA had the lowest mutation efficiency, with very low number of grains showing any mutations on Acid-PAGE while having the highest number of T0 plants expressing Cas9. Construct α2_sgRNA had the lowest number of T0 plants expressing Cas9 but the highest percentage of grains showing clear changes. Construct γ3_sgRNA generated the highest number of grains with any sort of mutations. Construct α2γ3_sgRNA gave only 3 plants with clearly mutated grains but these plants gave 3 or 4 mutant grains each, which is the highest ratio of mutated grain per plant (Table 4).
It was observed that construct γ3_sgRNA and construct α2γ3_sgRNA produced grains which relatively often had differences in ω-gliadin expression although these constructs do not target this gliadin family. This phenomenon was rarely observed with construct α2_sgRNA and seen only once with construct α1_sgRNA.

No correlation was found between transgene copy number present in a plant and the number of grains presenting an altered gliadin protein profile, nor in the number of mutations seen per grain. Since the Cas9 expression was observed by RT-PCR, but was not quantified, no correlation can be made between its expression level and the number of gliadin protein changes observed on the Acid-PAGE.

Of the T1 grains obtained with α1_sgRNA construct, 3 embryos were grown to mature plants and from the obtained T2 grains 8 were tested. These grains all had a similar profile to the wild type. An additional 100 T2 grains were tested from the most promising α1-14_G1 T1 plant from which the T1 grain clearly displayed abnormal protein profile, but also these had a normal profile. T2 grains from other CRISPR/Cas9 constructs remain to be tested.

Table 4. Summary of T1 Fielder-CRISPR grains, and associated T0 plants, with modified gliadin protein profile

CRISPR/Cas9 constructs	T0 plants expressing Cas9	**T1 grains tested on Acid-PAGE**	T0 plants giving T1 mutant grains with clear changes	**T1 grains with clear changes in gliadin profile**	T0 plants giving T1 mutant grains with any changes	**T1 grains with any changes in gliadin profile**
Construct α1_sgRNA sgRNA_α87	38	325	3 (7.9%)	4 (1.2 %)	3 (7.9%)	8 (2.5 %)
Construct α2_sgRNA sgRNA_α213 + sgRNA_α324	12	288	7 (58.3%)	13 (4.5%)	7 (58.3%)	19 (6.6 %)
Construct γ3_sgRNA sgRNA_γ86 + sgRNA_γ272 + sgRNA_γ603	35	280	10 (28.6%)	10 (3.6 %)	15 (42.9%)	23 (8.2 %)
Construct α2γ3_sgRNA sgRNA_α213 + sgRNA_α324 + sgRNA_γ86 + sgRNA_γ272 + sgRNA_γ603	32	256	3 (9.4%)	11 (4.3 %)	11 (34.4%)	20 (7.8 %)
TOTAL	117	**1149**	23 (19.6%)	**38 (3.3 %)**	36 (30.8%)	**70 (6.1 %)**

In this table are displayed the number of T0 plants and associated T1grains tested on Acid-PAGE, as well as the number of T0 plants giving T1 mutant grains with "clear changes" or "any changes" and the number of T1 mutant grains showing "clear changes" and "any changes". For each construct mutations were observed. Mutations occurring in both the targeted and non-targeted gliadin families were included in "any changes" here.

Discussion

In this study, CRISPR/Cas9 was deployed in hexaploid bread wheat to target gliadin genes. The Cas9-induced mutations were designed to remove gliadin gene copies or disrupt immunogenic epitopes in gliadins, preventing them from triggering the human immune system and causing coeliac disease upon wheat gluten consumption. At the start of this study, CRISPR/Cas9 had not been used to target large clusters of repeated gene families, such as the gliadins, in polyploid crops such as hexaploid wheat. Therefore, the challenge was not only to generate such mutant plants but also to develop screening methods to identify them and characterise their mutations further.

Gliadin alignments, sgRNA design and Fielder-CRISPR lines generation

In the absence of α- and γ-gliadin genomic or proteomic sequences from the hexaploid wheat cultivar Fielder used for gliadin gene editing experiments, and considering the time and resources necessary to fully clone and sequence these gene families, an alternative approach has been used mainly based on publically available data. Only seven α-gliadins from Fielder were cloned, sequenced and merged with all α-gliadin genomic sequences from Genbank. The Fielder sequences were different to one another but similar to some sequences already present in Genbank. Deduced protein sequences were manually aligned to enable inferring gliadin epitope positions and conserved regions in gliadin gene sequences across cultivars. Using this approach, it is possible to design sgRNA targeting as many genes as possible near their epitope region, not only in Fielder but also in many other different cultivars. It also revealed the presence of groups based on sequences patterns, some associated to specific homoeologous genomes, similarly to what was observed by Ozuna *et al.* (2015). The same approach was used for γ-gliadins, although no Fielder γ-gliadins were sequenced. The association of CD epitopes and genomes of origin was possible for α-gliadins whereas it was not straightforward for γ-gliadins.

The sequences downloaded from Genbank originated from different hexaploid varieties. When sequences from one cultivar are uploaded into the database, there is no guarantee that the set of sequence variants per cultivar is complete. Furthermore, gliadin genes are present in each homoeologous genome and clustered in specific loci as multiple copy repeats, with identical or variable sequences. A gene sequence found multiple times in one cultivar, is uploaded only

once as new sequence to avoid redundancy. However, if an identical sequence was found in different cultivars or in the same cultivar different studies by different groups, it will be present several times in the database. Therefore, the frequency with which a sequence is present in the alignments generated for this study does not reflect the frequency at which a gene sequence is present in the genome of a cultivar such as Fielder. As a result, the proportion of sequences targeted by the different sgRNA designed does not reflect the real percentages of matching sequences present in Fielder, especially since the evaluation was made on 100% match while in reality, sgRNA are known to target sequences with slightly lower similarity as well (Endo *et al*., 2015).

Four CRISPR/Cas9_sgRNA constructs, containing 1 to 5 different sgRNAs, were designed and used to target α- and/or γ-gliadin gene families, in hexaploid wheat. The constructs were stably transformed into bread wheat cultivar Fielder. A total of 117 regenerated lines expressed Cas9, regardless of which construct they carried. The T1 grains generated by these plants potentially contain mutations at a number of target sites which are segregating independently. However, it is possible that the Cas9-induced mutation is heterozygous in a cell. It is also possible that Cas9 did not cut the same genes in different cells of the T0 plant generating chimerical plants or even that Cas9 remained active in the gametes after meiosis. This means that, following segregations events, each grain generated by a T0 plant could have a unique assortment of gliadin mutations. For this reason, each grain was cut into three parts, the embryo and two identical pieces of endosperm, to run further complementary types of analysis (sequencing and advanced proteomics) on the identical fractions of endosperm isolated from the same T1 or T2 grain.

Pre-screening for wheat grains with mutated gliadins

Due to the high complexity of the large gliadin gene families that are only expressed in wheat grains, the traditional screening methods such as restriction site loss or sequencing are not appropriate for pre-screening and identification of potential mutants. Instead, these methods would be relevant for in depth study and characterisation of the mutations occurring in interesting mutants identified using different pre-screening method. Acid-PAGE was successful for identifying gliadin protein profile differences (Cooke, 1987) and we therefore employed it for high-throughput pre-screening of the grains from mutant wheat plants. This method was first optimised using CS and associated set of deletion lines and

nullisomic-tetrasomic lines. These lines were previously characterised as missing chromosome pairs or arms that carry gliadin genes from specific homoeologous genome. These lines enabled identifying the homoeologous chromosome arms and the sub-genome that most likely was altered in a mutant, based on the absence or shift in position of the band.

The optimised Acid-PAGE protocol was then validated, as it was able to identify 14 lines out of 360 (3.9%) Paragon γ-irradiated lines that showed gliadin proteins expression changes compared to Paragon. Lines showed changes in one or two gliadin families but never in all three families simultaneously. Irradiation mutation is known to trigger large deletions, up to several mega-bases, which can explain the deletion of a complete gliadin gene locus. The nature of the homoeologous chromosomes altered by γ-ray were inferred based on the results obtained with the CS line set, since similarly large deletions are expected in both line sets. The percentage of mutations observed on Acid-PAGE was around 4%, for all visible changes in the three large gliadin gene families, each counting more than dozens of members.

Changed gliadin profiles were observed with each construct, implying that each construct contained at least one sgRNA that successfully generated mutations in some copies of the targeted α- or γ-gliadin gene family. More importantly, it indicates that CRISPR/Cas9 can edit a sufficient number of genes within large family in polyploid plants to actually generate a different phenotype in the progeny. The acid-PAGE analysis of the Fielder-CRISPR lines revealed differences in gliadin profiles in 70 T1 grains (6.1% of the total number of grains tested) harvested from 36 T0 plants (30.8% of the total number of T0 plants expressing CRISPR/Cas9 constructs) across all four CRISPR/Cas9_sgRNA constructs used. Therefore, a single T0 plant that produced T1 mutant grains, gave on average 2 mutant grains out of 8 grains tested.

Constructs targeting γ-gliadins sometimes trigger shift of protein bands in the untargeted ω-gliadins. This phenomenon could be explain by the deletion of ω-gliadins present in between targeted γ-gliadins since both gene families are suspected to have some overlap on the short arm of chromosome 1 (Dong *et al.*, 2016). Moreover, it is known that knocking out some gene in a gluten family triggers the compensation by other gene families (Pistón *et al.*, 2011, Galili *et al.*, 1986).

Variation in efficiency could be observed between the constructs, with construct α1_sgRNA being significantly less efficient. This could be explained by the presence of only one sgRNA in this construct while several are present in other constructs. It could also be due to a presence of secondary structure in the sgRNA – identified after use, using RNAfold software - that could decrease the target binding efficiency. The number of sgRNA in a construct appears to slightly increase the ratio of mutant grains obtained, but no direct correlation was observed. Indeed, construct α2_sgRNA and construct γ3_sgRNA generated 19 and 23 mutant grains respectively while construct α2γ3_sgRNA – combining both previous sgRNA guides in a single construct - gave only 20 mutant grains, which does not show a cumulative efficiency. However, the Cas9 expression level in different plants as well as the actual number of mutations generated at the DNA level would be necessary to give a robust answer regarding the variation of efficiency of the different constructs.

Inheritance of gliadin mutations observed in T1 grain endosperm

Three T1 embryos, associated with three T1 endosperm samples with gliadin mutations, were planted and T1 plants grown. From each of the three T1 plants, 8 T2 grain were harvested and run on Acid-PAGE, revealing no mutations. An additional 100 T2 grains from T1 α1-14_G1 mutated plants, which the T1 endosperm showed three missing α-gliadin protein bands associated to 6AS, did not reveal mutations either. This suggests that the mutations identified in the T1 generation do not appear to have been inherited in the T2 generation and that no new mutations were observable in the second generation, in contrast to CRISPR/Cas9 mutations generated in wheat (Wang *et al.*, 2018).

The first hypothesis for the lack of inheritability that one would think of is segregation issues, especially in case of potential heterozygous mutations, in gene families in hexaploid plants. Meanwhile, looking at the biology of the T1 endosperm and its associated embryo - which is the origin of the future T1 plant -, the absence of inheritance of mutations associated with missing bands on the Acid-PAGE is not straight forward. Wheat endosperm is formed by the fecundation of doubled-haploid female (2x ABD_f) central cell by the second haploid male gamete (1x ABD_m) while the first haploid male gamete (1x ABD_m) fecundate the haploid female (1x ABD_f) egg cell in order to form the embryo (Wegel and Shaw, 2005). Wheat endosperm is therefore triploid (2x ABD_f + 1x ABD_m) although carrying the similar genetic material than the diploid embryo (1x

ABD_f + 1x ABD_m). Mutations present in the endosperm are identical to the one present in the embryo, only the dosage of female mutation is doubled in the endosperm. In case of T1 α1-14_G1, for the endosperm to display three α-gliadin band missing on Acid-PAGE, corresponding to genes in short arm of chromosome 6A, this would imply that the mutations were present in both female and male gametes. Therefore, the associated embryo subsequently growing into T1 plant should be homozygous for these mutations on 6AS. As a consequence, all gametes formed by T1 α1-14_G1 plant should also carry these mutations and the resulting T2 grain should have a triploid endosperm and a diploid embryo, all homozygous for these α-gliadin mutations on chromosome 6A, with potential new mutations generated afterwards. The T2 endosperms should therefore show at least similar double α-gliadin band missing associated to chromosome 6A, which was never observed in more than 100 tested T2 endosperms from α1-14_G1 T1 mutant.

The lack of inheritable phenotypes associated with the mutations in T2 grains could potentially be explained by epigenetic modification and compensation within the targeted gene family. Indeed, mutations leading to suppression of some gliadin protein expression in the T1 may have triggered the expression of remaining intact copies that were initially not expressed. Such compensation mechanisms within a gliadin family could be plausible considering that it already takes place between different gliadin families (Pistón *et al.*, 2011; Galili *et al.*, 1986). It could explain that the mutant phenotype cannot be visualised on Acid-PAGE despite the actual inheritance of the gliadin mutations themselves, and thus the detection requires further analysis with different methods. However, the T2 grains tested were obtained with the less efficient sgRNA and the T2 grains generated with the other, more efficient, constructs are yet to be tested.

Comparison with other groups also using CRISPR/Cas9 to target gluten genes in hexaploid wheat

Sánchez-León *et al.* (2018), who used a sgRNA, namely sgRNA_α2, having 13 overlapping bases with our sgRNA_α213 and targeting α-gliadins 6 base-pairs upstream, reported a higher success rate with T1 grains from one T0 CRISPR/Cas9 line being mutated in a similar way, and inheriting those mutations in T2 grains. These differences of results may be related to a low Cas9 efficiency in our case. Cas9 sequences have quickly been improved to increase the mutation efficiency and the

Cas9 version we used (different from the one used by Sánchez-León *et al.* (2018)) has been reported as having a low efficiency (Ma *et al.*, 2015, Wang *et al.*, 2016). In addition, in monocots, most research report the use of maize ubiquitin promoter (Sánchez-León *et al.,* 2018; Ma *et al*., 2015) while we used the rice actin promotor. Those factors could explain not only the low number of T1 mutant grains per T0 plants obtained but also the low level of gene copies mutated within each mutant wheat grain. It could also explain the absence of mutated phenotype inheritance, assuming that with a higher efficiency and most of the targeted genes mutated, compensation of the mutated gene copies becomes much more difficult for the crop.

Comparison of CRISPR/Cas9 targeted mutations with γ-irradiation random mutagenesis

Using either CRISPR/Cas9 or γ-irradiation method, the mutation efficiency revealed by the Acid-PAGE is comparable. However, the mutation rate obtained using CRISPR should be much higher. Fielder-CRISPR lines tested correspond to T1 generation grain potentially heterozygous for mutations while Paragon γ-irradiated correspond to M4 generation grain where most mutations are homozygous and easier to visualise. In addition, Acid-PAGE will probably not reveal amino-acid substitutions or small in-frame indels potentially generated by CRISPR/Cas9 since proteins with different sequences but having similar molecular weight/charge ratio can be represented by a single band on the gel (Anderson and Greene, 1997). Moreover, knocking out a gliadin gene may not supress any protein band if another similar gene has not been knocked out. Alternatively, 2D-gels could give a higher resolution but other methods such as deep DNA sequencing are needed to reveal and characterise more subtle types of mutations and to get a better idea of the differences in efficiency between both methods.

Interestingly, some T1 CRISPR-Fielder grains showed a similarly altered gliadin protein profile as some M4 Paragon γ-irradiated grains (Fig. 6a, 6e and 5a, 5f)). The difference, however, is that in the Fielder-CRISPR grains, only the α- or γ-gliadin gene family were targeted and are probably modified whereas in the Paragon γ-irradiated grains it is possible that not only gliadin genes but also unrelated flanking genes were deleted as a side effect. Furthermore, wheat lines with small deletions in the epitope region possibly generated by CRISPR/Cas9 would be more interesting than multiple gene copy deletions generally generated by γ-irradiation for two reasons. First of all, the gliadin copies would remain under a 'safer' form while retaining the backing properties more easily than if being totally removed

using irradiation. In addition, having small deletions might avoid compensation, by other potentially immunogenic genes, whose expression is triggered when a gene family is partly knocked down. However, following the actual regulation regarding gene-editing and mutation breeding, the Fielder-CRISPR wheat could not be grown yet due to strict GM regulation while the Paragon γ-irradiated wheat could directly be grown and used for breeding programs without any restrictions (*Chapter 5*).

Conclusion

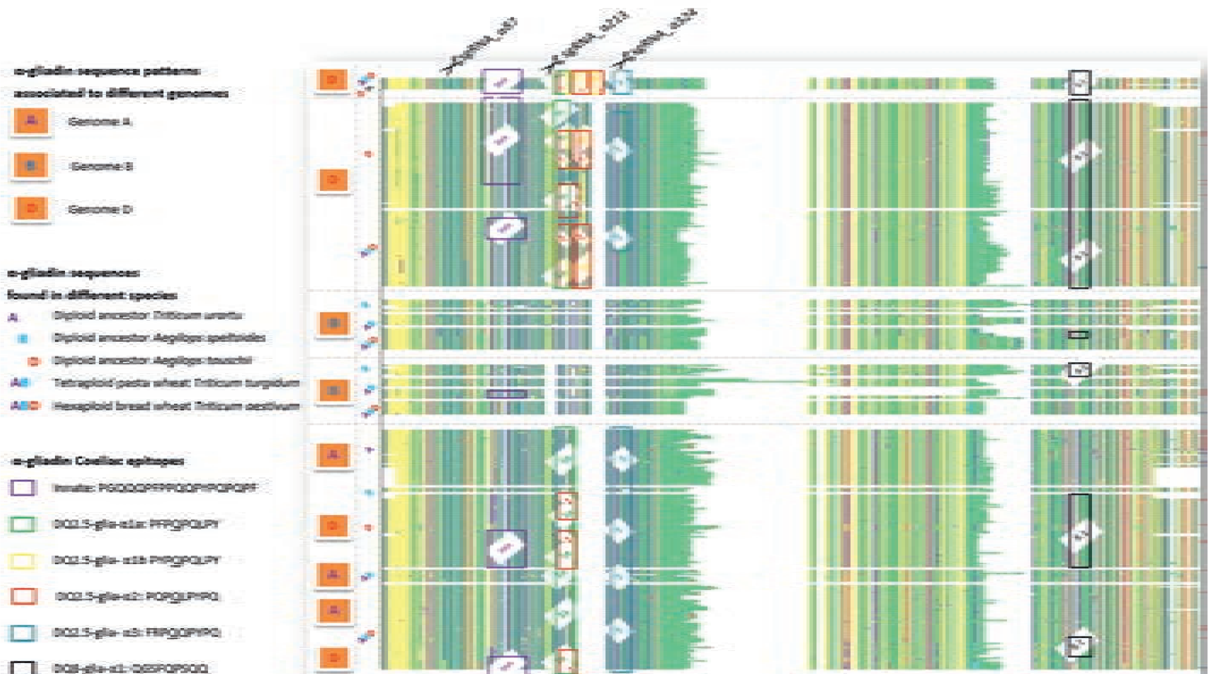
This pilot study aimed to mutate the large α- and γ-gliadin gene families in hexaploid wheat to decrease gluten immunogenicity for coeliac patients. We succeeded in generating wheat grains that contained gliadin protein profiles altered for the targeted gene family in the bread wheat cultivar Fielder by targeted mutagenesis using CRISPR/Cas9 and in Paragon germplasm induced by random mutagenesis using γ-irradiation. As many gliadin genes have not been altered, optimisation concerning uses of different promoter and different Cas9 or new Base Editor systems needs to be done to obtain wheat plants which are safe for CD patients.

The pre-screening and identification of the mutations was performed using Acid-PAGE, which identifies mostly the non-sense mutations and large deletions. However, the final purpose is to modify the epitopes into safe versions without knocking out the complete gliadin genes, in order to avoid compensation by other gliadins and to maintain baking quality. Therefore, reliable high throughput methods for small in-frame mutation detection will be important. Methods such as droplet digital PCR, enrichment and sequencing as well as advanced proteomics studies will be needed to identify also these subtle modifications and characterise them further. This should increase the percentage of plants in which mutations have been induced, even further and enable the identification of mutations in T2 plants.

Segregating out the CRISPR/Cas9 construct from promising lines and subsequently self-pollinating these wheat lines to make all mutations homozygous would be the next steps. Potentially interesting lines would then need extensive investigation including immunological tests using monoclonal antibodies to determine their immunogenicity level and rheological studies to evaluate the bread dough quality obtained using these generated “hypoimmunogenic-gluten” wheat lines.

Acknowledgements

AJ was funded by FP7-PEOPLE-2013_ITN-607178. This manuscript does not necessarily reflect the Commission's views and does not anticipate the European Commission's future policy in this area. Melanie Craze, Sarah Bowden and Ruth Bates from the NIAB Crop Transformation team are thanked for creating the CRIPSR/Cas9 wheat lines. Helen Appleyard's Analytical Services team, especially Sofia Alves and Pravina Solanki, are thanked for technical assistance with Acid-PAGE gels. The Paragon γ-irradiated population and the Chinese Spring nullisomic-tetrasomic lines were kindly provided by Dr. Clare Lister and Dr. Simon Griffiths of JIC, Norwich, UK, respectively. The Cas9 endonuclease and the sgRNA sequences used were kindly provided by Professor Caixia Gao, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, and described in Shan *et al.*, 2013.

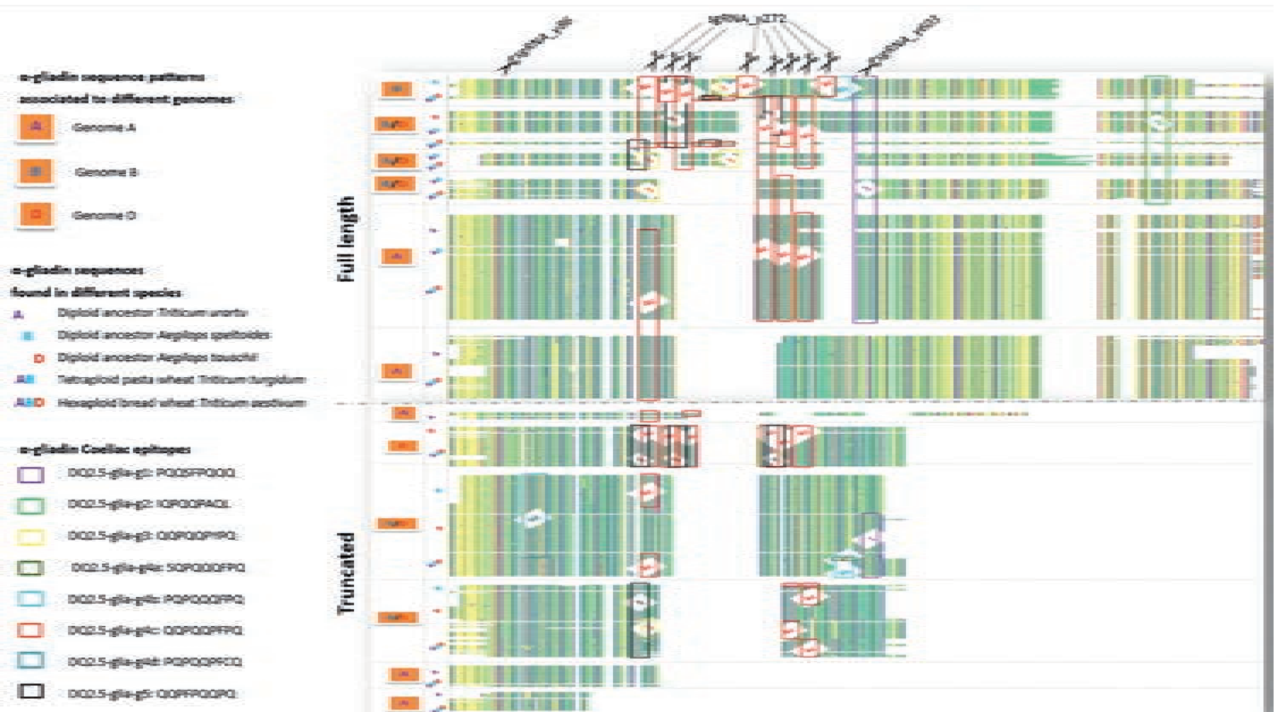


***Supplementary Figure 1.* α-gliadin protein sequences alignment**

A selected subset of the 1273 aligned α-gliadin protein sequences from cultivated wheat and wild relative is shown here. Five patterns based on variation at the DQ2.5 epitope region were identified and are separated by horizontal dash-lines. Each pattern has been associated to a dominant genome, indicated by large A, B or D letters on the first column, with which its sequence s seem to be related to. Within a pattern, sequence originating from different wheat species diploid, tetraploid or hexaploid are separated by an empty line and the genomes present in the species are indicated by smaller A, B and D letters in the second column. The canonical sequences of CD epitopes, which often overlap with one another, are framed in different colours with their category number indicated as well. On top of the figure, the scissors indicates the position at which the sgRNA designed are cutting, relatively to the position of the CD epitopes.

Some patterns have specific CD epitopes combinations and are clearly associated to a genomes while some are not. The 5th pattern, has actually an amino-acid substation in genome D compared to genome A, making it safer for CD patient than in genome D. Sequences from genome B have naturally occurring amino-acids deletions within the epitope regions that prevent their recognition by the immune system in comparison to proteins from other genomes.

2



Supplementary Figure 2. γ-gliadin protein sequences alignment
A selected subset of the 1273 aligned γ-gliadin protein sequences from cultivated wheat and wild relatives is shown here. Six main patterns based on variation at the DQ25 epitope region were identified and are separated by horizontal dash-lines. Each pattern has been associated to a dominant genome, indicated by large A, B or D letters on the first column, with which its sequences seem to be related to. Within a pattern, sequence originating from different wheat species diploid, tetraploid or hexaploid are separated by an empty line and the genomes present in the species are indicated by smaller A, B and D letters in the second column. The canonical sequences of CD epitopes, which often overlap with one another, are framed in different colours with their category number indicated as well. On top of the figure, the scissors indicates the position at which the sgRNA designed are cutting, relatively to the position of the CD epitopes. A clear distinction was made by a thick dash line between the full length protein and truncated ones, usually rising from pseudogenes.
Some patterns have specific CD epitopes combinations and are clearly associated to a genomes while some are not.

Chapter 3

Optimization of droplet digital PCR for determining genetic variation in the large α-gliadin gene family in hexaploid wheat

Aurélie Jouanin [1,2], Rubén Tenorio Berrio [1], Fiona J. Leigh [2], Jan G. Schaart [1], Richard G.F. Visser [1], Marinus J.M. Smulder [1]

[1] *Wageningen University and Research, Plant Breeding, Wageningen, The Netherlands*

[2] *The John Bingham Laboratory, NIAB, Huntingdon Road, Cambridge, United Kingdom*

Submitted

Abstract

Droplet digital PCR (ddPCR) enables the accurate quantification of nucleic acids of interest. This technique has been proposed as a rapid and cost-efficient alternative for high-throughput screening of gene copy number variation (CNV). The sensitivity of ddPCR has yet to be tested for the quantification of the gene copy number in large families in polyploid plant species, where the relative differences between copy numbers are smaller. The α-gliadin gene family in hexaploid bread wheat varieties and lines was used here as a case study to test the feasibility and accuracy of ddPCR for CNV analysis in large (> 50 gene copies) gene families. We designed and optimised duplex ddPCR assays that amplify a reference gene that occurs as a single copy gene on each of the six homoeologous chromosomes, along with α-gliadin genes for which the copy number is to be determined, and demonstrate their accuracy for measuring gene copy numbers in four sets of genetic material. First, the sum of α-gliadin gene copy number measured with the ddPCR assays in diploid and tetraploid wheat accessions corresponded to the copy number measured in Synthetic Hexaploid Wheat lines produced from them, demonstrating the linearity of the detection up to 95 gene copies. Second, a comparison of α-gliadin CNV of a set of deletion and nullisomic-tetrasomic lines from variety Chinese Spring enabled determining the number of α-gliadin genes present on the three pairs of homoeologous chromosomes 6A, 6B, and 6D. Third, ddPCR assays detected a reduction of 18 α-gliadin gene copies between variety Paragon and some of its derived γ-irradiated mutant lines. Finally, the comparison of two different duplex ddPCR assays demonstrated the feasibility to distinguish, quantitatively, small indels from larger deletions induced by a CRISPR/Cas9 construct targeting α-gliadin genes in variety Fielder. We showed that a single Fielder-CRISPR line could contain small indels (1-50bp) in up to 10 α-gliadin genes and large deletions (>300bp) in 20 α-gliadin genes out of the 87 α-gliadin genes present in Fielder. We demonstrate that ddPCR is suitable and sensitive for high-throughput screening of gene CNV and gene-editing induced mutations in large gene families, in polyploid organisms. We also show a feasible ddPCR strategy to distinguish small indels from large deletions generated by CRISPR/Cas9 sgRNAs, in large gene families. In case of wheat, ddPCR thus represents a valuable tool for high throughput screening for lines with low gliadin gene copies, to be used in breeding programs toward hypoimmunogenic gluten, representing an important step forward for people with Coeliac disease.

Key words: *α-gliadins, CNV, CRISPR/Cas9, ddPCR, Gene editing, Gluten, Gene family, Mutation breeding, Polyploid, Wheat, Coeliac disease.*

3

Introduction

Gene families are composed of several genes that often vary in terms of sequence content and copy number across varieties, influencing diverse crop traits. Due to the high variability, it is challenging to qualitatively or quantitatively assess a gene family and predict its impact on the phenotype or performance of a plant. In the 60's, gene copy number variation (CNV) was estimated using the southern blot technique, which is very tedious (Collier *et al.*, 2017). Since the 2000's, qPCR was used to estimate CNV quickly and using less DNA (Collier *et al.*, 2017). However, the measurement accuracy is limited for high gene copy number (Karlen *et al.*, 2007) and it varies significantly among replicates (Weaver *et al.*, 2010). Whole genome sequencing can be used to have a hint on the gene copy number by using the coverage variation. However, data analysis is difficult for large and often highly similar gene families and cannot yet be applied for high-throughput screening in case of large polyploid genomes, for instance for gliadin genes in hexaploid wheat. Recently, droplet digital PCR (ddPCR), which is more precise than qPCR, has been suggested as a reliable alternative for high throughput gene copy number quantification (Hindson *et al.*, 2011, Manoj 2016).

ddPCR is a PCR-based assay, using primers for amplification of specific DNA fragments and probes or a fluorescent intercalating dye for detection. The power of ddPCR relies on the partitioning of the DNA fragments to amplify into water-in-oil droplets, enabling multiple simultaneous endpoint PCR reactions per sample and increasing the reliability and precision of the quantification. In the first step of the ddPCR workflow (Fig. 1), the PCR reaction and droplet generating oil are loaded into separated wells of an injection moulding cartridge. An emulsion is created by applying a vacuum, generating around 20.000 droplets in each sample. Then the samples are transferred into a thermocycler to undergo a standard PCR. Each droplet constitutes a partition in which an individual PCR reaction will occur. Some of these partitions, called positive droplets, contain a copy of the DNA fragment to amplify and will yield a fluorescent PCR product, others do not and these form the negative droplets. After the amplification, samples are transferred to the reader. Here, droplets are aspirated, separated and aligned toward the detector where they are detected one by one. Positive droplets are distinguished by a strong fluorescence signal while negative droplets are distinguished by their low intrinsic fluorescence due to imperfect quenching. The negative droplets form the

background, which is essential to establish a contrast with positive droplets. The concentration of DNA fragment to amplify is determined by calculating the ratio of positive and negative droplets in the sample. The random distribution of the templates across the droplets allows correction and calculation of the confidence interval of the concentration using Poisson statistics (Basu, 2017). To calculate the absolute gene copy number, it is necessary to know the number of genomes relative to the amount of DNA used. For more precision, a single copy gene can be included as reference. This method can be used for high throughput screening, giving accurate and reproducible results in a quick and relatively cheap manner (Hindson *et al.*, 2011). Nevertheless, the accuracy of ddPCR for CNV has never been tested on organisms having both a large gene copy number and a large genome size, such as the α-gliadin gene family (> 50 members) from hexaploid wheat (16 Gb genome size).

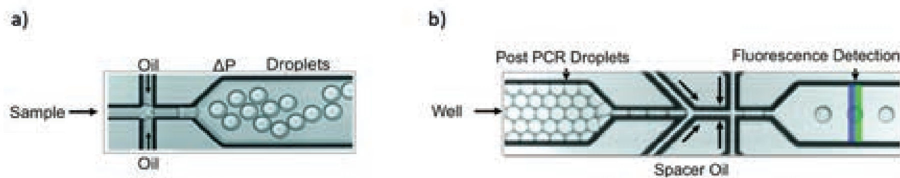


Figure 1. Main steps of droplet digital PCR workflow.
a) Droplet generation: A vacuum is applied to the well, aspirating the sample and the oil and generating around 20.000 surfactant-stabilized droplets. b) Read droplet fluorescence: After PCR, droplets are extracted from each well and individually streamed through a fluorescence detector. Adapted from Hindson *et al.*, 2011.

Hexaploid wheat (*Triticum aestivum*, 2n = 6x = 42, AABBDD) is a staple food crop with grains containing gluten in their endosperm. Gluten includes two major storage protein types: the glutenins and the gliadins. Gliadins consist of three gene families: α-, γ-, and ω- gliadins. The α-gliadin family was estimated to contain between 60 and 150 gene copies in *T. aestivum* (Anderson *et al.*, 1997; Ozuna et *al.*, 2015), located at *Gli-2* loci on the short arm of homoeologous chromosome 6 (Payne, 1987). The γ- and ω-gliadins were estimated to contain around 40 and 20 gene copies, located at *Gli-1* and *Gli-3* loci respectively on the short arm of homoeologous chromosomes 1 (Sabelli and Shewry, 1991). The number of α-gliadins and their sequence composition is relevant for Coeliac disease (CD) in humans. This autoimmune reaction is triggered by the ingestion of gluten proteins carrying immunogenic epitopes that can be recognised as antigens by the immune

system. CD affects around 1-2 % of the human population (Koning, 2012) and produces a chronic inflammation of the small intestine, leading to other symptoms including bowel disorders and malnutrition. Up to date, the only remedy is a strict gluten-free diet. Alternatively, “low-gliadin” products would constitute a great step forward toward “hypoimmunogenic-gluten” food products for CD patients. Assessing the α-gliadin gene copy number of hexaploid wheat using ddPCR is therefore of great interest in the context of developing healthier alternatives to gluten-free products for CD patients.

In this paper, we describe the design, optimization and validation of a ddPCR protocol for measuring near-absolute and relative α-gliadin gene CNV. We first used a mixture of DNA from three hexaploid wheat cultivars for the protocol optimisation. Then, we analysed diploid, tetraploid and resulting synthetic hexaploid wheat (SHW) DNA, in order to check the linear range of the ddPCR detection of α-gliadins. Subsequently, we tested the sensitivity and accuracy of the ddPCR for CNV in a range of materials with decreasing number of gene copy changes: Chinese Spring (CS), the model cultivar for hexaploid wheat, and a series of selected deletion lines and nullisomic-tetrasomic lines, previously characterised as lacking parts or complete chromosome arms with multiple α-gliadin genes; Selected variety Paragon γ-irradiated lines that are known to have deletions in the α-gliadins; and CRISPR/Cas9 lines of variety Fielder, in which it is assumed that there are deletions in the α-gliadins as visualized by missing specific α-gliadin bands on Acid-PAGE.

