Mol Breeding
DOI 10.1007/s11032-006-9035-0

Distribution of β-amylase I haplotypes among European cultivated barleys

Elena Chiapparino · Paolo Donini · James Reeves · Roberto Tuberosa · Donal M. O'Sullivan

Received: 26 April 2006 / Accepted: 13 July 2006
© Springer Science+Business Media B.V. 2006

Abstract The barley *β-amylase I* (*Bmy1*) locus encodes a starch breakdown enzyme whose kinetic properties and thermostability are critical during malt production. Studies of allelic variation at the *Bmy1* locus have shown that the encoded enzyme can be commonly found in at least three distinct thermostability classes and demonstrated the nucleotide sequence variations responsible for such phenotypic differences. In order to explore the extent of sequence diversity at the *Bmy1* locus in cultivated European barley, 464 varieties representing a cross-section of popular varieties grown in western Europe over the past 60 years, were genotyped for three single nucleotide polymorphisms chosen to tag the four common alleles found in the collection. One of these haplotypes, which has not been explicitly recognised in the literature as a distinct allele, was found in 95% of winter varieties in the sample. When release dates of the varieties were considered, the lowest thermostability allele (*Bmy1-Sd2L*) appeared to decrease in abundance over time, while the highest thermostability allele (*Bmy1-Sd2H*) was the rarest allele at 5.4% of the sample and was virtually confined to two-row spring varieties. Pedigree analysis was used to track transmission of particular alleles over time and highlighted issues of genetic stratification of the sample.

Keywords Barley · β-amylase · Genetic diversity · Haplotype · Pedigree · Single nucleotide polymorphism (SNP)

E. Chiapparino · P. Donini · J. Reeves ·
D. M. O'Sullivan (✉)
National Institute of Agricultural Botany (NIAB), Huntingdon Road, Cambridge CB3 0LE, UK
e-mail: donal.osullivan@niab.com

R. Tuberosa
Department of Agroenviromental Sciences and Technology, University of Bologna, Viale Fanin 44, 40137 Bologna, Italy

Present Address:
E. Chiapparino
Istituto Agronomico per l'Oltremare, via A. Cocchi 4, 50131 Florence, Italy

Introduction

β-Amylase I (1,4-alpha-glucan maltohydrolase; EC 3.2.1.2), in combination with the activities of limit dextrinase, α-amylase and α-glucosidase, is a key enzyme in the degradation of starch during the barley malting process. Its activity level is strongly correlated with the industrial measures of diastatic power (DP) and fermentability (apparent attenuation limit, AAL), which are

important parameters used in estimating the quality of malt for beer production (Evans et al. 1995; Eglington et al. 1998). However, in most commercial malts, β-amylase (Bmy1) activity is insufficient, and the enzyme appears to be rate limiting for starch hydrolysis during mashing. The lack of activity is due to its thermostability during the mashing process, which is performed at elevated temperature (Thacker et al. 1992).

Interest in genetic improvement of β-amylase thermostability and activity levels for commercial malting purposes has led to a series of studies on variation in the enzyme kinetics and molecular structure of the β-amylase enzyme encoded by the *Bmy1* locus on chromosome 4H (Forster et al. 1991).

Initial classification of allelic diversity was made on the basis of separation of isoforms by isoelectric focusing (IEF). Forster et al. (1991) recognised two IEF banding patterns which, they called β-Amy-1a and β-Amy-1b. Eglington and Evans (1996) also found two isoforms designated Bmy1-Sd1 and Bmy-Sd2 and showed their equivalence to the two isoforms observed by previous authors. Eglington et al. (1998) studied β-amylase thermostability from a range of germplasm and made several new findings. Firstly, high (Sd2H) and low (Sd2L) thermostability subgroups within the Sd2 IEF class were shown to correspond to alleles differing in no less than three amino acid positions based on a comparison of partial sequences from representative Sd2L varieties with that of Haruna Nijo (Yoshigi et al. 1994b). Secondly, a novel source of β-amylase with high thermostability and a unique IEF banding pattern was found (Sd3). Thus *Sd3* was proposed as a fourth *Bmy1* allele. The Sd1, Sd2L, Sd2H, Sd3 nomenclature has not been used by other authors. Kaneko et al. (2000) defined A (high), B (medium) and C (low) type thermostability classes, which are equivalent to the Sd2H, Sd1 and Sd2L classes described above.

The difference in β-amylase thermostability was suggested to reside in the structure of the enzyme itself, which revealed a number of sequence polymorphisms and amino acid exchanges when allele sequences are compared (Eglington et al. 1998; Erkkila et al. 1998; Kaneko et al. 2000; Ma et al. 2001). Ma et al. (2001) showed that the functional differences between the three allelic *Sd1, Sd2L* and *Sd2H* forms reside mainly in three amino acid substitutions, R115C, V233A and L347S.

Our aim in the present work was to use single nucleotide polymorphism (SNP) genotyping to make a comprehensive survey of diversity in the *Bmy1* coding sequence in European barley germplasm.

Four hundred and sixty-four cultivated barley varieties, representing a large proportion of the acreage sown to barley in the EU over the last 60 years, have been surveyed for the three known *Bmy1* alleles by genotyping the varieties at the three SNPs, whose nature cause the three functional amino acid changes described by Ma et al. (2001). In addition, in order to explore novel nucleotide variation at the locus, still unexplored by the published studies, nine European barley varieties were sequenced for *Bmy1* gene fragments. The sequence results show that, in addition to the expected haplotype which described the *Sd1*, *Sd2L* and *Sd2H* alleles, a novel haplotype (previously undescribed combinations of the three functional mutations assayed) was observed. The distribution of the *Bmy1* alleles in the cultivated germplasm will be discussed. We conclude that genotyping known functional polymorphisms shows some promise for the early detection of barley lines that carry superior *beta*-amylase alleles, and may represent a useful tool for selection strategies for breeders.

Materials and methods

Plant material and DNA extraction

Four hundred and sixty-four spring and winter barley commercial varieties were obtained from germplasm collections around Europe and from academic collections including those from BBSRC (JIC, Norwich, UK), NIAB (Cambridge, UK), the University of Helsinki, IPK (Gatersleben, Germany), and CSIC (Zaragoza, Spain).

Material was selected to represent the major varieties from the cultivated germplasm pool grown in the EU in the last 60 years. These comprise 279 two-row spring, 74 two-row winter, 52 six-row winter and 43 six-row spring varieties

which broadly reflects the relative size and importance of these respective crop types across the geographic zone and time period covered. The varieties originate from 13 EU countries and have been released in different decades (Table 1). To select the most diverse and representative material, pedigree relationships, molecular marker-based dendograms, variety catalogues (such as European Common Catalogue, UK National List and NIAB Cereal Recommended List) were used.

