Theor Appl Genet (2006) 112: 738–743
DOI 10.1007/s00122-005-0178-x

ORIGINAL PAPER

Catherine Ravel · Sébastien Praud · Alain Murigneux
Laurent Linossier · Mireille Dardevet
François Balfourier · Philippe Dufour
Dominique Brunel · Gilles Charmet

Identification of *Glu-B1-1* as a candidate gene for the quantity of high-molecular-weight glutenin in bread wheat (*Triticum aestivum* L.) by means of an association study

Received: 23 June 2005 / Accepted: 30 November 2005 / Published online: 16 December 2005
© Springer-Verlag 2005

Abstract A previous study in wheat (*Triticum aestivum* L.) identified two candidate genes controlling a quantitative trait locus (QTL) for high-molecular-weight glutenin subunit (HMW-GS) GluBx. These candidates were *Glu-B1-1*, the structural gene coding for Glu1Bx, and the B homoeologous gene coding for SPA (*spa-B*), a seed storage protein activator. The goal of this study was to identify the best candidate gene for this QTL. Single nucleotide polymorphisms (SNPs) are an abundant source of DNA polymorphisms that have been successfully used to identify loci associated with particular phenotypes. As no linkage disequilibrium was detected between *Glu-B1-1* and spa-B, we performed an association study to identify the individual gene responsible for the QTL. Six SNPs, three located in *Glu-B1-1* and three in *spa-B*, were genotyped by mass spectrometry in a collection of 113 bread wheat lines. These lines were also evaluated for protein content as well as the total quantity of HMW-GSs and of each HMW-GS in seed samples from two harvest years. Significant associations were detected only between *Glu-B1-1* polymorphism and most of the traits evaluated. *Spa-B* was unambiguously discarded as a candidate. To our knowledge, this is the first report on an association study that was successfully used to discriminate between two candidate genes.

Communicated by R. Bernardo

C. Ravel (✉) · M. Dardevet · F. Balfourier · G. Charmet
Amélioration et Santé des Plantes (UMR 1095),
63100 Clermont-Ferrand, France
E-mail: ravel@clermont.inra.fr
Tel.: +33-473-624310
Fax: +33-473-624453

S. Praud · A. Murigneux · P. Dufour
Biogemma, Les Cézeaux, 63170 Aubière, France

L. Linossier
ULICE, Zac des portes de Riom, 63200 Riom, France

D. Brunel
INRA-Centre National de Génotypage, CP 5708,
91057 Evry, France

Introduction

In cereal grains, nitrogen is stored in proteins. The storage proteins, also called prolamins, are accumulated in the endosperm. The wheat prolamin gene family is divided into two different groups: gliadins and glutenins, which display a high level of allelic variability. Glutenins are divided into low-molecular-weight (LMW) and high-molecular-weight (HMW) subunits (Payne 1987). In hexaploid wheat, HMW-glutenin subunits (GSs) are encoded by three loci, *Glu-A1*, *Glu-B1* and *Glu-D1*, located on the long arms of homeologous group-1 chromosomes (Payne 1987). Each locus comprises two tightly linked genes: *Glu-D1-1* and *Glu-D1-2*, which encode the *x*- and *y*-subunits, respectively. Gliadins and glutenins are important for bread-making quality since they are involved in the formation of the gluten network (Branlard and Dardevet 1985; Metakovsky et al. 1997; Branlard et al. 2001).

During cereal endosperm development, the expression of seed storage protein genes is controlled by a network of regulatory proteins that bind specifically to conserved DNA elements called *cis*-elements, and modulate their transcription. In plants, one of the first described (and probably one of the best studied) transcription factors involved in this regulation is the bZIP protein from maize (*Zea mays*) OPAQUE2 (O2). Mutations in either the O2 gene or in its DNA binding sites within the zein promoters result in a dramatic reduction in the level of zein transcription (Schmidt et al. 1992). In bread wheat, an O2-like factor, called storage protein activator (SPA), has been identified (Albani et al. 1997). SPA appeared to trans-activate the expression of a gene coding for LMW-GS (Albani et al. 1997).

