March 2014

Project Report No. 528

Association genetics of UK elite barley (AGOUEB)

by

W Thomas[1], J Comadran[1], L Ramsay[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] and R Waugh[1]

[1]SCRI, Invergowrie, Dundee DD2 5DA
[2]NIAB, Huntingdon Road, Cambridge CB3 0LE
[3]University of Birmingham, Edgbaston, Birmingham B15 2TT
[4]KWS UK, 56 Church Street, Thriplow, Nr Royston, Herts SG8 7RE
[5]LS Plant Breeding, North Barn, Manor Farm, Milton Road, Cambridge CB24 9NG
[6]Syngenta Seeds Ltd, Market Stainton, Market Rasen, Lincs LN8 5LJ
[7]Nickerson-Advanta Ltd, Joseph Nickerson Research Centre, Rothwell, Market Rasen, Lincs LN7 6DT
[8]Secobra/Dalgety, Throws Farm, Stebbing, Great Dunmow, Essex CM6 3AQ
[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
[12]The Scotch Whisky Research Institute, The Robertson Trust Building, Research Park North, Riccarton, Edinburgh EH14 4AP
[13]Campden BRI, Lyttel Hall, Nutfield, Surrey RH1 4HY

This is the final report of a 54 month project (RD-2004-3043) which started in October 2005. The work was funded by the Sustainable Arable Link programme and a contract for £161,066 from HGCA.

CONTENTS

1. ABSTRACT

Our aim was to DNA fingerprint 1000 barley genotypes using a panel of 3000 molecular markers to characterise the variation that exists amongst UK elite barley varieties and associate variation in marker profiles i.e. DNA fingerprints 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 in a series of trials to 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.

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 individual 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 similarity 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 to predict the favourable combinations of alleles as a short-cut to directly measuring performance, a technique known as Marker-Assisted Selection, we still need to determine how all these genes act together to produce enhanced performance.

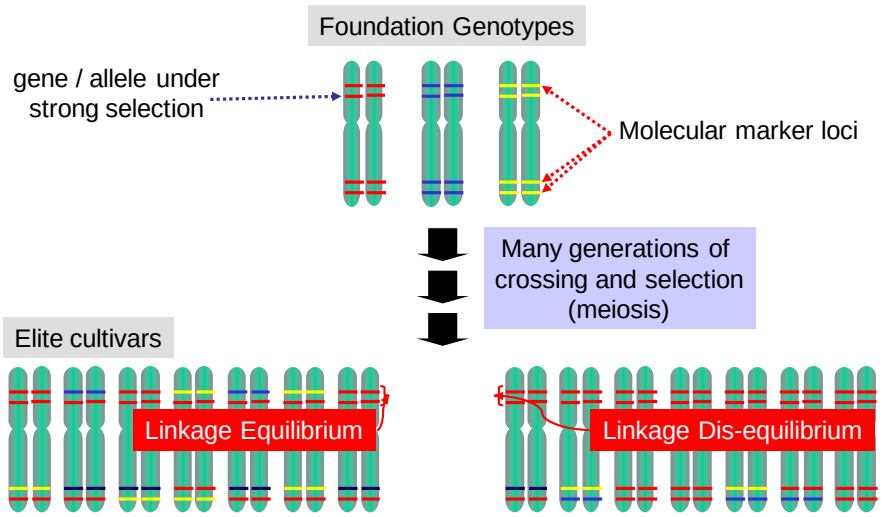
Finally, the extent of the DNA fingerprinting that we have carried out means that future experiments can be based upon the same material that we have studied but could access the existing molecular marker information, removing the need and expense of large-scale genotyping material in the experiment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

2. SUMMARY

2.1. Background

Plant breeding is a lengthy and expensive Research and Development exercise with selection for key performance characters such as yield and quality based upon direct measurements from trials that are generally grown in the later stages of the selection process. The use of molecular markers that are 'genetic signposts' for genetic regions associated with improvements in performance can be applied in the early stages of the selection process (Marker Assisted Selection) and provide an enriched gene pool from which the best of the best lines can be selected by direct measurements. Marker Assisted Selection therefore offers the promise of improved efficiency of plant breeding but, despite years of research, has yet to be realised on a large scale for small-grained cereals. Previous plant research to identify markers to use in Marker Assisted Selection have concentrated on progenies from specific crosses due to their practical ease and genetical and statistical power to detect significant associations of markers with characters, so-called Quantitative Trait Loci (QTL). Crosses were frequently 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 significant results, the applicability of such findings in general plant breeding was very limited because the crosses were far from typical of those being made in current breeding programmes. Commercial plant breeding programmes are generally based upon inter-crosses of quite similar lines where the differences are not so marked but nevertheless result in genetic improvements. In contrast, mammalian genetic studies cannot utilise such special populations and have developed approaches based upon assembling a diverse collection of individuals and collecting phenotypic data upon them. The collection is also genotyped with an extensive set of DNA markers, allelic variation at each of which is tested to determine if it is significantly associated with the observed phenotypic differences. This approach is termed association mapping and potentially can be used to assess all the differences that exist within a collection of lines, e.g. all the winter barley lines that have been recommended over the past 40 years. Effectively, the approach relies upon detecting DNA markers that are closely linked to chromosomal segments that are under selection because they contain a gene (or genes) that control important characters such as yield. In such cases, a specific allele of the closely linked marker tends to be in excess in the higher yielding lines and another allele in excess in lower yielding lines. Such a case is termed Linkage Disequilibrium as the expectation is that marker alleles would be equally distributed amongst higher and lower yielding lines if there was no association – Linkage Equilibrium. In plants, the challenge is to generate sufficiently detailed genetic fingerprints in order to detect significant marker/character associations before recombination means that allelic differences at a marker return to equilibrium (Summary Figure 1).

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 is a variable that 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 processors as well as end-users. It follows from the previous paragraph that for association genetics analyses to be effective, LD must persist to a greater degree than the average marker density. For instance, barley chromosomes should have a genetic map length of 1200 centimorgans (cM) and a typical survey with 120 Simple Sequence Repeat (SSR) markers means that LD would need to persist beyond 10cM to have a chance of detecting significant association of an SSR with a performance trait. Whilst this is possible in barley, 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 help break down adverse tight linkages.



Summary Figure 1. Derivation of inbred lines (represented by a single pair of chromosomes) with markers either associated with a character under selection (Linkage Disequilibrium, lower RHS) or not associated (Linkage Equilibrium, lower LHS).

2.2. Methodology and objectives

We viewed association genetics analyses as the most promising methodology to identify the location of genes controlling important characters for growing and processing barley and thereby identify molecular markers that could be used by barley breeders to select varieties that better meet these two needs. We therefore needed to combine a genetic fingerprinting methodology with high quality phenotypic data.

2.2.1. Genetic fingerprinting

Previous studies had shown that most of the variation amongst elite spring barley varieties originated from relatively few 'founder' varieties. This means that there is a relatively small number of alleles at any one gene and very limited variation around it. A consequence of this is that different varieties tend to share the same pattern of allelic differences, or haplotype, in any one chromosomal segment. Preliminary studies showed that linkage disequilibrium can persist for distances of up to several cM. Association genetics studies with markers spaced on average 1cM apart should therefore permit the identification of chromosomal regions affecting most characters. Prior to the project, such a scan was limited by the lack of a high throughput and low-cost genotyping assay of over 1000 markers that could be applied to a large number (1000) genotypes at the same time.

Through international collaboration, we were able to identify a large number of Single Nucleotide Polymorphisms (SNPs) in individual barley genes, which we then used to develop a high-throughput SNP genotyping platform for barley, using the Illumina Goldengate assay. Three pilot-phase 1536 SNP GoldenGate Oligo Pooled Arrays (OPAs) assays referred to as POPA1, POPA2 and POPA3 were developed and two 1536 SNP production scale OPAs, referred to as BOPA1 and BOPA2, were developed from SNPs tested on the pilot OPAs that then formed the basis of the genotyping undertaken in the project. BOPA1 consisted of single SNPs within 1536 separate genes and BOPA2 consisted of single SNPs within additional genes, additional SNPs within some of the genes on BOPA1 and specific SNPs designed to sample the known variants at key developmental loci such as Ppd-H1 and Vrn-H2, as well as the *mlo* mildew resistance locus.

2.2.2. Phenotypic data

Barley breeders have continued to improve the yield and agronomic potential of barley as witnessed by the increases offered by new additions to the HGCA/CEL Recommended Lists (RL) of varieties for growing in the UK. The phenotypic data underlying these improvements is very extensive with over 20 fungicide treated trials for spring and winter barley each year. There are generally 7-8 treated trials each year for the stage before RL, National List (NL) trials, so that the mean performance of a line that has completed two years of NL will be based upon at least 14 site

x season combinations. Most research studies of yield and agronomic performance are based upon 4-8 site x season combinations so the combination of data from National and Recommended List trials represents a major improvement in the range of environments sampled. Examination of the HGCA Recommended Lists show that the standard errors associated with the means for characters such as yield are very small compared to those achieved in normal QTL studies and the extra precision of these data increased our chances of identifying genetic regions controlling them.

No one genotype was in trials throughout the period of study, so it was necessary for some varieties to be grown over overlapping periods of at least five years to provide reliable predictions of genotypic means for material grown in different seasons. This was generally provided by the designated control varieties for National List trialling, augmented by other successful varieties that remained on the Recommended List for a long period. To improve the accuracy of these predictions, we identified a number of winter and spring barley lines for growing in trials at a number of sites for the first three harvest years of the project. These included some that had been market successes, some failures, some that failed to gain a recommendation and some key progenitors. As the micro-malting protocol had changed during our survey period, these trials also provided material for end user quality analyses through collaboration with the Maltsters Association of Great Britain, the Scotch Whisky Research Institute and Campden Brewing Research International.

2.2.3. Objectives

We therefore set the following seven objectives for the project:

1. To genotype 1000 elite barley lines with 3000 SNP marker loci developed from 1000-1500 genes (i.e. 2-3 SNPs/gene) to give the 1-2cM resolution necessary for association genetics studies in barley.
2. To assemble and analyse phenotypic data for economically important traits from extant official trial and breeders' data.
3. To generate novel phenotypic data through additional field, glasshouse and controlled environment studies. In particular these will include a detailed characterisation of disease resistance.
4. To analyse the genotypic data to determine the level and extent of linkage disequilibrium across the genome, to predict haplotype structures and to test for various factors such as population structure that shape the observed linkage disequilibrium patterns.
5. To investigate the association of the genotypic and phenotypic data to determine the genetic control of qualitative and quantitative traits. The latter will include highly heritable traits such as heading date (flowering) for which several candidate genes/genomic regions are known and other biologically and agronomically important traits such as yield and yield components that have proved less tractable to date using traditional linkage studies.

