



## SNP markers for early identification of high molecular weight glutenin subunits (HMW-GSs) in bread wheat

Catherine Ravel, Annie Faye, Sarah Ben-Sadoun, Marion Ranoux, Mireille Dardevet, Cécile Dupuits, Florence Exbrayat-Vinson Exbrayat, Charles C. Poncet, Pierre Sourdille, Gerard G. Branlard

### ► To cite this version:

Catherine Ravel, Annie Faye, Sarah Ben-Sadoun, Marion Ranoux, Mireille Dardevet, et al.. SNP markers for early identification of high molecular weight glutenin subunits (HMW-GSs) in bread wheat. TAG Theoretical and Applied Genetics, 2020, 133 (3), pp.751-770. 10.1007/s00122-019-03505-y . hal-02528930

**HAL Id: hal-02528930**

**<https://hal.inrae.fr/hal-02528930>**

Submitted on 19 Oct 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# **SNP markers for early identification of high-molecular-weight glutenin subunits (HMW-GSs) in bread wheat**

Catherine Ravel<sup>1</sup> · Annie Faye<sup>1</sup> · Sarah Ben-Sadoun<sup>1</sup> · Marion Ranoux<sup>1</sup> · Mireille Dardevet<sup>1</sup> ·  
Cécile Dupuits<sup>1</sup> · Florence Exbrayat<sup>1</sup> · Charles Poncet<sup>1</sup> · Pierre Sourdille<sup>1</sup> · Gérard. Branlard<sup>1</sup>

<sup>1</sup>INRA, UMR1095, Genetics Diversity and Ecophysiology of Cereals, F-63100, Clermont-Ferrand, France

Clermont Auvergne University, UMR1095, Genetics Diversity and Ecophysiology of Cereals, F-63100, Aubière, France

✉ [catherine.grand-ravel@inra.fr](mailto:catherine.grand-ravel@inra.fr) (ORCID 0000-0003-2528-4532)

## Abstract

**Key message** A set of eight SNP markers was developed to facilitate the early selection of HMW-GS alleles in breeding programmes.

In bread wheat (*Triticum aestivum*), the high-molecular-weight glutenin subunits (HMW-GS) are the most important determinants of technological quality. Known to be very diverse, HMW-GS are encoded by the tightly linked genes *Glu-1-1* and *Glu-1-2*. Alleles that improve the quality of dough have been identified. Up to now, sodium-dodecylsulfate–polyacrylamide gel electrophoresis (SDS-PAGE) of grain proteins is the most widely used for their identification. To facilitate the early selection of HMW-GS alleles in breeding programmes, we developed DNA-based molecular markers. For each accession of a core collection (n = 364 lines) representative of worldwide bread wheat diversity, HMW-GSs were characterized by both genotyping and SDS-PAGE. Based on electrophoresis, we observed at least 8, 22 and 9 different alleles at the *Glu-A1*, *-B1*, and *-D1* loci, respectively including new variants. We designed a set of 17 single nucleotide polymorphism (SNP) markers that were representative of the most frequent SDS-PAGE alleles at each locus. At *Glu-A1* and *Glu-D1*, two and three marker-based haplotypes respectively captured the diversity of the SDS-PAGE alleles rather well. Discrepancies were found mainly for the *Glu-B1* locus. However, statistical tests revealed that two markers at each *Glu-B1* gene and their corresponding haplotypes were more significantly associated with the rheological properties of the dough than were the relevant SDS-PAGE alleles. To conclude, this study demonstrates that the SNP markers developed provide additional information on HMW-GS diversity. Two markers at *Glu-A1*, four at *Glu-B1* and two at *Glu-D1* constitute a useful tool-box for breeding wheat to improve end-use value.

**Key words** Bread wheat (*Triticum aestivum* L.) . End-use value . High-molecular-weight glutenin subunit . Sodium-dodecylsulfate–polyacrylamide gel electrophoresis . Single nucleotide polymorphism marker

### **Abbreviations**

Bx7OE        overexpressed Bx7 allele

HMW-GS     High-molecular-weight glutenin subunit

KASP        Kompetitive allele-specific PCR

LMW-GS     Low-molecular-weight glutenin subunit

SDS-PAGE   Sodium-dodecylsulfate–polyacrylamide gel electrophoresis

SNP        Single nucleotide polymorphism

SSP        Seed storage proteins

## Introduction

Wheat is one of the three most important crops in the world with production of about 729 million tonnes in 2014 (<http://faostat3.fao.org>). In this context, wheat includes tetraploid species ( $2n = 28$ ) such as durum wheat (*Triticum turgidum* spp. *durum*) and hexaploid species ( $2n = 42$ ) such as bread wheat (*T. aestivum* spp. *aestivum*). Wheat is thus a major component of the human diet worldwide, often being the main source of energy. Carbohydrates from bread alone contribute about 20% of energy intake (Shewry and Hey 2015). Wheat is also an important plant source of protein providing on average 20% of the total protein in the human diet. Almost all the world's wheat production is used after industrial processing. Each type of wheat end-product requires particular qualities for processing that are mainly based on the properties of dough, determined by unique combinations of cohesiveness and viscoelasticity due to gluten. As described in Shewry et al. (2002), gluten is a continuous network formed when wheat seed storage proteins (SSP) are mixed with water. Therefore, SSP concentration and composition are largely responsible for wheat end-use quality. While gluten is necessary for processing wheat-based products, it also triggers gluten-related disorders in humans, like allergies, coeliac disease and non-coeliac gluten sensitivity (Sapone et al. 2012).

The wheat SSP represent about 80% of the total protein in the grain. They mainly consist of polymeric glutenins and monomeric gliadins. According to their electrophoretic mobility, gliadins and glutenins are subdivided into several fractions. Glutenins are classified as high-molecular-weight or low-molecular-weight glutenin subunits (HMW-GS and LMW-GS, respectively). Glutenins, in particular HMW-GS, confer dough elasticity while gliadins confer viscosity (MacRitchie 1999; Shewry et al. 2002). Dough quality results from the balance between these two properties, or in other words, the gliadin-to-glutenin ratio. A pioneer study (Payne et al. 1979) reported that HMW-GS influence dough strength and this has been

confirmed by numerous studies, reviewed by Shewry (2009). The effects of HMW-GS on the rheological properties of dough are mainly explained by their  $\beta$ -spiral structure, which has intrinsic elasticity (Shewry et al. 2001), and their ability to form large polymers stabilized by interchain disulfite and hydrogen bonds (see the review of Shewry et al. 2002). HMW-GSs are therefore essential for dough visco-elasticity as they can form an elastic network that acts as the backbone of gluten, making them an attractive target for genetic engineering.

LMW-GS are encoded by multigene families located at the orthologous *Glu-3* loci. HMW-GS are encoded by the *Glu-1* loci, named *Glu-A1*, *Glu-B1* and *Glu-D1*, located on the long arms of the homoeologous chromosomes of group 1 (Payne 1987). Each locus comprises the two tightly linked genes *Glu-1-1* and *Glu-1-2* that encode x-type and y-type HMW-GS, respectively. In wheat cultivars, three to five of the six genes of this small multigene family are generally expressed, as some alleles are known to be null. For example, *Glu-A1-2* is rarely expressed, and silent alleles at *Glu-A1-1* and *Glu-B1-2* have been reported. In addition, duplication of *Glu-B1-1* may arise as in the overexpressed Bx7 allele (Bx7OE) leading to over-expression (Ragupathy et al. 2008).

The glutenin coding sequence is composed of a central repetitive domain and two unique-sequence termini. The size of the repetitive domain is highly variable, which explains why the proteins can be easily distinguished by differences in electrophoretic mobility in sodium-dodecyl-sulphate polyacrylamide gel electrophoresis (SDS-PAGE). According to Shewry et al. (1992), size separation of HMW-GS makes it possible to identify 3 (called a, b and c), 11 (a to k) and 6 (a to f) allelic forms at *Glu-A1*, *Glu-B1* and *Glu-D1* loci, respectively. Less frequent alleles have been reported, for instance, three x-type HMW-GS at the *Glu-A1* locus (Margiotta et al. 1996; Gobaa et al. 2007; Ribeiro et al. 2013). The AAC glutenin allele database (<http://www.aaccnet.org/initiatives/definitions/Pages/Gluten.aspx>) contains records of the alleles of over 8,500 wheat genotypes from around the world, giving an overview of the

wide diversity of HMW-GS. Each HMW-GS has different effects on end-use quality. Thus, to breed for end-use value, several studies have ranked HMW-GS alleles in order of their influence on flour quality (Branlard and Dardevet 1985; Payne et al. 1987), and highlighted the *Glu-D1d* (Dx5 + Dy10), *Glu-B1b* (Bx7 + By8) and *Glu-B1c* (Bx7 + By9) alleles as being strongly associated with high quality (Pirozi et al. 2008).

For plant breeding purposes, SDS-PAGE separation is still used to identify HMW-GS as well as LMW-GS and select those associated with high quality. The SDS-PAGE method requires flour in order to extract the grain protein, which means that plants need to be grown to a late development stage before testing them. The method is non-destructive in that only half a grain is needed for testing. The interpretation of gel images or electrophoregrams is not completely accurate, even when performed by experienced staff. Some HMW-GS may be confounded due to their quasi-identical electrophoretic mobility. Classifying unambiguously LMW-GS by SDS-PAGE method is also difficult because of their large numbers and overlapping mobility with gliadins in the gel. This method is time consuming and not suitable for high-throughput analysis. Typing glutenins with molecular markers would solve many of the drawbacks of the SDS-PAGE method. Typing with molecular markers requires low amounts of DNA that can be easily extracted from leaves of seedlings. In addition, high-throughput genotyping methods based on single nucleotide polymorphisms (SNPs) are now available, even for polyploid wheats (Bérard et al. 2009; Akhunov et al. 2009; Wang et al. 2014; Rimbart et al. 2017).

Since the 1990s, many attempts have been made to develop molecular markers from glutenins sequences, generally based on presence/absence or amplicon-size polymorphisms, as reviewed by Gale (2005) and later by Liu et al. (2012). Despite the difficulty to capture the complexity of these gene families or to interpret the results obtained, especially for LMW-GS, some of these markers were successfully used. For instance, Jin et al. (2011), Liu et al. (2010)

and more recently by Iba et al. (2018) reported that the identification of HMW-GS or LMW-GS genes by DNA-based methods agrees with the results from SDS-PAGE. However, despite continued work (see for instance Liu et al. 2008; Xu et al. 2008), other than the G-A change used to discriminate between the Dx2 and Dx5 alleles (Schwarz et al. 2003), few SNPs are available to distinguish between different HMW-GS. This may be due to the difficulty of developing gene-specific markers for a multigenic family. The repetitive domain of HMW-GS genes also limits the choice of site for marker development.

The objective of our study was to develop and validate a set of SNPs to identify the main HMW-GSs. We decided to type SNPs by a flexible method based on competitive allele-specific PCR (KASP-assay). Although the design of allele-specific primers is constrained by the polymorphism location, the common allele can be designed at a gene-specific location to facilitate base-calling in the wheat hexaploid genome. The SNP markers designed were genotyped in a wheat core collection (Balfourier et al. 2007). The clustering observed differed slightly from that based on the SDS-PAGE profiles. The relationship between these markers and end-use quality was established by measuring alveographic parameters. Therefore, the set of markers from HMW-GS sequences developed in this study is of significant value to wheat breeders, who will be able to characterize HMW-GS routinely and unambiguously from DNA.