Materials and methods

Plant materials

Genomic DNA was extracted using a CTAB-based extraction protocol from young leaves of various wheat species, varieties and accessions (Table 1). Fielder, Paragon and Chinese Spring (CS) hexaploid bread wheat DNA was pooled together for performing optimisation experiments. Three synthetic hexaploid (AABBDD genome, Ogbonnaya et *al*., 2013) wheat lines (SHW) developed by NIAB (Cambridge, UK) were analysed, along with the three different *Aegilops tauschii* lines (DD genome) and two *Triticum turgidum* (AABB genome) lines from which it was generated. CS, five CS nullisomic-tetrasomic lines (Sears,

Table 1. Wheat genotypes studied

Species	Lines	Description
Aegilops taushii	Ent 392 Ent 404 Ent 405	DD parental lines
Triticum turgidum	Biensur Hoh 501	AABB parental lines
Synthtic Hexaploid Wheat (SHW)	NIAB SHW 54	Ent 392 (DD) x Biensur (AABB)
	NIAB SHW 75	Ent 404 (DD) x Hoh 501 (AABB)
	NIAB SHW 76	Ent 405 (DD) x Hoh 501 (AABB)
***Triticum aestivum* cultivar 'Chinese Spring' (CS)**	WT	Control
	6AS-1 6BS-4/5BS-2 6DS-4 6DS-4/1BS-19	Kansas Deletion lines
	N6AT6D = 6DDDDBB N6AT6B = 6BBBBDD N6DT6A = 6AAAABB N6BT6D = 6DDDDAA	Nullisomic-tetrasomic lines
***Triticum aestivum* cultivar 'Paragon'**	WT	Control
	P3-75 P6-43 P6-57 P6-59	γ-irradiated lines M4 generation
***Triticum aestivum* cultivar 'Fielder'**	WT	Control
	α1-14_G1 α2-1_G10 α2-1_G15 α2-15_G1 α2-19_G9 α2-19_G12 α2-21_G10 α2-26_G14 α2-28_G7	CRISPR/Cas9 lines α-gliadins targeted
	γ3-1_G6 γ3-12_G4 γ3-13_G7 γ3-17_G8 γ3-18_G2	CRISPR/Cas9 lines γ-gliadins targeted
	α2γ3-38_G1 α2γ3-38_G4 α2γ3-38_G7 α2γ3-38_G8	CRISPR/Cas9 lines α- and γ- gliadins targeted

1966; obtained from Germplasm Resources Unit, JIC, UK) and four CS Kansas deletion lines (Endo, 1988; Endo and Gill, 1996; obtained from Kansas State University) that had modified gliadin protein profiles with bands missing (obtained from Wheat Genetic and Genomic Resources Centre, Kansas State University, USA), were analysed individually. Four Paragon γ-irradiated lines developed by JIC (Norwich, UK; Shaw *et al.*, 2013) and selected for changes in

gliadin protein profiles (*Chapter 2*) were also included. Eight Fielder T1 lines containing CRISPR/Cas9-induced mutations in α-gliadin protein profiles as visualized on Acid-PAGE, 5 lines containing CRISPR/Cas9-induced mutations in γ-gliadins and 4 lines containing CRISPR/Cas9-induced mutations in α- and γ-gliadins (*Chapter 2*) were tested as well.

DNA digestion

α-gliadins are clustered as repeated sequences in one region on the homoeologous chromosomes 6. Separating the DNA of the consecutive genes is required to avoid introducing bias into the ddPCR binary detection system, as both one or multiple consecutive copies in a single droplet would produce one positive droplet. ApoI was identified as a unique restriction site located within the α-gliadin gene sequence and highly conserved across the different copies (Fig. 2). Wheat genomic DNA was digested using 0.5 µl of *Apo*I restriction enzyme (NEB), 2 µl of 10x NEB buffer, 500 ng of genomic DNA and Milli-Q water to adjust to the final reaction volume to 20 µl. The digestion mix was incubated at 37°C for 2 hours and inactivated at 65°C for 20 min, following the ddPCR protocol of Bio-Rad (rather than 80°C as recommended by NEB for *ApoI*-HF). The digestion was then diluted ten times in all the samples, to reduce the salt concentration.

Primer design for ddPCR multiplexing

Primers designed for ddPCR must have a length between 17 and 25 nucleotides, an annealing temperature between 50°C and 70°C and an amplicon length between 50 and 200 nucleotides (Hindson *et al.*, 2011; Bio-Rad 2018). Multiplexing primer pairs with varying amplicon length necessitates that all primer pairs should have a similar annealing temperature, be designed on highly conserved regions of the genes to amplify, and that each amplicon preferably has ~50bp difference from the other amplicons (Bio-Rad, 2018). The degenerated primers for the multiplex were therefore manually designed, taking into account only existing sequence combinations present in the gliadin database (*Chapter 2*). Each primer was checked for absence of self-complementarity and primer dimer formation with other primer pairs using the online tool Multiple Primer Analyzer (Thermofisher). Primer specificity was checked by blasting in Ensembl! Plant against the *Triticum aestivum* database. The absence of ApoI restriction sites between the primer binding sites was verified.

Three primer pairs could be designed fulfilling all the criteria (Table 2). For the *TaPFT1* reference gene, the primer pair Ref_181 was designed to amplify all three homoeologous variants, giving amplicons of 181bp, based on the Chara sequences used by Fitzgerald *et al.* (2010). For the α-gliadin genes, two different primer pairs were designed. Primer pair Epi_61 binds immediately upstream of the epitope regions at the position targeted by the sgRNA_α213 in most Fielder CRISPR/Cas9 lines, giving amplicons of 61 bp in nearly all gliadin genes if the Cas9 has not induced a mutation. Primer pair UD_136 is complementary to the unique domain in the centre of the gene, a region that is not targeted by any sgRNA in Fielder CRISPR/Cas9 lines, located more than 240 bp downstream Epi_61 reverse primer. It produces an amplicon of 136 bp (Fig. 2).

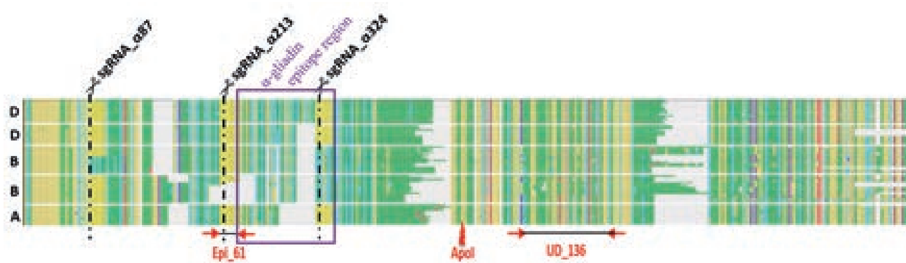


Figure 2. Alignment of α-gliadin sequences, localisation of ddPCR primer pairs relatively to sgRNAs targets and ApoI restriction site.
In this depiction of a subset of the sequence alignment of over 1200 sequences (*Chapter 2*), each block corresponds to a group of α-gliadin sequences identified across the three homoeologous chromosomes 6A, 6B, and 6D. It is unknown how many genes and of which group are present and/or expressed in each accession or variety used in this study. Therefore, the ApoI restriction site was selected as it is located in a conserved part of the proteins, and the Epi_61 and UD_136 ddPCR primer pairs were designed as to bind to the most conserved parts of the genes, although even with the use of a few degenerate sites it was impossible to design primer pairs that would anneal all possible genes. Epi_61 binds close to sgRNA_α213 target site, while UD_136 binds at a non-targeted site over 240 bp downstream, in order to be able to discriminate between small indels (which will affect only Epi_61) and large deletions over 240 bp (which will affect both primer pairs). The picture is adapted from *Chapter 2*.

This design was aimed at enabling a comparison of the number of amplicons generated at sgRNA-targeted and non-targeted sites in order to deduce the (relative) number of small indels generated by sgRNA_α213 with a local effect, compared to the number of large deletions. In the case of CRISPR/Cas9 lines, the decrease of amplicon number correlates with the presence of small indels or larger deletions, but as we do not know the length of the large deletions this does not necessarily indicate a decrease of full-length gene copy number. When

used to study deletion lines or γ-irradiated lines, known to contain deletions of several megabases, there is a direct correlation between the decrease of amplicon number and the decrease of gene copy number present in the plant.

Table 2. ddPCR primers used for CNV detection

Primer name	Primer sequence (5'- -3')	Tm (°C)	Primer length (bp)	CG%	Amplicon length (bp)
F-Epi_61	GCAACCATTTCCATCACAACA	59.9	21	42.9	61
R1-Epi_61	AAGGCGTCGGCGTTGAT	61.5	17	58.8	61
R2-Epi_61	AAGGtGTCGGCGTTGAT	58.1	17	52.9	61
F1-UD_136	TTGCAACAACACAGCATAGC	59.8	20	63.0	136
F3-UD_136	TTGCAgCAACACAaCATAGC	59.8	20	63.0	136
R2-UD_136	GGTcCGGTAgGTgTTACAaC	61.7	20	55.0	136
R3-UD_136	GGTtCGGTAgGTgTTACAaC	59.9	20	50.0	136
F_Ref_181	GCAGAGTGGAGTGGACATTT	60.1	20	50.0	181
R_Ref_181	CTCTATGGTGGCTGTTGCTA	59.4	20	50.0	181

Primer sequences with only upper-case characters cover the most common haplotype while lower-case characters mark haplotypes with a lower frequency.

The two degenerate primers in the same orientation were mixed 1:1, and forward and reverse primers were then pooled in equal volume, representing a degenerated primer pair mix with a concentration of 1µM. All degenerated primer pairs were then tested individually and multiplexed, using basic PCR with a gradient ranging from 55°C to 65°C, on ApoI digested DNA pool from Chinese Spring, Fielder and Paragon.

ddPCR protocol

The ddPCR reaction mixture was prepared using 2-3 µl of degenerated primer pair mix (depending on the primer combination) at 1 µM, 12 µl of ddPCR EvaGreen Supermix (Bio-Rad), 10 ng of 1:10 diluted digested DNA and Milli-Q water in a final volume of 23.6 µl per sample. Of this mix, 20 µl was transferred into a sample well of a DG8 Cartridge for QX200 droplet generator™ and 70 µl of Droplet Generation Oil for EvaGreen chemistry (Bio-Rad) was loaded into the corresponding oil well of the cartridge, for each sample. The cartridge was sealed with a rubber gasket (Bio-Rad) and transferred into the QX200 droplet generator (Bio-Rad). A final volume of 40 µl of droplet-partitioned samples, containing approximately 20,000 droplets, was generated for each sample. Each 40 µl sample was transferred into a 96-well plate (Bio-Rad), sealed and

transferred onto a regular Thermocycler (Bio-Rad). The PCR protocol used was: 5 min at 95°C, 40 cycles of 30 sec at 95°C and 60 sec at 56°C, followed by one step of 5 min at 4°C and a final step of 5 min at 90°C followed by infinite hold at 2°C. The ramp rate for the entire run was 2°C/s. After the PCR, the plate was transferred in the QX200 reader (Bio-Rad) for reading and counting positive and negative droplets.

Data acquisition

Droplet count was performed sample per sample by the QX200 reader (Bio-Rad) linked to the QuantaSoft Software™ (Bio-Rad) for data acquisition and posterior analysis. The software was run on "absolute gene copy number" with detection channel adapted for EvaGreen QX200 ddPCR florescence. Default parameters were used, except for the "Experiment Type" that was set to "Copy Number Variation 6" and "the ploidy" of the sample was entered for the reference for each well of the 96-well plate. The first "target" was set as unknown and the second one as reference. Samples with fewer than 10.000 droplets were excluded. In the output, each droplet from a sample is represented by a dot and each dot is plotted according to its fluorescence amplitude. Dots therefore clustered in clouds or populations based on their fluorescent amplitude, which is determined by the presence or absence of specific DNA template to amplify in each droplet.

Analysis of ddPCR output

After acquiring the data in the QuantaSoft Software™ (Bio-Rad), the representation of the amplitude of the droplets signal as a 1D dot-plot was used to check whether there was clear distinction between the background (negative droplets without the specific DNA fragment to amplify) and the specific DNA fragment amplified (positive droplets with signal amplitude related to the size of the amplicon associated to each fragment to amplify) (Fig. 3a). The thresholds of discrimination between background and amplicon groups were adjusted when necessary. The representation of the ampliude of the droplets signal as a 2D dot-plot was used to identify four dot-clusters associated to droplets containing no amplicon, Epi_61 or UD_136 amplicons, Ref_181 amplicons and multiple-amplicons, based on their respective signal amplitude correlated with their length (Fig. 3b). Once the populations are defined, Poisson statistics are applied and yield an absolute measure of the different amplicon concentration, necessary to calculate the gene copy number present in the sample. The 95%

confidence intervals given by the software (Hindson *et al*., 2011) are used, and the standard deviation of the confidence intervals difference was calculated.

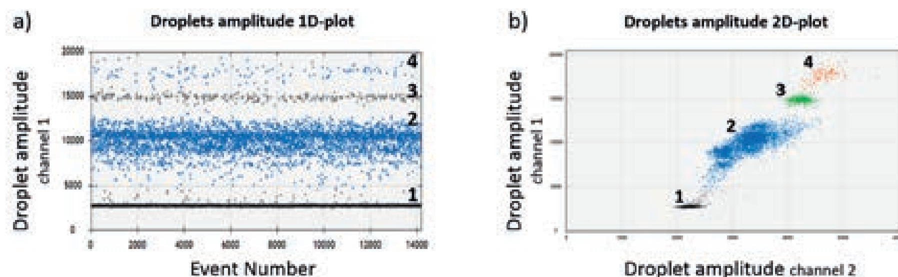


Figure 3. Example of the multiplex assay using two primer pairs, Epi_61 and Ref_181.
a) 1D plot view. b) 2D plot view. Effect of amplicon length on the signal amplitude of droplets with (1) no-template, (2) Epi_61, (3) Ref_181 and (4) Epi_61+ Ref_181.

Results

ddPCR optimisation experiments

Four experiments were carried out using pooled gDNA of the varieties Chinese Spring, Fielder and Paragon and digested by *ApoI* in order to test the possibility of multiplexing the reference gene primer pair with the α-gliadin genes primer pairs, to define the optimal primer annealing temperature, to determine the optimal amount of digested gDNA, and to optimize the distinction of the amplicons, using different primer concentrations.

Multiplexing primer pairs

Multiplexing primer pairs generating different amplicons is often challenging in ddPCR. Hence, we first tested and compared the three primer pairs Epi_61, UD_136 and Ref_181, individually and together. Two duplex experiments were performed, combining either Epi_61 or UD_136 with Ref_181, and a triplex experiment combining all three primer pairs. The same quantity of each degenerated primer pair mix at 1μM was used in each experiment (triplex thus has 3 time more primers concentration). Standard ddPCR protocol was followed, but using 10 ng of gDNA and an annealing temperature of 60°C.

Each primer pair tested individually resulted in good DNA fragment amplification and a clear distinction from the background signal (Fig. 4d). The longer the

amplicons the higher the signal amplitude of the droplets. Epi_61 or UD_136, that both generate α-gliadin gene amplicons, showed a much larger cloud of dots than Ref_181. This is related to a high variability of signal amplitude per droplet caused by the large sequence length variation of this gene family. When duplexing the reference with one of the α-gliadin primer pair, it was possible to distinguish both amplicons from each other and from the multiple amplicon-containing droplets with an artefact signal (Fig. 4a, b, d). The absolute amplicon copy numbers obtained in simplex and duplex experiments were not significantly different. When triplexing the reference with both α-gliadin primer pairs, the signal amplitude of the droplets overlapped (Fig. 4d, last column), which impeded distinguishing the different amplicons from one another quantitatively (Fig. 4c). Therefore, we decided to use duplexing of the reference primer pair and one of the α-gliadin primer pairs throughout the study.

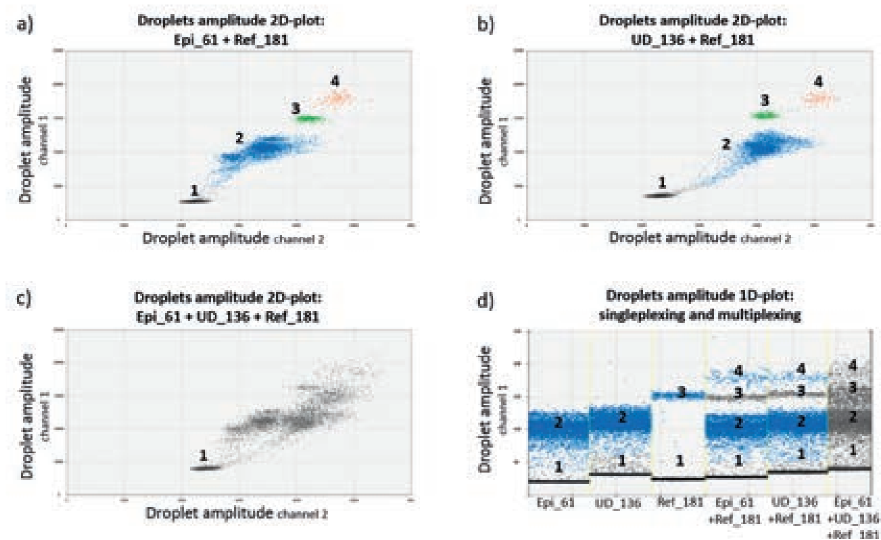


Figure 4. Simplex, duplex and multiplex assays.
a) and b) 2D views of the two duplexing primer combinations (gliadin (blue) and Ref_181 (green)). c) 2D view of multiplexing Epi_61, UD_136 and Ref_181. d) 1D plot of all simplex, duplex and multiplex combinations. (1) Negative droplets, (2) α-gliadin positive droplets, (3) reference positive droplets and (4) multiple amplicons-containing droplets.

Optimal annealing temperature

To determine the optimal annealing temperature for duplex ddPCRing of reference and either α-gliadin primer pairs, a thermal gradient experiment was performed with eight different annealing temperatures (Fig. 5). The standard ddPCR protocol was used with 10 ng of gDNA. The best separation of the different

amplicon groups was observed with an annealing temperature between 55°C and 57°C, for both Epi_61 + Ref_181 and UD_136 + Ref_181 duplexing assays (Fig. 5). 56°C was selected as optimal annealing temperature for further reference and α-gliadin duplex ddPCR experiments.

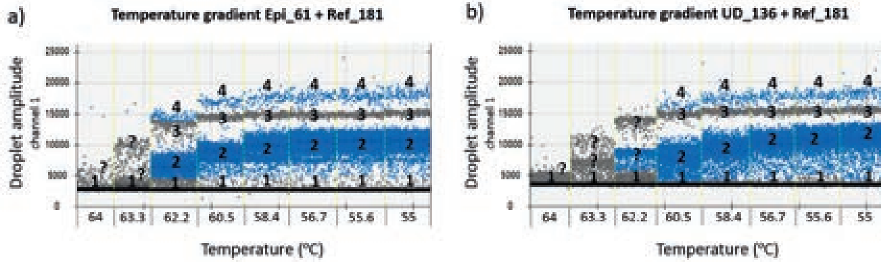


Figure 5. 1D plot view of the annealing temperature thermal gradient.
Representing the negative droplets (1), amplicon positive droplets (2), reference positive droplets (3), double positive droplets (4) and undetermined droplets (?). Each sample annealed at 64, 63.3, 62.2, 60.5, 58.4, 56.7, 55.6 and 55 ° C, from left to right, duplexing with a) Epi_61 and Ref_181 or b) UD_136 and Ref_181.

Optimal DNA input

In ddPCR, the amount of input DNA is determined by the optimal copy number of DNA fragment to amplify, also called "sweet spot". The sweet spot stands for around 1,500 DNA fragment copies per µl or 30,000 copies for 20 µl of reaction. We calculated that 10.4 ng of DNA per reaction corresponds to the theoretical sweet spot, based on the estimated hexaploid wheat genome mass (34.66 pg, according to Kim *et al.*, 2015; Garcia *et al.*, 2013) and the estimated α-gliadin gene copy number (100 copies, according to Ozuna *et al.*, 2015), while including sufficient single copy reference gene in the reaction. To confirm this sweet spot, ddPCR experiments were performed for the range of 2 ng to 50 ng of digested DNA per sample. It was possible to distinguish enough reference gene amplicon and to remain within the detection limit of the droplet reader using any tested concentration. Nevertheless, extremely low or high DNA concentration hampered the visualization of the output and the distinction of amplicons in duplex experiments (not shown). The α-gliadin gene copy number measured using duplex ddPCR was stable between 10 and 30 ng of DNA input per assay (Table 3). The amount of digested wheat gDNA was therefore set to 10 ng per 20 µl of ddPCR reaction.

Table 3. Comparison of α-gliadin gene copy number measured with duplex ddPCR with different amounts of DNA input.

DNA (ng)	α-gliadin gene copies
50	50.9
40	54.3
30	58.2
20	61.5
10	58.2
5	56
2	41

Optimal primer concentration

The general primer concentration used was 100 nM of each primer pair. Increasing the concentration of the primer rises the amplitude of the respective amplicons and can be used for optimizing the difference between the amplicon signals. When multiplexing UD_136 and Ref_181, the distinction between both amplicon groups was insufficient in some genotypes with high α-gliadin amplicon size variability. Increasing the Ref_181 primer pair concentration to 150 nM improved the discrimination of the populations (Fig. 6a) without significantly altering the copy number detected (Fig. 6b). Meanwhile, increasing primer concentration rises the signal amplitude of the background. Subsequently, a Ref_181 concentration of 150 nM instead of regular 100 nM was used for those genotypes in which the Ref_181 signals overlap with those of the UD_136 α-gliadin amplicons.

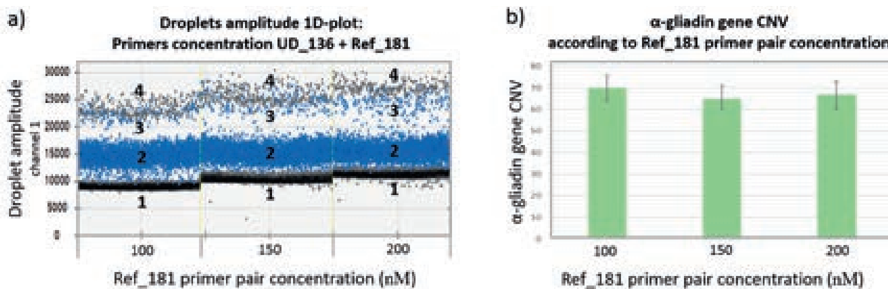


Figure 6. Primer concentration optimization.
a) 1D plot view of the effect of increasing Ref_181 primer pair concentration. Representing negative droplets (1), UD_136 positive droplets (2), Ref_181 positive droplets (3) and multiple-amplicons positive droplets (4). Increasing the Ref_181 concentration increase the separation between UD_136 and Ref_181 positive droplets. Increasing Ref_181 concentration also increases the amplitude of the negative droplets. b) α-gliadin gene CNV was independent of the Ref_181 primers concentration tested. Error bars indicate the Poisson 95% confidence intervals for each copy number determination.

Technical replicates

Technical replicates were used within a plate and across plates during the optimisation run. Replicates within a plate showed no significant differences,

although some failed for technical reasons. Replicates across plates showed higher variation than within a plate, which were sometimes significant. Although this could be corrected for using a "landmark sample" present in all plates, we decided to limit the comparisons to samples on the same plate for higher accuracy.

Measuring the α-gliadin gene copy number in wheat for the three homoeologous chromosomes

The duplex ddPCR reveals the α-gliadin amplicon number present in a hexaploid wheat line. This amplicon number directly correlates with the α-gliadin gene copy number in case of wheat lines, unless they would contain small indels induced in the amplicon region (see below). With the ddPCR method we can thus determine the total number of α-gliadin genes in bread wheat, but using selected genetic material we should also be able to determine the number of genes on each three pairs of homoeologous chromosomes 6 (written as 6AA, 6BB, 6DD, to emphasise that the numbers concern the chromosome pairs and not the individual chromosomes). For this we compared the number of α-gliadin genes in Synthetic Hexaploid Wheat (SHW) and their parental lines, as well as in variety Chinese Spring and various deletion and nullisomic-tetrasomic lines derived from it.

Synthetic Hexaploid Wheat lines

Synthetic Hexaploid Wheat (SHW) lines have been created by crossing tetraploid *Triticum turgidum (*AABB) with diploid *Aegilops tauschii* (DD) followed by embryo rescue and spontaneous chromosome doubling to obtain hexaploid synthetic lines (AABBDD) (Ogbonnaya *et al*., 2013). The α-gliadin gene copy number of the five parental lines and their three resulting SHW lines were analysed using duplex ddPCR. The α-gliadin gene copy number was 14-17 in the diploid DD genome accessions, 70-76 in the tetraploid AABB genome, and 86-95 in the hexaploid SHW lines. Generally, the sum of gene copy number measured in the tetraploid parent and the diploid parent lines was not significantly different from the copy number measured in the resulting SHW plant. This consistency in gene copy number measured reflects the accuracy and the robustness of ddPCR for gene CNV studies in large gene families in polyploids (Table 4).

Chinese Spring Wheat deletion lines

Chinese Spring (CS) deletion lines (Endo, 1988; Endo and Gill, 1996) carry a homozygous large deletion in one pair of homoeologous chromosome 6 (short arm),

Table 4. Comparison of measured and expected CNV of the SHW lines (AABBDD).

SHW lines name	Measured in Parent AABB	Measured in Parent DD	Measured in SHW AABBDD	Sum of measured Parental AABB + DD
NIAB SHW 54	76	15.5	95	91.5
NIAB SHW 75	70	17.2	86	87.2
NIAB SHW 76	70	13.5*	87	83.5

Measured α-gliadin gene copy number in parental AABB and DD lines and their corresponding AABBDD SHW lines compared to the sum of α-gliadin gene copy number measured in both parental AB and D lines. *CNV calculated using the average Ref_181 concentration of the two other DD lines as the reference gene did not amplify in Ae. tauschii Ent 405, presumably due to reference sequence variations.

where α-gliadin genes are located. The 6DS-4/1BS-19 line was generated by a cross between 6DS-4 and 1BS-19 deletion lines which lacks γ- and ω-gliadin genes on the short arm of chromosome 1. Using duplex ddPCR, 61 α-gliadin gene copies were found in hexaploid (AABBDD) CS, while all deletion lines showed a significant reduction in α-gliadin gene copy number compared to CS (Table 5), with 42 to 52 copies remaining (Fig. 7).

The 15 genes missing in the 6DS deletion are consistent with the number estimated in the diploid DD genome donor of the SHW lines (Table 4). We estimate from the data that in 6AS-1, 6BS4/5BS-2 and 6DS-4 deletion lines respectively 23, 17 and 15 α-gliadin genes were missing (Table 5). The addition of these genome-specific α-gliadin genes is 55 copies while 61 copies were measured in CS. This difference could be explained by variations of the confidence interval or by copies still present at *glia-2* loci, if the deletions did not include all genes.

Remarkably, deletion line 6DS-4/1BS-19 lacks 11 gene copies more than 6DS-4. If the identification of the line is correct, it would suggest 11 α-gliadin-like sequences on 1BS (Table 5).

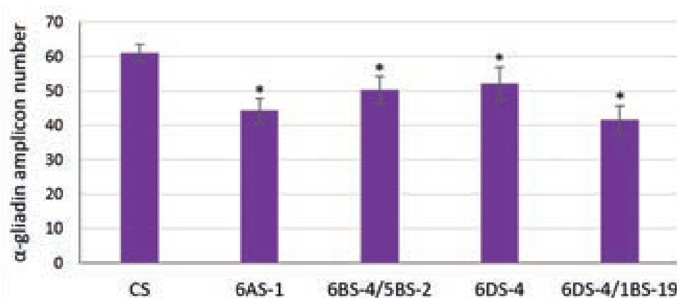


Figure 7. α-gliadin gene copy number in CS and in four CS deletion lines.
Duplex PCR using Epi_61 and Ref_181 primer pairs on hexaploid CS and CS lines carrying a deletion in one of the homoeologous chromosome 6. Error bars indicate the Poisson 95% confidence interval. Stars indicate significant amplicons reduction compared to Paragon based on standard deviation of confidence intervals difference.

Chinese Spring Wheat nullisomic-tetrasomic lines

CS nullisomic-tetrasomic lines (Sears, 1966) were made by removing one pair of an homoeologous chromosome and replacing it by one pair of another homoeologous chromosome. We used the lines that miss one pair of homoeologous chromosome 6 but have, instead, two pairs of one of the other homoeologous chromosomes 6. For better readability, we renamed the nullisomic-tetrasomic lines based on the homoeologous genomes they contain. For instance, the line Nullisomic for 6A but Tetrasomic for 6D is originally called N6AT6D and contains 6DDDDBB which we renamed 6DDB. Using ddPCR, CS revealed to have 61 α-gliadin gene copies, consistent with the deletion lines study. In the nulli-tetra lines, gene copies were estimated to be 47 in 6DDB, 50 in 6BBD, 66 in 6AAB and as many as 111in 6DDA (Fig. 8).

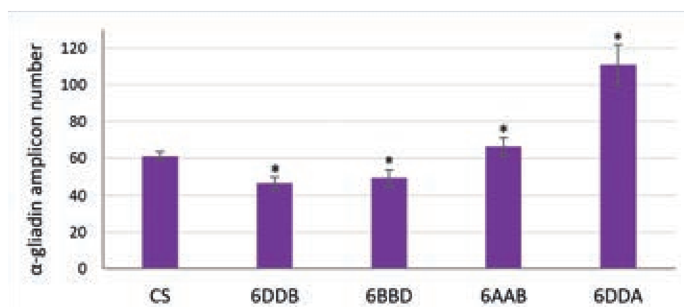


Figure 8. α-gliadin gene copy number in CS and in four related CS nullisomic-tetrasomic lines.
Duplex PCR using Epi_61 and Ref_181 primer pairs on CS lines lacking one of the homoeologous pairs of chromosome 6 and possessing two pairs of another homoeologous chromosome. Error bars indicate the Poisson 95% confidence. Stars indicate significant amplicons reduction compared to Paragon based on standard deviation of confidence intervals difference.

The α-gliadin copy number associated to each homoeologous chromosome in CS was calculated by solving the three functions with three variables resulting in the total copy number measured for three different combinations of nulli-tetra lines (Calculation 1).

DDB = 47
B = 47 – 2D
B = 47 – 2 * (50 - 2B)
B = 47 – 100 + 4B
(100-47) / 3 = B
17.6 = B

BBD = 50
D = 50 – 2B
D = 50 – 2 * (47 – 2D)
D = 50 – 94 + 4D
(94 – 50) / 3 = D
14.6 = D

AAB = 66
66 – B = 2A
66 – 17.6 =2A
24.2 = A

Calculation 1. Calculation of α-gliadin gene copy number in each homoeologous genome.

In nulli-tetra line 6DDA – that contains 6DDDDAA -, the reference did not amplify well, most likely due to a mutation in the reference sequence. Therefore the average reference copy number of the other three lines was used and 52 α-gliadin gene copies were calculated. This is consistent with 6AA + 6DD + 6DD = 53.4 α-gliadin gene copies, based on Calculation 1 values.

No significant differences were found between the results obtained with nulli/tetra CS lines and CS deletion lines (Table 5), meaning that the full complement of genes on the *glia-2* loci was deleted from the specific homoeologous chromosome in the deletion lines. The 15 copies obtained for the DD genome using the nulli-tetra lines was similar to the number obtained using the 6DS-4 deletion line containing only deletion in the DD genome. However, it did not match with the 26 copies from 6DS-4/1BS-19 double deletion line.

Table 5. α-gliadin gene copies per genome estimated in CS nulli/tetra and missing in CS deletion lines compared to CS

Diploid genome	Nulli-Tetra lines	Deletion lines
AA	24.2	23.2
BB	17.6	17.1
DD	14.6	15.3 (25.6*)
AA + BB + DD (CS WT estimated)	56.4	55.6
AABBDD (CS WT measured)	61	61

* Includes additional deletion in chromosome 1 short arm.

3

Quantification of the intact α-gliadin gene copy number after random or targeted mutagenesis

Mutagenesis was used to mutate or delete α-gliadin genes, in order to reduce the immunogenicity for people with coeliac disease (*Chapter 2*). Using Acid-PAGE, lines were selected in which the α-gliadin profile was modified, most frequently in the form of bands that are absent. However, this is a qualitative method, and it does not see small insertions and deletions that only change a few amino acids. With ddPCR method we set out to determine the quantity of the mutations and deletions induced, both in lines from random mutagenesis – obtained using γ-irradiation - and in lines produced with targeted mutagenesis – generated using CRISPR/Cas9-. In the case of the Paragon γ-irradiated lines, where the deletions may be megabases in size, the decrease in α-gliadin amplicon number is correlated with

the decrease of the α-gliadin full-length gene copy number. In contrast, in case of the Fielder-CRISPR lines, where mutations often consist of relatively small indels (up to a few hundred bases), the decrease of amplicons generated correlates to the presence of mutations or deletions and can but does not necessarily indicate a decrease of the full-length gene copy number. The term amplicon number is therefore used here, instead of gene copy number.

Random mutagenesis: Paragon γ-irradiated lines

Four Paragon γ-irradiated lines that showed loss of bands in the α-gliadin protein profile on the Acid-PAGE, were tested for α-gliadin gene copy number loss. Two duplex ddPCR assays were implemented using the internal reference (Ref_181) and either the primer pair located on (Epi_61) or downstream (UD_136) the α-gliadin epitope region.

All four Paragon γ-irradiated lines showed a significant reduction of α-gliadin gene copies (Fig. 9). In comparison to the 61 α-gliadin gene copies measured in Paragon, P3-75 and P6-57 showed a 16-18 copy number reduction while P6-43 and P6-59 carried as many as around 30 copies less. There are no significant differences between results obtained with both ddPCR primer pairs in Paragon, which indicates that they do not differ in specificity with regard to producing amplicons in this multigene family. The absence of significant differences between both primer pairs in irradiated lines indicate a similar reduction of amplicons both at epitope and conserved region of the α-gliadin genes, which is likely caused by full gene or loci deletion.

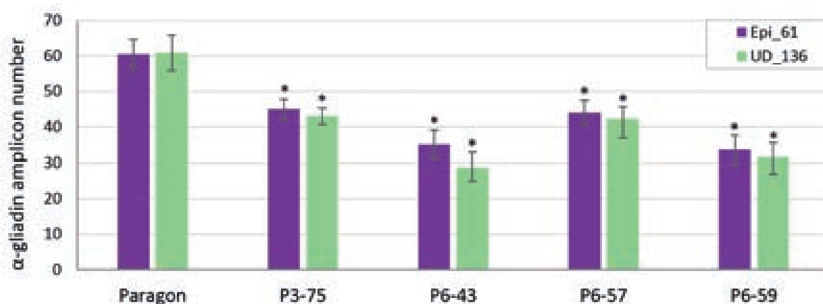


Figure 9. ddPCR results of α-gliadins gene copy number in Paragon wild type and γ-irradiated lines.
Two multiplex ddPCR assays, with Epi_61 or UD_136 α-gliadin primer pairs and Ref_181, were performed on Paragon and 4 of the γ-irradiated lines pre-screened for α-gliadin protein profile alterations. The absolute α-gliadin gene copy number estimations with the two primer pairs were not significantly different within all samples. Error bars indicate the Poisson 95% confidence intervals. Stars indicate significant amplicons reduction compared to Paragon based on standard deviation of confidence intervals difference.

Targeted mutagenesis: Fielder CRISPR/Cas9 lines

Targeted mutagenesis using CRISPR/Cas9 was done with sgRNAs that focussed on regions near or in coeliac disease epitopes, or upstream of these epitopes. For a total of 18 Fielder CRISPR/Cas9 lines that showed an altered gliadin protein profile on Acid-PAGE, the α-gliadin gene copy number was determined quantitatively using ddPCR with two different primer pairs, one near the epitope region (Epi_61) and one 240 nt downstream (UD_136) (Fig. 10). Both would detect a complete loss of the gene, but the first one is also expected to be affected by small indels and medium deletions as it is near the target site from the two sgRNAs targeting α-gliadins, present in most tested plants.