DNA was extracted from bulk flour from 30 seeds for each accession using the Qiagen DNeasy® 96 Plant Kit according to the manufacturer's protocols.

Primer design and PCR

Locus specific amplification primers and sequencing primers for the *Bmy1* gene were designed from the published sequences of the *beta*-amylase complete gene (GeneBank accession D49999 and AF061203) using the BioOligo™ program (Bio/Gene Limited, UK). Sequence were first aligned using ClustalX (Thompson et al. 1997) and primers designed in the conserved region to amplify fragments between 600 and 1,000 bp in coding and non-coding regions (Table 2).

Each reaction was performed in a total volume of 20 μl, containing 50 ng of template DNA, 100 nM of each primers, 200 μM dNTPs, 2.5 mM of $MgCl_2$, 1× buffer and 1 U *Taq* polymerase (Roche).

DNA sequencing, alignment and polymorphism identification

To explore novel nucleotide variation at the *Bmy1* locus, four *Bmy1* gene fragments, spanning coding and non-coding regions, were sequenced from nine samples, (two winter barleys: Igri and Maris Otter and seven spring barleys: Berac, Bomi, Carmen, Decanter, Erkki, Golf and Triumph) chosen to be the most diverse based on pedigree analysis. The fragments sequenced cover 3.2 kb of the 3.8 kb *Bmy1* gene.

PCR products were treated with 1.5 U of *Exonuclease I* (New England BioLab) and 1 U of Shrimp Alkaline Phosphatase (Promega) and then sequenced directly using the primers given in Table 2 using an ABI 3100 Sequencer (Applied Biosystems). The sequencing reactions were performed using the ABI PRISM® BigDye™ Terminator v.3.1 Cycle Sequencing Kit (Applied Biosystems) according to the manufacturer's protocol. The sequences were aligned and mutations identified using the Staden package v.2002.0 (http://www.staden.sourceforge.net/). An alignment of the 11 database sequences (see Table 3 for details) and of nine sequences obtained by direct sequencing was performed using ClustalX (Thompson et al. 1997). DNA polymorphism

Table 1 Barley varieties used

	Pre 1940	1940	1950	1960	1970	1980	1990	ND	Total
Austria	1		1	3	1	1			7
Belgium				2			1		3
Czech Rep.		1	1	1		2			5
Denmark	3	2	3	4	8	2			22
Finland		1	2	5	8	8	8		32
France	1			6	19	25	3		54
Germany	13	4	11	27	20	27	6	1	109
Ireland	1			1	1				3
Netherlands	1		4	8	21	9	1		44
Norway	2	2		2					6
Spain			1			2	5	1	9
Sweden	3	2	4	8	8	6			31
UK	3	6	8	6	23	47	46		139
Total	28	18	35	73	109	129	70	2	464

ND Not defined

Table 2 List of primers used for locus amplification, sequencing and SNP genotyping

Primer name and location in GenBank acc. no. AF061203	Purpose	5′-3′ sequence	PCR[a]
Bamy1-Forward432	SNAPshot locus amplification	TGTTTGAGCTGGTGCAGAAGGCTGG	(1)
Bamy1-Reverse634	SNAPshot locus amplification	CGGCAGATCTTCCATGGAAGAGAGGC	
Bamy1-F2300	SNAPshot locus amplification	CAAGCAGACTTCAAAGCA	(2)
Bamy1-R2936	SNAPshot locus amplification	TTGGACTAGTTCTTCTGGTG	
C115R	SNaPshot extension primer	CAGTGGGTGCGGGACGTCGGCACG	NA
V233A	SNaPshot extension primer (tail underlined)	GACT CCTGAGTGGGAATTTCCTAACGATG	NA
S347L	SNaPshot extension primer (tail underlined)	AAAAGACT TTCACTTGCGCGGAGATGAGGGATT	NA
Bamy1-F17	Sequencing locus amplification and sequencing	AAGGCAACTATGTC CAAGTC	(2)
Bamy1-R1193	Sequencing locus amplification and sequencing	GGAAACAAACACTCACGATG	
Bamy1-F1210	Sequencing locus amplification and sequencing	CCAAACTAATAATCTTTCCT	(3)
Bamy1-R2289	Sequencing locus amplification and sequencing	GTCTGCTTGTAGGTATTTAT	
Bmy1-F2169	Sequencing locus amplification and sequencing	TGAAATAGTTAAAACAAATGCACGAT	(4)
Bmy1-R3036	Sequencing locus amplification and sequencing	GGCCGGGGCTAAGCCTCCTTTTCTAA	
Bamy1-F3116	Sequencing locus amplification and sequencing	CGTCAAAATAATTTCTCAGGT	(2)
Bamy1-R3726	Sequencing locus amplification and sequencing	CCACCCATGCCACTAGTAG	
Bmy1-R732	Sequencing	TTGGCACCTTTAACCTTTTGCTTGTA	
Bmy1-F1506	Sequencing	TGGCAGTTTCAGATTAGGACAAGGCG	

[a] The numbers in parentheses represent one of the four PCR conditions used: (1) A touchdown PCR consisting of 2 min at 94°C and 35 cycles of 1 min at 94°C, 1 min annealing and 1 min extension at 72°C, ending with 2 min at 72°C. Annealing temperatures were progressively decreased by 1°C/cycle from 72 to 65°C over the first 8 cycles. (2) A similar touchdown PCR condition except that the annealing temperature were decreased by 1°C/cycle from 60 to 55°C over the first 6 cycles. (3) A touchdown PCR consisting of 2 min at 94°C and 35 cycles of 1 min at 94°C, 1 min annealing and 1 min extension at 72°C, ending with 2 min at 72°C. Annealing temperatures were progressively decreased by 1°C/cycle from 55 to 50°C over 6 cycles. (4) A regular PCR profile consisting of an initial denaturation of 94°C for 2 min, 35 cycles of 1 min at 94°C, 1 min at 55°C, and 1 min at 72°C, ending with 2 min at 72°C to complete extension

Table 3 SNPs in the *Bmy1* coding region for 11 previously sequenced genotypes compared with nine European varieties

