ORIGINAL PAPER

High level of conservation between genes coding for the GAMYB transcription factor in barley (*Hordeum vulgare* L.) and bread wheat (*Triticum aestivum* L.) collections

**Grit Haseneyer · Catherine Ravel · Mireille Dardevet · François Balfourier ·
Pierre Sourdille · Gilles Charmet · Dominique Brunel · Sascha Sauer ·
Hartwig H. Geiger · Andreas Graner · Silke Stracke**

Abstract The transcription factor GAMYB is involved in gibberellin signalling in cereal aleurone cells and in plant developmental processes. Nucleotide diversity of *HvGAMYB* and *TaGAMYB* was investigated in 155 barley (*Hordeum vulgare*) and 42 wheat (*Triticum aestivum*) accessions, respectively. Polymorphisms defined 18 haplotypes in the barley collection and 1, 7 and 3 haplotypes for the A, B, and D genomes of wheat, respectively. We found that (1) *Hv*- and *TaGAMYB* genes have identical structures. (2) Both genes show a high level of nucleotide identity (>95%) in the coding sequences and the distribution of polymorphisms is similar in both collections. At the protein level the functional domain is identical in both species. (3) *GAMYB* genes map to a syntenic position on chromosome 3. *GAMYB* genes are different in both collections with respect to the Tajima *D* statistic and linkage disequilibrium (LD). A moderate level of LD was observed in the barley collection. In wheat, LD is absolute between polymorphic sites, mostly located in the first intron, while it decays within the gene. Differences in Tajima *D* values might be due to a lower selection pressure on *HvGAMYB*, compared to its wheat orthologue. Altogether our results provide evidence that there have been only few evolutionary changes in *Hv*- and *TaGAMYB*. This confirms the close relationship between these species and also highlights the functional importance of this transcription factor.

G. Haseneyer · A. Graner · S. Stracke
Leibniz Institute of Plant Genetics and Crop Plant Research
Gatersleben (IPK), Corrensstr. 3, 06466 Gatersleben, Germany

C. Ravel (✉) · M. Dardevet · F. Balfourier · P. Sourdille ·
G. Charmet
INRA, UMR1095, 63100 Clermont-Ferrand, France
e-mail: ravel@clermont.inra.fr

D. Brunel
INRA, Centre National de Génotypage, CP5708,
91057 Evry, France

S. Sauer
Max Planck Institute for Molecular Genetics,
Ihnestr. 73, 14195 Berlin, Germany

H. H. Geiger
Institute of Plant Breeding, Seed Science and Population
Genetics, University of Hohenheim, 70599 Stuttgart, Germany

Present Address:
G. Haseneyer
Plant Breeding, Technische Universitaet Muenchen/
Centre of Life and Food Sciences Weihenstephan,
Am Hochanger 4, 85350 Freising, Germany

Present Address:
S. Stracke
Department of Crop Sciences, Quality of Plant Products,
University of Goettingen, Carl-Sprengel-Weg 1,
37075 Goettingen, Germany

Introduction

Gibberellins (GAs) are hormones which play crucial roles in plant growth and development such as seed germination, stem and hypocotyl elongation, leaf expansion, and floral

and fruit development (Hooley 1994). Some GA-regulated genes have been analysed in detail showing the presence of GA-responsive *cis*-elements in their promoters. *Trans*-acting factors able to bind to the *cis*-elements have been identified. One such factor identified in barley (*Hordeum vulgare* L.) is GAMYB, a GA-dependent MYB transcription factor (Gubler et al. 1995). Other than in barley, GAMYB homoeologs have been isolated from rice (*Oryza sativa*), wheat (*Triticum aestivum*), *Lolium temulentum* and *Arabidopsis thaliana* (Gubler et al. 1997; Gocal et al. 1999; Chen et al. 2001; Stracke et al. 2001). The Poaceae appear to have a single copy gene of GAMYB, whereas *Arabidopsis* has a small family of *GAMYB*-like genes with three copies (Stracke et al. 2001).

The physiological functions of the *GAMYB* gene and its interactions have been studied. In rice and barley, the GAMYB protein induces the expression of genes encoding hydrolytic enzymes needed for germination, such as α-amylases, proteinases and cell-wall degrading proteins, through direct binding to a highly conserved 21 bp GA-responsive element (GARE, TAACAA/GA) (Gubler and Jacobsen 1992; Gubler et al. 1997, 1999; Cercós et al. 1999). GAMYB also activates gene expression during endosperm development (Diaz et al. 2002). Moreover, the involvement of GAMYB as a *trans*-activator of GA signalling may not be restricted to the cereal aleurone. *GAMYB* genes are also involved in flower development (Gocal et al. 1999; Murray et al. 2003), seed maturation (Diaz et al. 2002), and stem elongation (Gocal et al. 2001; Chen et al. 2001). GAMYB thus plays multiple roles in the GA signalling cascade throughout plant growth. In rice, Tsuji et al. (2006) showed that GAMYB function differs in aleurone cells and flower organs. These diverse functions appear not to depend on GAMYB alone and they might result from differences in the organ-regulation of GAMYB expression (Tsuji et al. 2006) and interactions with other proteins (Diaz et al. 2002, 2005; Gubler et al. 2002; Isabel-LaMoneda et al. 2003; Washio 2003; Rubio-Somoza et al. 2006).

In animals, the DNA-binding domain characteristic of MYB proteins consists of three repeats of about 50 residues (R1, R2 and R3). Plant MYB homologs are structurally and functionally more variable. Their MYB domain usually contains two imperfect repeats (R2 and R3) characterized by three regularly spaced tryptophan residues which play a role in the folding of the hydrophobic core of the domain and are thus generally conserved (Kanei-Ishii et al. 1990; Martin and Paz-Ares 1997). Members of the MYB family of transcription factors have a highly conserved N-terminus corresponding to the MYB domain, but are very variable in the C-terminal region. In barley GAMYB, the typical R2 and R3 repeats are located in the N terminal region and are

followed by two transcriptional activation domains (see Woodger et al. 2003).

To describe and compare sequence diversity of GAMYB within and between wheat and barley, we sequenced *GAMYB* genes from two collections of barley and wheat that represent a cross section of the genetic diversity of these species. Our results show, that *HvGAMYB* and *TaGAMYB* share a high similarity on the nucleotide level. Nevertheless, there are marked differences in nucleotide and haplotype diversity between the three homoeologs in wheat and highly divergent patterns of linkage disequilibrium in wheat and barley.

Materials and methods

Plant material

In total 155 accessions were selected from the Barley Core Collection (BCC) and the Gaterslebener Genebank (HOR) (Table S1a) originating from Europe ($N = 77$), East Asia ($N = 28$), America ($N = 24$), and West Asia and North Africa ($N = 26$). This germplasm set includes two-rowed and six-rowed barleys and is part of a larger germplasm that is described in more detail in Haseneyer et al. (in preparation). Nine accessions were classified as "breeding/research material", 33 accessions as "traditional cultivars/landrace" and 113 accessions represented "advanced/improved cultivars".

A wheat collection of 42 lines from different origins (Table S1b) was chosen from the worldwide bread wheat core collection of 372 accessions (Balfourier et al. 2007) to represent the variability of this core collection as described by Ravel et al. (2006). The wheat lines used here were sampled by a strategy reported to maximise the diversity (Gouesnard et al. 2001). For each wheat line, seeds used in this work were provided by INRA-Clermont-Ferrand Genetic Resource Centre for Cereal Crops and came from a single, self-pollinated head. All plantlets for a given accession are thus considered as genetically identical.

