Theor Appl Genet (2006) 113:1063–1079
DOI 10.1007/s00122-006-0366-3

ORIGINAL PAPER

Novel haplotype description and structural background of the eventual functional significance of the barley β-amylase gene intron III rearrangements

T. G. Sjakste · A. F. Zhuk

Received: 22 March 2006 / Accepted: 6 July 2006 / Published online: 19 August 2006
© Springer-Verlag 2006

Abstract To extend the knowledge on the haplotype variability of the *Bmy1* gene, region of the intron III was sequenced in a set of 20 Latvian accessions and Danish variety Maja, the data were compared to the previously reported allelic variants of the structural gene. Taking into account the polymorphisms of 59 loci and the microsatellite (MS) motif, 11 Latvian varieties turned out to have haplotype similar to cultivar Adorra, 1 – to Haruna Nijo, and 8 – to the newly described Abava *Bmy1* intron III haplotype. High level of polymorphisms of $(TG)_m$ as well as $(G)_n$ component of MS was revealed for all the haplotypes studied. We conclude that the MS motif rather than the MS size length polymorphism correlates with mutations in the coding region of the β-amylase gene. Five graphical haplotype-specific intron III structures were constructed on the basis of the co-localization of the transcription factor binding sites (TFBSs), remnants of the transposable elements, and intron III polymorphic loci. Inter- and intrahaplotype variability was analyzed on the eventual functional significance of the *Bmy1* intron III rearrangements. Novel data on the intron III nucleotide sequences of the *Bmy1* gene were deposited in the GenBank (http://www.ncbi.nlm.nih.gov/) under accession numbers DQ316895–DQ316905.

Communicated by F. Ordon

T. G. Sjakste (✉) · A. F. Zhuk
Genomics and Bioinformatics, Institute of Biology
of the University of Latvia, Miera Str. 3,
Salaspils, 2169, Latvia
e-mail: tanja@email.lubi.edu.lv

Introduction

The highly polymorphic barley *Bmy1* gene encoding endosperm specific β-amylase provides an excellent model for comparative genomic analysis of the allelic variants and evaluation of the mechanisms of the gene structural rearrangements and their eventual functional significance. Four allelic forms of β-amylase exhibit increasing levels of thermostability in the range of Sd1, Sd2L, Sd2H, and Sd3 alleles (Eglinton et al. 1998). Structural gene *Bmy1* consists of seven exons and six introns (Yoshigi et al. 1995). The gene is located on the long arm of chromosome 4H (Kreis et al. 1987) and is characterized by a high level of polymorphisms related to the thermostability and kinetic properties of the enzyme (Erkkilä et al. 1998; Erkkilä and Ahokas 2001; Kaneko et al. 2000; Ma et al. 2001; Paris et al. 2002). To distinguish the *Bmy1* alleles, Paris et al. (2002) proposed using two non-synonymous single nucleotide polymorphisms (SNPs) of the coding region of the gene (cSNPs) linked to the enzyme thermostability ($G^{496} \rightarrow C$ and $C^{698} \rightarrow T$ in exon III and exon IV correspondingly). Despite only two cSNPs in sight, this approach was successfully used in several genotyping studies (Polakova et al. 2003; Malisheva et al. 2004; Sjakste and Röder 2004). Discovery of synonymous cSNP $G^{702} \rightarrow A$ enabled Sd1 alleles to be subdivided into two subgroups (Polakova et al. 2003).

Comparison of the enzymatic activity of different allelic forms suggested that rearrangements of the promoter, intron II and especially intron III regions of the structural gene could be associated with the β-amylase properties together with amino acid substitutions (Erkkilä and Ahokas 2001). Actually, sequences of four *Bmy1* allelic variants are publicly accessible including genes of Haruna Nijo (GenBank D4999), Adorra (GenBank AF061203),

Hordeum spontaneum strain PI 296897 (GenBank AF061204), and Finnish landrace line HA52 (GenBank AJ301645). Simple alignment of the mentioned genomic sequences suggests existence of numerous polymorphic loci in coding and non-coding regions of the gene. Some of them could influence enzymatic properties or/and could be used for more precise variety discrimination.

Interesting data were accumulated about the intron III polymorphisms. A 126 bp insertion/deletion event (indel) in the 5′ region is associated with allelic variants of the gene encoding enzymes of low or high thermostability correspondingly (Erkkilä and Ahokas 2001). Genotyping of the 3′ region of the intron in 55 North European barley varieties revealed the linkage between a 1 + 6 bp indel (S/L allele) and $C^{698} \rightarrow T$ (V233A) allelic variations of exon IV (Sjakste and Röder 2004) responsible for the highest and the lowest thermostability of the enzyme correspondingly (Paris et al. 2002). High intron III microsatellite (MS) length polymorphism was identified in varieties from all over the world (Sjakste and Röder 2004; Malisheva et al. 2004). However, no significant correlations were revealed between MS size and S/C^{698} genotype (Sjakste and Röder 2004; Malisheva et al. 2004). At that time, we described several linkage blocks on three polymorphic events (1 + 6 bp indel, MS size, and $C^{698} \rightarrow T$ variation) that were transmitted through the generations of different independent pedigrees (Sjakste and Röder 2004). Our analysis suggested the existence of certain associations between MS locus variability and other polymorphisms of the structural gene. Comparative analysis of the publicly available genomic sequences of the four allelic variants could not reveal the details of these associations due to the lack of data.

Therefore, the first goal of this work was to obtain more details about *Bmy1* intron III polymorphisms in different barley varieties. This task was achieved by extending the sequencing data using a set of Latvian barley cultivars. Secondly, we tried to develop approaches for more precise haplotype classification and characterization of inter- and intrahaplotype variability. Finally, we made an attempt to use intron III sequences of different haplotypes as a model system for comparative computational evaluation of the structural and eventual functional reorganization of the non-coding gene region.

Materials and methods

Detection and genotyping of polymorphisms

The set of 21 spring barley accessions used in the polymorphisms detection and genotyping study was composed of 20 Latvian commercial varieties and a well-known Danish variety Maja (Table 1). Genotypes included in the set were previously analyzed on MS or/and *Bmy1* alleles' genetic diversity and inheritance through the generations of known pedigrees (Sjakste et al. 2003; Sjakste and Röder 2004). The seeds of all accessions studied were obtained from the Latvian Gene Bank of Cultivated Plants in Salaspils, Latvia. Genomic DNA was extracted according to previously described procedures (Plaschke et al. 1995) using young leaves of one plant of each accession per sample. Two or three individual DNA samples from each accession or bulk DNA from ten accessions was used in further analysis.

The intron III region was screened for detection of polymorphisms by direct sequencing of the amplified fragment of the gene. The strategy of the amplification and sequencing is presented in Fig. 1. Primer design was performed by the Primer 3.0 program using the highly conserved regions from exon III and exon IV as well as from intron III generated from the alignment of database available genomic sequences of four barley accessions: Haruna Nijo (GenBank D4999), Adorra (GenBank AF061203), *Hordeum vulgare* subsp. *spontaneum* NPGS PI 296897 (GenBank AF061204), and the Finnish landrace line HA52 (GenBank AJ301645).

PCR resulted in amplification of the region encompassing the whole intron III (primers *Bmy1* AF: 5′-tgacagatgtatgccgatta-3′ and *Bmy1* AR: 5′-ttgttggagtaccatgcaag-3′) or intron 3′ region encompassing the MS portion (*Bmy1* S1F: 5′-cggataccatgtaaaactgcac-3′ and *Bmy1* S2R: 5′-tttttctgtaatggcaatggt-3′). Amplification was followed by direct sequencing with the same primers plus two others specific for the 5′ region of intron III (*Bmy1* S3F: 5′-ggcgtggtaaacctgacttg-3′ and *Bmy1* S4R: 5′-tcaggtttaccacgccttg-3′) in both forward and reverse orientation of each PCR product (Fig. 1). Two individual DNA samples were sequenced per each accession. The 126 bp indel in the 5′ region was additionally genotyped on the size difference of the amplified products (primers *Bmy1* AF and *Bmy1* S4R) using agarose gel electrophoresis. Several DNA samples or/and bulk DNA were used in this case for each accession analyzed.

Fluorescence was measured using an ABI Prism 7000 sequence detection system and analyzed with the ABI PRISM 7000 SDS software version 1.0.

Bioinformatics analysis

Alignments of the previously published *Bmy1* intron III genomic sequences and sequences produced in this study were generated by the multiple alignment service

Table 1 Description of the accessions including year of release, origin/breeding station, immediate pedigree, form of spike, β-amylase intron III haplotype, and details on MS polymorphism