## **Materials and methods**

### **Plant materials and phenotyping**

The INRA core collection includes 366 accessions from 70 different countries, chosen to represent the genetic diversity present in cultivated wheat (Balfourier et al. 2007). These accessions are either landraces from the 19th century or cultivars from the 20th century. In this



work, 364 lines of the core collection were used. All the seeds used for DNA extraction, provided by the Biological Resources Centre on Small Grain Cereals ([http://www6.clermont.inra.fr/umr1095\\_eng/Teams/Experimental-Infrastructure/Biological-Resources-Centre](http://www6.clermont.inra.fr/umr1095_eng/Teams/Experimental-Infrastructure/Biological-Resources-Centre)), were obtained from self-pollinated ears. Fresh leaves of five plants per accession were pooled and bulk genomic DNA was extracted from 100 mg of frozen leaves with the Biosprint 96 kit using a Biosprint workstation (Qiagen).

All accessions of the core collection were grown at Clermont-Ferrand in 2006 as described in Bordes et al. (2008). Seeds were harvested in bulk and phenotyped for quality traits (Bordes et al. 2008). In the present work, we used average single grain dry mass ( $\text{mg DM grain}^{-1}$ ), the total quantity of protein per grain ( $\text{mg N grain}^{-1}$ ) calculated from the average single grain dry mass and wholemeal flour protein concentration ( $\text{mg protein g DM}^{-1}$ ), grain hardness (dimensionless), and alveographic parameters, i.e. dough strength ( $W, 10^{-4} \text{ J}$ ), tenacity ( $P, \text{ mm H}_2\text{O}$ ), and extensibility ( $L, \text{ mm}$ ). Grain hardness and wholemeal flour protein concentration were determined by near-infrared spectroscopy. These data were provided by Bordes et al. (2008).

For each accession, proteins were extracted from 10 mg of wholemeal flour ground in a Cyclotec mill (sieve 0.75 mm, mill 6800, FOSS Electric A/S, Hilleroed, Denmark). HMW-GS were fractionated in vertical slabs using SDS-PAGE with a modified protocol based on that of Singh et al. (1991). They were identified according to the numbering system developed by Payne and Lawrence (1983) with the names available at <https://shigen.nig.ac.jp/wheat/komugi/>.

## **SNP discovery and genotyping**

Sequences from the main alleles of the six HMW-GS genes were retrieved from GenBank (<http://www.ncbi.nlm.nih.gov/genbank/>). Many sequences are available for these genes, so a first set was collated to include at least one sequence per gene encompassing, where possible, the entire coding DNA sequence (CDS), the 5' untranslated regions (UTR), the promoter region (about 1-kb region upstream of the start codon) and the 3' UTR (Table 1, Fig. S1). Then for each HMW-GS gene class, except the silent *Glu-A1-2*, another set of sequences was collated either from GenBank or from data obtained by Ravel et al. (2006; 2014) (Table 1). Each of the latter five files was aligned with Clustal X (Larkin et al. 2007) and alignments were improved manually. Allelic SNPs were detected in each of these alignments. The most relevant SNPs detected to discriminate HMW-GS were retained for genotyping. SNPs within the repetitive domain were discarded as designing specific primers in this region was deemed impracticable.

KASPar assays (LGC Genomics, LLC, Beverly, MA, USA) were developed to genotype relevant SNPs. This method is based on competitive annealing of two labelled allele-specific primers (Gao et al. 2016). The 3'-end of each allele-specific primer was designed to target the allelic SNPs detected in the alignments for each HMW-GS gene class. A third common primer was designed in the vicinity of the other two to be copy-specific (for A, B, or D chromosomes). The 3'-end of the common primer targeted polymorphisms between the six HMW-GS genes deduced from the first alignment made with one sequence per HMW-GS gene. All the primers were designed with Primer3 (Untergasser et al. 2012).

Primer mix was prepared as recommended by LGC Genomics and described in Gao et al. (2016). The total reaction volume was 5  $\mu$ L in a 384-well plate, composed following LGC Genomics recommendations. The concentration of MgCl<sub>2</sub> was adjusted to 2.2 mM when the mean GC percent of the common primer sequence was between 33% and 55% (Table 2). The amplification reaction was performed in a Veriti® thermal cycler (Applied Biosystems)

programmed as recommended by LGC Genomics with a modification in the 10 touchdown cycles, so annealing temperature was decreased by 0.8°C per cycle from 65°C to 57°C for 1 min, followed by 40 cycles of 94°C for 20 s and 57°C for 1 min. The LightCycler® 480 System (Roche) was used as a plate reader to detect the fluorescence signal.

### **Statistical analysis and haplotype representation**

The distribution observed for the HMW-GS identified by SDS-PAGE was compared in a  $\chi^2$  test with that reported in Shewry et al. (1992), which was derived from the 300 accessions characterized by Payne and Lawrence (1983).

The relevance of the main factors, i.e. marker, haplotype (i.e. the combination of alleles at each marker for a given locus) and SDS-PAGE profiles, was tested independently through ANOVA analyses conducted with the linear model (lm) function of R v3.3.0 (R Core Team, 2016). In these analyses, heterozygous markers were treated as missing data. In each analysis, the model included covariates to estimate the part played by the different markers (or by the haplotypes) in the explanation of the phenotypic traits under study. These covariates were the structure components of the core collection (Bordes et al. 2008), the grain hardness and the protein concentration. The structure components of the core collection were considered to avoid spurious association (Flint-Garcia et al. 2003). The grain hardness and the protein concentration were taken into account because these traits are known to modify dough viscoelasticity (Branlard et al. 2001; Eagles et al. 2006). Rare haplotypes (frequency < 2.5%) were discarded from the analysis. F tests were considered significant at  $\alpha = 0.01$ . For haplotypes and SDS-PAGE profiles, means were compared by the Student-Newman-Keuls test function from the R

library *Agricolae*. Means were judged to be significantly different when the *P*-values of the Student-Newman-Keuls test were < 0.05.

For loci with a high level of polymorphism, the SNP combination (haplotype) was represented using sequence logos, plotted with seqLogo available in Bioconductor (Bembom, 2016), with the overall height of the stacks proportional to the information content at the position considered and the height of letters within the stack indicating the relative frequency of each nucleotide.

## Results

In this work, we developed a series of PCR-based assays to readily identify wheat HMW-GS based on competitive annealing of primers to SNPs using the KASP approach (Semagn et al. 2014). The HMW-GS of 364 accessions were identified by SDS-PAGE profiling and by molecular markers. The results obtained with the two methods were compared to each other. At *Glu-A1* and *Glu-D1*, the molecular marker captured the diversity of the SDS-PAGE alleles rather well. Most discrepancies were found at *Glu-B1*. However, two markers at each *Glu-B1* gene and their corresponding haplotypes were more significantly associated with the rheological properties of the dough than were the relevant SDS-PAGE alleles.

### Allelic diversity of HMW-GS assessed by SDS-PAGE

For 364 accessions of the INRA core collection (n=366), HMW-GS were characterized by SDS-PAGE. The results obtained were statistically compared to previously published data (Shewry et al. 1992) and significant differences were pointed out between both collections. In the core collection, 4, 18 and 7 known HMW-GS alleles were detected at *Glu-A1*, *Glu-B1* and *Glu-D1*,

respectively (Table 3, Table S1). In addition, 10, 19 and 11 lines had previously undefined SDS-PAGE profiles at *Glu-A1*, *Glu-B1* and *Glu-D1* loci respectively (Table S1) that may be indicative of novel alleles. Up to 7.7% of the accessions are heterozygous at least at one of the three loci.

At *Glu-A1*, the three main alleles a, b and c were detected in 340 lines. In a  $\chi^2$  test, no significant difference was observed ( $\chi^2 = 0.15$ ,  $P = 0.92$ ) between the allele frequencies in the INRA core collection and those reported by Shewry et al. (1992). As expected, the Glu-A1c (A<sub>x</sub>null) allele found in 41% of the core collection is the major allele. The rare allele *Glu-A1t* characterizes a single line, accession number 2301 from Australia (Table S1). We observed 10 lines with undefined patterns that could correspond to four novel alleles (Table S1).

At *Glu-B1*, 249 (74 %) out of the 336 accessions which are not heterozygous have one of the four main alleles, i.e. Glu-B1a (B<sub>x</sub>7 + B<sub>y</sub>Null), Glu-B1b (B<sub>x</sub>7 + B<sub>y</sub>8), Glu-B1c (B<sub>x</sub>7 + B<sub>y</sub>9), or Glu-B1d (B<sub>x</sub>6 + B<sub>y</sub>8). The major allele in the core collection is Glu-B1b with a frequency of about 32%, while Shewry et al. (1992) previously described Glu-B1c as the major allele at a frequency of 30%. This difference significantly skewed the  $\chi^2$  test ( $\chi^2 = 53.53$ ,  $P < 0.001$ ) when our results for these four alleles were compared with those from Shewry et al. (1992). The rare alleles Glu-B1e, f, g, h, i, k, p, aj, al and am were also present in the INRA core collection (Table 3). Undefined patterns were observed for 18 lines indicative of at least four novel alleles (Table S1).

At *Glu-D1*, 323 lines (89%) were assigned to one of Glu-D1a (D<sub>x</sub>2 + D<sub>y</sub>12), Glu-D1b (D<sub>x</sub>3 + D<sub>y</sub>12), Glu-D1c (D<sub>x</sub>4 + D<sub>y</sub>12) or Glu-D1d (D<sub>x</sub>5 + D<sub>y</sub>10) alleles, while seven lines had less frequent alleles (Table 3). The remaining lines were either heterozygous (23 lines) or showed novel profiles (11 lines assigned to at least two novel alleles (Table S1)). As previously found (Shewry et al. 1992), Glu-D1a was the most frequent allele present in 184 lines of the

INRA core collection. However, when the four main alleles (Glu-D1a, b, c and d) were taken into account, the frequencies observed in the core collection differed significantly ( $\chi^2 = 39.47$ ,  $P < 0.001$ ) from those reported by Shewry et al. (1992). The comparison of both collections showed an excess of lines with the Glu-D1b allele at the expense of the Glu-D1c in the INRA core collection. Rare alleles were also detected like Glu-D1f (Dx2.2 + Dy12). Lontoi from Finland (accession number 4343) is null at *Glu-D1-1*.

### **Development and validation of SNP markers to establish HMW-GS composition**

To establish HMW-GS composition from genomic DNA, we developed SNP markers for genotyping by a strategy based on competitive allele-specific PCR (KASP). These markers were developed based on the polymorphisms found after aligning the nucleic sequences of HMW-GS genes. The markers all target non-repetitive regions of the genes, i.e. promoters, UTRs, or non-repetitive regions of the CDS (Table 2, Figure S1).

Twenty SNPs and 2 indels were found in the alignment of sequences at *Glu-A1-1*, which was restricted to the CDS (Figure S2). The first two SNP assays developed target G/A differences in the non-repetitive 5' region of the CDS (Table 2). The A allele at the first marker identified Glu-A1a (Ax1) while the G allele at the second locus identified Glu-A1c (Axnull). Three haplotypes are thus expected: AA-AA, GG-AA and GG-GG for the HMW-GS Ax1, Ax2\* (Glu-A1b), and Axnull, respectively.

To identify SNPs in *Glu-D1-1*, sequences up to about 1,250 nucleotides upstream of the start codon were aligned (Figure S3). This region contains seven SNPs. We designed two primer pairs to discriminate the alleles encoding the main HMW-GS (Dx2, Dx3 and Dx5) (Table 2). The T polymorphism (or A in the alignment as allele-specific primers were designed on the

reverse strand) at the marker Mk-D1-1-4 is specific for the Dx5 allele and the G nucleotide at Mk-D1-1-8 characterizes the Dx3 allele. We observed 25 SNPs and two insertion-deletions (indels) between both *Glu-D1-2* sequences comprising the promoter region and the CDS (Figure S4). Allele-specific primers of the marker developed for this SNP, Mk-D1-2-1, were designed on the reverse strand. The A and G polymorphisms (T and C in the alignment, Figure S4) characterize Dy10 and Dy12 alleles, respectively. Considering these SNPs, the haplotypes expected at *Glu-D1* by combining alleles at Mk-D1-1-8, Mk-D1-1-4 and Mk-D1-2-1 are AA-CC-GG for Glu-D1a (Dx2 + Dy12), GG-CC-GG for Glu-D1b (Dx3 + Dy12), and AA-TT-AA for Glu-D1d (Dx5 + Dy10). The rare *Glu-D1e* allele (Dx2 + Dy10) could be distinguished with this set of markers by the haplotype AA-CC-AA.

