"Le Sélectionneur Français" 2011 (62), **31-44**

GENETIQUE D'ASSOCIATION CHEZ UNE CEREALE : EXEMPLE DE L'ORGE (ASSOCIATION GENETICS IN BARLEY)

WTB Thomas[1], L Ramsay[1], J Comadran[1], P Shaw[1] D Marshall[1], A Newton[1], D O'Sullivan[2], J Cockram[2], I Mackay[2], R Bayles[2], J White[2], M Kearsey[3], Z Luo[3], M Wang[3], C Tapsell[4], D Harrap[4], P Werner[4], S Klose[5], P Bury[6], J Wroth[6], O Argillier[6], R Habgood[7], M Glew[7], A-M Bochard[7],
P Gymer[8], D Vequaud[8], T Christerson[9], B Allvin[10], N Davies[11], R Broadbent[11], J Brosnan[12], T Bringhurst[12], C Booer[13] & R Waugh[1]

[1]SCRI, Invergowrie, Dundee DD2 5DA, UK; School of Biosciences
[2]NIAB, Huntingdon Road, Cambridge CB3 0LE, UK
[3]University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK
[4]KWS UK, 56 Church Street, Thriplow, Nr Royston, Herts SG8 7RE, UK
[5]LS Plant Breeding, North Barn, Manor Farm, Milton Road, Cambridge CB24 9NG, UK
[6]Syngenta Seeds Ltd, Market Stainton, Market Rasen, Lincs LN8 5LJ, UK
[7]Nickerson-Advanta Ltd, Joseph Nickerson Research Centre, Rothwell, Market Rasen, Lincs LN7 6DT, UK
[8]Secobra/Dalgety, Throws Farm, Stebbing, Great Dunmow, Essex CM6 3AQ, UK
[9]SW Seed, Svalöf Weibull AB, SE-268 81 Svalöv, Sweden
[10]Perten Instruments AB, P.O. Box 5101, SE-141 05 Huddinge, Sweden
[11]The Maltsters Association of Great Britain, 31B Castlegate, Newark, Notts NG24 1AZ, UK
[12]The Scotch Whisky Research Institute, The Robertson Trust Building, Research Park North, Riccarton, Edinburgh EH14 4AP, UK
[13]Campden BRI, Lyttel Hall, Nutfield, Surrey RH1 4HY, UK

ABSTRACT

The development of a high throughput SNP genotyping platform enabled the genetic fingerprinting of elite UK barley varieties to a sufficient density of coverage to allow the possibility of association analyses. Using morphological and agronomic data collected during official UK trials, association genetic analysis has allowed the dissection of the genetic control of Mendelian traits resulting in the cloning of DUS character. This methodology is now being applied to more quantitative agronomic traits and the results already obtained means that the genotyping technology has now been taken up and applied by breeding companies.

(**Key Words:** Genome wide association scan, DUS, anthocyanin,)

1 –INTRODUCTION

Plant breeding is a lengthy and expensive Research and Development exercise with selection for key performance characters such as yield and quality generally occurring towards the end of the selection process. The use of markers for the early identification of genetic regions associated with improvements in performance (Marker Assisted Selection) offers the promise to improve the efficiency of plant breeding but, despite years of research, has yet to be realised on a large scale for small-grained cereals. Plant genetic studies have generally concentrated on controlled crosses often derived from inbred stocks due to their practical ease and genetic and statistical power. Often, crosses have been chosen that maximise parental differences in order to facilitate whole genome map construction and the detection of genetic regions affecting characters such as yield and agronomic performance. Whilst this produces results, the applicability of the findings in general plant breeding is limited because the crosses do not match the concentration on crossing within the elite gene-pool that is generally adopted to produce a high mid-parental value and thus a high probability of producing

improved recombinant inbred lines from a cross. In contrast, mammalian genetic studies cannot utilise such special populations and have developed alternative approaches to phenotype and genotype collections of individuals in an approach termed association mapping. Here, the challenge is to generate detailed genetic fingerprints for each of the lines that has been phenotyped and then detect individual genetic markers that are in Linkage Disequilibrium (LD) with the phenotype (ROSTOKS *et al*, 2006).

It follows that the power and discrimination of association studies is largely determined by the underlying patterns of linkage disequilibrium (LD) in the chosen population, which can be affected by one or more of the following factors; a.) past history of the population, b.) domestication bottle-necks, c.) selection, d.) population admixture, and e.) population structure. In addition, detection will also depend upon aspects of the phenotype, such as heritability, penetrance, size of the effect etc. If, however, LD can be used in association genetics of pools of elite germplasm that barley breeders are working with then the results will be immediately applicable and facilitate more effective selection with downstream benefits to growers and processor as well as end-users.

This paper describes the initial findings of a project in the UK entitled 'Association Genetics of UK Elite barleys' (AGOUEB) that has attempted to address some of these issues. This was made possible by advances in molecular fingerprinting of barley that in turn have relied on the past development of genomic resources in the crop. These advances enabled a dense enough marker coverage to make it practicable to attempt association genetics analysis. As the seven barley chromosomes have a total genetic map length of around 1200 centimorgans (cM) and LD persists generally 1-2 cM in elite material the project needed around 1200 spaced SNPs in order to have a chance of detecting significant association of a marker with a performance trait. In addition increasing the marker density would not only increase the reliability of detection but also resolve the chromosomal region controlling a performance trait to a much narrower region and thus provide tools to break down adverse tight linkages.

2 - BACKGROUND

2.1 - Germplasm

The aim of the project was to detect genetic regions associated with improvement in the agronomic potential of barley varieties available to the UK farmer. Barley breeders make many crosses each year, each of which is designed to produce progeny that will meet the needs of the farmer and the end-user. The progeny from these crosses are selected in trials and nurseries over seasons and sites until the most promising have been identified for entry into official trials to determine whether or not any should be granted Plant Breeders Rights (PBR). PBR is granted for a variety after it has successfully completed the two years of National List (NL) trials, when it is tested for Distinctness, Uniformity and Stability (DUS) and Value for Cultivation and Use (VCU) and it is placed on the National List, enabling seed of the line to be traded. The best varieties emerging each year from the second year of National List trials are selected as candidates for Recommended List (RL) trials; securing recommended list status is generally a major pre-requisite for a commercially successful variety.