H[a]	Genotypes[c]	References Position[b]	**343 (R115C)**	471	495 (D165E)	531	591	666	**698 (V233A)**	702	736 (F246L)	741	868	945	**1,040 (L347S)**	1,289 (V430A)	1,293	1,581 (M527I)
I (*Sd1*)	Franklin (AF300800)	Ma et al. (2000b)	**T (C)**	A	G (E)	A	C	T	**T (V)**	G	T (F)	A	C	C	**C (S)**	C (A)	T	G (M)
	Morex (AF414081)	Clark et al. (2003)	**T (C)**	A	G (E)	A	G	T	**T (V)**	G	T (F)	A	C	C	**C (S)**	C (A)	T	G (M)
	Harrington (AB048949)	Li et al. (2002)	**T (C)**	A	G (E)	A	G	T	**T (V)**	G	T (F)	A	C	C	**C (S)**	C (A)	T	G (M)
	Triumph	This work	**T (C)**	A	G (E)	A	ND	ND	**T (V)**	G	T (F)	A	C	C	**C (S)**	C (A)	T	ND
	Golf	This work	**T (C)**	A	G (E)	A	ND	ND	**T (V)**	G	T (F)	A	C	C	**C (S)**	C (A)	T	ND
Ia	HA52 (AJ301645)	Erkkila et al. (2001)	**T (C)**	A	G (E)	A	G	T	**T (V)**	G	C (L)	A	C	C	**C (S)**	C (A)	T	G (M)
II (*Sd2H*)	Haruna nijo (D49999)	Yoshigi et al. (1995)	**C (R)**	A	C (D)	G	G	C	**C (A)**	A	T (F)	G	T	T	**C (S)**	T (V)	C	G (M)
	Berac	This work	**C (R)**	A	C (D)	G	ND	ND	**C (A)**	A	T (F)	G	T	T	**C (S)**	T (V)	C	ND
	Carmen	This work	**C (R)**	A	C (D)	G	ND	ND	**C (A)**	A	T (F)	G	T	T	**C (S)**	T (V)	C	ND
IIa	*H. spontaneum* PI296897 (AF061204)	Erkkila et al. (1998)	**C (R)**	A	C (D)	G	G	C	**C (A)**	A	T (F)	G	C	C	**C (S)**	T (V)	T	G (M)
III (*Sd2L*)	Hiproly (X52321)	Kreis et al. (1987)	**C (R)**	A	C (D)	G	G	T	**T (V)**	A	T (F)	A	C	C	**T (L)**	T (V)	T	A (I)
	Adorra (AF061203)	Erkkila et al. (1998)	**C (R)**	A	C (D)	G	G	T	**T (V)**	A	T (F)	A	C	C	**T (L)**	T (V)	T	G (M)
	Schooner (AF300799)	Ma et al. (2000b)	**C (R)**	C	C (D)	G	C	T	**T (V)**	A	T (F)	A	C	C	**T (L)**	T (V)	T	A (I)
	Bomi	This work	**C (R)**	A	C (D)	G	ND	ND	**T (V)**	A	T (F)	A	C	C	**T (L)**	T (V)	T	G (M)
	Decanter	This work	**C (R)**	A	C (D)	G	ND	ND	**T (V)**	A	T (F)	A	C	C	**T (L)**	T (V)	T	G (M)
IV (*Sd4*)	Orca (AY454398)	S.E. Clark et al., unpublished	**C (R)**	A	C (D)	G	G	T	**T (V)**	A	T (F)	A	C	C	**C (S)**	T (V)	T	G (M)
	Steptoe (AF414082)	Clark et al. (2003)	**C (R)**	A	C (D)	G	G	T	**T (V)**	A	T (F)	A	C	C	**C (S)**	T (V)	T	G (M)
	Igri	This work	**C (R)**	A	C (D)	G	ND	ND	**T (V)**	A	T (F)	A	C	C	**C (S)**	T (V)	T	ND
	Erkki	This work	**C (R)**	A	C (D)	G	ND	ND	**T (V)**	A	T (F)	A	C	C	**C (S)**	T (V)	T	ND
	Maris Otter	This work	**C (R)**	A	C (D)	G	ND	ND	**T (V)**	A	T (F)	A	C	C	**C (S)**	T (V)	T	ND

[a] Haplotypes. Allele equivalent in brackets cf. comments in Results section

[b] The position of the nucleotide substitution is numbered according to the β-amylase coding sequence of Harrington (GenBank acc. num. AB048949). When the substitution causes an amino acid exchange the identity and its relative position in the protein sequence is given below in brackets. C: Cysteine; R: Arginine; D: Aspartic; E: Glutamic; V: Valine; A: Alanine; F: Phenylalanine; L: Leucine; S: Serine; M: Methionine; I: Isoleucine

[c] Genotypes from database are given with respective references and GenBank accession numbers in brackets. Amino acid composition of the genotypes is given in brackets

analysis was performed using DNAsp v.4.0 (Rozas et al. 2003).

SNP genotyping

All the 464 barley samples were assayed for three SNPs detecting amino acid (a.a.) substitutions affecting Bmy1 substrate affinity and thermostability (Ma et al. 2001). The two target regions containing the three SNPs were amplified using the primers given in Table 2 and SNP genotyping was performed by dideoxy single-base primer-extension chemistry using the ABI PRISM® SNaPshot™ Multiplex Kit (Applied Biosystems) following the manufacture's protocol. The two PCR products were purified as recommended and mixed for the subsequent extension reaction. The three specific extension primers (Table 2) were combined and reactions performed as per the manufacture's instructions. The fragments were run on an ABI PRISM 3100 Sequencer (Applied Biosystems) and samples were analysed using GeneScan® 3.7 and Genotyper® 3.7 software.