Cultivar	Year of release	Origin/breeding station	Pedigree	Form	Haplotype	MS region sequence data	MS size (bp)
Maja	1934	Denmark	Binder/Gull	2	HN	$(TG)_9(G)_{16}TG$	150
Latvijas Vietejie	–	LV	Latvian local (landrace)	2	HN	$(TG)_{10}(G)_{18}TG$	154
Ansis	1995	LV/Stende	Jarek/Taifun	2	AD	$TG(G)_2(TG)_4TT(G)_{12}$	140
Balga	1990	LV/Priekuli	Gunilla/KM 1192	2	AD	$TG(G)_2(TG)_4TT(G)_{12}$	140
Kombainieris	1950	LV/Stende	Maja/Latvian (Talsu) local	2	AD	$TG(G)_2(TG)_4TT(G)_{12}$	140
Linga	1985	LV/Priekuli	Gunilla/KM 1192	2	AD	$TG(G)_2(TG)_4TT(G)_{12}$	140
Rasa	1991	LV/Stende	Frankengold/KM-R-54/72	2	AD	$TG(G)_2(TG)_4TT(G)_{12}$	140
Sencis	1994	LV/Stende	Rupal/Ofir/Torkel	2	AD	$TG(G)_2(TG)_4TT(G)_{14}$	140
Druvis	1999	LV/Stende	Dobrij/HVS 115440	6	AD	$TG(G)_2(TG)_4TT(G)_{15}$	143
Malva	2001	LV/LLU	Stende 8142/Stende 7542	2	AD	$TG(G)_2(TG)_4TT(G)_{14}$	143[a]
Idumeja	1998	LV/Priekuli	Imula/Ilga	2	AD	$TG(G)_2(TG)_4TT(G)_{18}$	146
Imula	1985	LV/Stende	Abava/2 × Akka	2	AD	$TG(G)_2(TG)_4TT(G)_{18}$	146
Priekulu 1	1954	LV/Priekuli	Selection from Norwegian local	2	AD	$TG(G)_2(TG)_4TT(G)_{18}$	146
Ilga	1983	LV/Priekuli	KM-1192/Hadmersleben 70197/70	2	AB	$(TG)_6TT(G)_{14}$	140
Klinta	1993	LV/LLU	Torkel/CF-42	2	AB	$(TG)_6TT(G)_{15}$	140
Abava	1978	LV/Stende	Mari/Elsa//Domen	2	AB	$(TG)_6TT(G)_{16}$	143
Gate	1995	LV/Priekuli	Emir/2 × Nadja//HE-497/Hadmersleben 70197/70	2	AB	$(TG)_6TT(G)_{15}$	143
Dzintars	1930	LV/Stende	Selection from Latvian local	6	AB	$(TG)_6TT(G)_{16}$	143
Ruja	1992	LV/Priekuli	Abava/Kombainieris/Trumph	2	AB	$(TG)_6TT(G)_{16}$	143
Vairogs	1930	LV/Priekuli	Selection from Priekulu local	6	AB	$(TG)_6TT(G)_{16}$	143
Agra	1984	LV/Priekuli	Priekulu 1/Otra	6	AB	$(TG)_6TT(G)_{18}$	146

LV indicates Latvian origin, *LLU* indicates Latvian University of Agriculture

[a] Data on MS size are indicated according to Sjakste and Röder (2004) or to the data of this paper

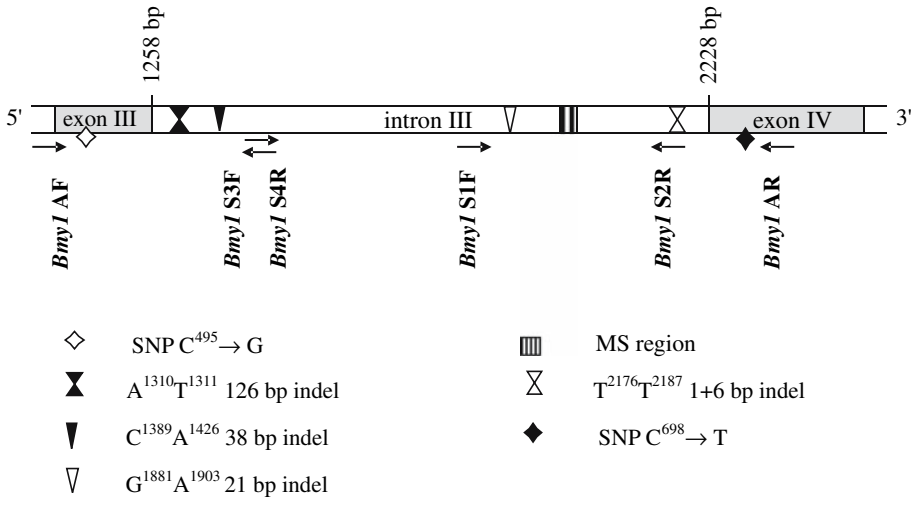

Fig. 1 Scheme of the intron III region of the *Bmy1* gene and strategy of the experiment. The relative positions of some polymorphisms are indicated according to the genomic sequence of accession Haruna Nijo (GenBank D4999)

ClustalW (http://www.clustalw.genome.jp/). Identification of transposable elements and transcription factor binding sites (TFBSs) was performed using Blast against TREP (http://www.wheat.pw.usda.gov/ITMI/Repeats/blastrepeats3.html) and Genomatix software (DiAlign TF, Release 3.1, and MatInspector, Release 7.4 tools, at http://www.genomatix.de/), respectively.

Results

Genotyping results

Bmy1 intron III sequences of 20 Latvian (Table 1) and 1 Danish variety Maja were compared by alignment with the corresponding regions of 4 allelic variants of the *Bmy1* gene (Haruna Nijo, Adorra, *H. spontaneum* strain PI 296897, and Finnish landrace line HA52). Data on polymorphic loci revealed are summarized in Tables 1 and 2 and Fig. 2.

All the previously described (Erkkilä and Ahokas 2001; Erkkilä et al. 1998; Sjakste and Röder 2004) polymorphic loci of the *Bmy1* intron III were detected in our study. Data on the high level of MS length polymorphism obtained earlier for Latvian varieties (Sjakste and Röder 2004) were confirmed by the present results. Details on MS motif variations are reported for the first time (Table 1; Fig. 2).

Our study revealed two variable components of the repeated portion of the intron. Variable 5′-MS component of MS could be represented either by $(TG)_m$ (Haruna Nijo, *H. spontaneum* strain PI 296897, Maja, Latvijas Vietejie, and Abava-like Latvian accessions), or by the restructured motifs of Finnish landrace HA52, Adorra, and Adorra-like Latvian accessions (Table 1; Fig. 2). The 5′-$(TG)_m$ repeat was shown to be followed immediately by a $(G)_n$ repeat in genes of Maja and Latvijas Vietejie similarly to Haruna Nijo, and *H. spontaneum* strain PI 296897. The 5′-TG-reach region is separated from the 3′-$(G)_n$ repeat by a TT motif in Finnish landrace HA 52, Adorra, and all Latvian accessions besides Latvijas Vietejie (Table 1; Fig. 2). A high level of the 3′-$(G)_n$ repeat polymorphism as well as several SNPs in the 5′- and the 3′-MS flanks were revealed in the accessions analyzed (Table 2; Fig. 2).

Summarizing the data on MS motif and length polymorphism, we describe here all variations of the region by five formulas (Fig. 2), and classify them as HN (Haruna Nijo-like), HS (*H. spontaneum* strain PI 296897-like), HA (Finnish landrace HA52-like), AB (Latvian cultivar Abava-like), and AD (Adorra-like) MS variants (Fig. 2).

Haplotype classification

All the 16 indels, 38 SNPs, 3 double SNPs (dSNPs), 1 fragment substitution (FR) and MS sequence motif [in contrast to the number of $(TG)_m$ and $(G)_n$ repeats] were used for the characterization of the haplotypes of the intron III of the Bmy1 gene. Five haplotypes were classified according to their MS motif as HN, HS, HA, AB, and AD haplotypes (Table 2; Figs. 2, 3).

The Danish variety Maja and the old Latvian cultivar Latvijas Vietejie turned out to have HN haplotype. Intron III sequences of both accessions differ from HN only in three SNPs (allele A^{1630} is common with HS, alleles T^{2020} and G^{2044} are common with HS, HA, AB, and AD), and a repeat number of the MS components that formed $(TG)_9$ for Maja and $(G)_{18}$ for Latvijas Vietejie in contrast to the $(TG)_{10}$ and $(G)_{16}$ for Haruna Nijo.

Eleven Latvian varieties (Table 1) were shown to have AD haplotype of the intron III with MS motif $TG(G)_2(TG)_4TT(G)_n$ of high level of the $(G)_n$ variability

Table 2 Haplotypes and eventual functional significance of polymorphisms of the intron III of *Bmy1* gene. Function of variation was scored as generation (G) or loss (L) of TFBS. Numbering of polymorphic loci is given according to the genomic sequence of cultivar Haruna Nijo (GenBank D49999). TFBSs family/matrix names is given according to Genomatix Matrix Family Library (Version 5.0)

Variation	Haplotype					TFBSs fate as a result of variation comparing with Haruna Nijo haplotype			
	HN	HS	HA	AB	AD	Fate	Family/matrix × number	Str. +/−	Abbr.
Indel	–	–	–	–	$A^{1310}T^{1311}$ → A-126 bp-T	G	P$EINL/ TEIL.01 × 2	+/−	E
						G	P$LREM/ ATCTA.01 × 3	+	LR
						G	P$STKM/STK.01	+	ST
SNP	–	–	–	–	$A^{1318} \rightarrow C$	L	P$LREM/ ATCTA.01	−	LR
SNP	–	–	$C^{1350} \rightarrow T$	$C^{1350} \rightarrow T$	–	No			
SNP	–	$G^{1384} \rightarrow A$	$G^{1384} \rightarrow A$	$G^{1384} \rightarrow A$	$G^{1384} \rightarrow A$	No			
Indel	–	C^{1389}-38 bp-$A^{1426} \rightarrow CA$	–	–	–	L	P$DOFF/PBF.01	+	D1
						L	P$GAPB/GAP.01	+	GA
						L	P$NCS1/NCS1.01	+	N
SNP	–	–	$A^{1407} \rightarrow G$	$A^{1407} \rightarrow G$	$A^{1407} \rightarrow G$	L	P$AHBP/ATHB5.01	−	A
						L	P$NCS1/NCS1.01		N
						G	P$DOFF/PBF.01		D1
SNP	–	–	$C^{1505} \rightarrow G$	$C^{1505} \rightarrow G$	–	No			
SNP	–	–	$T^{1523} \rightarrow A$	$T^{1523} \rightarrow A$	$T^{1523} \rightarrow A$	No			
dSNPs	–	–	$T^{1560}T^{1561} \rightarrow AG$	$T^{1560}T^{1561} \rightarrow AG$	$T^{1560}T^{1561} \rightarrow AG$	G	P$TBPF/TATA.01	+	T1
FR	–	–	C^{1564}ACATTCGTAT^{1574} → AAATGATATT	C^{1564}ACATTCGTAT^{1574} → AAATGATATT	C^{1564}ACATTCGTAT^{1574} → AAATGATATT	G	P$GTBX/SBF1.01	+	B1
						G	P$NCS1/NCS1.01	+	N
SNP	–	–	$C^{1577} \rightarrow A$	$C^{1577} \rightarrow A$	$C^{1577} \rightarrow A$	No			
SNP	–	–	$T^{1584} \rightarrow A$	$T^{1584} \rightarrow A$	$T^{1584} \rightarrow A$	No			
SNP	–	–	$T^{1600} \rightarrow C$	$T^{1600} \rightarrow C$	$T^{1600} \rightarrow C$	No			
SNP	–	$T^{1605} \rightarrow G$	$T^{1605} \rightarrow G$	$T^{1605} \rightarrow G$	$T^{1605} \rightarrow G$	L	P$IBOX/GATA.01	−	I
						L	P$DOFF/PBOX.01	−	D2
SNP	–	–	$T^{1625} \rightarrow A$	$T^{1625} \rightarrow A$	$T^{1625} \rightarrow A$	G	P$AHBP/ ATHB5.01	−/+	A
						G	P$GTBX/SBF1.01	+	B1
Indel	–	–	A^{1628}CCT$^{1631} \rightarrow$ AT	A^{1628}CCT$^{1631} \rightarrow$ AT	A^{1628}CCT$^{1631} \rightarrow$ AT	G	P$AHBP/ ATHB5.01	−/+	A
						G	P$GTBX/SBF1.01	+	B1
						L	P$WBXF/ERE.01	+	W1
SNP	$C^{1630} \rightarrow A$	$C^{1630} \rightarrow A$	–	–	–	L	P$NCS1/NCS1.01	+	N
						L	P$WBXF/ERE.01	+	W1
Indel	–	–	$A^{1674}A^{1675} \rightarrow ACA$	–	–	G	P$SALT/ ALFIN1.01	−	SA
						L	P$CCAF/CCA1.01	+	C
Indel	–	–	$T^{1691}CG^{1693} \rightarrow TG$	–	–	No			

