Theor Appl Genet (2012) 125:1735–1749
DOI 10.1007/s00122-012-1950-3

ORIGINAL PAPER

Evaluation of diagnostic molecular markers for DUS phenotypic assessment in the cereal crop, barley (*Hordeum vulgare* ssp. *vulgare* L.)

James Cockram · Huw Jones · Carol Norris · Donal M. O'Sullivan

Received: 25 April 2012 / Accepted: 16 July 2012 / Published online: 17 August 2012
© Springer-Verlag 2012

Abstract The deployment of genetic markers is of interest in crop assessment and breeding programmes, due to the potential savings in cost and time afforded. As part of the internationally recognised framework for the awarding of Plant Breeders' Rights (PBR), new barley variety submissions are evaluated using a suite of morphological traits to ensure they are distinct, uniform and stable (DUS) in comparison to all previous submissions. Increasing knowledge of the genetic control of many of these traits provides the opportunity to assess the potential of deploying diagnostic/perfect genetic markers in place of phenotypic assessment. Here, we identify a suite of 25 genetic markers assaying for 14 DUS traits, and implement them using a single genotyping platform (KASPar). Using a panel of 169 UK barley varieties, we show that phenotypic state at three of these traits can be perfectly predicted by genotype. Predictive values for an additional nine traits ranged from 81 to 99 %. Finally, by comparison of varietal discrimination based on phenotype and genotype resulted in correlation of 0.72, indicating that deployment of molecular markers for varietal discrimination could be feasible in the near future. Due to the flexibility of the genotyping platform used, the genetic markers described here can be used in any number or combination, in-house or by outsourcing, allowing flexible deployment by users. These markers are likely to find application where tracking of specific alleles is required in breeding programmes, or for potential use within national assessment programmes for the awarding of PBRs.

Communicated by P. Hayes.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-012-1950-3) contains supplementary material, which is available to authorized users.

J. Cockram (✉) · H. Jones · C. Norris · D. M. O'Sullivan (✉)
John Bingham Laboratory, National Institute of Agricultural Botany (NIAB), Huntington Road, Cambridge CB3 0LE, UK
e-mail: james.cockram@niab.com

D. M. O'Sullivan
e-mail: donal.osullivan@niab.com

Introduction

Barley is ranked fourth in worldwide cereal production, and is used for a variety of end uses, including animal feed, human consumption and malting. The economic importance of barley has meant that marker-assisted breeding approaches are of considerable interest. Deployment of genetic markers that predict the phenotypic trait of interest with 100 % accuracy ('perfect markers') allows efficient tracking of favourable genetic variants through the breeding process, without the need for phenotypic evaluation. However, development of such markers has been slow, partly due to the large size of the barley genome (5,500 Mbp). Accordingly, many of the genetic markers developed for marker assisted selection (MAS) are actually 'diagnostic markers', which predict phenotype with varying degrees of accuracy. Nevertheless, the deployment of genetic markers within breeding programmes can be of considerable economic benefit. Barley has several advantages over related temperate cereal crops for the investigation of the genetic basis of phenotypic diversity. Perhaps the most significant of these is that unlike the related hexaploid cereal crop, wheat (*Triticum aestivum* L., $2n = 6x = 42$), barley is diploid ($2n = 2x = 14$). Furthermore, an increasing array of genomic tools such as bacterial artificial chromosome (BAC) libraries (Yu et al. 2000), chromosome addition/deletion lines (e.g., Islam,

1983), high density SNP arrays (Close et al. 2009), sequenced ESTs (e.g., http://tigr.org/) and full-length cDNA libraries (Matzumoto et al. 2011) are allowing the genetics of barley phenotypic diversity to be investigated with increasing precision. The pace of advance has been aided by next generation sequencing (NGS) approaches, allowing the development of increasingly dense SNP arrays (Ramsay et al. 2011) and novel genotyping and genetic mapping platforms and approaches such as genotyping-by-sequencing (GbS) (Maughan et al. 2010) and genome-wide association mapping (Waugh et al. 2010). Furthermore, using a variety of available genomic resources, the International Barley Sequencing Consortium (IBSC) has recently released a draft 28× coverage barley genome sequence assembly (Mayer et al. 2011) (available at http://mips.helmholtz-muenchen.de/plant/barley/).

A relatively low number of barley phenotypes are currently understood at the causative DNA variant level, and a large proportion of these relate to traits measured as part of the international legal framework established by the International Union for the Protection of New varieties of Plants (UPOV) for awarding Plant Breeders' Rights (PBR, similar to patent or intellectual property rights) to new crop varieties. As part of this evaluation process in the UK, all barley submissions are phenotypicaly assessed using 29 morphological traits to establish entries are distinct, uniform and stable (DUS), in comparison to all previously released varieties (http://www.upov.int/). The potential of genetic markers to replace traditional characteristics has been recognised within UPOV's Biochemical and Molecular Techniques (BMT) Working Group, where three models have been proposed. Here, we will consider Model 1: "Molecular characteristics as a predictor of traditional characteristics: Use of molecular characteristics which are directly linked to traditional characteristics (gene specific markers)". In addition, there is considerable overlap in the traits used for DUS assessment and those deployed to manage, annotate and establish non-redundancy in germplasm collections. While some characters are simply observable, many take a full growing season and dedicated field trials before they manifest, or can only be detected in the homozygous condition. Thus, collectively these characters represent an important tool in germplasm management, breeding and statutory assessment for the awarding of PBR.

At the molecular genetic level, most information is known for the character 'ear row number': depending on whether the two lateral spikelets present at each floret are infertile or fertile, barley varieties are described as 2 or 6 rowed, respectively. Ear row number is controlled by the *VRS1* locus (Komatsuda and Tanno 2004), with three independent mutations in the underlying *HvHOX1* gene conferring the 6-rowed phenotype (Komatsuda et al. 2007). In addition, partial filling of lateral spikelets in lines with dominant 2-row *Vrs1* alleles is controlled by a second locus, *INTERMEDIUM-C* (*INT-C*) (Komatsuda and Mano 2002), recently found to encode an orthologue of the maize meristem identity gene *TEOSINTINE BRANCHED 1* (*TB1*) (Ramsay et al. 2011). The *CLEISTOGAMY 1* (*CLY1*) locus controls open/closed flowering, which within DUS phenotypic assessment is recorded as the 'disposition of lodicules' (fleshy structures at the base of the floret which control floret opening). Two independent synonymous mutations within mir172 mRNA binding sites within the underlying *HvAP2* gene prevent microRNA-directed degradation of *HvAP2* mRNA, resulting in closed flowering due to the failure of the lodicules to expand (Nair et al. 2010). 'Flowering time' is recorded as part of DUS assessment, with four major genes known to largely account for its genetic control: the photoperiod pathway genes *PPD-H1* and *PPD-H2*, and the vernalization response genes *VRN-H1* and *VRN-H2* (Laurie et al. 1995). A natural genetic variant within the *PSEUDO RESPONSE REGULATOR* (*PRR*) gene underlying *PPD-H1* is thought to alter sensitivity to day length (Turner et al. 2005), while *PPD-H2* is thought to encode the putative flowering pathway gene *HvFT3*, with deletion of the gene associated with mutated non-sensitive *ppd-H2* alleles (Faure et al. 2007; Cockram et al. 2010a). The flowering time loci *VRN-H1* and *VRN-H2* also control the DUS trait 'seasonal growth habit' (Cockram et al. 2009).*VRN-H1* is widely viewed to encode a MADS-box transcription factor, with an allelic series of deletions spanning an intronic 'vernalization critical' region thought to confer vernalization insensitive alleles (Cockram et al. 2007a, b, c, 2008; Dubcovsky et al. 2005; Karsai et al. 2005; von Zitzewitz et al. 2005), while *VRN-H2* is thought to be encoded by one ofthree *ZCCT-H* genes that map to the locus (Karsai et al. 2005). Abolishment of vernalization sensitivity at the *VRN-H2* locus is associated with a deletion of all three *ZCCT-H* genes in all germplasm surveyed to date (Cockram et al. 2007b, 2008, 2009, 2010a). Diagnostic PCR/agarose gel markers for both *VRN-H1* (Cockram et al. 2009) and *VRN-H2* (Karsai et al. 2005) have been previously developed.