Results

Nucleotide variation and haplotype structure at the *Bmy1* locus

In order to determine whether the coding sequence variants observed in the mainly Japanese, Australian and American varieties represented in Genbank accurately captured allelic diversity in European cultivated barley, we determined nine new *Bmy1* coding sequences from a representative cross section of the European barley germplasm we intended to genotype and compared them with existing *Bmy1* sequences. A summary of the polymorphic positions in the resulting sequence alignments is shown in Table 3. Sixteen nucleotide polymorphisms were observed in total in the coding sequence (CDS), comprising nine synonymous substitutions and seven amino acid exchanges (Table 3). The 11 published coding sequences resolve into four clear haplotype groups, each of which is characterised by a unique amino acid signature. Haplotype group I (Table 3) differs from all the other haplotypes by three amino acid substitutions (R115C, D165E and V430A) as well as two silent nucleotide substitutions at positions 531 and 702 of the cDNA. Because this group includes Franklin and Harrington, designated in previous studies as carrying the *Sd1* allele (Ma et al. 2000b; Li et al. 2002), we henceforth refer to this as the ''Sd1'' haplotype. Haplotype group II is identified by six characteristic base changes that differentiate it from all other haplotypes at positions 666, 698, 741, 868, 945 and 1293 of the cDNA, just one of which encodes an amino acid substitution (V233A). This group includes the Japanese variety Haruna Nijo, which carries the high thermostability allele *Sd2H* (Eglinton et al. 1998). We will henceforth refer to this group as the ''Sd2H'' haplotype. Haplotype group III is identified by two diagnostic base changes at position 1,040 and 1,581 of the cDNA, which lead to the substitution of the consensus Serine at a.a. position 347 with a Leucine and of the consensus Methionine at a.a. position 527 with a Isoleucine. This group contains Hiproly, Adorra and Schooner (all previously characterised as carrying the *Sd2L* allele; Eglinton et al. 1998; Erkkila et al. 1998) and will therefore be identified as the ''Sd2L'' group. The haplotype group IV is identical to the ''Sd2L'' haplotype except for one amino acid substitution (S347L) caused by the SNP at cDNA position 1,040. This haplotype is more difficult to label; since prior to the work of Clark et al. (2003), Steptoe was thought to carry the *Sd2L* allele on the basis of an intron III indel polymorphism (Erkkila et al. 1998; Erkkila 1999). However, based on consideration of its unique amino acid haplotype, its high frequency of occurrence (this work), and recent evidence described by Clark et al. (2003) demonstrating its inherently higher activity compared with the *Sd1* allele carried by Morex, from which it can be inferred that it also has higher activity and thermostability with respect to the low thermostability *Sd2L* allele, we suggest that this group needs to be considered as a previously unrecognised allele or haplotype which we refer to provisionally as *Sd4* to distinguish it from the *Sd3* allele, which has in the

published literature only been observed in wild barley accessions.

Two further haplotypes (Ia and IIa) were defined by single representative sequences from landrace and wild genepools, respectively. The Finnish landrace HA52 is identical to the "Sd1" haplotype except for one amino acid substitution (F246L) caused by the SNP at cDNA position 736. The *H. spontaneum* accession PI 296897 presents complete a.a sequence identity to "Sd2H" haplotype yet carries three silent SNPs towards the 3′ end of the gene. We also considered that the silent nucleotide substitutions observed at cDNA positions 471 and 591 in cv. Schooner did not merit creation of new haplotype grouping for two reasons: (1) These SNPs did not affect the amino acid sequence, which was identical to all the other genotypes belonging to the "Sd2L" haplotype; (2) β-amylase activity measured from Schooner placed it in the low thermostability group "Sd2L". Likewise, the M527I a.a. substitution in Adorra is disregarded since it is otherwise identical to other Sd2Ls and the mutation occurs in the C-terminal region which is cleaved after germination to give the active enzyme, and the differences in the thermostability and kinetic proprieties of β-amylase allelic form are retained after germination (Ma et al. 2000a, b). In summary, we found no new sequence haplotypes among the nine varieties we sequenced compared to the haplotypes previously observed in cultivated barleys from around the world. Although it is not by any means excluded that novel polymorphisms and recombinant alleles can be found within cultivated EU barleys, we reasoned that the already observed haplotypes are likely to account for the common haplotypes encountered in the majority of EU elite varieties, and that in any case, even where not recognised, any novel haplotypes present in the sample would be grouped with their most related common allele, and therefore we decided to genotype the full set of 464 varieties with the minimum set of SNPs required to distinguish haplotype groups I–IV.

For this purpose, we chose the SNPs underlying the three mutations R115C, V233A and L347S (shown in bold in Table 3). Apart from identifying which of the four alleles known to occur in cultivated barley was present in a particular variety, these three SNPs correspond to those identified by Ma et al. (2001) as affecting the thermostability and kinetic properties of the three allelic β-amylase forms then known.

Distribution of *Bmy1* alleles in European cultivated barley germplasm

The three-SNP haplotype tag set identified above was assayed on the full variety set by a triplex single-base primer extension reaction. Of the eight possible combinations of the three SNPs, just the four predicted from sequencing results, corresponding to the three previously described alleles plus the new *Sd4* haplotype, were observed. In other words, no novel alleles representing putative recombinant products of known alleles were observed. The frequency of occurrence of each allele/haplotype in the germplasm under study is shown in Fig. 1a. The *Sd1*, *Sd2L* alleles and the *Sd4* haplotype are predominant and are present in similar proportions, while the high thermostability allele *Sd2H* was detected in just 25 varieties. When allele frequencies are shown sorted by country of release (Fig. 1b) a few exceptions to the general observation of predominance of the *Sd1*, *Sd2L* alleles and *Sd4* haplotype may be noted. In the Finnish material assayed, the *Sd2H* allele was not detected, while the *Sd4* haplotype was near absent in the Swedish sample and not observed at all in the Danish varieties analysed, although it should be noted that the number of varieties selected from each country varies considerably. Finally, observing the changes in relative allele frequency over the decades, the *Sd1*, *Sd2L* alleles and *Sd4* haplotype are still predominant although it appears that the representation of the least thermostable allele (*Sd2L*) has decreased rapidly over recent decades (Fig. 1c).

Although the high thermostability allele *Sd2H* was present, it was at 5.4% by far the least widespread of the four alleles observed and was virtually confined to spring, two-row varieties. In order to gain retrospective insight into how this potentially useful allele had been deployed, we looked at the pedigree relationships between all the 26 *Sd2H* varieties. Pedigrees of nineteen of