Eight seeds from each accession were grown in the greenhouse and leaves from 2-weeks-old seedlings were harvested and pooled for genomic DNA extraction using the methods described in Stein et al. (2001) for barley and Tixier et al. (1998) for wheat. The different sample sizes chosen for barley and wheat were based on the different ploidy levels. Thus, the wheat sample is smaller than the barley sample but the number of sequences analysed is approximately identical.

Physical assignment of the *GAMYB* fragments obtained to chromosomes was done using a set of Chinese Spring aneuploid lines (Sears 1966; Endo and Gill 1996).

Gene amplification and sequencing

The oligonucleotides used for amplification and sequencing (Table S2) were designed using Primer 3 (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3.cgi) based on the reference sequences for wheat (AY615200) and barley (AY008692). As wheat is a hexaploid species, direct sequencing of genes from PCR products requires the design of locus-specific PCR primers to avoid co-amplification of the different copies (Ravel et al. 2006). PCR profiles are summarized in Table S2. For PCR from barley DNA a mix of $10 \times$ Qiagen® PCR Buffer, deionised H_2O, 5 μM of each primer, 0.25 units of *Taq* DNA Polymerase (Qiagen) and 20 ng DNA was prepared. For wheat, PCR reactions were performed in a final volume of 25 μl containing 25 ng of genomic DNA, 250 μM of each dNTP, 0.4 μM of each primer, 1 unit of *Taq* polymerase (Qiagen) and $1 \times$ *Taq* polymerase buffer.

Sequencing of the barley and wheat amplicons was done according to the ABI PRISM® BigDye™ Terminator Cycle Sequencing protocol using AmpliTaq®. Both strands of each fragment were re-sequenced on the ABI3730xl DNA Analyzer system (Applied Biosystems). As the *TaGAMYB-D1* fragment (Table S2) was approximately 2,000 bp long, an internal primer (5′-GAGCTGGATGATGAGCCTCT-3′) was used to obtain the complete sequence.

We used the Sequencher™ program Version 4.5 (Gene Codes Cooperation) and the Staden package (Staden et al. 2000) for sequence alignment and editing in barley and wheat, respectively. Barley and wheat sequences were aligned by ClustalW (Thompson et al. 1994). The alignment was manually corrected using BioEdit software (Hall 1999). All positions given in the text correspond to the position in the haplotype sequence alignment relative to the start codon (Fig. S1).

Diversity analysis

Haplotype diversity ($\hat{h}$), nucleotide diversity estimated as Pi (π; Nei 1987), the mean pair-wise differences, and Theta (θ; Watterson 1975), the number of segregating sites were computed using the DNA-SP version 4.10 software package (Rozas et al. 2003). This software does not take into account the alignment gaps that can cause an underestimation of π and θ values. For this, insertion–deletion events (indels) were treated as single sites. To analyse the pattern of diversity we applied the sliding window method with a window size of 100 bp and a step size of 25 bp. The π and θ values were compared by the Tajima's *D* test (Tajima 1989) implemented in DNA-SP.

The π values obtained for each species were compared by resampling. For both species, 100 resampled files were generated using the bootstrap method implemented in seqboot (Phylip 3.6; Felsenstein 2005). Each file was used to calculate its π value. Such sampling allowed us to estimate the average and the standard deviation of the empirical distribution for π in each species. The averages obtained were compared using the SAS TTEST procedure (SAS Institute, Inc., USA).

Linkage disequilibrium was estimated using squared allele–frequency correlations, R^2 (Hill and Robertson 1968), for pairs of polymorphic sites with minor allele frequencies over 5%. The Fisher's exact test was used to determine whether the associations between polymorphisms were significant. The analyses were performed by applying TASSEL version 9.3.1 (Buckler et al. 2006).

Genetic mapping

HvGAMYB was genetically mapped in the Steptoe × Morex mapping population developed by Kleinhofs et al. (1993). Positions were determined on an updated Steptoe × Morex map (Stein et al. 2007; http://pgrc.ipk-gatersleben.de/transcript_map/momavis.php). We designed a cleaved amplified polymorphic sequence (CAPS) marker that requires the use of the restriction enzyme *Msp*I based on the single nucleotide polymorphism (SNP) at position 2542 (Fig. S1). In wheat, genetic mapping was done using the International Triticeae Mapping Initiative (ITMI) segregating population. We genotyped the SNP at position 2037 (Fig. S1) using primer-extension chemistry to generate allele-specific products analysed by mass spectrometry as explained in Balfourier et al. (2006).

Linkage was analysed using Mapmaker/exp 3.06 (Lander et al. 1987) in wheat and Joinmap 3.0 (Kyazma, The Netherlands) in barley. The Kosambi mapping function was applied to transform recombination frequencies into map distances given in centiMorgans (cM).

Phylogenetic analysis

Phylogenetic analyses were conducted using MEGA version 2.1 (Kumar et al. 2001). A pair-wise Kimura (1980) distance matrix among the haplotype sequences of *GAMYB* was calculated and represented by a neighbour-joining tree. The reliability and goodness of fit of dendrograms obtained from sequences were tested through resampling (1,000 bootstrap samples).

Results

Gene structure

GAMYB genes were amplified from 155 barley and 42 wheat lines from two germplasm collections that are
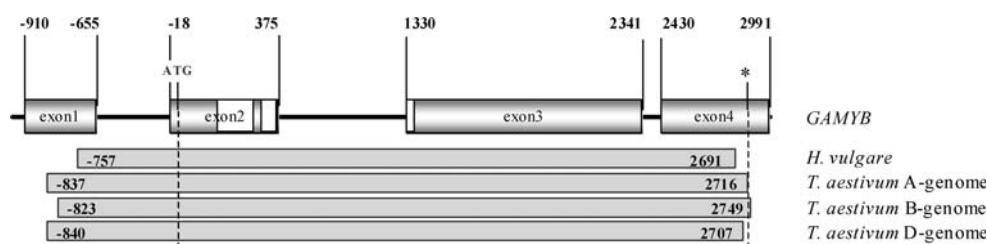
Fig. 1 Structure and fragments of the *GAMYB* gene that were sequenced in *Hordeum vulgare* and from each genome of *Triticum aestivum*. *Numbers* indicate nucleotide position relative to the start codon (ATG) which is 1; positions left to the initiation site are counted as negative starting with –1. The *asterisk* shows the stop codon (alignment position 2722–2724 bp). The start and end points of the four sequences in the alignment are shown. The *white boxes* indicate the functional MYB domain

representative of the diversity of these two species. The alignment of the different barley *GAMYB* sequences gave a total gene length of 3,332 bp including a gap of 25 bp, although sequences for exon1 and exon4 were incomplete (Fig. 1). For wheat, *TaGAMYB*-A, -B and -D correspond to A, B, and D homoeologous copies of *GAMYB*, respectively. The lengths of the consensus sequences obtained for *TaGAMYB*-A, -B and -D were 3,380, 3,395 and 3,368 bp, respectively. These three consensus sequences contained the entire coding sequence.

The structure of *TaGAMYB* and *HvGAMYB* was identical containing four exons with the coding sequence spanning from exon2 to exon4 (Fig. 1). However, the length of exons and introns differed slightly between *HvGAMYB* and *TaGAMYB*-B (Fig. S1). The alignment of sequences corresponding to the different haplotypes is provided as supplementary material (Fig. S1).

Nucleotide diversity

Within the 155 barley accessions studied, we detected 45 polymorphisms (Table S3a) including 13 indels and an average polymorphism density of one polymorphic site per 74 bp. Two (position 1546, position 2017) of the seven single nucleotide polymorphisms (SNPs) in the coding sequence caused a change in the amino acid composition. Most of the remaining polymorphisms ($N = 38$) were in the noncoding sequence.