6. Development of theoretical approaches for modelling linkage and linkage disequilibrium between polymorphic markers and quantitative trait loci in the populations with and without genealogical information, and development of data-specific statistical methods and computational tools for analyzing the data generated in the project.
7. To develop a GERMINATE based web accessible public/private database that will store the genotypic and phenotypic data and the results of association genetic analysis.

2.3. Results

Objective 1 – Collection of genotypic data. We genotyped 1,000 spring and winter barley lines with 1536 BOPA1 and 1536 BOPA2 SNP markers plus another 24 with BOPA1 only. These genotypes comprise 547 in a publicly available set (the AGOUEB Public Set of RL and NL lines), 117 progenitor and disease differential lines and 360 breeders 'private' lines. Results from BOPA1 were generally better than those from BOPA2 in terms of the success rate of the markers. In addition, we identified some problems in the consistency of the genotypic data sets from BOPA1 and BOPA2, indicating some seed and/or DNA and/or data handling problems, resulting in the elimination of BOPA2 genotypes for a number of lines.

Objective 2 – Phenotypic data. We assembled data from the means of 849 fungicide treated trials grown over harvest years 1988-2006 as part of the National List and Recommended List Trial series for 282 spring and 325 winter lines for a number of variates.

Objective 3 – Collection of new phenotypic data. We selected 67 spring and 63 winter barley varieties to represent key varieties and important parents from the 1960s to the 2000s. These were grown in trials by SCRI, NIAB and the breeder partners to provide eight spring and seven winter barley sites for each of the harvest years 2006-2008.

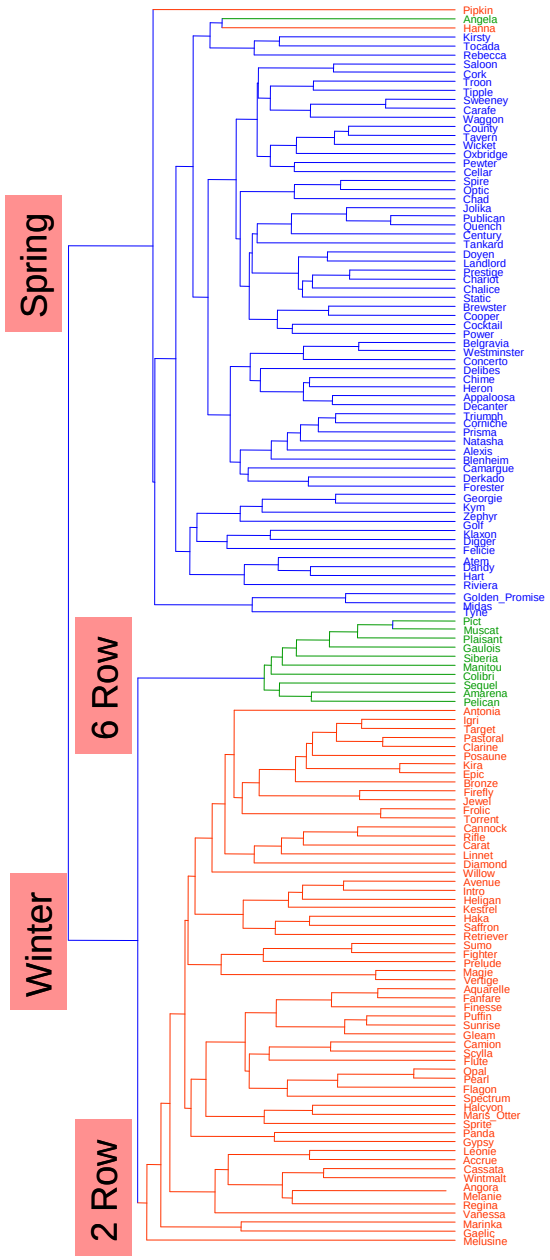
Apart from agronomic yield, the data coverage in the National and Recommended List Trials was surprisingly poor for other variates and we did not consider any variates for which there were less than 1000 data points. This left us with 44 and 39 variates from spring and winter barley trials respectively from which we derived overall means for each genotype in a mixed model analysis where we also accounted for differences between years and sites and the various interactions between the three main effects. By including the trials grown for Objective 3 as well as the data assembled under Objective 2 in the same analysis, we obtained means for up to 295 spring and 333 winter lines. We were able to obtain seed of 547 of these 628 lines and this was therefore the maximum number of lines for association analysis of the phenotypic data. These lines had all been in multi-site trials for at least two years and the data from them was shared amongst all project partners with the aim of making the data public at some time after the end of the project. We therefore called these lines the AGOUEB Public Set whereas data on lines contributed by

individual breeders remained private to the contributing company and was therefore called Breeder X's private set. The average heritability of the spring traits was 35.6% and 44.7% for the winters with a minimum of 5.6% for Germinative Energy in 4ml of water and a maximum of 83.4% for grain width:length ratio, both found for winter barley.

As part of the National Listing process, each line submitted has to be shown as distinct from any other line on the National List, that seed from the line are all the same (uniform) or that there are fewer than a defined limit of off-types, and that seed from the line breed true and thus are stable. This is established by assessing each submission over a period of two years for a range of discrete morphological characters in Distinctness, Uniformity and Stability (DUS) tests. As with the traits collected from the trials, coverage of these traits was not only highly variable but also subject to changes in the scoring for the genotypes surveyed in the project. After unifying scoring scales where necessary and possible, we had 33 DUS traits with scores for at least 200 of the genotypes under study, which we considered to be the minimum set for association analyses. Whilst eight of these were characters that fell into just two classes, e.g. two/six row ears, fourteen fell into nine different classes on a graduated scale, e.g. absence of wax to very strong wax deposition on the ear. One of the traits, disposition of the lodicules, was virtually monomorphic with only five variant phenotypes and was therefore excluded from analyses as the allele frequency was too low.

Objective 4 – Analysis of the population structure. Multi-variate analysis of the genotypic data collected upon the AGOUEB Public set revealed up to ten groups. However most of the differences between the genotypes were described by the three main developmental classes that they can be categorised by: spring two row, winter two row, and winter six row.

The three main categories can be seen in the dendrogram derived from BOPA1 genotyping shown in Summary Figure 2 where the 138 varieties from the AGOUEB Public Set that had also been placed upon the UK barley Recommended Lists are displayed. The 65 spring varieties are clearly separated from the 73 winter varieties and, within the latter, the two rows are clearly separated from the six rows.



Summary Figure 2. Dendrogram produced by BOPA1 genotyping of 138 barley varieties recommended between 1965 and 2006. Spring varieties are in blue, winter two row in red and winter six row in green.

The two most similar lines were the winter barley varieties Melanie and Angora, which were two sister selections from the same cross by the same breeder. Melanie and Angora were morphologically indistinguishable and could only be separated by examining their hordein strorage protein banding patterns. These two varieties differed at 24 of over 2800 loci where we had marker data for both, i.e. the two lines are over 99% similar. More importantly, we found that these differences were predominantly located on the short arm of barley chromosome 1, the distal portion of the long arm of 5 and the central portion of 6. The region of chromosome 1 is where genes controlling variation in the hordein storage proteins are located and thus the SNP differences in the region are consistent with hordein storage protein differences. The clear genetic differences detected by the SNP markers are in two genomic regions that are not covered by the

DUS markers that we have detected (Summary Figure 3). Whilst this difference seems small, we have sampled probably less than 10% of all barley genes yet over 16 million different genetic combinations can be produced from the 24 segregating loci that we have detected.

The next two most similar lines were the winter varieties Pearl and Opal, which again were sister lines from the same cross by the same breeder. The most similar spring varieties were Publican and Quench, which were selections by the same breeder from a cross made between the same parents but reciprocally. These two varieties differed at over 140 out of over 2800 marker loci where we have genotypic data for both, i.e. they are 95% similar.

Whilst this is a representation of the picture seen in the whole of the AGOUEB Public set, there are three winter varieties that are mis-classified on Summary Figure 2. One is Pipkin which, whilst classified as a winter variety, is a selection from a cross between Maris Otter and Sergeant and thus is likely to have a high proportion of spring alleles. The other two varieties are the six row Angela and the two row Hanna. Both these are varieties with names that have been used previously and closer inspection of the genebank accessions used to source the plant for genotyping revealed that we had accessed the earlier versions of both lines, which were spring two row types. This illustrates not only the problem in being careful to source the right accessions from gene-banks but also highlights the need to ensure that unique identifiers, such as breeders codes are stored in genebanks for all lines entered into National List trials.

The extent of Linkage Disequilibrium was variable across the genome. In centromeric regions, where recombination in barley is known to be limited, Linkage Disequilibrium persisted for over 10cM in some cases, e.g. chromosome 3H. Elsewhere in the genome, Linkage Disequilibrium decayed more rapidly but generally persisted for 2cM so that detection of marker trait associations with our marker density was truly feasible.

2.3.1. Associations with DUS characters

We detected significant associations of 41 markers with 19 of the 33 DUS traits. Most of the DUS traits where we found significant associations were associated with just one genomic location (Summary Table 1). The notable exceptions were Winter vs Spring growth habit (Season) and number of rows on the ear (Rows), where we detected 10 and 7 significant locations respectively. Whilst we correctly identified the locations of the two major loci known to affect ear row type on chromosomes 2H and 4H, some of the other predicted locations were surprising. Similarly, we correctly predicted the location of two of the seasonal growth habit loci, VrnH1 and VrnH2 on chromosomes 5H and 4H respectively but the finding of five additional associations in the region of VrnH1 was surprising.