Comparisons of varietal discrimination based on DUS phenotypic assessment and 'random' molecular markers such as simple sequence repeats (SSRs) has been undertaken in a number of crops (e.g., Gunjaca et al. 2008; Noli et al. 2008; Ibáñez et al. 2009a, b). Investigation of the validity of replacing morphological DUS traits for varietal discrimination with large numbers of molecular markers distributed throughout the genome is termed an 'Option 2' approach by UPOV (UPOV document INF/17/1 http://www.upov.int/). However, to our knowledge, like-for-like comparison of large numbers of DUS morphological traits with the corresponding diagnostic/perfect

genetic markers has not been carried out to date (UPOV 'Option 1' approach). With the aim of determining the genetic control of DUS phenotypes in barley, we recently undertook a GWA approach, utilising a panel of ~600 varieties genotyped with 1,536 SNPs (Cockram et al. 2010b). Using a suite of 32 historic DUS traits, significant marker-trait associations were identified for just under half of all phenotypes investigated. Indeed, mapping resolution was sufficient to fine map a selected trait (anthocyanin pigmentation) to within a 140-kb interval, identifying an exonic deletion resulting in a severe truncation of the predicted *HvbHLH1* protein as the likely causative genetic variant controlling lack of anthocyanin (Cockram et al. 2010b). In addition, GWA scans identified genomic loci controlling eleven of the remaining 29 DUS traits (Cockram et al. 2010b). The high proportion of DUS traits with diagnostic/perfect genetic markers provides an opportunity for the development of a suite of markers of immediate practical use in barley breeding, as well as for addressing the feasibility of a UPOV 'Option I' approach to DUS testing. Here, we describe the deployment of a unified platform to genotype genetic markers for phenotypic traits relevant to the awarding of PBR and germplasm management, and assess their value for application in breeding in an independent panel of barley germplasm.

Materials and methods

Germplasm, DNA extraction and phenotypic data

A collection of ninety barley varieties of predominantly European origin (the 'validation panel') was collated (Online Resource 1). The panel consists of a wide range of varieties of predominantly north-western European origin, many of which have been prominent in the pedigrees of modern UK cultivars. The collection includes varieties belonging to different end-use categories (malting or animal feed) and to the major agronomic groupings (spring-/winter-sown and 2-/6-row ear types), with the aim of ensuring a good representation of the likely allelic variants found in the current elite UK gene pool. The UK barley germplasm panel of 169 varieties (released between 1980 and 2005) is predominantly as described by Cockram et al. (2010b), with the addition of the National Listed varieties Maris Otter, Golden Promise, Triumph, Golf, Klaxon, Halycon, Blanche, Graphic, Tiffany, Acorn, County and NFC Tipple. Seed from each variety was grown to the two-leaf stage, and genomic DNA extracted from single leaves using the DNA Easy 96 Extraction Kit (Qiagen). DNA quality was assessed by running 2 µl aliquots of each extraction on an ethidium bromide stained 1.5 % agarose gel, and visualised under UV light. In addition, DNA quantity was determined using a Nanodrop 200 spectrophotometer (Thermo Scientific). Samples were diluted to a final concentration of 7 ng/µl using sterile water. Phenotypic data for DUS traits in UK germplasm were collated with breeder permission from records held at NIAB.

Molecular data, genotyping and data analysis

Genetic markers tagging barley phenotypic traits were identified by searches of published literature, and are represented either by single nucleotide polymorphisms (SNPs) or insertion/deletions (InDels). Reference DNA sequences sourced from public databases were manipulated and analysed using the Vector NTI Advance package v10.1.1 (Invitrogen). Genotyping was performed using the KASPar genotyping system, based on single-plex technology employing a universal fluorescent reporting system (http://www.kbioscience.co.uk/). For each KASPar assay, the following minimum details were recorded: (a) annotated genomic DNA sequence (b) the GenBank or HarvEST U32 Unigene accession number of the reference allele DNA sequence (c) a PMID number, linking to the relevant scientific publication describing allelic variants (d) genetic map position of the gene assayed (e) Information describing the genetic variants and their associated phenotypes. The polymorphic DNA features to be assayed, along with ~100 bp of surrounding sequence, were submitted to KBioscience for assay design using the format shown below, where the two allelic states at the target SNP are separated by the symbol '/' and enclosed within square brackets, known polymorphic base-pairs are indicated using standard nomenclature, and bases to avoid during primer design are replaced with the letter 'N':

GCGGGGCCGGTGCCGAACAACGCTGCCGCCGCCGCAGCAGCAGC**M**GCAGC**[A/G]**TCATC**C**CGATTCCCACCCTACATCGCCA**R**GCAGGCGCAGAGCTGGCTC**NN.**

The resulting KASPar assays were assessed using the 'validation panel'. Successfully converted markers were subsequently genotyped on the panel of 190 UK varieties. For all assays, 7 ng genomic DNA was used as a template. A negative water control was also included for each genotypic assay. Genotypic data for markers found to be convertible to the KASPar system were returned as .csv files, and viewed using SNP Viewer v1.99 (https://kbiosciences.co.uk/). DNA sequences and additional information for all markers developed are available at http://www.niab.com/mas/. In addition to SNP loci, KASPar assays may be designed to genotype InDels, where insertions are detected when the base adjacent to the deletion is called, and deletions called as 'unknown' data points (grouping with the negative water control) due to the failure of primers to anneal. Such markers are termed as 'pseudo-SNPs.

Diagnostic PCR/agarose gel assays for *VRN-H1* (Cockram et al. 2009) and *VRN-H2* (Karsai et al. 2005) are previously described. For these two markers, genotypes were sourced from Cockram et al. (2009) where common varieties/DNAs were utilized; for all other varieties, genotyping was performed de novo. Inter-varietal genetic distances were calculated using Roger's distance, implemented in PowerMarker v3.25 (Liu and Muse 2005). For phenotypic data, Euclidean distance was calculated using the statistics package R v2.9.0 (http://www.R-project.org/). The resulting matrices were used to calculate neighbour-joining trees using R. Phenotypic and genotypic distance matrices were compared by calculating correlation coefficients using R.

Results

Identification of genetic polymorphisms

Of the 29 phenotypes evaluated under barley DUS assessment (Online Resource 2), surveys of scientific literature show that genetic loci for 17 have previously been identified, either in bi-parental or association mapping populations (Table 1). Three traits are controlled by cloned genes, or are thought to be encoded by candidate genes for which diagnostic markers are available: characters 13 (controlled by *VRS1* and *INT-C*), 27 (*CLY1*) and 29 (*VRN-H1, VRN-H2*). In addition, GWA scans have previously genetically mapped loci controlling eleven additional DUS traits to a resolution of ~5 cM (Cockram et al. 2010b): 'Plant growth habit' (UPOV character 1), 'Hairiness of leaf sheaths' (2), Flag leaf: anthocyanin colouration of auricles' (3), 'Flag leaf: intensity of anthocyanin colouration of auricles' (4D), 'Awns: intensity of anthocyanin colouration of tips' (8), 'Awns: intensity of anthocyanin colouration' (9D), 'Ear: development of sterile spikelets' (CPVO character 18D), 'Sterile spikelet: attitude' (20), 'Grain: rachilla hair type' (22), 'Grain: anthocyanin colouration of lemma nerves' (24), 'Grain: spiculation of inner lateral nerves' (25), 'Grain: hairiness of ventral furrow (26), and 'Kernel: colour of aleurone layer' (28). Indeed, a subset of characters were identified as possessing significant GWA in genomic locations estimated to correspond to known morphological loci identified in bi-parental mapping populations. These include hairiness of leaf sheath (trait 2, *HSH1* locus, Lundqvist et al. 1996), sterile spikelet morphology (CPVO trait 18D, UPOV trait 20, *SLS* locus, Franckowiak, 1995), rachilla hair type (trait 22, *SRH*, Lundqvist et al. 1996), grain lateral nerve spiculation (trait 25, *GTH1,* Lundqvist et al. 1997), and aleurone colour (trait 28, *BLX*, Lundqvist et al. 1996). Finally, we hypothesise that 'time of ear emergence' (trait 7) and 'plant length' (trait 12) may be at least partly controlled by known flowering time (*VRN-H1*, *-H2*; *PPD-H1*, *-H2*, reviewed by Cockram et al. 2007a) and dwarfing (*HvBRI1*, Chono et al. 2003) loci, respectively. The genetic loci controlling the remaining 12 traits (characters 5–6, 10D–11, 14–19, 21, 23) are currently unknown in UK germplasm, and are not investigated further in this study.

Marker development and genotypic analysis

The genetic polymorphisms identified in literature searches were converted for use with the KASPar SNP genotyping platform. Marker details and information regarding interpretation of SNP calls are listed in Table 1. Of the 30 polymorphisms originating from 19 genes, assay design returned 25 putatively usable assays (83 % design success), all of which were found to work when applied in practice (Table 1). Genotyping success rate was found to be ~99 %, and included SNP, as well as pseudo-SNP, assays (Fig. 1). Allele frequencies and percent call rate for all markers are listed in Table 1. The 25 validated markers were subsequently genotyped across a panel of 169 UK barley varieties, returning ~4,000 high quality data-points (Online Resource 3), with a mean missing score rate of <1 %. A wide range of minor allele frequencies (MAF) were observed, ranging from 0.01 to 0.49 (mean MAF = 0.23, median MAF = 0.24). Three DUS-related markers were found not to be polymorphic in the UK set: (1) marker HvBRI1_A2570G, a putative diagnostic SNP at the *uzu* dwarfing gene (Chono et al. 2003), common in Asian barley. (2) Marker HvSdw1_AG, which tags a polymorphism within the candidate gene for the *Sdw1* locus controlling height (Jia et al. 2009). (3) HvCly1_A2664C, which represents one of the two perfect markers for cleistogamy (trait 26). Lack of polymorphism at this marker in the UK panel appears to be due to the observation that open-flowering varieties with the 'frontal' bib disposition are very rare in modern UK varieties, showing that when 'frontal' disposition is present, the alternative polymorphism has been deployed. Of the three markers listed above, HvSdw1_AG and HvCly1_A2664C were found to be polymorphic in the 'validation panel'.