Recently, Guillaumie et al. (2004) reported that the B homoeologous gene coding for SPA (called *spa-B*) and *Gu-B1-1*, which is the structural gene coding for the HMW-GS GluBx, are located on the 1BL chromosome at a distance of 1.3 cM from each other, both within the confidence interval of a quantitative trait locus (QTL) for the quantity of GluBx. These two genes were thus considered as putative candidate genes for this phenotypic trait. Separating the respective effects of these two candidates using classical linkage studies can be difficult as a large genetic population is needed.

Association studies based on correlations between alleles at different sites or linkage disequilibrium (LD) can provide high resolution for the identification of genes that contribute to phenotypic variation in natural populations. LD patterns play a key role in the statistical power of association studies. For instance, in maize, LD was generally observed to decline rapidly with an increase in physical distance (Tenaillon et al. 2001; Remington et al. 2001). This provides a favourable background for association studies which have been used successfully in this species (Thornberry et al. 2001; Remington et al. 2001). In a previous report, we reported several single nucleotide polymorphisms (SNPs) along the sequences of both *Glu-B1-1* and *spa-B* (Ravel et al. 2004). We also showed that there is no LD between these two genes.

The goal of this study was to determine whether *Glu-B1-1* or *spa-B* is the most appropriate candidate for a QTL for the quantity of GluBx. The absence of LD observed between these two genes enabled the use of an association study to approach the question. This required genotyping of some SNPs present in the two closely located genes in a collection of bread wheat lines that we also evaluated for protein composition.

Materials and methods

Plant materials

In this study, 113 genotypes of *Triticum aestivum* were genotyped and evaluated for protein composition (list of genotypes available on request). To avoid misinterpretation of results due to poor yield, which could lead to biased protein composition, the genotypes used in field trials were selected for their adaptation to our environmental conditions. The accessions we studied originated in 15 different countries (Argentina, Australia, Canada, China, France, Germany, Great Britain, Hungary, Mexico, Netherlands, New Zealand, Sweden, Switzerland, USA and former Yugoslavia). The origin of one line, Palio, is unknown. These lines were also selected to represent diverse status (landraces, old or recent cultivars). Their date of release ranged from 1842 for Noé to 2000. In addition, data concerning their neutral polymorphism based on a set of 42 microsatellite loci (one per chromosome arm) were available (Roussel et al. 2004, 2005).

DNA extraction and SNP genotyping

All the seeds used for DNA extraction came from ears that were self-pollinated to avoid possible cross-pollination. Leaves were harvested from a pool of five to seven 3-week old seedlings per accession and bulk genomic DNA was extracted from fresh leaves using the Sigma GenElute Plant Genomic DNA Kit (G2N-350).

Glu-B1-1 and *spa-B* showed a high level of nucleotide polymorphism (Ravel et al. 2004). For genotyping, we selected non-synonymous SNPs (leading to an amino acid change) or SNPs located in putative regulatory boxes (*cis*-motifs). SNPs were genotyped by mass spectrometry using the GOOD assay (Sauer et al. 2000). This method was performed with the PCR-products obtained with locus-specific primers (Table 1) according to PCR conditions described in Ravel et al. (2004).

Field trials

Seeds were sown in a nursery design (three rows of 1 m) at the INRA plant improvement station in Clermont-Ferrand (France) in 2001 and 2002. The traits measured were protein content, predicted by near-infrared spectrum (NIRS), the quantity of HMW-GS and protein fractions for each HMW-GS quantified using capillary electrophoresis. Extraction and separation procedures were adapted from those described by Carceller and Aussenac (1999). The results were expressed as proportions of dry weight.

Association study

To avoid spurious associations, the genetic structure among the lines studied in the field trial was inferred from marker data at the 42 microsatellite loci using the admixture option in STRUCTURE software (Pritchard et al. 2000). Five independent replicates of 300,000 Markov Chain iterations were used for each parameter set tested.