We chose to concentrate our survey on the period from 1993, when the funding of RL trials transferred from the public to the private purse with the use of levy payers' money to continue the funding of recommended list cereal trials. The funding of the National List (NL) trials was also gradually transferred to the private pocket with commercial breeders effectively funding the trials through submission fees. Between 1993 and 2005, barley breeders submitted an annual average of 38 and 40 lines of spring and winter barley lines for NL trials; the majority originated from UK based breeding programmes but submissions regularly included lines originating from Danish, French and German programmes. We therefore had 251 and 328 spring and winter barleys that had least completed 2 years of NL trials from the lines submitted between 1993 and 2005. A further 18 spring

and 28 winter barleys were already in or were selected for recommended list trials in 1993 and 1994 making a total of 625 different barley lines that were available during the survey period. We supplemented this list with some varieties that had been commercially successful during the 1980's and some other key progenitor lines to bring the total to 663 lines. We did, however, find several problems in sourcing seed of all the lines, especially those originating from the continental breeding programmes, so that we were eventually able to utilise 543 lines for genotyping in what we term the AGOUEB Public Set. Each of the breeders participating in the project submitted an additional 60 lines for genotyping and these were augmented with some key progenitor lines to provide 950 lines for analysis. Whilst we analysed all lines together, the information gathered upon the lines submitted by each breeder remained 'Private' to that breeder and is not publicly available. Thus we had seven separate databases; the Public Set, all of which will become publicly available and 6 individual breeder data bases of the Public Set augmented by their own 'Private' set.

Of the lines that completed NL from those that entered during our sample period, an average of 6 spring and 6 winters were entered into RL trials each year, resulting in totals of 31 and 40 new recommendations for spring and winter barley respectively from the set. This included varieties such as Optic, Pearl and NFC Tipple that have all had a significant impact upon the malting market over the past 15 years.

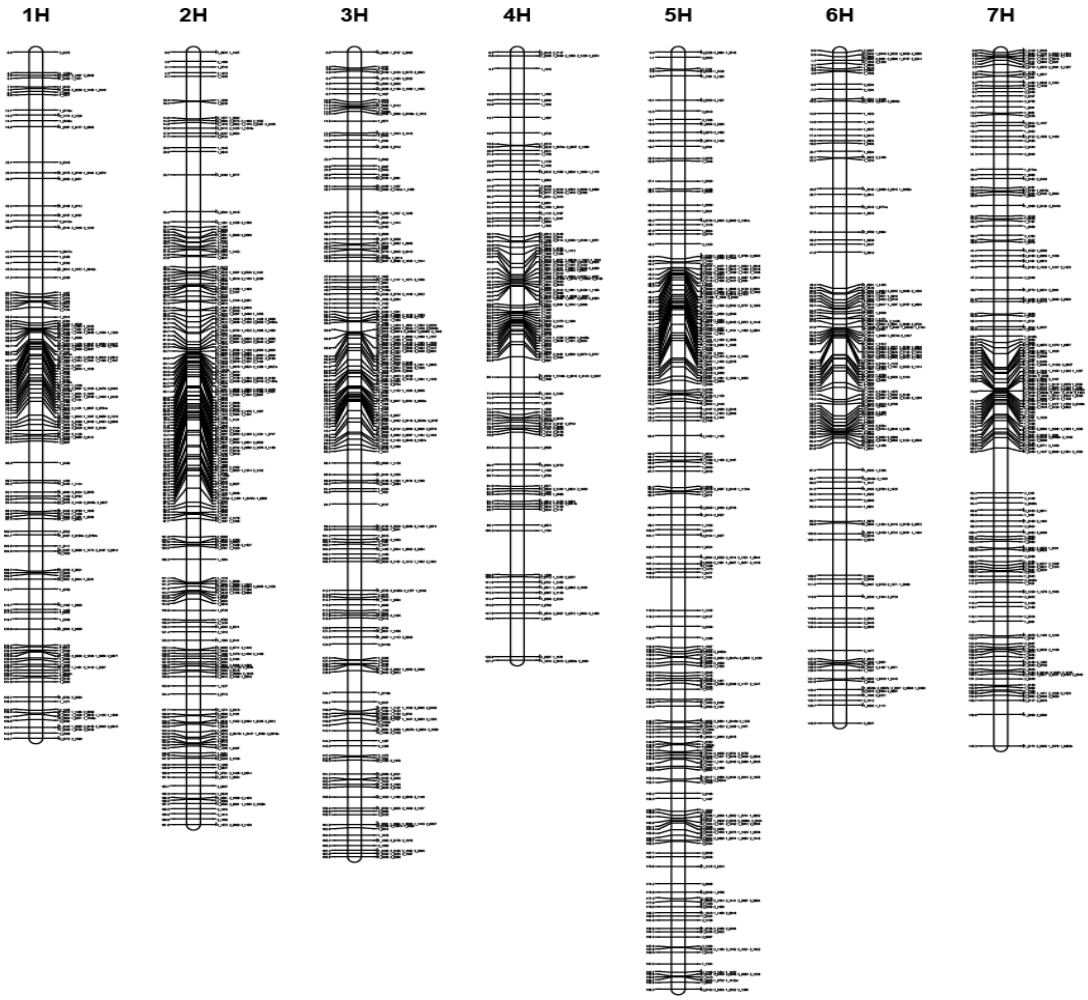

Figure 1 : Integrated SNP-based gene map of 2943 BOPA1 and BOPA2 SNP Markers.

2.2 - Genotyping

The genotyping for the AGOUEB was based upon the use of single nucleotide polymorphisms (SNPs) that were discovered through the comparison of sequence data derived from the genic regions of the barley genome. Such comparisons highlighted the presence of single nucleotide variants (SNPs) and therefore formed the basis of the genotyping platform used. The development of this high throughput genotyping platform was pivotal to the success of the project and conversely the project played an important role in the development and utilisation of the platform. This development was itself dependent on previous development of barley genomic resources and also the close collaboration within the world barley genetics community.

Two 1536-SNP production-scale OPAs, referred to as BOPA1 and BOPA2, were developed from SNPs tested on three 1536 SNP pilot assays. All sequences used as SNP sources were generated using the Sanger dideoxy chain termination method. Full details have recently been published by CLOSE *et al*., (2009). The pilot OPAs were used to genotype three barley mapping populations (Steptoe x Morex, Morex x Barke and the Orgeon Wolfe Barleys OWB). The three resulting maps together with a fourth map produced from a population genotyped with BOPA1 were fused using MergeMap to form a consensus map containing 2943 SNP loci with a total map length of 1099 cM (Figure 1). The identity and polarity of linkage groups were determined by integrating 110 previously mapped bin markers into the SxM and consensus maps. In all maps, chromosome 5H has the greatest length, a mean of 198 cM, consistent with previously published linkage maps. Chromosome 5H is also the most populated with 535 SNP loci and is subdivided into the largest number of marker bins (180). On the lower end of the spectrum chromosome 4H has only 338 SNP loci distributed among 113 marker bins covering 125 cM. The relationship of nearly one marker bin per cM holds for all seven linkage groups.