Analysis of DUS marker genotypes

To determine the predictive value of the genetic markers investigated, we constructed a phenotypic database for the current set of 29 DUS traits, scored across the 169 UK varieties investigated (Online Resource 2). Each trait differed in the number of records that were available, ranging from 76 % (trait 20) to 95 % fill (traits 13, 22 and 26). The resulting predictive values (based on the percentage correctly called trait scores, as predicted by marker genotype) of genetic markers for their relevant trait varied relatively

Table 1 Conversion and validation of diagnostic/perfect markers for DUS traits on the 'validation panel'

Marker names	UPOV character	Putative genetic locus	Gene/candidate gene	Gene[b]	Cr	Reference accession[c]	PM	Alleles	PMID	Allele 1, allele 2, het, missing	Notes
HvFT3_FC816A	1	*PPD-H2*	*HvFT3*	L	1H	DQ411319	InDel	PS	17339225	0.84, 0.16, N/A, N/A	Associated with erect (C) or prostrate (Del) growth habit
Hv11_11299	2	*HSH1*	N/A	L	4H	U32 9715	SNP	C/G	21115826	0.61, 0.36, 0.03, 0	Associated with presence (C) or absence (G) of hairs
HvANT2_3583InDel	3, 4D, 8, 9, 24	*ANT2*	*HvbHLH1*	CG	2H	HM163343	InDel	PS	21115826	N/A	Associated with presence (Ins) or absence (Del) of pigment
HvANT2_C4289T	3, 4D, 8, 9, 24	*ANT2*	*HvbHLH1*	CG	2H	HM163343	SNP	C/T	21115826	0.89, 0.10, 0.01, 0	Associated with presence (C) or absence (T) of anthocyanin
HvPPDH1_T3081G	7	*PPD-H1*	*HvPRR7*	UG	2H	AY943294	SNP	T/G	16284181	0.72, 0.27, 0, 0.01	Perfect marker: LD responsive (G) or non-responsive (T)
HvPPDH1_A2721G	7	*PPD-H1*	*HvPRR7*	UG	2H	AY943294	SNP	A/G	18669581	0.72, 0.26, 0, 0.02	Jones et al. SNP: LD responsive (C) or non-responsive (T)
HvVRNH3_A471T	7, 29	*VRN-H3*	*HvFT1*	UG	7H	BGS213	SNP	A/T	17158798	0.11, 0.84, 0, 0.04	Associated with spring (A) or winter (T) allele
HvVRNH3_G585C	7, 29	*VRN-H3*	*HvFT1*	UG	7H	BGS213	SNP	G/C	17158798	0.11, 0.88, 0.01, 0	Associated with spring (A) or winter (T) allele
HvVRNH1_SNP2	7, 29	*VRN-H1*	*HvBM5A*	UG	5H	AY750993	SNP	A/G	17713756	0.27, 0.6, 0.07, 0.07	Associated with hap 1A/1B (A) or 2/3/4A/4B/5A/5B/5C (G)
HvVRNH1_8P9_InDel	7, 29	*VRN-H1*	*HvBM5A*	UG	5H	AY750993	InDel	PS	17713756	0.12, 0.88, N/A, N/A	Perfect marker: spring (-:-) or winter (all other calls)[a]
HvVRNH1_0P5_InDel	7, 29	*VRN-H1*	*HvBM5A*	UG	5H	AY750993	InDel	PS	17713756	0.97, 0.03, N/A, N/A	Perfect marker: spring (-:-) or winter (all other calls) vern resp
HvVRNH1_various[e]	7, 29	*VRN-H1*	*HvBM5A*	UG	5H	AY750993	InDel	PS	17713756	N/A	N/A
ZCCT-Ha	7, 29	*VRN-H2*	*ZCCT-H*	CG	4H	DQ492695	InDel	PS	17220272	N/A	Associated with winter (A) or spring (Del) vernalization allele
ZCCT-Hb	7, 29	*VRN-H2*	*ZCCT-H*	CG	4H	AY485978	InDel	PS	17220272	N/A	Associated with winter (T) or spring (Del) vernalization allele
ZCCT-Hc	7, 29	*VRN-H2*	*ZCCT-H*	CG	4H	AY750993	InDel	PS	17220272	N/A	Associated with winter (A) or spring (Del) vernalization allele
HvBRI1_A2570G	12	*UZU*	*HvBRI1*	CG	4H	AB048949	SNP	A/G	14551335	0.97, 0, 0, 0.03	Perfect marker: wild-type (A) or dwarf (G) height
HvSdw1_AG	12	SDW1	*Ga20-ox*	CG	3H	N/A	SNP	A/G	19280236	0.47, 0.50, 0, 0.03	Associated with short (A) or tall (G) in bi-parental pop
HvVRS1_TINS243	13	*VRS1*	*HvHOX1*	UG	2H	EF067844	InDel	PS	17220272	N/A	Perfect marker: 2- (Del) or 6- (Ins) row. Pseudo-SNP
HvVRS1_C349G	13	VRS1	*HvHOX1*	UK	2H	EF067844	SNP	C/G	17220272	0.94, 004, 0, 0.1	Perfect marker: 2- (C) or 6- (G) row
HvVRS1_GINS681	13	*VRS1*	*HvHOX1*	UG	2H	EF067844	InDel	PS	17220272	0.93, 0, 0, 0.07	Perfect marker: 2- (Ins) or 6- (del) row. Pseudo-SNP

Table 1 continued

Marker names	UPOV character	Putative genetic locus	Gene/ candidate gene	Gene[b]	Cr	Reference accession[c]	PM	Alleles	PMID	Allele 1, allele 2, het, missing	Notes
11_20606	13	*INT-C*	*HvTB1*	L	4H	U32 3687	SNP	C/G	21115826	0.16, 0.83, 0, 0.01	Associated with *INT-C*.b1/.b2 (C) and *INT-C*.a (G) alleles
HvINTC_C124G	13	*INT-C*	*HvTB1*	UG	4H	d	SNP	C/G	21217754	0.81, 0.11, 0, 0.08	Associated with *INT-C*.b1/.b2 (C) and *INT-C*.a (G) alleles
HvINTC_C498T	13	*INT-C*	*HvTB1*	UG	4H	d	SNP	C/T	21217754	0.83, 0.10, 0, 0.07	Associated with *INT-C*.b2 (T) and *INT-C*.a/.b1 (C) alleles
Hv11_10933_GC	20	*VRS3*?	N/A	L	1H	U32 7800	SNP	G/C	21115826	0.68, 0.31, 0, 0.01	Associated with parallel (G) or divergent (C) spikelet attitude
Hv11_20850_AG	21	*SRH*	N/A	L	5H	U32 5004	SNP	G/A	21115826	0.49, 0.51, 0, 0	Associated with short (G) or long (A) rachilla hair
Hv11_10818_CA	25	*GTH1*	N/A	L	2H	U32 6023	SNP	C/A	21115826	0.42, 0.54, 0, 0.04	Associated with strong (C) or no/ increasing (A) spiculation
Hv2g01490_G607A	26	N/A	N/A	L	6H	JQ356757	SNP	G/A	21115826	0.64, 0.33, 0, 0.02	Associated with absence (G) or presence (A) of hairs
HvCly1_A2604G	27	*CLY1*	*HvAP2*	UG	2H	GQ403050	SNP	A/G	20018663	0.96, 0. 04, 0, 0	Perfect marker: open (A) or closed (G) flowering
HvCly1_A2664C	27	*CLY1*	*HvAP2*	UG	2H	GQ403050	SNP	A/C	20018663	0.97, 0, 0.01, 0.02	Perfect marker: open (A) or closed (C) flowering
HvOs03g14380_G125A	28	*BLX1*	N/A	L	4H	JQ356758	SNP	G/A	21115826	0.69, 0.28, 0, 0.03	Associated with white (G) or coloured (A) aleurone

DUS traits are coded as listed in Online Resource 2. Markers that failed to convert to KASPar are highlighted in grey. "-:-" Represents insertions (i.e., lack if 8.9 kb deletion). All others call assay the 8.9 kb deletion (spring allele) described by Cockram et al. (2007b)

Cr chromosome, *Ref* reference, *PM* polymorphism, *PMID* PubMed identifier for reference manuscript, *het* heterozygotes, *UG* underlying gene, *CG* candidate gene, *L* linked gene, *InDel* insertion/deletion, *PS* pseudo-SNP, *N/A* (not applicable/unknown)

[a] Manual interpretation of allele calling needed to call alleles

[b] Gene from which the assayed genetic marker is derived

[c] HarvEST U32 accessions are prefixed with 'U32', all others are NCBI accession numbers

[d] None available, see PMID 21217754

[e] Numerous *VRN-H1* InDel pseudo-SNPs failed to convert to KASPar (not shown here), assaying for the 4.1, 5.2, 6.3 intron I deletions and the 0.7 kb insertion, described by Cockram et al. (2007b)