The model used to test the association between polymorphisms and phenotypic traits was a combined analysis across years using the following linear model:

$$Y = \mu + Y_i + \sum G_j + P_k + Y \cdot P_{ik} + \varepsilon_{ikl}$$

Table 1 PCR primers used to product specific amplicons for genotyping

Gene fragment	Forward primer	Reverse primer
Glu-B1-1[a]	cctatgttaattttagacatgactgg	tactgccgcaaagaggaccagg
Glu-B1-1[b]	accgagcaccccaatctaca	tgttccgactgctgcgaagt
Spa-B	ggcatctccgtcctctcttc	tccaacttccgtctgagcat

[a]PCR primers to amplify a 750 bp length region upstream of the start codon
[b]PCR primers to amplify a 1,200 bp length region comprising 100 bp upstream of the start codon and 800 pb in the coding sequence

μ is the grand mean, Y the year main effect, G_j the contribution of the jth ancestor group to a given line (regressor), P_k the main effect of polymorphism (P being either a single SNP or a combined haplotype), $Y{\cdot}P$ the year × polymorphism combination and ε the residual.

This model made it possible to study the effect of each polymorphism on the genotypic part of the phenotypic variation. Two kinds of analysis were carried out: the first by considering each individual SNP as a factor (one degree of freedom), the second by every SNP from a given gene combining in haplotypes. These analyses were also performed using the GLM procedure of SAS. The LSMEAN procedure of SAS was used to estimate the least-square-adjusted mean (LSM) for each polymorphism. For haplotypes, LSMs were compared using the PDIFF option that gives probability values for the null hypothesis of no difference between haplotype means.

Results

SNP genotyped

We genotyped six SNPs (three SNPs per candidate gene). For *Glu-B1-1*, the first SNP genotyped (Glu-411) was located 411 bases upstream from the start codon and corresponded to a guanine (G)–Adenine (A) change. The second, Glu-4, was located four bases upstream from the start codon in a region similar to one which is known to be highly conserved in eukaryotes (Kozak 1987) and has been identified in genes coding for HMW-GS (Anderson and Green 1989). It led to a Cytosine (C)–Thymine (T) change. The last SNP, Glu98, was located 98 bases downstream from the ATG codon and corresponded to a G–A change. This non-synonymous mutation led to an arginine–histidine change. Three non-synonymous mutations detected in *spa-B* were also genotyped. The first one changed a TCA coding for a serine to a TGA leading to a stop codon. This mutation, called SPA20, was located 20 bases downstream from the initiator codon. The second mutation (SPA74) changed a CTG codon (leucine) to CCG (proline), while the last one, SPA306, changed GAG (glutamic acid) to GAC (asparagine). The physical distance between SPA20 and SPA74 was 54 bp, and between SPA74 and SPA306, it was 232 bp.

Together, the mutations genotyped in *Glu-B1-1* defined three haplotypes (GTA, ACG, GCG), the most frequent haplotype being GTA, which was found in 71 lines out of 113. Five haplotypes were detected for *spa-B*. The most frequent was the haplotype GCG, which characterized 59 lines, followed by the haplotypes CCG (29 lines) and CTC (4 lines). The two other haplotypes appeared to be extremely rare and were found in a single line only. Due to missing data, we were unable to assign 19 lines to one of these haplotypes. As expected, from this data we confirmed that LD between the two candidates, estimated as r^2, was around 0 for any possible comparison. This confirmed the almost complete decay of LD in this region as we previously reported (Ravel et al. 2004).