The barley source sequence underlying each OPA SNP was compared to the rice (*Oryza sativa*) version 5 and version 6 gene models using BLASTX, and the top hit was taken as the most similar rice gene. These rice best hit coordinates were used as the basis of alignments of each of the seven barley chromosomes with the twelve rice chromosomes.. The simplest relationships are essentially total synteny between barley 3H versus rice 1 (3HS = 1S, 3HL = 1L) and barley 6H versus rice 2 (6HS = 2S, 6HL = 2L). Four remaining barley chromosomes each are composed of ancestors of two rice chromosomes, in each case having one ancestral chromosome nested within the pericentric region, flanked by segments of the other syntenic chromosome. Chromosome 5H has a slightly more complex pattern of synteny having major regions of co-linearity with three rice chromosomes (Figure 2).

The two production OPAs, BOPA1 and BOPA2, had somewhat different design elements. These differences have been reflected in the performance of BOPA1 and BOPA2 for the genotyping of breeding germplasm in the USA within the BarleyCAP project (www.barleycap.org). A comparison of BOPA1 and BOPA2 in relation to both SNP representation and the performance within a subset of the BarleyCAP project indicated that BOPA1, which was designed using only SNPs with a minor allele frequency (MAF) of at least 0.08 in the design germplasm, yielded MAF values less than 0.05 for only 164 SNPs (10.7%). In contrast BOPA2, which targeted 615 SNPs with MAF less than 0.08 in the design germplasm, yielded MAF values less than 0.05 for 585 SNPs (38.1%). This included about three times as many SNPs with MAF = 0 (301 versus 99) and 4.4 times as many SNPs (284/65) with MAF between 0 and 0.05. Thus, BOPA2 has greater sensitivity to detect rare alleles than BOPA1, some of which may be important for the development of new varieties containing uncommon alleles of certain genes. But, this increased sensitivity is counterbalanced by a compromise in the reduced frequency of informative SNPs in general.

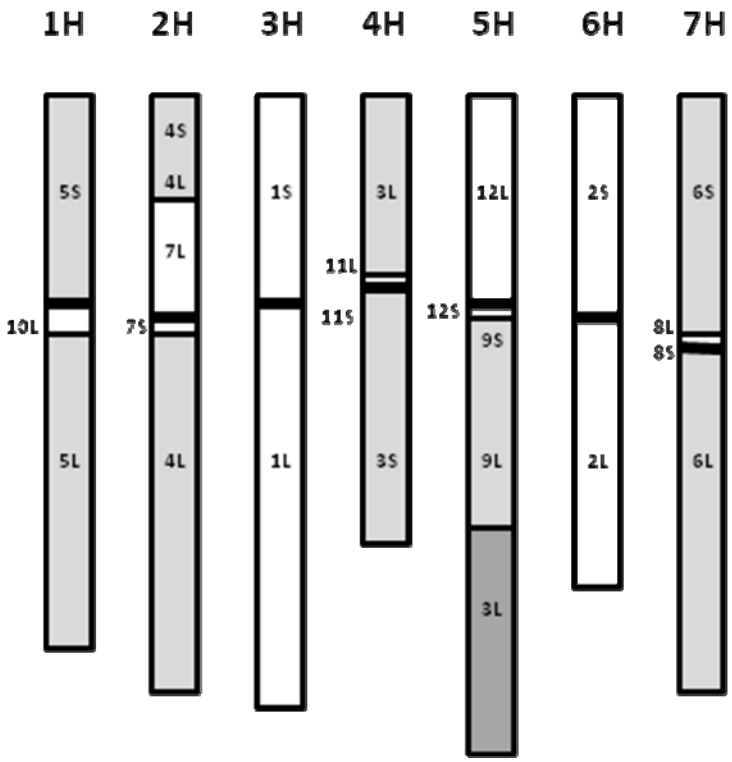

Figure 2 : Diagram showing the syntenic relationship of the genetic map of barley (1H to 7H) to the twelve chromosomes of rice after CLOSE *et al*., (2007).

The choice of SNPs for BOPA2 allowed the inclusion of polymorphisms that had previously been reported with genes known or suspected to be involved in important agronomic and economic traits. Some of the major genes included those involved in the control of flowering time and vernalisation that differentiate winter and spring sown crops. Thus SNPs within PPD-H1, VRN-H1, VRN-H2 and VRN-H3 (TURNER *et al.*, 2005, SZUCS *et al*., 2007, YAN *et al*., 2006, VON ZITZEWITZ *et al*., 2005) were included on the BOPA2 chip thus allowing additional functional information to be inferred from the genotyping results. For PPD-H1 the SNPs included from the Pseudo-Response Regulator gene include one of the single base variants that cause coding differences in the gene that have been shown to be completely associated with the allelic difference between response to long day length that underpins the winter/spring sown adaptation (TURNER *et al*., 2005). However for other genes the functional information is not necessarily straightforward as the diagnostic nature of the SNPs is determined by their relationship to the causal differences between the 'winter' and 'spring' alleles at these genes. For VRN-H2 the allelic differences are due to the presence and absence of ZCCT genes that are reflected in the presence or absence of the SNP rather than an alternative nucleotide, whereas for VRN-H1 the BOPA2 SNPs are derived from sequence information within the first intron of the BM5A MADS-Box transcription factor that are in tight linkage disequilibrium with the deletion that is believed to be causal to the functional difference (SZUCS *et al*., 2007)) It is worth noting that further investigation within relevant germplasm has generally shown that the relationship of diagnostic markers to the traits is complicated by multiple causal variants and different breeding histories (COCKRAM *et al*., 2008). Such complications can be an issue with the use of the SNP genotyping platform for diagnostics but are due to a biological rather than a technological issue. .

Another example of the diagnostic markers included on BOPA2 are the SNPs within *HvHox1* DNA binding protein that has been shown to be *VRS1* (KOMATSUDA *et al*., 2007), the major determinant of the two-rowed/sixed-rowed ear in barley. The five SNPs are not diagnostic in themselves but, in combination, they do allow the determination of the haplotype present at the gene.