*Glu-B1* is known for its high level of diversity, which makes it difficult to identify HMW-GS encoded at this locus. Unsurprisingly, the alignment of sequences (including the promoter region, UTRs and CDS) reveals up to 124 SNPs plus 8 indels (>1bp) in the CDS of *Glu-B1-1* and 94 SNPs plus 4 indels in the CDS of *Glu-B1-2*. Six markers were designed for each gene to reflect this diversity as much as possible (Table 2, Figures S5 and S6). The decision tree deduced from the alignment at *Glu-B1-1* is as follows: T at Mk-B1-1-1 specifies the HMW-GS Bx7 or Bx17 alleles (Ravel et al. 2014); T at Mk-B1-1-4 specifies the Bx6 and Bx6.1 alleles; A (that is T on the reverse strand) at Mk-B1-1-8 specifies the Bx13 allele; and A at Mk-B1-1-9 specifies the Bx20 or Bx14 alleles. The two remaining markers Mk-B1-1-11 and Mk-B1-1-OE were designed in the promoter region to distinguish the Bx7 overexpressed allele (Bx7<sup>OE</sup>). The C polymorphism at Mk-B1-1-11 should therefore indicate this HMW-GS. However, only a few sequences were available at this location for our alignment, making it difficult to predict the behaviour of this marker with confidence. In addition, we noticed that its allele-specific primers did not match the sequence of the Bx23 HMW-GS and missing data at this marker may

characterize this allele. Mk-B1-1-OE was designed at the junction of both repeats of *Glu-B1-1* to again characterize Bx7<sup>OE</sup>. At *Glu-B1-2*, the alignment illustrated the difficulty of identifying SNP markers for By types. Indeed, there are polymorphisms even between HMW-GS sequences of a given By type. We designed six markers. Mk-B1-2-19, -11, -14 and -20 should be able to discriminate the By8 alleles. G (that is C on the reverse strand) at Mk-B1-2-19 indicates the By8 allele associated with Bx6, G at Mk-B1-2-15 should specify the By15 alleles (and some By8 alleles), while T at Mk-B1-2-18 should characterize By18. However, behaviour of these markers is difficult to predict due to the intra-By variability.

### **Comparison of molecular markers and SDS-PAGE profiling for HMW-GS identification**

(Table S2)

At *Glu-A1-1*, 340 accessions have one of the three main SDS-PAGE alleles Glu-A1a, b or c (for Ax1, Ax2\* and Axnull, respectively). Of these, SNP calling with at least one of the two molecular markers produced missing or heterozygous data in only 3 and 10 accessions, respectively. In the 327 accessions with complete data, results from SDS-PAGE and molecular markers converged to specify identical alleles for all but six lines (1.8%; Fig. 1). Accession 8233 has the haplotype GG-GG diagnostic of Glu-A1c even though it has a typical Glu-A1b profile. The three lines 1236, 1332, and 2153 are AA-AA (Glu-A1a) although the electrophoretic profile was characteristic of the null allele. The accessions 236 and 4482 are both scored as GG-AA and hence as Glu-A1b, but in electrophoresis appear as Glu-A1c and Glu-A1a alleles, respectively. Each marker was responsible for half of the misclassifications (3 each) indicating that the error rate per marker is less than 1%.



At *Glu-D1*, 314 lines were assigned as having one of the main alleles, Glu-D1a (Dx2 + Dy12), Glu-D1b (Bx3 + Dy12), or Glu-D1d (Dx5 + Dy10). In these lines, we only observed one instance of missing data at Mk-D1-1-4. As expected, lines with the haplotypes AA-TT and GG-CC have the Dx5 and Dx3 HMW-GS, respectively (Fig. 2). However, the lines with Dx3 (n = 28) are distributed between two haplotypes, GG-CC and AA-CC (7 and 21 lines, respectively). Thus Mk-D1-1-8 allows 25% of the lines with Dx3 to be detected. No marker was designed for the Dx4. Consequently, the lines with this allele cannot be discriminated. No data was missing for the marker Mk-D1-2-1 at *Glu-D1-2* and results were perfectly consistent with SDS-PAGE identification (Fig. 2).

The SDS-PAGE allele at *Glu-B1* was unambiguously identified for 318 lines. Genotyping with the six SNP markers designed at *Glu-B1-1* gave missing data in one to seven lines per marker. The percentage of missing data for these markers was thus low (no more than 2%). As predicted, the marker Mk-D1-1-11 gave three alleles (Table 4). Mk-D1-1-11 data was missing for all the lines with Bx20 or Bx23. All these lines formed a third cluster with Mk-B1-1-11 showing its reproducibility. This third group is likely due to sequence dissimilarity between the allele-specific primers and the target genome region as observed for Bx23. Therefore, lines in this cluster can be considered as off-target variants (OTVs). Taking into account the OTVs at Mk-B1-11, the markers at *Glu-B1-1* defined nine haplotypes, most of them restricted to a given Bx type (Fig. 3). The haplotypes called H6, H7, H8 and H9 in Table 4 specify the Bx13, Bx14, Bx20 and Bx14, respectively. Four haplotypes (H1, H2, H3 and H4) contains all the lines with the Bx7 and Bx7<sup>OE</sup>. The haplotypes H1 and H4 include each two HMW-GS types. In H1, we observed all the Bx7<sup>OE</sup> lines plus two Bx7 lines. H4 includes the Bx7 and Bx17 types. All the Bx17 lines are found in this haplotype (Fig. 3). When the y type is considered, all the Bx7 + By9 lines, except for accession 1005, are also in H4 (Fig. 3) and, all

the lines with Bx7 and silent for *Glu-B1-2* are in H2 (except for accession 7276). Therefore, lines with the Bx7 + By8 combination are either in H2 (62 lines, 58%) or H4 (45 lines, 41%). The haplotype H5 corresponds to the Bx6 and Bx6.1. Mk-B1-1-8 and Mk-B1-1-9 are diagnostic markers as one of their alleles is strictly associated with a given Bx type. All the lines (n = 13) with T at Mk-B1-1-8 have the Bx13 allele and those with A at Mk-B1-1-9 have the Bx20 allele (n = 15) except accession 1232. The T allele at Mk-B1-1-4 signed all the lines with the HMW-GS Bx6 or Bx6.1. At Mk-B1-1-1 230 lines have a T polymorphism specifying Bx7 or Bx17 HMW-GSs. With this marker 98% of the lines with these two HMW-GSs were correctly identified, apart from the accession 1236, which has C instead. In addition, accessions 1232 and 23896 have the T polymorphism despite having neither the Bx7 nor the Bx17 allele (Table S1). The misclassification of accession 1232 had already been noted with Mk-B1-1-9, so may have been due to experimental error. As expected, Mk-B1-1-OE discriminated the lines with the Bx7<sup>OE</sup> allele, which all have the C polymorphism.

There were only a few instances of missing data (for 1 to 5 lines) with four out of the six markers at *Glu-B1-2*. However, with Mk-B1-2-11 and Mk-B1-2-18 there was missing data for 31 lines. None of the 14 haplotypes (Table 4) defined by these markers matched the groups defined by SDS-PAGE profiles. Deducing the SDS-PAGE allele at this locus through SNP markers is clearly difficult (Fig. 4). It is nevertheless possible to say that all the lines which are A at Mk-B1-2-14 or G at Mk-B1-2-19 have the HMW-GS By8, except for two lines with the A at Mk-B1-2-14 that are silent at *Glu-B1-2*. As expected, all the lines with the T at Mk-B1-2-18 are By18, but 41% of the lines with By18 do not have T at this site. Although Mk-B1-2-15 was designed from a sequence representative of By15, neither of the two lines in our collection with this HMW-GS had the expected G allele. All the lines with By16 have the T polymorphism at Mk-B1-2-20 as do 34 other lines (28 of which have the silent allele at *Glu-B1-2*). Puzzlingly,

both alleles at Mk-B1-2-11 were observed in all the By types, except By22 in six lines. Thus, this marker specifies two groups without any correspondence to a particular HMW-GS.

The markers at *Glu-B1-2* were analysed in terms of the haplotype at *Glu-B1-1*. Mk-B1-2-19 is a diagnostic marker as the G allele is strictly associated with the By8 of *Glu-B1d* (Bx6 + By8). This marker could be used to specify the electrophoretic alleles of lines sharing identical haplotypes at *Glu-B1-1* like Bx6 and Bx6.1. In this case, T at Mk-B1-2-19 clearly indicates By22 associated with Bx6.1. Similarly, haplotypes AA-CC-CC-TT-GG-CC (H2) and AA-GG-CC-TT-GG-CC (H4) at *Glu-B1-1* correspond to several SDS-PAGE alleles. For these two haplotypes, taking markers at *Glu-B1-2* into account could be conclusive. For H4, as expected, T at Mk-B1-2-18 discriminates 59% of lines (10 out 17) with the *Glu-B1i* (Bx17 + By18) allele. For both haplotypes, A at Mk-B1-2-14 is clearly associated with the By8 allele of *Glu-B1b*. This allows the assignment of 66% of lines with the *Glu-B1b* allele. Finally, for haplotype H2, G at Mk-B1-2-14 and T at Mk-B1-2-20 indicates *Glu-B1a* (Bx7 + ByNull).

Thus, correct assignment was possible with markers at *Glu-A1* and *Glu-D1*, but some discrepancies remain at *Glu-B1*. Because we focused on the main SDS-PAGE alleles, the less frequent alleles could not be typed with the markers developed here. Generally, rare electrophoretic alleles are confounded with more frequent ones, such as in Comet (accession 2301), which has the rare *Glu-A1t* allele and was classified as *Glu-A1b* using molecular markers, or in lines with Dx4, Dx2.2 or Dxnull, which are all confused with Dx2 (Fig. 2). Because of the discrepancies found at *Glu-B1-1*, the question arose as to whether SNP markers at this locus are as relevant as the SDS-PAGE alleles in terms of technological quality.

### **SNP markers in HMW-GS genes to breed for technological quality**

For *Glu-B1*, both techniques reveal similar but not identical groupings. To see whether molecular markers at this locus could be useful in improving the technological value of bread, their effects on rheological parameters of dough were measured and statistically compared to those of the electrophoretic alleles using a linear model with the structure components of each line, the grain hardness and the grain protein content as covariates. Such an analysis was also performed using the haplotypes resulting from the combination of individual markers at each gene or at the locus to reflect as best as possible the polymorphism found at the *Glu-B1* locus as a factor.

Significant associations at 1% between some markers and technological quality traits (alveograph parameters and total quantity of protein per grain) were detected. Two markers per locus were significantly associated with at least one of the traits studied, Mk-B1-1-1 and Mk-B1-1-8 at *Glu-B1-1* and Mk-B1-2-11 and Mk-B1-2-14 at *Glu-B1-2* (Table 5). The polymorphism T at Mk-B1-1-8 increased the quantity of protein per grain by 20%. The polymorphism T at Mk-B1-2-11 strongly increased strength, tenacity, and the ratio of tenacity to extensibility. Respectively, strength, tenacity, and the tenacity to extensibility ratio were on average 1.6, 1.35 and 1.4 times higher for the T than for the C polymorphism at this marker. This marker did not influence dough extensibility suggesting that its action on strength is explained by the significant changes in tenacity alone.