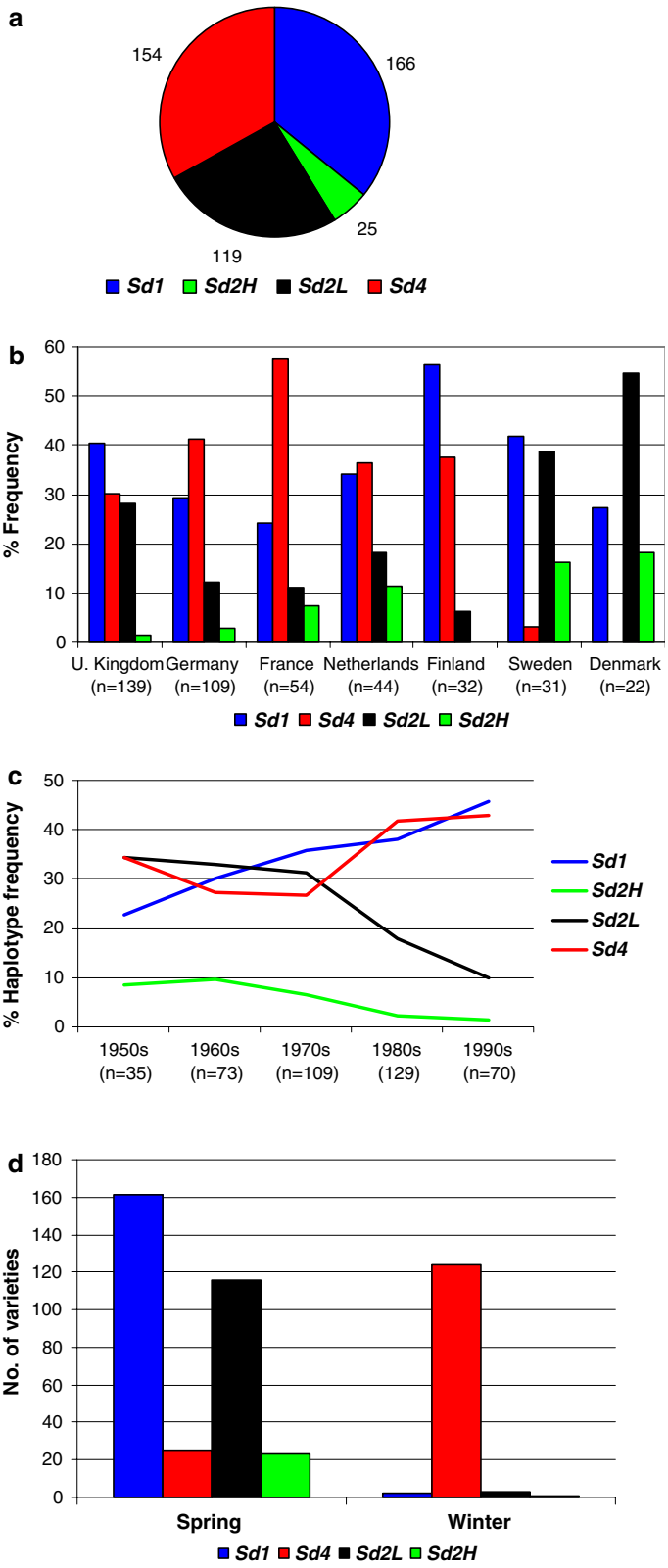

Fig. 1 **(a)** Frequency of each *Bmy1* haplotype in the germplasm assayed. The figure shows the number of accessions observed for each haplotype. Colour coding throughout figures follows the following convention: *blue*, *Sd1*; *green*, *Sd2H*; *black*, *Sd2L*; *red*, *Sd4*. **(b)** Percentage frequency of the four *Bmy1* haplotypes in the seven most represented European countries in the sample. Sample number in each category is shown. **(c)** Temporal trends of the four *Bmy1* haplotypes over five decades. As in (**b**), sample number in each category is stated. **(d)** Distribution of the *Bmy1* haplotypes over growth habit categories

the varieties were linked by common progenitor material and the complex relationships between these are represented in Fig. 2. From this diagram we see that Maja, a Danish variety derived from a Binder × Gull cross and released in the 1930s is the oldest variety included in the study which carries the *Sd2H* allele. The Maja *Sd2H* allele is directly transmitted to Balder and Rigel, as well as indirectly to Sultan and Amsel, which derive from more complex crossing schemes (dashed lines) that include Maja. Balder in turn, transmits the *Sd2H* allele to Cambrinus, Mosane and Berac. Both Rika and Herta carry *Sd2H*, which is further transmitted from Rika to Crusander, Ariel and Janes and from Herta to Herta 8, Tyra and Anita. Since both these latter *Sd2H* donors derive from a cross between Kenia and Isaria, genotyped in this work as *Sd2L* and *Sd1* respectively, it seems likely that one of these parental genotypes is incorrect. The most likely explanations for this are the possible heterogeneity in the original varieties, which have been fixed into distinct derived lines during maintenance in germplasm collections over many years, or erroneous passport data. Such inconsistencies, heterogeneities and errors have been encountered by previous authors at a relatively high frequency (Sjakste et al. 2003; Lund et al. 2003).

An interesting distribution of the alleles is observed when samples are sorted by growth habit (Fig. 1d): the *Sd4* haplotype dominates the winter varieties at a frequency of 94.8% (128/135), while the maximum allele frequency in the spring pool is 49.7% (*Sd1*), with none of the four alleles showing less than a 7% frequency. Although at a first glance the lack of polymorphism at the *Bmy1* locus in winter germplasm seems highly suggestive, it is worth noticing that there is a small but significant number of occurrences of the *Sd4* haplotype in the spring germplasm pool. Previous authors have suggested based on isoenzyme (Forster et al. 1991) and RFLP patterns (Igartua et al. 1999) that certain alleles of *Bmy1* are diagnostic of the winter growth habit. However, based on the higher discriminatory power of the haplotyping and the larger sample of this study, we can demonstrate that the correlation between winter growth habit and *Sd4* haplotype is far from perfect. If the pedigrees of the 24 spring *Sd4* varieties are examined in detail, it is observed that 12 out of 24 are Finnish varieties. Of these, nine are six-row spring varieties (the dominant type of barley grown in Finland) and it is instructive to examine the pedigree relationships between all the *Sd4* six row Scandinavian varieties used in this study and their progenitors (Fig. 3). From this diagram, it can be seen that the inheritance of the *Sd4* haplotype of no less than ten varieties (Finnish varieties Tammi, Otra, Hankkija 673, Eero, Pohto, Potra, Loviisa and Erkki; Norwegian varieties Herse and Lise) can be traced directly back to a single Scandinavian progenitor Asplund (released in Sweden in the 1920s). Only the origin of the

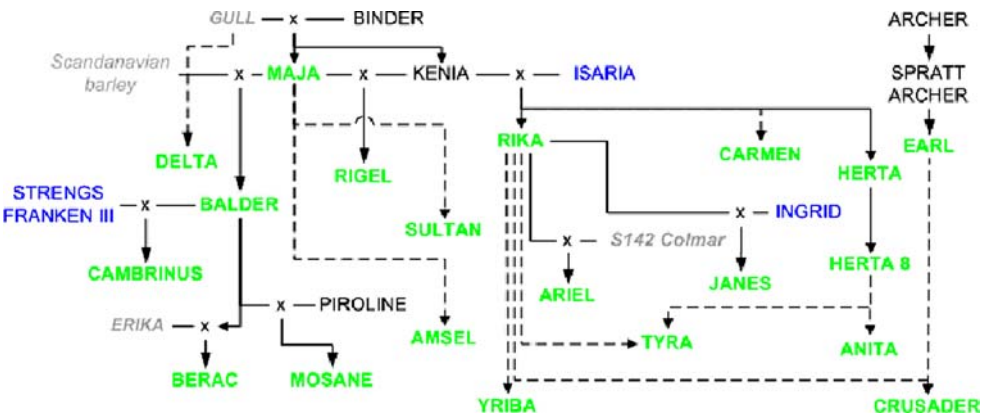

Fig. 2 Pedigree relationships among spring barley varieties carrying the *Bmy1-Sd2H* allele. Variety names are colour coded according to which *Bmy1* allele was observed in that variety, using the same convention as in Fig. 1, except that varieties not genotyped as part of this study are shown in *grey*