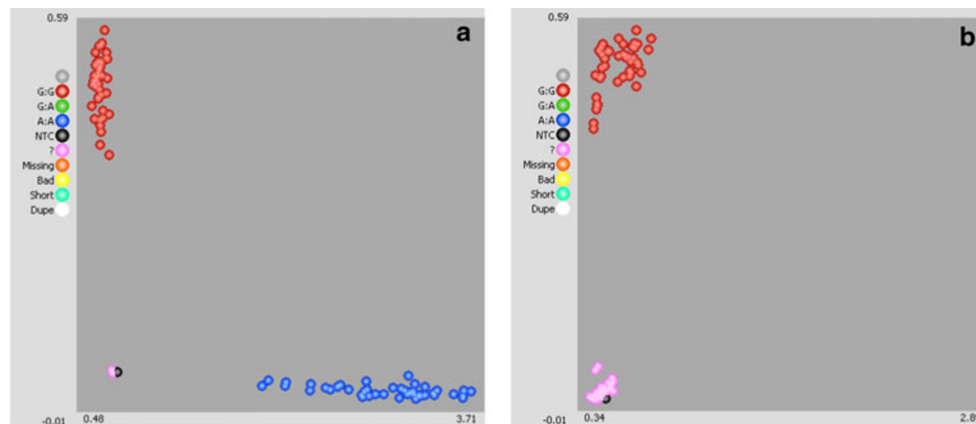

Fig. 1 Examples of allele calling for different classes of genetic polymorphism. **a** SNP (sterile spikelet attitude, marker Hv11_20850_AG) assaying for a G/A SNP. **b** InDel pseudo-SNP (seasonal growth habit, marker HvVRNH1_hap2_InDel), assaying for a 0.7 kb InDel within the vernalization critical region of the flowering time locus *VRN-H1*. G, Deletion; ?, Insertion (called as negative results, grouped with the water control). *Black data points* water negative control

widely (Table 2), and are divided into four broad categories.

Traits with perfect (100 %) phenotypic prediction

Molecular markers for three traits ('Ear: number of rows', 'Grain: disposition of lodicules' and 'Seasonal growth habit') were found to be 100 % predictive of phenotypic state. All three traits are scored as binary characters and the genetic markers deployed originated from map-based cloned genes. 'Ear row number' (trait 13) is controlled by one of the three mutations in *VRS1* (chromosome 2H), requiring a combined haplotype to perfectly predict phenotype. However, only two of the *VRS1* mutations were found to be convertible to the KASPar platform (HvVRS1_C349G and HvVRS1_GINS681) resulting in a predictive value of just 95 %. However, marker Hv11_20606_GC, in close linkage with the *INT-C* locus on chromosome 4H, displayed a predictive value of 100 %. Similarly, haplotype analysis determined from two SNPs within the gene encoding *INT-C* perfectly predicted row number, with haplotypes CT/CC and GC diagnostic for 2- and 6-row types, respectively. A marker prediction score of 100 % was also observed for 'Grain: disposition of lodicules' marker HvCLY1_A2604G, which assays for one of the two mutations at the *HvAP2* gene previously shown to control this trait (Nair et al. 2010). Only one variety within the UK panel was recorded as possessing the frontal "bib" type lodicules disposition (AFP 2/1091), which was predicted by the presence of a G nucleotide. Although the second mutation at *CLY1* was also assayed (marker HvCLY1_A2662C), it was found to be monomorphic in the UK varietal panel. Similarly, the genes underlying genetic control of "Seasonal growth habit" are relatively well characterised. Genotypic analysis using two KASPar markers from the *VRN-H1* locus (HvVRNH1_SNP2 and HvVRNH1_0P5_InDel) results in a predictive power of 99.3 %. Genetic markers assaying for the InDel of each of the three candidate genes underlying *VRN-H2* failed to convert to the KASPar platform.

Traits with very good (90–99 %) phenotypic prediction

This group predominantly contains traits with binary or three-state phenotypic scores, for which the gene/genetic variant underlying Mendelian genetic locus has yet to be cloned. In most cases, the genetic markers deployed originate from genes closely linked to the underlying locus, resulting in high predictive values, and provide good potential targets for future fine mapping. For character 'lower leaves: hairiness of leaf sheaths' (trait 2), marker Hv11_11299_GC returns a 96 % predictive value, with SNP G and C predictive of the absence (score 1) and presence (score 9) of hairs, respectively. Similarly high marker predictions were achieved for the three state character 'kernel: colour of aleurone layer' (trait 28), for which marker HvOs03g14250_C82T displays a predictive value of 92 %, based on SNP G predicting score 1 (white) and SNP A predictive for scores 2–3 (weakly–strongly coloured). Finally, although traits scored on a more continuous scale of $\geq$3 character states are more problematic for the development of molecular markers as diagnostic tools, fitting binary genetic markers does in some cases result in the identification of markers significantly associated with the trait. In practice, this often means that one allele is associated with a single strait score (e.g., absence), while the alternative SNP is associated with the remaining trait scores (e.g., increasing presence). This is true of the three anthocyanin intensity-related phenotypes (traits 3, 9D and 24). Failure to convert the diagnostic

Table 2 Predictive value of a subset of the genetic markers relevant to DUS traits

UPOV No.	DUS character	Genetic marker	No. Vars genotype and phenotype	No. correct pred.	% Correct pred.
1	Plant growth habit[a]	HvFT3_FC816A	122	107	87.7
2	Lower leaves: hairiness of leaf sheaths	Hv11_11299_GC	156	151	96.2
3	Flag leaf: anthocyanin colouration of auricles[b]	HvANT2_C4289T	146	144	98.0
9D	Awns: intensity of anthocyanin colouration of awn tips[b]	HvANT2_C4289T	148	145	97.3
24	Grain: anthocyanin colouration of lemma nerves[b]	HvANT2_C4289T	153	142	92.2
13	Ear: number of rows[c]	HvVRS1_C349G	160	152	94.4
13	Ear: number of rows[c]	HvVRS1_GINS681	160	145	90.1
13	Ear: number of rows[d]	HvVRS1_C349G & HvVRS1_GINS681	159	152	95.0
13	Ear: number of rows[e]	Hv11_20606_GC	157	157	100
13	Ear: number of rows[d]	HvINTC_C124G & HvINTC_C498T	156	156	100
20	Sterile spikelet: attitude (mid 1/3 of ear)[f]	Hv11_10933_GC	128	113	87.6
22	Grain: rachilla hair type	Hv11_10622_GA	152	104	68.0
22	Grain: rachilla hair type	Hv11_20850_AG	160	111	68.9
25	Grain: spiculation of inner lateral nerves[g]	Hv11_10818_CA	157	92	58.2
26	Grain: hairiness of ventral furrow[h]	HvOs02g01490_G607A	161	132	81.5
27	Grain: disposition of lodicules[i]	HvCly1_A2604G	155	155	100
27	Grain: disposition of lodicules[j]	HvCly1_A2664C	156	155	98.7
28	Kernel: colour of aleurone layer[k]	HvOs03g14380_G125A	158	146	92.4
29	Seasonal type[l]	VRN-H1 Multiplex PCR	143	143	100.0
29	Seasonal type	HvVRNH1_SNP2	137	129	94.2
29	Seasonal type[m]	HvVRNH1_0P5_InDel	142	4	2.8
29	Seasonal type[n]	HvVRNH1_SNP2 & HvVRNH1_0P5_InDel	135	134	99.3

Good phenotypic predictions (>90 % accuracy) are obtained by molecular markers for eight DUS traits

[a] SNP C associated with scores 1–4 (erect–semierect/intermediate), Del associated with scores 6–9 (intermediate/semiprostrate–prostrate. Varieties with score 5 (intermediate) were removed from the analysis

[b] In LD with causative InDel. SNP T predictive of score 1 (absence of anthocyanin), SNP C predictive of scores 2–9 (increasing presence)

[c] Causative SNP (1 of 3)

[d] Haplotype

[e] Not causative, linked to *INT-C* locus: row number ideotype

[f] 6-row varieties excluded from analysis. SNP G predictive of scores 1 (parallel) and 2 (parallel–divergent), SNP C predictive of score 3 (divergent)

[g] SNP A predictive of score 1 (absent/very weak), SNP C predictive of scores 2–9 (very weak–very strong)

[h] SNP G predictive of score 1 (absent/very weak), SNP A predictive of scores 2–9 (very weak–very strong)

[i] Causative locus. Only 1 example of frontal ‘bib’ type lodicule disposition (score 1)

[j] Causative locus. Only 1 example of frontal ‘bib’ type lodicule disposition (score 1). No polymorphism in UK lines assayed

[k] SNP G predictive of score 1 (white), SNP A predictive of scores 2–3 (weakly–strongly coloured)

[l] *VRN-H1* multiplex PCR assay

[m] Not diagnostic, as SNP A can confuse winter *VRN-H1* haplotype 1A with spring haplotype 1B, and wrongly predicts winter haplotype 5C (as Cockram et al. 2007b)

[n] Haplotype. As for HvVRNH1_SNP2, but with ‘Del’ diagnostic for winter haplotype 5C; will not discriminate between winter haplotype 1A and spring haplotype 1B (as Cockram et al. 2007b)

HvBLH1 InDel (Cockram et al. 2010b) to the KASPar platform meant that a closely linked SNP within the gene was utilized (HvANT2_C4289T). For the three related anthocyanin-related traits, SNP T is predictive of score 1 (absence of anthocyanin), while SNP C is predictive of scores 2–9 (increasing anthocyanin intensity), returning

predictive values for traits 3, 7 and 23 of 98, 97 and 92 %, respectively.