Again the diagnostic is complicated as the six-rowed state has arisen at least on three separate occasions each with a different mutational event. These can be followed in the relevant germplasm, though there are complications in distinguishing some six-rowed variants at VRS1 from the *deficiens* allele (extreme two-rowed form) that share a common underlying haplotype despite having very contrasting phenotypes (Saisho et al 2009). Other known genes on BOPA2 include several disease resistance genes including Rpg1, Rph7, Mla and mlo, genes potentially involved in malting quality such as Aglu2 (α-glucosidase), Bmy1 (β-amylase), Glb (1->3)-B-glucan 3-glucanohydrolase) and genes potentially involved in winter-hardiness and abiotic stress resistance (FRANCIA *et al*., 2007). For disease resistance the SNP genotyping data set has been supplemented by additional genotyping published diagnostic markers for mlo11 (PIFFANELLI *et al*., 2004) and Rrs2 (HANEMANN *et al*., 2009).

It is worth stressing the diagnostic potential of the SNP genotyping platform as ongoing work relates trait variation to particular genic SNPs. The information on rice/barley synteny and known candidate genes is allowing connections to be made within the project of relevant polymorphism in UK germplasm to functional variation in genes that potentially underpin the control of the trait.

3 - RESULTS – DATA COLLECTION

3.1 - Databasing & visualisation

A custom database schema was developed using the Germinate 2 platform in order to store and query genotypic and phenotypic data obtained throughout the duration of the AGOUEB project. The size and complexity of these datasets has meant that particular considerations have had to be made in relation to performance of the system. In addition we have spent considerable time ensuring necessary links and integration with external analysis programs such as Flapjack were maintained and developed. The Germinate 2 AGOUEB database (http:/bioinf.scri.ac.uk/agoueb/) was also heavily customised in order to provide additional tools required for the storage of SNP genotype, categorical and field trial data which would be available as part of this project.

We obtained classificatory data such as names, breeders codes, originating breeder, agent, pedigree and, where applicable AFP number, for 829 elite barley lines, comprising the AGOUEB Public Set, key progenitors and disease differentials. The information was gathered from a variety of sources with names, breeders codes, breeder and agent being derived from information published in the Plant Varieties and Seeds Gazette (http://www.fera.defra.gov.uk/plants/publications/gazette.cfm). Pedigrees were assembled from information that was published by the National Institute of Agricultural Botany (NIAB) in the series 'Varieties in Trial', from information associated with Recommended Lists and by contacting individual breeders. In addition, images of varieties grown in the AGOUEB trial series were stored and linked with the appropriate lines. We also provided a facility for users to add notes and comments about individual lines. One problem that we encountered was the re-cycling of names. For instance, we had two varieties named Aquarelle and two named Saffron within the AGOUEB Public set with the earlier submissions not being commercialised and the later ones proceeding to Recommended List status. We were able to solve this issue by including the AFP number with the name (or breeders code if un-named) in a two-part identifier. The breeders private lines were stored with codes that preserved the anonymity of the lines but clearly separated the individual breeders. No classificatory data was available for the breeders lines, although each breeder can subsequently add in information for their own lines.

Data from the Illumina GoldenGate assay for the 3072 SNPs in BOPA1 and BOPA2 is first analysed in Bead Studio and markers with good quality scores are then exported in a format for input into the AGOUEB database, where they are linked with classificatory data upon the markers themselves, e.g. map position, primers, associated barley contig and putative gene annotation. This information can then be exported in suitable formats for other analysis and visualisation software, such

as Genstat and FlapJack. For the AGOUEB public set and progenitors, we had 2,239,488 genotypic data points distributed across 729 lines.

The 33 DUS characters with sufficient data fill for 579 lines in the AGOUEB Public set were obtained from NIAB and stored within the database for output on the summary page for each variety. In addition, we obtained information from the National List and Recommended List fungicide treated and untreated Trials databases from the British Society of Plant Breeders (BSPB) members and Crop Evaluation Limited respectively via NIAB, augmented it with information gathered from the AGOUEB fungicide treated trials and disease nurseries to store as a set of raw data in the AGOUEB database. In summary, we stored data for 76 distinct phenotypes measured with varying frequency upon trials grown between 1988 and 2008, which resulted in 622,522 distinct phenotypic datapoints. The overall genotypic, site and year means derived from the analysis of the complete data set together with their standard errors and the components of variation for each character were stored in the database together with the analysis model utilised. Similarly, the kinship matrix and genome wide association scans for the DUS characters and VCU traits were stored in the database together with the model used in the analysis.

3.2 - Phenotyping: DUS Data description

Distinctness, uniformity and stability (DUS) phenotypic data for the public set of barley accessions were sourced from archives maintained at NIAB. For each submission, DUS assessment is carried out over the two years of National List Trials. The portfolio of DUS characters assessed has changed over the period covered by the accessions under study, as governed by changes in CPVO guidelines (http://www.cpvo.europa.eu/). A total of 94 DUS characters have been recorded over this time, 23 of which represent 2-state characters, 22 are 3-state characters, and 49 are recorded as ≥4 state characters. Currently, 28 phenotypes are scored during DUS assessment (http://www.cpvo.europa.eu/). Of the 609 public barley accessions investigated (as of 09/05/2008), 579 were found to possess associated historical DUS phenotypic data. However, due to the changes in the DUS phenotypes utilised over the years, the percentage of barley lies with available phenotypic data varied considerably. For effective association mapping, low population sizes will results in a considerable loss of power. Accordingly, we selected those traits which had a fill of over 33 % (≥ 200 barley lines) for subsequent analysis. The data for these 33 traits were subsequently investigated for errors and outliers. The primary obstacle encountered were changes in the scoring systems used, as determined by evolution of CPVO protocols over the time period studied. For example, up until 1980 the trait “awn length cf. glume” (character 22) was scored as 3, 5 and 7 (shorter, equal and longer, respectively). From 1981 onwards, this was changed to a scoring system of 1, 2 and 3 (shorter, equal and longer, respectively). Another category of change to the scoring system is exemplified by the trait “Sterile spikelet development”, where the system up until 2002 scored on a continuous scale from 1 (none/deficiens) to 9 (v long), after which it was scored as a binary trait: 1 (none/deficiens) or 9 (rudimentary – v long). Finally, cross-referencing of related traits (eg “auricle anthocyanin coloration” and “auricle anthocyanin intensity”) allowed identification and correction of additional errors. Where outliers in the database were identified, these were resolved by reference to the original paper records where available. In instances where this was not possible, outliers were replaced by missing values. The final DUS phenotypic dataset (v1.0), along with the codes for all character states, were databased using Germinate v2, and is currently accessible to consortium partners via the SCRI website (http://www.scri.ac.uk/).