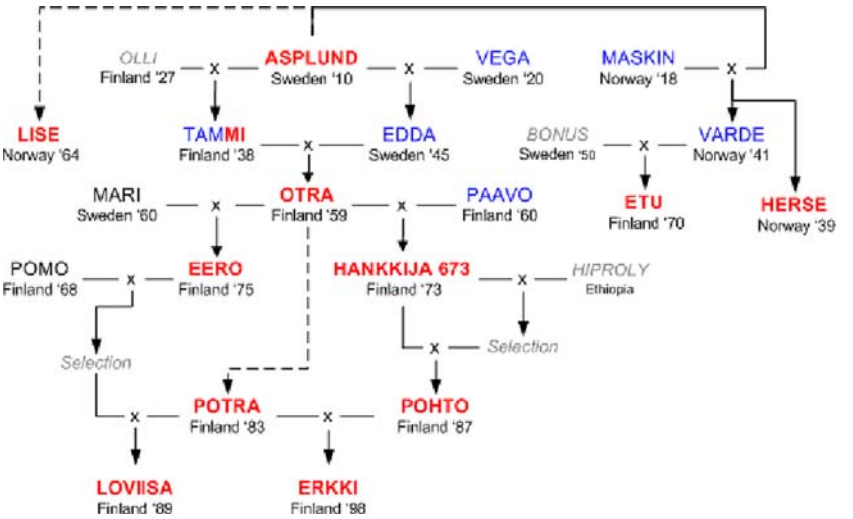

Fig. 3 Pedigree relationships among Scandanavian six-row spring barley varieties, which carry the *Sd4* haplotype. The country and year of release is shown for all varieties. Variety names are colour-coded according to *Bmy1* allelic status as in Fig. 2

Sd4 haplotype from varieties Silja and Etu cannot be directly traced to Asplund based on the available data, although Etu does have Asplund in its pedigree. Despite the inability to unequivocally trace all the occurrences of the *Sd4* haplotype, it is nevertheless clear that an allele introduced into a regional genepool via a widely used foundation genotype may persist in multiple varieties over a considerable period of time.

Discussion

Our analysis of coding sequence diversity in the *Bmy1* gene on chromosome 4H has revealed the existence of a new haplotype distinct from the three alleles already recognised to occur in cultivated barley and whose first exemplar in the nucleotide databases was that of cultivar Steptoe, submitted to Genbank in 2003 (accession number AF414082). We have provisionally designated this the *Sd4* haplotype. Although several surveys of allelic variation have preceded this one, none have used a combination of polymorphisms that distinguished the *Sd4* haplotype from *Sd2L*. Paris et al. (2002) used SNPs at positions 495 and 698 of the cDNA to distinguish the three forms of Bmy1—*Sd1*, *Sd2L* and *Sd2H* recognised at the time. Polakova et al. (2003) employed both the latter SNPs as well as an adjacent SNP at position 702 to recover essentially the same three alleles, with the addition of one novel haplotype found in a single Czech variety. Malysheva et al. 2004 used just the SNP at position 698 to distinguish high thermostability *Sd2H* and *Sd3* alleles from the rest. The combinations of polymorphisms used in all three studies detect identical haplotypes for both *Sd2L* and *Sd4* forms, with the result that no association of a *Bmy1* allelic form with the winter type was observed in these surveys. A recent biochemical study comparing activity of purified recombinant Bmy1 protein from Steptoe (*Sd4* haplotype) with that of Morex (*Sd1*) demonstrated that the Steptoe *Bmy1* allele encodes a protein with significantly higher activity than that of *Sd1* and that this can be attributed to the amino acid differences encoded by SNPs in the translated region of the gene (Clark et al. 2003). Taken together, the functional difference between the *Sd4* haplotype and *Sd1* alleles, combined with the widespread occurrence of the *Sd4* haplotype in cultivated barley (this work), constitute persuasive arguments for the recognition of this haplotype as a new allele. As detailed in the introduction, nomenclature of *Bmy1* alleles is not consistent between authors, and thus we propose that the list of phenotypically distinct *Bmy1* alleles and their names should be revised in the near future pending studies to clarify the canonical sequence and phenotype associated with each distinct allele. Such a nomenclature should follow the recommended rules for nomenclature and

gene symbolization in barley as reported in BGN 21:11–13.

The SNP genotyping assay we have described here has allowed us to observe in more detail than before the distribution of *Bmy1* alleles in European cultivated barley varieties over more than six decades of breeding. A major conclusion of this survey is that the most thermostable form of endosperm β-amylase, encoded by the *Sd2H* allele, is by far the least frequently occurring. This can be partly elucidated by pedigree analysis, which showed that, where enough information was available, the transmission of *Sd2H* could in all cases be traced back to the progeny of a single putative source (the Gull × Binder cross) amongst all the recognised foundation genotypes of modern spring barley (Russell et al. 2000). Similar results were obtained by Malysheva et al. (2004) and Sjakste and Roder (2004) by genotyping, respectively, a set of 889 accessions of cultivated barley and a set of 55 Latvian and Scandinavian barley accessions for a single mutation that discriminates *Sd2L* and *Sd1* (low and medium thermostability) alleles from *Sd2H* and *Sd3* (high thermostability) alleles. Furthermore, over the period of 50 years separating the release of Maja (Denmark, 1934) and Yriba (France, 1984), *Sd2H* has remained confined within the two row spring barley group.

Its deployment in varieties released over most of the time period covered by the study and originating from at least 8 of the 12 EU countries sampled can therefore be understood to reflect more the widespread utilisation of 'Gull × Binder'-derived varieties than any putative selective value of the high thermostability allele itself. It is indeed possible that *Sd2H* does not have a significant selective value in breeding terms. Its experimentally demonstrated high inherent thermostability (Eglington et al. 1998) may be balanced by poorer overall expression which would result in a neutral or even unfavourable effect on malting potential as measured in terms of diastatic power (D.P.) or fermentability (AAL) by breeders. Detailed comparative promoter studies would be needed to further investigate this possibility.