Table 2 continued

Variation	Haplotype					TFBSs fate as a result of variation comparing with Haruna Nijo haplotype			
	HN	HS	HA	AB	AD	Fate	Family/matrix × number	Str. +/−	Abbr.
SNP	–	–	–	–	$T^{1700} \rightarrow G$	G	P$MYCL/MYCRS.01	–	MY
SNP	–	–	$G^{1744} \rightarrow C$	$G^{1744} \rightarrow C$	$G^{1744} \rightarrow C$	No			
SNP	–	–	$C^{1746} \rightarrow T$	$C^{1746} \rightarrow T$	$C^{1746} \rightarrow T$	G	P$GBOX/TGA1A.01	–	G1
Indel	–	–	A^{1757}-11 bp-$C^{1769} \rightarrow AC$	A^{1757}-11 bp-$C^{1769} \rightarrow AC$	$A^{1757}T$-11 bp-$C^{1769} \rightarrow AC$	No			
SNP	–	–	$C^{1772} \rightarrow T$	$C^{1772} \rightarrow T$	$C^{1772} \rightarrow T$	No			
SNP	–	–	$T^{1775} \rightarrow A$	$T^{1775} \rightarrow A$	$T^{1775} \rightarrow A$	No			
SNP	–	–	$A^{1782} \rightarrow T$	$A^{1782} \rightarrow T$	$A^{1782} \rightarrow T$	No			
SNP	–	–	$T^{1792} \rightarrow G$	$T^{1792} \rightarrow G$	$T^{1792} \rightarrow G$	No			
SNP	–	–	$G^{1798} \rightarrow C$	$G^{1798} \rightarrow C$	$G^{1798} \rightarrow C$	No			
dSNPs	–	–	$C^{1800}A^{1801} \rightarrow TG$	$C^{1800}A^{1801} \rightarrow TG$	$C^{1800}A^{1801} \rightarrow TG$	No			
SNP	–	–	$T^{1818} \rightarrow C$	$T^{1818} \rightarrow C$	$T^{1818} \rightarrow C$	L	P$NCS1/NCS1.01		N
Indel	–	–	$C^{1849}C^{1850} \rightarrow C^{1849}$-4 bp-$C^{185°}$	$C^{1849}C^{1850} \rightarrow C^{1849}$-4 bp-$C^{185°}$	$C^{1849}C^{1850} \rightarrow C^{1849}$-4 bp-$A^{185°}$	No			
SNP	–	–	–	–	$C^{1850} \rightarrow A$[a]	No			
Indel	–	$C^{1862}CA^{1864} \rightarrow CA$	–	–	–	No			
Indel	–	–	G^{1881}-21 bp-$A^{1902} \rightarrow GA$	G^{1881}-21 bp-$A^{1902} \rightarrow GA$	–	G	P$MYBS/OSMYBS.01	–	M1
						G	P$GTBX/S1F.01	+	B2
						L	P$L1BX/ATML1.01	–	L
						L	P$TBPF/TATA.02	–	T2
SNP	–	–	–	–	$G^{1882} \rightarrow A$	No			
SNP	–	–	$A^{1918} \rightarrow G$	$A^{1918} \rightarrow G$	$A^{1918} \rightarrow G$	No			
dSNPs	–	–	$T^{1927}T^{1928} \rightarrow GG$	$T^{1927}T^{1928} \rightarrow GG$	$T^{1927}T^{1928} \rightarrow GG$	No			
Indel	–	–	–	–	$A^{1943}T^{1944} \rightarrow AAT$[b]	No			
SNP	–	–	$A^{1962} \rightarrow T$	$A^{1962} \rightarrow T$	–	L	GARP/ARR10.01	+	GR
						L	MYCL/ICE0.1 × 2	+	ML
SNP	–	–	–	–	$T^{1967} \rightarrow G$	L	MYCL/ICE0.1	+	ML
MS	HN	HS	HA	AB	AD	L	SALT/ALFIN1.01	+	SA
						G	CAAT/CAAT.01	+	CA
						G	MYBL/GAMYB.01	+	MG
						G	MIIG/NTLIM1.01	+	MI
SNP	–	–	$T^{2001} \rightarrow G$	$T^{2001} \rightarrow G$	$T^{2001} \rightarrow G$	L	SALT/ALFIN1.01	+	SA
SNP	–	$G^{2002} \rightarrow T$	–	–	–	G	MYBL/GAMYB.01	+	MG
Indel	–	–	–	$G^{2002}ATGC^{2006} \rightarrow G^{2002}C^{2006c}$	–	L	P$EINL/TEIL.01	+	E
Indel	–	$A^{2003}TG^{2005} \rightarrow AG$	–	–	–	L	P$EINL/TEIL.01	+	E
						G	P$MYCL/MYCRS.01	–	MY
Indel	–	–	–	–	$A^{2013}T^{2014} \rightarrow AAT$[b]	No			
SNP	–	–	$G^{2019} \rightarrow T$	$G^{2019} \rightarrow T$	$G^{2019} \rightarrow T$	No			
SNP	$C^{2020} \rightarrow T$	$C^{2020} \rightarrow T$	$C^{2020} \rightarrow T$	$C^{2020} \rightarrow T$	$C^{2020} \rightarrow T$	No			
SNP	–	–	$T^{2041} \rightarrow C$	–	–	No			
SNP	$C^{2044} \rightarrow G$	$C^{2044} \rightarrow G$	$C^{2044} \rightarrow G$	$C^{2044} \rightarrow G$	$C^{2044} \rightarrow G$	No			

Table 2 continued

Variation	Haplotype					TFBSs fate as a result of variation comparing with Haruna Nijo haplotype			
	HN	HS	HA	AB	AD	Fate	Family/matrix × number	Str. +/−	Abbr.
SNP	–	$G^{2055} \rightarrow A$	–	–	–	No			
SNP	–	–	$A^{2061} \rightarrow G$	$A^{2061} \rightarrow G$	$A^{2061} \rightarrow G$	G	P$LEGB/RY.01	+	LE
SNP	–	–	$T^{2066} \rightarrow C$	$T^{2066} \rightarrow C$	$T^{2066} \rightarrow C$	No			
Indel/SNP	–	–	$G^{2071}CT^{2073} \rightarrow GT$	$G^{2071}CT^{2073} \rightarrow GT$	$C^{2072} \rightarrow T^{b}$ $G^{2071}CT^{2073} \rightarrow GT$	No			
SNP	–	–	$T^{2105} \rightarrow A$	$T^{2105} \rightarrow A$	$T^{2105} \rightarrow A$	No			
SNP	–	–	$G^{2162} \rightarrow A$	$G^{2162} \rightarrow A$	$G^{2162} \rightarrow A$	No			
Indel	–	–	$A^{2176}T^{2177} \rightarrow AAT$	$A^{2176}T^{2177} \rightarrow AAT$	$A^{2176}T^{2177} \rightarrow AAT$	G	P$MADS/MADSA.01	+	M
						G	P$PSRE/GAAA.01	+	P
Indel	–	–	$T^{2186}T^{2187}$ $\rightarrow T^{2186}$-6 bp-T^{2187}	$T^{2186}T^{2187}$ $\rightarrow T^{2186}$-6 bp-T^{2187}	$T^{2186}T^{2187}$ $\rightarrow T^{2186}$-6 bp-T^{2187}	No			
SNP	–	–	–	–	$A^{2219} \rightarrow T$	No			

[a] Variation $C^{1850} \rightarrow A$ is the character of Adorra and AD-like sequence variants besides Druvis, Idumeja, and Imula

[b] Variations $A^{1943}T^{1944} \rightarrow AAT$, $A^{2013}T^{2014} \rightarrow AAT$, $C^{2072} \rightarrow T$ belongs only to Adorra intron III sequence but not to AD-like variants

[c] Deletion $G^{2002}ATGC^{2006} \rightarrow G^{2002}C^{2006}$ was revealed only in the intron III sequence of the cultivar Ilga

between cultivars in repeat number, ranging from $n = 12$ (Ansis, Balga, Kombainieris, Linga, Rasa) to $n = 18$ (Idumeja, Imula, and Priekulu 1, Table 1).