In the 42 wheat lines, we detected 0, 23, and 2 polymorphisms in *GAMYB*-A, -B and -D, respectively. All sequence information obtained for wheat is available at http://urgi.infobiogen.fr/Gnp. In the B genome, we found 19 polymorphisms and 4 indels (Table S3b) and observed on average one polymorphism per 178 bp. As for barley, most of the polymorphims (17 out of the 23) were present in introns. Only the SNP (position 2468) in the fourth exon leads to an amino-acid change (glutamic acid to glutamine).

In barley, π and θ values were calculated using the sequence information of all 155 accessions and amounted to $\pi = 2.23 \times 10^{-3}$ and $\theta = 2.37 \times 10^{-3}$. These π and θ values were not significantly different. When classified according to geographical origin the nucleotide diversity ranged from $\pi = 1.74 \times 10^{-3}$ (East Asia) to $\pi = 2.78 \times 10^{-3}$ (America, Table S4).

In the B genome of wheat, π and θ values were 2.68×10^{-3} and 1.53×10^{-3}, respectively. A significantly positive value was found for Tajima's D test ($P < 0.05$). When classed by geographical region, the nucleotide diversity values of lines ranged from $\pi = 2.51 \times 10^{-3}$ (America) to $\pi = 3.74 \times 10^{-3}$ (West Asia, Table S4). Due to their lack of polymorphism, *TaGAMYB*-A and -D were not considered in this way. The mean values calculated for π by the bootstrap strategy were $2.36 (\pm 0.49) \times 10^{-3}$ and $2.68 (\pm 0.57) \times 10^{-3}$ for *HvGAMYB* and *TaGAMYB*-B, respectively. For θ, we obtained values of $2.51 (\pm 0.37) \times 10^{-3}$ and $1.52 (\pm 0.30) \times 10^{-3}$ for barley and the wheat B genome, respectively. The nucleotide diversity of *TaGAMYB-B* estimated by the π value was significantly higher than that of *HvGAMYB* ($P < 0.0001$). By contrast, the θ value of *TaGAMYB-B* was significantly lower than that of *HvGAMYB* ($P < 0.0001$).

The distribution of *GAMYB* polymorphisms was similar in both collections (Fig. 2). In barley and wheat, most of the polymorphisms detected were in the first two introns, while the third intron was totally conserved. Exons 1 and 2 were conserved in wheat; but showed one polymorphism in barley. The third exon contained no polymorphisms. In both species, we observed a high conservation of the functional domain containing no polymorphism in wheat and one synonymous SNP in barley.

Haplotype analysis

Within the barley germplasm set 18 haplotypes were classified (Table S3a; Fig. 3) resulting in 4 amino acid sequences. One dominant haplotype (haplotype 1) occured at a frequency of 47.7%. Haplotypes 2–11 were less frequent but each represented more than 1% of lines. Haplotypes 12–18 were singletons. Haplotype 11 was only represented by American accessions. Almost 50% of the European accessions were haplotype 1. The diversity within the barley set was a result of diverse alleles being evenly

Fig. 2 Nucleotide diversity values Pi (π, *black line*) and Theta (θ, *grey line*) in *GAMYB* using the sliding window approach (step size, 25 bp; window size, 100 bp) in *H. vulgare* (**a**) and *T. aestivum* B-genome (**b**). The structure of *GAMYB* is given between Fig. **a** and **b**; the *open boxes* show the position of the functional MYB domain
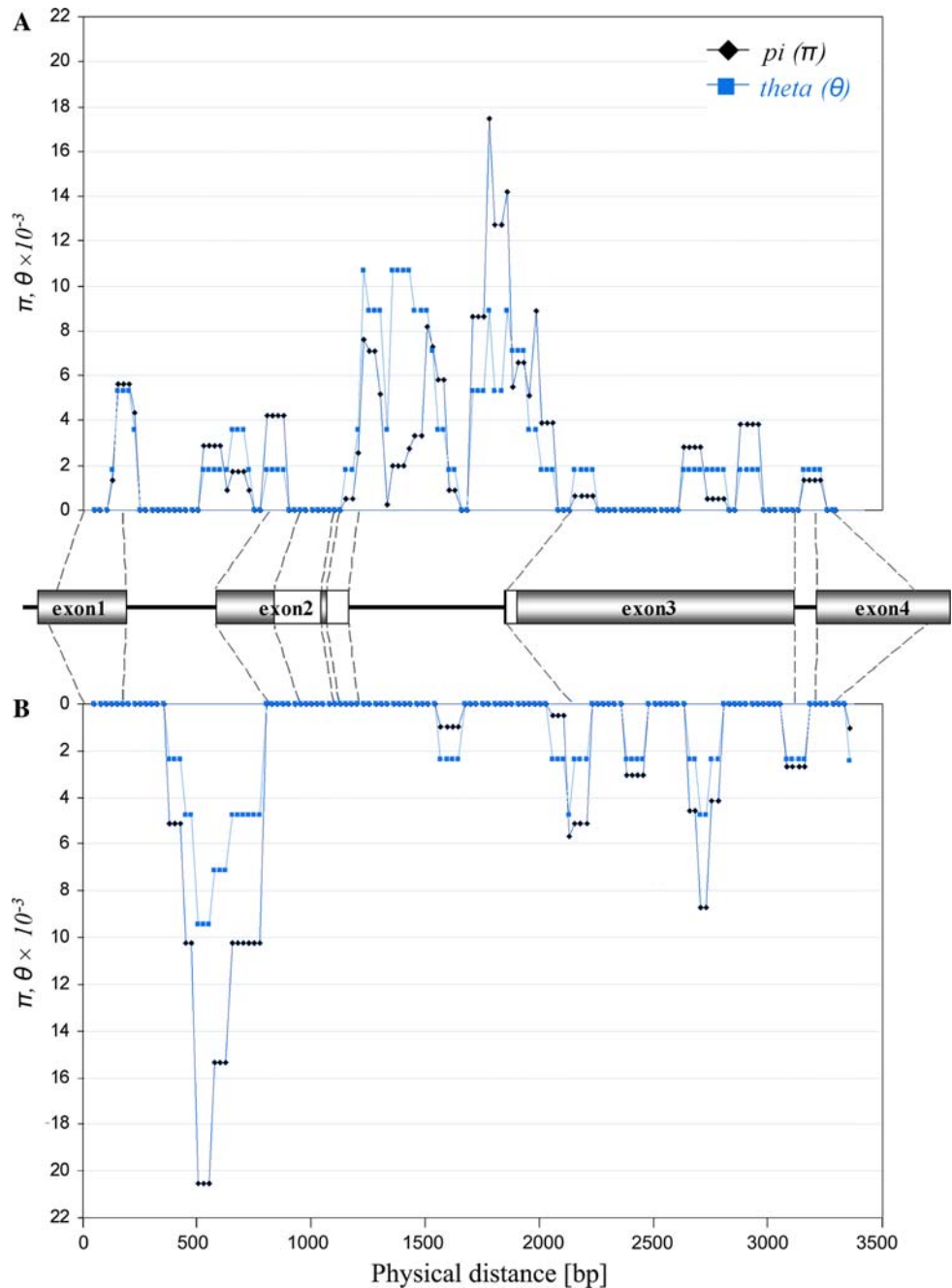

distributed across all geographical origins. Despite the lower number of East Asian accessions ($N = 45 = 29.0\%$), these were classified in 8 of the 18 haplotypes and three haplotypes (10, 12, and 17) were specific to this region. The haplotype diversity based on geographical origins ranged from $\hat{h} = 0.622$ (East Asia) to $\hat{h} = 0.837$ (West Asia, Table S4). Overall, we observed a high haplotype diversity within the *HvGAMYB* gene with $\hat{h} = 0.735 \pm 0.033$.