A missing value at a single SNP marker generates an undefined haplotype. To bypass this problem, we decided to study haplotypes derived from the four significant markers only. Results from haplotype analysis confirmed preceding results (Table 6). When significant, haplotypes explained either a similar or even a higher proportion of the phenotypic variance observed than individual markers. As *Glu-B1-1* and *Glu-B1-2* are tightly linked, we also defined global haplotypes from the combination of haplotypes at each locus. Global haplotypes

significantly affected all the traits studied except for extensibility. Global haplotypes generally affected the traits more strongly than the haplotypes made only with markers at each HMW-GS gene. For example, for dough strength the global haplotype explained the highest proportion of the phenotypic variance (13%; Table 6). On average, dough strength of the best global haplotype (TT-CC at *Glu-B1-1* and TT-GG at *Glu-B1-2*) was more than twice that of the worst (CC-CC and CC-GG, respectively).

SDS-PAGE alleles at *Glu-B1* loci significantly influenced the total quantity of protein per grain, the dough strength and tenacity (Table 7). In accordance with results from global haplotype analysis, the strongest effect is observed for dough strength with up to 7% of the total phenotypic variance explained by the SDS-PAGE profile at *Glu-B1-1* and *Glu-B1-2*. *Glu-B1c* (Bx7 + By9), *Glu-B1b* (Bx7 + By8) and *Glu-B1i* (Bx17 + By18) are the best alleles with strength values on average 70% higher than those of the worst group made up of the *Glu-B1a* (Bx7 + Bynull), *Glu-B1d* (Bx6 + By8), *Glu-B1e* (Bx20 + By20), and *Glu-B1f* (Bx13 + By16) alleles.

At *Glu-B1-2*, Mk-B1-2-11 appeared to be an interesting marker to improve quality. It is one of the most associated markers without being related to a given SDS-PAGE allele. This marker could be seen as a novel means to improve quality. Remarkably, the proportion of phenotypic variance explained by the model based on molecular markers combined into global haplotypes for both HMW-GS genes was about twice as high as that obtained with the model based on SDS-PAGE information.

## Discussion

### HMW-GS loci in the core collection are highly polymorphic

Using the SDS-PAGE technology on a collection of 364 lines representing the worldwide diversity of bread wheat, and taking into account novel alleles detected, we observed at least 8, 22 and 9 alleles at the *Glu-A1*, *-B1*, and *-D1* loci, respectively. We found ~5% of lines exhibited heterozygosity at least at one of the six *Glu-1* loci. Heterozygosity may reflect the fact that the collection includes some lines that are really a mix of seeds, which may occur when landraces are used and the seed sets are not pure enough. Heterozygosity may also result from duplication of a locus within the genome, a phenomenon that is frequent in the wheat genome (Choulet et al. 2014; Glover et al. 2015). The number of alleles detected was more than was described by Shewry et al. (1992), who reported the presence of 3 (a, b, c), 11 (from a to k) and 6 alleles for *Glu-A1*, *-B1*, and *-D1* loci, respectively, in a collection of 300 lines. The latter alleles, apart from *Glu-Bj* (Bx21 + By21), are also present in the INRA core collection we studied. In both collections, the most frequent alleles at each locus were the same although we observed different frequencies of alleles at *Glu-B1* probably due to the different genetic structures of these two collections. Indeed, the set of lines used by Shewry (1992) was not representative of the worldwide wheat diversity. In addition, we observed novel HMW-GSs at each locus illustrating the high level of diversity of the INRA collection (Balfourier et al. 2007; Horvath et al. 2009) and validating the sampling strategy that was based on maximising the number of alleles at neutral markers (Schoen and Brown 1993) and the number of geographical origins. The novel HMW-GSs found should be confirmed before being added to the catalogue of *Glu-1* alleles started by Payne and Lawrence (1983), available through databases (e.g. the database of the National Bioresource Project at <https://shigen.nig.ac.jp/wheat/komugi/genes/macgene/2013/GeneSymbol.pdf>), and named according to current nomenclature. The high level of polymorphism found in the INRA core collection proved a valuable source of material to reach our objective. This material is likely to be much more complex than that generally used by breeders.

## **How well do KASPar assays identify HMW-GS compared to SDS-PAGE?**

HMW- and LMW-GS strongly influence dough functionality (Shewry et al, 2002). For their identification, molecular markers have many advantages over biochemical markers. Therefore, HMW- and LMW-GS gene markers have been developed (see for instance Zhang et al. 2004; Wang et al. 2009, Wang et al. 2010, Iba et al. 2018). In addition, different systems based on PCR-markers have been developed as reported for LMW-GS by Zhang et al. (2011). In our work, for an easy and early identification of HMW-GSs, we focused our marker development strategy on SNPs since such markers are now recognized worldwide as the best markers for high-throughput genotyping in most species, even in polyploids. Numerous strategies for SNP genotyping, reviewed by Paux et al. (2012), have given rise to several high-throughput wheat genotyping tools like high-density arrays (for instance, see Wang et al. 2014, Rimbart et al. 2017) or KASP assays (Allen et al. 2011). Here, the set of SNP markers developed is based on the KASPar technology, which requires two allele-specific primers and a common primer. The latter primer has to target a locus-specific region. This makes KASPar technology flexible and well-suited for use in polyploid species or when genotyping members of multigene families. This flexibility and the relatively low cost of KASPar technology explain why it has been adopted for plant genotyping for research or breeding purposes. In wheat, it has been successfully used for fingerprinting diverse collections of accessions (Gao et al. 2016) or for marker-assisted breeding as demonstrated by Neelam et al. (2013), who developed a diagnostic KASPar assay for a locus involved in leaf rust resistance.

A set of KASPar markers to identify HMW-GS should provide breeders with a more convenient and efficient tool than SDS-PAGE to select HMW-GS associated with a high technological quality. To reach this objective, we developed 17 KASPar assays and observed a high degree

of concordance between the HMW-GS identified by SDS-PAGE and molecular markers despite a few discrepancies between the two methods. Indeed, just two and three SNP markers respectively can be used to differentiate the three main alleles at *Glu-A1* and up to three alleles at *Glu-D1*. However, identification of *Glu-D1b* is partial. The six markers at *Glu-B1-1* discriminated the main Bx types known, while six markers at *Glu-B1-2* can help to define alleles at *Glu-B1* once the Bx type is determined. The differences observed are mainly explained by the fact that SNP markers distinguished more alleles than electrophoresis of proteins. This is the case for the HMW-GS Bx7, By8, and By18. Based on markers, the By8 associated with Bx7 differs from the By8 associated with Bx6. SNP markers discriminated three groups including lines with Bx7 in addition to the group with Bx7<sup>OE</sup>. This result is not surprising since the four alleles *Glu-B1b*, *Glu-B1u*, *Glu-B1ak*, and *Glu-B1al* (Bx7 + By8, Bx7\* + By8, Bx7\* + By8\*, and Bx7<sup>OE</sup> + By8\*, respectively) are difficult to distinguish using SDS-PAGE (Marchylo et al. 1992). Similarly, markers indicate that By18 does not correspond to a unique sequence. This is also the case for the Dx3 allele found in at least two molecular haplotypes. One of them is identical to that of Dx2, making it impossible to detect all the lines having the Dx3 allele with the SNP markers developed here. Discrepancies could also result from a lack of discriminating SNP markers as for Bx17, which has a molecular haplotype similar to those of some lines carrying Bx7. All SDS-PAGE silent alleles gave clear SNP alleles confirming that they do not derive from deletions. Indeed, the molecular basis of the silent *Glu-A1-2* gene may involve stop codons that would lead to premature termination of translation (Forde et al. 1985; Bustos et al. 2000), transposon-like insertion in the coding sequence (Harberd et al. 1987), or transcriptional inactivation due to alteration probably in the distal 5' or 3' regions (D'Ovidio et al. 1996). Discrepancies observed in typing HMW-GS by SDS-PAGE or molecular markers were expected as electrophoresis is based on the length polymorphism of proteins while KASPar assays on nucleotide sequence polymorphism. The latter were more frequent at *Glu-B1* as this



locus is known for its high level of diversity with more than 69 SDS-PAGE alleles already reported (Lei et al. 2006; MacIntosh et al. 2013; Janni et al. 2017). At this locus, for *Glu-B1-2*, Mk-B1-2-11 clearly identified alleles which did not match any known SDS-PAGE allele. This marker thus brings novel information and confirms that the level of polymorphism observed at the gene level is higher than at the protein level. Resolution at *Glu-B1* is improved by taking into account the haplotypes of both genes *Glu-B1-1* and *-2*. However, the question then arises as to whether molecular haplotypes at *Glu-B1* would be useful for breeding purposes even if they do not perfectly reflect the diversity profiled by SDS-PAGE.

#### **A set of SNP markers at *Glu-B1* to improve end-use quality**

Association analysis was used to investigate the impact of SNP markers at *Glu-B1* on the total quantity of protein per grain and the alveographic parameters (dough strength, tenacity, extensibility, and the tenacity to extensibility ratio). Two markers per HMW-GS gene, and their respective haplotypes, affect all these traits except extensibility. The strongest effect is observed on dough strength. This confirms that HMW-GSs have a key role in determining dough tenacity and strength by forming the backbone of the gluten network (Shewry et al. 2001). Extensibility was not modified by HMW-GSs, as was expected, as this parameter depends mainly on LMW-GS (Rasheed et al. 2014). The global haplotype defined by the combination of significant markers at *Glu-B1-1* and *-2* explained a larger proportion of the phenotypic variance than was found with the SDS-PAGE alleles. Breeding for processing quality using these four SNP markers at *Glu-B1*, plus the two at *Glu-A1* and three at *Glu-D1* (markers for the latter two genes matched the SDS-PAGE alleles well) appears more efficient than using SDS-PAGE alleles. In addition, using molecular markers will offer breeders a simple low-cost tool for selection at the

plantlet stage. This time-saving is a great advantage compared to SPS-PAGE that can only be done on grains.

To conclude, the SNP markers developed here provide an easy tool for typing HMW-GSs. Generally, SNP-defined haplotypes fit SDS-PAGE allele profiles well despite a few differences. The minimum number of markers needed to select for improved processing qualities is eight, namely Mk-A1-1-12 and -13 at *Glu-A1*, either Mk-D1-1-4 for *Glu-D1-1* or Mk-D1-2-1 for *Glu-D1-2* (redundant markers probably due to a high level of linkage disequilibrium between the two loci), and Mk-D1-1-8 at *Glu-D1*, plus at least four markers (out of 12) associated with rheological traits at *Glu-B1*. Despite absence of polymorphism at Glu-B-1-2-11, similar associations have been detected in a file of elite lines (n=245) showing the consistency of our results. However, HMW-GSs are not the sole determinants of end-use performance of flour as breeders have to deal with other traits, especially, as far as storage proteins are concerned, with low-molecular-weight glutenins and gliadins. The HMW-GS markers have been tested here on a large diverse set of wheat germplasm, so are ready for use in large-scale and high-throughput HMW-GS screening in breeding programs.

#### **Author contribution statement**

CR, PS, GB designed the experiment. CR analysed the sequences and designed the primers for SNP genotyping. AF extracted grain proteins and performed electrophoresis (SDS-PAGE). FE carried out DNA extraction. MR, MD, CD, FE genotyped the SNPs. CP supervised SNP genotyping. CR and SBS analysed the data. CR wrote the manuscript. All authors commented on the manuscript.

## Acknowledgements

The authors thank François Balfourier and Jacques Bordes for providing all the accessions of the INRA core collection and alveographic data, as well as Rachel Carol from Emendo Bioscience for English corrections. The genotyping project was conducted on the genotyping platform GENTYANE at INRA Clermont-Ferrand ([gentyane.clermont.inra.fr](http://gentyane.clermont.inra.fr)).