Intron III of the *Bmy1* gene of other eight Latvian varieties studied were classified as novel AB haplotype (Table 1; Fig. 2) with MS motif $(TG)_6TT(G)_n$. As for the AD haplotype, the MS component $(G)_n$ was revealed as highly polymorphic in repeat number, ranging from $n = 14$ (Ilga) till $n = 18$ (Agra). Absence of three insertions A^{1310}-126 bp-T^{1311}, $A^{1943}AT^{1944}$, and $A^{2013}AT^{2014}$, as well as the presence of alleles A^{1318}, T^{1700}, C^{1850}, G^{1882}, T^{1967}, A^{2219} are the AB features common with HN, HS, HA haplotypes in contrast to AD. Alternatively, alleles A^{1384} and G^{1605} are common with HS, HA, and AD haplotype in contrast to HN. The presence of three insertions A^{1389}-38 bp-T^{1426}, $C^{1862}CA^{1864}$, $A^{2003}TG^{2005}$, and alleles G^{2002} and G^{2055}, are similar to HN, HA, and AD in contrast to HS. Absence of the 21-bp fragment G^{1881}-21 bp-T^{1902} and three alleles T^{1350}, G^{1505}, and T^{1962} are features common only to HA in contrast to HN, HS, and AD haplotypes. Alternatively, indel variants $A^{1674}A^{1675}$, and $T^{1691}CG^{1693}$, and allele T^{2041} are features common with HN, HS, and AD, in contrast to HA. Thirty-one mutations including six indels ($A^{1628}CCT^{1631} \rightarrow AT$, A^{1757}-11 bp-$C^{1769} \rightarrow AC$, $C^{1849}C^{1850} \rightarrow$ C-4 bp-C, $G^{2071}CT^{2073} \rightarrow GT$, $A^{2176}T^{2177} \rightarrow AAT$, and $T^{2186}T^{2187} \rightarrow$ T-6 bp-T), one fragment substitution $C^{1564}ACATTCGTAT^{1574} \rightarrow$ AAATGATATT, and twenty-four SNPs/dSNPs were revealed as identical between AB and both HA and AD haplotypes (Table 2; Fig. 3).

Interhaplotype variability

Table 3 summarizes the publicly available and our present data on interhaplotype variability of the *Bmy1* allelic variants focusing on several exon III and exon IV cSNPs as well as on intron III polymorphisms. Five cSNP were taken into account including three synonymous and two non-synonymous substitutions previously used in *Bmy1* allele classification (Paris et al. 2002) and in several genotyping studies (Polakova et al. 2003; Sjakste and Röder 2004; Malisheva et al. 2004). Four exonic haplotypes including $G^{495}/T^{666}/T^{698}/G^{702}/A^{741}$, $C^{495}/T^{666}/T^{698}/A^{702}/A^{741}$, $C^{495}/C^{666}/C^{698}/A^{702}/G^{741}$, and $G^{495}/T^{666}/C^{698}/A^{702}/G^{741}$ are considered to belong to Sd1, Sd2L, Sd2H, and Sd3 *Bmy1* alleles correspondingly (Table 3). However, the newly described haplotypes $G^{495}/N^{666}/T^{698}/A^{702}/N^{741}$ (Polakova et al. 2003) and $N^{495}/C^{666}/C^{698}/G^{702}/G^{741}$ (Sjakste and Röder 2004) suggest more allelic variants of the gene. Exonic haplotype $N^{495}/T^{666}/T^{698}/G^{702}/A^{741}$ belongs to accessions of both AB-like (varieties Abava and Agra) and

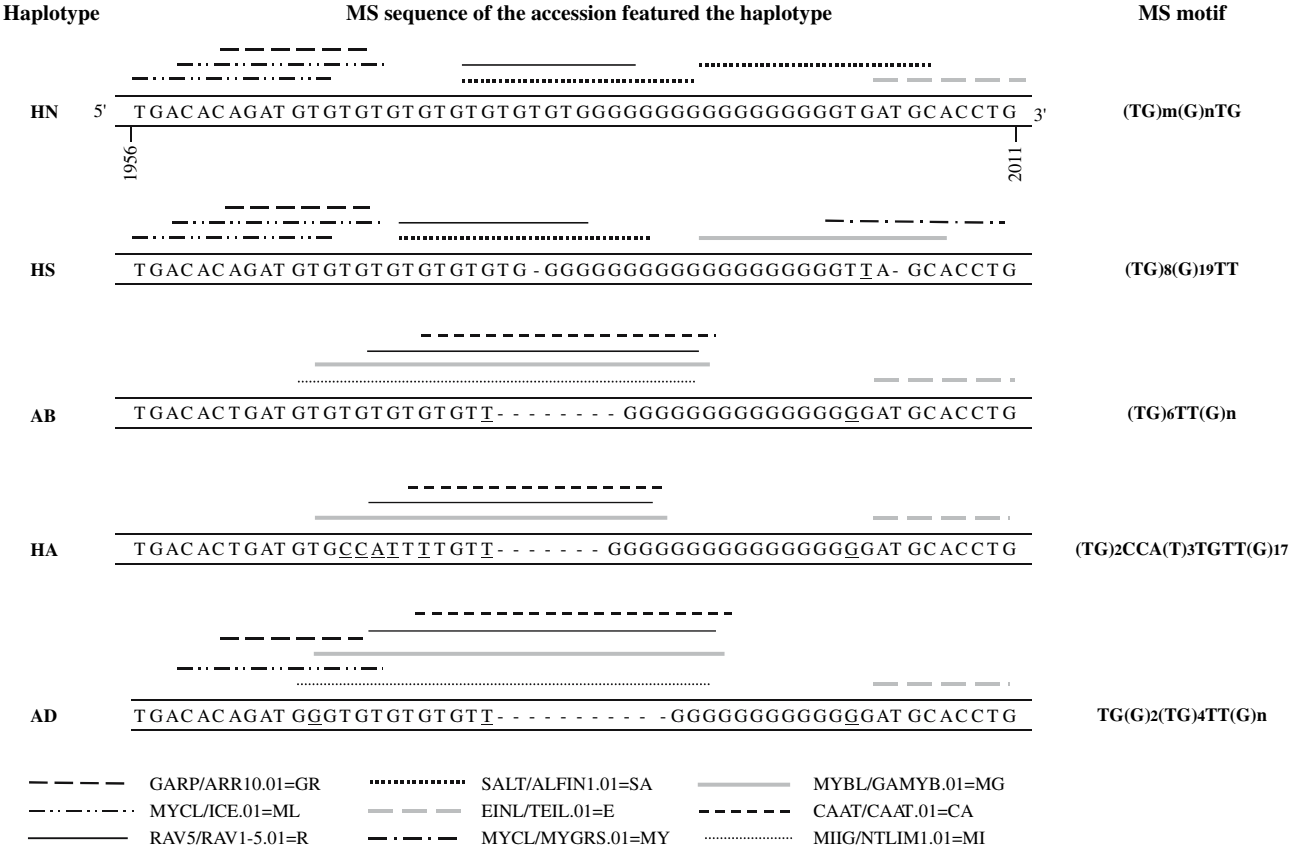

Fig. 2 β-Amylase intron III MS sequence alignment, formulas of the haplotype-specific MS motifs and the eventual TFBSs position. HN, HS, HA, and AD haplotypes are represented by the intronic sequences of the β-amylase gene of Haruna Nijo (D49999), *H. vulgare* subsp. *spontaneum* NPGS PI 296897 (AF061204), Finnish landrace line HA52 (AJ301645), and Adorra (AF061203) correspondingly. AB haplotype is represented by the intron III sequence obtained for Latvian cultivar Abava. The numbering of the MS position is indicated according to the genomic sequence of accession Haruna Nijo (D49999). SNPs determined the differences in MS motif compared to HN haplotype are *underlined*. Predicted TFBSs are shown by the *individual lines* of different types under the MS sequences. Description of *TF* is given in Table 4

to HA52 haplotypes of intron III. Alternatively, intronic HN and HS haplotypes belong to accessions of the same exonic haplotype $C^{495}/C^{666}/C^{698}/A^{702}/G^{741}$.

Data presented in Table 3 illustrate the insufficiency of the *Bmy1* allele discrimination based only on two cSNPs in positions 495 and 698, as well as the necessity to take into consideration the structural rearrangements in the non-coding gene regions.

Immediate pedigree analysis

Data on several pedigree stories of Latvian barley varieties (Table 1) permit analysis of transmittance of the intron III polymorphisms through the generations to be performed. For example, Balga and Linga both descended from the same cross Gunilla/KM 1192 and possess the same MS $TG(G)_2(TG)_4TT(G)_{12}$ and whole intron III AD-like haplotype. Abava, as one of the immediate parents, passes its haplotype to Ruja. Three of the Latvian varieties analyzed including Abava and Ilga of AB haplotype, and Imula of AD haplotype are involved in the Idumeja (AD haplotype) pedigree (Fig. 4). The scheme illustrates the transmittance of AD-like haplotype from Imula to Idumeja. The presence of 126 bp insertion as the main AD haplotype feature allows us to speculate that Akka also has AD haplotype.

Functional analysis

Comparative analysis was performed to identify both conservative and variable functional elements in the intron III of the *Bmy1* gene.

Consensus and conserved sequences

No polymorphisms were revealed in splice donor and branch sites of intron III. AD haplotype appeared to have one SNP $A^{2219} \rightarrow T$ in the splice acceptor site, unlike other haplotypes studied.