Population structure

The highest log likelihood for inferred population structure from SSR data was observed for five ancestor groups. The structure of the 113 lines studied could thus be explained by different levels of admixture from five ancestor groups. If each line is assigned to the group with the largest admixure contribution, 42 lines belong to group 1, 31 lines to group 2, 9 lines to group 3, 16 lines to group 4 and, 15 lines to group 5. In 71 of the 113 lines studied (63%), more than 70% of their origin came from a single group; in 30 lines, the contribution of the main ancestor group ranged from 50 to 70%, in the others, the contribution of any single ancestor group was less than 50%. Information found at http://www.genbank.vurv.cz showed that group 1 contained the most recent cultivars (the mean date of release was 1987) from Western Europe while the second group contained lines from the same origin but released earlier (mean 1971). These two mean dates of release were significantly different ($P < 0.01$). Eight of the nine lines belonging to group 3 were closely related to the French cultivar Etoile de Choisy but no data on the genealogy of this line are available. Group 4 was mostly characterized by lines that originated directly from eastern Eurasia or were related to lines from this geographic region while group 5 was related to germplasm from America. Interestingly, three lines out of the four harbouring the rare haplotype CTC for *spa-B* belonged to this group. The five groups observed were explained either by different origins or different dates of release (groups 1 and 2) probably due to different selection objectives. The structure analysis predicted 42 lines to be substantially admixted (i.e. less than 70% genome composition inherited from a single ancestor population), which suggests a low level of differentiation among the breeding material studied.

Association analysis

For *Glu-B1-1*, the main year effect was generally highly significant ($P < 0.0001$) and no significant year × genotype interaction was detected (the main year effect should remain the same, whatever the polymorphism tested subsequently). A significant effect of every SNP of *Glu-B1-1* was observed for protein content, and for quantities of HMW-GS, GluBx and GluBy (Table 2). The GluAx, GluDx and GluDy fractions were not affected by any of these polymorphisms. As expected, the strongest effect of these mutations was observed on the quantity of GluBx, the HMW-GS coded by *Glu-B1-1*. For this trait, each SNP explained about 30% of total variation. These results were confirmed by analysis of

Table 2 *P* value of the effect of nucleotide polymorphisms in *Glu-B1-1* on protein content, quantity of HMW-GS and each HMW-GS fraction

SNP	Protein content	HMW-GS	GluAx	GluBx	GluBy	GluDx	GluDy
Glu-411	0.05080	< 0.0001	0.6782NS	< 0.0001	0.0010	0.8496NS	0.8597NS
Glu-4	0.0005	< 0.0001	0.1648NS	< 0.0001	0.0470	0.7001NS	0.7009NS
Glu96	< 0.0001	< 0.0001	0.1530NS	< 0.0001	0.017	0.4163NS	0.3404NS
Haploglu	< 0.0001	< 0.0001	0.1961NS	< 0.0001	0.0710	0.6018NS	0.4283NS

The model comprised the main effect year, five ancestor groups used as covariates, the main effect nucleotide polymorphism and year × polymorphism interaction. The *P*-values for each SNP are given

Glu-411, Glu-4, Glu96 correspond to SNP genotyped in *Glu-B1-1*. HaploGlu corresponds to the main haplotype effect for the gene *Glu-B1-1*

HMW-GS, GluAx, GluBx, GluBy, GluDx, GluDx: quantity of glutenin, high molecular weight glutenin subunits and of each HMW-GS fraction

NS Not significant

the haplotype which had the strongest effect on the quantity of GluBx (this effect explained 37% of the total variation). Generally, no polymorphism in *spa-B* had a significant effect on the traits studied. Surprisingly, even the SNP named SPA20, which leads to a stop codon in the transcription factor, did not affect any variable.

Least-square-adjusted mean comparisons enabled the identification of favourable alleles or haplotypes (Table 3). For the Glu-411 mutation, the G allele significantly improved protein content and quantities of HMW-GS and GluBx in both years, while a significant decrease in the quantity of Glu1 By was observed. Similar results were obtained for the T and A alleles at Glu-4 and Glu98 loci respectively. In all these genotypes, a significant, albeit slight, decrease in the quantity of Glu1By was observed. Consequently, if the breeding goal is higher protein and HMW-GS content, the best haplotype of *Glu-B1-1* is 'GTA', and the worst is GCG. In this study, the cultivars characterized by subunits Bx7 or Bx17 were always characterized by the best haplotype 'GTA'.

Discussion

The method used in this study to genotype SNPs in hexaploid wheat proved to be efficient. The rapid breakdown of LD was confirmed in the region studied which contained a QTL for the quantity of GluBx with two candidate genes (Guillaumie et al. 2004). This was a priori a favourable situation for an association study whose aim was to identify the best candidate.