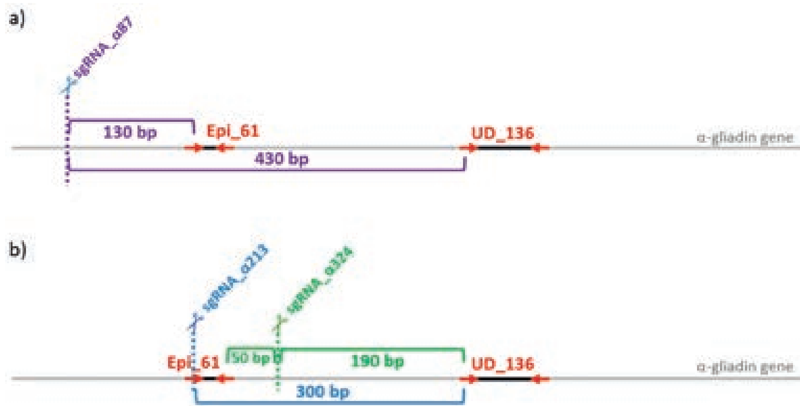


Figure 10. Distance from sgRNA target sites to Epi_61 or UD_136 primer pairs in α-gliadin gene.
The scheme represents an α-gliadin gene, the sgRNA target sites and the position of Epi_61 and UD_136 ddPCR primer pairs. a) α1 construct with only sgRNA_ α87; b) α2 and α2γ3 construct sgRNA_ α213 and sgRNA_ α324 targeting α-gliadin genes. The distance between each sgRNA target site and the two ddPCR primer pairs is reported. These distances correspond to the minimal length of a deletion induced by sgRNAs that would prevent at least one primer to anneal to the DNA, causing amplicon number reduction compare to the wild type. Relative amplicon number reduction detected by duplex ddPCR assays is therefore interpreted as number of deletions from a minimal size. The primer pair adjacent to a sgRNA target site will detect small indels as well as larger deletions while the primer pair distant from the sgRNA target site will only detect large deletions. The difference of amplicon number reduction between UD_136 and Epi_61 gives the number of smaller mutations. However, in lines harbouring the α2 and α2γ3 constructs (panel b), UD_136 detects large deletions but these may occur with (when initiated from sgRNA_ α213) or without (sgRNA_α324) affecting the Epi_61 amplicon. A subtraction of the two amplicons may thus underestimate the number of small indels.

Unlike in Paragon, the number of gene copies detected in Fielder was slightly different for the two amplicons: 87 with Epi_61 and 82 with UD_136. This may be due to SNPs in the primer annealing positions for UD_136 in some of the gliadin genes. The number of gene copies of the samples were subsequently

standardized to 87 copies in Fielder for both primer pairs. The standardization was also applied to all the CRISPR lines by adding 5 copies from the number obtained with UD_136.

Among the Fielder-CRISPR lines tested, one was transformed with sgRNA_α87 only (α1 construct), targeting α-gliadins far upstream the epitope region. The target site is around 130 bp of the Epi_61 and 430 bp upstream UD_136 ddPCR primer annealing sites (Fig. 10a). The significant reduction of 26 amplicons detected with Epi_61 duplex ddPCR and 14 amplicons with UD_136 duplex ddPCR showed that more than 1/3 of the initial 87 α-gliadin copies in Fielder underwent a large deletion of at least 130 bp, half of which have a deletion larger than 430 bp.

Eight Fielder-CRISPR lines were derived from plants transformed with both sgRNA_α213 and sgRNA_α324 (α2 construct), which both target α-gliadin epitopes. The Epi_61 primer pair annealing sites are located at the sgRNA_α213 target and 50 bp upstream of sgRNA_α324, therefore amplicon number variation detected with the Epi_61 duplex ddPCR concerns both small indels as well larger deletions (Fig. 10b). The UD_136 primer annealing site is located 300 bp downstream of that of sgRNA_α213 and 190 bp downstream of sgRNA_α324, therefore amplicon number variation in UD_136 duplex ddPCR would only reflect large deletions (Fig. 10b). These eight lines containing sgRNA_α213 and sgRNA_α324 showed a significant reduction of 17 to 24 Epi_61 amplicons detected compared to the 87 present in Fielder, suggesting 20% of the α-gliadin copies having mutations that can be small indels or larger deletions. A lower but still significant reduction of 7 to 13 UD_136 amplicon was observed in six of the eight lines, indicating the presence of 12% of large deletions in these lines. The significant reduction of amplicons detected with Epi_61 and UD_136 duplex ddPCR indicates that about 1/2 of the α-gliadin mutations are relatively small indels.

Four CRISPR lines were derived from plants transformed with the α2γ3 construct, which included an additional three sgRNA targeting γ-gliadin epitopes. All four lines revealed a significantly lower amplicon number measured by both duplex ddPCR compared to plants with the α2 construct. Epi_61 amplicons decreased as much as 32 to 38 copies, while UD_136 amplicons decreased 19 to 23 copies (Fig. 11). The significant reduction of amplicons detected with Epi_61 and

UD_136 duplex ddPCR indicates that roughly 40-50% of the deletions may be large, while the remainder may be smaller indels.

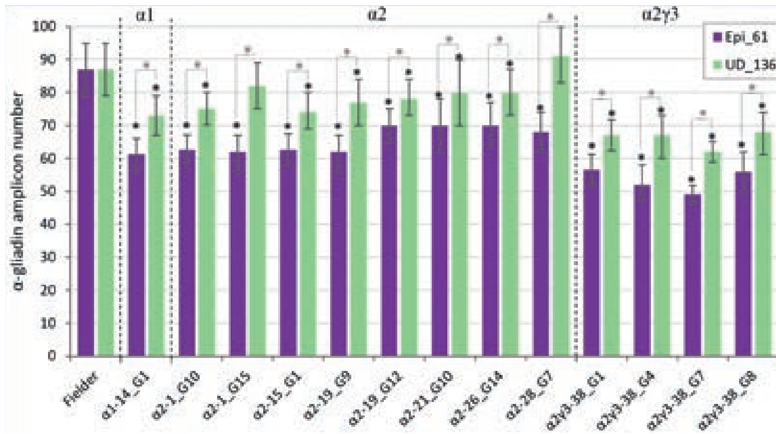


Figure 11. ddPCR results of α-gliadin gene amplicon number in Fielder and CRISPR/Cas9 lines with targeted α-gliadins.
Two multiplex ddPCR assays, with Epi_61 or UD_136 α-gliadin primer pairs and Ref_181, were performed on Fielder and 13 of the CRISPR/Cas9 lines with sgRNA targeting α-gliadins (α1 and α2 constructs) or both α- and γ-gliadins (α2γ3 construct), having been pre-screened for α-gliadin protein profile alterations on Acid PAGE. Error bars indicate the Poisson 95% confidence intervals. Black stars indicate significant amplicons reduction compared to Paragon and grey stars indicate significant difference between amplicons reduction obtained with both primer pairs, based on standard deviation of confidence intervals difference.

Five additional Fielder-CRISPR mutant lines were analysed that contained the γ3 construct, composed only of the three sgRNAs targeting γ-gliadins. Even though no α-gliadins were targeted, the α-gliadin amplicon number was significantly lower for both α-gliadin ddPCR primer pairs in all lines, with no significant differences between both primer pairs in most cases (Fig. 12). An average decrease of 20 α-gliadin Epi_61 amplicons was observed in the mutant lines, which is ~20% of the total α-gliadin copy number found in Fielder. Line γ3-1_G6 lacked 29 and 24 α-gliadin amplicons when using Epi_61 and UD_136 primer pairs, respectively. These reductions of the number of amplicons is close to the 30 copies reduction observed with Epi_61 and the 20 with UD_136 seen in line α2γ3-38_G1 (Fig. 11), in which both α- and γ-gliadins were targeted.

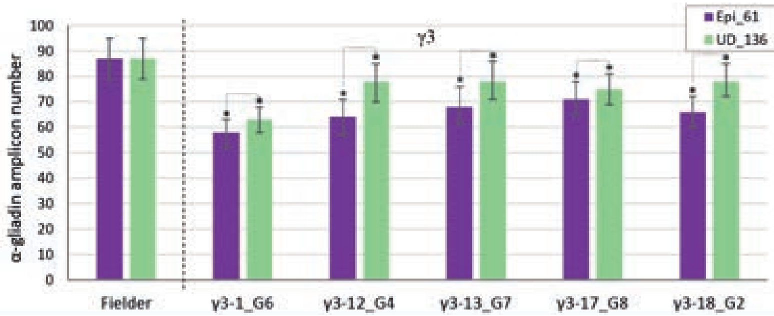


Figure 12. ddPCR results of α-gliadin gene amplicons number in Fielder and CRISPR/Cas9 lines with targeted γ-gliadins.
Two multiplex ddPCR assays, with Epi_61 or UD_136 α-gliadin primer pairs and Ref_181, were performed on Fielder wild type and 5 of the CRISPR/Cas9 lines targeting only γ-gliadins and no α-gliadins (γ3 construct), having been pre-screened for α-gliadin protein profile alterations. Error bars indicate the Poisson 95% confidence intervals. Black stars indicate significant amplicons reduction compared to Paragon and grey stars indicate significant difference between amplicons reduction obtained with both primer pairs, based on standard deviation of confidence intervals difference.

Discussion

The implementation of ddPCR is well established for accurate, reliable and fast gene copy number variation studies. However, the feasibility of this method was yet to be proven for CNVs in large gene families, or for quantifying the total number of gene copies in complex polyploid organisms. Challenges exist with regard to the linearity of the detection and to the sensitivity to detect the relatively smaller differences in copy number when the absolute number is large. We developed the methodology here for the α-gliadin family - previously estimated at approximately 100 copies -, in hexaploid (AABBDD) bread wheat. The linearity was assessed using sets of genetic material with chromosome deletions or replacements, the sensitivity was approached by analysing plants derived from random and targeted mutagenesis. The size of deletions and hence the nature of the gene edits was assessed by using two different amplicons in the α-gliadins.

ddPCR optimisation: internal reference, multiplexing and near-absolute copy number quantification

For ddPCR of α-gliadin genes in hexaploid wheat, we calculated that 10 ng of gDNA would be the "sweet spot", based on information known for single gene copy in humans. This result was confirmed experimentally using a range of input gDNA concentrations. To reinforce the accuracy of CNV, we have included the

*Ta*PFT1 as an internal reference. It is an orthologue of the *Arabidopsis thaliana* Phytochrome Flowering Time 1 gene, which is present once per homoeologous chromosome 5, meaning 6 times per hexaploid wheat genome (Fitzgerald *et al*., 2010). The Ref_181 primer pair was designed on a region conserved across the three homoeologous genomes. The results indicate that the internal standard gene approach works well for wheat, so that this primer pair is advisable for use in ddPCR assays in wheat. However, it must be kept in mind that some accessions may present SNPs in the sequences at the primer annealing positions, impeding the amplification. This was corrected using the average amplicon number of other samples, which appeared to work well. It could also be corrected with degenerated primers after identifying sequence variants.

Multiplexing α-gliadin primer pairs with a reference primer pair increases the accuracy of a ddPCR assay by enabling normalisation and direct amplicon comparisons within and across samples. It also decreases the costs. However, multiplexing requires much stricter experimental design than basic reactions, in order to be able to always clearly distinguish the multiple amplicons in the assay from each other and from the internal reference. Although fluorescently tagged probes are the most efficient for multiplexing assays, they are not compatible with highly variable gene sequences such as the α-gliadin family. We therefore used intercalating fluorescent dye assays to detect the α-gliadin gene sequence variants amplified with degenerate primer pairs. The use of different annealing temperatures to distinguish amplicons was discarded, as degenerated primers already have slightly different annealing temperatures. Increasing the concentration of one of the primer pairs (McDermott *et al*., 2013) increased both amplicons and background signal, increasing the probability of signal overlap. Therefore, we varied amplicon length to be able to multiplex the *Ta*PFT1 reference gene and the α-gliadin studied genes. As our attempt of triplexing the reference primer pair with two different α-gliadin primer pairs failed because of overlap in amplicon signal amplitudes, probably due to some variation in length of both α-gliadin genes, we settled on duplexing the Ref_181 internal reference with either of two α-gliadin primer pair Epi_61 or UD_136. The presence of the internal reference in both reactions increased the accuracy of the measurement while it also enabled the direct comparison of both α-gliadin amplicon numbers in the paired duplex reactions.

The higher accuracy generated by duplex ddPCR of an α-gliadin amplicon together with the reference amplicon was negatively affected by the occurrence of droplets containing multiple-amplicons. When working with amplicon size differences, like signal differences, it is difficult to define the content of the multiple-amplicons droplets correctly. Since the signal of the multiple-amplicons droplets overlapped with that of the reference droplets, they have been counted as references. This bias was present in all samples, thus did not significantly affect the relative gene copy number comparisons. To estimate the absolute copy number bias generated by this phenomenon, samples in which multiple-amplicons could be distinguished were first counted as such and then as references. The consequence was an absolute α-gliadin gene copy number reduction by 4-5 copies (which is in the order of ~5%). In addition, the high sequence variation among α-gliadin genes within a single locus and within a genotype, makes it difficult to amplify all copies with sequence variants. Although degenerated primers were designed to match the most α-gliadin variants using a multiple alignment of over 1200 sequences from various *Triticeae* species and accessions (*Chapter 2*), these primers may not amplify all sequence variants present in the sample. Together, the signal overlap of droplets containing multiple-positive or reference amplicon and the high variability of α-gliadin gene sequences indicate that the duplex ddPCR may give an underestimation of the absolute α-gliadin gene copy number present in the wheat accession studied.

Duplex ddPCR to determine gene copy numbers per homoeologous chromosome pair

The linearity and the detection limit of the method were first tested by duplex ddPCR reactions on Synthetic Hexaploid Wheat (SHW) and their respective parental lines. The diploid *Ae. tauschii* species (DD genome) revealed 14-17 α-gliadin gene copies, depending on the accession. The tetraploid *T. turgidum* (AABB genomes) varieties showed 70-76 α-gliadin gene copies. The small differences among genotypes were expected. Based on these results, the theoretical α-gliadin gene copy number of the SHW (AABBDD) based on the sum of that in the respective diploid (DD) and tetraploid (AABB) parental lines was calculated. The number was experimentally determined by duplex ddPCR on the actual SHW lines. Similar results of 87-95 α-gliadin gene copies were observed using both methods (Table 6), confirming the accuracy of duplex ddPCR and demonstrating the linearity of the duplex ddPCR up to almost 100 gene copies without saturating

the signal detection of the droplet reader. It also shows that the absolute α-gliadin gene copy number could be determined, if there would be certainty that the primers designed amplify all the sequence variants present in that genotype.

The linearity and sensitivity of the duplex ddPCR method were further tested on CS and related deletion and nullisomic-tetrasomic lines, which have different α-gliadin gene compositions as the gliadin loci on the homoeologous chromosomes 6 are missing or present twice. The method was sensitive enough to detect a decrease in copy number of 5-25 genes relative to 61 α-gliadin gene copies in wild type hexaploid CS (AABBDD), which is consistent with the 60 copies estimated in CS using Southern blot (Anderson *et al*., 1997). The reduction of α-gliadin copies in each deletion line enabled us to estimate the number of gene copies present on each pair of homoeologous chromosomes, and this was also confirmed with calculations based on the nullisomic-tetrasomic lines. A total of 24, 18 and 15 α-gliadin gene copies where determined for homoeologous chromosome pair 6 AA, 6 BB and 6 DD, respectively, using both methods. The number of copies per homoeologous chromosome was, for a long time, an open question, as even a genome sequence may not be accurately assembled in these regions with tandemly duplicated genes with highly similar sequences.

Dual duplex ddPCR to identify, quantify and distinguish between indels and large deletions

The validation of duplex ddPCR for α-gliadin CNV in hexaploid wheat was performed using Paragon lines randomly mutagenized with γ-irradiation, and Fielder CRISPR/Cas9 lines, gene edited using sgRNAs targeting α- and/or γ-gliadin gene families. All lines used had an altered gliadin protein profile on Acid-PAGE (*Chapter 2*). Two connected duplex ddPCR assays were employed, which amplified different parts of each α-gliadin genes, to identify, quantify and distinguish between small indels and large deletions.

The four Paragon γ-irradiated lines tested showed a significant reduction of 16-30 α-gliadin gene copies using both duplex PCR compared to the 61 copies found in the Paragon. Even when considering that some α-gliadin genes may not have been amplified by the ddPCR primer pairs, the γ-irradiation has caused a 25% to 50% reduction in the α-gliadin gene copy number. The α-gliadin gene copy numbers reduction measured using the two primer pairs were the same, confirming the

occurrence of large deletions that affected both primer pairs and, very likely, the whole gene. The reductions in copy number were as large as or higher than the number of α-gliadin genes on a single homoeologous chromosome pair in CS (6 AA=24, 6 BB=18, 6 DD=15) which is consistent with the deletion of at least one complete set of α-gliadin genes on one of the homoeologous chromosomes. This is plausible since γ-irradiation generates mostly deletions of large DNA fragments (Morita *et al.,* 2009).

In case of the Fielder-CRISPR lines, the interpretation of the two connected duplex PCR assays is more complex. The concept is based on the fact that small indels would only affect the amplification by the primer pair on target or adjacent to the sgRNA target site while larger deletions would decrease the amplicons number generated by both adjacent and distant primer pairs. The size of large deletions is then estimated as being at least of the distance between the sgRNA target site and the annealing position of the closest primer of a pair (Fig. 10). One Fielder-CRISPR line tested contained sgRNA_α87 (α1 construct) targeting α-gliadins at the beginning of the gene sequence. It showed that 1/3 of the initial 87 gene copies revealed large deletions of at least 130 bp, half of which were larger than 430 bp. The eight Fielder-CRISPR lines with sgRNA_α213 and sgRNA_α324 (α2 construct) showed on average 17 small indels or medium-sized deletions (10 - 50 bp). In addition, in six of the eight lines, large deletions (> 190 bp) were observed, in up to 13 of the 87 α-gliadin genes. The differential effect of both assays appears to work well, as the results were consistent across all lines. On average, small indels (20%) are usually more abundant than large deletions (8.5 %). Four Fielder-CRISPR lines contained the previous two sgRNAs targeting α-gliadins together with an additional three sgRNAs targeting γ-gliadins (α2γ3 construct). These lines all showed on average an additional decrease of 14 small deletions and 16 large deletions, compared to the lines where only α-gliadins were targeted. Thus, the presence of sgRNAs targeting γ-gliadins influences α-gliadin amplicons number. A similar result was found for five additional Fielder-CRISPR lines that only contained the previous three sgRNAs targeting γ-gliadins and no guide RNAs that targeted α-gliadins (γ3 construct). In these five plants, around 20% of α-gliadin small indels or large deletion where detected. In line γ3-1_G6. In this particular line, containing only sgRNA targeting γ-gliadins, the mutations measured is close to the one observed in α2γ3-38_G1 (Fig. 10) where both α- and γ-gliadins were targeted. This result implies that α-gliadin

mutations in this α2γ3-38_G1 lines might only be due to the presence of sgRNA actually targeting γ-gliadins.

The fact that sgRNAs targeting γ-gliadins affect similarly the α-gliadin Epi_61 and UD_136 amplicon number is surprising. One possible explanation is that the ddPCR primers partly amplify γ-gliadins, despite the specificity of the primer sequences checked using genome Blast. Indeed, due to duplexing constrains, the primer annealing temperature used in the duplex ddPCR was only 56°C, but on gel only one specific band of the expected length of 61 bp and 136 bp respectively are seen after amplification with annealing temperature between 50°C and 60°C. Another explanation is sgRNA off-target, where γ-gliadin-targeting sgRNAs would wrongly target α-gliadins, despites the absence of complementarity of the sgRNA with other gliadin gene families checked using genome Blast.

Independent evidence for an effect of γ-gliadins on chromosome 1 comes from the comparison between CS deletion lines 6DS-4 and 6DS-4/1BS-19. The latter one, obtained from crossing 6DS-4 missing α-gliadin with 1BS-19 missing gliadins from other families, unexpectedly showed less α-gliadin copies than in its parental line 6DS-4. The only known difference between the lines is an additional deletion in chromosome 1B, short arm, and this suggests the presence there of sequences that are amplified by Epi_61 and UD_136 primer pairs. Furthermore, a Blast of the primers into Ensembl plant *T. aestivum* online database gave only two hits on chromosome 1 with 95% homology but these were not significant (E value = 2.00). However, many hits were found in scaffolds that had not been not assigned yet to any chromosome, with significant E-values as low as $8.4\ 10^{-3}$. These results suggest the possible presence of α-gliadins-like sequences on chromosome 1, short arm near *glia-α1* loci encoding for γ-gliadins. This is not unrealistic from an evolutionary point of view, since α-gliadins were originally located on chromosome 1, short arm before being translocated to chromosome 6, short arm (Payne *et al.*, 1982; Shewry *et al.*, 1984). A hypothesis for the mechanism observed here is that large deletions of γ-gliadins lead to losses of intercalated or adjacent α-gliadin-like genes.

To test whether the hypothetic presence of α-gliadin-like sequences on chromosome 1BS occurs for each homoeologous genome, CS deletion lines that related to deletions on chromosome 1 and that show gliadin protein profile changes

on Acid-PAGE, should be analysed using duplex ddPCR with the α-gliadin Epi_61 primer pair. Another way of verifying the theory would be using duplex ddPCR with primer pairs amplifying γ-gliadins. A reduction of γ-gliadins amplicons in Fielder-CRISPR lines containing the α2 construct targeting only α-gliadins may also lead to deletion of surrounded γ-gliadin genes on chromosome 1.

Based on our experience and in view of the diversity of γ-gliadins, duplex ddPCR to measure γ-gliadin genes or amplicon copy number variation should be designed using a slightly different approach than for α-gliadin. Indeed, γ-gliadin gene sequences are so much more diverse than α-gliadins (Salentijn *et al.*, 2012; *Chapter 2*) that several primer pairs should be designed, each specific to one group of sequence patterns. This set of primer pairs, used in CS deletion lines, would give precise information regarding the sequence patterns on the homoeologous genomes, and this would assist the genome assembly in these complex regions.

Regarding the development of hypoimmunogenic wheat, ddPCR was proven to be a valuable pre-screening tool. In hexaploid wheat lines, this method enabled us to (1) assess near-absolute α-gliadin gene copy number in CS, (2) estimate their repartition over the homoeologous genomes AA, BB and DD, (3) detect α-gliadin gene copy number reduction in CS deletion lines and Paragon γ-irradiated lines as well as (4) quantifying and distinguishing small indels and larger deletions generated in α-gliadin genes by CRISPR-Cas9 constructs in Fielder lines. Therefore, wheat lines identified by ddPCR to contain a lower number of intact gliadin genes could be combined, by simple crossing or generation of synthetic hexaploid wheat, to produce lines with a reduced level of gluten immunogenicity. These lines should be further tested with deep sequencing and advanced proteomics methods in order to characterise the mutations generated, their expression and predict their immunogenic potential based on the sequence of detected epitopes. Further *in-vitro* immunogenic tests including ELISA and T cells stimulation should be done to prove the gluten hypoimmunogenicity of the newly generated bread wheat lines.

Conclusion

Gene copy number variation (CNV) in large gene families can represent an important feature for plant phenotypes. In this study, we developed two duplex ddPCR assays, each amplifying a reference gene with a known copy number per genome and one amplicon in the genes for which the gene copy number was to be determined. These ddPCR assays proved to be accurate and reliable to assess near-absolute α-gliadin gene copy number in hexaploid wheat up to almost 100 gene copies. Duplex ddPCR assays on sets of deletion and nullisomic-tetrasomic lines could quantify relative gene copy number variation and therefore enables to calculate the gene copy number associated to homoeologous chromosome pairs.

In addition to gene copy number variation, ddPCR proved to be able to quantify α-gliadins gene-edited within the gene family and to distinguish between small indels and large deletion induced by the sgRNA-Cas9, in hexaploid wheat. Indeed, we showed that a single Fielder-CRISPR line could contain small indels (1-50bp) in up to 10 α-gliadin genes and large deletions (>300bp) in 20 α-gliadin genes out of the 87 α-gliadin genes present in Fielder. These positive results, generated in a large gene family of a hexaploid plant, attest further the potential of ddPCR assays for high throughput gene-editing screening as shown by Gao *et al.* (2018) for single genes in tetraploid *Medicago sativa*. We conclude that ddPCR is a suitable method for pre-screening of gene-edited plants when gene families or multiple genes are targeted, and one that can be used for high-throughput screening.

3

Acknowledgements

AJ was funded by FP7-PEOPLE-2013_ITN-607178. This manuscript does not necessarily reflect the Commission's views and does not anticipate the European Commission's future policy in this area. The Paragon γ-irradiated population and the Chinese Spring nullisomic-tetrasomic lines were kindly provided by Dr. Clare Lister and Dr. Simon Griffiths of JIC, Norwich, UK, respectively. We would like to thank Dr. Guido Hooiveld for giving us access to the ddPCR and for his advice and support throughout the experimental process.

Chapter 4

Development of GlutEnSeq system to enrich for and sequence gluten gene families in hexaploid bread wheat lines with induced mutations in gluten genes

Aurélie Jouanin[1,2], Theo Borm[1], Bruno A. Santos[2], James Cockram [2], Lesley A. Boyd [2], Richard G.F. Visser[1], Marinus J.M. Smulders[1]

[1]Wageningen University and Research, Plant Breeding, Wageningen, The Netherlands

[2]NIAB, Huntingdon Road, Cambridge, United Kingdom

Submitted

Abstract

Genetic variation studies of large gene families in polyploid crops are challenging. A genomic reduction step such as target gene enrichment makes it possible to use sequencing to characterize all gene members more effectively. Here we report on the development of an in-solution gluten exome capture system called GlutEnSeq for Gluten gene Enrichment and Sequencing, which fully covers the sequence variation present in thousands of available prolamin gene sequences originating from various *Triticeae* species and cultivars. We assessed the efficacy of this capture system, followed by Illumina sequencing, to localise and characterise the mutations generated in gliadin gene families of hexaploid wheat. The on-target regions were determined based on the Chinese Spring (CS) reference genome RefSeq v1.0. Gluten gene sequences were generally enriched around 10,000-fold. For validation, the loss of gluten genes in CS deletion line 1BS-19/6DS-4 deletion was detected in the form of the absence of gluten gene coverage on chromosome 1B and Un (the α-gliadin *Gli-2* loci from chromosome 6D are not yet anchored in the CS reference genome but instead are part of the Unknown chromosome). Two γ-irradiated lines of cultivar Paragon, known to be affected in their gliadin protein profile, were shown to contain homozygous deletions for the α-gliadin *Gli-2* locus on 6A and the γ-gliadin *Gli-1* locus on 1B, respectively. Four CRISPR/Cas9 gliadin gene-edited lines of the cultivar Fielder revealed homozygous deletions of the γ-gliadin *Gli-1* locus on 1B and heterozygosity for the α-gliadin *Gli-2* locus on 6A. We also detected a decrease of gluten gene sequence coverage within parts of some gluten genes on 6A, 6B and 1B in some lines, most likely associated with heterozygous medium-sized deletions within the genes. GlutEnSeq thus enables identifying the homoeologous chromosome pair on which homozygous or heterozygous mutations occurred, such as large deletions across multiple consecutive genes or medium-size mutations within individual genes, in hexaploid wheat. The bioinformatics pipeline will be optimised to also enable efficient characterisation of small indels responsible for lack of coverage within individual gluten genes, in order to fully analyse the Fielder-CRISPR mutant lines for their decrease in gluten immunogenicity for Coeliac patients.

Key words: *GlutEnSeq, Enrichment and sequencing, Gene editing, CRISPR/Cas9, Mutation breeding, γ-irradiation, Deletion lines, Wheat, Polyploid, Gene family, Gluten, Prolamins, α-gliadin, γ-gliadin, Epitope, Coeliac Disease.*

Introduction

Bread wheat (*Triticum aestivum*) is a hexaploid crop (2n=6x=42) with 7 pairs of chromosomes for each of the homoeologous genomes A, B and D. The genome size of hexaploid wheat is approximately 16 Gb (Arumuganathan and Earle, 1991; IGWS, 2014), which is more than five times larger than the human genome. The hexaploid wheat genome consists of only 2% of protein coding sequences (Paux *et al.*, 2006) whereas over 80% are repetitive sequences (Li *et al.*, 2004, Wicker *et al.*, 2011), of which 70% are retroelements (class I; Clavijo *et al.*, 2017). Therefore, the size and the complexity of the wheat genome makes it a challenging puzzle to assemble the genome sequences (Wulff and Moscou, 2014). The first hexaploid wheat genome assembly of the reference cultivar Chinese Spring (CS) was released in 2012 (Brenchley *et al.*, 2012) but it was very fragmented and incomplete, with only 5 Gb -one third of the genome- successfully sequenced and roughly assembled. Since then, new assemblies have been released, based on different sequencing technologies and bioinformatic tools. In 2017, Zimin *et al.* published an assembly of 15 Gb, representing 96% of the genome, with a much better resolution.

Due to the large size of the wheat genome and the difficulty to assemble it, screening a wheat population for variation in complex gene families is currently not feasible using whole genome shotgun sequencing (WGS). Although using RNA sequencing (RNAseq) would reduce the genetic information to be sequenced, the amount of data generated is still high and difficult to handle. Furthermore, not every gene from a gene family is expressed, and expression also varies during development and/or under influence of environmental conditions. To circumvent these issues, Jupe *et al.* (2013), while studying a large and complex family of disease resistance genes (NB-LRRs) in potato, developed an exome capture-based approach called REnSeq for Resistance genes Enrichment and Sequencing.

Based on the REnSeq approach, we developed an exome capture system that we have called GlutEnSeq (Gluten gene Enrichment and Sequencing), in order to study the complex gluten gene families. Gluten consists of two types of proteins, the glutenins and the gliadins. Glutenins are divided into two groups, High Molecular Weight (HMW) and Low Molecular Weight (LMW) glutenins, estimated to have 20 and 35 gene copies respectively (Forde, 1985; Sabelli and Shewry, 1991) while

gliadins are divided into three groups, α- γ- and ω- gliadins and estimated to have 60-100, 39 and 16 gene copies respectively, in hexaploid wheat (Sabelli and Shewry, 1991; Anderson *et al.*, 1997; Ozuna *et al.*, 2015). The exact gene copy number varies among cultivars (Anderson *et al.,* 1997). Genes within each gene family are clustered at specific loci, present on the three homoeologous genomes A, B and D. HMW-glutenins are located at the *Glu-1* loci on the long arm of chromosomes 1. LMW-glutenins, γ- and ω- gliadins are clustered at loci *Glu-3*, *Gli-1* and *Gli-3* respectively, on the short arm of chromosomes 1. However, α-gliadins are located at *Gli-2* on the short arm of chromosomes 6. Within one gene, there are repetitive domains, made of simple and complex repeat motifs. Within a gene family, various groups can be distinguished based on sequence similarity. In addition, a large portion of them may be pseudogenes (up to 90% of the α-gliadin genes, as estimated by van Herpen *et al.,* 2006).

Our aim was to develop GlutEnSeq as a tool to facilitate the study of genetic variation in all gluten gene families simultaneously, and applicable for various species (wheat, barley, rye, or relatives). Here we focus on hexaploid wheat, with the intention to characterise mutations generated by γ-irradiation or CRISPR/Cas9 in gliadin gene families, working towards the development of wheat lines with hypoimmunogenic gluten for Coeliac patients.

Materials and methods

Development of solution-based gluten genes exome capture system

The exome capture system is based on in-solution sequence hybridization, to selectively enrich a DNA sample for gluten gene-like sequences. It has been developed using the MYbaits® technology of Arbor Bioscience (formerly MYcroarray; Ann Arbor, Michigan) in which multitudes of micro-beads are coated with biotinylated single-stranded RNA probes designed to capture DNA fragments containing gluten gene-like DNA sequences.

The probes, also referred to as baits, were designed based on all sequences from gluten genes (i.e. glutelins and prolamins) and flanking regions (e.g. promoters) from any Gramineae containing gluten-like proteins (97 species), as were available on NCBI in June 2015. Some gene sequences cloned and sequenced

in house were added. A total of 7,468 sequences of gDNA and (since the genes of interest do not have introns) cDNAs and ESTs from genes and pseudogenes were used to design the MYbaits® solution-based gluten exome capture system. Integrating all gluten gene sequences enabled the creation of a complete exome capture library that can be used as a resource for projects focusing on glutenins as well as gliadins, for any of the species included. In addition, sequences from the wheat Phytochrome and Flowering Time 1 (*Ta*PFT1) gene were integrated in the exome capture system design to be used as endogenous reference sequences. One copy of *Ta*PFT1 is present per haploid genome on chromosome 5, and the gene copies include homoeologous genome-specific SNPs in both Chinese Spring and in cultivar Chara (Fitzgerald *et al.*, 2010). This gene can therefore be used for data normalisation, giving a better estimation of the enrichment, the sequence coverage and the absolute gluten gene copy number sequenced.

Several successive steps were necessary to optimize the MYbaits® gluten exome capture design, and these steps were performed by Arbor Bioscience based on our criteria. The initial design of the baits was made by splitting each of the input sequences into 120 nucleotide sub-sequences using a moving window approach with steps of 40 nucleotides, giving over 65,000 bait sequences with 3x tiling density, to cover the integral sequences multiple times and to ensure that all gene copy variants would be captured. Bait sequences targeting wheat and having 100% sequence homology to another bait sequence were considered redundant and removed to avoid over-representation of some bait sequences. In case of other *Gramineae* species the threshold for discarding redundant bait sequences was set at 90% sequence homology, due to exome capture design restrictions. Baits usually capture DNA fragments with as low as 70-80% sequence homology (Jupe *et al.*, 2013). The bait sequence coverage was checked by BLASTing each designed bait sequence against all initial full-length sequences. Bait sequences containing signal peptide-like sequences were discarded to avoid capturing off-target genes also expressed in the grain endosperm. All bait sequences were also BLASTed against the Triticeae Repeat Element Database (TREP on http://wheat.pw.usda.gov/ITMI/Repeats/, which contains 1,717 transposon sequences) and bait sequences showing 1 or more BLAST hits with a Tm>40C were discarded in order to avoid capturing unwanted transposon-containing sequences. However, bait sequences from the gluten-specific repetitive regions were retained. Redundant complementary bait sequences were identified by performing a bait-sequence-

to-bait-sequence BLAST. Bait sequences complementary to other bait sequences were discarded.

Plant materials

In total seven mutated hexaploid wheat lines from three different cultivars, showing altered gliadin protein profile on Acid-PAGE compared to their wild type parents (*Chapter 2*), were selected for GlutEnSeq. Chinese Spring (CS) and its 1BS-19/6DS-4 deletion line (Endo, 1988; Endo & Gill, 1996; obtained from Kansas State University), missing α-gliadin gene copies from chromosome arm 6DS and gliadins from chromosome arm 1BS, showing α- and ω-gliadin bands missing, were used. Cultivar Paragon and two Paragon γ-irradiated lines P3-75 and P5-20 (obtained from JIC, Norwich, UK; Shaw *et al*., 2013), selected for altered expression of α-gliadin and γ-gliadin proteins, respectively, were also used. Fielder and four Fielder-CRISPR gliadin mutant lines generated using sgRNA targeting of α-gliadins and/or γ-gliadins (*Chapter 2*) were tested as well. Line α1-14_G1 contains one sgRNA targeting α-gliadins and showed the absence of two α-gliadin protein bands on Acid-PAGE gels. The γ3-20_G6 line contains three sgRNA targeting γ-gliadins and it showed protein bands missing for γ-gliadins and bands shifted for ω-gliadins. The α2γ3-38_G4 and α2γ3-39_G7 lines contain two sgRNA targeting α-gliadins and three sgRNA targeting γ-gliadins, and they showed fainter α-gliadin bands and missing γ-gliadin bands, respectively.

Library preparation and enrichment for Illumina MiSeq 2*250 paired-ends sequencing

DNA was extracted from leaves of the ten mutated seedlings (generation M4 for Paragon and T1 for Fielder-CRISPR mutant lines), using a CTAB-based protocol. DNA samples were sent to Arbor Biosciences which performed all the steps from library preparation, capture and enrichment with GlutEnSeq to Illumina MiSeq 2*250 paired-end sequencing using one sequencing lane (for a total of 12 samples).