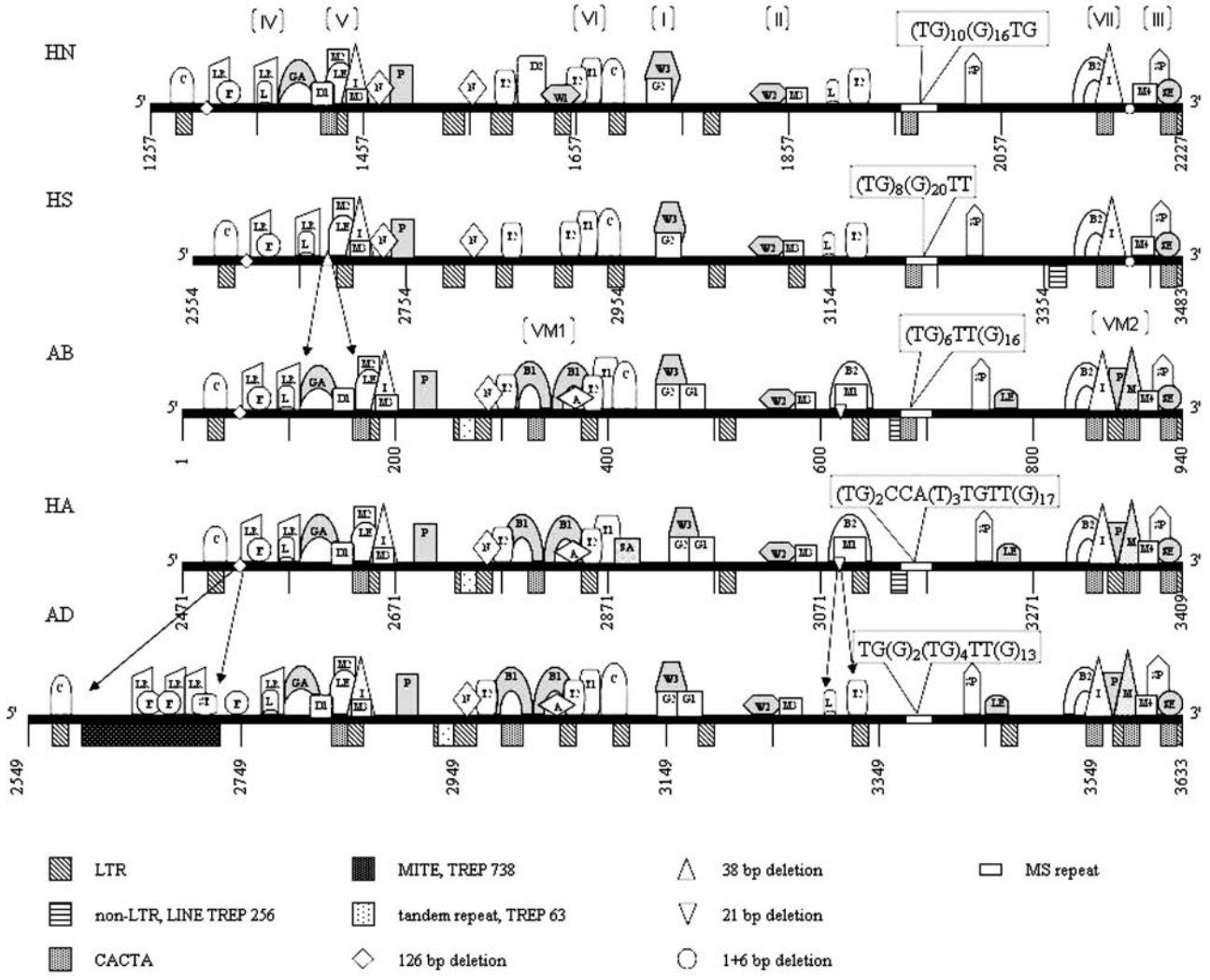

Fig. 3 Haplotype-specific co-localization of the intron III polymorphisms, TFBSs and remnants of mobile elements. HN, HS, AB, HA, and AD haplotypes are represented by the same accessions as in Fig. 2. Numbering is indicated in base pairs according to the genomic sequences of the mentioned accessions, sequence of AB haplotype is numbered starting from the first G of the intron. Each TFBS site is shown as a *symbol* with inserted abbreviated name. Correspondence of abbreviated names to the full TF names is explained in Table 4. Predicted positions of the conservative (*I–VII*) and variable (*VM1* and *VM2*) modules are indicated in *square brackets*. Three pairs of *arrows* compare the 126, 38, and 21 bp indel loci between different haplotypes. MS motif sequences are given in frames

Several other highly conserved sequences of different length were derived from the alignment (see Table 3; Fig. 3). Those longer than 30 bp included a 53 bp sequence in the 5′ flank of the intron upstream the site of the 126 bp indel, regions from A^{1427} till A^{1504}, from T^{1631} till A^{1674}, from T^{1701} till A^{1743}, and from T^{2187} till A^{2218}. Only two SNPs were revealed in the 73 bp region from T^{1311} till A^{1383} (Table 2).

Analysis on the TFBSs

Sequences of all five haplotypes studied were analyzed for the eventual presence of TFBS using the Genomatix software. Binding sites (BSs) for 37 plant transcription factors (TFs) of 27 families were taken into account according to the chosen parameters (core/matrix similarity≥0.85/0.85). Comments on TFs family/matrix information and their abbreviations (used in Fig. 3) are given in Table 4. The family/matrix name and its associated abbreviation (MYBST1.01 = M2, for example) are sometimes used in the text to avoid misunderstanding.

Three of the predicted TFs belong each to the W Box family (WBXF) and the MYB-like proteins (MYBL). Two TFs represent each the plant GT-box elements (GTBX), DNA binding with one finger (DOFF), G-box/C-box BZIP proteins (GBOX), MYB proteins with single DNA-binding domains (MYBS),

Table 3 Summary of the previously published and present data on the interhaplotype variability of the *Bmy1* allelic variants

Exon III/exon IV five SNPs haplotype	Reference on SNPs data		DNA source	Intron III haplotype features on MS motif and indel presence (+)/absence (-)					
	Allele/cultivar	GenBank, author		MS motif	126 bp	38 bp	21 bp	$A^{1674}CA^{1675}$	1+6 bp
$G^{495}/T^{666}/T^{698}/G^{702}/A^{741}$	Sd1, Franklin, Morex	AB048949, AF300800, AF414081	cDNA						
$G^{495}/T^{666}/T^{698}/G^{702}/A^{741}$	HA52	AJ301645	Genomic	HA	–	+	+	+	+
$N^{495}/T^{666}/T^{698}/G^{702}/A^{741}$	Abava[a], Agra[a], Kenia, Tammi	Sjakste and Röder (2004)	Genomic	AB	–	+	+	–	+
$G^{495}/N^{666}/T^{698}/G^{702}/N^{741}$	Cultivars of Sd1b[b]	Polakova et al. (2003)	Genomic						
$G^{495}/N^{666}/T^{698}/A^{702}/N^{741}$	Cultivars of Sd1[b]	Polakova et al. (2003)	Genomic						
$C^{495}/T^{666}/T^{698}/A^{702}/A^{741}$	Adorra	AF061203	Genomic	AD	+	+	–	–	+
$C^{495}/T^{666}/T^{698}/A^{702}/A^{741}$	Sd2L, Stepto, Orca, Schooner	X52321, AF414082, AY54398, AF300799	cDNA						
$N^{495}/T^{666}/T^{698}/A^{702}/A^{741}$	Akka, Imula[a], Linga[a], Vega	Sjakste and Röder (2004)	Genomic	AD	+	+	–	–	+
$C^{495}/N^{666}/T^{698}/A^{702}/N^{741}$	Sd2L	Polakova et al. (2003)	Genomic						
$C^{495}/C^{666}/C^{698}/A^{702}/G^{741}$	Sd2H	D21349, E08321	cDNA						
$C^{495}/C^{666}/C^{698}/A^{702}/G^{741}$	Haruna Nijo	D49999	Genomic	HN	–	+	–	–	–
$C^{495}/C^{666}/C^{698}/A^{702}/G^{741}$	*H. spont.*	AF061204	Genomic	HS	–	–	–	–	–
$N^{495}/C^{666}/C^{698}/A^{702}/G^{741}$	Denso, Rigel, Svanhals, Latvijas Vietieje[a]	Sjakste and Röder (2004)	Genomic	HN	–	+	–	–	–
$G^{495}/T^{666}/C^{698}/A^{702}/G^{741}$	Sd3	AF012343	cDNA						
$N^{495}/T^{666}/C^{698}/A^{702}/G^{741}$	Monte Cristo	Sjakste and Röder (2004)	Genomic						
$N^{495}/C^{666}/C^{698}/G^{702}/G^{741}$	Drost	Sjakste and Röder (2004)	Genomic						

[a] Indicates accessions characterized both on exon IV cSNPs (adapted from Sjakste and Röder 2004) and intron III polymorphisms

[b] Classification was used in Polakova et al. (2003)

Table 4 Description of the TFs which possess BSs in intron III of the *Bmy1* gene