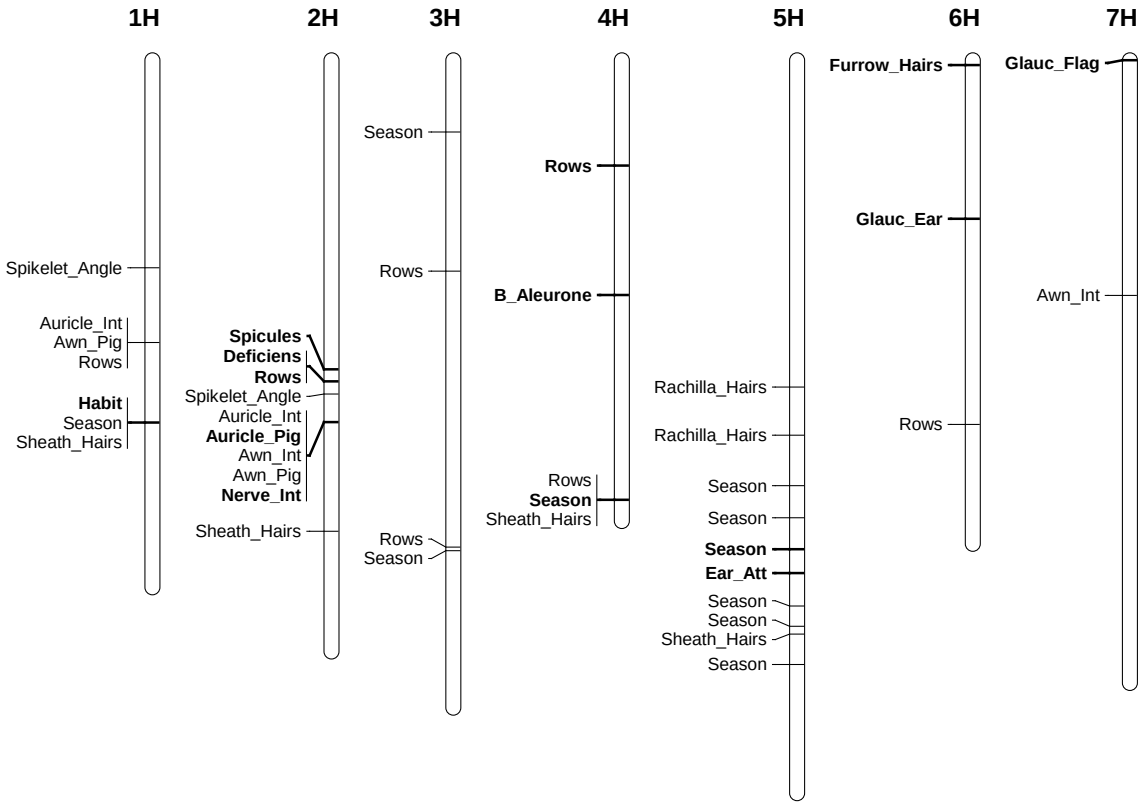
Summary Table 1. Number of significant associations detected for 19 morphological characters used to establish Distinctness, Uniformity, and Stability (DUS) in barley

Character (Abbreviation)	**Number of markers significantly associated with a character**
Intensity of Auricle Pigmentation (Auricle_Int)	2
Auricle Pigmentation (Auricle_Pig)	1
Intensity of Awn Pigmentation (Awn_Int)	2
Awn Pigmentation (Awn_Pig)	2
Blue/white Aleurone Colour (B_Aleurone)	1
Absence/Vestigial Lateral Florets (Deficiens)	1
Ear Attitude (Ear_Att)	1
Ventral Furrow Hairs (Furrow_Hairs)	1
Ear Glaucosity (Glauc_Ear)	1
Flag Leaf Glaucosity (Glauc_Flag)	1
Plant Habit (Habit)	1
Intensity of Lemma Nerve Pigmentation (Nerve_Int)	1
Length of Rachilla Hairs (Rachilla_Hairs)	2
2/6 Row Ears (Rows)	7
Winter or Spring Seasonal Habit (Season)	10
Presence/Absence of Leaf Sheath Hairs (Sheath_Hairs)	4
Spicules on Inner Lateral Nerves of Grain (Spicules)	1
Spikelet Angle	1
Grand Total	41

The distribution of the loci controlling the DUS traits is shown in Summary Figure 3. We clearly identified a major locus affecting pigmentation on chromosome 2H and used this information to identify the causal gene and the polymorphism responsible for non-production of anthocyanin. Whilst this appeared a reasonable sampling of the barley genome and many of the loci identified were in plausible regions, e.g. Leaf Sheath Hairs and Blue Aleurone on chromosome 4H, there were also some surprising associations. For instance, the association of Rows and Season on the long arm of chromosome 3H was in the region of the *sdw1* dwarfing gene, which is only found in spring two-rowed types in UK elite germplasm. It seems unlikely that this region of the genome affects these characters and the finding may be due to the assumption of additivity for characters known to be subject to gene interactions (epistatis).

Nonetheless, we have identified major loci controlling many DUS characters and this knowledge will help barley breeders develop new varieties that are distinct from each other. There are also large regions of the barley genome that are not involved in the genetic control of DUS characters

and thus allow barley breeders unconstrained opportunities for recombining to improve phenotypic performance.

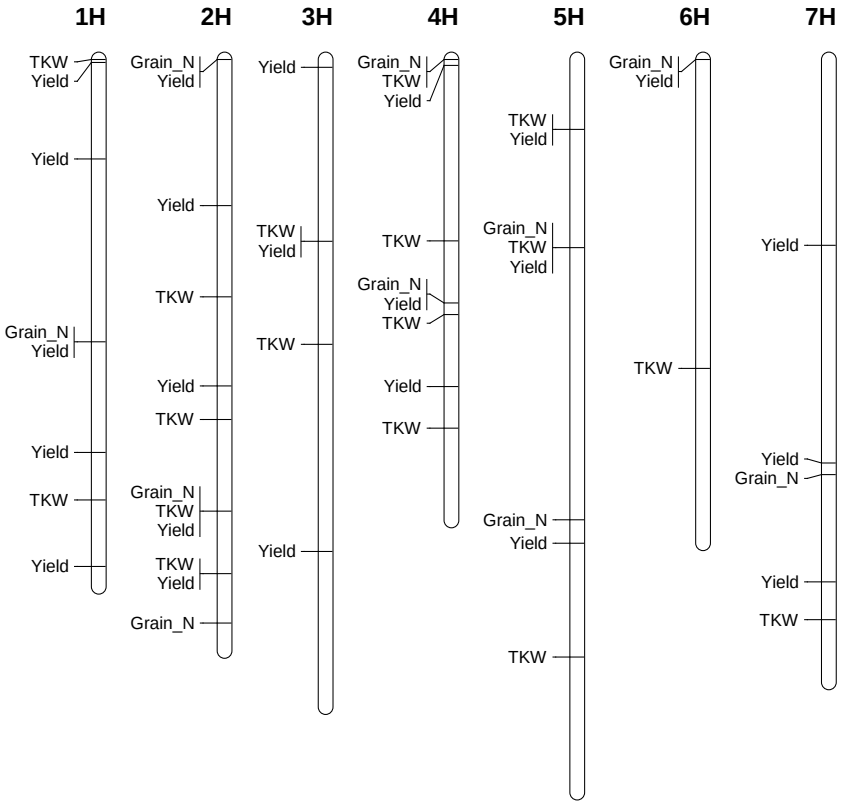


Summary Figure 3. Chromosomal locations of marker associations with 19 DUS characters. Characters in bold indicate locations where just one marker is associated with a trait or where a location corresponds to the map position of the morphological character.

2.3.2. Associations with VCU characters

We used a kinship matrix constructed from the BOPA1 genotypes to correct for population sub-structure in the spring genotypes and then conducted single marker genome scans to identify significant associations with VCU phenotypic data. Here, we have summarised the findings from studies of the BLUPs derived from an overall analysis of the NL and RL databases (see above). Such an analysis will detect overall mean effects for each locus but, because of the unbalanced nature of the data-set, we cannot detect QTL x Environment interactions, which are likely to be important for some characters. Nevertheless, we detected at least one significant marker/phenotype association for the characters that we studied and the genomic locations of markers associated with Yield, Thousand Kernel Weight and Grain Nitrogen are shown in Summary Figure 4. Most of these associations can be found amongst the large body of published QTL locations from pair crosses and are therefore validated by these reports. Crucially, we have identified all these associations from studying existing data for just one population, whereas the published reports have been accumulated from a large number of separate studies. Thus, our approach has not only provided valuable data but also been extremely cost-effective.

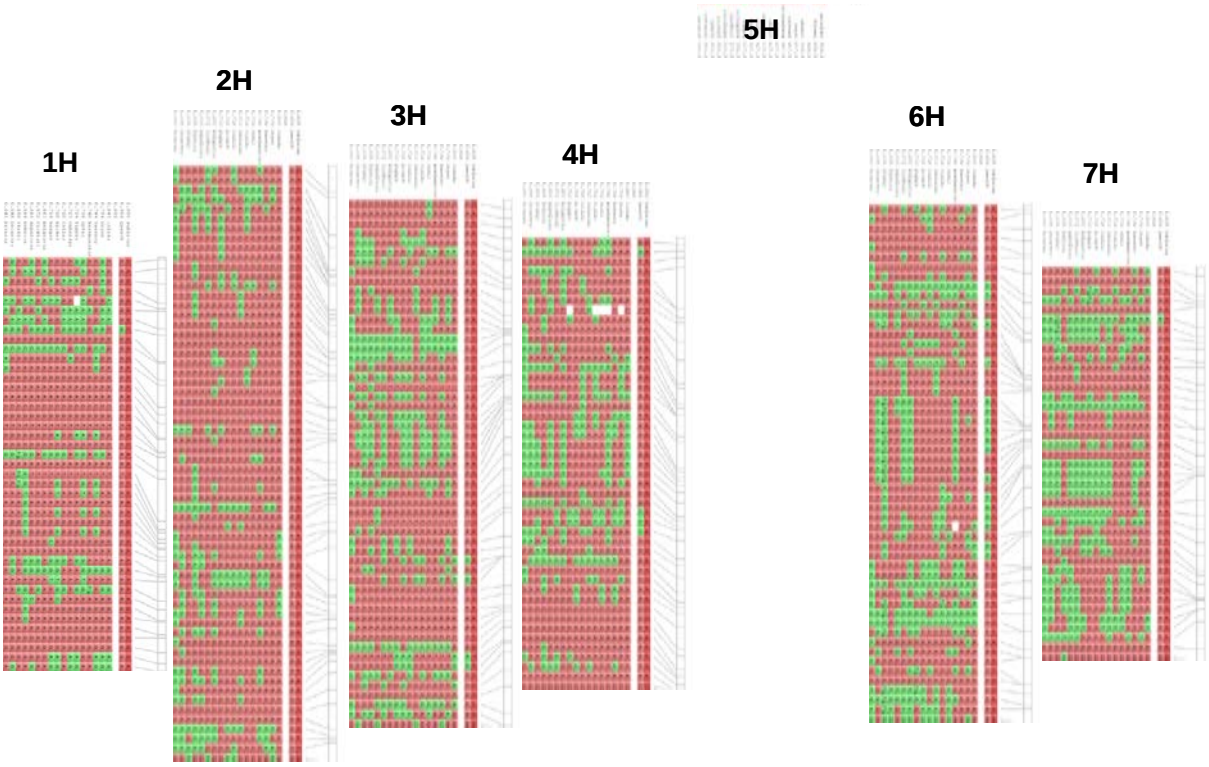
Summary Figure 4 also shows that many of the associations are co-located, reflecting the correlations between characters. For instance, a marker towards the centromeric region of chromosome 5H is associated with all three characters such that an increase in thousand grain weight results not only in an increase in yield but also a decrease in Grain Nitrogen content. Curated BOPA2 data were not available in time to conduct analyses to report here but a fuller analysis of VCU data, especially BLUPs obtained by including the data generated from the AGOUEB trial series, using BOPA1 and BOPA2 markers will be carried out and the scans lodged on the AGOUEB web site (www.agoueb.org).