## Compliance with ethical standards

**Conflict of interest:** On behalf of all authors, the corresponding author states that there is no conflict of interest.

## References

- Akhunov E, Nicolet C, Dvorak J (2009) Single nucleotide polymorphism genotyping in polyploid wheat with the Illumina GoldenGate assay. *Theor Appl Genet* 119:507-517. <https://doi.org/10.1007/s00122-009-1059-5>
- Allen AM, Barker GL, Berry ST et al (2011) Transcript-specific, single-nucleotide polymorphism discovery and linkage analysis in hexaploid bread wheat (*Triticum aestivum* L.). *Plant Biotechnol J* 9:1086–1099. <https://doi.org/10.1111/j.1467-7652.2011.00628.x>
- Balfourier F, Roussel V, Strelchenko P, Exbrayat-Vinson F, Sourdille P, Boutet G, Koenig J, Ravel C, Mitrofanova O, Beckert M, Charmet G (2007) A worldwide bread wheat core collection arrayed in a 384-well plate. *Theor Appl Genet* 114:1265-1275. <https://doi.org/10.1007/s00122-007-0517-1>
- Bembom O (2016) seqLogo: Sequence logos for DNA sequence alignments. R package version 1.40.0
- Bérard, A, Le Paslier MC, Dardevet M, Exbrayat-Vinson F, Bonnin I, Cenci A, Haudry A, Brunel D, Ravel C (2009) High-throughput single nucleotide polymorphism genotyping in wheat (*Triticum* spp.). *Plant Biotechnol J* 7: 364-374. <https://doi.org/10.1111/j.1467-7652.2009.00404.x>
- Bordes J, Branlard G, Oury FX, Charmet G, Balfourier F (2008) Agronomic characteristics, grain quality and flour rheology of 372 bread wheats in a worldwide core collection. *J Cereal Sci* 48:569–579
- Branlard G, Dardevet M (1985) Diversity of grain protein and bread wheat quality .2. Correlation between high molecular-weight subunits of glutenin and flour quality characteristics. *J Cereal Sci* 3:345-354

- Branlard G, Dardevet M, Saccomano R, Lagoutte F, Gourdon J (2001) Genetic diversity of wheat storage proteins and bread wheat quality. *Euphytica* 119:59-67
- Bustos AD, Rubio P, Jouve N (2000) Molecular characterization of the inactive allele of the gene *Glu-A1* and the development of a set of AS-PCR markers for HMW glutenins of wheat. *Theor Appl Genet* 100:1085-1094
- Choulet F, Alberti A, Theil S, Glover NM, Barbe V, Daron J et al (2014) Structural and Functional Partitioning of Bread Wheat Chromosome 3B. *Science* 345(6194):1249721. <https://doi.org/10.1126/science.1249721>
- D'Ovidio R, Masci S, Porceddu E (1996) Sequence analysis of the 5' non-coding regions of active and inactive 1Ay HMW glutenin genes from wild and cultivated wheats. *Plant Sci.* 14: 61-69
- Eagles HA, Cane K, Eastwood RF, Hollamby GJ, Kuchel H, Martin PJ, Cornish GB (2006) Contributions of glutenin and puroindoline genes to grain quality traits in southern Australian wheat breeding programs. *Aust J Agric Res* 57:179–186
- Fido R, Bekes F, Gras P, Tatham A (1997) Effects of  $\alpha$ -,  $\beta$ -,  $\gamma$ - and  $\omega$ -gliadins on the dough mixing properties of wheat flour. *J Cereal Sci* 26:271-277
- Flint-Garcia SH, Thornsberry JM, Buckler ES IV (2003) Structure of linkage disequilibrium in plants. *Annu Rev Plant Biol* 54:357–374. <https://doi.org/10.1146/annurev.arplant.54.031902.134907>
- Forde J, Malpica J-M, Halford NG, Shewry PR, Anderson OD, Greene FC, Mifflin BJ (1985) The nucleotide sequence of a HMW glutenin subunit gene located on chromosome 1A of wheat (*Triticum aestivum* L.). *Nucleic Acids Res* 13:6817-6832
- Gale K (2005) Diagnostic DNA markers for quality traits in wheat. *J Cereal Sc* 41: 181-192. <https://doi.org/10.1016/j.jcs.2004.09.002>
- Gao LF, Jia JZ, Kong XY (2016) A SNP-Based Molecular Barcode for Characterization of Common Wheat. *PLoS One* 11:e0150947. <https://doi.org/10.1371/journal.pone.0150947>
- Glover NM, Daron J, Pingault L, Vandepoele K, Paux E, Feuillet C, Choulet F (2015) Small-scale gene duplications played a major role in the recent evolution of wheat chromosome 3B. *Genome Biol* 16:188. <https://doi.org/10.1186/s13059-015-0754-6>
- Gobaa S, Kleijer G, Stamp P (2007) 2<sup>+</sup>, a new high molecular weight glutenin subunit coded by *Glu-A1*: its predicted structure and its impact on bread-making quality. *Plant Breeding* 126:1-4. <https://doi.org/10.1111/j.1439-0523.2006.01313.x>
- Harberd NP, Flavell RB, Thompson RD (1987) Identification of a transposon like insertion in a *Glu-1* allele of wheat. *Mol Gen Genet* 209:326-332
- Horvath A, Didier A, Koenig J, Exbrayat F, Charmet G, Balfourier F (2009) Analysis of diversity and linkage disequilibrium along chromosome 3B of bread wheat (*Triticum aestivum* L.). *Theor Appl Genet* 119: 1523-37. <https://doi.org/10.1007/s00122-009-1153-8>
- Iba M.I, Kiszonas A, Morris CF. (2018) Development of haplotype-specific molecular markers for the low-molecular-weight glutenin subunits. *Mol Breeding* 38:68. <https://doi.org/10.1007/s11032-018-0827-9>

- Janni M, Cadonici S, Pignone D, Marmioli N (2017) Survey and new insights in the application of PCR-based molecular markers for identification of HMW-GS at the *Glu-B1* locus in durum and bread wheat. *Plant Breeding*. <https://doi.org/10.1111/pbr.12506>
- Jin H, Yan J, Peña RJ, Xia XC, Morgounov A., Han LM, Zhang Y, He ZH (2011) Molecular detection of high- and low-molecular-weight glutenin subunit genes in common wheat cultivars from 20 countries using allele-specific markers. *Crop & Pasture Sc* 62:746-754. <https://doi.org/10.1071/CP11134>
- Larkin M, Blackshields G, Brown N, Chenna R, McGettigan P, McWilliam H, Valentin F, Wallace I, Wilm A, Lopez R, Thompson J, Gibson T, Higgins D (2007) Clustal W and clustal X version 2.0. *Bioinformatics* 23:2947-2948. <https://doi.org/10.1093/bioinformatics/btm404>
- Lei ZS, Gale MR, He ZH, Gianibelli C, Larroque O, Xia XC, Butow BJ, Ma W (2006) Y-type gene specific markers for enhanced discrimination of high-molecular weight glutenin alleles at the *Glu-B1* locus in hexaploid wheat. *J Cereal Sci* 43:94-101. <https://doi.org/10.1016/j.jcs.2005.08.003>
- Liu SX, Chao S, Anderson J (2008) New DNA markers for high molecular weight glutenin subunits in wheat. *Theor Appl Genet* 118:177-183. <https://doi.org/10.1007/s00122-008-0886-0>
- Liu L, Ikeda TM, Branlard G, Pena RJ, Rogers WJ, Lerner SE, Kolman MA, Xia X, Wang L, Ma W, Appels R, Yoshida H, Wang A, Yan Y, He Z. (2010) Comparison of low molecular weight glutenin subunits identified by SDS-PAGE, 2-DE, MALDI-TOF-MS and PCR in common wheat. *BMC Plant Biol*. 10: 124. <https://doi.org/10.1186/1471-2229-10-124>
- Liu Y, He Z, Appels R, Xia X (2012) Functional markers in wheat: current status and future prospects. *Theor Appl Genet* 125:1-10. <https://doi.org/10.1007/s00122-012-1829-3>
- Lu G, Moriyama EN (2004) Vector NTI, a balanced all-in-one sequence analysis suite. *Brief Bioinform* 5:378-388. <https://doi.org/10.1093/bib/5.4.378>
- McIntosh R, Dubcovsky J, Rogers J, Morris C, Appels R, Xia XC (2013) Catalogue of gene symbols for wheat. In: 12th International Wheat Genetics Symposium, Yokohama, Japan
- MacRitchie F (1999) Wheat proteins: Characterization and role in flour functionality. *Cereal Foods World* 44(4):188-193
- Marchylo BA, Lukow OM, Kruger JE (1992) Quantitative variation in high molecular weight glutenin subunit 7 in some Canadian wheats. *J Cereal Sci* 15:29-37
- Margiotto B, Urbano M, Colaprico G (1996) Detection of y-type subunit at the *Glu-A1* locus in some Swedish bread wheat lines. *J. Cereal Sci.* 23:203-211
- Neelam K, Brown-Guedira G, Huang L (2013) Development and validation of a breeder-friendly KASPar marker for wheat leaf rust resistance locus Lr2. *Mol Breeding* 13:233-237. <https://doi.org/10.1007/s11032-012-9773-0>
- Paux E, Sourdille P, Mackay I, Feuillet C (2011) Sequence-based marker development in wheat: advances and applications to breeding. *Biotechnology Advances* 30:1071-1088. <https://doi.org/10.1016/j.biotechadv.2011.09.015>
- Payne PI (1997) Genetics of wheat storage proteins and the effect of allelic variation on bread-making quality. *Ann Rev Plant Physiol* 38:141-153

- Payne PI, Corfield KG, Blackman JA (1979) Identification of a high-molecular-weight subunit of glutenin whose presence correlates with bread-making quality in wheats of related pedigree. *Theor Appl Genet* 55:153-159
- Payne PI, Lawrence GJ (1983) Catalogue of alleles for the complex gene loci, *Glu-A1*, *Glu-B1*, and *Glu-D1* which code for high molecular-weight subunits of glutenin in hexaploid wheat. *Cereal Res Commun* 11:29-35
- Payne P I, Nightingale MA, Krattiger AF, Holt LM (1987) The relationship between HMW glutenin subunit composition and the bread-making quality of British-grown wheat varieties. *J Sci Food Agric* 40 (1):51–65
- Pirozi M, Margiotta B, Latiandra D, MacRitchie F (2008) Composition of polymeric proteins and bread-making quality of wheat lines with allelic HMW-GS differing in number of cysteines. *J Cereal Sci* 48: 117-122
- Ragupathy R, Naeem HA, Reimer E, Lukow OM, Sapirstein HD, Cloutier S (2008) Evolutionary origin of the segmental duplication encompassing the wheat GLU-B1 locus encoding the overexpressed Bx7 (Bx7OE) high molecular weight glutenin subunit. *Theor Appl Genet* 116:283-296. <https://doi.org/10.1007/s00122-007-0666-2>
- Rasheed A, Xia X, Yan Y, Appels R, Mahmood T, He Z (2014) Wheat seed storage proteins: Advances in molecular genetics, diversity and breeding applications. *J Cereal Sci* 60:11-24. <https://doi.org/10.1016/j.jcs.2014.01.020>
- Ravel C, Praud S, Murigneux A, Canaguier A, Sapet F, Samson D, Balfourier F, Dufour P, Chalhou B, Brunel D, Beckert M, Charmet G (2006) Single-nucleotide polymorphism frequency in a set of selected lines of bread wheat (*Triticum aestivum* L.). *Genome* 49:1131-1139. <https://doi.org/10.1139/g06-067>
- Ravel C, Fiquet S, Boudet J, Dardevet M, Vincent J, Merlino M, Michard R, Martre P (2014) Conserved cis-regulatory modules in promoters of genes encoding wheat high-molecular-weight glutenin subunits. *Front Plant Sci*.5:621. <https://doi.org/10.3389/fpls.2014.00621>
- R Core Team (2016) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Ribeiro M, Bancel E, Faye A, Dardevet M, Grand Ravel C, Branlard G, Igrejas G (2013) Proteogenomic Characterization of Novel x-Type High Molecular Weight Glutenin Subunit 1Ax1.1. *Int J Mol Sci* 14(3):5650-67. <https://doi.org/10.3390/ijms14035650>
- Rimbert H, Darrier B, Navarro J, Kitt J, Choulet F, Leveugle M et al (2017) High throughput SNP discovery and genotyping in hexaploid wheat. *PLoS One* 13(1) e0186329 <https://doi.org/10.1371/journal.pone.0186329>
- Sapone A, Bai JC, Ciacci C, Dolinsek J, Green PHR, Hadjivassiliou M, Kaukinen K, Rostami K, Sanders DS, Schumann M, Ullrich R, Villalta D, Volta U, Catassi C, Fasano A (2012) Spectrum of gluten-related disorders: consensus on new nomenclature and classification. *BMC Med* 10:13. <https://doi.org/10.1186/1741-7015-10-13>
- Semagn K, Babu R, Hearne S, Olsen M (2014) Single nucleotide polymorphism genotyping using Kompetitive Allele Specific PCR (KASP): overview of the technology and its application in crop improvement. *Mol Breeding* 33: 1-14. <https://doi.org/10.1007/s11032-013-9917-x>
- Shoen DJ, Brown AHD (1993) Conservation of allelic richness in wild crop relatives is aided by assessment of genetic markers. *Proc Natl Acad Sci U S A* 90:10623–10627. <https://dx.doi.org/10.1073%2Fpnas.90.22.10623>