3.3 - Phenotyping: Historical VCU Data

Each year, breeders submit their most promising selections for first year National List trials (NL1). The poorer-performing lines from NL1 trials are discarded or withdrawn and the remainder progress to second year National List trials (NL2) at the end of which, a variety is placed on the National List if it passes Distinctness, Uniformity and Stability (DUS) and Value for Cultivation and Use (VCU) tests. Up to 2002, the British Society of Plant Breeders (BSPB) also ran a parallel series of trials to NL1 and NL2 at selected member sites, where NL1 and NL2 entries were generally grown together. Between 1993 and 2002, several of the BSPB sites were officially licensed NL sites and the NL1 and NL2 entries were grown in separate trials at these sites so we have assigned them to the NL1 and NL2 trial series, rather than the BSPB. At the end of NL2, the agronomic data from BSPB, NL1 and NL2 trials is then combined to identify the best performing lines for entry into Recommended List trials (RL). The remaining lines from NL2 are eventually withdrawn from the National List, unless they show promise as a variety in another country. At the end of the first year of RL, candidates that show some advantage over existing varieties on the RL will be give a provisional recommendation (P1). If the provisional recommendation continues to show merit after a second year of RL it continues in trial with a provisional recommendation (P2) and will gain full recommendation after a third year of trials if it continues to show merit that is also reflected in increasing seed sales.

During the project, we supplemented the above data set with data obtained from growing a sample of 64 spring and 60 winter varieties at eight and seven sites respectively for each of the three harvest years from 2006 to 2008 inclusive, which we have called the AGOUEB series. For each crop type, the lines grown represented market successes and failure over the sample period as well as some lines that failed to gain recommendation and some progenitors. The trials were grown according to the current VCU protocol for fungicide treated trials and scored for the same phenotypes as the official trials plus some additional phenotypes that related to grain quality. This trial not only enabled us to make un-ambiguous estimates of breeding progress for the phenotypes that we measured but could also be added into to the official data set as an extra trial series to provide a means of stabilizing the predictions of genetic performance of lines that were grown over a relatively short time-scale.

In our project, we only considered lines that had advanced to the end of NL2 and so each line would have been trialled over a minimum of two years and a number of phenotypic scores collected on it. We considered 66 of the phenotypic scores present in the 1988-2006 database to be of interest for analysis but, apart from yield, the amount of data that had been collected for these variates was limited. For instance, the winter barley data set contains over 19,000 records for treated yield but less than 7,500 records for lodging and leaning, all of which were variates that should have been scored on all trials according to current protocols. When the data from the AGOUEB series was added to the official set, we had 46 phenotypes for spring barley where there were over 1000 data values which we considered worthwhile for data analysis and 38 for winter barley. In addition, winter hardiness data from an INRA (France) site was available for each year from 1997 to 2005, apart from 2001. We also included two variates obtained from these trials making 40 in total for winter barley.

None of the varieties had been used in trials throughout the period 1988 to 2008 and most were in trials for only two or three years. The winter barley Pastoral was in trial for each year apart from 2004 and 2005 and the spring barley Optic has been in trial each year from 1992.

3.4 - Diversity in Public material

Considering BOPA1 and 2 genotypes, 463 SNPs were excluded from further analyses because they had >10 % missing values. With the remaining data we built two data matrices to be used in further analyses: (1) a data matrix of 523 lines x 2610 SNP markers containing all the SNP data and (2) a data matrix of 523 lines x 890 SNP markers without non mapped and map position redundant SNP markers.

The rationale behind removing position redundant SNP markers was to avoid bias towards low recombinogenic gene rich heterochromatic regions such as centromeres. If a big portion of SNP markers cluster together, there is the possibility that too much weight is given to that region in particular. It is important to note that genetic dissection of the mapping position of each SNP is dependent upon the polymorphism in the 4 mapping populations used to build the consensus map and that SNP markers excluded from this data set are (1) not necessarily redundant genetically and (2) not limited to chromosomal centromeric regions.

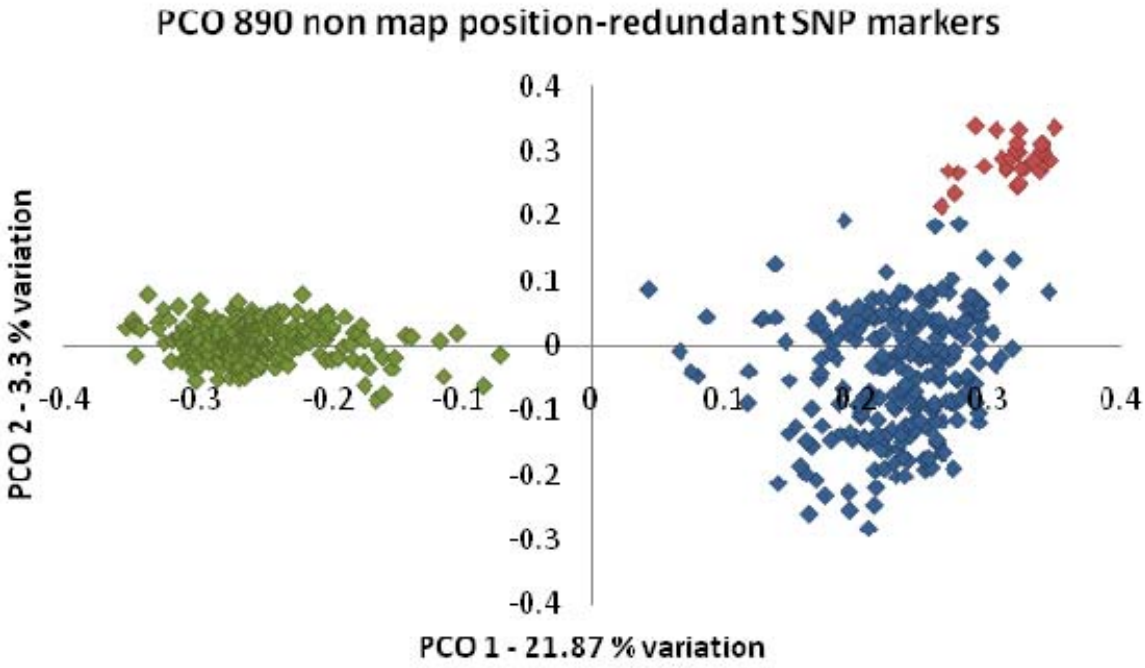

Figure 3 : Diversity of AGOUEB Public set. Green diamonds represent spring lines, Blue represents winter two-row lines and Red represents winter 6 row lines.