Bioinformatic analysis of enriched gluten sequences

Analysis of the sequences resulting from gluten gene enrichment followed by Illumina sequencing was conducted using the wheat genome reference assembly of the International Wheat Genome Sequencing Consortium (IWGSC) RefSeq v1.0, before its full annotation was released (IWGSC, 2018). Gluten gene families are known to have varying copy number in different wheat cultivars (Anderson

et al., 1997). In addition, gluten genes contain low complexity repeat domains made of consecutive "CCA" or "CAG" codons, which translate into glutamine (Q) repeats that are characteristic of gluten proteins. The length of these glutamine-coding domains varies between gene copies, within and between wheat varieties. For example, the length of PolyQ domain 1 in α-gliadins (*General Introduction*) ranges from 24 bp to 210 bp (between 8 and 70 consecutive repeats of glutamine codon), as based on a multiple sequence alignment (*Chapter 2*). Since these short sequence repeats are also abundantly present throughout in the genome as trinucleotide microsatellite repeats, our first task was to define our actual target regions within the reference genome framework.

The target regions within the reference genome were identified by mapping back the bait sequences to the reference genome using two slightly different approaches: A 'high stringency' approach wherein 10 copies of each of the bait sequences and 10 copies of all baits sequences with every possible single nucleotide change were mapped back to the reference, and a 'low stringency' approach wherein a number of random changes R was introduced in 10 copies of each bait sequence for each R in the range of 10-29 (inclusive). For clarity, this means that in the 'high stringency' approach, each 120-nucleotide long bait sequence was represented using a dataset of 1+120*3=361 sequences with small variations, and in the 'low stringency' approach, each bait sequence was represented by a dataset of 200 sequences with larger variations. Read mapping of both the 'low stringency' and 'high stringency' datasets was performed separately using "bwa mem" (Li, 2013) with default settings and post-processed using 'samtools sort' (Li *et al*., 2009).

After read mapping, genome sequence coverage graphs were obtained using "bedtools genomecov" (Quinlan and Hall, 2010), and converted to bed files using a coverage threshold of 1 using a custom perl script. Another custom script was subsequently used to expand the regions in the bed files by 500bp left and right and to merge them when appropriate to accommodate the fact that properly captured DNA fragments were expected to exhibit up to 500bp overhang beyond the actual probe sequences. Subsequently, regions in the bed files were filtered using a custom script to remove regions shorter than 1200 bp after expansion, as such regions were surmised to arise due to short sequence motifs present throughout the genome. After this, lines in both bed files were tagged as '1' or '2' as either 'low stringency' or 'high stringency' using the unix 'awk' command-line utility. Then, they were

combined into a single bed file using the 'bedtools merge' (Quinlan and Hall, 2010) utility, using the option to sum fields in this file, producing a bed file with regions tagged as either '1', '2' or '3' for regions found using either the 'low stringency' approach only, the 'high stringency' approach only, or using both approaches, after some minor manual editing. This bed file containing the final set of target regions was subsequently used in downstream analyses, either directly or in a modified form, for visualisation purposes, for which the regions were expanded slightly by 200bp on either side and with coordinates converted to compact the regions to visualize into a set of adjacent blocks using a custom script. Similar coordinate conversions were applied, as appropriate, to other files to maintain consistency in visualisations. In addition to visualisation purposes, the bed file containing the combined ('high stringency' and 'low stringency') regions was used to extract histograms of the on/off target coverages using a custom script.

Reads obtained from each variety were mapped back against the reference genome using 'bwa mem' (Quinlan and Hall, 2010) using default options, post-processed using 'samtools sort' and 'samtools rmdup' (Quinlan and Hall, 2010) before being converted into nucleotide coverage graphs using 'bedtools genomecov'. For visualisation purposes, regions from these genome coverage graphs were extracted and their coordinates converted to place-defined target regions in adjacent blocks using a custom perl script. Custom perl scripts were used to generate the final GNUPlot (Racine, 2006) scripts and annotation datasets that were used in the visualisations of data.

Results

Solution-based gluten exome capture: final design

To finalise the gluten exome capture design, the specificity of the bait sequences toward their target gene family was checked based on genome-wide hit counts. Each bait sequence was BLASTed against the entire Chinese Spring hexaploid wheat reference genome *(Ensembl! version release 29 of Triticum_aestivum. IWGSC1.0+popseq.29.dna_sm.toplevel.fa.gz,).* The stringent "1 or more BLAST hit with a Tm>40C" cut off was applied. To get an accurate estimate of bait specificity, the number of hits in the wheat genome obtained by a bait sequence was plotted against the number of bait sequences that showed a specific number of hits in the

wheat genome. It enabled selecting the most specific bait sequences (Fig. 1). This value should correspond to the number of targets found in one diploid genome, which, multiplied by three, was comparable to the estimated gene copy number in the targeted family, in a hexaploid genome.

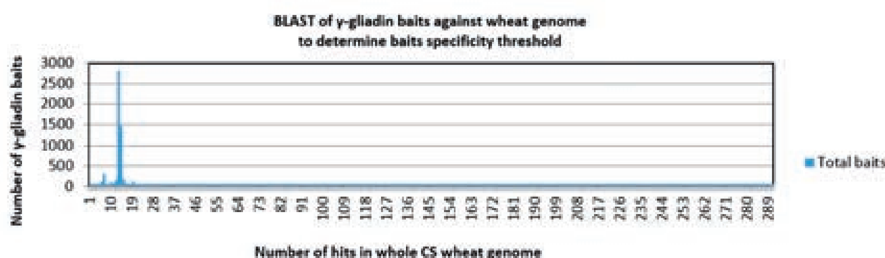


Figure 1. Distribution of γ-gliadin-targeting baits matching different number of sequences in CS wheat genome.
The number of BLAST hits in the CS genome obtained by a bait sequence was plotted against the number of bait sequences that show a specific number of hits in the wheat genome. It enables selecting the most specific bait sequences. Here, in the case of baits targeting γ-gliadins, most of the bait sequences hit 13 sequences in the wheat genome. This means that bait sequences having between 1 and 20 hits in the wheat genome should be specific. If we assume that most primer pairs specifically catch 13 or 14 gene copies in only one of the three homoeologous chromosomes, the total gene copy number estimate is around 3*13=39, which corresponds to gene copy number estimations made using RFLP (Sabelli and Shewry, 1991).

After this step, the number of bait sequences was 40,559. It was reduced by applying a more stringent cut-off regarding the bait sequence specificity to 39,997 (Table 1), to fit the 40,000 baits design offered by Arbor Biosciences.

Table 1. Repartition of the baits targeting each gene family present in the MYbaits® gluten exome capture

Genes	Initial sequences used	Unique 120bp baits generated	Specific baits remaining	Final baits
HMW-glutenins	947	10.567	4.361	
LMW-glutenins	1.876	12.120	9.275	
α-gliadins	2.622	24.576	22.937	
γ-gliadins	1.073	9.802	6.421	No data concerning the final number of baits per family
ω-gliadins	282	2.641	1.124	
***Ta*PFT1 ref**	16	83	72	
Precurser	15	75	n/a	
Promotors/Flaking	637	n/a	n/a	
TOTAL	**7.468**	**59.864**	**44.190**	**39.997**

This table summarizes the evolution of the representation of the different gene families during the process of designing the exome capture system. The final number of bait sequences for each family is not known since the steps discarding signal peptide and transposon containing bait sequences were performed after merging the gene family files for simplicity.

GlutEnSeq specificity and efficiency

To assess the efficacy of GlutEnSeq as a method for characterisation of mutations induced in gluten genes, hexaploid wheat lines tested as mutant lines were pre-screened for gliadin protein profile alteration comparing to the wild type, using Acid-PAGE (*Chapter 2*). Based on these results, it was possible to infer the chromosome and sometimes even the homoeologous genome on which the mutations occurred. For twelve lines, DNA samples were captured and enriched with GlutEnSeq and sequenced using Illumina MiSeq 2*250 paired-end sequencing. We describe here the analysis of 10 of these samples.

As a first step in the analysis, the bait target regions were identified in the Chinese Spring reference genome, leading to the identification of 573 target regions in the reference genome, 12 of which were identified using only the high-stringency target region identification method and 292 which were only identified using the low-stringency target region identification method (Table 2). On-target regions refer to regions where the baits from the exome capture system map on CS reference genome, some of which are on the target chromosomes (i.e. 1 and 6 expected to contain gluten genes) and others map to remaining chromosomes (2, 3, 4, 5 and 7) whereas off-target regions refer to the rest of the genome. Note that not all contigs have been mapped to the chromosome scaffolds of RefSeq v1.0, as a part is included as Un (Unknown).

Based on the identified on-target regions, the Chinese Spring reference genome was subdivided into 4 compartments: 1) Off-target regions, 2) On-target regions found using both low- and high-stringency target identification methods, 3) On-target regions only identified using the low-stringency method and 4) On-target regions only identified using the high-stringency method (these were relatively rare). The sequence reads of the different lines studied here were mapped back to the Chinese Spring reference genome, after which nucleotide coverage tables were obtained. Coverage histograms were extracted for each of the four compartments, which were used to compute average coverage figures for each of the compartments as shown in Table 3.

Table 2. Number and size of target regions identified in the Chinese Spring v1.0 reference genome using high- and low-stringency methods

	Number of regions identified using high/low stringency				Length (bp) of regions identified using high/low stringency			
	Both	Only high	Only low	Total	Both	Only high	Only low	Total
Chr1A	**26**	**1**	**8**	**35**	**82.925**	**1.228**	**13.507**	**97.660**
Chr1B	**35**	**0**	**21**	**56**	**90.932**	**0**	**31.432**	**122.364**
Chr1D	**27**	**1**	**10**	**38**	**86.849**	**1.227**	**16.990**	**105.066**
Chr2A	11	1	10	22	19.792	1.225	15.541	36.558
Chr2B	7	0	14	21	12.939	0	19.981	32.920
Chr2D	4	1	8	13	7.038	1.517	13.514	22.069
Chr3A	5	1	21	27	7.558	1.217	31.735	40.510
Chr3B	8	1	12	21	12.472	1.230	18.473	32.175
Chr3D	5	0	6	11	8.299	0	9.329	17.628
Chr4A	7	0	12	19	12.823	0	18.162	30.985
Chr4B	2	0	12	14	3.086	0	18.831	21.917
Chr4D	2	0	13	15	3.746	0	20.535	24.281
Chr5A	10	1	17	28	20.919	1.231	26.924	49.074
Chr5B	7	1	20	28	12.228	5.376	29.004	46.608
Chr5D	4	0	5	9	11.445	0	7.518	18.963
Chr6A	**16**	**0**	**9**	**25**	**55.600**	**0**	**13.270**	**68.870**
Chr6B	**32**	**2**	**17**	**51**	**103.558**	**3.186**	**25.626**	**132.370**
Chr6D	**3**	**0**	**1**	**4**	**4.349**	**0**	**1.780**	**6.129**
Chr7A	6	2	23	31	10.556	2.454	36.740	49.750
Chr7B	7	0	20	27	10.066	0	28.990	39.056
Chr7D	3	0	13	16	5.761	0	20.490	26.251
ChrUn	42	0	20	62	104.437	0	29.808	134.245
Overall	**269**	**12**	**292**	**573**	**687.378**	**19.891**	**448.180**	**1.155.449**

Table 3. Enrichment efficiency: ratio between on- and off-target coverage of enriched sequences

	Ratio between average coverage in on- and off-target regions (enrichement in folds)									
	Chinese Spring	**1BS-19/6DS-4**	**Paragon**	**P3-75**	**P5-20**	**Fielder**	**α1-14_G1**	**γ3-20_G6**	**α2γ3-38_G4**	**α2γ3-39_G7**
Chr1A	164	157	11.294	11.136	7.512	9.869	9.821	10.225	10.412	7.271
Chr1B	187	51	10.865	10.478	2.721	10.152	10.466	9.865	10.774	4.112
Chr1D	197	198	15.288	14.963	10.189	13.978	14.597	14.591	14.914	9.495
Chr2A	457	87	6.031	6.268	4.944	6.940	7.485	6.922	7.094	5.277
Chr2B	400	74	6.039	5.971	5.221	6.569	7.050	6.192	7.801	5.158
Chr2D	104	31	2.829	2.568	2.208	3.060	3.844	3.370	3.525	2.692
Chr3A	180	16	1.376	1.244	1.104	1.479	1.735	1.410	1.611	1.202
Chr3B	374	58	4.633	4.920	3.955	5.468	5.876	5.606	5.640	4.301
Chr3D	72	33	2.270	2.385	1.641	2.501	2.588	2.521	2.774	1.829
Chr4A	565	100	6.703	6.745	6.201	8.293	9.571	8.236	8.662	7.119
Chr4B	252	73	5.987	6.091	5.317	5.956	6.075	5.595	5.925	4.503
Chr4D	150	55	3.964	4.052	3.032	4.222	4.583	4.498	4.566	2.956
Chr5A	168	36	3.377	3.255	2.444	3.234	3.789	3.424	3.701	2.463
Chr5B	130	40	3.428	3.245	2.670	3.762	4.139	3.524	3.934	2.849
Chr5D	51	39	3.386	3.185	2.434	3.274	3.252	3.288	3.526	2.400
Chr6A	243	216	14.767	849	10.553	14.565	15.593	14.749	8.496	10.151
Chr6B	274	239	7.984	7.909	5.477	16.708	17.254	12.506	17.403	11.387
Chr6D	540	83	7.549	6.961	6.207	9.268	10.164	8.866	9.354	7.378
Chr7A	221	35	2.413	2.515	2.036	2.829	3.094	2.799	3.065	2.265
Chr7B	259	58	3.552	3.598	2.969	4.287	4.506	4.213	4.391	3.398
Chr7D	61	19	1.418	1.404	1.134	1.583	1.699	1.547	1.607	1.236
ChrUn	169	79	9.336	8.216	6.569	9.341	9.454	8.678	9.017	6.908
Overall	**225**	**109**	**8.852**	**7.735**	**5.675**	**9.702**	**10.104**	**8.825**	**9.784**	**6.439**

While a significant proportion of the sequence data was off-target, in particular for Chinese spring and CS deletion line 1BS-19/6DS-4 where it was systematically below 600-fold, the general on-target enrichment was above 1000-fold while the on-target enrichment on targeted chromosomes was above 6000-fold. Overall, the enrichment worked well, with at least a factor 3 increase of coverage in on-target regions versus off-target regions, even though the fraction of on-target nucleotides is relatively low since on-target regions constitute only a very small part of the total genome (data not shown).

GlutEnSeq coverage profile

Sub-selection of sequence coverage data (obtained from the read mapping data) around only the identified target regions (i.e., essentially removing the off-target regions from the data) resulted in what we refer to as condensed on-target coverage profiles or, for short, "coverage profiles" for each of the accessions under study. Below, these coverage profiles are compared to each other in various combinations, as appropriate – CS versus a CS deletion line (Fig. 2), Paragon versus Paragon γ-irradiated lines (Fig. 3) and Fielder versus Fielder CRISPR gliadin-edited lines (Fig. 4).

In these coverage profiles (Fig. 2, 3 and 4) three main types of peaks can be observed: Very narrow, very high peaks most likely corresponding to incidental repeats, low level peaks most likely corresponding to noise caused by mis-hybridization or mis-mapping of reads, and broader peaks essentially covering most of a target region, most likely corresponding to our actual genes of interest. Coverage variation within and between these latter peaks may correspond to variations in hybridization efficiency or in copy number. The physical origin of each region in a condensed coverage profile is depicted beneath each coverage profile, giving information on the genomic position of the different target regions. The target regions are color-coded according to their respective compartment: red for regions only identified using the low-stringency method, blue for regions identified using only the high-stringency method and grey for regions identified using both methods.

GlutEnSeq to study gliadin-induced mutations in hexaploid wheat

CS and gliadin deletion line

Figure 2 shows coverage profiles for CS and CS deletion line for two chromosomes, 1B in Figure 2a and 6D/Un in Figure 2b. Several target regions in this figure exhibit a broad coverage peak, which we associate with proper mapping of target gene sequences. Often these target regions are grouped in a small physical region, reinforcing the possibility that each region in a group represents a single gene in a gluten genes cluster.

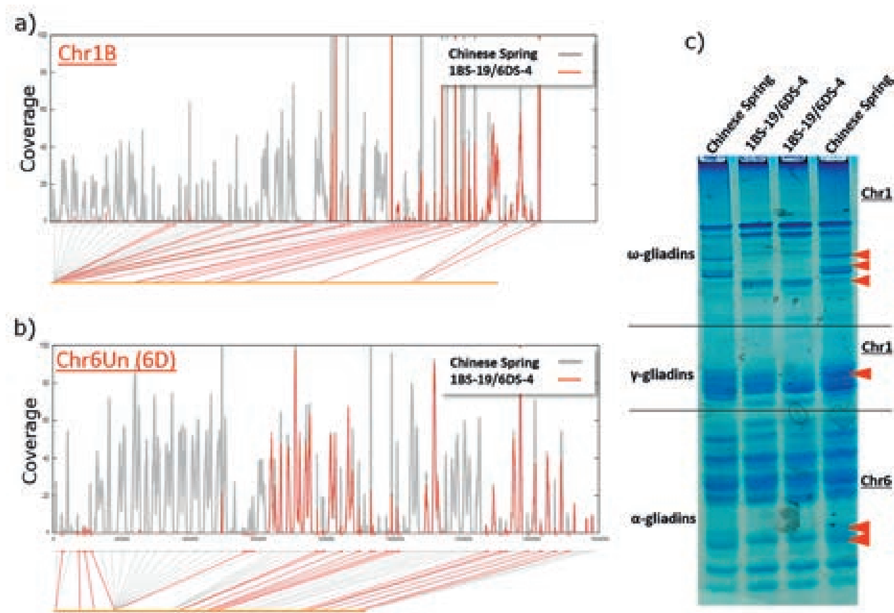


Figure 2. Gliadin sequence coverage profile and Acid-PAGE of Chinese Spring and a Chinese Spring deletion line.
Panel a) and b) display superimposed coverage profiles of CS (grey) and CS deletion line 1BS-19/6DS-4 (red) for chromosome 1B and Un, respectively. The coverage profiles show the number of sequenced nucleotides covering the target regions, which the exome capture system enriched for. Absence of red peaks hence indicates the absence of coverage in the target regions in the deletion line. The track below the figure shows the physical position of each target region on the CS reference genome (chromosome 1B and Un, where Un is not properly assembled). The coverage profiles themselves are a condensed representation of only the target regions, leaving out most of the off-target regions of the genome. In the track below the figure, grey lines represent target regions identified using both high- and low-stringency methods, red lines represent target regions identified using only the low-stringency method. Panel c) shows the Acid-PAGE gliadin protein profile of CS (outer lanes) in comparison to the CS deletion line (in the two central lanes, in duplo). The deletion line lacks protein bands of each of the three gliadin families, as indicated with red arrows.

Almost no bait sequences were mapping to chromosome 6D although it carries many α-gliadins. However, the U chromosome that combines all non-anchored contigs had a lot of bait sequences mapping to it (Fig. 2b).

The coverage profile of CS deletion line 1BS-19/6DS-4, which lacks a part of chromosome 1B short arm that carries γ- and ω-gliadin genes as well as part of chromosome 6D short arm that carries α-gliadin genes, is shown superimposed on the coverage profile of CS. An absence of coverage in the large gluten gene regions in chromosome 1B and Un was observed in the CS deletion line (Fig. 2a, 2b). The absence of coverage in the CS deletion line on chromosome 1B and Un correlates with the characterisation of the deletion line. The lack of coverage on Un in the deletion line and the lack of target regions on 6D in CS lead us to infer that the α-gliadins from 6D are not yet anchored in 6D, but present on Un. No differences in coverage profile between CS and the CS deletion line were observed on other chromosomes. The Acid-PAGE of this deletion line visualises the impact of removing parts of gluten loci on the proteins produced in then endosperm (Fig. 2c).

Paragon and γ-irradiated gliadin mutant lines

Two Paragon γ-irradiated lines were selected as they lacked some gliadin bands on Acid-PAGE. Line P5-20 lacks bands associated to γ-gliadins (which are located on chromosome 1 short arm) while line P3-75 lacks bands related to α-gliadins (located on chromosome 6 short arm; Fig. 3b, 3d). A coverage profile of Paragon together with the two Paragon γ-irradiated mutants P5-20 and P3-75 of chromosomes 1B and 6A was generated (Fig. 3a, 3c).

The comparison between sequence coverage profiles of Paragon and mutants shows the absence of coverage in a large part of the gluten locus on chromosome 1B for the P5-20 mutant line (Fig. 3a). In case of line P3-75, there is no coverage for the complete *Gli-2* locus on chromosome 6A for α-gliadins (Fig. 3c). These results correlate with the bands missing on the gels.

Paragon line P5-20 and CS deletion line 1BS-19/6DS-4 lack the same peaks in their coverage profiles and have similar alteration of γ-gliadins on the protein gels, although they show different changes in the ω-gliadins.

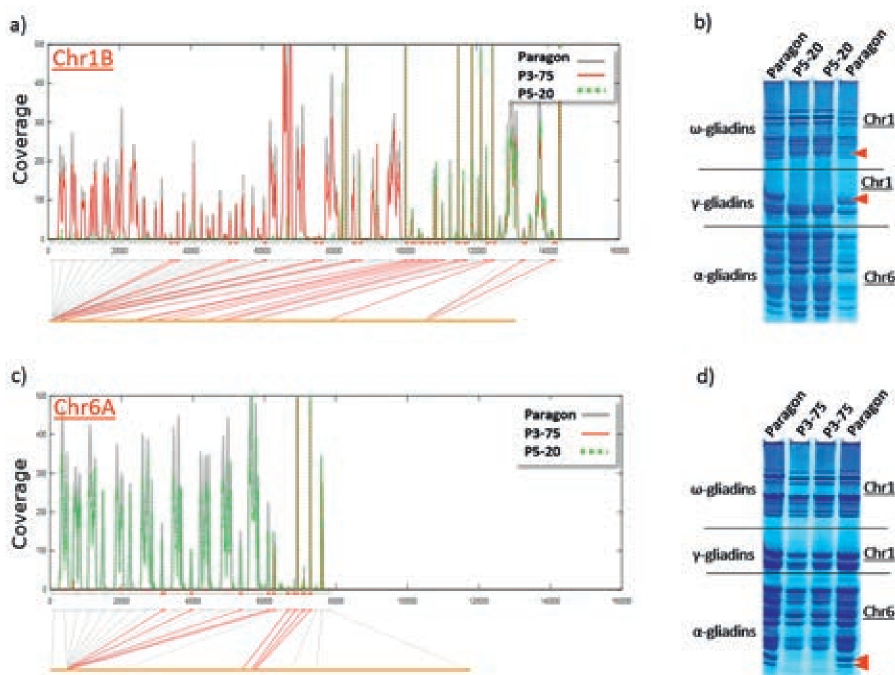


Figure 3. Gliadin sequence coverage profile and Acid-PAGE of Paragon and two Paragon γ-irradiated lines
Panel a) and c) display superimposed sequence coverage profiles of Paragon (grey) and Paragon γ-irradiated lines P3-75 (red) and P5-20 (dotted green line) for chromosome 1B and 6A, respectively. The absence of green peaks in Chr1 and red peaks in Chr6 indicate the absence of coverage of gluten genes and α-gliadin genes in lines P5-20 and P3-75 respectively. Panel b) and d) are the Acid-PAGE gliadin protein profiles of Paragon (outer lanes) in comparison to the Paragon mutant (central lanes, in duplo). The deletion line displays protein bands missing in corresponding gliadin families.

Fielder and Fielder-CRISPR gliadin mutant lines

Four Fielder-CRISPR lines were selected as they lacked some gliadin bands on Acid-PAGE. Line α1-14_G1 contains one sgRNA targeting α-gliadins and showed the absence of two α-gliadin protein bands from the gel. Line γ3-20_G6 contains three sgRNA targeting γ-gliadins and showed protein bands missing for γ-gliadins and bands shifted for ω-gliadins. The α2γ3-38_G4 and α2γ3-39_G7 lines contain two sgRNA targeting α-gliadins and three sgRNA targeting γ-gliadins, and they showed faint α-gliadin bands (Fig. 4d) and missing γ-gliadin bands (Fig. 4b), respectively.

Coverage profiles were generated for Fielder together with two Fielder-CRISPR mutants, α1-14_G1 and γ3-20_G6 (data not shown) and together with two other Fielder-CRISPR mutants, α2γ3-38_G4 and α2γ3-39_G7 (Fig. 4a, 4c).

The comparison between coverage profiles of Fielder and the gene-edited mutants shows the absence of coverage in a large part of the gluten locus on chromosome 1B in line α2γ3-39_G7 (Fig. 4a), which is similar to the CS deletion line (Fig. 2a) and P5-20 Paragon γ-irradiated line (Fig. 3a). The Acid-PAGE profile is in accordance with this result (Fig. 4b) and is similar to the one of P5-20 (Fig. 3b) where a band is missing in the γ-gliadins.

In case of line α2γ3-38_G4, the coverage across the complete *Gli-2* locus on chromosome 6A is approximate half the coverage observed across the same locus for Fielder (Fig. 4c). This is in correlation with the gliadin protein profile where the α-gliadin bands are present but less bright than in Fielder (Fig. 4d).

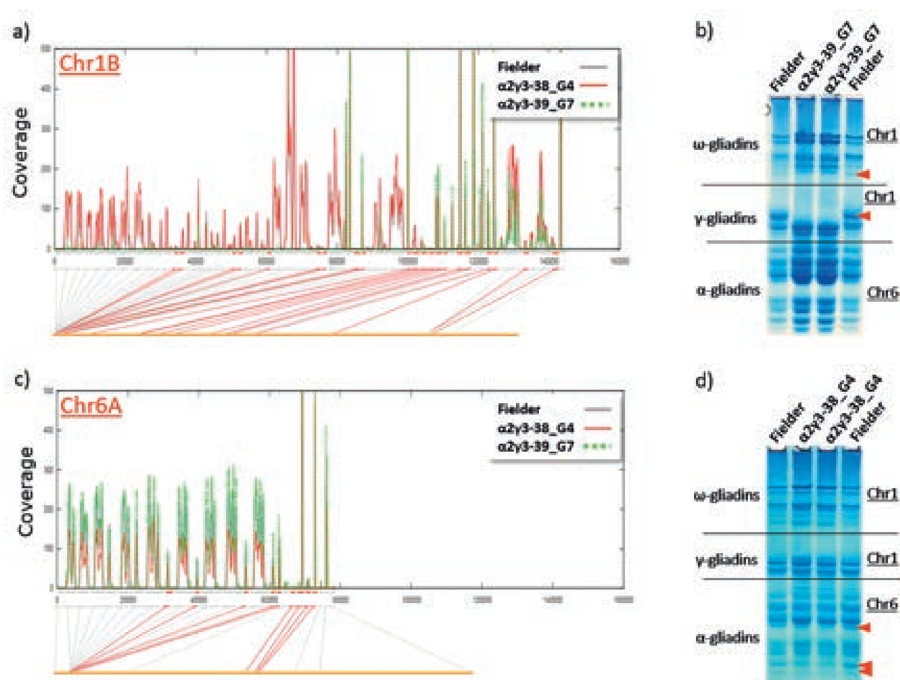


Figure 4. Gliadin sequence coverage profiles and Acid-PAGE of Fielder and two Fielder-CRISPR lines. Panel a) and c) display superimposed sequence coverage profiles of Fielder (grey) and two Fielder CRISPR lines: α2γ3-38_G4 (red) and α2γ3-39_G7 (dotted green line) for chromosome 1B and 6A, respectively. For line α2γ3-39_G7, the absence of green peaks in Chr1 indicates the absence of coverage of gluten genes while for line α2γ3-39_G7 the lower height of red peaks in Chr6 indicates a decrease of coverage by half for α-gliadin genes. Panel b) and d) are the Acid-PAGE gliadin protein profiles of Fielder (outer lanes) in comparison to the Fielder-CRISPR mutant (central lanes, in duplo). The deletion line displays protein bands missing in corresponding gliadin families.

In the γ3-20_G6 line, coverage for some genes from chromosome 1B and 6B show an approximately two-fold reduction (data not shown). These results are consistent with the Acid-PAGE where bands are faint in α-gliadins, one band is missing in γ-gliadins and bands shifted in ω-gliadins.

For line α1-14_G1 the coverage profiles are hard to interpret. It seems that on chromosome 6A, less coverage is present in the middle part of some α-gliadin genes (data not shown). This result would correlate with the Acid-PAGE where two bands associated with α-gliadin genes are extremely faint.

Discussion

GlutEnSeq was developed to identify and characterise mutations generated in gluten gene families from hexaploid wheat species in order to select hypoimmunogenic wheat lines for Coeliac patients. The development of a gluten gene enrichment method was necessary due to the large size of the wheat genome and to the complexity of the gluten gene families.

GlutEnSeq development and enrichment efficiency

The choice of including all gluten gene families from various gluten-containing species in the design of the capture system was made to enable studying several gluten families together and to be able to use the same capture design for different studies in wheat or other species, such as glutenin studies for bread dough quality or secalin studies for rye with low Coeliac immunogenicity. The choice to retain glutamine-rich repetitive sequences specific to gluten genes was made to bind as many gluten sequences as possible, since these sequences were not similar to retro-transposon sequences. Retaining these repetitive sequences may lead to enrichment for some off-target sequences or biased the homogeneity of the enrichment efficiency (Henry *et al.*, 2014), and indeed many repeat-containing sequences were present in the enriched samples, which effectively decreased the sequencing depth of gluten gene target sequences.

The success of the enrichment obtained using GlutEnSeq was determined by looking at the ratio between average sequence coverage in on-target and off-target regions, resulting in an ‘enrichment ratio’. On-target regions refer to regions where

the baits from the exome capture system map on CS reference genome, many of which are on the target chromosomes (i.e. 1, 6 and Un containing gluten genes) and others map to remaining chromosomes (2, 3, 4, 5 and 7). Together this is a small proportion of the genome. Off-target regions refer to the rest of the genome. For all but two samples (CS and 1BS-19/6DS-4), the enrichment ratio on target chromosomes was around 10,000-fold on average, except of course for regions in mutants where target gene copies were missing. On the other chromosomes, where additional potential target regions were found in our bait sequence analysis, the enrichment ratio was below 4000-fold. This result supports the hypothesis that these are actually off-target regions coincidentally captured due to conserved sequence motifs shared with gluten genes, e.g. from other genes of the prolamin superfamily, which occur on all wheat chromosomes (Juhász *et al.*, 2018).

The CS and 1BS-19/6DS-4 samples had a much lower enrichment ratio (an average of 200) than all other samples. The exact cause of this is unknown, but it was noted that sequence library prep for these accessions was done separately from the others, and the insert size distribution of the resulting sequences was different (data not shown). Evidently, this lower enrichment ratio is also reflected in the fact that for these accessions the vast majority of the sequenced nucleotides (>95%, data not shown) was located in off-target regions for these accessions. Nevertheless, because the target represents only a small fraction of the genome, the remaining average coverage in the target regions (15x for CS and 5.5x for 1BS-19/6DS-4) still allowed qualitative assessments.

Bait mapping and mis-assignment of 6D α-gliadin locus sequences

Few baits mapped to chromosome 6D in all 10 lines, even though this chromosome contains the 6D α-gliadin locus (van Herpen *et al.*, 2006). Instead, a cluster of baits was found to map to the unplaced sequences in the reference which are combined into pseudo-chromosome “Un”. Juhász *et al.* (2018, Suppl. S2) also noted that a part of chromosome “Un” in RefSeq v1.0 contains several α-gliadins not yet anchored in 6D. In line with this, CS deletion line 1BS-19/6DS-4, which is known to lack the 6D α-gliadin locus, lacks coverage for several on-target genes on chromosome “Un”.

Comparison between gliadin sequences in wheat cultivars and mutated lines

GlutEnSeq proof of concept to assess gluten gene loss was achieved by comparing the gluten mapping profile of CS with its deletion line 1BS-19/6DS-4, which lacks parts of chromosomes 1BS and 6DS that carry gliadin genes. The deletion line showed partial absence of sequence coverage for chromosomes 1B and Un (which contains the 6D α-gliadin locus). This confirm that this approach can be used to determine which homoeologous gluten loci has been mutated. In theory, the absence of reads mapping to a locus can have two different reasons: either the sequence was completely absent from the sample or the sequence was too different to map to the reference in the correct locus. For that reason, GlutEnSeq should also be used for the reference cultivar that was used for mutagenesis. When no reads map to a full locus, while they are present in the cultivar before mutagenesis, it is most likely that large deletions occurred at the genomic level. In contrast, when only reads mapping in the middle of the gene are absent, it is possible that only small mutations were generated within the gene.

GlutEnSeq was validated using Paragon and two of its γ-irradiated mutants, selected for altered gliadin profiles. Both mutants revealed lack of sequence coverage in a large part of the gluten loci, confirming a large DNA fragment deletion in 6AS for P3-75 and 1BS for P5-20. It is known that using γ-irradiation can cause large deletions. In addition, the genes missing are consistent with the gliadin bands missing on the protein gels (*Chapter 2*), namely α-gliadins in P3-75 and γ-gliadins in P5-20.

GlutEnSeq was then applied to Fielder and its CRISPR/Cas9 mutants, which display differences in the protein profiles on Acid-PAGE. Absence of sequence coverage for a series of gliadins on 1B was detected in CRISPR line α2γ3-39_G7, similarly to the absence observed in the CS deletion line and in one of the Paragon mutants. Line α2γ3-39_G7 contains sgRNA constructs targeting both α- and γ-gliadins and the phenotype shows that γ-gliadin bands are missing on the gel. We can therefore conclude that CRISPR/Cas9 with multiplex sgRNA can generate large deletions in genes that are present in tandem repeats, as reported in rice (Zhou et *al.*, 2014). Moreover, the CRISPR/Cas9-induced deletions were homozygous in the T1 generation used here for GlutEnSeq.

Line α2γ3-38_G4, which contains the same CRISPR/Cas9 construct, showed an approximately two-fold decrease in sequence coverage along the full series of α-gliadin genes on 6A. Two hypotheses are possible. The first one is that the full *Gli-2* locus has been removed, but from only one of the two copies of chromosome 6A, and the T1 grain is still heterozygous for this deletion. The second hypothesis is the T1 grain is homozygous for a deletion, but this deletion only affected about half of the genes in the locus, and the reads obtained from the remaining genes of this locus were mapped back to all genes in the CS reference sequence. On the gel no gliadin band is missing in α2γ3-38_G4 but three bands are fainter. These are the same bands as were completely missing in γ-irradiated mutant P3-75, which was homozygous for the 6A locus deletion. Therefore, it is most likely that α2γ3-38_G4 is heterozygous for the deletion of α-gliadin *Gli-2* locus on chromosome 6A. A study of the T2 seeds from this line will be needed to confirm this.

In the α1-14_G1 mutant line, in which one α-gliadin band is missing and two other are faint (data not shown), fewer reads seem to map to the middle part of four α-gliadin genes on chromosome 6A. In this case, we are probably not dealing with a deletion of a large part of the α-gliadin *Gli-2* locus but instead with smaller mutations within several of the α-gliadin genes. Therefore, CRISPR/Cas9-induced medium size mutations in several genes within a locus may trigger similar gliadin protein profile changes on the Acid-PAGE as when the full locus is gone, if the deletions knock out gene expression, e.g. by generating premature stop codons. Several studies have shown that CRISP/Cas-induced small deletions may produce stop codons and abolish gene expression, in various species. Sánchez-León *et al.* (2018) detected them as a result of several small deletions in α-gliadin genes in wheat. Whether our line α1-14_G1 contains gliadins with new stop codons, will be determined by zooming in on the reads for those regions, but it may explain the difficulties to observe the mutations in the T2 grains.

Limitations of GlutEnSeq

One limitation of GlutEnSeq is that, although the baits are designed on the basis of gluten sequences from many species, the initial mapping is made to the CS reference genome. As a consequence, sequence reads obtained from genes that are not present in the CS reference genome cannot be mapped to their proper location. If a sufficiently similar sequence is present in the CS reference genome, then such reads may accidentally be mapped to the incorrect locus, otherwise

such sequences may remain completely unmapped. This significantly complicates analysis of mutants of a strain for which no proper reference sequence is available: if reads do not map or map to the wrong locus, then conclusions about structural variations may be wrong. One example of the difference among wheat cultivars is the number of α-gliadin genes that we estimated using ddPCR (*Chapter 2*), where CS and Paragon have around 60 genes, but Fielder has more than 87 genes.