- Schwarz G, Sift A, Wenzel G, Mohler V (2003) DHPLC scoring of a SNP between promoter sequences of HMW glutenin x-type alleles at the *Glu-D1* locus in wheat. *J Agr Food Chem* 51: 4263-4267. <https://doi.org/10.1021/jf0261304>
- Shewry PR (2009) Wheat. *J Exp Bot* 60:1537-53. <https://doi.org/10.1093/jxb/erp058>
- Shewry P, Halford N, Tatham A (1992) High-molecular-weight subunits of wheat glutenin. *J Cereal Sci* 15:105-120
- Shewry P, Tatham A (1997) Disulphide bonds in wheat gluten proteins. *J Cereal Sci* 25:207-227. <https://doi.org/10.1006/jcrs.1996.0100>
- Shewry PR, Popineau Y, Lafiandra D, Belton P (2001) Wheat glutenin subunits and dough elasticity: findings of the EUROWHEAT project. *Trends Food Sci Technol* 11:433-441. [https://doi.org/10.1016/S0924-2244\(01\)00035-8](https://doi.org/10.1016/S0924-2244(01)00035-8)
- Shewry PR, Halford NG, Belton PS, Tatham AS (2002) The structure and properties of gluten: an elastic protein from wheat grain. *Philos Trans R Soc Lond B Biol Sci* 357:133-142. <https://doi.org/10.1098/rstb.2001.1024>
- Shewry PR, Hey S (2015) The contribution of wheat to human diet and health. *Food Energy Secur* 4:178-202. <https://doi.org/10.1002/fes3.64>
- Singh N, Shepherd K, Cornish G (1991) A simplified SDS-PAGE procedure for separating LMW subunits of glutenin. *J Cereal Sci* 14:203-208
- Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG (2012) Primer3 - new capabilities and interfaces. *Nucleic Acids Res* 40(15):e115. <https://doi.org/10.1093/nar/gks596>
- Wang LH, Zhao XL, He ZH, Ma W, Appels R, Pena RJ, Xia XC (2009) Characterization of low-molecular-weight-glutenin sub-unit *Glu-B3* genes and development of STS markers in common wheat (*Triticum aestivum* L.). *Theor Appl Genet* 118:525-539. <https://doi.org/10.1007/s00122-008-0918-9>
- Wang LH, Li GY, Pena RJ, Xia XC, He ZH (2010) Development of STS markers and establishment of multiplex PCR for *Glu-A3* alleles in common wheat (*Triticum aestivum* L.). *J Cereal Sci* 51:305-312. <https://doi.org/10.1016/j.jcs.2010.01.005>
- Wang S, Wong D, Forrest K et al. (2014) Characterization of polyploid wheat genomic diversity using a high-density 90 000 single nucleotide polymorphism array. *Plant Biotechnol J* 12:787-796. <https://doi.org/10.1111/pbi.12183>
- Xu Q, Xu J, Liu C, Chang C, Wang C, You M, Li B, Liu G (2008) PCR-based markers for identification of HMW-GS at *Glu-B1x* loci in common wheat. *J Cereal Sci* 47: 394-398. <https://doi.org/10.1016/j.jcs.2007.05.002>
- Zhang W, Gianibelli MC, Rampling LR, Gale KR (2004) Characterization and marker development for low molecular weight glutenin genes from *Glu-A3* alleles of bread wheat (*Triticum aestivum* L.). *Theor Appl Genet* 108:1409-1419. <https://doi.org/10.1007/s00122-003-1558-8>
- Zhang X, Liu D, Yang W, Liu K, Sun J, Guo X, Li Y, Wang D, Ling H, Zhang A (2011) Development of a new marker system for identifying the complex members of the low-molecular-weight glutenin subunit gene family in bread wheat (*Triticum aestivum* L.). *Theor Appl Genet* 122:1503-1516. <https://doi.org/10.1007/s00122-011-1550-7>

## List of tables

**Table 1** List of HMW-GS sequences aligned to detect gene-specific regions and allelic SNPs

**Table 2** Primers designed for KASP assays to characterize the diversity of HMW-GS genes

**Table 3** Occurrence of HMW-GS proteins and corresponding *Glu-1* alleles in 364 accessions from the INRA core collection

**Table 4** Distribution of haplotypes at *Glu-B1* in 318 lines from the INRA core collection

**Table 5** Degrees of freedom, F and P values, and percentage of the phenotypic variance explained by the SNP markers at *Glu-1* associated to rheological traits.

**Table 6** Degrees of freedom, F and P values, and percentage of the phenotypic variance explained by haplotypes deduced from SNP markers at *Glu-1* associated to rheological traits.

**Table 7** Degrees of freedom, F and P values, and percentage of the phenotypic variance explained by SDS-PAGE profiles at *Glu-1* loci for rheological traits.

## Figure legends

**Fig. 1** Haplotypes based on molecular markers in relation to the SDS-PAGE allele at *Glu-A1-1* in accessions of the INRA core collection. Haplotypes combine alleles for Mk-A1-1-12 and Mk-A1-1-13 at *Glu-A1-1*.

**Fig. 2** Haplotypes based on molecular markers in relation to the SDS-PAGE allele at *Glu-D1* in accessions of the INRA core collection.

Haplotypes combine alleles for Mk-D1-1-8 and Mk-D1-1-4 at *Glu-D1-1*, followed by the allele for Mk-D1-2-1 at *Glu-D1-2*. Mk-D1-1-4 and Mk-D1-2-1 are designed on the reverse strand.

**Fig. 3 A.** Haplotype details at *Glu-B1-1* in accessions of the INRA core collection based on molecular markers. Haplotypes combine alleles for Mk-B1-1-OE, -11, -4, -1, -9 and -8 at *Glu-B1-1* (Table 4).

**A.** Haplotypes in relation to the SDS-PAGE allele at *Glu-B1-1* in accessions of the INRA core collection.

**B.** Pie charts indicating the Bx + By allele composition given by the SDS-PAGE method for haplotypes H2 and H4.

**Fig. 4** Logo sequences of haplotypes for each HMW-GS at *Glu-B1-2* represented by at least two lines in the INRA core collection. The order of markers is according to the order in the sequence.

## Supplementary materials

**Table S1.** HMW-GS alleles identified by SDS-PAGE in the 364 lines of INRA core collection

**Table S2.** HMW-GS alleles identified by SNP markers in the 364 lines of INRA core collection



**Fig S1.** Alignment of one sequence per *Glu-1* gene. This is a bioedit file with tags. Start and stop codons are indicated in yellow. The putative TSS (transcription start site) is in blue. Grey shading indicates the repetitive domain. This domain was not used for designing primers so alignment in this region was not manually improved.

**Fig. S2.** Alignment of CDS at *Glu-A1-1*. This is a bioedit file with tags. Start and stop codons are indicated in yellow. SNPs chosen for typing HMW-GS are in red. SNPs in the repetitive domains are in grey.

**Fig. S3.** Alignment of the region upstream of the start codon for *Glu-D1-1*. This is a bioedit file with tags. Start codons, the chosen SNPs and remaining SNPs are indicated in yellow, red and grey, respectively.

**Fig. S4.** Alignment of the promoter region and CDS for *Glu-D1-2*. This is a bioedit file with tags. Start and stop codons, the chosen SNP and the remaining SNPs are indicated in yellow, red and grey, respectively.

**Fig. S5.** Alignment of the promoter region and CDS for *Glu-B1-1*. This is a bioedit file with tags. Start and stop codons, the chosen SNP and the remaining SNPs are indicated in yellow, red and grey, respectively.

**Fig. S6.** Alignment of the promoter region, CDS and UTRs for *Glu-B1-2*. This is a bioedit file with tags. Start and stop codons are indicated in yellow while the chosen SNPs are indicated in red.

**Table 1** List of HMW-GS sequences aligned to detect gene-specific regions and allelic SNPs

Gene	Accession number	HMW-GS	Cultivar	Alignment to design	
				gene-specific primers	allele-specific primers
<i>Glu-A1-1</i>	DQ537335.1	Ax2*	Renan	Yes	Yes
<i>Glu-A1-1</i>	AF145590.1	Axnull	Pane 247	No	Yes
<i>Glu-A1-1</i>	X61009	Ax1	Hopea	No	Yes
<i>Glu-A1-2</i>	DQ537335.1	null	Renan	Yes	No
<i>Glu-B1-1</i>	DQ537336.1	Bx7	Renan	Yes	Yes
<i>Glu-B1-1</i>	KM116481	Bx6	Récital	No	Yes
<i>Glu-B1-1</i>	HQ731653	Bx6.1	Kotte	No	Yes
<i>Glu-B1-1</i>	EU157184.1	Bx7OE	Glenlea	No	Yes
<i>Glu-B1-1</i>	EF413002.1	Bx13	Jimai 20	No	Yes
<i>Glu-B1-1</i>	EF540764	Bx13	unknown	No	Yes
<i>Glu-B1-1</i>	AY367771.1	Bx14	unknown	No	Yes
<i>Glu-B1-1</i>	AB263219.1	Bx17	Haruyutaka	No	Yes
<i>Glu-B1-1</i>	KM116484	Bx20	Ralet	No	Yes
<i>Glu-B1-2</i>	DQ537336.1	By8	Renan	Yes	Yes
<i>Glu-B1-2</i>	FJ561336.1	By8	Soisson	No	Yes
<i>Glu-B1-2</i>	HQ731654.1	By8	Kotte	No	Yes
<i>Glu-B1-2</i>	JN255519.1	By8	Uri	No	Yes
<i>Glu-B1-2</i>	KM116487	By8	Récital	No	Yes
<i>Glu-B1-2</i>	X61026.1	By9	Cheyenne	No	Yes
<i>Glu-B1-2</i>	EU137874.1	BY15	Xiaoyan 54	No	Yes
<i>Glu-B1-2</i>	DQ086215	BY15	Xiaoyan 6	No	Yes
<i>Glu-B1-2</i>	EF540765.1	BY16	Atlas66 or Jimai20	No	Yes
<i>Glu-D1-1</i>	DQ208971.1	Dx2	Chinese Spring	No	Yes
<i>Glu-D1-1</i>	KM116493	Dx3	A4	No	Yes
<i>Glu-D1-1</i>	DQ537337.1	Dx5	Renan	Yes	Yes
<i>Glu-D1-2</i>	DQ537337.1	Dy10	Renan	Yes	Yes
<i>Glu-D1-2</i>	BK006459.1	Dy12	Chinese Spring	No	Yes

**Table 2** Primers designed for KASP assays to characterize the diversity of HMW-GS genes. Markers are ordered according to the gene and the marker name.