Principal Coordinate Analysis, PCO, based on simple matching of SNP alleles was performed with Genstat 13 using the second matrix of 890 non-redundant (by map position) SNP markers. PCO analysis partitioned the AGOUEB public dataset as expected, into two-rowed spring barley varieties, two-rowed winter barley varieties and six-rowed winter barley varieties (Figure 3.). PCO 1, which separates spring and winter growth habit, accounts for 21.87 % of the genetic variation while PCO 2, which separates winter barley varieties in relation to ear morphology, accounts for 3.3 % of the genetic variation.

Whilst there was considerable diversity amongst the individual lines within each grouping, there were some lines that were different from another by less than 1% of the 3000 SNPs studied. The most similar recommended list lines were the winter two rows Angora and Melanie, which were sister lines from a cross made by the Breun breeding programme in Germany. We did not have any genotypes for one or both at 238 of the 3072 SNPs but only 24 of the remaining were polymorphic between Angora and Melanie. These varieties were indistinguishable from morphological tests (RUSSELL et al., 1997) and could only be separated by hordein profiling and six of the 21 with known map location were on chromosome 1H, where loci for Hordein A and B bands are located. Five and four SNPs were clustered at the distal and centromeric regions of chromosomes 5H and 6H respectively; regions where we did not detect any associations of markers with DUS characters. The closest spring varieties were Prestige and Class with just 29 polymorphisms from the 2754 SNP loci where we had genotypes for both. Nineteen of the polymorphisms were located in two clusters on chromosome 2H; one in a centromeric region and one towards the end of the long arm. Another cluster

of five SNPs was located towards the end of the long arm of chromosome 7H. Whilst these two varieties were distinguishable by morphological DUS characters, only one character differed and the molecular marker polymorphisms were concentrated in regions where we did not detect any associations with DUS characters. Publican and Quench are two varieties released by the same breeder at the same time from crosses between the same parents (Sebastian and Drum) but made reciprocally. Whilst Quench is an accepted brewing variety, Publican is an accepted distilling variety in the UK. There are 149 SNP differences between the two lines with 37 and 68 located on chromosomes 5H and 6H respectively, indicating that the end-user differences between the two are likely to be located in these regions.

The first data matrix of 523 x 2610 SNPs was used to determine genetic diversity present in the spring and winter elite gene pools separately. We expected that the higher selection pressure experienced by the spring gene pool would be reflected in the amount of genetic diversity present in both the spring and winter datasets. We found 48.4% of the 2610 BOPA1 and BOPA2 markers (1262) had minimum allele frequencies (MAF) lower than 5 % amongst the 253 spring barley lines present in the AGOUEB public set (Fig. 2.d.2a). In comparison, 45.6% of the 2610 markers (1190) had minimum allele frequencies (MAF) lower than 5 % amongst the 242 two rowed winter barleys in the AGOUEB Public set(Fig. 2.d.2b). Whilst the numbers are similar, by using a contingency chi-square, we detected a significant association between spring and winter habit and minor allele frequency, suggesting that the winters were more diverse. The number of 6 rowed winter varieties (<30 lines) is not large enough for a meaningful interpretation of genetic diversity in this sub-population. The presence of fixed or nearly fixed alleles at individual loci is as important as the polymorphic markers in the population, because they may represent traces of past or recent strong signatures of selection on loci of vital agronomic importance.

3.5 - Genotype data and correction of population substructure for DUS analysis

Genotyping of the public barley lines with the 1,536 feature BOPA1 SNP array provided the preliminary genotype file. After SNP-calling and primary quality checking by SCRI, this provided the basis from which subsequent data processing and cleaning were made. Markers with minor allele frequency < 0.1 or genotyping success rate ≤ 0.95 were removed from the dataset, as were cultivars with success rate ≤ 0.84. The final dataset consisted of 490 cultivars and 1,111 markers (mean nucleotide diversity = 0.41; mean, median and mode distance between markers = 1.0, 0.5, 0.0 cM respectively; 5.7 % markers ≥ 4 cM spacing) with a call rate of 0.997. Using this data matrix, we investigated genetic substructure within the association panel. Principal component analysis showed 24 % of the genetic variation can be described by the first two components; overlaying phenotypic information for the two major agronomic classes (‘ear row-number’ and ‘seasonal growth habit’) suggests these categorical trait combinations are largely responsible for the major genetic divisions observed . To account for the strong genetic stratification observed, we used the DUS phenotype/BOPA1 dataset to investigate various methodologies (STRUCTURE, Eigenstrat, Mixed Linear Model. Genomic Control was applied alone, and in combination with all other methods), finding a mixed linear regression model, with coefficients of kinship estimated using a matrix of between-individual genetic correlation, to perform best.

3.6 - Genome-wide association mapping of DUS traits and marker enrichment

Of the 33 traits with a fill of ≥200 varieties, the trait “lodicules disposition” was removed from further analysis (as there were only 5 instances of a line possessing the lower frequency character state), leaving 32 DUS characters for subsequent analysis. Initially, uncorrected genome-wide association (GWA) analyses were implemented. Analysis of quantile-quantile plots of expected and observed associations indicate that for all traits, the expectation of strong confounding due to a heavily structured population was realized in the observed excess of associations when tested without correction, and that while power to detect significant associations is retained, efficient correction for the extensive genetic substructure observed has been achieved. Subsequent GWA analysis using the mixed model to correct for population substructure identified eighteen genomic locations associated (5

$\leq$ -$\log_{10}$ $p \leq 113$) with fifteen traits. The majority of traits with significant associations appeared to identify a single genetic locus : 'seasonal growth habit' (chromosome 1H), 'grain lateral nerve spiculation' (2H), 'awn anthocyanin coloration', 'awn anthocyanin intensity', 'auricle anthocyanin coloration', 'auricle anthocyanin intensity', 'lemma nerve anthocyanin intensity' (identifying overlapping regions on chromosome 2H), 'grain aleurone color' (4H), 'hairiness of leaf sheath' (4H), 'rachilla hair type (5H), 'ear attitude' (5H) and 'grain ventral furrow hair' (6H). The ear morphology characters 'sterile spikelet attitude' and 'ear-row number' both identified two regions of association (1H and 2H; 3H and 4H, respectively). Traits for which significant associations were detected had a mean heritability of $h^2 = 0.58$, while the proportion of phenotypic variation (V_P) accounted for by peak markers was between 0.08 and 0.84 (mean $V_P = 0.41$).