In addition, we chose for the data analysis to start by visualising the power of GlutEnSeq in terms of revealing large deletions in gluten loci or medium-sized mutations in individual gluten genes. This analysis, however, does not allow the identification and characterisation of subtle indels that may be responsible for gene knock-out and newly functional gene sequences. In-frame indels in gliadins may supress their immunogenicity for CD patients while remaining functional for bread dough rheology. Additional ways of analysing the data set (inspired from MutRenSeq (Steuernagel *et al*., 2017)) are therefore necessary in order to reveal the full potential of GlutEnSeq for characterising small gluten mutations, despite the level of complexity associated to wheat gluten gene families.

Conclusions

Genetic variation studies of large gene families in polyploid crops are often challenging and require a genomic complexity reduction step such as target gene enrichment and sequencing. Here we developed an in-solution gluten exome capture system called GlutEnSeq for Gluten gene Enrichment and Sequencing, which enable doing various studies of the large gluten gene families present in *Triticeae*. We assessed the efficacy of the method to localise and characterise the type of mutations generated in gliadin gene families of hexaploid wheat, with the aim to develop bread wheat varieties with hypoimmunogenic gluten for Coeliac patient.

The in-solution gluten exome capture system successfully enriched hexaploid wheat DNA samples in gluten genes by around 10,000-fold. GlutEnSeq enabled identifying homoeologous chromosome pairs in which large deletion occurred across gluten gene clusters, such as in CS deletion lines, Paragon γ-irradiated lines but also in some Fielder-CRISPR lines, by analysing differences in gluten gene coverage profile. In the lines made with targeted gene editing, it was possible

to distinguish between lack of coverage of a full gluten gene locus – most likely associated with complete deletion of gluten gene loci – and changes in coverage within genes, as expected for small indels and medium-sized deletions. It was also possible to distinguish between homozygous and heterozygous deletions based on variations in gluten coverage profiles compared to what was observed in the original cultivar, Fielder.

So far, gluten gene coverage profile variation enabled us to infer whether mutations were large deletions of gluten gene loci or medium-sized mutations occurring within a gluten gene. Further developments, on the bioinformatics side, are being made to enable characterisation of small indels responsible of lack of coverage within individual gluten genes, in order to fully analyse the Fielder-CRISPR edited lines for their modification of gliadin genes, and the associated decrease in gluten immunogenicity for Coeliac patients.

Acknowledgements

AJ was funded by FP7-PEOPLE-2013_ITN-607178. This manuscript does not necessarily reflect the Commission's views and does not anticipate the European Commission's future policy in this area. The Paragon γ-irradiated population was kindly provided by Dr. Clare Lister from JIC, Norwich, UK. We would like to sincerely acknowledge Dr. Brande Wulff at JIC for suggesting the use of the REnSeq approach and for advice on technical aspects, together with Dr. Burkhard Steuernagel. We also are grateful to Professor Jonathan Jones and his team, Dr. Kamil Witek, Dr. Agathe Jouet and Dr. Oliver Furzer from TSL, Norwich, UK, for hosting AJ in their lab to be trained on the REnSeq protocol in a version for PacBio Sequencing, which was used for a pilot run of GlutEnSeq.

Chapter 5

Development of wheat with hypoimmunogenic gluten obstructed by the gene editing policy in Europe

Aurélie Jouanin[1, 2], Lesley Boyd[2], Richard G.F. Visser[1], Marinus J.M. Smulders[1]

[1]Wageningen University and Research, Plant Breeding, Wageningen, The Netherlands
[2]NIAB, Genetics and Breeding, Cambridge, United Kingdom

Published in Frontiers in Plant Science, 9, 1523 (2018)
https://doi.org/10.3389/fpls.2018.01523

Abstract

Coeliac Disease (CD) is an auto-immune reaction to gluten in 1-2% of the human population. A gluten-free diet, excluding wheat, barley and rye, is the only remedy. This diet is difficult to adhere to, partly because wheat gluten is added to many processed products for their viscoelastic properties. In addition, gluten-free products are less healthy and expensive. Wheat products containing only hypoimmunogenic gluten proteins would be a desirable option. Various gluten peptides that trigger CD have been characterized. A single wheat variety contains around hundred gluten genes, producing proteins with varying numbers of epitopes. Gene editing using CRISPR/Cas9 can precisely remove or modify the DNA sequences coding for immunogenic peptides. Wheat with hypoimmunogenic gluten thus exemplifies the potential of gene editing for improving crops for human consumption where conventional breeding cannot succeed. We describe here, in relation to breeding hypoimmunogenic wheat varieties, the inconsistencies of applying GM regulation in Europe for gene-edited plants while mutation breeding-derived plants are exempted. We explain that healthy products derived from this new technology may become available in the US, Canada, Argentina and other countries but not in Europe, because of strict regulation of unintended GM risk at the expense of reduction the existing immunogenicity risks of patients. We argue that regulation of gene-edited plants should be based on scientific evidence. Therefore, we strongly recommend implementing the innovation principle. Responsible Research and Innovation, involving stakeholders including CD patient societies in the development of gene-editing products, will enable progress towards healthy products and encourage public acceptance.

Key words: *Coeliac disease, Gluten, Wheat, Mutation breeding, New Plant Breeding Technique, Policy, Public acceptance, Innovation principle, GM Regulation, Genetic Modification, Risk assessment.*

Wheat gluten and coeliac disease

Bread wheat (*Triticum aestivum*) is a staple crop consumed worldwide. The properties that make wheat flour suitable for bread-making are conferred by gluten, the glutenin and gliadin storage proteins present in the grain. High molecular weight (HMW) glutenins provide dough with elasticity, which is the most important property for bread quality, while gliadins provide viscosity (Shewry *et al.*, 2009).

Wheat gliadins, and to a lesser extend low molecular weight (LMW) glutenins, carry immunogenic peptides that can cause Coeliac Disease (CD) in 1-2% of the human population (Fasano, 2006). CD leads to an inflammation of the small intestine, which affects nutrient absorption and causes diverse symptoms (Husby *et al.*, 2012).

A gluten-free (GF) diet, excluding wheat, barley and rye, is the only way CD patients can avoid symptoms. It is difficult to adhere to as wheat gluten is added to many food products (Atchison *et al.*, 2010). Furthermore, current GF products are low in proteins and nutrients, high in salt and contain many additives to emulate the rheology of gluten-based dough (Caponio *et al.*, 2008; Capriles and Arêas, 2014; Belz, 2016; Horstmann *et al.*, 2016). Hence, healthier but safe products for CD patients are needed.

Breeding towards hypoimmunogenic wheat: a complex challenge

Breeding wheat without immunogenic epitopes (Gilissen *et al.*, 2014) would be a definitive solution for CD patients (Shewry and Tatham, 2016). Developing "hypoimmunogenic gluten" wheat varieties that retain baking quality is, however, very challenging. Firstly, gluten proteins are encoded by five gene families containing many immunogenic epitopes. Within these families, α-gliadins on chromosomes 6 trigger CD strongly, followed by γ-gliadins, ω-gliadins and LMW glutenins on chromosomes 1. Secondly, bread wheat is allohexaploid, with three sets of chromosomes referred to as genome A, B and D. Each of these genomes contains all gluten gene families. As a result, a single bread wheat variety has a

combination of gliadins and glutenins, some without any CD epitopes, others with one or more immunogenic epitopes (Van Herpen *et al.*, 2006; Tye-Din *et al.*, 2010; Salentijn *et al.*, 2013). No cultivated wheat or wild relative has been identified that contains only CD safe gluten epitopes (van den Broeck *et al.*, 2010a; 2010b). Consequently, conventional breeding alone cannot produce hypoimmunogenic varieties.

Gil-Humanes *et al.* (2010) used RNA interference to reduce the expression of the gliadin gene families by 97%, abolishing stimulation of T cells from CD patients while no major issues were reported regarding seed germination or dough quality (Gil-Humanes *et al.*, 2014). Becker *et al.* (2012) reduced the expression of up to 20 α-gliadins, but other storage proteins became more abundant. As the transgenic RNAi construct remains in the wheat genome to silence the genes, these plants are subject to GM regulation, which in the EU is expensive, takes a long time, and has an uncertain outcome (Laursen, 2016). In practice this precludes investments in what initially will be a niche product.

Another approach is mutation breeding. Exposure to γ-irradiation has been used to randomly remove large regions of chromosomes in wheat, among which the gluten genes on chromosomes 1 and 6 (van den Broeck *et al.*, 2009). Selected mutations in separate plants can be combined by crossing and selecting, as was done for “ultra-low gluten” barley (Tanner *et al.*, 2016). We screened a γ-irradiated population of variety Paragon (JIC, Norwich, UK) to identify relevant deletions in hexaploid bread wheat. Paradoxically, mutation breeding is regulated as conventional breeding based on a history of safe use, although it randomly alters or removes many other genes besides the intended ones.

CRISPR/Cas9 editing of gliadin genes towards hypoimmunogenic gluten

Gene editing (Baltes and Voytas, 2015), a prominent New Plant Breeding Technique (NPBT), can be used to develop wheat with hypoimmunogenic gluten (Sánchez-León *et al.,* 2018; Jouanin *et al.,* 2018). A Cas9 nuclease is directed by a guide RNA to a target region within the genome and generates a double strand break. Inaccurate DNA repair by the plant may result in mutations at the target site. As a pilot project, we focussed on mutating epitopes in α- and γ-gliadin

genes - which are the most immunogenic - separately and simultaneously using CRISPR/Cas9. We transformed immature embryos of the bread wheat variety Fielder with constructs with Cas9 and multiplex guide RNA constructs, and regenerated plants. Due to the contiguity of the gliadin genes on the chromosome, gene copies located between two DNA breaks may be lost from the genome as well. Sánchez-León *et al.* (2018) successfully targeted α-gliadins with CRISPR/Cas9, generating small deletions. The Cas9 construct is to be out-crossed in subsequent generations (Schaeffer and Nakata, 2015; Sprink *et al.*, 2015). Alternatively, Cas9 can be delivered through transient expression or as ribonucleoprotein (Zhang *et al.*, 2016; Liang *et al.,* 2017).

First, we tested grains of the plants produced for changes in gluten composition by acid-polyacrylamide gels, and determined the number of mutated or deleted gliadin genes using droplet digital PCR. Some γ-irradiated lines showed identical gliadin profile changes to gene-edited lines (Fig. 1). Sequencing data enabled determining the type of mutations generated, while proteomics analysis can identify changes in amino acid composition of modified gliadin proteins. These data will enable predicting whether a mutation in an epitope decreases its immunogenicity, as crucial residues have been determined experimentally (Mitea *et al*., 2010) and the affinity of the human receptors has been fully characterized (Petersen *et al*., 2014; 2016).

Second, gluten from selected lines should be tested for immunogenicity and dough rheology. These tests are designed in collaboration with gastroenterologists, immunologists, food scientists and CD patient associations. They comprise *in-vitro* studies using epitope-specific T-cell clones isolated from CD patients (Anderson *et al.,* 2000) and bread quality tests. Sánchez-León *et al.* (2018) made CRISPR/Cas9 mutant wheat lines with altered α-gliadin profiles and a reduction in immunogenicity, which retained acceptable dough quality.

As a third and final step, *in-vivo* studies are needed where gluten from mutant grains would be given to voluntary CD patients to confirm hypoimmunogenicity. Then, hypoimmunogenic wheat will be ready to be cultivated in a separate production chain, carefully controlled from field to packaging to avoid contamination with regular wheat, barely or rye, similar to a gluten-free oat chain (Smulders *et al*., 2018). It will likely be sold under a specific hypoimmunogenic gluten label.

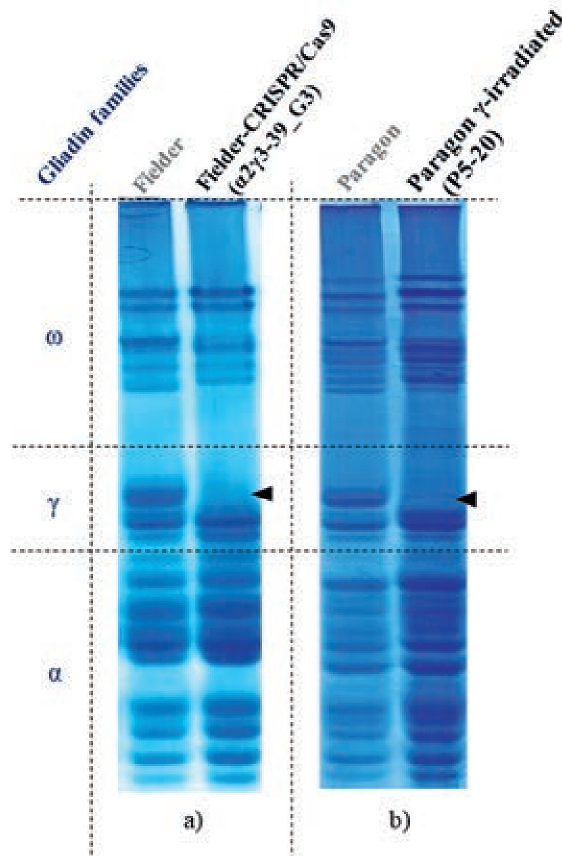


Figure 1. CRISPR/Cas9 gene editing (a) and γ-irradiation mutation breeding (b) generated similar changes in wheat γ-gliadin protein profiles (Chapter2).
Lanes are Acid-PAGE gliadin profiles from control lines (on the left), a T1 Fielder grain, stably transformed with a CRISPR/Cas9 construct, and a M4 Paragon mutant line. Similar amounts of protein extract were loaded.

Gene-edited plant varieties: regulation, safety, acceptance and policy in Europe

We describe here, in relation to hypoimmunogenic wheat, the inconsistencies of applying GM regulation for gene-edited plants in Europe while mutation breeding-derived plants are exempted. The EU regulation is based on the process used, not on the product generated, and follows the precautionary principle. Other countries have a product-based system (Canada) or a mixed product/process-based system (USA, Argentina).

The origin of GM regulation for gene editing plants in Europe

Competent Authorities of several EU countries, including the Swedish Board of Agriculture, as well as the European Food Safety Authority (EFSA, 2015) are in favour of adopting gene-edited products (Sprink *et al*., 2016a) with conventional breeding regulations or adapted regulations (Whelan and Lema, 2015). EFSA found a very low level of intended or unintended risks associated with site-directed mutated products (EFSA, 2012). Furthermore, the former Chief Scientific Advisor to the President of the European Commission (Simon, 2013) and the European Academies Science Advisory Council (EASAC, 2015) supported the regulation of gene editing plants as non-GM. However, the EC postponed a decision, mainly due to pressure from NGOs (Lawler, 2015). Recently, the European Court of Justice ruled that according to the text of the Directive 2001/18/EC, such products should be regulated as GM (ECJ, 2018a; 2018b).

Inconsistent regulation of mutated plants in Europe

Random versus targeted mutations

Mutation breeding deploys chemical mutagens or radiation. Because mutations occur randomly, large mutant populations must be screened to find a plant that contains the desired mutation, and each plant will contain many other mutations. These plants and products are considered as GM but exempted from GM regulation in most countries, including the EU (Directive 2001/18/EC, Annex 1B), due to a history of safe use and consumption since the 1930's. Over 3200 commercial crop varieties have been produced using mutation breeding (Ahloowalia *et al*., 2004; Bado *et al*., 2015).

Gene editing uses a nuclease to generate a double-strand break at a desired target site in the genome, and plants are selected in which a mistake during repair led to a mutation of the target site. Off-targets may occur at a low frequency, much lower than in mutation breeding. In a product-based approach, the fact that plants obtained via gene editing are similar to those obtained using mutation breeding, means that they will follow the regulation of conventionally bred plants due to history of safe use (Fig. 1). In contrast, in a process-based approach, as used by the EU, it has to go through the process of GM risk assessment.

Detrimental effects on costs and opportunities

GM regulation of gene edited plants in Europe implies time-consuming (6 years) and costly ($35M) GM safety tests and administrative processes (McDougall, 2011), with uncertain outcome as the final permission is still a political decision. GM regulation erases the core advantages of gene editing as a quick, precise and cheap method to develop high added-value plants to meet the needs of consumers and society.

In the US, were both mutation breeding and gene editing are exempted from GM regulation, the latter will be preferred since it is more precise, faster and versatile as it can produce homozygous mutations in several gene families simultaneously target (Fig. 2). Consequently, European companies move their research facilities to the US (Burger and Evans, 2018), and European researchers move to US start-ups focusing on gene editing, such as Calyxt, which develops reduced-gluten wheat for the US market. As hypoimmunogenic wheat will initially be a niche product, the costs of GM regulation will be too high for small and medium-sized companies in Europe. Thus, regulation of gene editing as GM will impede innovation, competitiveness and access to healthier food in Europe.

Detection, labelling and effects on trade

It is often impossible to distinguish products obtained using gene editing from those with mutation breeding or from 'natural', spontaneous mutations (Sprink *et al*., 2016b). The absence of distinctness will hamper control and labelling of gene editing-derived products, especially when it concerns material from outside Europe where gene-edited varieties are exempted from GM regulation. It represents a major issue. If Europe does not accept gene-edited products due to their lack of compliance with GM regulation applied in EU, this would block the import of any product that is not GM-labelled and tested. Indeed, any non-GM labelled product could potentially be produced with gene-editing, since there is no obligation of labelling gene-edited products in the US. As a consequence, world trade could be disrupted (Cheyne, 2012).

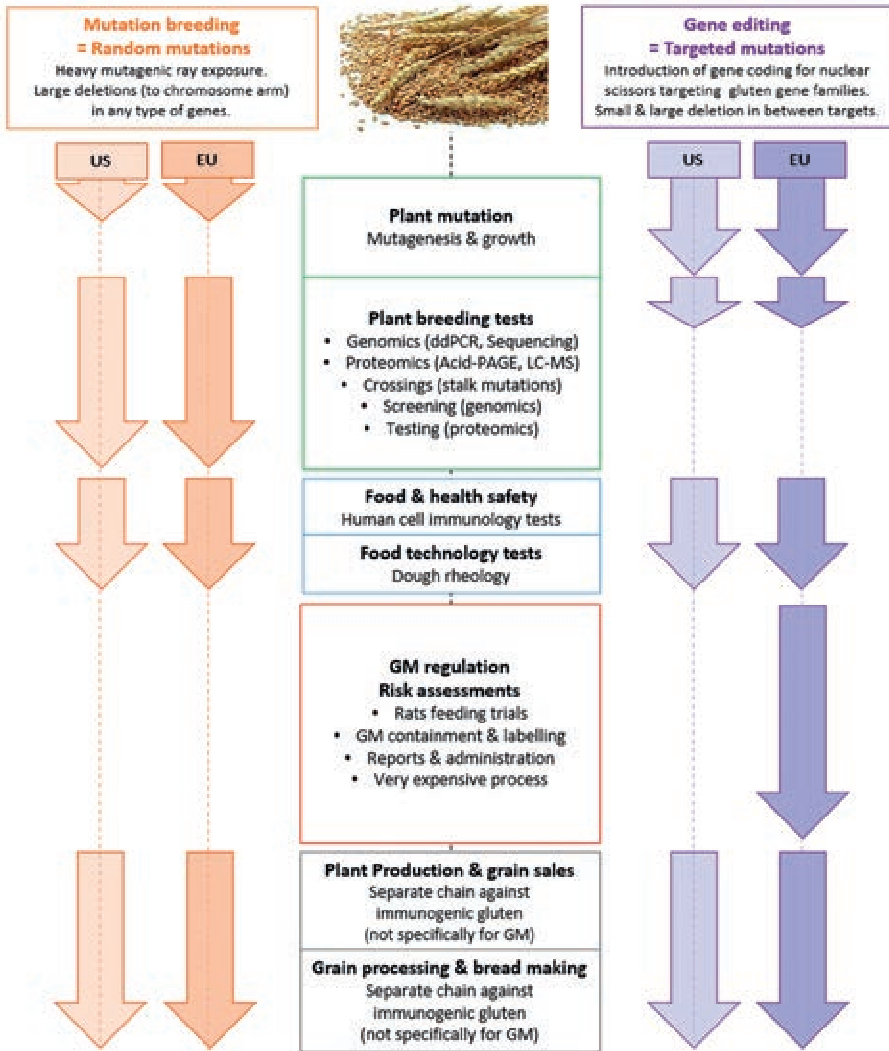


Figure 2. Comparison of the main technical and regulatory steps in the US and EU for breeding a hypoimmunogenic wheat variety using mutation breeding and gene editing.

5

Some gene-edited plants had similar targeted mutations as plants produced with γ-irradiation (Fig. 1) and they cannot be distinguished by their gluten profile. In case of hypoimmunogenic wheat, a separate production chain is always required to avoid contamination with regular immunogenic wheat. The traceability is guaranteed, and products will be labelled as hypoimmunogenic, so it would be relatively easy to label them as derived from gene-edited wheat, even in the US. For other products this will not be the case, as no separate production chain is necessary.

Food safety, environmental safety and food security tests under GM regulation

For each new technology one undertakes a cost, benefit and risk analysis. According to the European Food Safety Authority (EFSA, 2012), the scientific facts gathered so far show no higher food safety risks of gene-edited plants than mutation breeding plants that have a history of safe consumption. Furthermore, gene editing leads to plants with fewer off-targets modifications, making them at least as safe as conventionally-bred ones (Lucht, 2015). This implies, from a risk assessment perspective, that gene-edited plants should be regulated as conventionally bred ones (EPSO, 2015).

Food safety testing

The GM food safety risk assessment tests are related to the presence of foreign genes in the plants. These tests have not uncovered issues for over two decades (SNSF, 2012) and are not adapted for gene editing due to the absence of foreign genes introduced. In case of hypoimmunogenic wheat varieties, food safety issues will already thoroughly have been tested for coeliac patients, to ensure that epitope content is sufficiently low. However, to comply with the GM regulation for food safety, a rat feeding study has to be performed to test whether animals (that do not have CD) would develop unknown symptoms from eating hypoimmunogenic compared to regular wheat. On top of time, costs and animal welfare issues, there is no relevance for these tests.

Environmental safety testing

Regarding environmental risks, under GM regulation, gene-edited plants have to follow strict containment rules. With regard to outcrossing, wheat is a self-pollinated species, and there are no wild populations. Outcrossing to other varieties would introduce hypoimmunogenic gluten, which is safer for human

health, while bread quality would barely be affected. Gluten proteins are storage proteins in the grain and loss of gluten storage proteins did not lead to decreased fitness in gluten-free barley (Tanner, personal communication).

Food security

Considering food security, regulating gene editing as GM in Europe impedes the goals of increasing food production with fewer inputs (Ishii and Araki, 2016) for all types of agriculture, including integrated and organic farming (Andersen *et al.*, 2015). As the economy of many developing countries relies on food exports to the EU, regulating gene editing as GM in the EU consequently has a negative impact on the availability of the technology for local markets in these countries, affecting their food security (Heap, 2013).

Public acceptance and Responsible Research and Innovation

The public needs to be better informed about new food technologies, to enable educated choices about food consumption. Scientists should contribute to this knowledge transfer and creation of awareness. However, the complexity of science often confuses people's risk perception, decreasing their trust in scientific facts and increasing their fears, that they base on inaccurate information or wrong concepts from non-scientific sources (Lucht, 2015). This contributes to empower NGOs that influence the regulation-making process by claiming to protect consumers' safety on no scientific grounds.

In a context where scientific communication has proven to be insufficient, the Responsible Research and Innovation initiative (RRI) (Owen *et al.*, 2012) should be implemented as complementary approach. Targeted consumers should be asked for their interest in a potential product benefiting them and their trust in the methods used, in order to assess product acceptance prior its development, and they should remain involved during the whole process.

CD patients are the prime consumers for gene-edited hypoimmunogenic wheat. Following this RRI initiative, the idea of developing such a product has been discussed with CD patient associations early on. They understand the complexity of the challenge and appreciate the effort of scientists to develop a solution, even if the initial results are not perfect, as often when developing products concerning health issues (Schenk *et al.*, 2011).

Gene-edited hypoimmunogenic wheat fits into a strongly growing market of GF food for coeliacs and other consumers (Sapone *et al.*, 2012). In addition, it may contribute to preventing genetically predisposed children of developing CD, as quantity of exposure matters (Koning, 2012). Thus, there is a clear prospective gain in health which is held back in Europe by the GM regulation of gene editing. CD patients, relatives and others benefiting from gene-edited products should stand up and help the scientific community to convince politicians to adopt a science-based regulation of gene-edited plants and derived products.

Policy making: "innovation principle" instead of "precautionary principle"

Considering the incoherence of applying GM regulation in EU to gene-edited products that may be identical to conventional varieties and anticipating its consequences in terms of food and environmental safety, food security, as well as associated economic issues, we strongly suggest the EC to review its position on the matter. So far, "the precautionary principle" (EC, 2000) is being applied solely, although technically this principle, meant as a provisional measure to avoid discernible risks based on scientific evidence, is not valid anymore considering the history of safe use of GM (no evidence of hazards for 20 years (SNSF, 2012)). We argue that the "innovation principle" (EPSC, 2016) should be used instead where relevant risk assessment would be designed on a case-per-case base, to enable benefiting of gene-edited products while complying with relevant risks management. This would constitute an appropriate regulation for the future of food security, healthy food, as well as protection of the environment and economy.

Conclusions and recommendations

Gene editing has made it possible to remove CD epitopes from wheat gluten. It is expected that in America, derived-products from such wheat will be on the market soon. Due to their absence of compliance with GM regulation, these products will remain illegal in the EU, as long as gene-edited products will be regulated as GM, following a process-based approach.

In addition, these niche products would not be developed in EU either due to the lack of profitability associated to expensive GM regulation tests and labelling.

These GM tests, based on an precautionary principle, are required to detect unintended effects associated to transgenes, which are not present in the product.

We argue that, instead, gene-edited plants should be regulated as plants made with mutation breeding, on a product-based approach, and follow the innovation principle. This principle values benefits associated with the product while scientifically complying with trait-specific risk management. It should be part of a Responsible Research and Innovation initiative involving targeted consumers as stakeholders, to ensure their acceptance throughout the gene-edited product development process.

Food safety, environmental safety and food security in Europe will directly be affected by the regulation of gene editing as GM, and we expect politico-economic issues related to non-GM regulation of gene editing in other countries. Therefore, we strongly advise the EC to review its position on NPBT regulation by considering the present case and the regulatory advices provided.

Acknowledgements

AJ was funded by FP7-PEOPLE-2013_ITN-607178. This manuscript does not necessarily reflect the Commission's views and does not anticipate the European Commission's future policy in this area. C.C.M. van de Wiel is thanked for suggestions and critical reading of the manuscript.

Author contribution statement

AJ conceived this Perspective and wrote the first version. MS adapted the manuscript. LB and RV introduced arguments and edited the text. AJ and MS edited the final version. All authors approved the final version.

General Discussion

Overall goals and findings regarding gliadin gene editing in wheat

Why develop wheat varieties with hypoimmunogenic gluten?

Bread wheat contains two groups of gluten proteins. On the one hand, the glutenins - High Molecular Weight (HMW) and Low Molecular Weight (LMW) glutenins - which can form a network, provide elasticity and thus are essential for good bread dough quality (Weegels *et al.*, 1996; Shewry *et al.*, 2009). On the other hand, the gliadins (α-, γ-, and ω-gliadins) which can contribute viscosity to the network. All include immunogenic epitopes causing Coeliac disease (CD), an autoimmune reaction, affecting 1-2% of the human population worldwide, but the gliadins include the major epitopes (Mäki *et al.*, 2003; Rewers, 2005; Catassi *et al.*, 2014; Vriezinga *et al.,* 2015). The only remedy to prevent CD is a gluten-free (GF) diet excluding wheat, barley and rye, which is very difficult to adhere to since gluten is added to many processed food products for its visco-elastic properties. In addition, GF products contain various substitutes to improve product quality that are less healthy, the products are more difficult to find and more expensive than regular food products (Caponio *et al.*, 2008; Capriles and Arêas, 2014; Belz, 2016; Horstmann *et al.*, 2016). There is therefore a need to create healthier, wheat-based products that would nevertheless be safe for CD patients.

Why use gliadin gene editing?

One option to generate wheat-based products safer for CD patients, which is already being explored since decades, is the development of food processing strategies that modify or degrade gluten – especially gliadins – while retaining the nutrients and most of the other characteristics (*Chapter 1*). So far, no food strategies were yet found to produce wheat-based products that are safe for CD patients. For instance, during sourdough preparation the bacteria produce proteinases that degrade gluten proteins, but they only digest up to 95% of the gliadins (Loponen *et al.*, 2007) and the final taste is not appreciated by most people. Therefore, there is a need to implement, in parallel, breading strategies towards wheat with lower gliadin content or with safe gliadin epitopes (Ribeiro *et al.*, 2018). However, bread wheat is allohexaploid with three homoeologous genomes A, B and D, each containing the five gluten gene families accounting from 20 to 150 genes in total (Sabelli and Shewry, 1991; Anderson *et al.,* 1997; Ozuna *et al.,* 2015). Despite the presence of some genes without gliadin epitopes, conventional breeding

alone cannot succeed in generating wheat varieties with hypoimmunogenic gluten, let alone varieties with entirely safe gluten for CD patients. Therefore, genetic modification (GM) has been deployed to silence gliadin genes with RNAi affected (Gil-Humanes *et al*., 2010) while random mutation breeding can be used to remove gliadin loci from bread wheat (*Chapter 2, 3, 4*). Despite their relative success rate, implementation of these two methods face different types of hurdles: the first one, being a GM method, must follow substantial, expensive and time-consuming regulations before being allowed in the market, whereas the second one generates a large number of off-target mutations and is resource-consuming, as it requires identifying and combining the deletions of all gluten loci into a single wheat plant (*Chapter 5*). As a result, the use of gene editing through the CRISPR/Cas9 system – for which the regulation was not clear at the start of this PhD project - represented a promising alternative in order to precisely modify or delete gliadin epitopes or gliadin genes to prevent CD.

Can gene editing modify gliadin families in hexaploid wheat and how to analyse the mutations?

At the start of my PhD project, gene editing with CRISPR/Cas9 had not yet been used to target and produce knockout mutations in large gene families in polyploids, although it was reported to be successful in small gene families in – diploids - animals (Wang *et al*., 2013a) or for mutating a single copy gene across all six homoeologous loci in hexaploid wheat (Shan *et al*., 2013). Therefore, two main technical questions were addressed. The first one was about the feasibility of using CRISPR/Cas9 for large gene families with diverse sequences. The second one was regarding the choice of screening and analysis methods, with the aim to identify and characterise the gliadin mutations generated in order to pre-evaluate the level of gluten immunogenicity of the mutant wheat lines before clinical testing. Ideally, the screening methods should be amendable for high-throughput screening.

Generating gliadin gene-edited wheat and screening for gliadin protein profile changes using Acid-PAGE

I designed four CRISPR/Cas9 constructs containing multiple guide RNAs (sgRNA) targeting α-, γ-gliadins or both simultaneously at different positions within the genes, that were transformed into the hexaploid bread wheat cultivar Fielder (*Chapter 2*). I first used acid polyacrylamide gel electrophoresis (Acid-PAGE) as a method to screen for mutant Fielder-CRISPR wheat grains with modifications

D

in the gliadin protein profiles. Acid-PAGE is an established method for visualising gliadin protein patterns, and it has been used for many years to detect differences between cultivars (Cooke, 1987). Screening for potential mutant lines was labour-intensive. It did enable identifying the mutant lines in which some gliadin proteins were absent, confirming the feasibility of using CRISPR/Cas9 to edit gliadin gene families in a manner that eventually modifies the gliadins present in wheat grains (*Chapter 2*). The absence of specific proteins can either be due to the deletion of the complete gliadin gene locus or the generation of small or medium-sized off-frame indels in many gliadin genes. Small, in-frame mutations leading to gliadin proteins with amino-acid changes cannot be identified using this method.

Quantification of small mutations and large deletions in gliadin genes using ddPCR

Genomic analyses were performed on the lines grown from grains that showed altered gliadin protein profiles, in order to gather more knowledge regarding the generated mutations occurring, both in expressed and non-expressed genes. In *Chapter 3*, droplet digital PCR (ddPCR) was used to detect small mutations and to quantitatively distinguish them from large mutations (Gao et *al*., 2018), as CRISPR/Cas9 can generate both, especially when target genes are tandemly organised and multiplexed sgRNAs target the simultaneously several position of a gene (Qi *et al*., 2013; Zhou *et al*., 2014). We developed two duplex ddPCR assays which together showed that a single Fielder-CRISPR line could contain small indels (1-50bp) in up to 10 α-gliadin genes and large deletions (>300bp) in 20 α-gliadin genes out of the 87 α-gliadin genes present in Fielder (*Chapter 3*). This method revealed that in most Fielder-CRISPR gliadin mutant lines CRISPR/Cas9 generated more large deletions in α-gliadins than small indels in these genes. However, large deletions may affect individual genes or can span several of the tandemly arranged α-gliadin genes, thus removing all genes present at a gliadin locus at once.

Distinction between large deletions in individual gliadin genes and across gliadin loci using GlutEnSeq

To test further whether the large deletions were associated with individual gliadin gene deletions or with full gliadin loci deletions, next generation sequencing was used. However, as the wheat genome is massive in size and full of repeats it is currently not possible to easily perform whole genome resequencing and analysis.

Consequently, a complexity reduction step is necessary (Jupe *et al.,* 2013), and for this I developed a gluten genes enrichment and sequencing approach, which was called GlutEnSeq (*Chapter 4*). In most cases, the coverage profile of the Fielder-CRISPR gliadin mutant lines revealed the deletion of a full gliadin locus on one of the homoeologous chromosome (A, B or D), in some lines in homozygous form, in others possibly heterozygous. The analysis of the GlutEnSeq data does not yet reveal the gliadin sequence changes generated by small mutations. Although most of them are expected to generate frameshifts, which would be expected to lead to rapidly degraded mRNA or truncated proteins, it is necessary to identify them and evaluate the immunogenic potential of their epitopes in case of stable expression. Further bioinformatic analyses of the sequence reads produced by the GlutEnSeq is required to properly characterise the small mutations induced in gliadin genes.

Characterisation of new gliadin protein sequences end new epitope generated using LC-QTOF-MSMS

Other methods that may be used to characterise and assess the immunogenicity of newly generated and expressed gliadin sequences, include advanced proteomics (Sánchez-León *et al*., 2018; van den Broeck *et al*., 2015). Liquid-chromatography (LC) associated with quadruple time of flight (QTOF) and tandem mass-spectrometry (MSMS) on a gluten protein extract is a state-of-the-art approach to detect and identify specific gliadin peptides. I have been involved in a pilot experiment to qualitatively identify known CD epitope present in Fielder but absent in the Fielder-CRISPR gliadin mutant lines, using an untargeted approach. However, the amount of information generated had not yet been adequately analysed to generate results reliable enough to be presented in this thesis. Ultimately, one would also want to identify the alternative peptide and infer its sequence based on its MSMS spectrum. This approach would reveal the protein sequences of the gliadins expressed in the grain and would enable inferring their potential level of immunogenicity based on the knowledge of gliadin epitopes and their interaction with the human immune cell receptors.

CRISPR/Cas9 efficacy to edit gliadin families in wheat and outcomes of the screening methods used

CRISPR/Cas9 has been shown to generate, in the progenies of some transformed plants, mutations in up to 1/3 of the total α-gliadin gene copies present in hexaploid bread wheat variety Fielder (this thesis). Among these mutated genes,

an estimated 1/3 carried small indels while the remaining 2/3 were affected by large deletions, probably by the deletion of the full gliadin locus from one of the homoeologous chromosomes, in a homozygous or heterozygous manner. These mutations generated a different gliadin protein composition in the grains of the Fielder-CRISPR plants.

All the methods developed and used during my PhD project give complementary information, which is all required to properly assess the gliadin mutants in wheat. However, evaluating the strengths of each of them now that they have been developed, and based on the experiences while using them, future projects should implement them in a different order (Fig.1).