For *TaGAMYB,* one, seven and three haplotypes were identified for the A, B and D genomes, respectively. For the B genome, the seven haplotypes resulted in two different amino acid sequences. The haplotype diversity for this gene was high, $\hat{h} = 0.746 \pm 0.003$, which was almost identical to that observed in barley. The most frequent haplotype (haplotype 1, about 48%) included lines from all the origins studied (Fig. 3). Haplotypes 2 and 4 were each represented by one accession (Récital from France and A4 from Afghanistan). In this set of lines, the overall diversity was mainly due to the diversity present in European and Asian lines as reflected by the high haplotype diversity values for Europe (0.770 ± 0.043) and West Asia (1.000 ± 0.272, Table S4). For the D genome, we found

Fig. 3 Haplotype frequencies of the *GAMYB* gene in barley (**a**) and wheat B-genome (**b**). Pie charts show the composition of each germplasm with numbers (*N*) of accessions per geographical origin. The origins are symbolised by different patterns as shown in the key (*AM* America, *EA* East Asia, *EU* Europe, *OC* Oceania, *WA* West Asia only, *WANA* West Asia and North Africa)
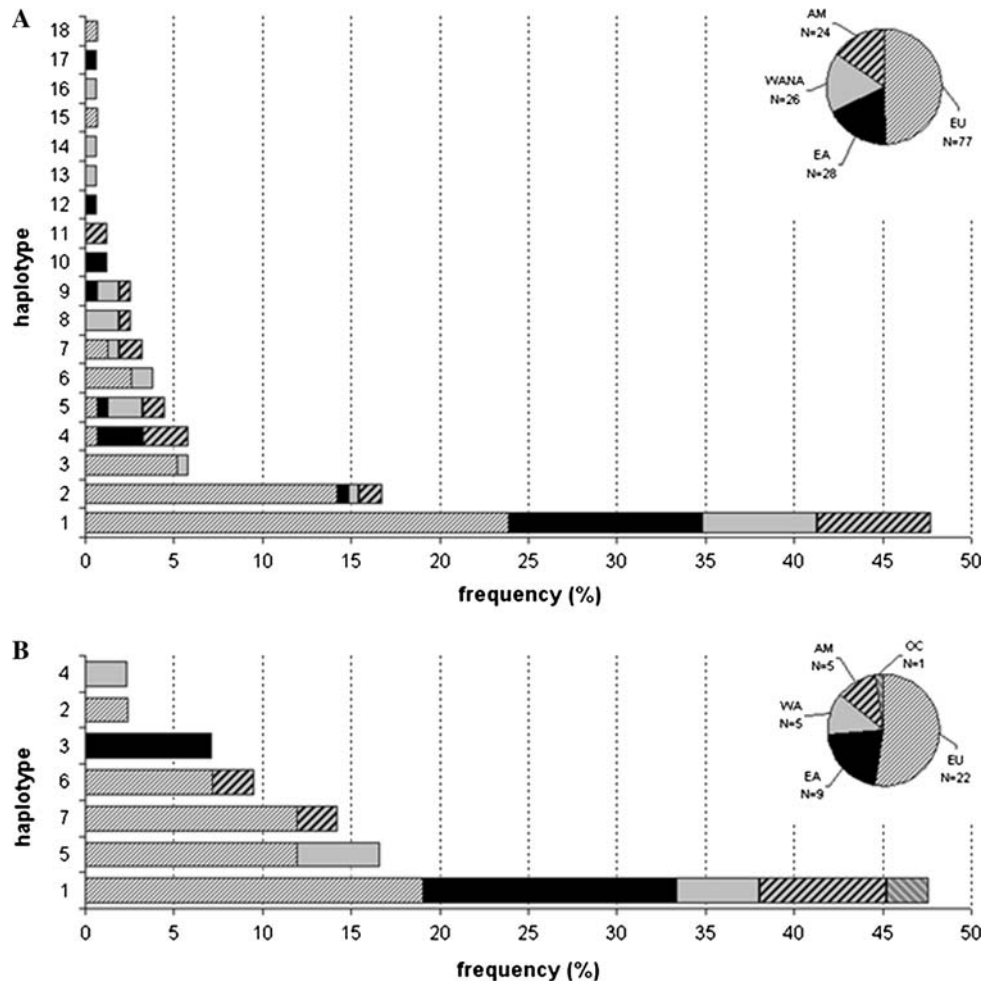
three haplotypes ($\hat{h} = 0.148 \pm 0.007$) that did not result in variation at the protein level.

Linkage disequilibrium (LD)

In barley, LD was generally significant along the sequenced part of the gene (on average $R^2 = 0.3$). Eighteen pairs of sites (among 120) revealed a high level of R^2 (>0.6) and 13 site pairs were in complete LD (Fig. 4). The remaining significant pairwise comparisons yielded moderate LD values. The plot of R^2-values as a function of the pairwise distance between polymorphic sites revealed a steep decay of LD within the *GAMYB* locus at a distance of about 2.5 kb (Fig. 5).

In the B genome of wheat (Fig. 4), we observed an absolute LD between all SNPs located in the first intron. Additionally, we detected a significant LD between pairs of polymorphic sites in the first intron and the third exon. A strong LD was found between the two polymorphic sites in the second intron and the site in the 3′UTR region, but no significant LD was found between these and other sites. No significant LD was detected between the site located in the

coding part of the fourth exon and other sites. This result suggested that LD is gradually lost along *TaGAMYB*-B (Fig. 5). Unlike in barley, the pattern of LD in wheat allowed the identification of a set of SNPs, which can be used to distinguish between different haplotypes (except for haplotypes 2 and 4). This set should include at least four SNPs: the first SNP being chosen from those at positions −389, −317, −293, −261, −260, −200, −166, −156, −65, −52 or 1515; the second SNP being at positions 863 or 1000; the third at position 1749; and the last SNP at position 2079 or 2037.

Genetic mapping of the *GAMYB* genes

Using a CAPS marker designed for the polymorphic site at position 2542 *HvGAMYB* was mapped on the long arm of chromosome 3H in the Steptoe × Morex mapping population, where it co-segregated with the marker ABG453 at position 100.6 cM (Stein et al. 2007) (Fig. 6).