Several sources in the literature concur with our analysis regarding the distribution of *Sd2H* in Europe. Kihara et al. (1999), for example, analysed 274 barley varieties from ten different regions of the world for the variation of β-amylase thermostability. Out of the 68 European varieties assayed, 60 belonged to the B and C thermostability classes, shown to represent the products of the alleles *Sd1* and *Sd2L* respectively (Kreis et al. 1987; Kaneko et al. 2000), while most of the varieties belonging to the thermostability class A (reported as the result of the expression of the allele *Sd2H*—Yoshigi et al. 1994a; Eglington et al. 1998) originated from Japan, China and Korean Peninsula. Since introduction of barley of Far Eastern origin into Scandanavia (the same geographical region where Gull and Binder were selected from landraces) has been documented (Plarre and Hoffman 1963), it is possible to speculate that *Sd2H* was introduced perhaps just on one single occasion into the European breeding pool from the East. Clearly, therefore, our data set, which inventories the distribution of superior alleles (in this case *Sd2H* for thermostability) in popular and widely used varieties from around Europe provides a powerful tool to breeders, who cannot only define backgrounds in which Bmy1 thermostability may be improved, but can also select from all the identified donor backgrounds carrying the *Sd2H* those most suited to their purposes.

Perhaps the most surprising result of our study has been the observation that over 90% of winter varieties in our study carried the *Sd4* haplotype. Again, however, an historical analysis of the development of modern winter barley may help to understand this phenomenon. Compared to spring barley, winter barley was of relatively minor importance in Europe until the 1950s and hence, had a much smaller basis of local landraces for the development of improved cultivars (Fischbeck 1992; Melchinger et al. 1994). Just four progenitors (Firlbecks 4zlg, Dea, Hauters and Vogelsanger Gold) were the predominant ancestors of six-row winter barley cultivars. The original genetic base of the two-row winter barley in Europe was probably even narrower than for six-row type. Intensive breeding work for two-row winter barley started only after Word War II, mainly by crossing six-row winter cultivars with two-row spring progenitors. Thus the narrow

genetic basis for the winter pool could explain the predominance of the *Sd4* haplotype over the other allelic forms.

However, modern winter barley is not dramatically lacking in genetic diversity across the whole genome despite the fact that winter and spring groups are readily distinguished using molecular markers (Melchinger et al. 1994; Russell et al. 1997; Backes et al. 2003). The most probable explanation for the dominance of the *Sd4* haplotype in the winter pool is the close genetic linkage of the *Bmy1* locus and the *Vrn-H2* (formerly known as *Sgh1)* locus on chromosome 4H. *Vrn-H2* is one of the three major genes (*Vrn-H1, Vrn-H2* and *Vrn-H3*) whose allelic status controls the winter/spring habit (Laurie et al. 1995; Karsai et al. 2005). *Vrn-H2* is orthologous to the recently cloned wheat *VRN-A2* vernalization gene (Yan et al. 2004). Thus, the 95% correlation between the dominant winter allele of *Vrn-H2* with the *Sd4* haplotype in the narrow base of winter progenitor materials may have been created and subsequently maintained simply by linkage drag around the *Vrn-H2* locus when the winter allele was selected and fixed. This hypothesis is supported by the results of a genome-wide RFLP diversity study (Backes et al. 2003), where winter and/or spring-specific RFLP profiles were detected preferentially for those RFLPs mapping on chromosomes 5H, 4H and 1H reported to carry genes related to the growth habit (5H, 4H) and to the photoperiod (1H) response respectively (Hayes et al. 1993; Pan et al. 1994; Laurie et al. 1995). The recent molecular characterisation of barley vernalization genes *Vrn-H1* (Danyluk et al. 2003; Trevaskis et al. 2003; von Zitzewitz et al. 2005), *Vrn-H2* (von Zitzewitz et al. 2005; Dubcovsky et al. 2005) and photoperiod response gene *Ppd-H1* (Turner et al. 2005) opens up the possibility to rigourously determine the nature and extent of linkage disequilibrium (LD) in the regions surrounding these genes and indeed gametic phase disequilibrium between such loci.

Although we have just taken one example in the present work, other examples of highly stratified markers have been noted in the literature. For example, the distribution of haplotypes of STS marker MWG699, very closely linked to the *vrs1* row number locus, are markedly different in two-row versus six-row cultivated material (Tanno et al. 2002). Even more dramatically, a unique haplotype group of STS marker sKT7 tightly linked to the *nud* locus controlling the hulled/naked character of barley was exclusively present in over 100 naked barleys of diverse geographic origin, indicating a monophyletic origin for the naked phenotype (Taketa et al. 2004). These examples of marked allelic confinement illustrate dual aspects of analyses of sequence diversity in cultivated backgrounds. On one hand, there is the expectation that analysis of associations between marker sequence haplotypes and phenotypes will reveal associations between markers and closely linked functional polymorphisms; this is the principal underlying association genetics (Rafalski 2002). In this instance, we chose to use functional polymorphisms in the β-amylase gene as a tool for viewing any changes in allele distribution over time and found that although the *Bmy1-Sd2L* allele appears to have significantly reduced over recent decades, the distribution of β-amylase alleles of different thermostability and activity levels is nevertheless profoundly affected by the footprint of selection at the *Vrn-H1* locus 1 cM away. Further studies are needed to elucidate whether linkage disequilibrium LD extends beyond *Bmy1*, and to clarify overall patterns of LD in cultivated germplasm, since high levels of LD will reduce the mapping resolution of association studies and make it more difficult for breeders to break up unwanted associations of marker alleles.

Acknowledgements This work was financed by the EU FP5 project GEDIFLUX QLRT-2000-00934. We thank Mike Ambrose (JIC, Norwich), Bob Jarman (NIAB, Cambridge), Alan Schulman (University of Helsinki), Marion Roeder (IPK, Gatersleben) and Ernesto Igartua (Spain) for supplying seed of barley accessions for this study and Alan Schulman for providing information on the pedigrees of Finnish varieties.