Summary Figure 4. Chromosomal location of markers significantly associated with yield, thousand kernel weight (TKW) and grain nitrogen content (Grain_N) for more than 200 spring barley lines grown in UK National and Recommended List treated trials 1988 - 2006.

Objective 6 – Alternative analytical methods for association genetics. Existing and novel approaches to account for population structure were tested and applied to the data generated from the project. Whilst some of the novel approaches appeared promising alternatives, we decided to adopt a standard approach for the analyses of associations with DUS and VCU characters. This was based upon use of a kinship matrix to account for population structure.

Objective 7 – Develop a database for data storage and retrieval within the project. We used a GERMINATE database (http://bioinf.hutton.ac.uk/germinate) to store all the data associated with the project as it is designed to hold a range of passport type data to classify entries and a range of additional data types including genotypic data, phenotypic data, and pedigrees. We have also developed a range of associated tools to visualise the data. For instance, we have developed FLAPJACK (http://bioinf.hutton.ac.uk/flapjack) to visualise the matrix of genotypes and markers. Summary Figure 5 has been produced from a FLAPJACK representation of the HGCA/CEL 2008 spring barley Recommended List to highlight the differences between the 19 varieties with Publican and Quench being the two rightmost varieties for each chromosome. Quench is a selection from a cross between the same parents as Publican and differs at 34 of a subset of 384 of the 1536 BOPA1marker loci. The question is whether any of these 34 differences are associated with the fact that Publican is an accepted distilling variety whereas Quench is an accepted brewing variety.

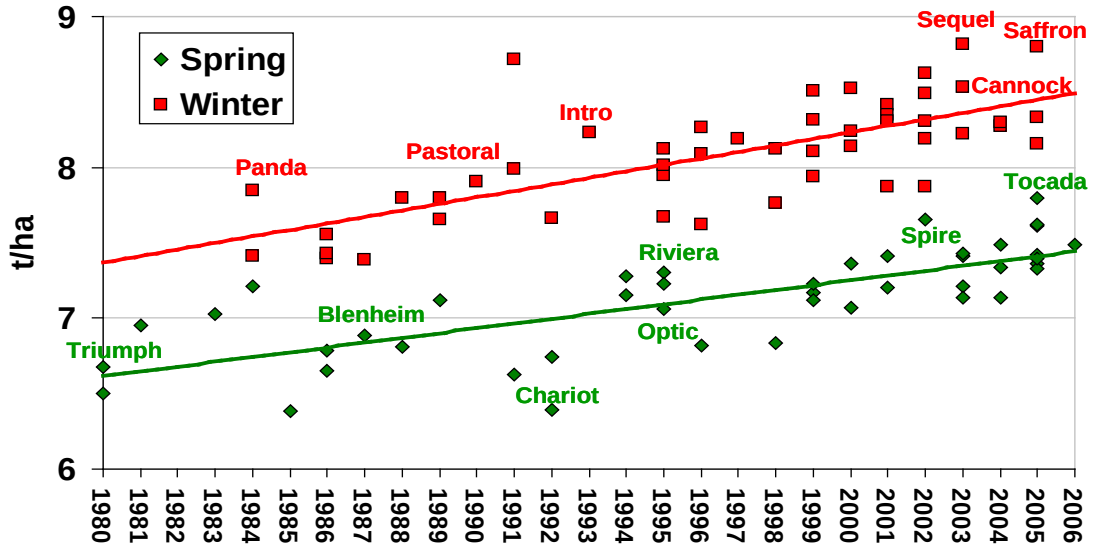


Summary Figure 5. Graphical genotypes of 19 spring barley varieties on the UK Recommended List for 2008. Varieties are organised in columns for chromosomes 1H to 7H. 384 BOPA1 SNP markers are organised by chromosomes into rows. Each little cell therefore represents the genotype of one variety for one marker and every one is coded with reference to Publican (right-most column for each chromosome) so that all red cells are the same as Publican and all the green cells indicate varieties carrying the alternative allele.

2.3.3. Genetic progress

The trials that were grown during the AGOUEB project not only provided a means of improving the prediction of means but also an unambiguous estimation of breeding progress. The varieties grown in the trials were first recommended over a range of dates between 1965 (Maris Otter) and 2006 (Appaloosa). For each crop type, we can therefore regress the mean of any variety for a given phenotype against the year in which it was first recommended and a significant relationship indicates significant genetic change over time. Summary Figure 6 shows significant improvement in yield of both the winter and spring crop over time with an average improvement of 38 kg/ha/yr for the spring and winter crop, equivalent to approximately 0.5% per annum.

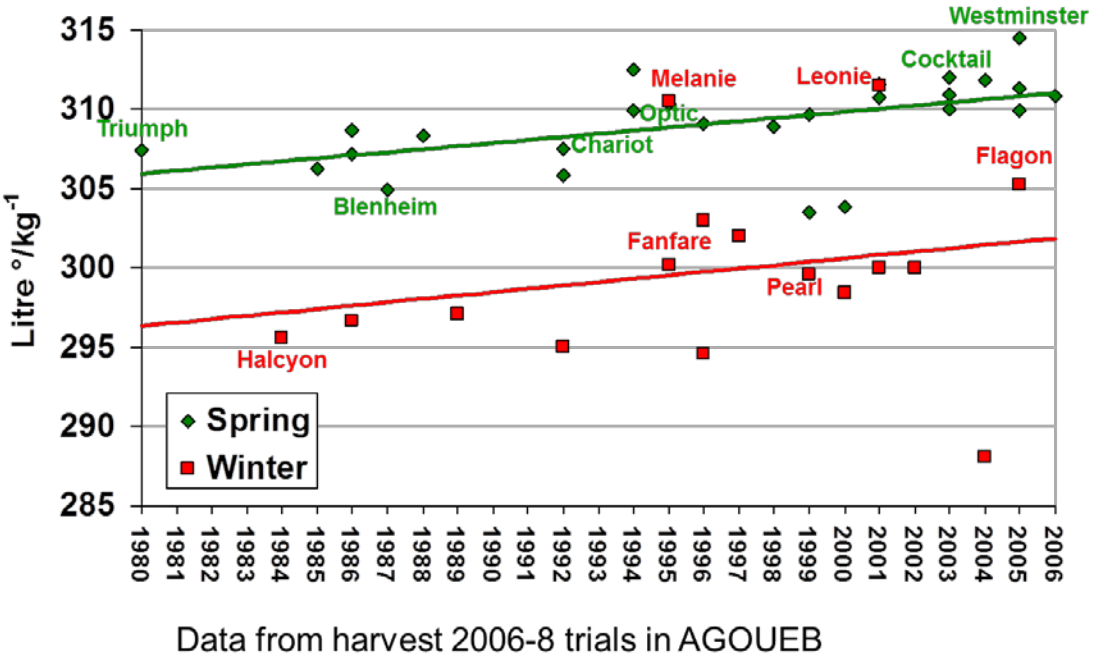
As noted in the full report, much of this increase in yield appears to be due to an increase in thousand grain weight in both crop types, which has increased by an average of 0.15g/yr over the same period. There has been no change in heading or maturity date for both crop types over the same period, indicating that improved efficiency of conversion of biomass into grain rather than an extension of the growing season was responsible for the yield increase. Despite the fact that most of the spring crop is semi-dwarf, there has been a significant decrease in plant height over the period of 0.4cm/yr but there has been no significant change in the winter crop over the same period.



Summary Figure 6. Overall mean fungicide treated yield derived from varieties grown together in over 20 trials from 2006-2008 plotted against their year of first recommendation. The red and green lines show the estimated linear relationship between yield for the winter and spring varieties respectively over time.

Similarly, we have also used the data derived from the analyses conducted by MAGB and SWRI to find a significant improvement in hot water extract of 0.2 litre°/kg for the spring crop over the period after the introduction of Triumph (Summary Figure 7). Whilst this does not seem large for a

character with a mean of over 300, it is highly significant for an industry consuming 1.75 million tonnes per annum in the UK. There is also an apparent increase in the hot water extract of the winter crop over the same period but this is not significant. It is biased by the apparently high extracts obtained for Melanie and Leonie, which were only assessed at one site but significantly out-performed the best control (Flagon).