In wheat, the A, B and D homoeologs were assigned to the long arm of group 3 chromosomes (data not shown) to the bins 3AL3-0.42–0.78, 3BL7-0.63–1.00 and 3DL3-

Fig. 4 Linkage disequilibrium matrix for SNPs within *GAMYB* in *H. vulgare* (*top*) and *T. aestivum* B-genome (*bottom*). SNPs with a minor allele frequency below 5% were not included. Polymorphic sites are plotted on the *x*-axis and *y*-axis. Pairwise comparison values of linkage disequilibrium are displayed as R^2 and the corresponding *P* values for Fisher's exact test are given. The structure of the gene is shown between the two heat plots. *Black lines* indicate the positions of SNPs in the gene; *open boxes* indicate the functional MYB domain
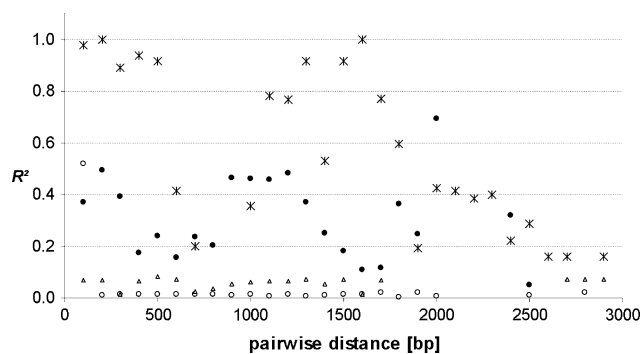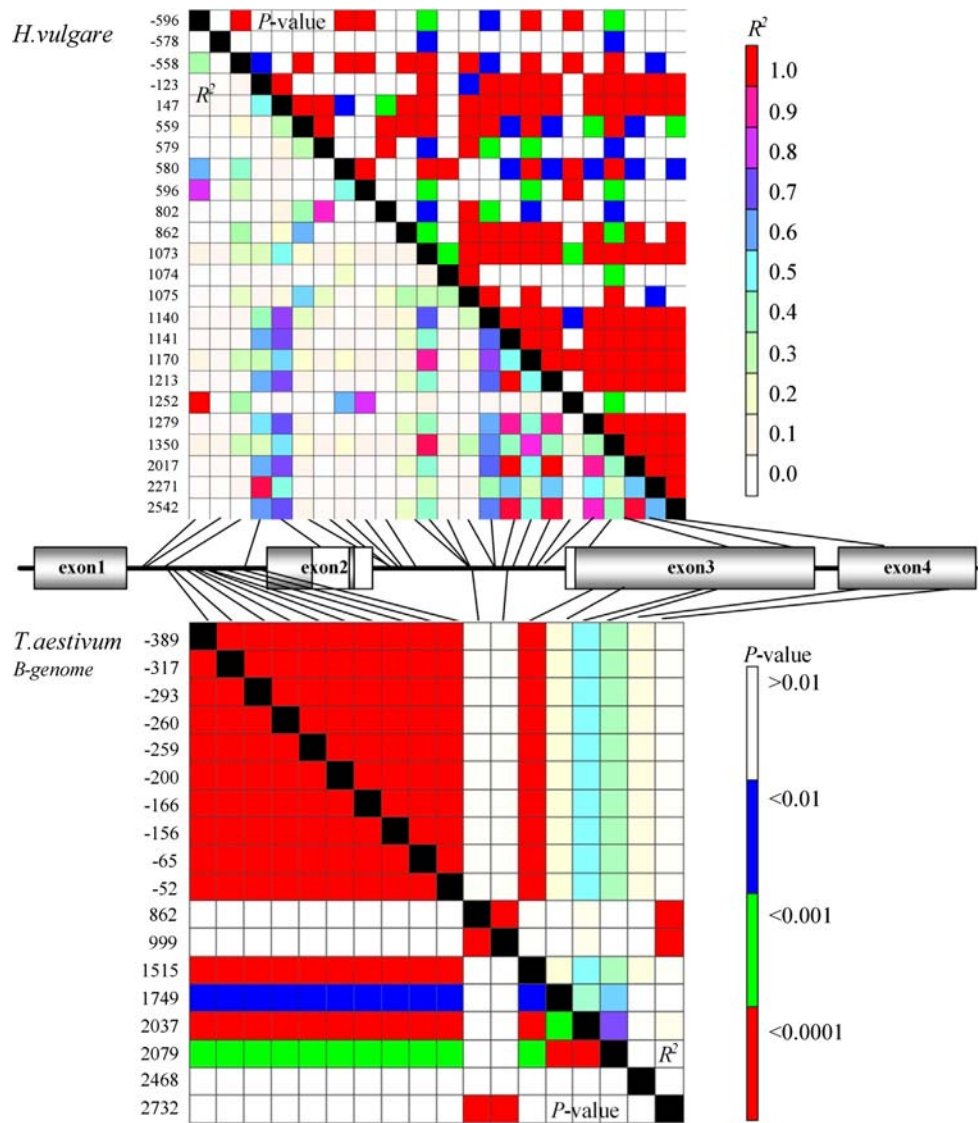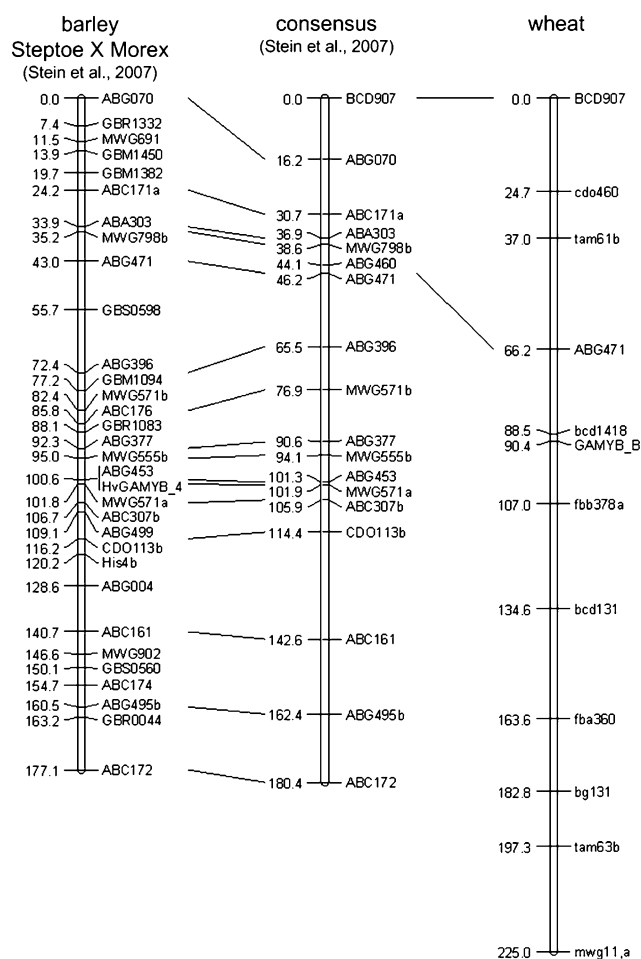

Fig. 5 Linkage disequilibrium structure in *HvGAMYB* (*filled circles*) and *TaGAMYB-B* (*crosses*). The plot shows the pair-wise LD measurement R^2 related to physical distance (bp). The data consist of 23 polymorphic sites in the germplasm set of 155 barley accessions and 21 polymorphic sites in 42 wheat accessions. Only sites with a minor allele frequency > 0.05 are considered. Non-significant R^2-values are indicated by *open symbols*

0.81–1.00. Our results suggested that these genes are likely to be located in the sub-telomeric region of the group 3 chromosomes. In the ITMI population, using the SNP at position 2037 *TaGAMYB*-B was mapped on the 3BL between the markers bcd1418 and fbb378a (Fig. 6).