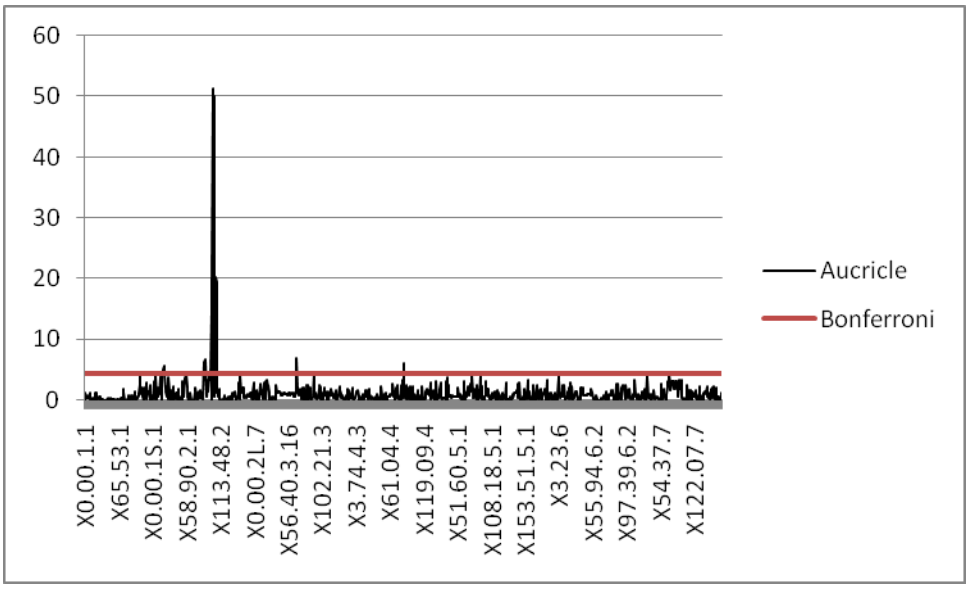

Figure 4. Genome wide association scan of the DUS character; pigmentation of auricles

3.7 - Fine-mapping the *ANT2* locus

Studies of anthocyanin biosynthesis have reported a Mendelian locus on chromosome 2H, termed *ANTHOCYANINLESS 2* (*ANT2*). To support the hypothesis that our GWA scans of anthocyanin pigmentation DUS characters (e.g. Figure 4) identified *ANT2*, we utilized a doubled haploid mapping population (n = 209) developed by KWS between two cultivars included in our association panel that differ for the presence ('Retriever') and absence ('Saffron') of anthocyanin pigmentation (subsequently referred to as 'red' and 'white' varieties), previously genotyped using the BOPA1 SNP array. Anthocyanin pigmentation was mapped as a single dominant Mendelian trait, co-segregating with markers 11_21007 and 11_21175 at 98.82 cM on the long arm of chromosome 2H, lying within the chromosomal interval identified in the association panel, and collinear with the previously mapped *ANT2* locus. We subsequently employed a composite phenotype with two character states: absence of anthocyanin coloration in all recorded tissues (awns, auricles and lemma nerves), or presence in one or more of these structures . GWA analysis found the genetic interval controlling this trait to lie between 93.5 and 103.67 cM on chromosome 2H, with the peak association (-$\log_{10}$ p = 51.7, marker 11_21175) at 96.82 cM. Towards developing additional genetic markers around the *ANT2* locus, we investigated the extent of macro-colinearity between the genetic map of barley chromosome 2H and the physical maps of rice chromosome 4 and brachypodium (*B. distachyon*) chromosome 5. Utilizing these comparative analyses, we developed genotypic assays for

six additional barley genes (HvOs04g47010, HvOs04g47020, HvOs04g47080, HvOs04g47110, HvOs04g47120, HvOs04g47170) close to the most significant markers identified during GWA analysis) and applied these across the complete panel. Subsequent association analysis shows the *ANT2* locus is defined within a ≤ 0.57 cM interval by recombination events distal to *HvOs04g47110* (Conserved Hypothetical Protein) and proximal to *HvOs04g47020* (Genetic Modifier). A contiguous barley physical map encompassing the flanking markers was constructed, and the minimum tiling path sequenced (BAC clones 77O02, 739E22, and 274B17, GenBank accession HM163343). The 260 kb sequenced interval contains eleven genes, of which eight are located at collinear positions in one or more related cereal genomes . Within the sequenced contig, three gene models were identified between the flanking markers, including a strong candidate gene encoding a protein containing a bHLH DNA-binding domain, a feature common among transcription factors known to regulate pigment synthesis in other plant species. Phylogenetic analysis of bHLH proteins from the anthocyanin pigmentation pathways of petunia, antirrhinum, maize and arabidopsis, as well as their rice homologues, shows that the barley *ANT2* candidate gene belongs to a clade containing bHLH proteins encoded by genes at the *R/B* loci, previously found to control anthocyanin pigmentation in maize. Semi-quantitative RT-PCR found *HvbHLH1* to be expressed in the target tissues of both 'Saffron' and 'Retriever. Sequencing a ~4.6 kb stretch of *HvbHLH1* from -343 to +4,628 bp in a subset of 90 cultivars (GenBank accessions HM370298 to HM370387) identified 69 polymorphisms arranged in 4 haplotypes, with haplotype 1 exclusive to 'white' varieties, while haplotypes 2-4 were associated with anthocyanin coloration in one or more tissues. The identified polymorphisms include eight synonymous and four non-synonymous variants, as well as a 16 bp deletion within exon 6 that results in truncation of the predicted protein upstream of the bHLH domain. Subsequent genotyping in the complete association panel established that the 16 bp deletion occurred in all cultivars lacking anthocyanin pigmentation, but was absent in cultivars in which anthocyanin is expressed in one or more tissues. We also found the deletion to perfectly co-segregate with 'white' *ant2* alleles in our bi-parental mapping population. Thus, the 16bp InDel is diagnostic for the ability of the plant to product anthocyanin, providing an easily applied PCR/agarose gel based barker for marker assisted selection. Further details are given in COCKRAM *et al*., (2010)**.**

4 - CONCLUSION

Our aim was to DNA fingerprint 1000 barley genotypes using a panel of 3000 molecular markers that represented sequence variation in known barley genes to characterise the variation that exists amongst UK elite barley varieties and associate sequence variants with differences in performance and morphological characters. Over 500 of these lines had been evaluated in spring and winter barley National and Recommended List trials between 1988 and 2006 and thus an extensive body of performance data (yield, height, disease resistance, quality etc.) already existed for these lines. Additionally we grew a subset of lines representing market successes and failures over our survey period to both provide an unambiguous estimate of breeding progress and additional data to improve the prediction of means of varieties that generally were not grown in the same trials.