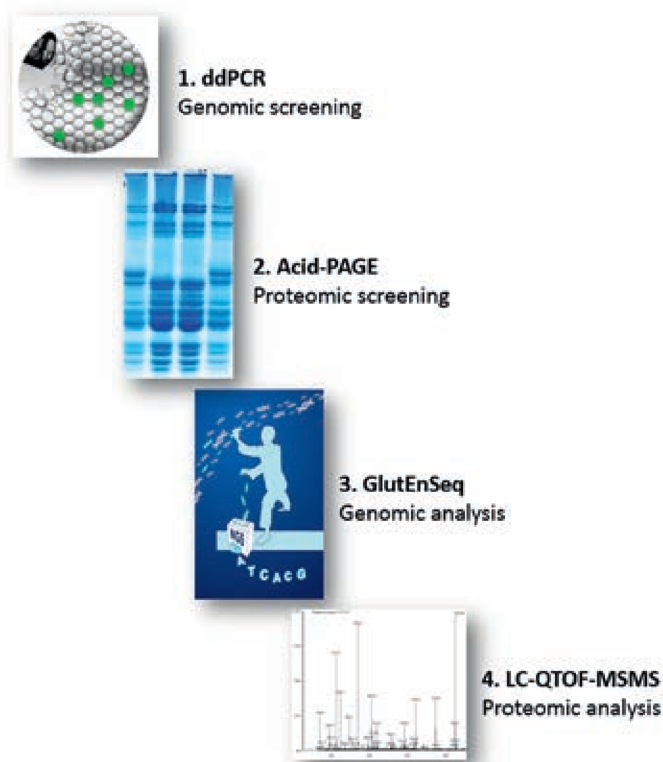


Figure 1. Optimised pipeline to screen for and analyse CRISPR/Cas9-induced mutations in wheat gliadins. To optimise the identification of CRISPR/Cas9-induced mutations in gliadin genes and characterise them, the four methods should be used in the following order: 1) ddPCR enables high throughput quantification of small mutations and large deletions in gliadins, 2) Acid-PAGE enables the visualisation of absence of gliadin protein expression, 3) GlutEnSeq enables the genomic characterisation of deletions and soon of the edited sequences, and 4) LC-QTOF-MS enables identification of modified gliadin protein peptides and eventually their sequences.

The ddPCR, which is a fast, easy to use and relatively cheap method to screen DNA extracts of grains, in a 96-wells format that only takes a couple of hours to analyse, should be used as a high-throughput pre-screening method for identifying gliadin gene mutations (Fig. 1). This would considerably reduce the number of mutant grains to be screened using Acid-PAGE, which is a time-consuming method to evaluate the major impact of the mutations at the protein level (Fig. 1). GlutEnSeq, time-consuming but expensive as well, could then be performed on the most interesting plants to gather information on the size-range of the deletions and on the sequences of the remaining genes, enabling estimating the reduction of immunogenicity compared to the initial variety used (Fig. 1). This also enables estimating what would remain to be edited in a next round. Finally, LC-QTOF-MSMS, which is by far the most complex method applied in this pipeline, should be performed on the most interesting mutant lines only, to get an exact idea of the gluten extract content in term of gliadin epitopes, and to predict its immunogenic level (Fig. 1).

Once lines have been developed that are devoid of the major toxic epitopes -in one or more rounds of gene editing-, the next steps would be to extract gluten and perform tests with epitope-specific monoclonal antibodies (as far as these are available). Then, T cells specific for immunogenic gliadin epitopes would be used for additional testing before to set up an *in vivo* test offered to CD patients who volunteer to test whether the new wheat varieties have no or a very low immunogenic effect on them (Anderson *et al.,* 2000). In parallel, dough rheology tests have to be performed in order to guarantee a good bread dough quality for bread baking despite the modification of the total gluten content in wheat grains.

Reflecting on the gene editing method used towards hypoimmunogenic gluten

Is CRISPR/Cas9 still the best gene editing method to obtain hypoimmunogenic gluten in wheat?

D

CRISPR for generating targeted mutations

When my PhD project started, CRISPR associated with the Cas9 endonuclease or nickase, to generate double stand breaks in the target DNA to induce non-homologous end joining type of DNA repair, were the only advanced CRISPR

systems available in plants. As revealed by the different screening and analysing methods developed during this PhD project, the outcome of CRISPR/Cas9 with multiplexed sgRNA for editing gliadin genes in wheat turned out to produce some small deletions, as was expected, but the most successful lines included large deletions, usually of an entire α-gliadin locus with tandemly repeated genes. This type of full-locus deletions are similar to the ones obtained using γ-irradiation, only more precise since they have been induced in a directed manner. Thus it can be expected that flanking non-gliadin genes have not been affected in the gene-edited lines, while the large deletions induced by γ-irradiation, when they remove a gliadin locus, would remove considerably more genes, in a random fashion. In addition, multiplexing sgRNA enables targeting several gene families simultaneously, which speeds up the possibility of combining the deficiency of multiple gluten loci.

However, fully removing gluten loci is often triggering a phenomenon called compensation, visible as overexpression of some of the remaining gluten genes, which may carry other immunogenic epitopes or impede bread dough rheology. Sánchez-León *et al.* (2018) were quite successful in editing α-gliadins, in terms of reduction of protein produced, but they observed compensation in the form of a higher synthesis and accumulation of ω-gliadins, which, in some of their lines, actually increased the estimated immunogenicity for Coeliac patients. One solution is to remove them all: once all nine gliadin gene loci are removed, gliadin compensation can no longer occur, preventing any unwanted increase of immunogenicity. Only glutenins, which are most important for bread dough quality and -generally- devoid of major CD epitopes, would remain. The potential loss of viscosity due to the absence of gliadins could be compensated with naturally Coeliac-safe gliadin-like proteins, for instance from oat – called avenins – in order to restore bread dough quality, if necessary (Londono *et al*., 2013; Smulders *et al*., 2018).

The literature shows that worldwide researchers are developing alternative approaches using the CRISPR system, with higher efficiency, providing more gene editing opportunities. Below, I discuss the application of a couple of them towards hypoimmunogenic gluten. However, with new developments being published almost every month, these options may quickly become outdated, only meaning that the goal of engineering wheat to be safe for Coeliac patients, is probably closer at hand than we currently expect.

CRISPR for gene correction or replacement

CRISPR associated with a Cas9 endonuclease or nickase enzyme, or associated with the Cpf1 endonuclease, generates DNA cuts with overhangs which facilitate targeted template insertion - using homologous recombination -, in order to correct (SDN-2) or replace (SDN-3) DNA sequences. This methods was proven to work in monocots (Sun *et al.*, 2016). In case of gliadin gene families, this strategy would be very promising, as the loci generated would produce gliadin proteins, but these would be safe proteins as they are already naturally existing in low amounts in some wheat varieties. For α-gliadins such safe variants have been identified in *Gli-2* locus on chromosome 6B. This would have advantages, for screening lines, for subsequent safety testing, and for public acceptance of the resulting products. For instance, in one implementation one could target all gliadins, and templates corresponding to gliadins with these naturally existing CD-safe epitope variants would be provided alongside, and these could be integrated to replace all or some of the immunogenic variants. This approach avoids the possibility that novel epitopes are created, and it would be relatively straightforward to screen the resulting plants, as the desired changes are designed and defined beforehand. Technical issues may occur, related to the fact that we are dealing with multiple copies in a gene family that would need to be resolved. It could be that the original immunogenic sequence cut out in one gene integrates in the next one instead of the safe template, and safe templates could be cut out again by the Cas9.

CRISPR for base editing

Recently, CRISPR systems have been developed with a deactivated nuclease, which is then used as “transporter” for another active enzyme having different properties. One of these alternative enzymes can be a deaminase. For instance, CRISPR-dCas9-cytidine deaminase enables deamination and thus conversion of target base C into a T (and, on the other strand, A into G), with the same sequence specificity as other CRISPR applications, but without generating a double-strand break in the DNA. It has recently proven successful in wheat (Zong *et al.*, 2017). A fusion with an adenosine deaminase generated an adenine base editor (changing G into A) (Kim, 2018). Such systems may be used for DNA codon modification into a codon coding for a different amino acid or a premature stop codon. An amino-acid change in the gliadin epitopes could abolish their immunogenicity while a stop codon would knock-out the gene and prevent stable or any translation of the gene into a gliadin protein. This option would represent a much more controlled

induction of mutations, enabling desired generation of mutations for predicted safe epitopes generation, and avoiding gluten compensation by other gene families. However, the detection of such subtle changes and the verification of their occurrence in each immunogenic epitope would represent a large challenge. This issue may be tackled with the capture array developed here in combination with Illumina sequencing, but it would require a tailored and optimised analysis pipeline to digest the amount of data efficiently and sufficiently precise. Moreover, although safe on the basis of existing knowledge of how the disease develops, there is no certain way of predicting the actual effect of newly generated epitopes on CD or on induction of the disease in susceptible people who do have yet to develop the disease.

CRISPR for gene silencing or overexpressing

A similar system, using CRISPR-dCas9-methyltransferase, enables the methylation of targeted cytosines in promotor regions of genes, which may prevent gene transcription and therefore it is a way to silence gene expression (Putcha, 2016). Using this system, we may be able to silence only gliadin genes with hypoimmunogenic epitopes, therefore limiting the risks of compensation by increased expression of other genes, and maintaining bread dough quality. For this to be possible it would require that we can identify distinct promotor sequences that would be different between (some) safe and immunogenic genes. Most likely, when some or all gliadin genes could be silenced, this would bring about the same results as when using active Cas9 to remove gluten loci, only more difficult to screen for.

Following the idea of working on the regulatory elements (Putcha, 2016) of gliadin genes rather than the gene themselves, another alternative would be over-expression of a transcription factor (TF) –using acetyltransferase - that represses gliadin expression, or silencing a TF activating their expression. Moehs *et al.* (2018) combined stable mutations in three Lys3a-like transcription factors to reduce gluten gene expression by 50%. However, the transcription factors found so far similarly influence the expression of both gliadins and glutenins genes and are therefore not ideal candidates since the aim is to keep the glutenins expression – at least HMW-glutenins – and decrease the immunogenic gliadin expression. Wheat without gluten does not have any trace left of its unique rheological properties that we do not want to lose. This is the reason why engineering wheat that is safe

for CD patients is more complicated than for barley, which is mainly used for making alcoholic beverages, where gluten properties are not required.

Combining different CRISPR methods

Looking at the different CRISPR methods, each of them presents advantages, inconveniences and uncertainties for editing gliadin genes toward wheat with hypoimmunogenic gluten. An idea would be to combine several of these methods, subsequently, to generate the safest wheat variety for CD patients while retaining sufficient baking quality. As an example, in a single wheat plant, the α-gliadin gene family being on chromosome 6 and having the least diversity and being the best characterised, could have their immunogenic epitopes replaced by known safe ones using the Cas9/template approach while the complex γ- and ω-gliadin genes families – overlapping on chromosome 1 – could be deleted. As a result, no gliadin toxic epitope would remain but some gliadin proteins would still be produced to confer some viscosity for bread dough quality.

In case it would be impossible to generate wheat varieties without any gliadin immunogenic epitope, a solution could be to make different varieties lacking different types of epitopes. They would be hypoimmunogenic or safe for different categories of CD patients. However, this scenario would imply very complex food production separations, labelling, testing and systematically genetically screening CD patients to inform them of their set of dominant T cells and the type of epitopes they recognise.

Is stable transformation of CRISPR constructs still the best method to edit wheat gliadins?

Alternatively to the integration of the CRISPR constructs within the plant DNA, a ribonucleoprotein CRISPR/nuclease complex can be used to introduce the sgRNA and the nuclease inside plant cell nucleus, in order to perform gene editing (Woo *et al.*, 2015). The method is applicable by transfection of the complex into plant protoplasts, where it is directed to the nucleus using specific signal peptides to edit target genes before the first cell division, and it is degraded quickly. Although it is usually a very attractive solution, especially for correcting genes using templates, the CRISPR ribonucleoprotein complex may not be suitable for gliadin editing in wheat for two reasons. First, due to the short life time of the complex, it is less likely that all targeted copies would be modified. Second, regeneration of

plants from protoplasts remains challenging in monocots such as wheat, whereas regeneration of stably transformed immature embryos to form plantlets is more efficient.

The future of gene editing wheat varieties for hypoimmunogenic gluten

How far are we from a commercial release of gene-edited hypo-immunogenic gluten wheat?

From a technical point of view

Developing wheat lines with safe gliadin epitopes remains a challenge due to the complexity and the size of the gluten gene families, their aptitude to compensate for the decrease of expression of some gliadins (Galili *et al.*, 1986; Pistón *et al.,* 2011) and the uncertainty, on a long term, of the degree of hypoimmunocity of the newly generated varieties. Despite these issues, gene editing methods are being improved in terms of efficiency and accuracy making the objectives of hypoimmunogenic gluten more realistic. An intermediate goal is wheat lines with at least a great reduction of epitopes related to certain gliadin families if not all gliadin families (*Chapter 1*). These lines may not be sufficiently safe for sensitive CD patients, but they would reduce the induction of the disease in susceptible people (children as well as grown-ups), because dosage (how much toxic gluten do you eat) correlates with the induction of the disease (Koning, 2012).

A concrete example comes from Barro's group in Spain, where regular CRISPR/Cas9 was applied on hexaploid bread wheat lines – with similar sgRNA for α-gliadin – in the same period as this PhD research was carried out. They succeeded in generating lines with low α-gliadin content, lower immunogenicity for some T cells, and a decent bread dough quality following rheology studies (Sánchez-León *et al.*, 2018). This academic group is collaborating with a US company called Calyxt, which is focusing on the development and commercialisation of gene-edited plants and is currently working on commercialisations of these gluten hypoimmunogenic wheat lines (Barro, personal communication). Hence, the first gene-edited wheat lines with hypoimmunogenic gluten and derived products might appear in products on supermarket shelves in the US within a few years' time.

From an immunological point of view

All wheat that is consumed up to now, contains many gluten genes with several epitopes. Immunologists have studied the immune reactions in patients, and they have identified several highly immunogenic epitopes, based on the observation that many patients have T cells that react to these epitopes (Arentz–Hansen *et al*., 2002). These epitopes are called "major epitopes". They have been very well characterized and they are the prime target of my gene-editing approaches to remove immunogenic epitopes from gluten (Sollid *et al*., 2012). In addition, patients also have many T cells that recognise other gluten peptides, but these differ among patients. These are referred to as "minor" epitopes. They are less well characterized, and as the diversity is large, it may be that we do not know them all. One hypothesis is that they are secondarily generated, meaning after the disease has been induced and intestinal inflammation has occurred. If so, it may be that some Coeliac patients cannot tolerate hypoimmunogenic wheat if it still has one of the minor epitopes to which they react. Meanwhile, this wheat without major epitopes would be safe for susceptible children and other people in which the disease has not yet developed, as the epitopes that induce the disease, are absent. However, an alternative possibility is that these "minor epitopes" are as immunogenic as the "major" ones, just less efficient because less abundant, and thus wheat with only minor epitopes would also be able to induce CD. Once hypoimmunogenic wheat would be common, its consumption may still induce the onset of the disease, although, most likely, later in life. Therefore, although safe on the basis of existing knowledge of how the disease develops, there is no certain way of predicting the actual effect of only "minor epitopes", or newly generated epitopes, on CD or on induction of the disease in susceptible people who do not have developed the disease yet (Lundin, Personal communication). It is therefore important that the introduction of hypoimmunogenic wheat varieties into the diet of patients, is monitored.

From a regulatory point of view

For more than ten years, in most countries, no clear regulations were defined regarding new plant breeding techniques (NPBT), a suite of techniques to which gene editing using TALENs or CRISPR systems are counted. As a matter of fact, gene editing requires foreign DNA or proteins to be inserted in plant genomes or plant cells, as a temporary step to perform editing of target genes, and are then crossed out or naturally degraded, so that they are absent in the resulting gene-edited plant and derived products. The issue is thus, whether to regulate gene-

edited plants – and derived products –using a process-based or a product-based approach. A process-based approach would consider the temporary GM step necessary to induce the targeted mutation and they would therefore fall under GM regulations (*Chapter 5*). On the contrary, a product-based approach would consider the absence of foreign entities in the final gene-edited plant, and that it would carry similar mutations as plants obtained through spontaneous mutations or produced using mutation breeding with mutagenizing agents –e.g. γ-irradiation or EMS -, which are exempted from GM regulations and follow conventional breeding rules because of a history of safe use. They have been exempted by putting them on Annex 1B of the GMO Directive 2001/18/EC.

While in 2017 and 2018, several countries in North and South America decided not to regulate gene editing and some other NPBT techniques as GM (e.g. Argentina, USA) – or to decide on a case per case basis as in Canada -, the European Commission (EC) announced and postponed a decision several times. In the end they have not made any announcement on the matter. On July 25th 2018, the European Court of Justice, in a case brought before them from France, ruled that, according to the text of the Directive 2001/18/EC, gene editing is considered as GM technique and that is not exempted from GM regulation but should be subjected to regulation in the European Union (ECJ, 2018a; 2018b). Consequently, gene-edited wheat lines with hypoimmunogenic gluten and derived products can be easily developed and commercialised in the American continent and in countries adopting similar gene editing regulations but they will not be accepted in Europe without fulfilling all GM related tests and labelling. This will mean in practice that gene-edited varieties in general, and niche products in particular, will not be produced, as the regulatory process itself already costs a lot of money (estimates range from a few million to 100 million euro per event (McDougall, 2011)). Moreover, even if the regulatory process would be successfully concluded, European countries could still reserve the right to reject the cultivation within their respective territories. Therefore, the two remaining sources of hope for CD patients in the EU is a change of gene-editing regulation or a successful wheat line based on combining γ-irradiation-induced large mutations deletions of all nine gliadin gene loci (*Chapter 5*). Such a line would most likely not have a good baking quality, and it could harbour more deleterious mutations, notably in the form of several deleted genes that used to be adjacent to the gluten loci – but it would not be regulated in the EU.

What would be required in the production chain using wheat with hypoimmunogenic gluten?

If hypoimmunogenic wheat varieties make it to the market, there are two options for wheat breeding companies, farmers and food processing companies. The first option is to have two production chains, meaning two sets of completely separated facilities for hypoimmunogenic wheat and regular wheat at the breeding companies, in the production farms, at the processing factory, all the way to packaging and labelling, to avoid any risks of contamination with other grains containing immunogenic gluten epitopes.

The other option would be to only work with hypoimmunogenic gluten-containing grains throughout the whole chain, which would decrease even more the risks of contamination. Most likely the separate chain solution is the one chosen initially, as only few hypoimmunogenic varieties would exist. In case the products are well accepted and the market grows, because consumers prefer safe wheat over regular wheat, even if they are not Coeliac patients themselves, more products will be made with hypoimmunogenic gluten. At some point they may be considered as the new standard for gluten.

In the long run, this would slowly imply the replacement of regular wheat varieties by hypoimmunogenic ones, abolishing the need for separate production chains, although strict precautions should remain in place to avoid contamination with related species, still containing immunogenic gluten epitopes.

How to shift from gluten to immunogenic epitopes detection in food products?

Currently, products labelled as GF have to contain less than 20 ppm of gluten (Saturni *et al*., 2010), which is tested using R5 antibody tests. However, in case of hypoimmunogenic gluten on the market, what has to be measured is not only the gluten content (as GF products will remain to produced) but also the amount of immunogenic epitope present in the total gluten extract of wheat-containing products. So far, no method exists to achieve this task. Antibodies specific for immunogenic epitopes could potentially be developed, but it has turned out to be very difficult to develop highly specific antibodies for these sequences with a high degree of similarity. Only one antibody is commercially available that recognises

the deamidated form of a CD epitope. It was developed by the group of Prof. Koning at Leiden University in the Netherlands.

Proteomic methods distinguishing gliadin protein epitope sequences could be developed and implemented for pure grains, but for food product extracts it would be more difficult to implement protein-based assays to test for the absence of immunogenic gluten epitopes. An alternative for grains would be to implement genomics assays, using the GlutEnSeq system followed by a comparison of the tested wheat sequences with known gluten immunogenic epitopes, or by recognising the varieties used. An absence of immunogenic epitopes at the DNA sequence level would ensure the absence of immunogenic epitope in the gluten protein used to make food products. This solution, could not be repeated on food products due to a too small amount of gluten gene DNA in food products, which would require a high level of precaution against contamination by immunogenic gluten downstream the genomic tests performed on the grains.

What about public acceptance of gene-edited hypoimmunogenic gluten wheat?

In terms of public acceptance, it has been shown that people suffering from a food-related disorder are usually supporting the development of healthier products for their disorder, regardless of the methodology used (Schenk *et al.*, 2008, 2011). Thus, there are high chances that CD patients welcome food products with hypoimmunogenic gluten generated using gene editing technology (*Chapter 5*). For non-CD people who do not consume GF products, the chances are that if their fellow CD-sufferers accept gene-edited products, they would encourage their development but may not want to consume them themselves at first. Once more and more products would be labelled as containing hypoimmunogenic gluten, the non-CD consumer may start considering it as the new standard. Progressively all products could be made of hypoimmunogenic gluten and current high immunogenic wheat would no longer be commercialised at least in some parts of the world. In addition, the regulation used in different countries regarding gene-edited products is likely to impact on the public acceptance of products containing gene-edited hypoimmunogenic gluten.

Would wheat hypoimmunogenic gluten protein families be safe for everyone?

For individuals without wheat- or gluten-related discomforts

People not subject to any adverse effects triggered by wheat or gluten should not experience any differences consuming hypoimmunogenic gluten than when consuming regular (high-immunogenic) wheat or gluten, especially if the strategy used is complete removal of gliadin loci. However, in case of hypoimmunogenic gluten generated with gliadin epitopes editing, predictions cannot be made whether on a long term, epitopes so far safe would not suddenly reveal immunogenic potential due to the increase of their frequency in diet and consumption by larger numbers of people.

For CD patients

In case of a wheat variety with hypoimmunogenic gluten, several issues exists. Firstly, the term “hypoimmunogenic” is used because an absolute “safe” variety cannot be guaranteed. In case of a variety where all α-, γ-, and ω-gliadin loci have been fully removed, the variety would be relatively safe for patients carrying T cells recognising the epitopes present in these families. However, a low percentage CD patients react to some LMW-glutenins epitopes and HMW-glutenins (the key family for baking quality). Therefore, LMW-glutenins should also be removed or edited to ensure a better hypoimmunogenicity. As for the HMW-glutenins, it might be possible, giving the variety of gene editing tools now available, to modify very subtly its sequences, prevent its immunogenicity while retaining a similar amount expressed, having similar 3D configurations to retain all its baking properties.

Secondly, there is an issue about minor epitopes in gliadins, that would not all be removed when only a subset of gliadins, with major epitopes, are removed (see above). To which extent this will be a problem for a subset of the existing patients, or for the induction of the disease in healthy people with a susceptible background, remains to be determined once the varieties exist.

D

For individuals with non-coeliac wheat sensitivity

Non-coeliac wheat sensitivity (NCWS) or non-coeliac gluten sensitivity (NCGS), are terms that actually refers to individuals presenting symptoms while eating

wheat, which are not caused by gluten (Lundin and Alaedini, 2012). A better adapted terminology would be non-gluten wheat sensitivity, NGWS. These individuals – as well as irritable bowel syndrome (IBS) sufferers - are not sensitive to wheat gluten but to other components present in wheat. The exact nature of these compounds is subject to research and debate, but it is currently not known.

To these persons, ingestion of hypoimmunogenic or highly immunogenic gluten would not make a difference. Since the actual wheat grain components that trigger the disorder are not clearly identified yet, it is not possible to engineer wheat safer for NCWS sufferers. GF products are currently safe for them, it can be expected that the market for GF products will not disappear once hypoimmunogenic ("safe-gluten") products appear. In theory, one could provide them with healthier products by using similar ingredients as in current GF products but adding purified gluten extracts rather than additives to restore the viscoelasticity required for baking quality.

For individuals with allergies

In addition to CD and NCWS, several allergies are triggered by gluten, although the number of people suffering from these allergies is much lower. For these people, a wheat variety lacking or having modified gliadins would be safer, regardless of the presence or absence of other gluten families. Therefore, a gene-edited wheat hypoimmunogenic for all gliadin families would benefit them. Wheat also contains several other allergens (Juhász *et al.*, 2018). In theory all or a subset of these allergens could also be removed. Technically, this might even be easier than removing CD epitopes from gluten, as fewer genes are involved, but the market would be much smaller as well.

Would it be better to breed for baking quality in "safe" cereals?

Looking at the various adverse conditions triggered by wheat, one could wonder whether breeding for baking quality in alternative cereals and pseudo-cereals considered safe -e.g., buckwheat, sorghum, amaranth, quinoa, etc. - would be a better option. However, most of these plants currently represent niche products without much breeding history. Breeding programs, which take half a dozen

years for well-known crops – would have to be started almost from scratch, which could take several decades and very few companies would make such investments, without guaranties on a market share and return on investment in case of success. Moreover, it would imply modifying the whole food processing chain and convert consumers to new types of food products, which is a real challenge. In addition, plants that are now considered safe while being niche products may suddenly also trigger some health issues once consumed on a larger scale by more individuals. The uncertainty related to these crops are pretty similar to the ones rising about gene-edited hypoimmunogenic wheat. However, the knowledge and methods necessary to develop hypoimmunogenic or hypoallergenic wheat with gene editing are already available, minimising the time scale necessary to generate safe and healthy food product for people suffering from CD and other gluten-related disorders.

Did wheat breeding increased the gluten immunogenicity over time?

There is a widespread belief that modern wheat breading increased gluten immunogenicity and is therefore responsible for the current burst of CD cases. However, studies comparing gliadin epitopes presents in some current commercial wheat varieties and wild relatives show that some wild wheat relatives contain higher amounts of immunogenic epitopes or few epitope variants more immunogenic than in some modern wheat varieties (Schaart *et al*., *in prep*). Other studies have found very little changes in overall gluten immunogenicity of varieties over the last 50 years, although the level of immunogenicity is highly different between varieties (van den Broeck *et al*., 2010a; Kasarda, 2013; Ozuna and Barro, 2018). Breeding is therefore not a very likely reason for the steep increase in diagnosed CD cases. A better understanding of CD mechanism, the wider awareness of its existence and its symptoms by both physicians and consumers, as well as the development of more detection methods, has contributed to an increase of the number of CD diagnosed (Murray *et al.,* 2003). Nevertheless, the incidence of CD itself also appears to be on the rise (Lohi *et al*., 2007). The evolution of the diet over time and across the world, together with many other factors, has probably contributed to this increase in the incidence of CD (*Chapter 1*). In this situation, hypoimmunogenic wheat may not change some of these underlying processes, but it would help CD and NCWS patients.

Concluding remarks

Wheat varieties with hypoimmunogenic gluten represent a need for individuals suffering from Coeliac Disease. Developing such varieties is challenging due to the polyploidy of wheat and the complexity of gluten gene families. However, this pilot project revealed the potential of gene editing technology to achieve this goal. Meanwhile, additional years will be necessary, for improving gene editing efficiency and screening methods, in order to generate a wheat variety hypoimmunogenic for most CD patients. In addition, the wheat-chain will need some adaptations and new methods to distinguish "gluten-safe" food products from "gluten-free" ones.

References

References

Aaltonen K., Laurikka P., Huhtala H., Mäki M., Kaukinen K., Kurppa K. (2017). The Long-Term consumption of Oats in Celiac Disease Patients Is Safe: A Large Cross-Sectional Study. *Nutrients,* 9, 611.

Acevedo-Garcia, J., Spencer, D., Thieron, H., Reinstädler, A., Hammond-Kosack, K., Phillips, A.L., Panstruga, R. (2017). mlo-based powdery mildew resistance in hexaploid bread wheat generated by a non-transgenic TILLING approach. *Plant Biotechnology Journal*, 15, 367-378.

Ahloowalia, B., Maluszynski, M., Nichterlein, K. (2004). Global impact of mutation-derived varieties. *Euphytica*, 135, 187–204.

Andersen, M.M., Landes, X., Xiang, W., Anyshchenko, A., Falhof, J., Østerberg, J., Olsen, L.I., Edenbrandt, A.K., Vedel, S.E., Jellesmark-Thorsen, B., Sandøe, P.,Gamborg, C., Kappel, K., Palmgren, M.G. (2015). Feasibility of new breeding techniques for organic farming. *Trends in Plant Science*, 20, 426-434.

Anderson, O.D. (1991). Characterization of members of a pseudogene subfamily of the wheat α-gliadin storage protein genes. *Plant Molecular Biology*, 16(2), 335-337.

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

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

Anderson, R.P., Degano, P., Godkin, A.J., Jewell, D.P., Hill, A.V. (2000). In vivo antigen challenge in celiac disease identifies a single transglutaminase-modified peptide as the dominant A-gliadin T-cell epitope. *Nature Medicine*, 6, 337.

Andersson, M., Turesson, H., Nicolia, A., Fält, A.S., Samuelsson, M., Hofvander P. (2017). Efficient targeted multiallelic mutagenesis in tetraploid potato (*Solanum tuberosum*) by transient CRISPR-Cas9 expression in protoplasts. *Plant Cell Reports*, 36, 117-128.

Arentz-Hansen, H., Körner, R., Molberg, Ø., Quarsten, H., Vader, W., Kooy, Y.M., Lundin, K.E., Koning, F., Roepstorff, P., Sollid, L.M., McAdam, S.N. (2000). The intestinal T cell response to α-gliadin in adult celiac disease is focused on a single deamidated glutamine targeted by tissue transglutaminase. *Journal of Experimental Medicine*, 191, 603-612.

Arentz–Hansen, H., Mcadam, S.N., Molberg, Ø., Fleckenstein, B., Lundin, K.E., Jørgensen, T.J., Jung, G., Roepstorff, P., Sollid, L.M. (2002). Celiac lesion T cells recognize epitopes that cluster in regions of gliadins rich in proline residues. *Gastroenterology*, 123(3), 803-809.

Arumuganathan, K. and Earle, E.D. (1991). Nuclear DNA content of some important plant species. *Plant Molecular Biology Reporter*, 9(3), 208-218.

Atchison, J., Head, L., Gates, A. (2010). Wheat as food, wheat as industrial substance; comparative geographies of transformation and mobility. *Geoforum*, 41, 236-246.

Bado, S., Forster, B.P., Nielen, S., Ali, A.M., Lagoda, P.J.G., Till, B. J., Laimer, M. (2015). Plant mutation breeding: Current progress and future assessment, *Plant Breeding Reviews,* 39, 23-87.

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

Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., Moineau, S., Romero, D.A., Horvath, P. (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science*, 315, 1709-1712.

Bassi, S., Maningat, C.C., Chinnaswamy, R., Gray, D.R., Nie, L. (1997). Midwest Grain Products. Alcohol-free wet extraction of gluten dough into gliadin and glutenin. United States Patent, No 5,610,277.

Basu, A.S. (2017). Digital assays part I: partitioning statistics and digital PCR. SLAS TECHNOLOGY: *Translating Life Sciences Innovation*, 22(4), 369-386.

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 alpha-gliadin content obtained by RNA interference. *Journal of Applied Botany and Food Quality,* 85, 23-33.

Belhaj, K., Chaparro-Garcia, A., Kamoun, S., Nekrasov, V. (2013). Plant genome editing made easy: targeted mutagenesis in model and crop plants using the CRISPR/Cas system. *Plant Methods,* 9, 39.

Belz, M.C.E. (2016). Reduction of salt in yeasted wheat bread: impact on bread quality and solutions using sourdough fermented by functional lactic acid bacteria strains. Chapter 5. Thesis, University College Cork.

Bigiarini, L., Pieri, N., Grilli, I., Galleschi, L., Capocchi, A., Fontanini, D. (1995). Hydrolysis of gliadin during germination of wheat seeds. *Journal of Plant Physiology*, 147, 161-167.

Bio-rad (2018). Droplet Digital™ PCR Applications Guide. [online] Available at: http://www.bio-rad.com/webroot/web/pdf/lsr/literature/Bulletin_6407.pdf [Accessed 23 April 2018].

Brenchley, R., Spannagl, M., Pfeifer, M., Barker, G.L., D'Amore, R., Allen, A.M., McKenzie, N., Kramer, M., Kerhornou, A., Bolser, D., Kay, S. (2012). Analysis of the bread wheat genome using whole-genome shotgun sequencing. *Nature*, 491(7426), 705.

Brooks, C., Nekrasov, V., Lippman, Z.B., van Eck, J. (2014). Efficient gene editing in tomato in the first generation using the clustered regularly interspaced short palindromic repeats/CRISPR-associated9 system. *Plant Physiology*, 166, 1292-1297.

Burger, L. and Evans, D. (2018). Bayer, BASF to pursue plant gene editing elsewhere after EU ruling. [online] Available at: https://www.reuters.com/article/us-eu-court-gmo-companies/bayer-basf-to-pursue-plant-gene-editing-elsewhere-after-eu-ruling-idUSKBN1KH1NF. [Accessed 28 August 2018].

Caponio, F., Summo, C., Clodoveo, M.L., Pasqualone, A. (2008). Evaluation of the nutritional quality of the lipid fraction of gluten-free biscuits. *European Food Research and Technology*, 227, 135-139.

Capriles, V.D. and Arêas, J.A.G. (2014). Novel approaches in gluten-free bread making: interface between food science, nutrition, and health. *Comprehensive Reviews in Food Science and Food Safety*, 13, 871–890.

Catassi, C. and Fasano, A. (2014). Coeliac disease: the debate on coeliac disease screening—are we there yet? *Nature Reviews Gastroenterology and Hepatology*, 11(8), 457.

Catassi, C., Gatti, S., Fasano, A. (2014). The new epidemiology of Celiac Disease. *Journal of Pediatric Gastroenterology and Nutrition,* 59 (Suppl. 1), S7–S9.

Catassi, C. (2015). Gluten sensitivity. *Annals of Nutrition and Metabolism*, 67 (Suppl. 2), 16-26.

Cheyne, I. (2012). Consumer labelling in EU and WTO law. Liberalising trade in EU and the WTO: a legal comparison, *Cambridge University Press 2012*, 309-322.

Clavijo, B J., Venturini, L, Schudoma, C., Accinelli, G.G., Kaithakottil, G., Wright, J., …, Clark, M.D. (2017). An improved assembly and annotation of the allohexaploid wheat genome identifies complete families of agronomic genes and provides genomic evidence for chromosomal translocations. *Genome Research,* 27, 885-896.

Cockram, J., Mackay, I. J., O'Sullivan, D.M. (2007). The role of double-stranded break repair in the creation of phenotypic diversity at cereal *VRN1* loci. *Genetics,* 177, 1-5.

Collier, R., Dasgupta, K., Xing, Y.P., Hernandez, B.T., Shao, M., Rohozinski, D., Kovak, E., Lin, J., Oliveira, M.L.P., Stover, E., McCue, K.F. (2017). Accurate measurement of transgene copy number in crop plants using droplet digital PCR. *Plant Journal*, 90(5), 1014-1025.

Cooke, R. J. (1987). The classification of wheat cultivars using a standard reference electrophoresis method. *Journal of the National Institute of Agricultural Botany*, 17(3), 273-281.

Cooke, R.J. (1995). Gel electrophoresis for the identification of plant varieties. *Journal of Chromatography A*, 698, 281-299.

Di Cagno, R., De Angelis, M., Auricchio, S., Greco, L., Clarke, C., De Vincenzi, M., Giovannini, C., D'Archivio, M., Landolfo, F., Parrilli, G., Minervini, F. (2004). Sourdough bread made from wheat and nontoxic flours and started with selected lactobacilli is tolerated in celiac sprue patients. *Applied and Environmental Microbiology*, 70, 1088-1096.

Dicke, W.K., Weijers, H.A., van de Kamer J.H. (1953). Coeliac disease. II. The presence in wheat of a factor having a deleterious effect in cases of coeliac disease. *Acta Paediatrica*, 42, 34-42.

Directive 2001/18/EC of the European Parliament and of the Council of 12 March 2001 on the deliberate release into the environment of genetically modified organisms and repealing Council Directive 90/220/EEC—*Commission Declaration. Official Journal,* 106, 1–39.

Dong, L., Huo, N., Wang, Y., Deal, K., Wang, D., Hu, T., Dvorak, J., Anderson, O.D., Luo, M.C., Gu, Y.Q.

(2016). Rapid evolutionary dynamics in a 2.8-Mb chromosomal region containing multiple prolamin and resistance gene families in Aegilops tauschii. *The Plant Journal*, 87(5), 495-506.