References

Backes G, Hatz B, Jahoor A, Fischbeck G (2003) RFLP diversity within and between major groups of barley in Europe. Plant Breed 122:291–299

Clark SE, Hayes PM, Henson CA (2003) Effects of single nucleotide polymorphisms in β-amylase 1 alleles from barley on functional properties of the enzymes. Plant Physiol Biochem 41:798–804

Danyluk J, Kane NA, Breton G, Limin AE, Fowler DB, Sarhan F (2003) TaVRT-1, a putative transcription factor associated with vegetative to reproductive transition in cereals. Plant Physiol 132:1849–1860

Dubcovsky J, Chen CL, Yan LL (2005) Molecular characterization of the allelic variation at the VRN-H2 vernalization locus in barley. Mol Breed 15:395–407

Eglington JK, Evans DE (1996) Identification of the β-amylase isoelectric focusing band pattern in barley (*Hordeum vulgare* L.). Barley Genet Newsl 27:8–12

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

Erkkila MJ (1999) Intron-III specific markers for screening of β-amylase alleles in barley cultivars. Plant Mol Biol Rep 17:139–147

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

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

Evans DE, Lance RCM, Eglinton JK, Lougue SJ, Barr AR (1995) The influence of β-amylase isoform pattern on β-amylase activity in barley and malt. In: Abstracts of the 45th Australian Cer Chem Conf, pp 357–364

Fischbeck G (1992) Barley cultivar development in Europe. Success in the past and possible changes in the future In: Munk L (ed) Proceeding of 6th International Barley Genetics vol II. Munksgaard Int.Publ.Ltd Copenhagen, 887 pp

Forster BP, Thompson DM, Watters J, Powell W (1991) Water-soluble proteins of mature barley endosperm: genetic control, polymorphism, and linkage with β-amylase and spring/winter habit. Theor Appl Genet 81:787–792

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

Igartua E, Casas AM, Ciudad F, Montoya JL, Romagosa II (1999) RFLP markers associated with major genes controlling heading date evaluated in a barley germ plasm pool. Heredity 83:551–559

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

Karsai I, Szucs P, Meszaros K, Filichkina T, Hayes PM, Skinner JS, Lang L, Bedo Z (2005) The Vrn-H2 locus is a major determinant of flowering time in a facultative × winter growth habit barley (*Hordeum vulgare* L.) mapping population. Theor Appl Genet 110:1458–1466

Kihara M, Kaneko T, Ito K, Aida Y, Takeda K (1999) Geographical variation of β-amylase thermostability among varieties of barley (*Hordeum vulgare*) and β-amylase deficiency. Plant Breed 118:453–455

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

Laurie DA, Pratchett N, Bezzant JH, Snape JW (1995) Genome mapping of five major genes and eight quantitative loci controlling flowering time in a winter-spring barley cross. Genome 38:575–585

Li CD, Langridge P, Zhang XQ, Eckstein PE, Rossnagel BG, Lance RCM, Lefol EB, Lu MY, Harvey BL, Scoles GJ (2002) Mapping of Barley beta-amylase alleles in which an amino-acid substitution determines beta-amylase isoenzyme type and the level of free beta-amylase. J Cereal Sci 35:39–50

Lund B, Ortiz R, Skovgaard IM, Waugh R, Andersen SB (2003) Analysis of potential duplicates in barley gene bank collections using re-sampling of microsatellite data. Theor Appl Genet 106:1129–1138

Ma Y, Stewart DC, Eglinton JK, Logue SJ, Langridge P, Evans DE (2000a) Comparative enzyme kinetics of two allelic forms of barley (*Hordeum vulgare* L.) β-amylase. J Cereal Sci 31:335–344

Ma YF, Eglinton JK, Evans DE, Logue SJ, Langridge P (2000b) Removal of the four C-terminal glycine-rich repeats enhances the thermostability and substrate binding affinity of barley β-amylase. Biochemistry 39:13350–13355

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

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

Melchinger AE, Graner A, Singh M, Messmer M (1994) Relationships among European Barley Germplasm: I. genetic diversity among winter and spring cultivars revealed by RFLPs. Crop Sci 34:1191–1199

Pan A, Hayes PM, Chen F, Blake T, Wright S, Karsai I, Bedo Z (1994) Genetic analysis of the components of winter hardiness in barley. Theor Appl Genet 89:900–910

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

Plarre W, Hoffmann W (1963) The development of barley growing and barley breeding in Europe. In: Broekhuizen S (ed) 1st International Barley Genetics Symposium NIBEM and NGC. Wageningen, pp 7–59

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

Rafalski A (2002) Applications of single nucleotide polymorphisms in crop genetics. Curr Opin Plant Biol 5:94–100

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

Russell J, Fuller J, Young G, Thomas B, Taramino G, Macaulay M, Waugh R, Powell W (1997) Discriminating between barley genotypes using microsatellite markers. Genome 40:442–450

Russell JR, Ellis RP, Thomas WTB, Waugh R, Provan J, Booth A, Fuller J, Lawrence P, Young G, Powell W (2000) A retrospective analysis of spring barley germplasm development from 'foundation genotypes' to currently successful cultivars. Mol Breed 6:553–568

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

Sjakste TG, Roder MS (2004) Distribution and inheritance of beta-amylase alleles in north European barley varieties. Hereditas 141:39–45

Taketa S, Kikuchi S, Awayama T, Yamamoto S, Ichii M, Kawasaki S (2004) Monophyletic origin of naked barley inferred from molecular analyses of a marker closely linked to the naked caryopsis gene (nud). Theor Appl Genet 108:1236–1246

Tanno K, Taketa S, Takeda K, Komatsuda T (2002) A DNA marker closely linked to the vrs1 locus (row-type gene) indicates multiple origins of six-rowed cultivated barley (*Hordeum vulgare* L.). Theor Appl Genet 104:54–60

Thacker SP, Ramamurthy V, Kothari RM (1992) Characterisation of barley β-amylase for application in maltose production. Starch 44:339–341

Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res 25:4876–4882

Trevaskis B, Bagnall DJ, Ellis MH, Peacock WJ, Dennis ES (2003) MADS box genes control vernalization-induced flowering in cereals. Proc Natl Acad Sci USA 100:13099–13104

Turner A, Beales J, Faure S, Dunford RP, Laurie DA (2005) The pseudo-response regulator Ppd-H1 provides adaptation to photoperiod in barley. Science 310:1031–1034

Yan L, Loukoianov A, Blechl A, Tranquilli G, Ramakrishna W, SanMiguel P, Bennetzen JL, Echenique V, Dubcovsky J (2004) The wheat VRN2 gene is a flowering repressor down-regulated by vernalization. Science 303:1640–1644

Yoshigi N, Okada Y, Sahara H, Koshino S (1994a) Expression in *Escherichia coli* of cDNA encoding barley β-amylase and properties of recombinant β-amylase. Biosci Biotechnol Biochem 58:1080–1086

Yoshigi N, Okada Y, Sahara H, Koshino S (1994b) PCR cloning and sequencing of the β-amylase cDNA from barley. J Biochem 115:47–51

Yoshigi N, OkadaY, Sahara H, Tamaki T (1995) A structural gene encoding beta-amylase of barley. Biosci Biotechnol Biochem 59:1991–1993

von Zitzewitz J, Szucs P, Dubcovsky J, Yan L, Francia E, Pecchioni N, Casas A, Chen TH, Hayes PM, Skinner JS (2005) Molecular and structural characterization of barley vernalization genes. Plant Mol Biol 59:449–467