Traits with good (80–90 %) phenotypic prediction

Three DUS traits were found to result in 80–90 % successful prediction of phenotype. For the three state character 'Sterile spikelet: attitude' (trait 20), marker Hv11_10933_GC resulted in a predictive value of 88 %, with SNP G predictive of scores 1 (parallel) and 2 (parallel–divergent) and SNP C predictive of score 3 (divergent). Phenotypic state for the binary trait 'Grain: hairiness of ventral furrow' (trait 26) was predicted with 81.5 % accuracy using marker HvOs02g01490_G607A (previously shown to be in linkage with the putative underlying locus *HSH1* Cockram et al. 2010a), with SNP calls of A and G predictive of the presence or absence of hair, respectively. 'Plant growth habit' (trait 1), is scored on a 1–9 scale as erect (score 1)—prostrate (score 9). The putative flowering pathway gene HvFT3 has previously been shown to be associated with this trait (Cockram et al. 2010b), and the corresponding marker *HvFT3*_FC816_A was designed to assay for the insertion/deletion of the gene. SNP C is called when *HvFT3* is present, while a 'missing' score assays for deletion of the gene. This marker fitted the phenotypic data best when 'C' is predictive of scores 1–4 (erect–semierect/intermediate) and 'Del' is predictive of scores 6–9 (intermediate/semiprostrate–prostrate), returning a predictive value of 87.7 %.

Traits with <80 % phenotypic prediction

'Grain: rachilla hair type' (trait 22) was tagged by marker Hv11_20850_AG, which maps close to the underlying *SRH* locus (Cockram et al. 2010b). The 'A' (long hair) → 'G' (short) SNP returns in a predictive value of 68.9 % in the panel of UK cultivars investigated. For the trait 'grain: spiculation of lateral nerves' (trait 25), marker Hv11_10818_CA was found to return a 58.2 % predictive value. Finally, we analysed the genotypes and haplotypes obtained using KASPar markers for known flowering time genes *VRN-H1*, *VRN-H3*, *PPD-H1* and *PPD-H2* (Table 1), as well as the PCR/agarose-gel based makers for *VRN-H1* and *VRN-H2*. No association with flowering time within winter or spring cultivar pools was identified.

Comparison of cluster analysis based on phenotypes and molecular markers

To determine the effectiveness of varietal discrimination using selected DUS phenotypes (1, 2, 3, 9D, 13, 20, 22, 24, 26, 27, 28, 29) versus their corresponding genetic markers, we performed cluster analysis based on the twelve DUS traits for which genetic markers proved highly (>80 %) predictive, and their twelve corresponding genetic markers: traits 1 (marker HvFT3_FC816A), 2 (HvOs03g01380_A447G), 3 (HvANT2_C4289T), 9D (HvANT2_C4289T), 13 Hv11_20606_GC), 20 (Hv11_10933_GC), 22 (Hv11_20850_AG), 24 (HvANT2_C4289T), 26 (HvOs02g01490_G607A), 27 (HvCLY1_A2604G), 28 (HvOs03g14380_G125A) and 29 (HvVRNH1_SNP2_0P5_HAP, which represents a combined haplotype from two of the KASPar assays). All varieties with >50 % missing data in either set (phenotype or genotype) were removed, leaving 158 varieties for subsequent analyses. Inter-variety distances were calculated and the resulting matrices used to calculate neighbour-joining trees based on phenotypic and genotypic data (Fig. 2). The two distance matrices produced were compared by calculating a correlation coefficient. The correlation was both high and positive (0.72), and shown to be highly significant by permutation ($p < 0.001$, 1,000 permutations). Cluster analysis using the twelve phenotypes was able to uniquely identify 88 % of the varieties included. Cluster analysis using the corresponding molecular markers was not able to achieve comparable resolution, with two large groups of 22 and 34 spring 2-rowed varieties showing 100 % genotypic identity. However, broad agreement between clusters calculated from phenotypic and genotypic data was evident.

Discussion

The advent of NGS platforms and SNP arrays has meant that while the cost of SNP identification and genotyping has fallen dramatically, phenotyping costs remain relatively high. Thus, deployment of genetic markers within plant breeding programmes is likely to become increasingly beneficial. While exploration of the UPOV Option II approach has been relatively well documented in crop species, the Option I approach has received much less attention, predominantly due to the lack of available diagnostic/perfect genetic markers. However, advances in the molecular genetics underpinning DUS traits are now facilitating assessment of UPOV Option I approaches in crops. For example, Cockram et al. (2009) developed and assessed genetic markers for the barley vernalization loci controlling the DUS trait 'seasonal growth habit', finding perfect prediction of winter and spring type. More recently, Arens et al. (2010) evaluated a set of molecular marker assays from genes controlling four disease resistance traits assessed during DUS assessment in tomato, finding highly correlated results between biological and molecular results. In this study, we source genetic polymorphisms in barley relevant to traits scored during the awarding of PBR and germplasm collection characterization, test these using a

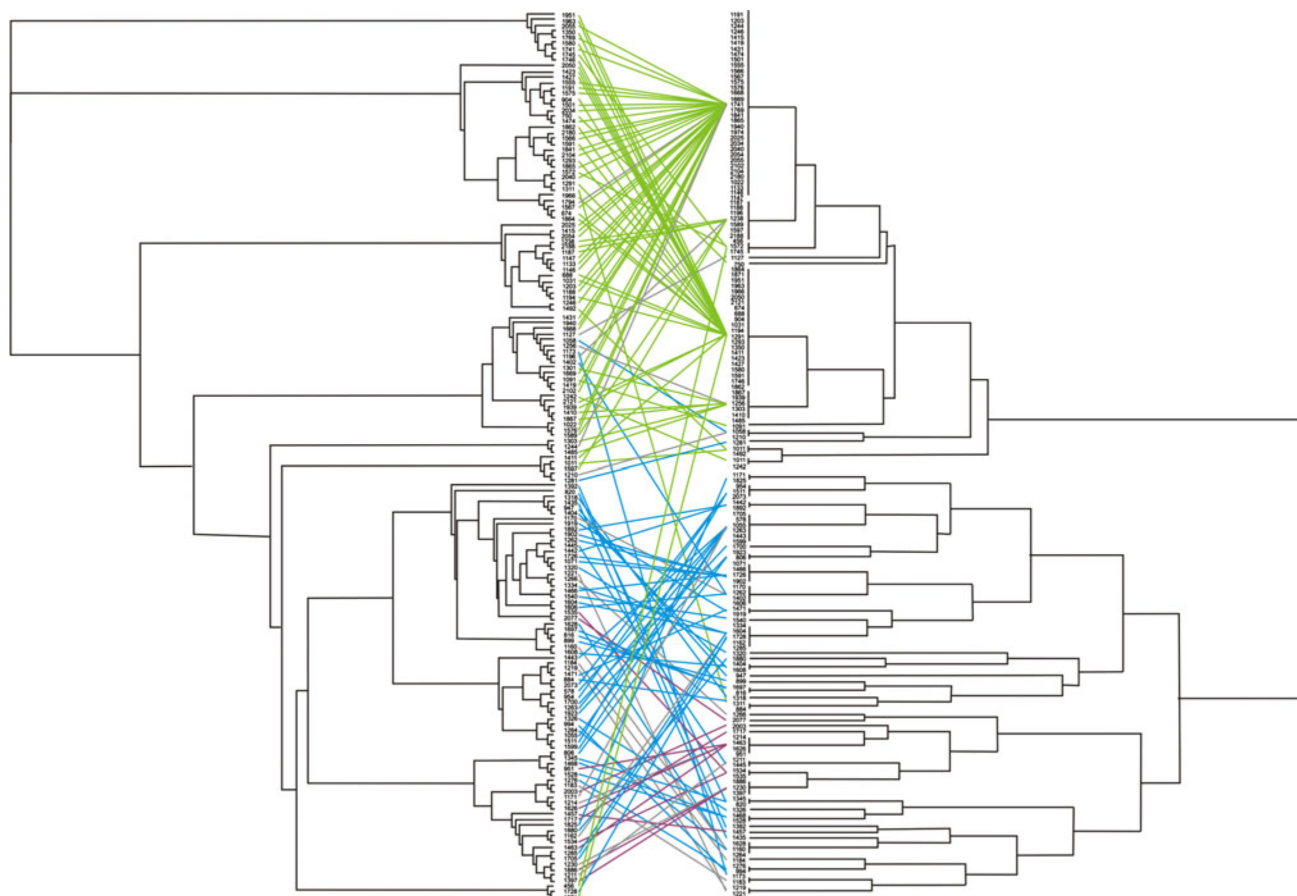

Fig. 2 Correspondence between phenotypic (*left of figure*, calculated using Euclidean distance) and genotypic (*right of figure*, calculated using Rogers' Distance) cluster analyses, based on DUS traits with >80 % marker prediction (1, 2G, 3, 7, 11G, 19, 21, 23, 25G, 26, 27, 28G), demonstrating broad agreement between clusters calculated using phenotypic and genotypic distance measures. Spring 2-row, winter 2-row and winter 6-row varieties are linked by *green*, *purple* and *blue lines*, respectively. Varieties for which seasonal type and/or row number are not known are linked by *grey lines*