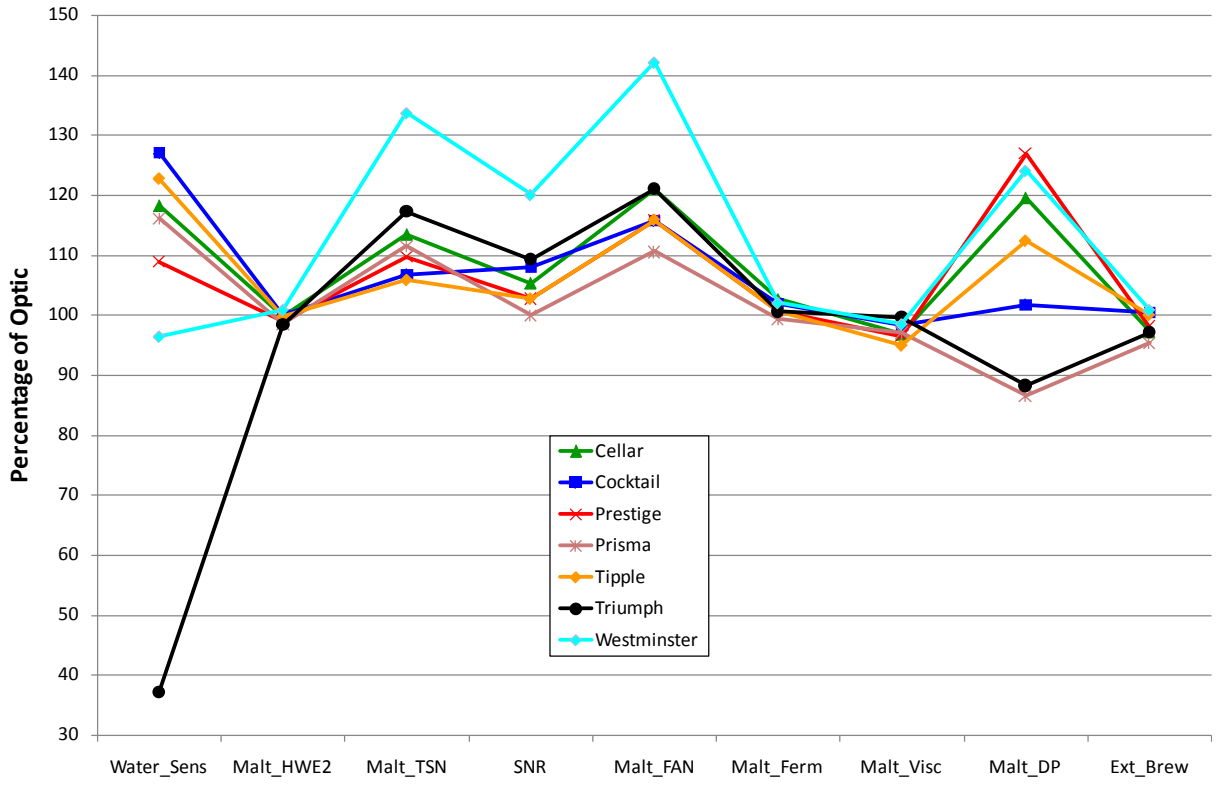
Summary Figure 7. Overall mean hot water extract of IBD approved spring and winter barley lines derived from analysis of micro-malted samples from trials of the varieties grown together in trials from 2006-8.

2.3.4. Brewing trials

We also conducted pilot brewing from 25kg lots of each of 8 spring and 4 winter barley varieties grown at two sites for each crop. Micro-malting analyses are typically carried out on small lots which may well behave differently to larger lots in commercial malting, brewing and distilling. We have analysed the data obtained from the pilot brewing to verify that genetic differences are still observable in the larger lots. All the varieties that we tested had been approved by the Institute of Brewing and Distilling so significant differences between varieties represent differences that are breeding targets to either maintain or improve upon for future varietal development.

Over 80 variates were measured on the samples from dry grain tests through to sensory attributes of the resulting brews. We detected genetic differences for 12 and 8 variates for the spring and winter barley samples respectively. Given the small number of varieties, especially for the winter crop, the finding of significant variation for some traits is highly relevant to both breeders and maltsters. It is noticeable that most of the genetic differences were found amongst the grain and processing characters and that there are few significant genetic differences for the beer and

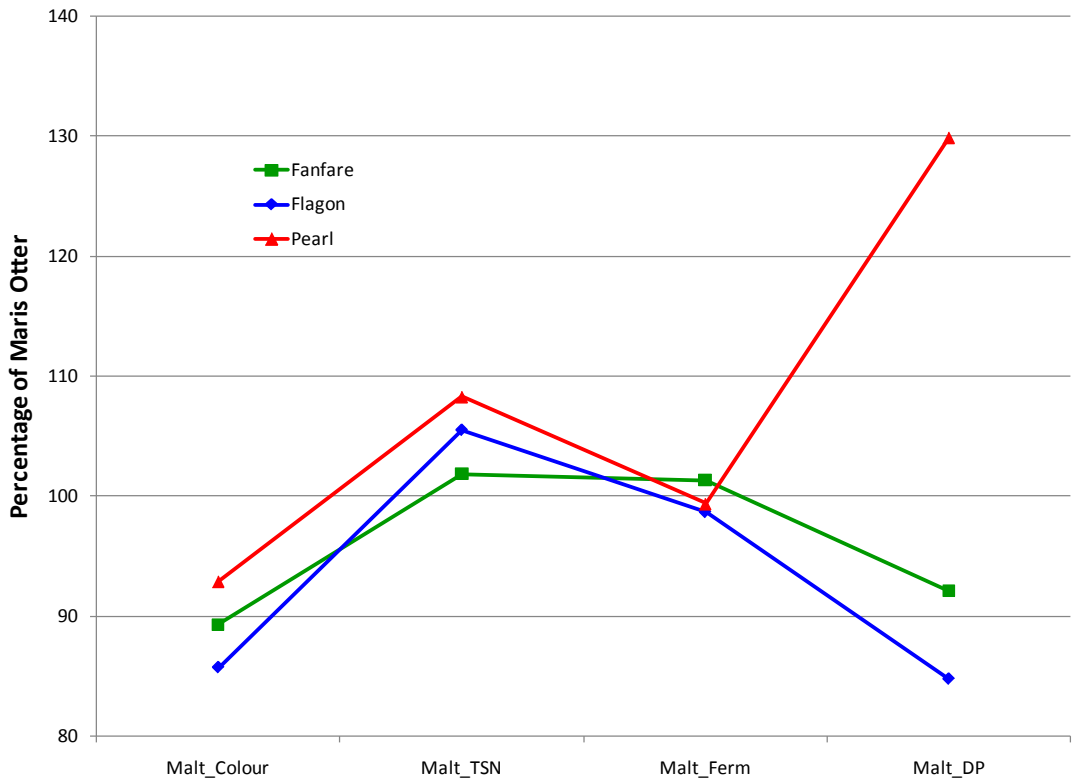
sensory characters, which illustrates the skill of the maltsters and brewers in adjusting their processes to achieve a uniform product. Amongst the springs, there was significant genetic variation for Di-methyl Sulphide (DMS) in the beer produced with the varieties Prisma and Triumph being significantly higher than the others. In general, Optic was one of the lower scoring varieties for the significant characters (Summary Figure 8), although this would partly reflect its reputation for low grain nitrogen accumulation for characters like Malt Total Soluble Nitrogen and Malt Free Amino Nitrogen. The newer varieties like Westminster and NFC Tipple tend to out-perform the older varieties like Prisma and Triumph, confirming that the genetic progress detected by micro-malting assays is reflected in an environment more akin to commercial processing.



Summary Figure 8. Means of 8 spring barley IBD approved varieties for nine characters relating to malt and brewing quality where there were significant genetic differences. All means are expressed as a percentage of Optic (100%).

The newer varieties have a significantly lower colour score than Maris Otter and tend to solubilise more nitrogen, whilst Pearl has a substantially higher diastatic power (Summary Figure 9). Other genetic differences, apart from three sensory attributes are, however, non-significant. Whilst not presented here, analysis of the lautering performance of Maris Otter shows that it processes remarkably well, even from a high nitrogen sample (>2%). By contrast, at least one of the two samples of the other three varieties, which were grown at the same sites, required raking for lautering to complete. Maris Otter clearly possesses attributes that have been recognised by

maltsters and brewers to the extent that some 5% of the tonnage purchased by English maltsters from harvest 2009 was Maris Otter.



Summary Figure 9. Means of 4 winter barley IBD approved varieties for four characters related to malt and brewing quality where there were significant genetic differences. All means expressed as a percentage of Maris Otter (100%).

2.4. Key conclusions

1. DNA fingerprinting has enabled the visualisation of the relationships between elite UK barley lines and their correct classification into morphological groups. Within these groups, two row malting winter barleys are clearly separated from two row feeding types.
2. We have identified strong marker trait associations for a range of DUS characters that for the first time provide an accurate view of the genomic regions sampled by such tests. The genetic resolution of the portion of the genome where the candidate genes for these characters is much finer than was possible by conventional bi-parental mapping and has permitted the identification of the major gene controlling anthocyanin pigmentation in various barley tissues.
3. Similarly, we have localised genomic regions controlling characters such as yield to a small number of candidate genes in several cases that provide real opportunities to not only utilise the associated markers in effective selection programmes but also to identify the causal genes.

4. We have also demonstrated that, despite the breeding concentration upon crosses within the elite gene pool, there remains a vast untapped reserve of unexploited genetic combinations, even when considering just recommended varieties.
5. Exploitation of new genetic combinations has sustained continued genetic progress in breeding for yield and is likely to continue to do so for the foreseeable future. The challenge is to design protocols to break down deleterious linkage blocks and re-assemble them in a targeted manner.
6. Whilst all current winter barley malting cultivars in the UK can be traced back to Maris Otter through their pedigrees, some of the processability of Maris Otter appears to have been lost in the improvement of other malting characteristics. The identification and genetic localisation of such a desirable characteristic unfortunately remains limited by the current lack of the ability to phenotype a large number of different samples, unless an appropriate surrogate test can be designed.
7. The genetic fingerprints and associated seed resources for the AGOUEB Public Set of genotypes form a biological and data resource that can and is being exploited in other projects, provided a cost-efficient means of identifying marker trait associations. For instance, projects on nitrogen use efficiency, Ramularia resistance, Fusarium resistance and second generation biofuels have all been funded and are exploiting resources generated within AGOUEB.

3. TECHNICAL DETAIL

3.1. Germplasm

3.1.1. Description

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.

3.1.2. Numbers and status of lines studied

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 547 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.

3.2. Genotyping

3.2.1. Description of markers

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.

Identification of Single Nucleotide Polymorphisms and development of GoldenGate Assays

Three pilot-phase 1536-SNP GoldenGate Oligo Pooled Arrays (OPAs) assays were developed. These "pilot OPAs" are referred to as POPA1, POPA2 and POPA3. Two 1536-SNP production-scale OPAs, referred to as BOPA1 and BOPA2, were developed from SNPs tested on the pilot OPAs. 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) and we have summarized the five OPAs below.

1. *POPA1 and POPA2*. The contents of POPA1 and POPA2 came from an initial list of SNPs comprised of the union of three intersecting lists from SCRI (1,658 SNPs), IPK (985 SNPs) and UCR (12,615 SNPs). SCRI and IPK SNPs were derived from PCR amplicon sequences, whereas UCR SNPs were derived nearly entirely from EST sequences. In the selection of SNPs for the OPAs, preference was given to SNPs derived from amplicon sequences. Nearly all SNPs on POPA1 and about 60% of the SNPs on POPA2 targeted stress-regulated genes. The composition of POPA1 included 1524 barley SNPs, one per gene, of which 1033 were derived from ESTs and 491 from amplicon sequences. The

composition of POPA2 included 1536 barley SNPs, one per gene including an additional SNP for 258 genes represented on POPA1, of which 1456 were from ESTs and 80 from amplicon sequences.