Family	Family information	Matrix name	Matrix information	No. of generated	No. of lost	Abbreviation
P$AHBP	*Arabidopsis* homeobox protein	ATHB5.01	HD Zip class I protein ATHB5	4	1	A[a]
P$GTBX	GT-box elements	SBF1.01	SBF-1	3	No	B1[a]
		S1F.01	S1F, site 1 binding factor of spinach rps1 promoter	1	No	B2[b]
P$CAAT	CCAAT binding factors	CAAT.01	CCAAT-box in plant promoters	1	No	CA[a]
P$CCAF	Circadian control factors	CCA1.01	Circadian clock associated 1	No	1	C[b]
P$DOFF	DNA binding with one finger (DOF)	PBF.01	PBF (MPBF)	1	1	D1[a]
		PBOX.01	Prolamin box, conserved in cereal seed storage protein gene promoters	No	1	D2[a]
P$EINL	Ethylene insensitive 3 like factors	TEIL.01	TEIL (tobacco EIN3-like)	3	2	E[b]
P$GBOX	Plant G-box/C-box BZIP proteins	TGA1A.01	TGA1A	1	No	G1[a]
		CPRF2.01	Constitutively expressed; light-induced nuclear import	No	No	G2[c]
P$GAPB	GAP-Box (light response elements)	GAP.01	*Cis*-element in the GAPDH promoters conferring light inducibility	No	1	GA[a]
P$GARP	GAP-Box (light response elements)	ARR10.01	Type-B response regulator (ARR10), member of the GARP-family of plant myb-related DNA-binding motifs	No	1	GR[a]
P$IBOX	Plant I-Box sites	GATA.01	Class I GATA factors	No	1	I[b]
P$L1BX	L1 box, motif for L1 layer-specific expression	ATML1.01	L1-specific homeodomain protein ATML1 (*Arabidopsis thaliana* meristem layer 1)	No	1	L[b]
P$LEGB	Legumin Box family	RY.01	RY and Sph motifs conserved in seed-specific promoters	1	No	LE[b]
P$LREM	Light responsive element motif, not modulated by different light qualities	ATCTA.01	Motif involved in carotenoid and tocopherol biosynthesis and in the expression of photosynthesis-related genes	3	1	LR[b]
P$MADS	MADS-box proteins	MADSA.01	Determines identity of floral meristem and sepal and petal development	1	No	M[a]
P$MIIG	MYB II-type BSs	NTLIM1.01	LIM domain protein binding to a PAL-box like sequence	1	No	MI[a]
P$MYBL	MYB-like proteins	AtMYB77.01	R2R3-type myb-like TF (I-type BS)	No	No	M3[c]
		GAMYB.01	GA-regulated myb gene from barley	2	No	MG[a]
		MYBPH3.01	Myb-like protein of *Petunia hybrida*	No	No	M4[c]
P$MYBS	MYB proteins with single DNA-binding domains	OSMYBS.01	Rice MYB proteins with single DNA-binding domains, binding to the amylase element (TATCCA)	1	No	M1[a]
		MYBST1.01	MybSt1 (Myb *Solanum tuberosum* 1) with a single myb repeat	No	No	M2[c]
P$MYCL	Myc-like binding factors	ICE.01	ICE (inducer of CBF expression 1), AtMYC2 (rd22BP1)	No	2	ML[a]
		MYCRS.01	Myc recognition sequences	2	1	MY[a]
P$NCS1	Nodulin consensus sequence 1	NCS1.01	Nodulin consensus sequence 1	1	4	N[b]
P$PSRE	Pollen-specific regulatory elements	GAAA.01	GAAA motif involved in pollen-specific transcriptional activation	1	No	P[b]
P$RAV5	5′-part of bipartite RAV1 BS	RAV1-5.01	5′-part of bipartite RAV1 BS, interacting with AP2 domain	No	No	R[c]
P$SALT	Salt/drought responsive elements	ALFIN1.01	Zinc-finger protein in alfalfa roots, regulates salt tolerance	1	1	SA[a]

Table 4 continued

Family	Family information	Matrix name	Matrix information	No. of generated	No. of lost	Abbreviation
P$SEF4	Soybean embryo factor 4	SEF4.01	Soybean embryo factor 4	No	No	SE[c]
P$SPF1	Sweet potato DNA-binding factor with two WRKY-domains	SP8BF.01	DNA-binding protein of sweet potato that binds to the SP8a (ACTGTGTA) and SP8b (TACTATT) sequences of sporamin and β-amylase genes	No	No	SP[c]
P$STKM	Storekeeper motif	STK.01	Storekeeper (STK), plant specific DNA-binding protein important for tuber-specific and sucrose-inducible gene expression	1	No	ST[a]
P$TBPF	TATA-binding protein factor	TATA.01	Plant TATA box	1	No	T1[b]
		TATA.02	Plant TATA box	No	1	T2[a]
P$WBXF	W Box family	ERE.01	Elicitor response element	No	2	W1[a]
		WRKY.01	WRKY plant specific zinc-finger-type factor associated with pathogen defense, W box	No	No	W2[c]
		ZAP1.01	*A. thaliana* zinc-dependent activator protein-1 (ZAP1)	No	No	W3[c]

Information is given according to the prediction by Genomatix Matrix Family Library (Version 5.0). Core/matrix similarity was taken as ≥0.85/0.85. Abbreviation of the family/matrix name was used in Fig. 3 for TFBSs indication

[a] Variable TFBSs

[b] Dual TFBSs

[c] Conservative TFBSs

Myc-like binding factors (MYCL), and TATA-binding protein factor (TBPF). Other families are represented only by one member (circadian control factor CCAF/CCA1.01, for example).

All TFBSs revealed could be divided into three groups. Group 1 is formed by BSs found exclusively in the conservative non-polymorphic DNA sequences (conservative TFBSs). Group 2 unites TFBSs that arise or disappear due to polymorphisms (variable TFBSs), and group 3 – BSs found both in conservative and variable regions of the intron (dual TFBSs).

The locations of the conservative, dual and some variable TFBSs within the intron III sequence of a particular haplotype are illustrated in Fig. 3. All other variable TFBSs (generated or lost as a result of sequence variations) are indicated in Table 2.

Conservative TFBSs

Eventually, eight types of conservative TFBSs could be found in portions of the intron III invariable between haplotypes (Table 4; Fig. 3) including TFBSs for the constitutively expressed BZIP protein responsible for light-induced nuclear import (CPRF2.01 = G2), Myb and Myb-like proteins (MYBST1.01 = M2, MYBPH3.01 = M4, and two sites for ATMYB77.01 = M3), soybean embryo factor 4 (SEF4.01 = SE), two sites for protein that binds SP8a and SP8b sequences of sporamin and β-amylase genes (SP8BF.01 = SP), and proteins of the W Box family (WRKY.01 = W2 and ZAP1.01 = W3). Besides the mentioned, conservative TFBS independent from interhaplotype variability of the region was predicted in the MS portion of intron for RAV1-5.01 (Fig. 2). Different TFs could interact with the same conservative regions of intron producing conservative modules consisting of several elements. Modules I and II were predicted to consist each of two conservative TFs (G2 and W3, and M3 and W2 correspondingly). The highly conservative 3′-end of the intron possesses binding capacity to three TFs of module III (SE, M4, and SP, Fig. 3).

Ten eventual TFBSs were evaluated as dual ones (Table 4). Haplotype independent dual TFBSs are illustrated in Fig. 3 including BSs of the circadian clock associated control factor 1 (CCA1.01 = C) and an ethylene insensitive factor (TEIL.01 = E) that flanks the 126 bp indel locus from its 5′- and 3′-ends correspondingly. The intron 5′-flank possesses the binding capacity to module IV (BSs of ATCTA.01 = LR and ATML1.01 = L), module V (two dual TFBSs of RY.01 = LE and GATA.01 = I, accompanied by two conservative TFBSs of MYBST1.01 = M2 and ATMYB77.01 = M3), and to GAAA motif element

involved in pollen-specific transcriptional activation (GAAA.01 = P). This binding capacity is present in all the five haplotypes. Further in the 3′ intron direction, conservative binding capacity was predicted for TFBSs of nodulin consensus sequence (NCS1.01 = N), plant TATA box (TATA.02 = T2), module VI (of two TATA-binding proteins), and module VII (GATA.01 = I and S1F.01 = B2).

Variation in the TFBSs pattern

The functional significance of all the polymorphisms revealed was analyzed by Genomatix software. Data were scored as generation or loss of TFBS sites as a result of sequence polymorphism, summarized in Table 2 and partly illustrated in Fig. 3. Among polymorphic events analyzed, at least 24 variations could change the binding capacity of the 10 dual and 18 variable TFBSs correspondingly (Table 4). Three indels of 126, 38, and 21 bp significantly changed the eventual pattern of TFBSs. The MITE element (126 bp insertion) adds six new BSs including three light responsive elements LR, two ethylene insensitive factors E and one storekeeper motif ST. Graphically, it looks like an expansion of LR + E modules in AD when compared to other haplotypes (Fig. 3). The 38 bp deletion is the feature of the HS haplotype and is responsible for the loss of TFBSs of the light response element (GAP.01 = GA), DOF factor PBF.01 = D1 (Fig. 4), and nodulin consensus sequence NCS1.01 = N (Table 2). The 21 bp insertion (present in AB and HA haplotypes) could generate additional TFBSs of both GT-box element (S1F.01 = B2) and MYB protein (OSMYBPH3.01 = M1) simultaneously with the loss of sequence affinity to the L1 box (ATML1.01 = L) and one of the plant TATA boxes (TATA.02 = T2). Deletion $A^{1628}CCT^{1631}$ and insertion $A^{2176}AT^{2177}$ results in the generation of new variable modules (VM1 and VM2 correspondingly) in AB, HA, and AD haplotypes (Fig. 3). Insertion $A^{1674}CA^{1675}$ results in eventual generation only in AB haplotype of the TFBS of the salt/drought responsible element ALFIN1.01 = SA.

In contrast to the mentioned mutations, the A^{1757}-11 bp-C^{1769} deletion, C^{1849}-4 bp-C^{1850} and T^{2186}-6 bp-T^{2187} insertion does not seem to be followed by the changes in sequence affinity to TFs (Table 2) in AB, HA, and AD haplotypes compared with HN and HS.

Eventual function of the MS variations

As mentioned above, the binding capacity to RAV1-5.01 was predicted for MS of all haplotypes studied independently of the restructuring of this portion of the intron III. The binding capacity to other TFs depends on the spectrum of sequence variations (Table 2) and determines the individual pattern of TFBSs for each haplotype studied (Fig. 2). For example, variations $T^{2001} \rightarrow G$ leads to the elimination of BSs with type-B response regulator (ARR10.01) of GARP-family and two inducers of CBF expression (ICE.01) in AB and HA haplotypes compared with the HN and HS variants. The presence of the TT motif in the border between $(TG)_m$ and $(G)_n$ components of MS eliminate binding capacity to the salt/drought responsible zinc-finger protein (ALFIN1.01) in favor of the CAAT box (CAAT.01), and MYB plant protein (GAMYB.01) in AB, HA, and AD haplotypes in comparison with HN and HS sequences. The 3′-flank of the HN haplotype possesses binding capacity to both salt/drought responsible zinc-finger protein (ALFIN1.01) and the ethylene insensitive factor TEIL.01. ALFIN1.01-site disappears in AB, HA, and AD haplotypes due to $T^{2001} \rightarrow G$ substitution. Restructuring of the 3′ MS flank in HS haplotype leads to the loss of both ALFIN1.01 and TEIL.01 BSs and the appearance of the binding capacity to GAMYB.01 and MYCRS.01.