Comparison between *HvGAMYB* and homoeologs of *TaGAMYB*

The sequences of *HvGAMYB* and *TaGAMYB* homoeologs were compared phylogenetically. A neighbour-joining tree revealed two distinct haplotype groups (Fig. S2). In the first group, all sequences from *HvGAMYB* were clustered together while the second group contained the *TaGAMYB* sequences. Within this second group, there were three subgroups corresponding to the homoeologous groups of wheat. Pairwise comparisons between the coding sequences of each haplotype yielded similar results. Polymorphism

Fig. 6 Genetic map of *GAMYB* on chromosome 3H of *H. vulgare* (*left* and *centre*) and *T. aestivum* (B genome, *right*). *Lines* between maps connect syntenic marker positions

Table 1 Range of sequence identity (%) from pairwise comparisons

	Haplotypes from			
	HvGAMYB	*TaGAMYB-A*	*TaGAMYB-B*	*TaGAMYB-D*
Haplotypes from				
HvGAMYB	98.8–100	95.9–96.1	95.1–95.6	95.7–96.0
	99.6–100			
TaGAMYB-A	96.4–96.5	100.0	96.4–96.5	98.2–98.2
	96.4–96.5	**100.0**		
TaGAMYB-B	95.8–96.2	96.0–96.2	99.8–99.9	96.4–96.6
	95.8–96.2	**96.0–96.2**	**99.8–100.0**	
TaGAMYB-D	96.4–96.5	98.0–98.0	95.7–95.8	99.9–100.0
	96.4–96.5	**98.0–98.0**	**95.7–95.8**	**100.0–100.0**

The coding sequences (and the amino acid sequences) of each *HvGAMYB* haplotype were compared to the coding sequences of each haplotype for the three *TaGAMYB* homoeologs. Nucleotide and amino acid identity are indicated in normal and bold characters, respectively

Nevertheless *GAMYB* genes showed a high level of conservation at the nucleotide level both within and between the two species. As expected, the level of conservation is higher in the exons than in the noncoding regions. We found that the nucleotide sequence of the functional MYB domain was almost completely conserved. Translation of the coding sequences revealed few amino acid differences within GAMYB from barley and the three wheat genomes. In agreement with Gubler et al. (1995) who reported that the N-terminus of GAMYB is highly conserved, most of the amino-acid polymorphisms in wheat and barley and between the species are at the C-terminus of the protein. However, the amino acid sequences of the two homologous tandem repeats R2 (42–94aa) and R3 (95–145aa), which constitute the functional MYB domain, are identical in both species indicating conservation due to function. This corresponds to findings in Arabidopsis, where the functional domains of GAMYB-like proteins where shown to have been conserved during natural and artificial selection (Gocal et al. 2001).

Despite the similar distribution of *GAMYB* polymorphisms in both collections, none of the polymorphisms identified in barley and wheat was located at the same position. This indicates that all the polymorphisms may have occurred after the ancestors of *H. vulgare* and *T. aestivum* diverged (10–14 million years, Wolfe et al. 1989). On average, we observed 1 polymorphic site per 74 bp for barley and 1 SNP per 441 bp for wheat when all the wheat genomes are considered together. In general, SNP frequency in different sets of barley germplasm and across different loci varies from 1 SNP every 27–78 bp (Bundock and Henry 2004; Russell et al. 2004) to 1 SNP every 200–240 bp (Rostoks et al. 2005). Hence the level of sequence polymorphism found here fits well with previous

between the individual *TaGAMYB* homoeologs was lower than between *HvGAMYB* and *TaGAMYB* alleles (Table 1). Despite a few differences in the coding sequence, the nucleotide identity between the consensus for *HvGAMYB* and *TaGAMYB* was high (about 95%, Table 1). In wheat, the A and D alleles were most similar.

Discussion

In the present study, a detailed comparison of *GAMYB* diversity in germplasm sets of wheat and barley is reported. The close relationship between wheat and barley is known from comparative genomics (Gale and Devos 1998, Feuillet and Keller 1999; Sandhu and Gill 2002) where differences in gene order and relative recombination in cereals were least between wheat and barley.

The germplasm analyzed in this study includes accessions from different geographical origins and represents a wide range of diversity known for both species.

reports based on a large number of lines including land races and wild barley (Bundock and Henry 2004; Russell et al. 2004). Results of the haplotype analysis confirm that the set of barley lines used here represents a wide range of genetic diversity of cultivated spring barley.

In wheat, the polymorphism density of *GAMYB* was lower than the value reported by Ravel et al. (2006), that is, 1 SNP per 212 bp for a set of 26 lines with different geographical origins, although the diversity of the B homoeolog is similar. In wheat, diversity within *GAMYB* was mainly caused by European and Asian accessions. European accessions were found in 5 of the 7 haplotypes. The two remaining haplotypes came from Asia. This agrees with reports on current worldwide wheat diversity showing the prevalent influence of lines of European and of Asian origin on diversity (Balfourier et al. 2007). The most frequent wheat haplotype comprised lines from all origins represented in the present collection. We observed that *TaGAMYB*-A was totally conserved and *TaGAMYB*-B had more sequence variation than its D homoeologs. This agrees with previous reports that indicate that the B genome is more polymorphic than the A and D genomes (e.g. Nelson et al. 1995; Ravel et al. 2006). However, despite the different levels of polymorphism observed between the three homoeologous genes of wheat, expression data from EST databases showed that the *GAMYBs* from all three genomes are expressed.

The Tajima's *D* test revealed no departure from the neutral equilibrium model for *HvGAMYB*. This indicates that DNA sequences of *GAMYB* in spring barley evolved randomly and were not under directional selection. In the wheat collection, a significantly positive Tajima *D* value for *TaGAMYB* was observed. This may reflect directional selection on *TaGAMYB* that was followed by a decrease of rare alleles and an excess of common alleles, respectively. This is supported by the reduced diversity and the higher frequency of minor alleles in wheat compared to *HvGAMYB*. Unlike barley, where only the total amount raw protein content is of importance, because of the requirements for malting quality (low protein levels) and feed (high protein level), a balanced composition of storage proteins is very important with respect to baking quality. To what extent the protein composition in wheat is dependent on the orchestrated expression of the three *TaGAMYB* homoeologs and the selective forces acting on the B-genome alleles needs to be further studied. From the present data we hypothesize that GAMYB effects rather the protein composition than the total protein content in the grain.

Recombination and mutation influence LD, which decreases with distance. In selfing species such as wheat and barley, effective recombination is rare and thus more extensive LD is expected than in outcrossing species. In the present study, the level of LD detected within *GAMYB* was lower than expected both for wheat and barley (10 cM, Kraakman et al. 2004; up to 5.5 cM, Stracke et al. 2007; up to 5 cM, Breseghello and Sorrells 2006). Caldwell et al. (2006) found different patterns of LD for four gene loci of barley and observed that the extent of LD depends strongly on the population under study. Therefore, the level of LD within *GAMYB*, may reflect the wide genetic basis of the germplasm as well as a high recombination frequency around the gene. Regarding the latter the cosegregating marker ABG453 has been genetically and physically mapped to a chromosomal segment on barley chromosome 3HL showing medium recombination activity (2.1 Mbp/cM). However, a recombination hotspot (0,3 Mbp/cM) is present only about 1 cM distal to the gene (Kunzel et al. 2000). A similar situation is expected for wheat, where a comparable pattern of recombination has been observed (Akhunov et al. 2003).

The structural organization of *GAMYB* has been conserved in wheat, barley. Conservation of gene structure in cereals has been described previously for other genes (Bennetzen and Ma 2003), e.g. the *Waxy* genes, the genes coding for the Opaque2 sub-family of bZIP factors (Vicente-Carbajosa et al. 1998), and the genes encoding class II starch synthase (Li et al. 2003).

The rice homolog *OsGAMYB* (Gubler et al. 1997) was mapped on rice chromosome 1 (Kaneko et al. 2002), which across its entire length shows a collinear marker order with barley 3H (Stein et al. 2007). Similarly, wheat group 3 chromosomes share the highest homology with rice chromosome 1 (Devos and Gale 1993) despite some rearrangements (Munkvold et al. 2004).