unified and flexibly deployable platform for their genotypic assessment, and assesses their diagnostic and discriminative power within a representative collection of UK germplasm. With the exception of the four pseudo-SNPs, all of the genetic markers developed here are co-dominant, allowing efficient identification and tracking of heterozygotes through breeding programmes, for varietal discrimination, and potentially for determination of varietal uniformity. The suite of markers investigated has capacity for phenotypic prediction, ranging from perfect predictive scores of 100 %, down to values of 58 %. Genetic markers for ten traits returned predictive scores of 80–100 %. UPOV Option 1 advocates the deployment of molecular markers for DUS traits directly tagging the underlying genetic determinants. Accordingly, markers derived from causative mutations prevent the risk of loosing marker-trait association due to recombination. However, genetic markers in linkage disequilibrium (LD) with causative mutations may also be deployed, ranging from additional (non-causative) polymorphisms within underlying gene (as exemplified by *INT-C*), to DNA polymorphisms in physical proximity to the causative polymorphism. The utility of the latter is largely determined by local rates of recombination, which vary depending on genomic location. For example, colinearity with rice shows the highly non-recombining regions that span barley centromeres' large physical distances are often colinear with whole rice chromosomes (Cockram et al. 2010b). Thus, the variation in predictive power afforded by 'non-perfect' markers investigated in this study is a function of physical proximity and local recombination rates, as well as allele frequency and underlying genetic complexity of the trait.

Interpretation of 'perfect' markers

The wide range of phenotypic prediction observed is likely to be due to a combination of factors. The best predictions were achieved for binary traits where the underlying genetics are relatively simple, and for which causative

polymorphisms have been identified. However, even in such apparently simple cases, factors such as epistatic interactions between genes, failure of conversion to the KASPar genotyping platform (predominantly for InDel polymorphisms) and the prospect of additional uncharacterized causative mutations at the underlying loci, mean that perfect prediction of phenotype is not necessarily guaranteed. This is exemplified by the trait 'Ear: number of rows', controlled by allelic interactions between the *VRS1* and *INT-C* loci. First, one of the three causative polymorphisms as *VRS1* (single nucleotide InDel, HvVRS1_TINS243) failed to convert to KASPar, resulting in >100 % predictive value based on haplotype analysis of the two remaining *VRS1* genetic markers. However, genetic markers diagnostic for the *INT-C* locus were found to perfectly predict ear row number. Although the causative polymorphisms underlying natural variation for *INT-C* have yet to be determined (Ramsay et al. 2011), wild-type and mutant alleles can be distinguished using two SNPs within the gene (HvINTC-C124G, HvINTC_C498T). Perfect allelic partitioning between 2- and 6-rowed barley for *INT-C* alleles that prevent or promote anther development in lateral spikelets, explains the perfect association with ear row number observed in this study. However, it should be noted that such partitioning of alleles between *VRS1* and *INT-C* may not necessarily be observed in non-elite barley germplasm. Furthermore, even in elite cultivars, recombination between the *INT-C* SNPs assayed and the presently unknown causative *INT-C* polymorphism(s) could lead to less than perfect prediction of ear row number. Similarly, although the genetics and genetic polymorphisms at the *VRN-H1* and *VRN-H2* loci controlling 'seasonal growth habit' are well defined, due to the number (≥5) and nature (InDels) of the different alleles at the *VRN-H1* locus that control spring seasonal growth habit (Cockram et al. 2007b, c), it is problematic to accurately predict phenotype using a single KASPar genetic marker. Despite this, we found a combination of two KASPar *VRN-H1* markers to result in a predictive power of 99 %. We note that this marker combination is known to be unable to distinguish the spring *VRN-H1* haplotype 1B from winter haplotype 1A (Cockram et al. 2007b), providing a possible explanation for the observed prediction rates <100 %. Furthermore, allelic state at *VRN-H2* also affects phenotype. However, as no UK spring varieties with winter alleles at *VRN-H1* have been identified to date (Cockram et al. 2007b, c, 2008, 2009), lack of a KASPar assay for *VRN-H2* may not always prove critical when applied in practice to UK germplasm. We note that perfect predictive power for seasonal growth habit in the UK germplasm collection investigated here was achieved using the previously developed PCR/agarose gel-based markers for *VRN-H1* (Cockram et al. 2009) and *VRN-H2* (Karsai et al. 2005) (Table 2).

Interpretation genetic markers with high phenotypic prediction

High marker-trait correlations are also predicted for traits with a low number (≤3) of character states for which the underling genetics are relatively simple, but where the underlying gene(s) have yet to be identified. Four DUS traits were found to be in this class, and all represent promising targets for map-based cloning of the underlying genes. The first, 'lower leaves: hairiness of leaf sheaths', returned predictive values of 96 % using marker Hv11_11299_GC. This SNP is currently the closest known genetic marker to the *HSH1* locus on the long arm of chromosome 4H that controls the trait (Cockram et al. 2010a). The high correlation between genotype and phenotype and well-established colinearity between barley chromosome 4H and colinear chromosomes from related sequenced cereal species means that this simply inherited trait which is physically problematic to score in the field is a good candidate for cloning (and hence, development of perfect markers). Similarly, marker-trait association for 'kernel: colour of aleurone layer' was 92 %. Comparative genomic analysis using previously identified flanking makers (11_21296 and 11_20453, orthologous to rice genes LOC_Os03g14040 and LOC_Os03g14690, respectively) suggests that the underlying barley *BLX* locus is located within a barley chromosomal region colinear to a ~365 kb region in rice containing a number of putative candidate genes. Third, comparative analysis using previously identified flanking makers (Cockram et al. 2010b) shows that the *SRH* locus underlying the Mendelianly inherited bi-modal trait 'grain rachilla hair type' is colinear with a ~290 Kb region of rice chromosome Os02, thus making *SRH* tractable to future map-based cloning and perfect marker development. Last, a predictive value of 88 % was returned for trait 'sterile spikelet: attitude' using marker Hv11_10933_GC (chromosome 1H, 55.49 cM), with this SNP likely to be in LD) natural variation at the *SLS* locus, known to control related lateral spikelet size and morphology (Franckowiak, 1995). The location of this genetic marker close to the highly non-recombining region associated with the centromere (data not shown) suggests that while fine mapping and cloning of the gene may be problematic due to unfavourable physical-to-genetic length ratios, reduction of recombination in the region means that development of additional markers in the region is likely to be highly diagnostic, due to strong linkage disequilibrium with the underlying locus. Therefore, genetic markers for this trait are likely to be deployable within breeding

programmes for the tracking of appropriate alleles during the process of varietal purification.