Variations in the number of $(TG)_m$ and $(G)_n$ repeats do not produce significant differences in the affinity of MS sequence to the plant TFBSs according to the Genomatix database. However, interesting information was obtained from alignment with vertebrate TFBSs (not shown). The number of $(G)_n$ repeats (accessions of HN, AB, and AD haplotypes) could change the spectrum and/or number of BSs with different zinc-finger proteins. In contrast, intrahaplotype variations in the number of TG-repeats seems not to influence the TFs binding capacity of the region (Haruna Nijo, Maja, Latvijas Vietejie were compared).

Mobile element presentation

The position of the MITE element (126 bp indel) and the remnants of different retrotransposons within the intron III are illustrated in Fig. 3 for all the five haplotypes studied. The presence of the MITE element is a characteristic of AD haplotype only. The remnants of the retrotransposones of the several classes were identified within highly conserved regions of the intron III. Their sequence motifs and positions are identical between all haplotypes studied (remnants of the LTR gypsy position in 5′- and CACTA in 3′-flank of the intron sequence correspondingly). Spectrum and positions of remnants of other mobile elements were revealed as variable between haplotypes studied.

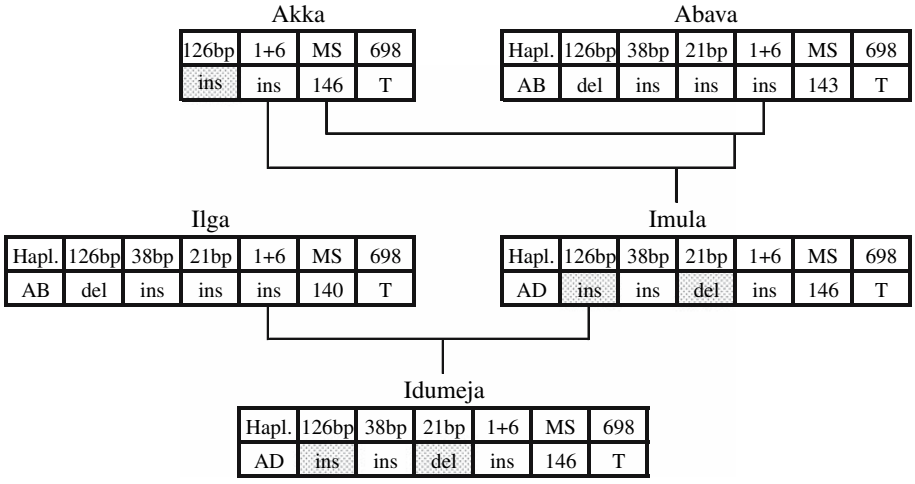

Fig. 4 Pedigree of the Latvian barley cultivar Idumeja. *Bmy1* intron III haplotype (*Hapl.*) together with allelic variants of indels are indicated according to the present data for cultivars Abava, Ilga, Imula, and Idumeja. Data on MS length, 1 + 6 bp indel, and $C^{698} \rightarrow T$ polymorphism of all the pedigree participants were published previously (adapted from Sjakste and Röder 2004), information on 126 bp insertion of the intron III of the *Bmy1* gene of cultivar Akka is present for the first time. The 126 bp insertion and the 21 bp deletion in AD haplotype are marked in *gray* to underline the transmittance of the polymorphic loci through generations

Discussion

Description of novel haplotypes

One of the goals of the current study was to extend data on polymorphisms of the intron III of the *Bmy1* structural gene. Besides the four haplotypes reported earlier (Haruna Nijo, Adorra, Finnish landrace HA52, and *H. spontaneum*), we have described a new AB-like haplotype of intron III. The absence of the 126 bp MITE element, three indels, nine SNPs, and specific MS motif belongs to polymorphisms that discriminate AB and AD haplotypes. Simultaneously, 44 other polymorphisms analyzed were revealed as shared between AB and AD haplotypes. These include $A^{2176}T^{2177} \rightarrow$ AAT, and $T^{2186}T^{2187} \rightarrow$ T-6 bp-T insertions in 3′ end sequences of both haplotypes previously reported as 1 + 6 bp insertion or L-allele linked to the T^{698} mutation of exon IV (Sjakste and Röder 2004). Allelic variants of other polymorphic loci of AB haplotype were found to be common with HN, HA, and HS haplotypes in different combinations. All together 59 polymorphisms including MS variations produce the unique *Bmy1* intron III AB haplotype. It should be considered that even more alleles and haplotypes could probably be detected by the analysis of a different collection of genotypes from over the world.

Haplotype classification

Paris et al. in 2002 proposed the high-throughput genotyping system for identification of the *Bmy1* allele taking into consideration two non-synonymous cSNPs in positions 495 and 698 of the *Bmy1* gene. This approach was successfully applied in several genotyping studies (Polakova et al. 2003; Sjakste and Röder 2004; Malisheva et al. 2004). However, identification and genotyping of additional cSNPs (Polakova et al. 2003; Sjakste and Röder 2004) revealed insufficiency of classification based only on two polymorphic loci. Polakova et al. (2003) subdivided the Sd1 allele in Sd1 and Sd1b. Exon IV haplotype $C^{666}/C^{698}/G^{702}/G^{741}$ of variety Drost. Sjakste and Röder (2004) suggest an additional *Bmy1* allelic variant different from SD2H. Assuming the present data on intron III interhaplotype polymorphism, we confirm here that patterns $T^{666}/T^{698}/A^{702}/A^{741}$ and $C^{666}/C^{698}/A^{702}/G^{741}$ (accessions Imula, Linga and Latvijas Vietejie; Sjakste and Röder 2004) belong to AD and HN haplotypes of intron III correspondingly. We show here that patterns $T^{666}/T^{698}/G^{702}/A^{741}$ could belong to both HA and AB haplotypes, as pattern $C^{495}/C^{666}/C^{698}/A^{702}/G^{741}$ is the trait of both HN and HS haplotypes. Discrimination of the *Bmy1* alleles of the similar exonic haplotypes could be achieved by taking into consideration intron III polymorphisms including the MS sequence motif. Our results clearly demonstrate the importance of information on the structural rearrangements in non-coding gene regions as well as in coding ones for allele identification.

Barley β-amylase is one of the key enzymes involved in the determination of malting quality. Structural reorganizations of the gene encoding the enzyme were shown to be responsible for kinetic properties, affinity, and thermostability of the gene product (Eglinton et al.

1998; Erkkilä and Ahokas 2001; Kaneko et al. 2000) that could contribute to barley malting quality. However, it is difficult to establish an association between particular structural variations of the gene-candidate and phenotypic manifestation of the polygenic trait. Therefore, we do not speculate here on the possibility of applying our findings directly to the evaluation of malting quality. On the contrary, we would like to demonstrate here the complexity and unreliability of such direct correlations. For example, the same G^{495}/T^{698}/G^{702} exonic haplotype belongs to Czech cultivars Olbram and Tolar of good and poor malting quality, respectively (Polakova et al. 2003), to Finnish landrace HA52 of high malting quality (Erkkilä and Ahokas 2001), and to Latvian accessions of AB haplotype (low to middle malting quality). Varieties Ansis, Rasa, and Sencis of AD haplotypes were the only accessions characterized as having middle to high malting quality in contrast to all other Latvian accessions of AD haplotype (M. Bleidere, Stende breeding station, Latvia, personal communication). Among the Latvian varieties analyzed, only the old local variety Latvijas Vietejie is considered to have high malting quality; however, this variety disappeared from the market at the beginning of the last century and hence was not available for quality control testing.

It was reported (Erkkilä and Ahokas 2001) that 126 bp indel in the 5′ region of intron III is associated with allelic variants of the gene encoding enzymes of low or high thermostability correspondingly. Enzyme thermostability, in turn, could influence malting quality (Eglinton et al. 1998). We confirm here that the presence of 126 bp MITE element in *Bmy1* intron III appears to be a feature indicating low to middle malting quality of barley genotypes (Adorra, Latvian varieties of AD haplotypes). However, the absence of 126 bp MITE element is not directly associated with high malting quality as it is observed in barley genotypes of both high malting quality (Haruna Nijo, Maja, Latvijas Vietejie, Finnish landrace HA52) and low to middle malting quality (Latvian barleys of AB haplotype).

In this respect we conclude that further studies on association of haplotypes to trait and quality remain of great topicality and importance for fundamental research and practical application.

Interhaplotype variability

Each of the intron III haplotypes analyzed possesses a unique spectrum of allelic variants of the polymorphic loci. Simultaneously, some polymorphisms are common for several haplotypes. The MS sequence motif and its flanking sequences are among the main features that enable the discrimination of all the haplotypes analyzed. Sequence polymorphisms of the 3′ end of MS in contrast to MS repeat number were used earlier to differentiate two α-amylase gene families in crustaceans (Van Wormhoudt and Sellos 2003). The high variability of MS and its flanking sequences may be the common mechanism for the evolution of at least some amylase genes in different taxons.

It seems that interhaplotype variability of intron III could be the consequence of the sequence structural reorganization during both natural evolution (HS compared to cultivated barley) and the breeding process. Modern Latvian barley accessions analyzed belong to only two of the described intron III haplotypes namely, AB and AD haplotypes. Pedigree analysis revealed the stability of the transmittance of this genomic portion through generations in different crosses in approximately 70 years of breeding in Latvia. It should be underlined that many fragments of the intron III sequence of different length are conserved between all haplotypes studied, and may have more ancient origin compared to polymorphic loci.