TaGAMYB colocalized with quantitative trait loci (QTL) influencing pre-harvest sprouting and nitrogen tolerance (Groos et al. 2002; Laperche et al. 2007). This suggests that *TaGAMYB*-B may be a candidate gene affecting these traits. In barley, QTLs for grain protein content and heading date near the *HvGAMYB* candidate gene have been reported (Hayes et al. 1993). A role of GAMYB in pre-harvest sprouting would be not surprising as GAMYB was previously shown to be expressed in barley during germination (Gubler et al. 1995). Likewise, a role of GAMYB in nitrogen metabolism and therefore grain protein content might be expected as this transcriptional protein is also involved in storage protein synthesis (Diaz et al. 2002).

To conclude, the high level of conservation observed in *GAMYB* coding sequences in both wheat and barley may reflect the importance of this transcription factor in several developmental mechanisms. Transcription factors often act pleiotropically and are sometimes modulated through protein–protein interactions, as is the case for GAMYB (Diaz et al. 2002). Their molecular structures and hence their primary sequences may thus be subject to strong

constraints which can explain why regulatory genes may show less diversity than structural genes. Therefore, the polymorphisms identified in the present study will provide an opportunity to studying the involvement of GAMYB in a variety of economically important phenotypic traits.

References

Akhunov ED, Goodyear AW, Geng S, Qi LL, Echalier B, Gill BS et al (2003) The organization and rate of evolution of wheat genomes are correlated with recombination rates along chromosome arms. Genome Res 13:753–763

Balfourier F, Ravel C, Bochard AM, Exbrayat-Vinson F, Boutet G, Sourdille P, Dufour P, Charmet G (2006) Développement, utilisation et comparaison de différents types de marqueurs pour étudier la diversité parmi une collection de blé tendre. Les Actes BRG 6:129–144

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 restricted to a full 384 deep well storage plate. Theor Appl Genet 114:1265–1275

Bennetzen JL, Ma J (2003) The genetic colinearity of rice and other cereals on the basis of genomic sequence analysis. Curr Opin Plant Biol 6:128–133

Breseghello F, Sorrells ME (2006) Association mapping of kernel size and milling quality in wheat (Triticum aestivum L.) cultivars. Genetics 172:1165–1177

Buckler E, Bradbury P, Kroon D, Ramdoss Y, Fink AJ (2006) Trait Analysis by Association, Evolution and Linkage—TASSEL. http://www.maizegenetics.net/tassel

Bundock PC, Henry RJ (2004) Single nucleotide polymorphism, diversity and recombination in the Isa gene of barley. Theor Appl Genet 109:543–551

Caldwell KS, Russell J, Langridge P, Powell W (2006) Extreme population-dependent linkage disequilibrium detected in an inbreeding plant species, Hordeum vulgare. Genetics 172:557–567

Cercós M, Gómez-Cadenas A, Ho THD (1999) Hormonal regulation of a cysteine proteinase gene, EPB1, in barley aleurone layers: Cis- and trans-acting elements involved in the coordinated gene expression regulated by gibberellins and abscisic acid. Plant J 19:107–118

Chen L, Nishizawa T, Higashitani A, Suge H, Wakeui H, Takeda K, Takahashi H (2001) A variety of wheat tolerant to deepseeding conditions: elongation of the first internode depends on the response to gibberellin and potassium. Plant Cell Environ 24:469–476

Devos KM, Gale MD (1993) Extended genetic maps of the homoeologous group 3 chromosomes of wheat, rye, and barley. Theor Appl Genet 85:649–652

Diaz I, Vicente-Carbajosa J, Abraham Z, Martinez M, Isabel-Lamoneda I, Carbonero P (2002) The GAMYB protein from barley interacts with the DOF transcription factor BPBF and activates endosperm-specific genes during seed development. Plant J 29:453–464

Diaz I, Martinez M, Isabel-LaMoneda I, Rubio-Somoza I, Carbonero P (2005) The DOF protein SAD interacts with GAMYB in plant nuclei and activates transcription of endosperm-specific genes during barley seed development. Plant J 42:652–662

Endo TR, Gill BS (1996) The deletion stocks of common wheat. J Hered 87:295–307

Felsenstein J (2005). PHYLIP (Phylogeny Inference Package) version 3.6. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle

Feuillet C, Keller B (1999) High gene density is conserved at syntenic loci of small and large grass genomes. Proc Natl Acad Sci USA 96:8265–8270

Gale MD, Devos KM (1998) Plant comparative genetics after 10 years. Science 23:656–659

Gocal GFW, Poole AT, Gubler F, Watts RJ, Blundell C, King RW (1999) Long-day up-regulation of a GAMYB gene during Lolium temulentum inflorescence formation. Plant Physiol 119:1271–1278

Gocal GFW, Sheldon CC, Gubler F, Moritz T, Bagnall DJ, MacMillan CP, Li SF, Parish RW, Dennis ES, Weigel D, King RW (2001) GAMYB-like genes, flowering, and gibberellin signaling in Arabidopsis. Plant Physiol 127:1682–1693

Gouesnard B, Bataillon T, Decoux G, Rozale C, Schoen DL, David JL (2001) An algorithm for building germplasm core collections by maximizing allelic or phenotypic richness. J Hered 92:93–94

Groos C, Gay G, Perretant MR, Gervais L, Bernard M, Dedryver F, Charmet G (2002) Study of the relationship between pre-harvest sprouting and grain color by quantitative trait loci analysis in a white × red grain bread wheat cross. Theor Appl Genet 104:39–47

Gubler F, Jacobsen JV (1992) Gibberellin-responsive elements in the promoter of a barley high-pI α-amylase gene. Plant Cell 4:1435–1441

Gubler F, Kalla R, Roberts JK, Jacobsen JV (1995) Gibberellin-regulated expression of a myb gene in barley aleurone cells: evidence for Myb transactivation of a high-pI alpha-amylase gene promoter. Plant Cell 7:1879–1891

Gubler F, Watts RJ, Kalla R, Matthews P, Keys M, Jacobsen JV (1997) Cloning of a rice cDNA encoding a transcription factor homologous to barley GAMYB. Plant Cell Physiol 38:362–365

Gubler F, Raventos D, Keys M, Watts R, Mundy J, Jacobsen JV (1999) Target genes and regulatory domains of the GAMYB transcriptional activator in cereal aleurone. Plant J 17:1–9

Gubler F, Chandler PM, White RG, Llewellyn DJ, Jacobsen JV (2002) Gibberellin signaling in barley aleurone cells. Control of SLN1 and GAMYB expression. Plant Physiol 129:191–200

Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for windows 95/98/NT. Nucleic Acids Symp Ser 41:95–98

Hayes PM, Liu BH, Knapp SJ, Chen F, Jones B, Blake T, Franckowiak J, Rasmusson D, Sorrells M, Ullrich SE, Wesenberg D, Kleinhofs A (1993) Quantitative trait locus effects and environmental interaction in a sample of North American barley germplasm. Theor Appl Genet 87:392–401

Hill WG, Robertson A (1968) Linkage disequilibrium in finite populations. Theor Appl Genet 38:226–231

Hooley R (1994) Gibberellins: perception, transduction and responses. Plant Mol Biol 26:1529–1555

Isabel-LaMoneda I, Diaz I, Martinez M, Mena M, Carbonero P (2003) SAD: a new DOF protein from barley that activates transcription of a cathepsin B-like thiol protease gene in the aleurone of germinating seeds. Plant J 33:329–340

Kanei-Ishii C, Sarai A, Sawazaki T, Nakagoshi H, He DN, Ogata K, Nishimura Y, Ishii S (1990) The tryptophan cluster: a hypothetical structure of the DNA-binding domain of the myb protooncogene product. J Biol Chem 265:19990–19995

Kaneko M, Ashikari M, Matsuoka M (2002) Mapping and expression of rice GAMYB. Rice Genet Newsl 19:60