Interpretation of genetic markers with medium phenotypic prediction

Lower genotype–phenotype predictions are expected where traits are recorded on a quantitative scale, are controlled by a larger number of genetic loci, and for which heritability is lower. In such cases, it could be expected that fitting of single genetic markers is unlikely to result in high correlations between genotype and trait scores. One such example is 'grain: spiculation of inner lateral nerves', where the more quantitative nature of the trait scoring system (1–9 scale) indicates that it is controlled by a larger number of loci of smaller effect. Indeed, in a *H. vulgare* x *H. spontaneum* mapping population, grain spiculation has previously been reported to be controlled by five QTL (Ellis et al. 1999), supporting the assumption of complex inheritance in UK cultivars, and the low marker-trait observation observed in this study. Similarly, trait 'plant growth habit' is scored on a 1–9 scale, with a genetic marker originating from the *HvFT3* candidate gene for *PPD-H2* resulting in 88 % marker-trait correlation. The only cloned locus controlling plant growth habit in cereals is the rice Cys_2–His_2 zinc-finger gene *PROSTRATE GROWTH 1* (*PROG1*), located on the short arm of chromosome Os02 (Tan et al. 2008). However, established colinear relationships between rice and the Triticeae (Devos, 2005) predict the barley orthologue to map to chromosome 6H, suggesting the orthologue is not responsible for the locus detected in the vicinity of *HvFT3* on 1H. Interestingly, a number of studies using bi-parental barley mapping populations have reported QTL for growth habit and early growth vigour (Boyd et al. 2003; von Korff et al. 2008) as well as for flowering time (Laurie et al. 1995, Faure et al. 2007) in this chromosomal region, although it is not known whether these represent pleiotropic effects or are due to close genetic linkage. Finally, while a putative causative mutation within *HvbHLH1* preventing the ability of barley to synthesise anthocyanin has been previously identified (Cockram et al. 2010a), within the DUS system, anthocyanin production is measured on a 1–9 intensity scale in three different plant tissues. However, while *HvbHLH1* has previously been shown to be diagnostic for the presence of anthocyanin production *per se*, in some instances tissue specificity appears to be controlled by an additional genetic determinants (Cockram et al. 2010b). In addition, failure of the putative causative *HvbHLH1* InDel polymorphism to convert to the KASPar platform meant that a closely linked SNP within the gene was utilized instead. Together, these factors are likely to account for the observed predictive values of >100 % in the three anthocyanin-related DUS traits. We note that while DUS traits have been purposefully selected to be environmentally stable, there nevertheless remains a degree of environmental influence. This is probably most true of the anthocyanin colouration traits, in which anthocyanin intensity within target tissues is known to be affected by growth season. As a result, intensity for a given variety often varies by one (or more rarely two) point on the phenotypic scale, depending on year. Furthermore, a degree of subjectivity is involved in phenotypic assessment. In terms of DUS phenotypic assessment, such affects are minimised by averaging scores taken across two growth seasons, and rigorous testing protocols which are calibrated against reference varieties. However, even if the genetic basis of such a quantitative trait was fully understood, there will still be an environmental component to its phenotypic expression. Finally, the lack of association between molecular markers for known major flowering time loci *VRN-H1, VRN-H2, PPD-H1* and *PPD-H2* within winter and spring germplasm pools indicates that these loci do not influence flowering time within winter and spring varietal pools. At least in the case of *VRN-H1* and *PPD-H2* in winter/.spring germplasm, and *VR-H2* and *PPD-H1* in winter germplasm, this is likely due to almost perfect allelic partitioning at these loci between spring and winter varieties.

Deployment of genetic markers for DUS assessment and future prospects

Of the 28 barley currently utilized DUS characters, five are classified as 'grouping traits' within the European Union (UPOV characters 2, 13, 22, 26 and 29, http://www.cpvo.europa.eu/), and are used to group candidate submissions and reference varieties to facilitate the assessment of distinctness. Of these, perfect genetic markers are currently available for characters 13 ('ear: row number') and 29 ('seasonal growth habit'). Deployment of perfect markers for the latter is especially beneficial, due to the need for a dedicated field trial to determine spring or winter growth habit within which no other characters are scored. The remaining three grouping traits ('lower leaves: hairiness of leaf sheaths', 'grain: rachilla hair type' and 'grain: hairiness of ventral furrow') are two-state characters controlled by single Mendelian factors, both of which are currently the focus of fine-mapping projects using GWA approaches. Thus, we predict that perfect markers for all four grouping traits should be available in the near future, paving the way for the replacement of phenotypic evaluation of grouping traits with diagnostic molecular tests. There is much interest in approaches that could reduce cost by eliminating unnecessary comparisons between existing and candidate varieties prior to more formal testing. While morphological

characters are the foundation of DUS testing, reduction of the number of reference varieties included in the growing trials could be achieved by incorporating molecular markers into the DUS test. The use of such markers within the framework of evaluation has the potential to increase the efficiency and speed of phenotypic assessment, whilst maintaining the strength and scope of protection provided by the system. By comparing the profiles of candidate varieties with those of existing varieties maintained in a central database, it would be possible both to eliminate from further testing those varieties which do not require comparison in a growing trial and to select the varieties most similar to the candidate for close comparison in field tests. A variation on this approach has already been implemented by French testing authorities, resulting in the utilization of reduced numbers of reference varieties and savings in time and cost.

Finally, we found genotypic cluster analysis to differentiate winter varieties more readily than spring varieties, with the largest undifferentiated winter cluster consisting of six varieties. This was due at least in part to (1) multi-character scores for a single phenotypic trait provide greater potential for differentiation compared to the equivalent bi-allelic genetic marker, (2) a single biallelic genetic marker was used to predict phenotype in three different DUS traits. Despite the shortcomings listed above, this study demonstrates the potential that deployment of larger numbers of diagnostic genetic markers may have in varietal discrimination. A major barrier for the deployment of genetic markers for the prediction of DUS traits is that while the former are essentially binary characters, the latter are often scored using three or more character states. However, the advent of NGS and high-density SNP arrays provides marker densities in the orders of magnitude greater than those deployed to date are beginning to be developed and applied to crops, including barley. The increased genetic marker coverage, coupled with larger mapping population sizes or novel designs such as Multi-parent Advanced Generation Inter-Cross (MAGIC) populations, will provide higher powered experiments, allowing detailed genetic dissection of DUS traits with complex genetic architecture (including those scored on a more quantitative basis, or for which current marker-trait associations are low). It is envisaged that the resulting sub-sets of diagnostic genetic markers (causative polymorphisms or haplotypes) will complement the perfect markers identified to date, resulting in a comprehensive set of markers with which to accurately predict DUS phenotypes in barley and other crops covered by the UPOV convention.

The approach followed in this study is to introduce novel and potentially streamlined methods for the scoring of DUS traits, which are embedded in both international treaties and breeding practice. However, distinctness and uniformity per se could be determined without any precise knowledge of cause and effect between DNA and phenotypic variants. NGS techniques are fast approaching the point where a whole-genome (or at least a whole exome) genetic distance-based distinctness test could be performed at reasonable cost and within days of receipt of an incoming candidate variety. In this scenario, reference sequence from all common knowledge varieties could be compared instantly, arguably improving the effectiveness of the distinctness test. The assessment of uniformity would require extending sequencing to a depth that permitted low frequencies of off-type alleles to be detected. However, the ability to describe a variety based on its DNA sequence poses additional problems, such as error rates in SNP calling (Nielsen et al. 2011) and setting statistically and biologically acceptable definitions of intra- and inter-varietal genome variability. Ultimately, for NGS-based assessment of DUS to be implemented, sequencing needs to become both faster and cheaper, the production of standard reports from the analysis of raw sequence needs to become highly automated, and lastly and perhaps most importantly, the replacement of traditional characters with sequence-based assessments would need to be understood and broadly accepted on its merits by the regulatory, breeding, seeds and producer communities. Even if all these dependencies were satisfied, a professionally conducted botanical description of each variety would still be needed at farmer level, so it is our opinion that genotype and phenotype-based assessments will most likely continue to co-exist for the foreseeable future.

Acknowledgments This work was funded by Defra grant 2009H. We thank the James Hutton Institute for provision of the European barley germplasm.

References

Arens P, Mansilla C, Deinum D, Cavellini L, Moretti A, Rolland S, van der Schoot H, Calvache D, Ponz F, Collonnier C, Mathis R, Smilde D, Caranta C, Vosman B (2010) Development and evaluation of robust molecular markers linked to disease resistance in tomato for distinctness, uniformity and stability testing. Theor Appl Genet 120:655–664. doi:10.1007/s00122-099-1183-2

Boyd WJR, Li CD, Grime CE, Cakir CR, Potipibol S, Kaveeta L, Men S, Jalal Kamali MR, Barr AR, Moody DB, Lance RCM, Logue SJ, Raman H, Read BJ (2003) Conventional and molecular genetic analyses of factors contributing to the variation in the timing of heading among spring barley (*H. vulgare* L.) genotypes grown over a mild winter growing season. Aust J Agric Res 54:1277–1301. doi:10.1071/ar03014

Chono M, Honda I, Zeniya H, Yoneyama K, Saisho D, Takeda K, Takatsuto S, Hoshino T, Wantanabe Y (2003) A semidwarf phenotype of barley *uzu* results from a nucleotide substitution in the gene encoding a putative brassinosteroid receptor. Plant Physiol 133:1209–1219. doi:10.1104/pp103.026195

Close TJ, Bhat PR, Lonardi S, Wu Y, Rostoks N, Ramsay L, Druka A, Stein N, Svensson JT, Wanamaker S et al (2009) Development and implementation of high-throughput SNP genotyping in barley. BMC Genomics 10:e582. doi:10.1186/1471-2164-10-582

Cockram J, Jones H, Leigh FJ, O'Sullivan D, Powell W, Laurie DA, Greenland AJ (2007a) Control of flowering time in temperate cereals: genes, domestication and sustainable productivity. J Exp Bot 58:1231–1244. doi:10.1093/jxb/erm042

Cockram J, Chiapparino E, Taylor SA, Stamati K, Donini P, Laurie DA, O'Sullivan DM (2007b) Haplotype analysis of vernalization loci in European barley germplasm reveals novel *VRN-H1* alleles and a predominant winter *VRN-H1*/*VRN-H2* multi-locus haplotype. Theor Appl Genet 115:993–1001. doi:10.1007/s00122-007-0626-x

Cockram J, Mackay IJ, O'Sullivan DM (2007c) The role of double-stranded break repair in the creation of phenotypic diversity at cereal *VRN1* loci. Genetics 177:1–5. doi:10.1534/genetics.107.074765