Gene	SNP location	Marker name	Type of primer <sup>(a)</sup>	Sequence (5'-3') <sup>a</sup>	HMW-GS target
<i>Glu-A1-1</i>	CDS	Mk-A1-1-12	ASP1	GAGCGCGAGCTCCAGGAG	Axnull Ax2*
			ASP2	GAGCGCGAGCTCCAGGAA	Ax1
			CP	GGCATGCCTTAAGCGAGTG	
<i>Glu-A1-1</i>	CDS	Mk-A1-1-13	ASP1	AAGTGTAACCTTCTCCGCAACA	Ax1 Ax2*
			ASP2	AAGTGTAACCTTCTCCGCAACG	Axnull
			CP	GGCCTGGATAGTATGAAACC	
<i>Glu-B1-1</i>	promoter	Mk-B1-1-OE	ASP1	AAAGAGGTAGGGGCAGGGA	Bx7 and all others
			ASP2	GAGCCCATCTCAGATTTTCAGC	Bx7OE
			CP	TGTCCATTATGCCCTCCTATG	
<i>Glu-B1-1</i>	promoter	Mk-B1-1-11	ASP1	GAATATTAGTGATGGCGTGAG	Bx7 and all others
			ASP2	GAATATTAGTGATGGCGTGAC	Bx7OE
			CP	GACGCACTTCTTCTCTCG	
<i>Glu-B1-1</i>	promoter	Mk-B1-1-4	ASP1	GCTCCTTACTTATCCAGCTTC	Bx7 Bx17 Bx20
			ASP2	GCTCCTTACTTATCCAGCTTT	Bx6
			CP	AATCGGTTGGAAAAGGGC	
<i>Glu-B1-1</i>	5'UTR	Mk-B1-1-1	ASP1	ATCAATTCAGTACAGTTCCT	Bx7 Bx17
			ASP2	ATCAATTCAGTACAGTTCACC	Bx6 Bx13 Bx20
			CP	CCGCAAAGAGGACCAGG	
<i>Glu-B1-1</i>	CDS	Mk-B1-1-9	ASP1	CAGTAGTCGTCGCCCTCG	Bx6 Bx7 Bx17
			ASP2	CAGTAGTCGTCGCCCTCA	Bx14 or Bx20
			CP	TTGCTGGTCCACCACCTGTT	
<i>Glu-B1-1</i>	CDS	Mk-B1-1-8 <sup>c</sup>	ASP1	TGCTGCGAAGGCGTAGTCTC	Bx7 Bx17 Bx20
			ASP2	TGCTGCGAAGGCGTAGTCTT	Bx13
			CP	AACAGGTGGTGGACCAGCAA	
<i>Glu-B1-2</i>	promoter	Mk-B1-2-19 <sup>c</sup>	ASP1	GCCTAAGAAGAAACGTGAGAA	Other By
			ASP2	GCCTAAGAAGAAACGTGAGAG	Some By8
			CP	ATTACGTGGCTTTAGCAGACT	
<i>Glu-B1-2</i>	promoter	Mk-B1-2-15	ASP1	CTTAGGCTAAACTAACCACA	Other By
			ASP2	CTTAGGCTAAACTAACCACG	By15, some By8
			CP	GACGTGGTGAAGGTTTCG	
<i>Glu-B1-2</i>	CDS	Mk-B1-2-11	ASP1	CGAGATGGCTAAGCGGT	Other By
			ASP2	CGAGATGGCTAAGCGGC	By16, some By8
			CP	AACTGTTGGTCCACGACCT	
<i>Glu-B1-2</i>	CDS	Mk-B1-2-14 <sup>c</sup>	ASP1	TTGGCTGACGGCGACGG	Other By
			ASP2	TTGGCTGACGGCGACGA	some By8
			CP	CAGCTCCGAGATGTTAGCGCT	
<i>Glu-B1-2</i>	CDS	Mk-B1-2-20	ASP1	CGTAAGACAATATGAGCAAACC	some By8
			ASP2	CGTAAGACAATATGAGCAAACCT	Other By
			CP	TTTGTGAAGATGTTCCCCA	
<i>Glu-B1-2</i>	3'UTR	Mk-B1-2-18	ASP1	ATCAATTGTACTCCTTTTCGA	Other By

<i>Glu-D1-1</i>	promoter	Mk-D1-1-8	ASP2	ATCAATTGTACTCCTTTCGT	By18
			CP	ACTGTTACTTCCAGTGCTTG	
			ASP1	TCGTGTTGCTGGAAATCCAA	Dx5 Dx2
<i>Glu-D1-1</i>	promoter	Mk-D1-1-4 <sup>c, d</sup>	ASP2	TCGTGTTGCTGGAAATCCAG	Dx3
			CP	AAATCTACGGCCACTCAAAG	
			ASP1	GCAGAAGTGTGGATGTTTGTATAT	Dx5
<i>Glu-D1-2</i>	CDS	Mk-D1-2-1 <sup>c</sup>	ASP2	GCAGAAGTGTGGATGTTTGTATAC	Dx2 Dx3
			CP	CCCAAGAAGGATAATCACTTT	
			ASP1	TAGAGGCCTCACCTTCAGCA	Dy10
			ASP2	TAGAGGCCTCACCTTCAGCG	Dy12
			CP	CCTCGTGGCTCTCACCA	

<sup>a</sup> sequence of allele-specific primers (ASP) without the FAM or VIC tails. Sequence of gene-specific common primer (CP).

<sup>b</sup> when possible

<sup>c</sup> the sequence of the reverse strand was used to design primers.

<sup>d</sup> used in Schwarz et al. (2003)

**Table 3** Occurrence of HMW-GS (proteins and corresponding *Glu-1* alleles) in 364 accessions from the INRA core-col

<i>Glu-A1</i>				<i>Glu-B1</i>				
Allele	Protein (Ax)	No of accessions	%	Allele	Proteins (Bx + By)	No of accessions	%	Allele
<i>a</i>	1	93	25.5	<i>a</i>	7	29	8.0	<i>a</i>
<i>b</i>	2*	98	26.9	<i>b</i>	7 + 8	117	32.1	<i>b</i>
<i>c</i>	null	149	40.9	<i>c</i>	7 + 9	63	17.3	<i>c</i>
<i>t</i>	21*	1	0.0	<i>d</i>	6 + 8	40	11.0	<i>d</i>
Undetermined		10	2.7	<i>e</i>	20 + 20	17	4.7	<i>f</i>
Heterozygous		13	3.6	<i>f</i>	13 + 16	11	3.0	<i>i</i>
Total		364		<i>g</i>	13 + 19	2	0.5	<i>l</i>
				<i>h</i>	14 + 15	2	0.5	Undetermined
				<i>i</i>	17 + 18	17	4.7	Heterozygous
				<i>k</i>	22	2	0.5	Total
					14 + 19	2	0.5	
				<i>p</i>	23 + 18	2	0.5	
					14	1	0.2	
					6.1 + 22	4	1.1	
					23 + 8	1	0.3	
				<i>aj</i>	8	1	0.3	
				<i>al</i>	7 <sup>OE</sup> + 8	6	1.6	
				<i>am</i>	18	1	0.3	
				Undetermined		18	4.9	
				Heterozygous		28	8.7	
				Total		364		

**Table 4** Distribution of haplotypes at *Glu-B1* in 318 lines from the INRA core collection. Markers are ordered according to their order in the gene sequence.

<i>Glu-B1-1</i>	Mk-B1-1- OE	Mk-B1-1-11	Mk-B1-1-4	Mk-B1-1-1	Mk-B1-1-9	<b>Mk-B1-1-8</b> (*)	ac
H1	CC	CC	CC	TT	GG	CC	
H2	AA	CC	CC	TT	GG	CC	
H3	CC	GG	CC	TT	GG	CC	
H4	AA	GG	CC	TT	GG	CC	
H5	AA	GG	TT	CC	GG	CC	
H6	AA	GG	CC	CC	GG	TT	
H7	AA	GG	CC	CC	GG	CC	
H8	AA	no call	CC	CC	GG	CC	
H9	AA	no call	CC	CC	AA	CC	
Undefined (**)							
							Total
<i>Glu-B1-2</i>	<b>Mk-B1-2-19</b>	Mk-B1-2-15	Mk-B1-2- 11	<b>Mk-B1-2- 14</b>	Mk-B1-2- 20	Mk-B1-2-18	ac
h1	AA	AA	CC	AA	CC	AA	
h2	AA	AA	CC	AA	TT	AA	
h3	AA	AA	TT	AA	CC	AA	
h4	AA	AA	TT	AA	TT	AA	
h5	AA	AA	CC	GG	CC	AA	
h6	AA	AA	CC	GG	TT	AA	
h7	AA	AA	TT	GG	CC	AA	
h8	AA	AA	TT	GG	TT	AA	
h9	AA	GG	CC	GG	CC	AA	
By10	AA	GG	TT	GG	CC	AA	
By11	AA	AA	CC	GG	CC	TT	
By12	AA	AA	TT	GG	CC	TT	
By13	GG	AA	CC	GG	CC	AA	
By14	GG	AA	TT	GG	CC	AA	
Undefined (**)							
							Total

(\*) In bold, markers with allele-specific primers designed on the reverse strand

(\*\*) Undefined: haplotypes with missing data or heterozygosity at one marker at least.

**Table 5.** df, F, P value, *and* percentage of the phenotypic variance explained by the SNP markers at Glu-1 associated to rheological traits. Means of each allele are compared.

Trait	Marker	df	F	<i>P-value</i>	% phenotypic variance explained	Alleles	Nb of lines	Means
Total quantity of protein per grain (mg nitrogen . grain-1)	Mk-B1-1-8	1	12.8	0.0004	1.92	CC	301	0.570
						TT	13	0.686
	Mk-B1-2-11	1	8.69	0.0035	1.46	CC	190	0.570
						TT	99	0.576
Strength (W, 10 <sup>-4</sup> J)	Mk-B1-1-1	1	16.42	6.5e-05	3.31	TT	230	204.13
						CC	83	141.08
	Mk-B1-2-11	1	58.89	0	11.21	CC	190	154.61
						TT	99	250.26
Tenacity (P, mm of water)	Mk-B1-2-11	1	35.49	7.6e-09	6.41	CC	190	51.29
						TT	99	68.78
Tenacity to extensibility ratio (P/L)	Mk-B1-2-11	1	10.78	0.0012	3.22	CC	190	0.468
						TT	99	0.633
	Mk-B1-2-14	1	10.6	0.0012	2.91	GG	237	0.511
						AA	77	0.578

The general linear model comprised the main effect marker, five ancestor groups, grain hardness and grain protein content as covariates.

**Table 6.** Df, F, P value, and percentage of the phenotypic variance explained by haplotypes deduced from S associated to rheological traits traits. Means of each allele are compared.