In this study, the structure of the 113 lines genotyped was based on five ancestor groups. However, our results showed a low level of differentiation among the material studied. This may be explained by the domestication and breeding history of bread wheat or the recent use of differentiated germplasms in breeding programmes since, to develop new breeding lines, wheat breeders usually crossed lines from diverse origins with complementary features (e.g. yield, quality and disease tolerance) in complex pedigrees. These crosses probably maintain a high level of admixture.

For association testing, all the SNPs genotyped were predicted to have a potentially functional influence on the candidate genes, *Glu-B1-1* and *spa-B*. All the lines studied were evaluated in two successive years for total protein content, total quantity of HMW-GS and quantity of each HMW-GS. These phenotypic traits are known to show genotype × environment interactions (Oury et al. 2003). This was taken into account in the statistical model to obtain reliable estimates of each SNP or haplotype factors. Our results indicated highly significant associations between *Glu-B1-1* polymorphisms

Table 3 Performance (proportion of dry weight indicated by LS mean values) of each nucleotide polymorphism of *Glu-B1-1* and haplotype on traits significantly affected

SNP	Nucleotide	Protein content	HPM-GS	GluBx	GluBy
Glu-411	A	13.60 b	1.23 b	0.18 b	0.20 a
	G	14.04 a	1.48 a	0.48 a	0.14 b
Glu-4	C	13.55 b	1.30 b	0.30 b	0.17 a
	T	14.18 a	1.50 a	0.49 a	0.15 b
Glu96	A	14.20 a	1.51 a	0.50 a	0.14 b
	G	13.20 b	1.25 b	0.26 b	0.18 a
HaploGlu	GTA	14.21 a	1.51 a	0.50 a	0.15 b
	ACG	13.47 b	1.19 b	0.19 c	0.19 a
	GCG	12.71 c	1.31 b	0.38 b	0.16 ab

Glu-411, Glu-4, Glu96 correspond to SNP genotyped in *Glu-B1-1* while SPA20, SPA74 and SPA306 correspond to SNPs genotyped in *spa-B*. HaploGlu and HaploSPA correspond to the main effect haplotype for the gene *Glu-B1-1* and *spa-B*, respectively

HMW-G, GluAx, GluBx, GluBy, GluDx, GluDx: quantity of high molecular weight glutenin, of each HMW-G fraction

LS means, compared for 1 year, followed by different letters (a and b) were significantly different at 5% as shown by a pairwise *T* test

and protein content, the quantity of HMW-GS and that of GluBx. In fact, despite a slightly significant decrease in the quantity of GluBy, the increase in the amount of GluBx led to an increase in the overall quantity of HMW-GS and in total protein content. These results suggest that *Glu-B1-1* is a relevant candidate for the QTL. As expected, due to the high level of LD observed within this gene (Ravel et al. 2004), all three SNPs genotyped significantly influenced phenotypic traits, and it is consequently difficult to decide which one is really functional. Two of these polymorphisms are located upstream from the start codon. According to the results obtained for the promoter of *Glu-D1-1* (Norre et al. 2002), the polymorphism called Glu-411 is located in a *cis*-element which may influence transcription. The mutation observed may thus modify the transcription of the gene leading to modification in the quantity of GluBx. This mutation is tightly associated with the T allele at locus Glu-4 resulting in a sequence similar to the Kozak sequence which is known to be involved in the translation (Kozak 1987). This underlines the importance of the perigenic region for the expression of genes. The haplotype GTA linked with higher protein and HMW-GS content corresponds to the HMW-GS Bx7 or Bx17. An over-expressing allelic form of the Bx7 subunit has been described and reported to have a major impact on dough strength (Butow et al. 2003). The source of this allelic form was a Uruguayan landrace, Americano 44D (Butow et al. 2004). Unfortunately, we did not type the material studied to see whether it contains this allelic form or not. However, the pedigrees of each line studied revealed that only one line descends from Americano 44D. This suggests that the over-expressing allelic form of the Bx7 may be rare in the sample studied.