2. *POPA3*. The sources of POPA1 and POPA2 were insufficient to produce 1536 SNPs to complete the design of POPA3 without reducing the SNP selection criteria. Additional SNPs were therefore derived from the following three sources: *a)* an extended list of 5,732 SNPs identified in SCRI amplicon sequences, *b)* colleagues who contributed SNPs from amplicon sequences of specific genes of biological interest and *c)* an expanded barley EST resource. The first two of these additional sources were exhausted for POPA3 design. In the selection of EST-derived SNPs, priority was given to genes previously classified as having interesting expression patterns during malting or upon exposure to pathogens, or relevant to malting, brewing quality, abiotic stress or phenology. POPA3 was composed entirely of barley SNPs, in many cases more than one per gene and in some cases including genes represented on POPA1 or POPA2. In total, 967 POPA3 SNPs were derived from ESTs and 569 from amplicon sequences.
3. *BOPA1*. BOPA1 represented 705 SNPs from POPA1 and 832 from POPA2, including one SNP in common. All BOPA1 SNPs had a satisfactory technical performance on POPA1 or POPA2 and a Minor Allele Frequency (MAF) of at least 0.08 from a survey of ca 200 lines in a pilot study. BOPA1 included 1414 mapped and 122 unmapped SNPs (see below).
4. *BOPA2.* BOPA2 represented 406 SNPs from POPA1, 178 from POPA2 and 952 from POPA3. The primary emphases of BOPA2 were representation of mapped SNPs that were not included on BOPA1 and inclusion of multiple SNPs for certain genes to reveal haplotypes at these loci, with some weight given to MAF. BOPA2 contained 921 SNPs with MAF >0.08, 256 SNPs with MAF >0.04 <0.08, 345 SNPs with MAF >0.005 <0.04, and 14 SNPs with only one allele (MAF = 0) in the germplasm examined in the pilot study. BOPA2 included 1263 mapped and 273 unmapped SNPs (see below).

Genetic linkage maps

The pilot OPAS were used to genotype three barley mapping populations (Steptoe x Morex, Morex x Barke and the Orgeon Wolfe Barleys OWB). For each of the mapping populations the linkage groups separated cleanly using MSTMap at LOD 4 or 5. MSTMap first identifies linkage groups, then determines marker order by finding the minimum spanning tree of a graph for each linkage group, then calculates distances between marker using recombination frequencies. JoinMap 4 was used to confirm linkage groups and marker order determined by MSTMap. These three maps together with a fourth map produced from a population genoptyped 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. Because the SNP data are

more complete and seem generally to be of higher quality than the SxM bin marker data, the 2943 "SNP-only" map and its distance coordinates are taken as the central point of reference in this paper. 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.

Once the SNP loci were arranged by position on the consensus map, graphical visualization enabled inspection of the distribution of recombination events that indicated that there were no singleton double recombinant loci in densely marked regions of any of the maps. Since such loci are often indicative of genotyping errors, the complete absence of suspicious double recombinants can be considered an indicator of high fidelity of the data from the 2943 SNP loci selected for linkage map production. Other quality metrics include the frequency of missing data or apparent heterozygosity; with practically all individuals in all three mapping populations having homozygous genotype calls for all loci and no missing data. This is 100% of 153,636 possible genotype calls in the MxB population, 99.999% of 145,266 possible genotype calls in the OWB population and 100% of 116,840 possible genotype calls in the SxM population. The high fidelity and lack of missing data among these selected 2943 SNPs facilitated the production of individual and consensus maps.

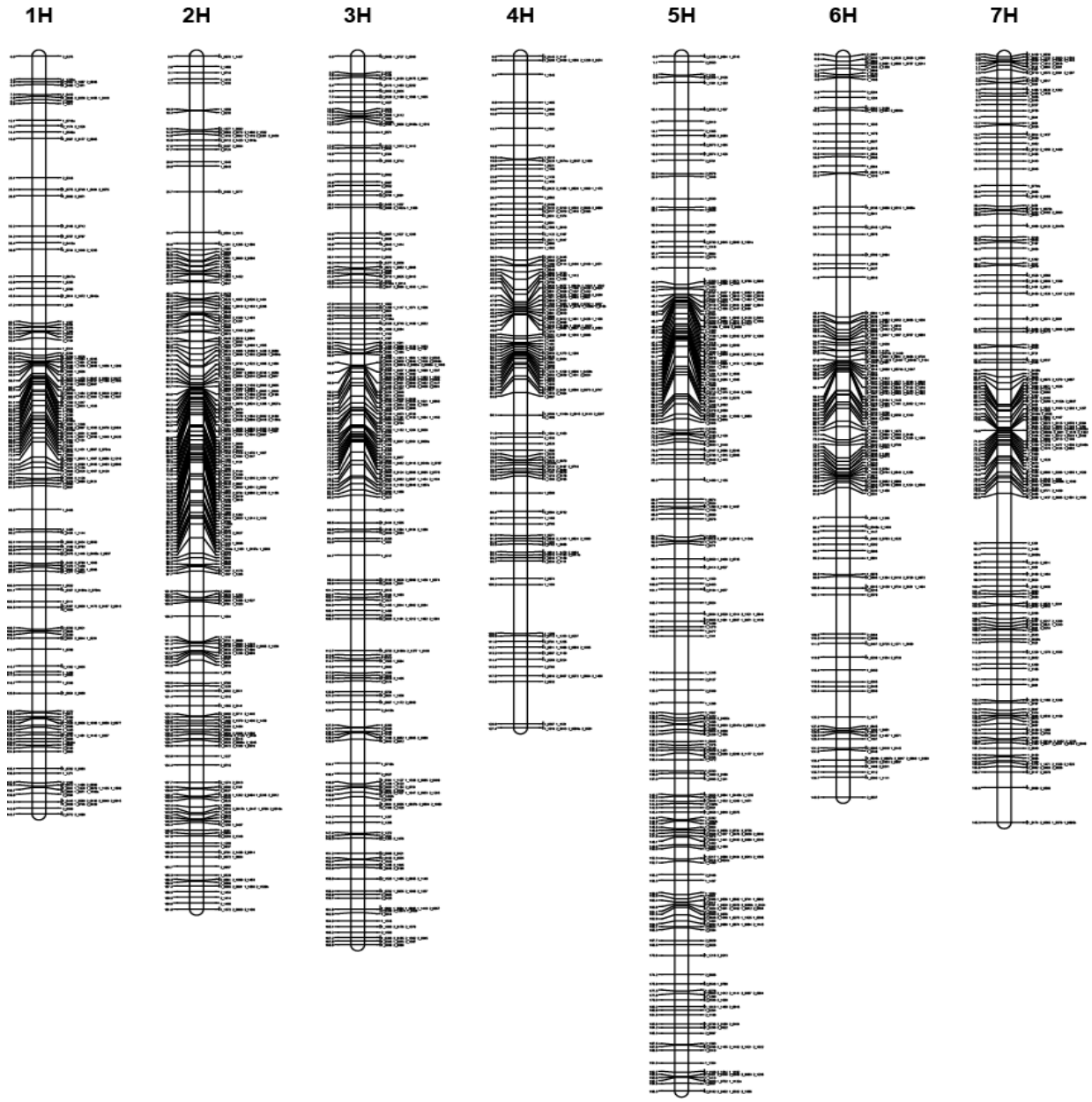


Figure 1. Integrated SNP-based gene map of 2943 BOPA1 and BOPA2 SNP Markers.

Synteny

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. An illustration of barley-rice synteny for all seven barley chromosomes is provided in Figure 2. 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.

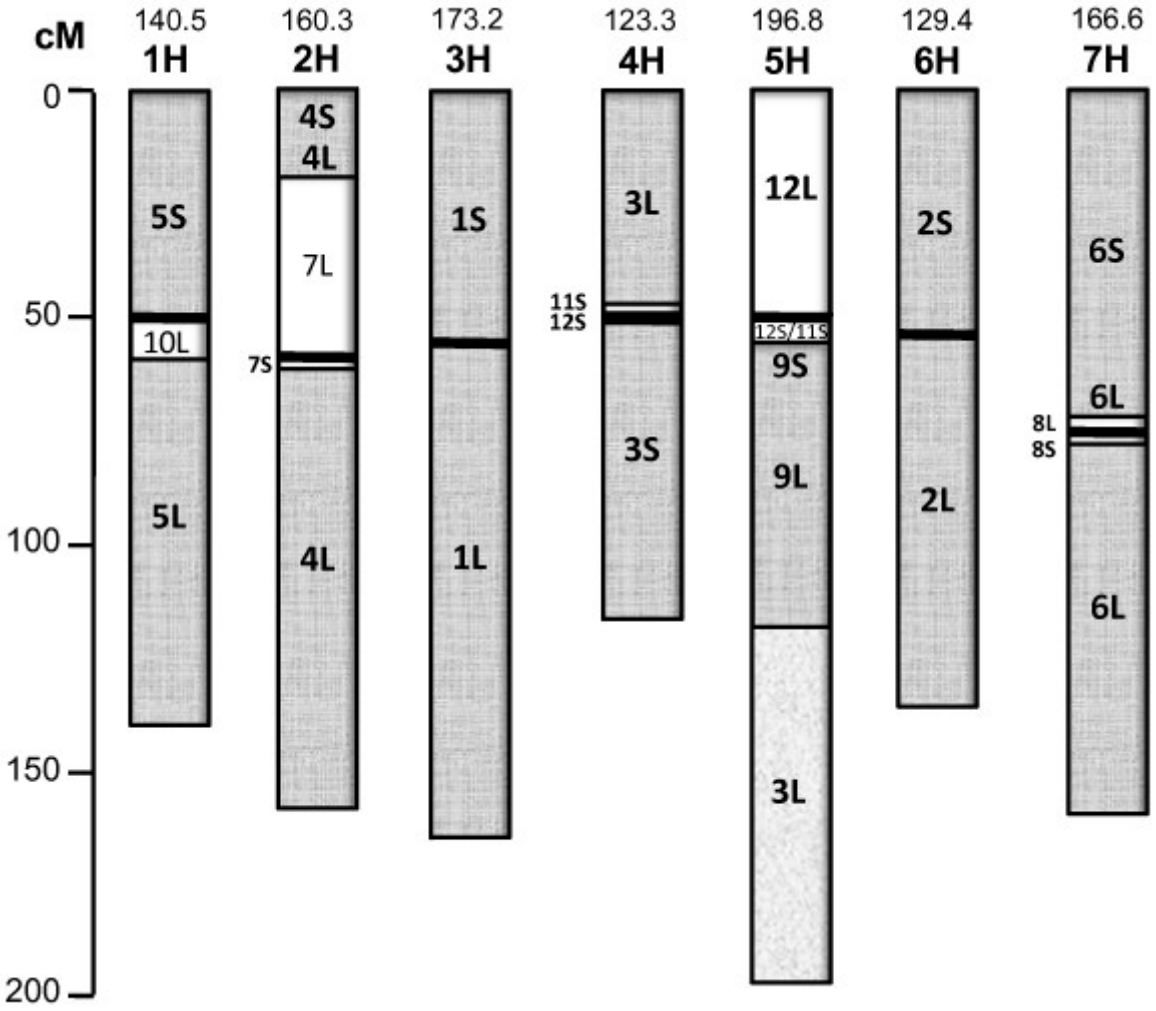


Figure 2. Barley-rice synteny summary. Seven barley linkage groups represented as rice synteny blocks. Numbers inside each barley chromosome indicate syntenic rice chromosome arm. (Close et al 2009).