Intrahaplotype variability and validity of MS markers

A complex MS repeat is present in intron III of the *Bmy1* gene. Five haplotypes analyzed possess a different sequence motif in this repeated portion. Variations in the repeat number of both $(TG)_m$ or/and $(G)_n$ MS components were shown in the present study for HN, AB, and AD haplotypes. Such variations could result in an apparent similarity in the size of the MS portion of the different haplotypes when it is determined as MS length polymorphism. This can give rise to mimicry of different haplotypes when the MS fragment size analysis is exclusively used for genotyping studies. As shown here, approximately half of the Latvian varieties studied belong to the AB-like haplotype, others to the AD haplotype. However, the spectrum of sizes of the MS sequences was similar in both haplotypes. Our previous results (Sjakste and Röder 2004) on the absence of the linkage between MS length polymorphism and the C^{698} mutation of exon IV is confirmed by the present study. Indeed, not the number of repeats but the MS motif itself is the main trait of the locus that correlates with mutations in the coding region of the gene. Therefore, *Bmy1* MS length polymorphism cannot be used for association studies but should be considered as the trait of the intrahaplotype variability of repeated portion of the gene. High level of intrahaplotype variability in the number of repeats was shown previously for MS of intron VI of the α-amylase gene of a crustacean *Litopenaeus vannamei* (Van Wormhoudt and Sellos

2003). Taking into account our own and published data, we stress the necessity to test the applicability of MS marker size as sole parameter of genomic region polymorphism in every particular association and linkage disequilibrium study.

Eventual functional significance of the intronic rearrangements

In recent years it has become clear that introns are functionally active participants of gene and genome function. They code for regulatory elements of autocatalytic activity, alternative splicing, gene transcription, and sponsors of mobility including endonucleases and reverse transcriptases (Lewin 2004). This means that not only conserved sequences within introns possess functionality as regulatory elements, but variable regions could also be adaptively active. However, the discovery of the functional role of intronic regulatory elements (IREs; Yutzey et al. 1989) is still the task for future investigations.

In plants, IREs, which could be involved in regulation of gene transcription, have been reported for rice (Fiume et al. 2004), *Lotus japonicus* (Kapranov et al. 2001), *Arabidopsis* (Gazzani et al. 2003; Michaels et al. 2003; Sheldon et al. 2002; Wang et al. 2002). Large deletions within the first intron of the locus of vernalization *VRN-1* are associated with spring growth habit in barley and wheat (Fu et al. 2005). The 126 bp deletion of the intron III of *Bmy1* gene encoding barley endosperm specific β-amylase is associated with Sd2L allele of the gene encoding the enzyme of low-thermostability (Erkkilä et al. 1998). 1 + 6 bp indel in 3′ region of the same intron is linked with exon IV functional SNP (Sjakste and Röder 2004).

In the present paper we made an attempt to use intron III sequences of different haplotypes of the *Bmy1* gene as a model system for comparative computational evaluation of the structural and eventual functional reorganization of the non-coding gene region. We show that highly conserved sequences within intron could possess binding capacity to regulatory elements of several families including plant circadian control factor, G/C BZIP proteins, GT-, I-, legumin- and W-boxes, light and pollen-specific elements, several MYB- and MYB-like proteins, L1BX, RAV5, SEF4, SPF1 TATA-BS, nodulin consensus sequence, ethylene insensitive factors. Six conservative modules consisting of two and more TFs were predicted for all haplotypes studied. Sequence variations could change the pattern of TFBSs with regulation factors of some mentioned TF families (besides RAV5, SEF4, SPF1), and other TF families including AHBP-, CCAAT-proteins, DOF-, GAP-, and MADS-boxes, MYC-like factors, salt/drought responsive element, and storekeeper motif. Several variable modules could be defined and addressed specifically to individual polymorphic loci. For example, insertion of the MITE element in the 5′ region of the intron results in the expansion of BSs to ATCTA-light responsible and ethylene insensitive factors in AD haplotype. In turn, 38 bp deletion in HS haplotype eliminates highly conserved in other haplotypes module of GAP- and DOF-proteins. The 21 bp insertion creates the additional site TATCCA to bind specific MYB protein involved in regulation of sugar metabolism (Lu et al. 2002).

Many of the TFs predicted in our analysis have been extensively studied, their biological functions span on the regulation of secondary metabolism, cellular morphogenesis, cell cycle, development, signal transduction, and disease resistance (Gubler et al. 1995; Gomez-Cadenas et al. 2001; Lu et al. 2002; Wullmout et al. 1998; Yanhui et al. 2006). Sequence variations could change affinity and/or quantity of TFBSs, change the pattern of functional regulatory elements and as a result, influence gene expression.

Taking into account the high levels of inter- and intrahaplotype variability demonstrated in the present study, intron III of the barley seed-specific β-amylase gene can be used as a perfect model for evaluation of the eventual functional significance of the conserved intron portions as well as for sequence rearrangements.

Acknowledgments The study was supported by a grant from the Latvian Council of Sciences. Sequencing was performed by the Sequencing Service of the Biomedical Research Centre (Riga). We thank M. Bleidere for explanations on the malting quality of Latvian barley varieties and N. Sjakste for helpful discussion of the manuscript.

References

Eglinton JK, Langridge P, Evans DE (1998) Thermostability variation in alleles of barley β-amylase. J Cereal Sci 28:301–309

Erkkilä MJ, Ahokas H (2001) Special barley β-amylase allele in a Finnish landrace line HA52 with high grain enzyme activity. Hereditas 134:91–95

Erkkilä MJ, Leah R, Ahokas H, Cameron-Mills V (1998) Allele-dependent barley grain β-amylase activity. Plant Physiol 117:679–685

Fiume E, Christou P, Giani S, Beviarion D (2004) Introns are key regulatory elements of rice tubulin expression. Planta 218:693–703

Fu D, Szücs P, Yan L, Helguera M, Skinner JS, von Zitzewitz J, Hayes PM, Dubkovskii J (2005) Large deletions within the first intron in *VRN*-1 are associated with spring growth habit in barley and wheat. Mol Genet Genomics 273:54–65

Gazzani S, Gendali AR, Lister C, Dean C (2003) Analysis of the molecular basis of flowering time variation in Arabidopsis accessions. Plant Physiol 132:1107–1114

Gomez-Cadenas A, Zentella R, Walker-Simmons MK, Ho T-HD (2001) Gibberellin/abscisic acid antagonism in barley aleurone cells: site of action of the protein kinase PKABA1 in relation to gibberellin signaling molecules. Plant Cell 13:667–679

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

Kaneko T, Kihara M, Ito K (2000) Genetic analysis of β-amylase thermostability to develop a DNA marker for malt fermentability improvement in barley, *Hordeum vulgare*. Plant Breed 119:197–201

Kapranov P, Routt SM, Bankaitis VA, de Bruiji FJ, Szczyglowski K (2001) Nodule-specific regulation of phosphatidylinositol transfer protein expression in *Lotus japonicus*. Plant Cell 13:1369–1382

Kreis M, Williamson M, Buxton B, Pywell J, Heigaard J, Svendsen I (1987) Primary structure and differential expression of β-amylase in normal and mutant barleys. Eur J Biochem 169:517–525

Lewin B (2004) Genes VIII. Pearson, Prentice Hall

Lu CA, Ho T-HD, Ho S-L, Yu S-M (2002) Three novel MYB proteins with one DNA binding repeat mediate sugar and hormone regulation of α-amylase gene expression. Plant Cell 14:1963–1980

Ma YF, Evans DE, Logue SJ, Langridge P (2001) Mutations of barley β-amylase that improve substrate-binding affinity and thermostability. Mol Genet Genomics 266:345–352

Malisheva L, Ganal MW, Röder MS (2004) Evaluation of cultivated barley (*Hordeum vulgare* L.) germplasm for the presence of thermostable alleles of β-amylase. Plant Breed 123:128–131

Michaels SD, He Y, Scortecci KC, Amasino RM (2003) Attenuation of FLOWERING LOCUS C activity as a mechanism for the evolution of summer-annual flowering behaviour in Arabidopsis. Proc Natl Acad Sci USA 100:10102–10107

Paris M, Jones MGK, Eglinton JK (2002) Genotyping single nucleotide polymorphisms for selection of barley β-amylase alleles. Plant Mol Biol Rep 20:149–159

Plaschke J, Ganal MW, Röder MS (1995) Detection of genetic diversity in closely related bread wheat using MS markers. Theor Appl Genet 91:1001–1007

Polakova K, Laurie D, Vaculova K, Ovesna J (2003) Characterization of β-amylase alleles in 79 barley varieties with pyrosequencing. Plant Mol Biol Rep 21:439–447

Sheldon CC, Conn AB, Dennis ES, Peacock WJ (2002) Different regulatory regions are required for the vernalization-induced repression of *FLOWERING LOCUS C* and for the epigenetic maintenance of repression. Plant Cell 14:2527–2537

Sjakste T, Röder M (2004) Distribution and inheritance of beta-amylase alleles in North European barley varieties. Hereditas 141:39–45

Sjakste TG, Rashal I, Röder MS (2003) Inheritance of microsatellite alleles in pedigrees of Latvian barley varieties and related European ancestors. Theor Appl Genet 106:539–549

Van Wormhoudt A, Sellos D (2003) Highly variable polymorphism of the alpha-amylase gene family in *Litopenaeus vannamei* (*Crustacea Decapoda*). J Mol Evol 57:659–761

Wang HY, Lee MM, Schiefelbein JW (2002) Regulation of the cell expansion gene *RHD3* during Arabidopsis development. Plant Physiol 129:638–649

Wullmout RL, Rushton PJ, Hooley R, Lazarus CM (1998) DNAase I footprints suggest the involvement of at least three types of transcription factors in the regulation of α-*Amy2/A* by gibberellin. Plant Mol Biol 38:817–825

Yanhui C, Xiaoyuan Y, Kun H, Meihua L, Jigang L, Zhaofeng G, Zhiqiang L, Yunfei Z, Xiaoxiao W, Xiaoming Q, Yunping S, Li Z, Xiaohui D, Jingchu L, Xing-Wang D, Zhangliang C, Hongya G, Li-Jia Q (2006) The MYB transcription factor superfamily of Arabidopsis: expression analysis and phylogenetic comparison with the rice MYB family. Plant Mol Biol 60:107–124

Yoshigi N, Okada Y, Sahara H, Tamaki T (1995) A structural gene encoding β-amylase of barley. Biosci Biotechnol Biochem 10:1991–1993

Yutzey KE, Kline RL, Konieczny SF (1989) An internal regulatory element controls troponin I gene expression. Mol Cell Biol 9:1397–1405