Trait	Haplotypes	df	F	P-value	% phenotypic variance explained	Haplotypes		N li
						code	Composition <sup>(4)</sup>	
Total quantity of protein per grain (mg nitrogen . grain-1)	at <i>Glu-B1-1</i> <sup>(1)</sup>	2	7.19	0.0009	2.14	3	CC-TT	7
						2	CC-CC	6
						1	TT-CC	2
	at <i>Glu-B1</i> <sup>(3)</sup>	6	4.95	8.04e-05	4.90	3/3	CC-TT/CC-GG	4
						2/3	CC-CC/CC-GG	4
						2/4	CC-CC/TT-GG	1
						1/3	TT-CC/CC-GG	7
						1/4	TT-CC/TT-GG	5
						1/2	TT-CC/TT-AA	1
						1/1	TT-CC/CC-AA	5
Strength (W, 10 <sup>-4</sup> J)	at <i>Glu-B1-2</i> <sup>(2)</sup>	3	18.48	1.0e-10	10.74	4	TT-GG	7
						2	TT-AA	1
						1	CC-AA	5
						3	CC-GG	1
	at <i>Glu-B1-1</i> <sup>(1)</sup>	2	8.64	0.0002	3.48	1	TT-CC	2
						3	CC-TT	1
						2	CC-CC	6
	at <i>Glu-B1</i> <sup>(3)</sup>	6	11.35	0	13.07	1/4	TT-CC/TT-GG	5
						1/2	TT-CC/TT-AA	1
						2/4	CC-CC/TT-GG	1
						1/3	TT-CC/CC-GG	7
						1/1	TT-CC/CC-AA	5
						3/3	CC-TT/CC-GG	4
						2/3	CC-CC/CC-GG	4
Tenacity (P, mm of water)	at <i>Glu-B1-2</i> <sup>(2)</sup>	3	13.51	3.2e-08	7.25	2	TT-AA	7
						4	TT-GG	1
						1	CC-AA	5
						3	CC-GG	1
	at <i>Glu-B1</i> <sup>(3)</sup>	6	6.36	2.88e-06	7.08	1/2	TT-CC/TT-AA	1
						1/4	TT-CC/TT-GG	5
						2/4	CC-CC/TT-GG	1
						1/1	TT-CC/CC-AA	5
						1/3	TT-CC/CC-GG	7
						3/3	CC-TT/CC-GG	4



						2/3	CC-CC/CC-GG	4
						2	TT-AA	3
	at <i>Glu-B1-2</i> <sup>(2)</sup>	3	7.25	0.0001	6.39	4	TT-GG	7
						1	CC-AA	5
						3	CC-GG	1
Tenacity to extensibility ratio (P/L)						1/2	TT-CC/TT-AA	3
						2/4	CC-CC/TT-GG	3
						1/4	TT-CC/TT-GG	5
	at <i>Glu-B1</i> <sup>(3)</sup>	6	3.67	0.0016	6.65	1/1	TT-CC/CC-AA	5
						2/3	CC-CC/CC-GG	4
						3/3	CC-TT/CC-GG	7
						1/3	TT-CC/CC-GG	7

The general linear model comprised the main effect haplotype, five ancestor groups, grain hardness and grain covariates. All haplotypes with a frequency<0.025 were discarded. Significant differences were judged at a=

(1) *Glu-B1*- 1 haplotypes generated with alleles at Mk-B1-1-1 and Mk-B1-1-8.

(2) *Glu-B1*- 2 haplotypes generated with markers Mk-B1-2-11 and Mk-B1-2-14

(3) Global haplotype at *Glu-B1* i.e. combination of haplotypes at *Glu-B1*- 1 /Glu-B1-2.

(4) The haplotype given are the combination of alleles at Mk-B1-1-1 and -8 at *Glu-B1*- 1 followed by the al and -14 at Glu-B1-2.

(5) Different letters after the mean values indicate significant difference at  $\alpha=5\%$ . P-values are adjusted with method.

**Table 7.** Df, F, P value, and percentage of the phenotypic variance explained by SDS-PAGE profiles at Glu rheological traits. Means of each allele are compared.

Trait	PAGE allele	df	F	<i>P-value</i>	% phenotypic variance explained	HMW-GS defined by SDS-PAGE	Nb of lines	Mea
Total quantity of protein per grain (mg nitrogen . grain-1)	PAGE_Bx <sup>(1)</sup>	4	3.79	0.0051	2.37	Bx13	13	0.6
						Bx20	17	0.6
						Bx6	40	0.5
						Bx7	209	0.5
						Bx17	17	0.5
	PAGE_By <sup>(2)</sup>	5	3.44	0.0049	2.65	By16	11	0.6
						By20	17	0.5
						By9	63	0.5
						ByNull	30	0.5
						By8	164	0.5
						By18	20	0.5
	PAGE_BxBy <sub>(3)</sub>	6	3.1	0.0059	2.89	Bx13 + By16	11	0.6
						Bx20 + By20	17	0.5
						Bx7 + By9	63	0.5
						Bx6 + By8	39	0.5
						Bx7 + ByNull	29	0.5
						Bx17 + By18	17	0.5
						Bx7 + By8	117	0.5
Strength (W, 10–4 J)	PAGE_Bx <sup>(1)</sup>	4	4.05	0.0033	3.58	Bx17	17	238
						Bx7	209	199
						Bx13	13	158
						Bx6	40	141
						Bx20	17	107
	PAGE_By <sup>(2)</sup>	5	5.41	8.99E-05	5.52	By9	63	239
						By18	20	219
						By8	164	182
						ByNull	30	145
						By16	11	135
						By20	17	107
	PAGE_BxBy <sub>(3)</sub>	6	5.13	5.07e-05	6.58	Bx7 + By9	63	239
						Bx17 + By18	17	238
						Bx7 + By8	117	191
						Bx7 + ByNull	29	143

						Bx6 + By8	39	139
						Bx13 + By16	11	135
						Bx20 + By20	17	107
						Bx17 + By18	17	67.
						Bx7 + By9	63	63.
						Bx7 + By8	117	58.
Tenacity (P, mm of water)	PAGE_BxBy (3)	6	2.89	0.009	3.42	Bx6 + By8	39	49.
						Bx13 + By16	11	46.
						Bx7 + ByNull	29	45.
						Bx20 + By20	17	44.

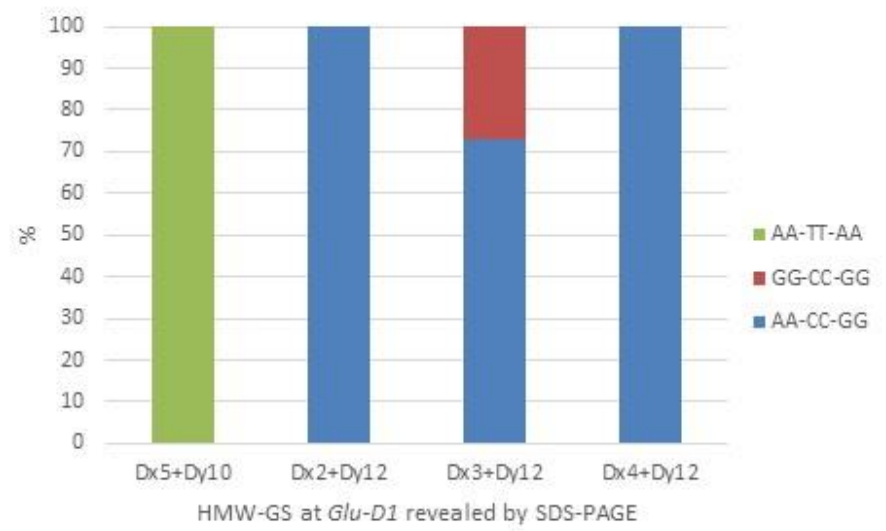
The general linear model comprised the main effect SDS-PAGE, five ancestor groups, grain hardness and grain content as covariates. All groups with a frequency < 0.025 were discarded. Significant differences were judged by Tukey's test.

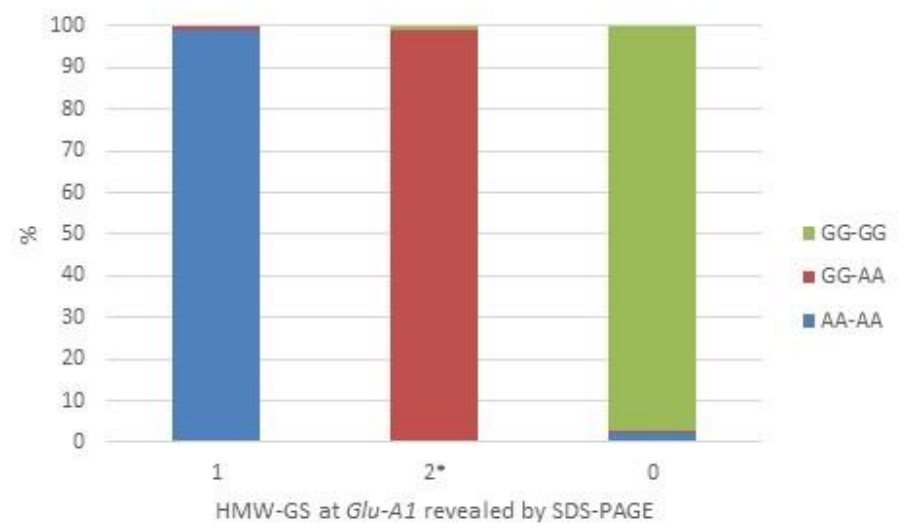
(1) PAGE\_Bx is the HMW-GS coded by *Glu-B1*- 1 revealed by the SDS\_PAGE method.

(2) PAGE\_By is the HMW\_GS coded by *Glu-B1*- 2 revealed by the SDS\_PAGE method.

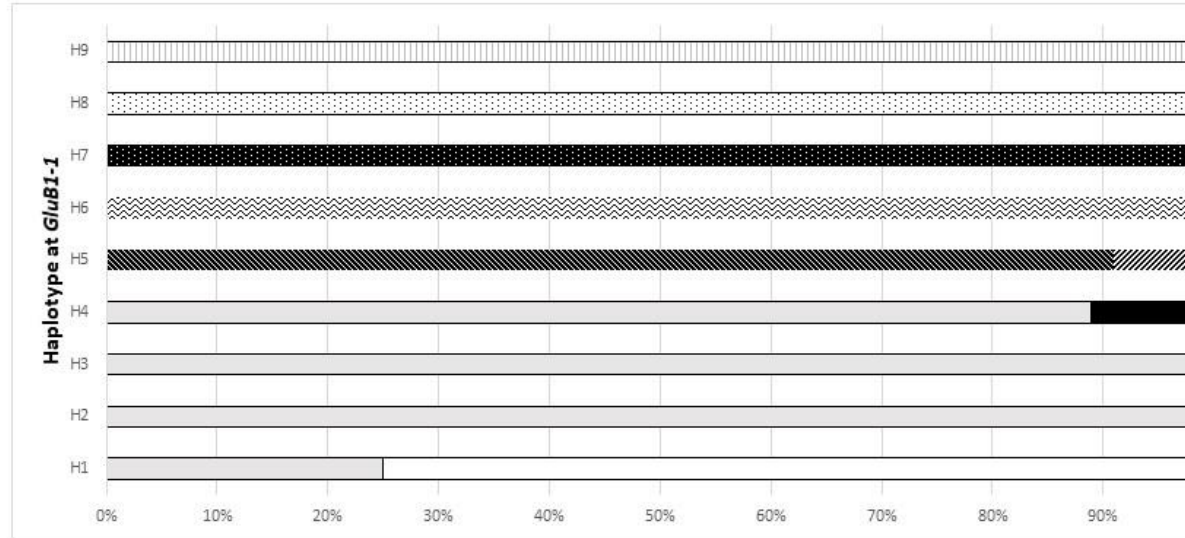
(3) PAGE\_BxBy is the combination of the Bx and By subunits.

(5) Different letters after the mean values indicate significant difference at  $\alpha=5\%$ . P-values are adjusted with the Bonferroni method.





A



B