BOPA1 and BOPA2 elements and performance

As discussed above, 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 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.

3.2.2. Diagnostic markers

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 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. 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). The issues, that there can be with use of the SNP genotyping platform for diagnostics, are thus due to the complexities of the underlying genetics rather than any practical issues with the technology.

Another example of the diagnostic markers included on BOPA2 are the SNPs within *HvHox1* DNA binding protein that has been shown to be *VRS1*, 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. 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 (Haneman 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.

3.2.3. Databasing and visualisation

Description

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.

Germinate AGOUEB has been deployed using the MySQL relational database management system and all user interfaces and export routines written using the Perl programming language. The modular and standardised nature of the Germinate 2 platform reduces the overheads of developing additional database analysis components and user interfaces. Other benefits include the ability to incorporate additional features which may have been developed for other Germinate 2 installations into the AGOUEB framework. At this moment Germinate 2 has been deployed publically for data from other projects such as pea (Jing et al. 2010), potato (http://bioinf.scri.ac.uk/germinate_cpc) and ryegrass (Kopecky et al. 2009). The schema designed for the storage of genotypic data was developed specifically for the high volume Illumina SNP data generated within this project.

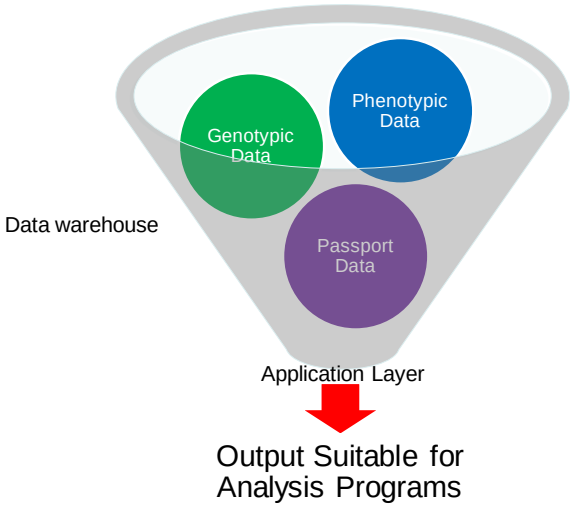


Figure 3. Overview of Germinate AGOUEB data base.

Germinate 2 AGOUEB can be better described as a data warehouse. This distinction is important and revolves around the proposed use of the system. While databases are primarily used for constantly changing data the warehouse architecture we have implemented here facilitates better the storage of historical data and is optimised for specific queries and analysis operations. While new data is easily added we assume this will come infrequently and in batches. Because of this we can make some assumptions and tailor the database to perform better under these circumstances by use of targeted de-normalisation and index generation. It also allows the easy inclusion of new data types should such a circumstance arise in the future. An alternative to this would have been to adopt a federated database model; this would however introduce new levels of complexity and an overhead burden of maintaining different database systems.

Modern association genetics based experiments rely on data from many sources therefore pooling data into one place offers many advantages over more traditional file based approaches. The most important of these is in the use of database features to ensure the validity and integrity of data being stored. The ability to view multiple data sources together offers an important aid in identifying potential problem data items which would not be easily spotted if the dataset was viewed in isolation.

Stored Data

We obtained classificatory data such as names, breeders codes, originating breeder, agent, pedigree and, where applicable AFP number (section 3.3.1), 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 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.

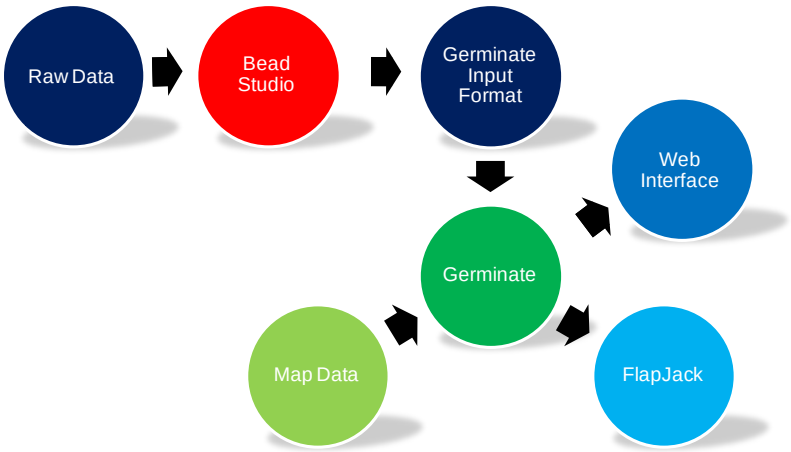


Figure 4. Genotypic data flow to and from the Germinate AGOUEB database.

The 33 DUS characters with sufficient data fill (see section 3.3.1) 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 Crop Evaluation Limited via NIAB, augmented it with information gathered from the AGOUEB fungicide treated trials and disease nurseries (see section 3.3.3) 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 (see section 3.3.2) 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 (see section 3.4.3) and VCU traits (see section 3.4.4) were stored in the database together with the model used in the analysis.

Interface

The Germinate AGOUEB database uses a custom designed web-based interface written in Perl. The interface allows users to browse datasets and generate data overviews and extract data from the varying datasets that are stored for use in external analysis packages. We have implemented several different export formats in Germinate 2 to ensure that the exported data is in the correct format for key applications such as Flapjack. Alternatively, we have provided the option to export plain text files that can then be manipulated for import into other applications. Additional formats may be supported by the development of conversion scripts which can be incorporated in the Germinate 2 application framework.

One of the main features of this interface is in the ability to generate custom groupings of lines. This enables users to only export, or view data for, a custom set of lines. Basic groupings are predefined such as all the AGOUEB lines but users can generate groups based on database queries or by uploading a text file containing AFP numbers of interest. While not mandatory the use of user logins with the Germinate AGOUEB database allows users to tag groups with their details. These groups are private only to the user who created them but may be made public by selecting a simple option from the database web interface.

3.2.4. Diversity in Public material

In total, 523 of the 547 AGOUEB barley lines from the public dataset have been genotyped with 3072 SNPs assembled in two Illumina Barley Oligo Pooled Arrays (BOPA1 and BOPA2) as 24 were genotyped with BOPA1 only. The accessions represent the available genetic variation in current elite North West European barley germplasm. 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.

Germplasm Relationships in the Public Set

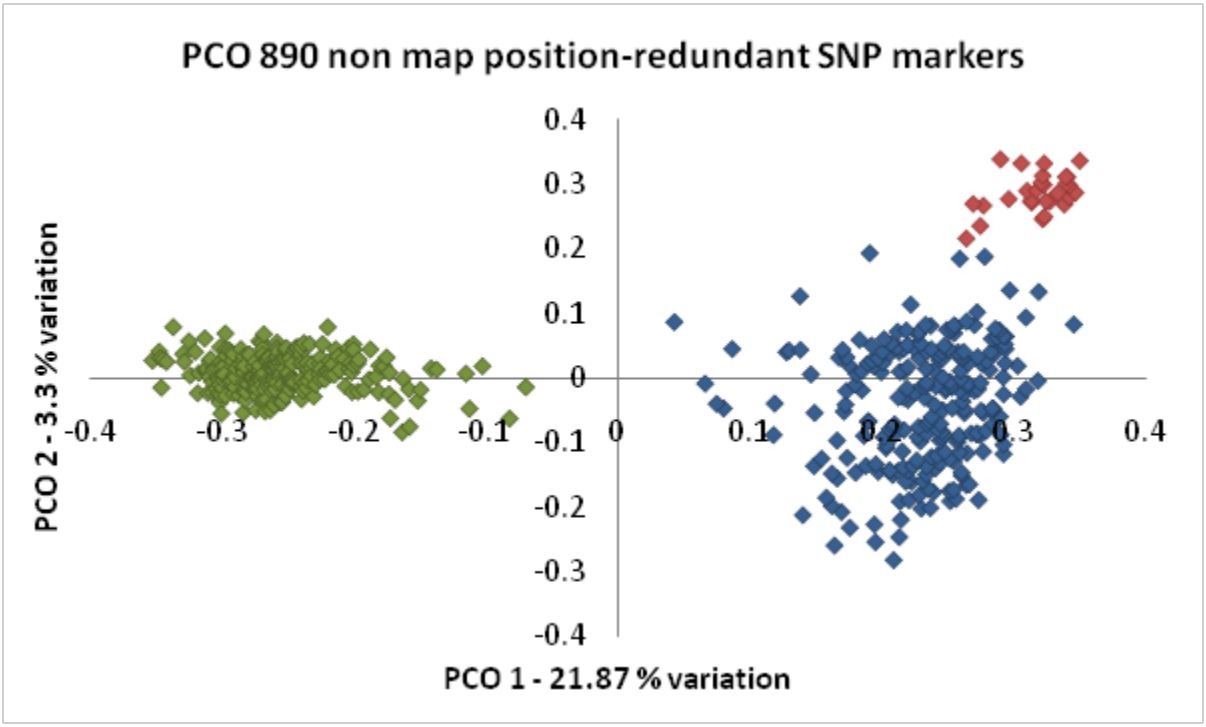


Figure 5. Diversity of AGOUEB Public set. Green diamonds represent spring lines, Blue represents winter two-row lines and Red represents winter 6 row lines.

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 5). 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.