Despite an SNP leading to a stop codon, no association was found between quantitative traits studied and polymorphisms along *spa-B*. This result eliminates this gene as a candidate for the QTL. SPA function has already been partly elucidated (Albani et al. 1997) and it has been demonstrated that this bZip protein can activate transcription of an LMW-GS gene *via* a conserved *cis*-element also found in promoters of other storage protein genes such as HMW-GS genes. Moreover, Norre et al. (2002) and Guillaumie et al. (2004) suggested that SPA might be involved in the activation of HMW-GS genes. The absence of an association between the quantity of HMW-GS and the polymorphism of the gene coding for SPA was thus unexpected. This may imply either that SPA does not activate *Glu-B1-1* or that the SNP, leading to a stop codon, does not dramatically alter SPA function. This SNP is close to the putative start codon defined by Albani et al. (1997). This creates a short upstream open reading frame (uORF) followed by another in-frame start codon (Guillaumie et al. 2004) that could give rise to a functional protein. The presence of several short uORFs is a frequent feature of the bZIP transcription factors related to the maize transcriptional activator O2. For example, it was also found in the barley transcriptional activator of the seed storage protein, *Blz2* (Oñate et al. 1999). Lohmer et al. (1993) showed that the presence of uORFs reduced the translation of O2. This suggests that an allelic form of *spa-B* could lead to a lower level of SPA which could reduce the level of expression of its targets. However, hexaploid wheat contains three homoeologous genes coding for SPA (Guillaumie et al. 2004). These copies may be expressed and this could counterbalance the effect of the polymorphism leading to a stop codon.

To conclude, this study demonstrates the utility of genetic association in bread wheat and its usefulness in discriminating between possible candidates. Such discrimination is also possible by fine mapping using a large mapping population, but about 4,000–6,000 recombinant lines would be needed to obtain the same 1 cM accuracy of QTL position that was obtained here at a much lower cost. Our results showed that one haplotype of *Glu-B1-1* improved protein content through an increase in the quantity of GluBx. This should provide breeders with a tool for early indirect selection of high protein genotypes. Moreover, to our knowledge, this is the first time in plants that association genetics has been used to successfully choose between two candidate genes for a QTL. However, other genes closely linked with *Glu-B1* which remain to be discovered, cannot be totally discarded from the list of candidates for the QTL.

Acknowledgements The authors thank D. Goodfellow and F. Vear for improving upon their English, B. Debote, B. Duperrier and their staff for high-quality experimentation, S. Giancola and A. Bérard for genotyping. This research was funded by the French federative genomic programme Génoplante.

References

Albani D, Hammond-Kosack MCU, Smith C, Conlan S, Colot V, Holdsworth M, Bevan M (1997) The wheat transcriptional activator SPA: a seed-specific bZIP protein that recognizes the GCN4-like motif in the bifactorial endosperm box of prolamin genes. Plant Cell 9:171–184

Anderson OD, Grenne FC (1989) The characterization and comparative analysis of high-molecular-weight glutenin genes from genomes A and B of a hexaploid bread wheat. Theor Appl Biol 77:689–700

Branlard G, Dardevet M (1985) Diversity of grain protein and bread wheat quality. II. Correlation between high molecular weigh 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

Butow BJ, Ma W, Gale KR, Cornish GB, Rampling L, Larroque O, Morell MK, Békès F (2003) Molecular discrimination of Bx7 alleles demonstrated that a highly expressed high-molecular-weight glutenin allele has a major impact on wheat flour dough strength. Theor Appl Genet 107:1524–1532

Butow BJ, Gale KR, Ikea J, Juhasz A, Bedö Z, Tamas L, Gianibelli MC (2004) Dissemination of the highly expressed Bx7 glutenin subunit (*Glu-B1a1* allele) in wheat as revealed by novel PCR markers and RP-HPLC. Theor Appl Genet 109:1525–1535

Carceller JL, Aussenac T (1999) Accumulation and changes in molecular size distribution of polymeric proteins in developing grains of hexaploid wheats: role of the dessication phase. Aust J Plant Physiol 26:301–310