Doudna, J.A. and Charpentier, E. (2014). The new frontier of genome engineering with CRISPR-Cas9. *Science,* 346(6213), 1258096.

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.

EASAC (European Academies Science Advisory Council) (2015). EASAC Statement on new breeding techniques. [online] Available at: http://www.epsoweb.org/file/2105. [Accessed 14 April 2018].

EC (European Commission) (2000). Communication from the Commission on the pre-cautionary principle: COM(2000)1 final. [online] Available at: http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52000DC0001. [Accessed 11 November 2017].

ECJ (European Court of Justice) (2018a). Judgement of the Court in Case C-528/16, 25 July 2018. [online] Available at: http://curia.europa.eu/juris/documents.jsf?num=C-528/16. [Accessed 1 August 2018].

ECJ (European Court of Justice) (2018b). Opinion of Advocate General Bobek in Case C-528/16, 18 January 2018. [online] Available at: http://curia.europa.eu/juris/documents.jsf?num=C-528/16. [Accessed 1 August 2018].

Eckardt, N.A. (2010). Evolution of Domesticated Bread Wheat. *The Plant Cell,* 22, 993.

EFSA (European Food Safety Authority) (2012). Scientific opinion addressing the safety assessment of plants developed using zinc finger nuclease 3 and other site-directed nucleases with similar function. *EFSA Journal*, 10, 2943.

EFSA (European Food Safety Authority) (2015). Genetically modified organisms UNIT. Mandate Number: M-2015-0183. [online] Available at: http://registerofquestions.efsa.europa.eu/roqFrontend. [Accessed 11 November 2017].

EFSA (European Food Safety Authority) GMO UNIT (2015). EFSA response to DG SANTE request to provide technical assistance on issues related to the legal analysis of new plant breeding techniques. Mandate Number: M-2015-0183. p 4. [online] Available at: http://registerofquestions.efsa.europa.eu/roqFrontend/ListOfQuestionsNoLogin?1 [Accessed 1 August 2018].

Endo, M., Mikami, M., Toki, S. (2015). Multigene knockout utilizing off-target mutations of the CRISPR/Cas9 system in rice. *Plant and Cell Physiology*, 56(1), 41-47.

Endo, T.R. (1988). Induction of Chromosomal Structural Changes by a Chromosome of *Aegilops cylindrica* L. in Common Wheat. *Journal of Heredity,* 79(5), 66-370.

Endo, T.R., and Gill, B.S. (1996). The deletion stocks of common wheat. *Journal of Heredity,* 87, 295-307.

EPSC (European Political Strategy Centre) (2016). Towards an innovation principle endorsed by better regulation (EPSC Strategic Notes No. Issue 14). [online] Available at: https://ec.europa.eu/epsc/sites/epsc/files/strategic_note_issue_14.pdf. [Accessed 1 August 2018].

EPSO (European Plant Science Organisation) (2015). Crop genetic improvement technologies. [online] Available at: http://www.epsoweb.org/file/2147. [Accessed 16 November 2017].

Fasano, A. (2006). Systemic autoimmune disorders in celiac disease. *Current Opinion in Gastroenterology,* 22(6), 674-679.

Fasano, A. (2009). Surprises from celiac disease. *Scientific American*, 301(2), 54-61.

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

Fitzgerald, T.L., Kazan, K., Li, Z., Morell, M.K., Manners, J.M. (2010). A high-throughput method for the detection of homoeologous gene deletions in hexaploid wheat. *BMC Plant Biology*, 10(1), 264.

Forde, B.G., Heyworth, A., Pywell, J., Kreis, M. (1985). Nucleotide sequence of a B1 hordein gene and the identification of possible upstream regulatory elements in endosperm storage protein genes from barley, wheat and maize. *Nucleic Acids Research,* 13, 7327-7339.

Galili, G., Levy, A.A., Feldman, M. (1986). Gene-dosage compensation of endosperm proteins in hexaploid wheat Triticum aestivum. *Proceedings of the National Academy of Sciences*, 83(17), 6524-6528.

Gänzle, M.G., Loponen, J., Gobbetti, M. (2008). Proteolysis in sourdough fermentations: mechanisms and potential for improved bread quality. *Trends in Food Science and Technology*, 19, 513-521.

Gao, R., Feyissa, B.A., Croft, M., Hannoufa, A. (2018). Gene editing by CRISPR/Cas9 in the obligatory outcrossing *Medicago sativa. Planta*, 247(4), 1043-1050.

Gao, S., Gu, Y.Q., Wu, J., Coleman-Derr, D., Huo, N., Crossman, C., Jia, J., Zuo, Q., Ren, Z., Anderson, O.D., Kong, X. (2007). Rapid evolution and complex structural organization in genomic regions harboring multiple prolamin genes in the polyploid wheat genome. *Plant Molecular Biology*, 65, 189-203.

Garcia Giménez, S., Leitch, I.J., Anadon Rosell, A., Canela, M.Á., Gálvez, F., Garnatje i Roca, T., Gras, A., Hidalgo, O., Johnston, E., de Xaxars, G.M., Pellicer, J., Siljak-Yakovlev, S., Vallès, J., Vitales, D., Bennett, M.D. (2014). Recent updates and developments to plant genome size databases. *Nucleic Acids Research,* 42, D1159-D1166.

Gil-Humanes, J., Pistón, 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. *Proceedings of the National Academy of Sciences of the U.S.A*., 107, 17023-17028.

Gil-Humanes, J., Pistón, F., Giménez, M.J., Martín, A., Barro, F. (2012). The introgression of RNAi Silencing of gamma-gliadins into commercial lines of bread wheat changes the mixing and technological properties of the dough. *PLoS One*, 7, e45937.

Gil-Humanes, J., Pistón, F., Altamirano-Fortoul, R., Real, A., Comino, I., Sousa, C., Rosell, C.M., Barro, F. (2014). Reduced-gliadin wheat bread: an alternative to the gluten-free diet for consumers suffering gluten-related pathologies. *PLoS One*, 9, e90898.

Gil-Humanes, J., Wang, Y., Liang, Z., Shan, Q., Ozuna, C.V., Sánchez-León, S., Baltes, N.J., Starker, C., Barro, F., Gao, C., Voytas, D.F. (2017). High-efficiency gene targeting in hexaploid wheat using DNA replicons and CRISPR/Cas9. *Plant Journal*, 89, 1251–1262.

Gilissen, L.J.W.J., van der Meer, I.M., Smulders, M.J.M. (2008). Beyond coeliac disease toxicity: Detoxified and non-toxic grains. *Frontiers in Celiac Disease*. Vol. 12. Karger Publishers, 2008, 139-147.

Gilissen, L.J.W.J., van den Broeck, H.C., Londono, D.M., Salentijn, E.M.J., Koning, F., van der Meer, I.M., Smulders, M.J.M. (2012). Food-related strategies towards reduction of gluten intolerance and gluten sensitivity. *Proceedings of the 25th Meeting of the Working Group on Prolamin Analysis and Toxicity*, 29-35.

Gilissen, L.J.W.J., van der Meer, I.M., Smulders, M.J.M. (2014). Reducing the incidence of allergy and intolerance to cereals. *Journal of Cereal Science*, 59, 337-353.

Gilissen, L.J.W.J., van der Meer, I.M., Smulders, M.J.M. (2016a). Why oats are safe and healthy for Celiac Disease patients. *Medical Sciences*, 4, 21.

Gilissen, L.J.W.J., van der Meer, I.M., Smulders, M.J.M. (2016b). Strategies to reduce or prevent wheat coeliac-immunogenicity and wheat sensitivity through food. *Proceedings of the 29th Meeting of the Working Group on Prolamin Analysis and Toxicity,* 41-54.

Goncharov, N. (2011). Genus *Triticum* L. taxonomy: the present and the future. *Plant Systematics and Evolution,* 295, 1-11.

Goryunova, S.V., Salentijn, E.M.J., Chikida, N.N., Kochieva, E.Z., van der Meer, I.M., Gilissen, L.J.W.J., Smulders M.J.M. (2012). Expansion of the γ-gliadin gene family in Aegilops and Triticum. *BMC Evolutionary Biology,* 12, 215.

Greco, L., Gobbetti, M., Auricchio, R., Di Mase, R., Landolfo, F., Paparo, F., Di Cagno, R., De Angelis, M., Rizzello, C.G., Cassone, A., Terrone, G. (2011). Safety for patients with celiac disease of baked goods made of wheat flour hydrolysed during food processing. *Clinical Gastroenterology and Hepatology,* 9, 24-29.

Gu, Y.Q., Crossman, C., Kong, X., Luo, M., You, F.M., Coleman-Derr, D., Dubcovsky, J., Anderson, O.D. (2004). Genomic organization of the complex α-gliadin gene loci in wheat. *Theoretical and Applied Genetics*, 109, 648-657.

Gupta, P.K., Mir, R.R., Mohan, A., Kumar, J. (2008). Wheat genomics: present status and future prospects. *International Journal of Plant Genomics,* 36, 896451.

Gustafson, K.L. (2016). Impact of ingredients on quality and sensory characteristics of gluten-free baked goods. Kansas state exchange research.

Hardy, M.Y., Tye-Din, J.A., Stewart, J.A. (2015). Ingestion of oats and barley in patients with celiac disease mobilizes cross-reactive T cells activated by avenin peptides and immuno-dominant hordein peptides. *Journal of Autoimmunity*, 56, 56–65.

Hartmann, G., Koehler, P., Wieser, H. (2006). Rapid degradation of gliadin peptides toxic for coeliac disease patients by proteases from germinating cereals. *Journal of Cereal Science*, 44, 368-371.

Hausch, F., Shan, L., Santiago, N.A., Gray, G.M., Khosla, C. (2002). Intestinal digestive resistance of immunodominant gliadin peptides. *American Journal of Physiology-Gastrointestinal and Liver Physiology*, 283(4), G996-G1003.

Heap, B. (2013). Europe should rethink its stance on GM crops. *Nature*, 498, 409.

Henry, I.M., Nagalakshmi, U., Lieberman, M.C., Ngo, K.J., Krasileva, K.V., Vasquez-Gross, H., Akhunova, A., Akhunov, E., Dubcovsky, J., Tai, T.H., Comai, L. (2014). Efficient genome-wide detection and cataloging of EMS-induced mutations using exome capture and next-generation sequencing. *The Plant Cell*, 26, 1382-1397.

Hindson, B.J., Ness, K.D., Masquelier, D.A., Belgrader, P., Heredia, N.J., Makarewicz, A.J., Bright, I.J., Lucero, M.Y., Hiddessen, A.L., Legler, T.C., Kitano, T.K. (2011). High-throughput droplet digital PCR system for absolute quantitation of DNA copy number. *Analytical Chemistry*, 83(22), 8604-8610.

Hollon, J., Puppa, E.L., Greenwald, B., Goldberg, E., Guerrerio, A., Fasano, A. (2015). Effect of gliadin on permeability of intestinal biopsy explants from celiac disease patients and patients with non-celiac gluten sensitivity. *Nutrients*, 7, 1565-1576.

Horstmann, S.W., Belz, M.C.E., Heitmann, M., Zannini, E., Arendt, E.A. (2016). Fundamental study on the impact of gluten-free starches on the quality of gluten-free model breads. *Foods*, 5, 30.

Huang, S., Sirikhachornkit, A., Su, X., Faris, J., Gill, B., Haselkorn, R., Gornicki, P. (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(12), 8133-8138.

Husby, S., Koletzko, S., Korponay-Szabó, I.R., Mearin, M.L., Phillips, A., Shamir, R., Troncone, R., Giersiepen, K., Branski, D., Catassi, C., Lelgeman, M.,Mäki, M, Ribes-Koninckx, C., Ventura, A., Zimmer, K.P. (2012). European Society for Pediatric Gastroenterology, Hepatology, and Nutrition Guidelines for the Diagnosis of Coeliac Disease. *Journal of Pediatric Gastroenterology and Nutrition*, 54, 136–160.

IWGSC (International Wheat Genome Sequencing Consortium) (2018). Shifting the limits in wheat research and breeding through a fully annotated and anchored reference genome sequence. *Science*, 361(6403), eaar7191.

Ishida, Y., Tsunashima, M., Hiei, Y., Komari, T. (2015). Wheat (*Triticum aestivum* L.) transformation using immature embryos*, Agrobacterium Protocols,* 189-198, 3rd Edn. Springer, New York.

Ishii, T. and Araki, M. (2016). Consumer acceptance of food crops developed by genome editing. *Plant Cell Reports*, 35, 1507-1518.

Jones, H., Gosman, N., Horsnell, R., Rose, G.A., Everest, L.A., Bentley, A.R., Tha, S., Uauy, C., Kowalski, A., Novoselovic, D., Simek, R. (2013). Strategy for exploiting exotic germplasm using genetic, morphological, and environmental diversity: the *Aegilops tauschii* Coss. example. *Theoretical and Applied Genetics*, 126, 1793-1808.

Jouanin, A., Gilissen, L.J., Boyd, L.A., Cockram, J., Leigh, F.J., Wallington, E.J., van den Broeck, H.C., van der Meer, I.M., Schaart, J.G., Visser, R.G., Smulders, M.J. (2018). Food processing and breeding strategies for coeliac-safe and healthy wheat products. *Food Research International*, 110, 11-21.

Juhász, A., Belova, T., Florides, C.G., Maulis, C., Fischer, I., Gell, G., Birinyi, Z., Ong, J., Keeble-Gagnère, G., Maharajan, A., Ma, W., Gibson, P., Jia, J., Lang, D., Mayer, K.F.X., Spannagl M., IWGSC, Tye-Din, J.A., Appels, R., Olsen, O.A. (2018). Genome mapping of seed-borne allergens and immunoresponsive proteins in wheat. *Science advances*, 4(8), p.eaar8602.

Jupe, F., Witek, K., Verweij, W., Śliwka, J., Pritchard, L., Etherington, G.J., Maclean, D., Cock, P.J., Leggett, R.M., Bryan, G.J., Cardle, L. (2013). Resistance gene enrichment sequencing (RenSeq) enables reannotation of the NB-LRR gene family from sequenced plant genomes and rapid mapping of resistance loci in segregating populations. *Plant Journal*, 76, 530-544.

Kanerva, P., Brinck, O., Sontag-Strohm, T., Salovaara, H., Loponen, J. (2011). Deamidation of gluten proteins and peptides decreases the antibody affinity in gluten analysis assays. *Journal of Cereal Science*, 53, 335-339.

Karlen, Y., McNair, A., Perseguers, S., Mazza, C., Mermod, N. (2007). Statistical significance of quantitative PCR. *BMC Bioinformatics*, 8(1), 131.

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? *Journal of Agricultural and Food Chemistry,* 61, 1155-1159.

Kaukinen, K., Collin, P., Huhtala, H., Mäki, M. (2013). Long-term consumption of oats in adult celiac

disease patients. *Nutrients*, *5*, 4380–4389.

Kim, J.H., Park, S.B., Roh, H.J., Park, S., Shin, M.K., Im Moon, G., Hong, J.H., Kim, H.Y. (2015). A simplified and accurate detection of the genetically modified wheat MON71800 with one calibrator plasmid. *Food Chemistry,* 176, 1-6.

Kim, J.S. (2018). Precision genome engineering through adenine and cytosine base editing. *Nature plants*, 1.

Koning, F. (2005). Celiac Disease: Caught Between a Rock and a Hard Place. *Gastroenterology*, 129, 1294 1301

Koning, F. (2012). Celiac disease: quantity matters. *In Seminars in immunopathology*, 34(4), 541-549, Springer-Verlag.

Koning, F. (2015). Adverse Effects of Wheat Gluten. *Annals of Nutrition and Metabolism,* 67 (suppl. 2), 8-14.

Krasileva, K.V., Vasquez-Gross, H.A., Howell, T., Bailey, P., Paraiso, F., Clissod, L., Simmonds, J., Ramirez-Gonzalex, R.H., Wang X., Borrill, P., Fosker, C., Ayling, S., Phillips, A.L., Uauy, C., Dubcovsky, J. (2017). Uncovering hidden variation in polyploid wheat. *Proceedings of the National Academy of Sciences of the U S A*, 114, E913–E921.

Kratzer, W., Kibele, M., Akinli, A., Porzner, M., Boehm, B.O., Koenig, W., Oeztuerk, S., Mason, R.A.,

Mao, R., Haenle, M.H. (2013). Prevalence of celiac disease in Germany: A prospective follow-up study. *World Journal of Gastroenterology*, 19, 2612-2620.

Kronberger, N., Wagner, W., Nagata, M. (2014). How natural is “more natural”? the role of method, type of transfer, and familiarity for public perceptions of cisgenic and transgenic modification. *Science Communication*. 36, 106-130.

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

Laass, M.W., Schmitz, R., Uhlig, H.H., Zimmer, K.P., Thamm, M., Koletzko, S. (2015). The prevalence of celiac disease in children and adolescents in Germany: results from the KiGGS study. *Deutsches Ärzteblatt International*, 112, 553-560.

Lagrain, B., Brunnbauer, M., Rombouts, I., Koehler, P. (2013). Identification of intact high molecular weight glutenin subunits from the wheat proteome using combined liquid chromatography-electrospray ionization mass spectrometry. *PLoS ONE*, 8, e58682.

Laursen, L. (2016). Will Europe toast GM wheat for gluten sufferers? *Nature Biotechnology*, 34, 369-371.

Lawler, M. (2015). Silencing the voice of scientific reason. *The Lancet Oncology*, 16, 4-5.

Lawrenson, T., Shorinola, O., Stacey, N., Li, C., Østergaard, L., Patron, N., Uauy, C., Harwood, W. (2015). Induction of targeted, heritable mutations in barley and *Brassica oleracea* using RNA-guided Cas9 nuclease. *Genome Biology*, 16, 258

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., 1000 Genome Project Data Processing Subgroup. (2009). The Sequence alignment/map (SAM) format and SAMtools, *Bioinformatics*, 25(16), 2078-9.

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM, *arXiv preprint arXiv,* 1303.3997.

Li, W., Zhang, P., Fellers, J.P., Friebe, B., Gill, B.S. (2004). Sequence composition, organization, and evolution of the core *Triticeae* genome. *The Plant Journal*, 40(4), 500-511.

Liang, Z., Chen, K., Li, T., Zhang, Y., Wang, Y., Zhao, Q., Liu, J., Zhang, H., Liu, C., Ran, Y., Gao, C. (2017). Efficient DNA-free genome editing of bread wheat using CRISPR/Cas9 ribonucleoprotein complexes. *Nature Communications*, 8, 14261.

Lohi, S., Mustalahti, K., Kaukinen, K., Laurila, K., Collin, P., Rissanen, H., Lohi, O., Bravi, E., Gasparin,

M., Reunanen, A., Mäki, M. (2007). Increasing prevalence of coeliac disease over time. *Alimentary pharmacology & therapeutics*, 26(9), 1217-1225.

Londono, D.M., van ‘t Westende, W.P.C., Goryunova, S.V., Salentijn, E.M.J., van den Broeck, H.C., van der Meer, I.M., Gilissen, L.J.W.J., Smulders, M.J.M. (2013). Avenin diversity analysis of the genus *Avena* (oat). Relevance for people with celiac disease. *Journal of Cereal Science*, 58, 170–177.

Loponen, J. (2006). Prolamin degradation in sourdoughs. Academic Dissertation. EKT series 1372. University of Helsinki.

Loponen, J., Sontag-Strohm, T., Venäläinen, J., Salovaara, H. (2007). Prolamin hydrolysis in wheat sourdoughs with differing proteolytic activities. *Journal of Agricultural and Food Chemistry*, 55, 978-984.

Loponen, J., Kanerva, P., Zhang, C., Sontag-Strohm, T., Salovaara, H., Gänzle, M.G. (2009). Prolamin hydrolysis and pentosan solubilization in germinated-rye sourdoughs determined by chromatographic and immunological methods. *Journal of Agricultural and Food Chemistry*, 57, 746-753.

Lucht, J.M. (2015). Public acceptance of plant biotechnology and GM crops. *Viruses,* 7, 4254–4281.Lundin, K.E. and Alaedini, A. (2012). Non-celiac gluten sensitivity. *Gastrointestinal Endoscopy Clinics*, 22(4), 723-734.

Ma, X., Zhang, Q., Zhu, Q., Liu, W., Chen, Y., Qiu, R., Wang, B., Yang, Z., Li, H., Lin, Y., Xie, Y., Shen, R., Chen, S., Wang, Z., Chen, Y., Guo, J., Chen, L., Zhao, X., Dong, Z., Liu, Y.-G. (2015). A robust CRISPR/Cas9 system for convenient, high-efficiency multiplex genome editing in monocot and dicot plants. *Molecular Plant*, 8(8), 1274-1284.

Mäki, M., Mustalahti, K., Kokkonen, J., Kulmala, P., Haapalahti, M., Karttunen, T., Ilonen, J., Laurila, K., Dahlbom, I., Hansson, T., Höpfl, P., Knip, M. (2003). Prevalence of Celiac Disease among children in Finland. *New England Journal of Medicine*, 348, 2517-2524.

Milner, M.J., Howells, R.M., Craze, M., Bowden, S., Graham, N. and Wallington, E.J. (2018). A PSTOL-like gene in wheat, *TaPSTOL*, controls a number of agronomically important traits in wheat. *BMC Plant Biology*, 18(1), 115.

Manoj, P. (2016). Droplet digital PCR technology promises new applications and research areas. *Mitochondrial DNA* Part A, 27(1), 742-746.

McCallum, C.M., Comai, L., Greene E.A., Henikoff, S. (2000). Targeting Induced Local Lesions IN Genomes (TILLING) for plant functional genomics. *Plant Physiology,* 123, 439–442.

McDermott, G.P., Do, D., Litterst, C.M., Maar, D., Hindson, C.M., Steenblock, E.R., Legler, T.C., Jouvenot, Y., Marrs, S.H., Bemis, A., Shah, P. (2013). Multiplexed target detection using DNA-binding dye chemistry in droplet digital PCR. *Analytical chemistry*, 85(23), 11619-11627.

McDougall, P. (2011). The cost and time involved in the discovery, development and authorisation of a new plant biotechnology derived trait. *Crop Life International*, 1-24. [online] Available at: https://croplife.org/wp-content/uploads/pdf_files/Getting-a-Biotech-Crop-to-Market-Phillips-McDougall-Study.pdf. [Accessed 1 August 2018].

Mitea, C., Salentijn, E.M., van Veelen, P., Goryunova, S.V., van der Meer, I.M., van den Broeck, H.C., Mujico, J.R., Monserrat, V., Gilissen, L.J., Drijfhout, J.W., Dekking, L. (2010). A universal approach to eliminate antigenic properties of alpha-gliadin peptides in celiac disease. *PloS One*, 5, e15637.

Moehs, C.P., Austill, W.J., Holm, A., Large, T.A., Loeffler, D., Mullenberg, J., Schnable, P.S., Skinner, W., van Boxtel, J., Wu, L., McGuire, C. (2018). Development of reduced gluten wheat enabled by determination of the genetic basis of the *lys3a* low hordein barley mutant. *bioRxiv*, 354548.

Molberg, Ø, Mcadam, S.N., Körner, R., Quarsten, H., Kristiansen, C., Madsen, L., Fugger, L., Scott, H., Norén, O., Roepstorff, P., Lundin, K.E.A., Sjöström, H., Sollid, L.M. (1998). Tissue transglutaminase selectively modifies gliadin peptides that are recognized by gut-derived T cells in celiac disease. *Nature Medicine,* 4, 713–717.

Molberg, Ø., Uhlen, A.K., Jensen, T., Flæte, N.S., Fleckenstein, B., Arentz–Hansen, H., Raki, M., Lundin, K.E.A., Sollid, L.M. (2005). Mapping of gluten T-cell epitopes in the bread wheat ancestors: Implications for celiac disease. *Gastroenterology*, 128(2), 393-401.

Morita, R., Kusaba, M., Iida, S., Yamaguchi, H., Nishio, T., Nishimura, M. (2009). Molecular characterization of mutations induced by gamma irradiation in rice. *Genes and Genetics Systems*, 84, 361-370.

Murray, J.A., Van Dyke, C., Plevak, M.F., Dierkhising, R.A., Zinsmeister, A.R., Melton III, L.J. (2003).

Trends in the identification and clinical features of celiac disease in a North American community, 1950–2001. *Clinical Gastroenterology and Hepatology*, 1(1), 19-27.

Mustalahti, K., Catassi, C., Reunanen, A., Fabiani, E., Heier, M., McMillan, S., Murray, L., Metzger, M.H., Gasparin, M., Bravi, E., Mäki, M. (2010). The prevalence of celiac disease in Europe: Results of a centralized, international mass screening project. *Annals of Medicine,* 42, 587-595.

Ogbonnaya, F.C., Abdalla, O., Mujeeb-Kazi, A., Kazi, A.G., Xu, S.S., Gosman, N., Lagudah, E.S., Bonnett, D., Sorrells, M.E., Tsujimoto, H. (2013). Synthetic hexaploids: harnessing species of the primary gene pool for wheat improvement. *Plant Breeding Reviews*, 37, 35-122.

Oladosu, Y., Rafii, M.Y., Abdullah, N., Hussin, G., Ramli, A., Rahim, H.A., Miah, G., Usman, M., (2016). Principle and application of plant mutagenesis in crop improvement: a review. *Biotechnology and Biotechnological Equipment*, 30, 1-16.

Owen, R., Macnaghten, P., Stilgoe, J. (2012). Responsible research and innovation: From science in society to science for society, with society. *Science and Public Policy*, 39, 751-760.

Ozuna, C.V., Iehisa, J., Giménez, 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 Journal,* 82(5), 794-805.

Ozuna, C.V. and Barro, F. (2018). Characterization of gluten proteins and celiac disease-related immunogenic epitopes in the Triticeae: cereal domestication and breeding contributed to decrease the content of gliadins and gluten. *Molecular Breeding*, 38(3), 22.

Paux, E., Roger, D., Badaeva, E., Gay, G., Bernard, M., Sourdille, P., Feuillet, C. (2006). Characterizing the composition and evolution of homoeologous genomes in hexaploid wheat through BAC-end sequencing on chromosome 3B. *The Plant Journal*, 48(3), 463-474.

Payne, P.I., Holt, L.M., Worland, A.J., Law, C.N. (1982). Structural and genetical studies on the high-molecular-weight subunits of wheat glutenin. Part 3. Telocentric mapping of the subunit genes on the long arms of the homoeologous group 1 chromosomes. *Theoretical and Applied Genetics*, 63(2), 129-138.

Payne, P.I., Jackson, E.A., Holt, L.M. (1984). The association between γ-gliadin 45 and gluten strength in durum wheat varieties: A direct causal effect or the result of genetic linkage? *Journal of Cereal Science,* 2(2), 73-81.

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

Perochon A., Jianguang J., Kahla A., Arunachalam C., Scofield S., Bowden B., Wallington E., Doohan F. (2015). TaFROG encodes a Pooideae orphan protein that interacts with SnRK1 and enhances resistance to the mycotoxigenic fungus *Fusarium graminearum*. *Plant Physiology*, 169, 2895-2906.

Petersen, J., Montserrat, V., Mujico, J.R., Loh, K.L., Beringer, D.X., van Lummel, M., Thompson, A., Mearin, M.L., Schweizer, J., Kooy-Winkelaar, Y., van Bergen, J. (2014). T-cell receptor recognition of HLA-DQ2–gliadin complexes associated with celiac disease. *Nature Structural and Molecular Biology*, 21, 480-488.

Petersen, J., Kooy-Winkelaar, Y., Loh, K.L., Tran, M., van Bergen, J., Koning, F., Rossjohn, J., Reid, H.H. (2016). Diverse T cell receptor gene usage in HLA-DQ8-associated Celiac Disease converges into a consensus binding solution. *Structure*, 24, 1643-1657.

Pinto-Sánchez, M.I., Causada-Calo, N., Bercik, P., Ford, A.C., Murray, J.A., Armstrong, D., Semrad, C., Kupfer, S.S., Alaedini, A., Moayyedi, P., Leffler, D.A., Verdú, E.F., Green, P. (2017). Safety of adding oats to a gluten-free diet for patients with celiac disease: Systematic review and meta-analysis of clinical and observational studies. *Gastroenterology*, 153(2), 395-409.

Pistón, F., Gil-Humanes, J., Rodríguez-Quijano, M., Barro, F. (2011). Down-regulating γ-gliadins in bread wheat leads to non-specific increases in other gluten proteins and has no major effect on dough gluten strength. *PLoS One*, 6, e24754.

Puchta, H. (2016). Using CRISPR/Cas in three dimensions: towards synthetic plant genomes, transcriptomes and epigenomes. *The plant journal*, 87(1), 5-15.

Qi, Y., Li, X., Zhang, Y., Starker, C.G., Baltes, N.J., Zhang, F., Sander, J.D., Reyon, D., Joung, J.K., Voytas, D.F. (2013). Targeted deletion and inversion of tandemly arrayed genes in *Arabidopsis thaliana* using zinc finger nucleases. *G3: Genes, Genomes, Genetics*, 8(10), 3-113.

Qiao, S.W., Sollid, L.M., Blumberg, R.S. (2009). Antigen presentation in celiac disease. *Current Opinionin Immunology*., 21, 111–117.

Quinlan, A.R. and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 26(6), 841-842.

Racine, J. (2006). gnuplot 4.0: a portable interactive plotting utility. *Journal of Applied Econometrics*, 21(1), 133-141.

Ratcliff, F., Harrison, B.D., Baulcombe, D.C. (1997). A similarity between viral defence and gene silencing in plants. *Science,* 276, 1558-1560.

Reif, J.C., Zhang, P., Dreisigacker, S., Warburton, M.L., van Ginkel, M., Hoisington, D., Bohn, M., Melchinger, A.E. (2005). Wheat genetic diversity trends during domestication and breeding. *Theoretical and Applied Genetics*, 110, 859-864.

Reunala, T. (1998). Dermatitis herpetiformis: coeliac disease of the skin. *Annals of Medicine,* 30(5), 416-418.

Rewers, M. (2005). Epidemiology of celiac disease: what are the prevalence, incidence, and progression of celiac disease? *Gastroenterology*, 128(4), S47-S51.

Ribeiro, M., Nunes, F.M., Rodriguez-Quijano, M., Carrillo, J.M., Branlard, G., Igrejas, G. (2018). Next-generation therapies for celiac disease: The gluten-targeted approaches. *Trends in Food Science and Technology*, 75, 56-71.

Risacher, T., Craze, M., Bowden, S., Paul, W., Barsby, T. (2009). Highly efficient *Agrobacterium* -mediated transformation of wheat via *In Planta* inoculation. In H.D. Jones, and P.R. Shewry (Eds.), *Methods in Molecular Biology, Transgenic Wheat, Barley and Oats,* 478, 115-124).

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

Salentijn, E.M.J., Goryunova, S.V., Bas, N., van der Meer, I.M., van den Broeck, H.C., Bastien, T., Gilissen, L.J.W.J., Smulders, M.J.M. (2009). Tetraploid and hexaploid wheat varieties reveal large differences in expression of alpha-gliadins from homoeologous *Gli-2* loci. *BMC Genomics*, 10, 48.

Salentijn, E.M.J., Mitea, D.C., Goryunova, S.V., van der Meer, I.M., Padioleau, I., Gilissen, L.J.W.J., Koning, F., Smulders, M.J.M. (2012). Celiac disease T-cell epitopes from γ-gliadins: immunoreactivity depends on the genome of origin, transcript frequency, and flanking protein variation. *BMC Genomics*, 13, 277.

Salentijn, E.M.J., Esselink, D.G., Goryunova, S.V., van der Meer, I.M., 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.

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

Sapone, A., Bai, J.C., Ciacci, C., Dolinsek, J., Green, P.H., Hadjivassiliou, M., Kaukinen, K., Rostami, K., Sanders, D.S., Schmann, M., Ullrich,R., Villalta, D., Volta. U., Catassi, C., Fasano, A. (2012). Spectrum of gluten-related disorders: consensus on new nomenclature and classification. *BMC Medicine*, 10, 13.

Saturni, L., Ferretti, G., Bacchetti, T. (2010). The gluten-free diet: safety and nutritional quality. *Nutrients*, 2, 16-34.

Schaart, J.G., van de Wiel, C.C. M., Lotz, L.A.P., Smulders, M.J.M. (2016). Opportunities for products of new plant breeding techniques. *Trends in Plant Science*, 21, 438–449.

Schaeffer, S.M and Nakata, P.A. (2015). CRISPR/Cas9-mediated genome editing and gene replacement in plants: transitioning from lab to field. *Plant Science*, 240, 130–142.

Schenk, M.F., Fischer, A.R.H., Frewer, L.J., Gilissen, L.J.W.J., Jacobsen, E., Smulders, M.J.M. (2008). The influence of perceived benefits on acceptance of GM applications for allergy prevention. *Health, Risk and Society*, 10, 263-282.

Schenk, M.F., van der Maas, M.P., Smulders, M.J.M., Gilissen, L.J.W.J., Fischer, A,H,R., van der Lans, I.A., E Jacobsen, E., Frewer, L.J. (2011). Consumer attitudes towards hypoallergenic apples that alleviate mild apple allergy. *Food Quality and Preference*, 22, 83–91.

Scherf, K.A., Wieser, H., Koehler, P. (2016). Improved quantitation of gluten in wheat starch for Celiac Disease patients by gel-permeation high-performance liquid chromatography with fluorescence detection (GP-HPLC-FLD). *Journal of Agricultural and Food Chemistry*, 64, 7622-7631.

Schwalb, T., Wieser, H., Koehler, P. (2012). Studies on the gluten-specific peptidase activity of germinated grains from different cereal species and cultivars. *European Food Research and Technology,* 235, 1161-1170.

Sears, E.R. (1966). Nullisomic-Tetrasomic Combinations in Hexaploid Wheat. In: Riley R, Lewis KR (Eds) *Chromosome Manipulations and Plant Genetics*, 29-45, Springer US.

Segovia V., Hubbard A., Craze M., Bowden S., Wallington E.J., Bryant R., Greenland A., Bayles R., Uauy C. (2014). *Yr36* confers partial resistance at temperatures below 18°C to UK isolates of *Puccinia striiformis*. *Phytopathology*, 104 (8), 871-878.

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

Shaw, L.M., Turner, A.S., Herry, L., Griffiths, S., Laurie, D.A. (2013). Mutant alleles of Photoperiod-1 in wheat (*Triticum aestivum L.*) that confer a late flowering phenotype in long days. *PLoS One*, 8(11), p.e79459.

Shewry, P.R., Miflin, B.J., Kasarda, D.D. (1984). The structural and evolutionary relationships of the prolamin storage proteins of barley, rye and wheat. *Philosophical Transactions of the Royal Society London B*, 304(1120), 297-308.

Shewry, P.R., Halford, N.G., Lafiandra, D. (2003). The genetics of wheat gluten proteins. In J.C. Hall, J.C. Dunlap, T. Friedman (Eds.), *Advances in Genetics,* 49, 111-184, Academic Press.

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

Shewry, P.R., D'Ovidio, R., Lafiandra, D., Jenkins, J.A., Mills, E.N.C., Békés, F. (2009). Wheat grain proteins. In K. Khan, P.R. Shewry (Eds.), *Wheat: Chemistry and Technology*, 223-298, (4th edn.), A.A.C.C, St. Paul, MN., U.S.A.

Shewry, P.R. and Tatham, A.S. (2016). Improving wheat to remove coeliac epitopes but retain functionality. *Journal of Cereal Science*, 67, 12–21.

Shu, Q.Y., Forster, B.P., Nakagawa, H. (2011). (Eds) *Plant Mutation Breeding and Biotechnology*, CABI.

Simon, F. (2013). Chief EU scientist backs damning report urging GMO 'rethink'. [online] Available at:https://www.euractiv.com/section/science-policymaking/news/chief-eu-scientist-backs-damning-report-urging-gmo-rethink/. [Accessed 28 August 2018].