Not surprisingly, multi-variate analysis of the marker data generated by the DNA fingerprinting resulted in three general groupings that represented spring barley, two row winter barley and six row winter barley. Whilst some varieties were genetically (and morphologically) quite similar, e.g. Angora and Melanie, there were still considerable genetic differences between varieties within the three major groupings and therefore plenty of genetic variation for breeders to continue to exploit. This was borne out by the results of the trials carried out within the project, where we clearly demonstrated breeding progress for yield, which appeared to be due to increased grain size, in the winter and spring crop and also for malt extract in the spring crop.

We have also been able to identify associations of molecular markers with morphological characters used to establish Distinctness, Uniformity and Stability (DUS) in National List testing. For the first time, we can therefore confirm that DUS characters are controlled by genes on each of barley's seven chromosomes. We have also been able to closely define the chromosomal region

harbouring these controlling genes and use the syntenic relationships between barley and the fully sequenced genomes of rice and *Brachypodium* to identify potential candidate genes for the characters. The success of this approach has been demonstrated by the cloning of the major gene responsible for the development of anthocyanin pigmentation in various barley tissues. The approach is also proving successful in identifying genes responsible for performance characters such as yield. Whilst this offers the prospect of utilising DNA markers as surrogates to assemble favourable performance alleles at a number of segregating loci in a targeted selection strategy, we still need to determine how all these genes act together to produce enhanced performance.

Finally, the extent of the DNA fingerprinting that we have carried out means future experiments can be based upon the same material that we have studied and can access the existing molecular marker information, obviating the need and expense of large-scale genotyping material in the experiment as well as phenotyping.

REFERENCES

CLOSE T.J., BHAT P.R., LONARDI S., WU Y., ROSTOKS N., et al. (2009) Development and implementation of high-throughput SNP genotyping in barley. BMC Genomics 10: 582.

COCKRAM J., WHITE J., LEIGH F.J., LEA V.J., CHIAPPARINO E., LAURIE D.A., MACKAY I.J., POWELL W., O'SULLIVAN D.M. (2008) Association mapping of partitioning loci in barley. *BMC Genetics* 9: 16

COCKRAM J., WHITE J., ZULUAGA D.L., SMITH D., COMADRAN J., MACAULAY M., LUO Z., KEARSEY M.J., WERNER P., HARRAP D., TAPSELL C., LIU H., HEDLEY P.E., STEIN N., SCHULTE D., STEUERNAGEL B., MARSHALL D.F., THOMAS W.T.B., RAMSAY L., MACKAY I., BALDING D.J., THE AGOUEB CONSORTIUM, WAUGH R., O'SULLIVAN D.M. (2010) Genome-wide association mapping to candidate polymorphism resolution in the unsequenced barley genome. *PNAS Early Ed* www.pnas.org/cgi/doi/10.1073/pnas.1010179107.

FRANCIA E., BARABASCHI D., TONDELLI A.., LAIDO G., RIZZA F. *et al*. (2007) Fine mapping of a HvCBF gene cluster at the frost resistance locus Fr-H2 in barley. *Theor Appl Genet* 115: 1083-1091.

HANEMANN A., SCHWEIZER G.F., COSSU R., WICKER T., ROEDER M.S. (2009) Fine mapping, physical mapping and development of diagnostic markers for the Rrs2 scald resistance gene in barley. *Theoret Appl Genet* 119: 1507-1522

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.

PIFFANELLI P., RAMSAY L., WAUGH R., BENABDELMOUNA A., D'HONT A., HOLLRICHER, K., JORGENSEN, J.H., SCHULZE-LEFERT P., PANSTRUGA R. (2004) A barley cultivation-associated polymorphism conveys resistance to powdery mildew. *Nature* 430: 887-891

ROSTOKS N., RAMSAY L., MACKENZIE K., CARDLE L., BHAT P., PRASANNA R., ROOSE M.L., SVENSSON J.T., STEIN N., VARSHNEY R.K., MARSHALL D.F., GRANER A., CLOSE T.J. & WAUGH R. 2006. Recent history of artificial outcrossing facilitates whole-genome association mapping in elite inbred crop varieties. *Proc Natl Acad Sci USA* 49: 18656-18661.

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

SAISHO D., POURKHEIRANDISH M., KANAMORI H., MATSUMOTO T., KOMATSUDA T. (2009) Allelic variation of row type gene Vrs1 in barley and implication of the functional divergence. *Breed Sci* 59: 621-628

SZUCS P., SKINNER J.S., KARSAI I., CUESTA-MARCOS A., HAGGARD K.G., COREY A.E., CHEN T.H.H., HAYES P.M. (2007) Validation of the VRN-H2/VRN-H1 epistatic model in barley reveals that intron length variation in VRN-H1 may account for a continuum of vernalization sensitivity. *Mol Genet Genom* 277: 249-261

TURNER A., BEALES J., FAURE S., DUNFORD R.P., LAURIE D.A. (2005) The pseudo-response regulator Ppd-H1 provides adaptation to photoperiod in barley. *Science* 310: 1031-1034.

VON ZITZEWITZ J., SZUCS P., DUBCOVSKY J., YAN L.L., FRANCIA E., PECCHIONI N., CASAS A., CHEN T.H.H., HAYES P.M., SKINNER J.S. (2005) Molecular and structural characterization of barley vernalization genes. *Plant Mol Biol* 59: 449-467.

YAN L., FU D., LI C., BLECHL A., TRANQUILLI G., BONAFEDE M., SANCHEZ A., VALARIK M., YASUDA S., DUBCOVSKY J. (2006) The wheat and barley vernalization gene VRN3 is an orthologue of FT. *Proc Natl Acad Sci USA* 103: 19581-19586.