Kraakman ATW, Niks RE, Van den Berg P, Stam P, Van Eeuwijk VA (2004) Linkage disequilibrium mapping of yield and yield stability in modern spring barley cultivars. Genetics 168:435–446

Kimura M (1980) A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16:111–120

Kleinhofs A, Kilian A, Saghai Maroof MA, Biyashev RM, Hayes PM, Chen FQ, Lapitan N, Fenwick A, Blake TK, Kanazin V, Ananiev E, Dahleen L, Kudma D, Bollinger J, Knapp SJ, Liu B, Sorrells M, Heun M, Franckowiak JD, Hoffman D, Skadsen R, Steffenson BJ (1993) A molecular, isozyme and morphological map of the barley (Hordeum vulgare) genome. Theor Appl Genet 86:705–712

Kumar S, Tamura K, Jakobsen IB, Nei M (2001) MEGA2: molecular evolutionary genetics analysis software. Bioinformatics 17:1244–1245

Kunzel G, Korzun L, Meister A (2000) Cytologically integrated physical restriction fragment length polymorphism maps for the barley genome based on translocation breakpoints. Genetics 154:397–412

Laperche A, Brancourt-Hulmel M, Heumez E, Gardet O, Hanocq E, Devienne-Barret F, Le Gouis J (2007) Using genotype × nitrogen interaction variables to evaluate the QTLs involved in wheat tolerance to nitrogen contrainsts. Theor Appl Genet 115:399–415

Lander ES, Geen P, Abrahanson J, Barlow A, Daly MJ, Lincoln SE, Newbur L (1987) MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1:174–181

Li Z, Sun F, Xu S, Chu X, Mukai Y, Yamamoto M, Ali S, Rampling L, Kosar-Hashemi B, Rahman S, Morell MK (2003) The structural organisation of the gene encoding class II starch synthase of wheat and barley and the evolution of the genes encoding starch synthases in plants. Funct Integr Genomics 3:76–85

Martin C, Paz-Ares J (1997) MYB transcription factors in plants. Trends Genet 13:67–73

Munkvold JD, Greene RA, Bermudez-Kandianis CE, La Rota CM, Edwards H, Sorrells SF, Dake T, Benscher D, Kantety R, Linkiewicz AM, Dubcovsky J, Akhunov ED, Dvorak J, Miftahudin Gustafson JP, Pathan MS, Nguyen HT, Matthews DE, Chao S, Lazo GR, Hummel DD, Anderson OD, Anderson JA, Gonzalez-Hernandez JL, Peng JH, Lapitan N, Qi LL, Echalier B, Gill BS, Hossain KG, Kalavacharla V, Kianian SF, Sandhu D, Erayman M, Gill KS, McGuire PE, Qualset CO, Sorrells ME. (2004) Group 3 chromosome bin maps of wheat and their relationship to rice chromosome 1. Genetics 168:639–650

Murray F, Kalla R, Jacobsen J, Gubler F (2003) A role for HvGAMYB in anther development. Plant J 33:481–491

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York

Nelson JC, Van Deynze AE, Sorrells ME, Lu YH, Atkinson M, Bernard M, Leroy P, Faris J, Anderson JA (1995) Molecular mapping of wheat: major genes and rearrangements in homoeologous groups 4, 5 and 7. Genetics 141:721–731

Ravel C, Praud S, Murigneux A, Canaguier A, Sapet F, Samson D, Balfourier F, Dufour F, Chalhoub B, Brunel D, Beckert M, Charmet G (2006) Single-nucleotide polymorphisms (SNPs) frequency in a set of selected lines of bread wheat (Triticum aestivum L.). Genome 49:1131–1139

Rostoks N, Mudie S, Cardle L, Russell J, Ramsay L, Booth A, Svensson J, Wanamaker S, Walia H, Rodriguez E, Hedley P, Liu H, Morris J, Close T, Marshall D, Waugh R (2005) Genome-wide SNP discovery and linkage analysis in barley based on genes responsive to abiotic stress. Mol Gen Genomics 274:515–527

Rozas J, Sánchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP: DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19:2496–2497

Rubio-Somoza I, Martinez M, Abraham Z, Diaz I, Carbonero P (2006) Ternary complex formation between HvMYBS3 and other factors involved in transcriptional control in barley seeds. Plant J 47:269–281

Russell J, Booth A, Fuller J, Harrower B, Hedley P, Machray G, Powell W (2004) A comparison of sequence-based polymorphism and haplotype content in transcribed and anonymous regions of the barley genome. Genome 47:389–398

Sandhu D, Gill KS (2002) Structural and functional organization of '1S0.8 gene-rich region' in Triticeae. Plant Mol Biol 48:791–804

Sears ER (1966) Nullisomic-tetrasomic combinations in hexaploid wheat. In: Riley R, Lewis KR (eds) Chromosome manipulations and plant genetics. Olivier and Boyd, Edinburgh, pp 29–45

Staden R, Beal KF, Bonfield JK (2000) The Staden package, 1998. Methods Mol Biol 132:115–130

Stein N, Herren G, Keller B (2001) A new DNA extraction method for high-throughput marker analysis in a large-genome species such as Triticum aestivum. Plant Breed 120:354–356

Stein N, Prasad M, Scholz U, Thiel T, Zhang H, Wolf M, Kota R, Varshney RK, Perovic D, Grosse I, Graner A (2007) A 1000-loci transcript map of the barley genome: new anchoring points for integrative grass genomics. Theor Appl Genet 114:823–839

Stracke R, Werber M, Weisshaar B (2001) The R2R3-MYB gene family in Arabidopsis thaliana. Curr Opin Plant Biol 4:447–456

Stracke S, Presterl T, Stein N, Perovic D, Ordon F, Graner A (2007) Effects of introgression and recombination on haplotype structure and linkage disequilibrium surrounding a locus encoding Bymovirus resistance in barley. Genetics 175:805–817

Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585–595

Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position specific gap penalties and weight matrix choice. Nucleic Acids Res 22:4673–4680

Tixier MH, Sourdille P, Charmet G, Gay G, Jaby C, Cadalen T, Bernard S, Nicolas P, Bernard M (1998) Detection of QTLs for crossability in wheat using double-haploid population. Theor Appl Genet 97:1076–1082

Tsuji H, Aya K, Ueguchi-Tanaka M, Shimada Y, Nakazono M, Watanabe R, Nishizawa NK, Gomi K, Shimada A, Kitano H, Ashikari M, Matsuoka M (2006) GAMYB controls different sets of genes and is differentially regulated by microRNA in aleurone cells and anthers. Plant J 47:427–444

Vicente-Carbajosa J, Oñate L, Lara P, Diaz I, Carbonero P (1998) Barley BLZ1: a bZip transcriptional activator that interacts with endosperm-specific gene promoters. Plant J 13:629–640

Washio K (2003) Functional dissections between GAMYB and Dof transcription factors suggest a role for protein–protein associations in the gibberellin-mediated expression of the RAmy1A gene in the rice aleurone. Plant Physiol 133:850–863

Watterson GA (1975) On the number of segregating sites in genetical models without recombination. Theor Popul Biol 7:256–276

Wolfe KH, Gouy M, Yang Y-H, Sharp PM, Li WH (1989) Date of the monocot-dicot divergence estimated from chloroplast sequence data. Proc Natl Acad Sci USA 86:6201–6205

Woodger FJ, Millar A, Murray F, Jacobsen JV, Gubler F (2003) The role of GAMYB transcription factors in GA-regulated gene expression. J Plant Growth Regul 22:176–184