Smulders, M.J.M., Jouanin, A., Schaart, J., Visser, R.G.F., Cockram, J., Leigh, F., Wallington, E., Boyd, L.A., van den Broeck, H.C., van der Meer, I.M., Gilissen, L.J.W.J. (2015). Development of wheat varieties with reduced contents of celiac-immunogenic epitopes through conventional and GM strategies, *Proceedings of the 28th Meeting of the Working Group on Prolamin Analysis and Toxicity*, 47-56.

Smulders, M.J.M., van de Wiel, C.C.M., van den Broeck, H.C., van der Meer, I., Israel-Hoevelakena, T.P.M., Timmer, R.D., van Dinter, B-J., Braun S., Gilissen, L.J.W.J. (2018). Oats in healthy gluten-free and regular diets: a perspective. *Food Research International*, 110, 3-10.

SNSF (Swiss National Science Foundation) (2012). Benefits and risks of the deliberate release of genetically modified plants. National Research Programme NRP 59. [online] Available at: http://www.nfp59.ch/e_index.cfm. [Accessed 11 Nov 2017].

Sollid, L.M., Qiao, S.-W., Anderson, R.P., Gianfrani, C., Koning, F. (2012). Nomenclature and listing of celiac disease relevant gluten T-cell epitopes restricted by HLA-DQ molecules. *Immunogenetics*, 64(6), 455-460.

Spaenij–Dekking, L., Kooy–Winkelaar, Y., van Veelen, P., Drijfhout, J.W., Jonker, H., van Soest, L.,

Smulders, M.J.M., Bosch, D., Gilissen, L.J.W.J., Koning, F. (2005). Natural variation in toxicity of wheat: potential for selection of nontoxic varieties for celiac disease patients. *Gastroenterology*, 129, 797-806.

Sprink, T., Metje, J., Hartung, F. (2015). Plant genome editing by novel tools: TALEN and other sequence specific nucleases. *Current Opinion in Biotechnology*, 32, 47–53.

Sprink, T., Eriksson, D., Schiemann, J., Hartung, F. (2016a). Regulatory hurdles for genome editing: process- vs. product-based approaches in different regulatory contexts. *Plant Cell Reports*, 35, 1493-1506.

Sprink, T., Metje, J., Schiemann, J., Hartung, F. (2016b). Plant genome editing in the European Union - to be or not to be - a GMO. *Plant Biotechnology Reports*, 10, 345-335.

Stenman, S.M., Lindfors, K., Venäläinen, J.I., Hautala, A., Männistö, P.T., Garcia-Horsman, J.A., Kaukovirta-Norja, A., Auriola, S., Mauriala, T., Mäki, M., Kaukinen, K. (2010). Degradation of celiac disease-inducing rye secalin by germinating cereal enzymes: diminishing toxic effects in intestinal epithelial cells. *Clinical and Experimental Immunology*, 161, 242-249.

Steuernagel, B., Witek, K., Jones, J.D., Wulff, B.B. (2017). MutRenSeq: A Method for Rapid Cloning of Plant Disease Resistance Genes. *Wheat Rust Diseases*, Humana Press, New York, NY, 215-229.

Sun, Y., Zhang, X., Wu, C., He, Y., Ma, Y., Hou, H., Guo, X., Du, W., Zhao, Y., Xia, L. (2016). Engineering herbicide-resistant rice plants through CRISPR/Cas9-mediated homologous recombination of acetolactate synthase. *Molecular Plant*, 9(4), 628-631.

Sutton, K.H., Simmons, L.D., Cummack, J.B., Roberts, S.J. (2015). Production of flours with reduced epitope content using milling technology. *Cereal Chemistry*, 93, 352-356.

Tanner, G.J., Blundell, M.J., Colgrave, M.L., Howitt, C.A. (2016). Creation of the first ultra-low gluten barley (*Hordeum vulgare* L.) for coeliac and gluten-intolerant populations. *Plant Biotechnology Journal*, 14, 1139-1150.

Tye-Din, J.A., Stewart, J.A., Dromey, J.A., Beissbarth, T., van Heel, D.A., Tatham, A., Henderson, K., Mannering, S.I., Gianfrani, C., Jewell, D.P., Hill, A.V. (2010). Comprehensive, quantitative mapping of T cell epitopes in gluten in celiac disease. *Science Translational Medicine,* 2, 41ra51-41ra51.

Uhde, M., Ajamian, M., Caio, G., De Giorgio, R., Indart, A., Green, P.H., Verna, E.C., Volta, U., Alaedini, A. (2016). Intestinal cell damage and systemic immune activation in individuals reporting sensitivity to wheat in the absence of coeliac disease. *Gut,* 311964.

Upchurch, H.F., Conway, E., Patterson Jr, M.K., Maxwell, M.D. (1991). Localization of cellular transglutaminase on the extracellular matrix after wounding: characteristics of the matrix bound enzyme. *Journal of cellular physiology*, 149, 375-382.

USDA (2017). Importation, interstate movement, and environmental release of certain genetically engineered organisms. A proposed rule by the Animal and Plant Health Inspection Service on 01/19/2017. Document Citation 82 FR 7008; Document Number 2017-00858. [online] Available at https://www.federalregister.gov/documents/2017/01/19/2017-00858/importation-interstate-movement-and-environmental-release-of-certain-genetically-engineered. [Accessed 1 August 2018].

van de Wal, Y., Kooy, Y., van Veelen, P., Peña, S., Mearin, L., Papadopoulos, G., Koning, F. (1998) Selective deamidation by tissue transglutaminase strongly enhances gliadin-specific T cell reactivity. *Journal of Immunology,* 161, 1585-1588.

van de Wiel, C.C.M., Schaart, J.G., Lotz, L.A.P., Smulders, M.J.M. (2017). New traits in crops produced by genome editing techniques based on deletions. *Plant Biotechnology Reports,* 11, 1-8.

van den Broeck, H.C., van Herpen, T.W.J.M., Schuit, C., Salentijn, E.M.J., Dekking, L., Bosch, D., Hamer, R.J., Smulders, M.J.M, Gilissen, L.J.W.J., van der Meer, I.M. (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.

van den Broeck, H.C., de Jong, H.C., Salentijn, E.M.J., Dekking, L., Bosch, D., Hamer, R.J., Gilissen, L.J.W.J., van der Meer, I.M., Smulders, M.J.M. (2010a). 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.

van den Broeck, H.C., Hongbing, C., Lacaze, X., Dusautoir, J.C., Gilissen, L.J.W.J., Smulders, M.J.M., van der Meer, I.M. (2010b). In search of tetraploid wheat accessions reduced in celiac disease-related gluten epitopes. *Molecular Biosystems,* 6, 2206-2213.

van den Broeck, H.C., Gilissen, L.J.W.J., Smulders, M.J.M., van der Meer, I.M., Hamer, R.J. (2011). Dough quality of bread wheat lacking alpha-gliadins with celiac disease epitopes and addition of celiac-safe avenins to improve dough quality. *Journal of Cereal Science*, 53, 206-216.

van den Broeck, H.C., Cordewener, J.H., Nessen, M.A., America, A.H., 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.

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

van Herpen, T.W.J.M, Riley, M., Sparks, C., Jones, H.D., Gritsch, C., Dekking, E.H., Hamer, R.J., Bosch, D., Salentijn, E.M.J., Smulders, M.J.M., Shewry, P.R. (2008). Detailed analysis of the expression of an alpha-gliadin promoter and the deposition of alpha-gliadin protein during wheat grain development. *Annals of Botany*, 102, 331-342.

Vartdal, F., Johansen, B.H., Friede, T., Thorpe, C.J., Stevanović, S., Eriksen, J.E., Sletten, K., Thorsby, E., Rammensee, H.-G., Sollid, L.M. (1996). The peptide binding motif of the disease associated HLA-DQ (α 1* 0501, β 1* 0201) molecule. *European Journal of Immunology*, 26, 2764-2772.

Vincentini, O., Borrelli, O., Silano, M., Gazza, L., Pogna, N., Luchetti, R., De Vincenzi, M. (2009). T-cell response to different cultivars of farro wheat, *Triticum turgidum ssp. Dicoccum*, in celiac disease patients. *Clinical Nutrition*, 28, 272-277.

Vriezinga, S.L., Schweizer, J.J., Koning, F., Mearin, M.L. (2015). Coeliac disease and gluten-related disorders in childhood. *Nature Reviews Gastroenterology and Hepatology,* 12, 527–536.

Walter, T., Wieser, H., Koehler, P. (2014). Degradation of gluten in wheat bran and bread drink by means of a proline-specific peptidase. *Journal of Nutrition and Food Sciences,* 4, 293.

Wang, H., Yang, H., Shivalila, C.S., Dawlaty, M.M., Cheng, A.W., Zhang, F., Jaenisch, R. (2013a). One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell*, 153(4), 910-918.

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

Wang, W., Akhunova, A., Chao, S., Akhunov, E. (2016). Optimizing multiplex CRISPR/Cas9-based genome editing for wheat. *BioRxiv.*

Wang, W., Pan, Q., He, F., Akhunova, A., Chao, S., Trick, H., Akhunov, E. (2018). Transgenerational CRISPR-Cas9 Activity Facilitates Multiplex Gene Editing in Allopolyploid Wheat. *The CRISPR Journal*, 1(1), 65-74.

Wang, Y., Cheng, X., Shan, Q., Zhang, Y., Liu, J., Gao, C., Qiu, J-L. (2014). Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew. *Nature Biotechnology,* 32, 947-951.

Waterhouse, P.M. and Helliwell, C.A. (2003). Exploring plant genomes by RNA-induced gene silencing. *Nature Reviews Genetics*, 4, 29-38.

Weaver, S., Dube, S., Mir, A., Qin, J., Sun, G., Ramakrishnan, R, Jones, R.C., Livak, K.J. (2010). Taking qPCR to a higher level: analysis of CNV reveals the power of high throughput qPCR to enhance quantitative resolution. *Methods*, 50(4), 271-276.

Weegels, P.L., van de Pijpekamp, A.M., Graveland, A., Hamer, R.J., Schofield, J.D. (1996). Depolymerisation and re-polymerisation of wheat glutenin during dough processing. I. Relationships between glutenin macropolymer content and quality parameters. *Journal of Cereal Science*, 23(2), 103-111.

Wegel, E. and Shaw, P.J. (2005). Chromosome organization in wheat endosperm and embryo. *Cytogenetic and genome research*, 109(1-3), 175-180.

Wen, S., Wen, N., Pang, J., Langen, G., Brew-Appiah, R.A., Mejias, J.H., Osorio, C., Yang, M., Gemini, R., Moehs, C.P., Zemetra, R.S. (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 USA,* 109, 20543-20548.

Whelan, A.I., and Lema, M.A. (2015). Regulatory framework for gene editing and other new breeding techniques (NBTs) in Argentina. *GM Crops Food*, 6, 253–265.

Wicker, T., Mayer, K.F., Gundlach, H., Martis, M., Steuernagel, B., Scholz, U., Šimková, H., Kubaláková, M., Choulet, F., Taudien, S., Platzer, M., Feuillet C., Fahima, T., Budak, H., Doležel, J., Keller, B., Stein, N. (2011). Frequent gene movement and pseudogene evolution is common to the large and complex genomes of wheat, barley, and their relatives. *The Plant Cell*, tpc-111.

Woo, J.W., Kim, J., Kwon, S.I., Corvalán, C., Cho, S.W., Kim, H.., Kim, S.G., Kim, S.T., Choe, S., Kim, J.S. (2015). DNA-free genome editing in plants with preassembled CRISPR-Cas9 ribonucleoproteins. *Nature Biotechnology*, 33(11), 1162.

Wu, C.H., Nakai, S., Powrie, W.D. (1976). Preparation and properties of acid-solubilized gluten. *Journal of Agricultural and Food Chemistry*, 24, 504-510.

Wu, H., Sparks, C., Amoah, B., Jones, H.D. (2003). Factors influencing successful Agrobacterium-mediated genetic transformation of wheat. *Plant Cell Reports*, 21, 659-668.

Wu, H., Doherty, A., Jones, H.D. (2008). Efficient and rapid Agrobacterium-mediated genetic transformation of durum wheat (Triticum turgidum L. var. durum) using additional virulence genes. *Transgenic Research*, 17, 425-436.

Wulff, B.B. and Moscou, M.J. (2014). Strategies for transferring resistance into wheat: from wide crosses to GM cassettes. *Frontiers in Plant Science*, 5, 692.

Xing, H.L., Dong, L., Wang, Z.P., Zhang, H.Y., Han, C.Y., Liu, B., Wang, X.C., Chen, Q.J. (2014). A CRISPR/Cas9 toolkit for multiplex genome editing in plants. *BMC Plant Biology*, 14(1), 327.

Zanini, B., Petroboni, B., Not, T., Di Toro, N., Villanacci, V., Lanzarotto, F., Pogna, N., Ricci, C., Lanzini, A. (2013). Search for a toxic cereals: a single blind, cross-over study on the safety of a single dose of *Triticum monococcum*, in patients with celiac disease. *BMC Gastroenterology*, 13, 92.

Zhang, X., Liu, D., Zhang, J., Jiang, W., Luo, G., Yang, W., Sun, J., Tong, Y., Cui, D., Zhang, A. (2013).

Novel insights into the composition, variation, organization, and expression of the low-molecular-weight glutenin subunit gene family in common wheat. *Journal of experimental botany,* 64(7), 2027-2040.

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

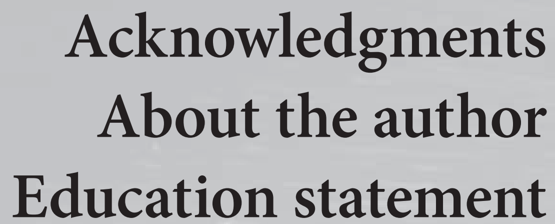
Zhou, H., Liu, B., Weeks, D.P., Spalding, M.H., Yang, B. (2014). Large chromosomal deletions and inheritable small genetic changes induced by CRISPR/Cas9 in rice. *Nucleic Acid Research*, 42, 10903-10914.

Zimin, A.V., Puiu, D., Hall, R., Kingan, S., Clavijo, B.J., Salzberg, S.L. (2017). The first near-complete assembly of the hexaploid bread wheat genome, *Triticum aestivum*. *GigaScience*. 6(11), 1-7.

Zipser, R.D., Patel, S., Yahya, K.Z., Baisch, D.W., Monarch, E. (2003). Presentations of adult celiac disease in a nationwide patient support group. *Digestive Diseases and Sciences*, 48(4), 761-764.

Zong, Y., Wang, Y., Li, C., Zhang, R., Chen, K., Ran, Y., Qiu, J.L., Wang, D., Gao, C. (2017). Precise base editing in rice, wheat and maize with a Cas9-cytidine deaminase fusion. *Nature biotechnology*, 35(5), 438.

Zuidmeer, L., Goldhahn, K., Rona, R.J., Gislason, D., Madsen, C., Summers, C., Sodergren, E., Dahlstrom, J., Lindner, T., Sigurdardottir, S.T., McBride, D. (2008). The prevalence of plant food allergies: a systematic review. *Journal of Allergy and Clinical Immunology,* 121, 1210-1218.


Acknowledgments
About the author
Education statement

Acknowledgments

Looking back, it is quite an impressive number of persons that I have come across, in the past 1500 days that contributed to make this project happen the way it did! From convincing me to apply for this PhD position to actively supporting and advising me, from sharing tips and ideas in the lab or at conferences to cheering me up over some drinks, so many people played a role, as minor as it was, that I would not possibly be able to name everyone within one or two pages. Therefore, please do not get offended if you do not find your name explicitly printed in here!

As an unusual way of starting PhD acknowledgments, I feel like expressing my gratitude to people without whom I would never have done any PhD! For that matter, I would like to thank deeply **Dr. Frans Krens**, who was my MSc. thesis and second-internship university supervisor, who succeeded in making me believe that I had a great scientific potential and should develop my academic career beyond a Master degree. He, together with **Dr. Jan Schaart**, were the ones who actively convinced me to apply for this PhD position. So, thank you both again, without your persistence, I would not be the candidate writing the acknowledgment related to this, or any, PhD thesis!

Before starting the acknowledgments regarding the actual PhD project, I would like to set up a tiny bit of context.

My PhD project was part of a Marie Curie Industrial Training Network (ITN) Program called "Maximising the potential of CROP researchers" (Max-CROP), financed by the European Commission through the 7th Framework Program (FP7-PEOPLE-2013_ITN-607178), which I really would like to thank. This program gave me great opportunities in terms of training, mobility and networking and I am very grateful for it.

Due to the design of the Max-CROP Program, I did half of my PhD at the Pre-breeding and Genetics department of the National Institute of Agricultural Botany (*NIAB*, as the industrial partner), Cambridge, UK and half at the Plant Breeding department of Wageningen University & Research (*WUR*, as the academic partner), Wageningen, The Netherlands. This explains the unusually large number of supervisors, officemates and colleagues I had over four years.

My first acknowledgments go to my five official supervisors.

At NIAB, I would like to thank **Dr. Fiona Leigh** especially for her support and kind words when I was feeling overwhelmed. I am grateful to **Dr. Emma Wallington**, for her expertise and for training me on various lab methods. I would like to extend my gratitude to the transformation team she leads; they really did a lot of work for generating all the transgenic wheat plants that constituted the basis of this PhD project. I am also grateful to **Dr. James Cockram** not only for bringing new angles for overcoming methodologic issues but also to have introduced me to some of his connections, notably at JIC and TSL, which have been very helpful for some aspect of the project.

At WUR, I would like to thank **Dr. Jan Schaart** for his great technical and lab support at the beginning of my PhD. I would like to express my profound gratitude to my main supervisor, **Dr. René Smulders**, not only for his guidance and all the knowledge he shared with me but also for his trust in my work and decisions, despite our frequent divergence of opinion over many aspects. René, I really would like to thank you for your support, for all the very interesting discussions we had and especially for dedicating some of your personal time for after-work Skype calls while I was oversees at NIAB. For this mater, I would like to extent my gratitude to your family to have accepted these additional working hours.

Additional researchers played a role in this project and would like to express them my gratitude as well.

At NIAB, I am grateful to the department of Seed Testing of **Ms. Helen Appleyard** and especially to Ms. **Pravina Solanki** and **Ms. Sofia Alves** who have turned me into a “Queen of Acid-PAGE” as well as for their great kindness during those intensive eight months in their lab. I also would like to acknowledge the bioinformaticians **Dr. Lawrence Percival-Alwyn** and more particularly **Dr. Bruno Santos** for his work and his patience to explain me the different analyses done.

At WUR, I also benefited from the expertise of bioinformaticians and I sincerely wish to acknowledge **Dr. Theo Borm**, without whom there would be no clear sequencing results presented in this thesis. I would like to thank as well **Dr. Elma Salentijn** for having introduced me to the complexity of dealing with sequence

alignments of gliadin gene families. Elma, now I get why you were happy moving on to projects without such large gene families! Thanks to the Bioscience department and particularly to **Dr. Hetty van den Broeck** and **Dr. Twan America** for performing the proteomics analysis and for their patience in explaining me all about it. My gratitude also goes to **Dr. Guido Hooiveld** of the department of Agrotechnology and Food Science for his guidance and valuable input regarding the ddPCR technique. Regarding this technique, I also would like to thank two **MSc. Students, Emiel Molenschot** who did pilot experiments and **Rubén Tenorio Berrio** who helped me developing a completely new assay and generated all the experimental data. Rubén, as the first MSc. student I really supervised, I hope I did a decent job but it was a great experience for me to work with a highly motivated and autonomous student, so really, thank you!

At the John Innes Centre (*JIC*, Norwich, UK), I would like to express my gratitude to **Dr. Brande Wulff**, who suggested us to use the gene enrichment and sequencing (RENSeq) approach. I am very grateful to him for his interest in my project and to his post-doc **Dr. Burkhardt Steuernagel** to have helped me into this direction and put me in contact with **Dr. Kamil Witek** of The Sainsbury Laboratory (*TSL*, Norwich, UK) who participated to the RENSeq protocol development. Consequently, I spend a month at TSL to be trained on this method by **Dr. Agathe Jouet** and **Dr. Oliver Furzer**, in **Prof. Jonathan Jones**' lab. I would like to deeply thank them all for having given me this opportunity and integrating me as if I was part of the group, I really appreciated it.

My gratitude also extends to the people who managed the project.

At NIAB, I thank **Dr. Lesley Boyd** who was the coordinator of the Max-CROP Program, for her general support and interest in my project.

At WUR, I would like to present my gratitude to my PhD promotor, **Prof. Richard Visser**, for his contribution and for being supportive throughout the project. I also would like to thank **Dr. Luud Gilissen** from Bioscience who played a greater role than only being my external supervisor, through his input, enthusiasm and his invitations to share my work with the scientific community on two different occasions, which were very productive experiences for me, I highly appreciated it.

My acknowledgments of course go to my fellow PhDs as well as post-docs and other colleagues.

Over the years and places, I had more than 30 different officemates and many more colleagues with whom I was very happy to have random discussions and jokes with, over lunches or coffees. Among them, I would specially like to thank **Xuan Xu, Laura Bouvet, Nicole Schatlowski and Clémence Marchal** for the endless laughs we had, the scientific discussions we shared and more especially your priceless support. You have been of an incredible help and I am deeply grateful to you!

We were five PhD candidates taking part in the Max-CROP Program, all facing the same issues for paperwork of any sort, commuting between NL and UK, having supervisors in at least two countries and attending common progress meetings and workshops. This created a special connection between us that I appreciated, despite some occasional divergence of opinions. I thank all four of you, **Mathilde Daniau, Tryntsje Boersma, Ioannis Baltzakis** and **Jasper Depotter** for the nice moments we shared together. I would like to extent my acknowledgments to **Mathilde** a little further, since we became close friends and were even housemates. Thank you for your support when all seemed to be going wrong with the PhD and thanks for the countless hilarious moments, I am glad we were in this adventure together.

Last but not least, I am deeply grateful to my friends and family.

I really would like to thank **my friends** from all around the world for their support, through small messages, over a drink, or hosting me for couple of days to cheer me up, it helped me a lot to keep motivation, so a big thanks to all of you guys! I especially would like to thank my best friend **Caroline** for pushing me never to give up and for having trust in my success all along. Heureusement que tu es là!

My “**Paranymphs**”, **Charlotte Prodhomme** and **Sri Sunarti**, also deserve to separately be thanked, and will deserve it even more after having helped me to handle my stress on the day of my defence!

The final expression of my gratitude goes obviously to **my family**, who had to hear me complaining a lot when my frustration level got high but who still did their very best to help me out, even when they did not really understand what was going on. Un grand merci à vous tous, et surtout à toi maman, pour m'avoir soutenue tout au long de ces quatre années, pour m'avoir écoutée me plaindre la plus-part du temps (je m'en excuse sincèrement) mais surtout pour avoir cru en moi et su trouver les mots justes pour me remotiver quand j'en avais besoin. Merci encore! Maintenant, c'est fini et j'ai ENFIN un « vrai » boulot!!!

About the author

From watering plants with a baby-bottle to becoming a plant scientist!

It was in the small town of Montluçon, in the volcanic region known as Auvergne, located in heart of France that Aurélie Jouanin was born, on the 9th of September 1988. As a kid, she developed interests in plants and especially ornamental ones. Later on, she got captivated by the genetics courses she received during her High School education. Therefore, molecular plant breeding and plant biotechnology became the direction she followed.

Aurélie started the first part of her Bachelor degree in 2006, at the IUT of Aurillac, France, on the broad topic of Agronomic Biology. To complete this degree, she did her first internship at the Plant Breeding department of Wageningen UR, where she learned about plant tissue culture and genetic transformation under the direction of Isolde Pereira-Bertram and Dr. Herma Koehorst-van Putten, working on Cassava root-specific promoter identification using reporter genes. In 2008, she started the second part of her Bachelor as a double degree between CAH in Dronten, The Netherlands and ESA in Angers, France, on Plant Breeding. This program gave her the chance to perform two internships. Therefore, she decided to get insight in conventional flower breeding through an internship at Orchideen-Holm, a familial Orchid Breeding company located in BedBurg-Hau, Germany, where she was supervised by Marko Holm and his father, founder of the company. Then, for her final Bachelor thesis, she decided to explore applied plant research from an industrial perspective, working on gynogenesis and Double Haploid in onion, under the supervision of Marcel Adriaanse and Dr. Witte van Cappellen, at Bejo Zaden, a Dutch vegetable seed company based in Warmenhuizen, The Netherlands.

As a logical continuation, Aurélie signed up for the Master of Plant Biotechnology, specialisation Molecular Plant Breeding and Pathology, at Wageningen University in 2010. She went back to the Plant Breeding group to do her Master thesis, molecular biology oriented this time, about the differential transcription of

anthocyanin-related genes in roses of different colours, under the supervision of Virginia Gitonga and Dr. Frans Krens. She, then, did a first Master internship under the supervision of Dr. Franck Lhuissier and Dr. Michiel de Both on improving the efficiency of KeyBase®, the ODN-based gene-editing technology developed at Keygene, a Dutch plant biotech company based in Wageningen, The Netherlands. Being given the opportunity of doing a second Master internship with the only ornamental breeding company working on genetic transformation in Europe at the time, she went working on improving petunia transformation rate as well as testing transgenic petunias for drought stress tolerance, under the supervision of Vanessa Lunke and Dr. Robert Boehm at Selecta Klemm, Stuttgart, Germany.

After having been convinced by several researchers and former supervisors to do a PhD and advised to apply for the ITN-Marie Curie PhD project that was focusing on applying mutagenesis techniques on bread wheat to decrease gluten immunogenicity for Coeliac Disease patients, Aurélie successfully got the position in 2014. She became, thus, part of the "Max-imising the potential of CROP researchers" (Max-CROP) Program funded by the European Commission via the 7th Framework Program. This International Training Network - PhD project was conducted between the Pre-breeding & Genetics department of the National Institute of Agricultural Botany, NIAB, Cambridge, UK – as industrial partner - and at the Plant Breeding department of Wageningen University – as academic partner -. The PhD supervision at NIAB was undertaken by Dr. Fiona Leigh, Dr. Emma Wallington and Dr. James Cockram while it was handled at WUR by Dr. Jan Schaart and Dr. René Smulders, Prof. Richard Visser being the PhD Promoter. Additionally, Dr. Lesley Boyd (NIAB) was the coordinator of the Max-CROP program and Dr. Luud Gilissen (Bioscience, WUR) was the PhD external supervisor. The present thesis book reports all the work done during these four years of PhD.

Since November 2018, Aurélie works as Research Scientist Plant Genome Editing, in the Research and Development department of KWS, a German crop breeding company, at its headquarters located in Einbeck, Germany.

Education Statement of the Graduate School Experimental Plant Sciences

Issued to:	**Aurélie Jouanin**
Date:	**28 January 2019**
Group:	**Laboratory of Plant Breeding**
University:	**Wageningen University & Research**

1) Start-Up Phase	*date*	*cp*
► **First presentation of your project**		
"Deployment of mutagenesis technics targeting gliadin gene families to eliminate Celiac Disease epitopes from bread wheat" *(Plant Breeding department, WUR, Wageningen, NL)*	19 Jan 2015	1,5
► **Writing or rewriting a project proposal**		
"Deployment of mutagenesis technics targeting gliadin gene families to eliminate Celiac Disease epitopes from bread wheat"	28 Jan 2015	6,0
► **Writing a review or book chapter**		
► **MSc courses**		
Programming in Python (INF-22306) *(WUR, Wageningen, NL)*	Sep-Oct 2014	3,0
Advanced Bioinformatics (BIF-30806) *(WUR, Wageningen, NL)*	Nov-Dec 2014	3,0
► **Laboratory use of isotopes**		
Subtotal Start-Up Phase		13,5

2) Scientific Exposure	*date*	*cp*
► **EPS PhD student days**		
EPS PhD student days: 'Get2gether' 2015 *(Soest, NL)*	29-30 Jan 2015	0,6
EPS PhD student days: 'Get2gether' 2016 *(Soest, NL)*	28-29 Jan 2016	0,6
► **EPS theme symposia**		
EPS Thema 4 symposium: "Genome Biology" 2014 *(WUR, Wageningen, NL)*	03 Dec 2014	0,3
EPS Thema 2 symposium: "Interactions between Plants and Biotic Agents" 2018 *(UvA, Amsterdam, NL)*	24 Jan 2018	0,3
► **National meetings (e.g. Lunteren days) and other National Platforms**		
► **Seminars (series), workshops and symposia**		
Symposium: All Inclusive Breeding, Integrating high throught put science *(WUR, Wageningen, NL)*	16 Oct 2014	0,3
Symposium: Omics Advances for Academia and Industry - Towards True Molecular Plant Breeding *(WUR, Wageningen, NL)*	11 Dec 2014	0,3
Workshop: Marie Curie ITN Max-CROP report 1 + Crop production & Genetics *(NIAB, Cambridge, UK)*	10-12 Feb 2015	0,8
Workshop: Marie Curie ITN Max-CROP report 2 + Crop pathology & Plant-pathogen interactions *(WUR, Wageningen, NL)*	06-08 May 2015	0,6
Workshop: REnSeq *(JIC, Norwich, UK)*	11-12 Jun 2015	0,3
Workshop: New Plant Breeding Technologies and other advances in plant genetic research for Africa *(NIAB, Cambridge, UK)*	09-10 Jul 2015	0,6
Symposium: BioRad ddPCR *(Cambridge, UK)*	25 Sep 2015	0,3

Workshop: GMOs Risk Assessment & Communication of Evidence *(Vienna, AT)*	05-08 Oct 2015	0,9
Workshop: Towards a sustainable future for wheat genomics: Plant genomics databases training *(TGAC, Norwich, UK)*	20-21 Oct 2015	0,6
Workshop: Health Grain Forum *(Wageningen, NL)*	16-17 Nov 2015	0,6
Workshop: Marie Curie ITN Max-CROP report 3, Mid-term assesment *(NIAB, Cambridge, UK)*	24 Nov 2015	0,3
Symposium: Monogram 2016 + Bioinformatics workshop *(Cambridge, UK)*	12-14 Apr 2016	0,7
Workshop: Whealbi, SNP discovery & analysis using next generation technologies training *(JHI, Dundee, UK)*	19-20 May 2016	0,6
Workshop: Marie Curie ITN Max-CROP report 4 *(NIAB, Cambridge, UK)*	21 Nov 2016	0,3
Workshop: Community Resource for Wheat transformation *(NIAB, Cambridge, UK)*	18 May 2017	0,3
Workshop: Your Plant science *(NIAB, Cambridge, UK)*	20-21 Nov 2017	0,6
Mini-Symposium EPS: Polyploidy, genetics & breeding *(WUR, Wageningen, NL)*	14 Jun 2018	0,2
► **Seminar plus**		
► **International symposia and congresses**		
Final conference of GMOs Risk Assessment & Communication of Evidence *(Potsdam, DE)*	09-10 Nov 2015	0,6
5th Plant Genomics & Gene Editing Congress_Europe *(Amsterdam, NL)*	16-17 Mar 2017	0,6
The Rank Prize Funds mini-symposia on The Role of Wheat in Diet, Health and Disease *(Grasmere, UK)*	24-27 Apr 2017	0,9
International Conference "Plant Transformation & Biotechnology IV" *(Vienna, AT)*	29-30 Jun 2017	0,6
International Conference "Plant Genome Editing & Genome Engineering" *(Vienna, AT)*	03-04 Jul 2017	0,6
Nestlé Conference: Planting seeds for the future of food II *(Nestlé HQ, Vevey, CH)*	06-07 Jul 2017	0,6
International Celiac Disease Symposium 2017 *(New Delhi, IN)*	08-10 Sep 2017	0,9
4th Conference of Cereal Biotechnology and Breeding, EUCARPIA Cereal *(Budapest, HU)*	06-09 Nov 2017	0,9
13th International Gluten Workshop, CIMMYT *(Mexico city, MX)*	14-17 Mar 2018	1,2
► **Presentations**		
Poster: Science week *(Cambridge, UK)*	14 Mar 2015	1,0
Talk: Health Grain Forum Workshop *(Wageningen, NL)*	16-17 Nov 2015	1,0
Poster + Flash talk: Monogram 2016 *(Cambridge, UK)*	12-14 Apr 2016	1,0
Poster: 5th Plant Genomics & Gene Editing Congres_ Europe *(Amsterdam, NL)*	16-17 Mar 2017	1,0
Talk: The Rank Prize Funds mini-symposia on The Role of Wheat in Diet, Health and Disease *(Grasmere, UK)*	24-27 Apr 2017	1,0
Talk: International Conference "Plant Genome Editing & Genome Engineering" *(Vienna, AT)*	03-04 Jul 2017	1,0
Poster: International Celiac Disease Symposium 2017 *(New Dehli, IN)*	08-10 Sep 2017	1,0
Talk: 4th Conference of Cereal Biotechnology and Breeding, EUCARPIA Cereal *(Budapest, HU)*	06-09 Nov 2017	1,0
Talk: Your Plant science *(NIAB, Cambridge, UK)*	20-21 Nov 2017	1,0
Talk: 13th International Gluten Workshop, CIMMYT *(Mexico city, MX)*	14-17 Mar 2018	1,0

		date	cp
►	**IAB interview**		
►	**Excursions**		
	Cereals 2015 *(Lincolnshire, UK)*	11 Jun 2015	0,3
	Syngenta visit, Crop Protection Technology Process *(Jealott's Hill, UK)*	25 Nov 2015	0,3
	Subtotal Scientific Exposure		27,6

3) In-Depth Studies		date	cp
►	**EPS courses or other PhD courses**		
	Plant CRISPR-Cas Workshop-training *(JIC/TSL, Norwich, UK)*	07-08 Sep 2015	0,4
	International CEPLAS summer School "Emerging Frontiers in Plants Sciences" *(Hennef/Cologne, DE)*	05-09 Jun 2017	1,5
►	**Journal club**		
►	**Individual research training**		
	Exome capture & Enrichment for PacBio sequencing *(TSL, Norwich, UK)*	Jun 2016	3,0
	Subtotal In-Depth Studies		4,9

4) Personal Development		date	cp
►	**Skill training courses**		
	EPS PhD introduction course *(WUR, Wageningen, NL)*	20 Jan 2015	0,2
	Intellectual Property: IP for Scientists and Engineers *(Cambridge, UK)*	12 Apr 2016	0,2
	Marie Sklodowska-Curie Action, ESOF Satellite Conference "Researcher and Society" *(Manchester, UK)*	28-29 Jul 2016	0,6
	Marie Sklodowska-Curie Action, Conference & General Assembly 2017 *(Salamanca, ES)*	24-25 Mar 2017	0,6
	International Summer School: Ethical, legal & societal aspects of genome editing in agriculture *(TTN, Munich/Schoenwag, DE)*	02-06 Oct 2017	1,5
	Entrepreneurship and how it works: workshop *(Judge Business School, Cambridge, UK)*	14-16 Jun 2017	0,9
	Scientific writing *(WUR, Wageningen, NL)*	Apr-Jun 2018	1,8
	Negotiating workshop *(WUR, Wageningen, NL)*	28 Jun 2018	0,1
►	**Organisation of PhD students day, course or conference**		
	Science week, outreach event *(Cambridge, UK)*	14 Mar 2015	0,3
	NIAB open days *(NIAB, Cambridge, UK)*	23-25 Jun 2015	0,3
►	**Membership of Board, Committee or PhD council**		
	Subtotal Personal Development		6,5

TOTAL NUMBER OF CREDIT POINTS*	**52,5**

Herewith the Graduate School declares that the PhD candidate has complied with the educational requirements set by the Educational Committee of EPS which comprises of a minimum total of 30 ECTS credits.

** A credit represents a normative study load of 28 hours of study.*

This PhD was part of a program called "Maximising the potential of CROP researchers" (Max-CROP) which was a Marie Curie Industrial Training Network (ITN) financed by the European Commission through the 7th Framework Program (FP7-PEOPLE-2013_ITN-607178). As part of this Program, the work was carried out at Plant Breeding, Wageningen University & Research, Wageningen, The Netherlands, as well as in the Department of Genetics & Pre-Breeding, National Institute of Agricultural Botany, Cambridge, The United Kingdom.

Cover, layout design: Ferdinand van Nispen tot Pannerden,
Citroenvlinder DTP&Vormgeving, *my*-thesis.nl

Printing: Proefschriftmaken, Vianen, The Netherlands || Proefschriftmaken.nl