Cockram J, White J, Leigh FJ, Lea VJ, Chiapparino E, Laurie DA, Mackay IJ, Powell W, O'Sullivan DM (2008) Association mapping of partitioning loci in barley (*Hordeum vulg*are ssp. *vulgare* L.). BMC Genet 9:e16. doi:10.1186/1471-2156-9-16

Cockram J, Norris C, O'Sullivan DM (2009) PCR markers diagnostic for seasonal growth habit in barley. Crop Sci 49:403–410. doi:10.2135/cropsci2008.07.0398

Cockram J, Howells RM, O'Sullivan DM (2010a) Segmental chromosomal duplications harbouring group IV *CONSTANS*-like genes in cereals. Genome 53:231–240. doi:10.1139/g09-101

Cockram J, White J, Zuluaga DL, Smith D, Comadran J et al (2010b) Genome-wide association mapping to candidate polymorphism resolution in the un-sequenced barley genome. Proc Natl Acad Sci USA 107:21611–21616. doi:10.1073/pnas.1010179107

Devos KM (2005) Updating the 'crop circle'. Curr Opin Plant Biol 8:155–162. doi:10.1016/j.pbi.2005.01.005

Dubcovsky J, Chen C, Yan L (2005) Molecular characterization of allelic variation at the *VRN-H2* vernalization locus in barley. Mol Breed 15:395–407. doi:10.1007/s11032-005-0084-6

Ellis RP, J Russell, L Ramsay, R Waugh (1999) Barley domestication–Hordeum spontaneum, a source of new genes for crop improvement. Scottish Crop Research Institute report. http://www.scri.ac.uk/scri/file/individualreports/1999/14BARLEY.PDF. Accessed 25 Jan 2012

Faure S, Higgins J, Turner A, Laurie DA (2007) The FLOWERING LOCUS-T-like family in barley (*Hordeum vulgare*). Genetics 176:599–609. doi:10.1534/genetics.106.069500

Franckowiak JD (1995) Notes on linkage drag in Bowman backcross derived lines of spring barley. Barley Genet Newsl 24:63–70

Gunjaca J, Buhinicek I, Jukic M, Sarcevic H, Vragolovic A, Kozic Z, Jambrovic A, Pejic I (2008) Discriminating maize inbred lines using molecular and DUS data. Euphytica 161:165–172. doi:10.1007/s10681-007-9518-z

Ibáñez J, Vargas AM, Palancar M, Borrego J, de Andrés MT (2009a) Genetic relationships among table-grape varieties. Am J Enol Vitic 60:35–42

Ibáñez J, Vélez MD, de Andrés MT, Borrego J (2009b) Molecular markers for establishing distinctness in vegetatively propagated crops: a case study in grapevine. Theor Appl Genet 119:1213–1222. doi:10.1007/s00122-009-1122-2

Islam AKMR (1983) Ditelosomic additions of barley chromosomes to wheat. In: Sakamoto S (ed) Proceedings of the 6th international wheat symposium. Maruzen Co. Ltd, Koyota, pp 223–238

Jia Q, Zhang J, Wescott S, Zhang XQ, Bellgard M, Lance R, Li C (2009) *GA-20 oxidase* as a candidate for the semidwarf gene *sdw1/denso* in barley. Funct Integr Genomics 9:255–262. doi:10.1007/s10142-009-0120-4

Karsai I, Szűcs P, Mészáros K, Filichkina T, Hayes PM, Skinner JS, Láng L, Bedö Z (2005) The *Vrn-H2* locus is a major determinant of flowering time in a facultative x winter growth habit barley (*Hordeum vulgare* L.) mapping population. Theor Appl Genet 110:1458–1466. doi:10.1007/s00122-005-1979-7

Komatsuda T, Mano Y (2002) Molecular mapping of the *intermedium spike-c* (*int-c*) and *non-brittle rachis 1* (*btr1*) loci in barley (*Hordeum vulgare* L.). Theor Appl Genet 105:85–90. doi:10.1007/s00122-001-0858-0

Komatsuda T, Tanno K (2004) Comparative high resolution map of the *six-rowed spike locus 1* (*vrs1*) in several populations of barley, *Hordeum vulgare* L. Hereditas 141:68–73. doi:0.1111/j.1601-5223.2004.01820.x

Komatsuda T, Pourkheirandish M, He C, Azhaguvel P, Kanamori H, Perovic D, Stein N, Graner A, Wicker T, Tagiri A, Lundqvist U, Fujimura T, Matsuoka M, Matsumoto T, Yano M (2007) Six-rowed barley originated from a mutation in a homeodomain-leucine zipper I-class homeobox gene. Proc Natl Acad Sci USA 104:1424–1429. doi:10.1073/pnas.0608580104

Laurie DA, Pratchett N, Bezant JH, Snape JW (1995) RFLP mapping of five major genes and eight quantitative trait loci controlling flowering time in a winter x spring barley (*Hordeum vulgare* L.) cross. Genome 38:575–585. doi:10.1139/g95-074

Liu K, Muse SV (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. Bioinformatics 21:2128–2129. doi:10.1093/bioinformatics/bti282

Lundqvist U, Franckowiak J, Konishi T (1996) New and revised descriptions of barley genes. Barley Genet Newsl 26:22–43

Lundqvist U, Franckowiak JD, Konoshi T (1997) New and revised descriptions of barley genes. Barley Genet Newsl 26:22–516

Matzumoto T, Tanaka T, Sakai H, Amano N, Kanamori H, Kurita A, Kamiya K, Yamamoto M, Ikawa H, Fujii N, Hiri K, Itoh T, Sato K (2011) Comprehensive sequence analysis of 24,783 barley full-length cDNAs derived from 12 clone libraries. Plant Physiol 156:20–28. doi:10.1104/pp110.171579

Maughan PJ, Yourstone SM, Byers RL, Smith SM, Udall JA (2010) Single-nucleotide polymorphism genotyping in mapping populations via genomic reduction and next-generation sequencing: proof of concept. Plant Genome 3:166–178. doi:10.3835/plantgenome2010.07.0016

Mayer KFX, Martis M, Hedley PE, Šimokova H, Liu H, Morris JA, Steuernagel B, Taudien S et al (2011) Unlocking the barley genome by chromosomal and comparative genomics. Plant Cell 23:1249–1263. doi:10.1105/tpc.110.082537

Nair SK, Wang N, Turuspekov Y, Pourkheirandish M, Sinsuwongwat S et al (2010) Cleistogamous flowering in barley arises from the suppression of microRNA-guided *HvAP2* mRNA cleavage. Proc Natl Acad Sci USA 107:490–495. doi:10.1073/pnas.0909097107

Nielsen R, Paul JS, Albrechtsen A, Song YS (2011) Genotype and SNP calling from next-generation sequencing data. Nat Rev Genet 12:443–451. doi:10.1038/nrg2986

Noli E, Teriaca MS, Sanguineti MC, Conti S (2008) Utilization of SSR and AFLP markers for the assessment of distinctness in durum wheat. Mol Breed 22:301–313. doi:10.1007/s11032-008-9176-4

Ramsay L, Comadran J, Druka A, Marshall DF, Thomas WT, Macaulay M, MacKenzie K, Simpson C, Fuller J, Bonar N, Hayes PM, Lundqvist U, Franckowiak JD, Close TJ, Muehlbauer GJ, Waugh R (2011) *INTERMEDIUM-C*, a modifier of lateral spikelet fertility in barley, is an ortholog of the maize domestication gene *TEOSINTE BRANCHED 1*. Nat Genet 43:169–172. doi:10.1038/ng.745

Tan L, Li X, Liu F, Sun X, Li C, Zhu Z, Fu Y, Cai H, Wang W, Xie D, Sun C (2008) Control of a key transition from prostrate to erect growth in rice domestication. Nat Genet 40:1360–1364. doi:10.1038/ng.197

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. doi:10.1126/science.1117619

Von Korff M, Grando S, Del Greco A, This D, Baum M, Ceccarelli S (2008) Quantitative trait loci associated with adaptation in Mediterranean dryland conditions in barley. Theor Appl Genet 117:653–669. doi:10.1007/s00122-008-0787-2

von Zitzewitz J, Szűcs P, Dubcovsky J, Yan L, Francia E, Pecchioni N, Casas A, Chen THH, Hayes P, Skinner J (2005) Molecular and structural characterization of barley vernalization genes. Plant Mol Biol 59:449–467. doi:10.1007/s11103-005-0351-2

Waugh R, Marshall D, Thomas B, Comadran J, Russell J, Close T, Stein N, Hayes P, Muehlbauer G, Cockram J, O'Sullivan D, Mackay I, Flavell A, Agoueb A, Barleycap RL (2010) Whole-genome association mapping in elite inbred crop varieties. Genome 53:967–972. doi:10.1139/G10-078

Yu Y, Tomkins JP, Waugh R, Frisch DA, Kudrna D, Kleinhofs A, Bruggeman RS, Meuhlbauer GJ, Wise RP, Wing RA (2000) A bacterial artificial chromosome library for barley (*Hordeum vulgare* L.) and the identification of clones containing putative resistance genes. Theor Appl Genet 101:1093–1099. doi:10.1007/s001220051584