Guillaumie S, Charmet G, Linossier L, Torney V, Robert N, Ravel C (2004) Co-location between a gene encoding for the bZip factor SPA and an eQTL for a high-molecular-weight glutenin subunit in wheat (*Triticum aestivum*). Genome 47:705–713

Kozak M (1987) An analysis of 5′-noncoding sequences from 699 vertebrate messenger RNAs. Nucleic Acids Res 15:8125–8148

Lohmer S, Maddaloni M, Motto M, Salamini F, Thompson RD (1993) Translation of the mRNA of the maize transcriptional activator Opaque-2 is inhibited by upstream open reading frames present in the leader sequence. Plant cell 5:65–73

Metakovsky EV, Félix I, Branlard G (1997) Association between dough quality (W value) and certain gliadin alleles in French common wheat cultivars. J Cereal Sci 26:371–373

Norre F, Peyrot C, Garcia C, Rancé I, Drevet J, Theisen M., Gruber V (2002) Powerful effect of an atypical bifactorial endosperm box from wheat HMWG-Dx5 promoter in maize endosperm. Plant Mol Biol 50:699–712

Oñate L, Vicente-Carbajosa J, Lara P, Diaz I, Carbonero P (1999) Barley BLZ2, a seed-specific bZIP protein that interacts with BLZ1 in vivo and activates transcription from the GCN4-like motif of B-hordein promoters in barley endosperm. J Biol Chem 274:9175–8182

Oury FX, Bérard P, Brancourt-Hulmel M, Depatureaux C, Doussinault G, Galic N, Giraud A, Heumez E, Lecomte C, Pluchard P, Rolland B, Rousset M, Trottet M (2003) Yield and grain protein concentration in bread wheat: a review and a study of multi-annual data from a French breeding program. J Genetic Breed 57:59–68

Payne PI (1987) Genetics of wheat storage proteins and the effects of allelic variation on breadmaking quality. Annu Rev Plant Physiol 38:141–153

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

Ravel C, Praud S, Canaguier A, Dufour P, Giancola S, Balfourier F, Chalhoub B, Brunel D, Linossier L, Dardevet M, Beckert M, Rousset M, Murigneux A, Charmet G (2004) DNA sequence polymorphism and their application in bread wheat. In: Vollmann J, Grausgruber H, Ruckenbauer P (eds) Genetic variation for plant breeding. Eucarpia, Tulln, pp 177–181

Remington DL, Thornberry JM, Matsuoka Y, Wilson LM, Whitt SR, Doebley J, Kresovich S, Goodman MM, Buckler ES (2001) Structure of linkage disequilibrium and phenotypic associations in the maize genome. Proc Natl Acad Sci USA 98:11479–11484

Roussel V, Koenig J, Beckert M, Balfourier F (2004) Molecular diversity in French bread wheat accessions related to temporal trend and breeding programmes. Theor Appl Genet 108:920–930

Roussel V, Leisova L, Exbrayat F, Stehno Z, Balfourier F (2005) SSR allelic diversity changes in 480 European bread wheat varieties released from 1840 to 2000. Theor Appl Genet (in press)

Sauer S, Lechner D, Berlin K, Plancon C, Heuermann A, Lehrach H, Gut IG (2000) Full flexibility genotyping of single nucleotide polymorphisms by the GOOD assay. Nucleic Acids Res 28:E100

Schmidt RJ, Ketudat M, Aukerman M, Hoschek G (1992) OPAQUE-2 is a transcriptional activator that recognizes a specific target site in 22 kD-zein genes. Plant Cell 4:689–700

Tenaillon M, Sawkins MC, Long AD, Gaut RL, Doebley JF, Gaut BS (2001) Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. *mays* L.). Proc Natl Acad Sci USA 98:9161–9166

Thornsberry JM, Goodman MM, Doebley J, Kresovich S, Nielsen D, Buckler ES (2001) *Dwarf8* polymorphisms associate with variation in flowering time. Nat Genet 28:286–289