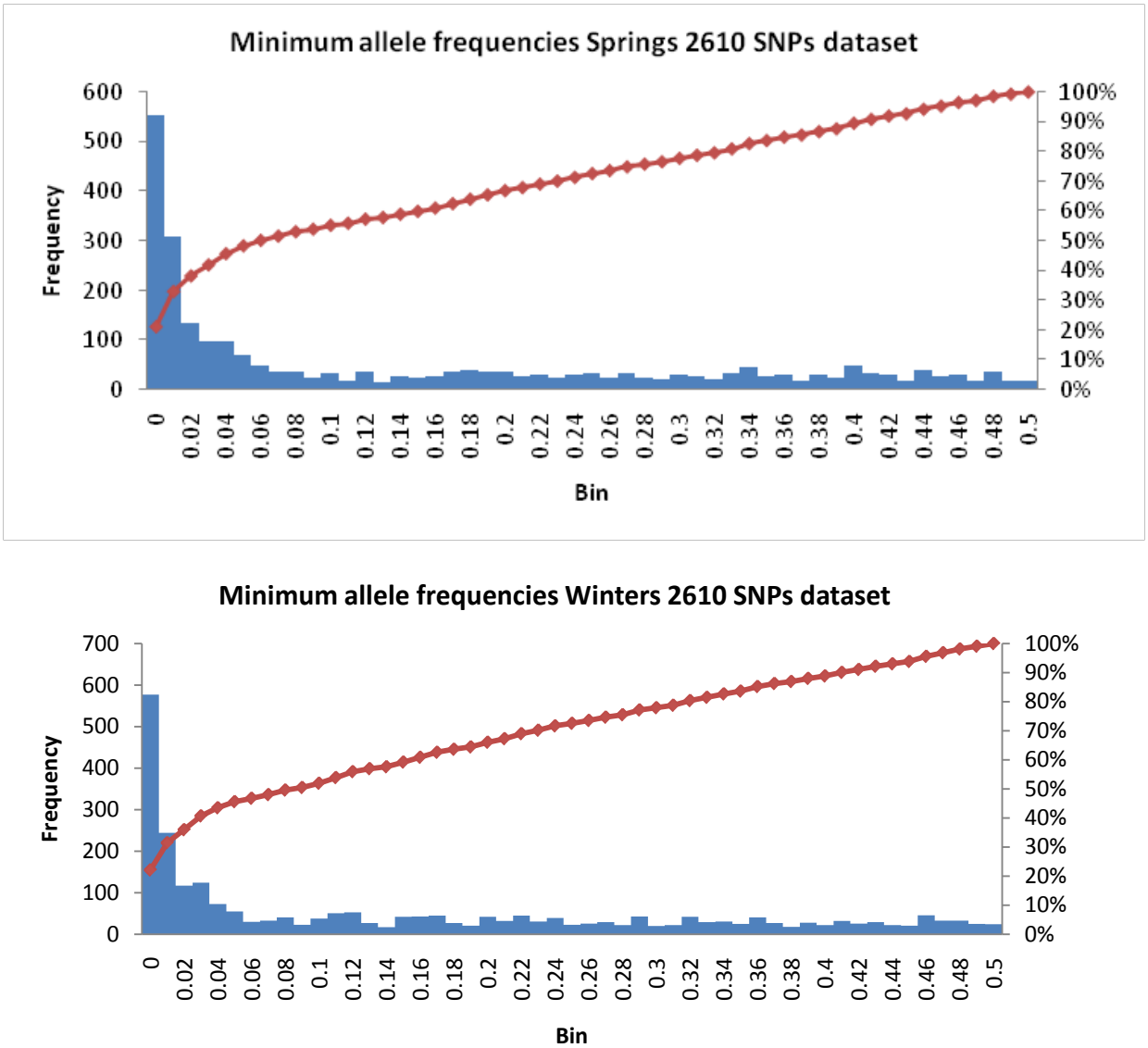


Figure 6. Number of markers in each of 51 bins of increasing Minor Allele Frequency (blue bars) with the cumulative total of markers plotted as red dots.

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 (Figure 6a). 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(Figure 6b). 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.

Footprints of Divergent Selection

To further investigate the importance of fixed / nearly fixed alleles in the population we first extended the 523 lines present in the public dataset by including 205 barley lines genotyped within AGOUEB as disease resistance progenitors to produce a dataset of 728 lines. This introduced lines with significant winter and spring admixture as well as increasing the number of 6 rowed lines so that we now had a total of 2780 SNPs with MAF >5%. We then performed a more detailed population structure analysis to assign lines to the ’spring’ and ‘winter’ growth habit classes but excluded 6 row types and lines with excess admixture of winter / spring genotypes arising from

recent inter gene-pool crossing. Lastly, all the SNP markers were screened locus-by-locus for signatures of divergent selection.

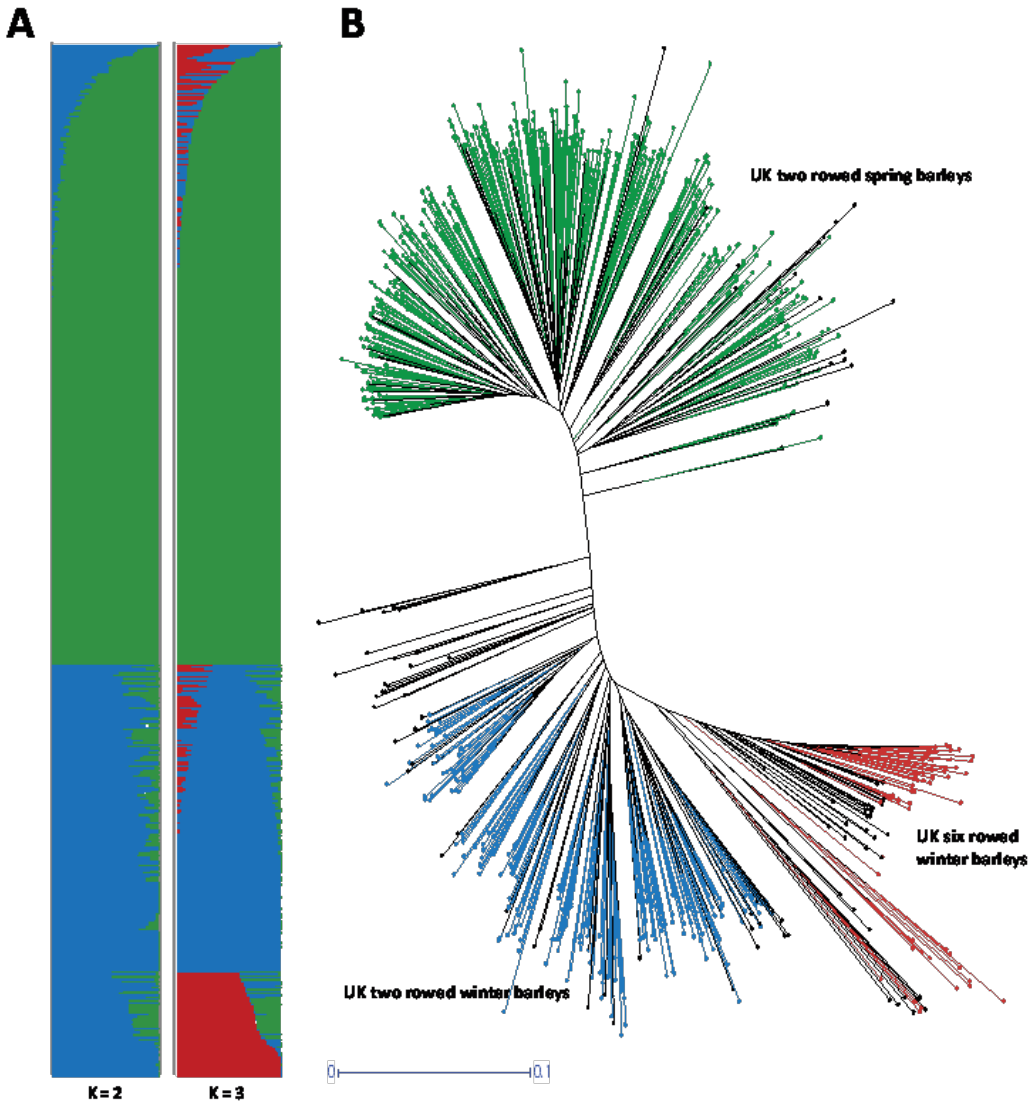


Figure 7. Graphical view of cluster analysis by STRUCTURE of 2780 SNPs used to genotype 728 barley lines with 2 groups (LHS). Green and Blue bar colouring indicates the relative probability of a genotype belonging to the spring and winter and winter grouping respectively. On the RHS, the analysis is repeated with 3 groups, where the winter is sub-divided into 2 and 6 row types with the red, blue and green bars indicating the relative probability of belonging to the 6 row winter, 2 row spring, and 2 row spring clusters respectively. The overall diversity is depicted in the Neighbour Joining tree shown in Figure 8.

We used the Bayesian approach implemented in the program STRUCTURE (Pritchard et al. 2000a; Pritchard et al. 2000b) (K = 2-3, burn-in period = 10,000, Mharkov Chain Monte Carlo iterations = 10,000, where K=the number of discrete classes) to obtain fractional memberships of the 728 barley genotypes to varying numbers of K subpopulations (Figure 7). Specifically, at K = 2, STRUCTURE separated the two main seasonal growth habit classes ‘spring’ and ‘winter’ types. Setting the STRUCTURE individual cluster assignment (K) to 3 matches the barley biotypes present within UK barley germplasm: two rowed spring, two rowed winter, and six rowed winter

barley types. We measured population differentiation by using the Fst estimator implemented in the STRUCTURE software. Both two rowed spring and winter sub-populations exhibited fixation index, Fst-value, of 0.51 and 0.52 consistent with estimates observed in a previous study with 105 and 51 two rowed springs and winters respectively (Comadran et al., 2009). The six rowed winter cluster exhibited a lower Fst-value of 0.36, indicative of recent gene flow with the other two sub-populations.

We also evaluated the genetic relationships among the accessions by generating a neighbour-joining population tree based on simple matching of allelic distances as implemented in DARwin (Perrier & Jacquemoud-Collet 2006). This analysis supported the same groupings as the Bayesian cluster analysis with two clearly separated branches corresponding to spring and winter types and a further subdivision of winter types related to ear row-number (Figure 8).

There were 44, 230 and 255 rare or nearly fixed SNP markers (minimum allele frequencies <0.05) that were exclusive to the 2 rowed winter, 2 rowed spring and 6 rowed winter germplasm pools respectively. Diversity summary statistics for the 2780 SNPs amongst the three groupings are summarized in Figure 9. The higher number of rare alleles present in the 6 rowed winter gene pool may be partly due to low sample size, and the higher number of rare alleles present in the spring pool relative to the 2 rowed winter pool may be due to the stronger breeding pressures experienced by the former for malting quality traits.

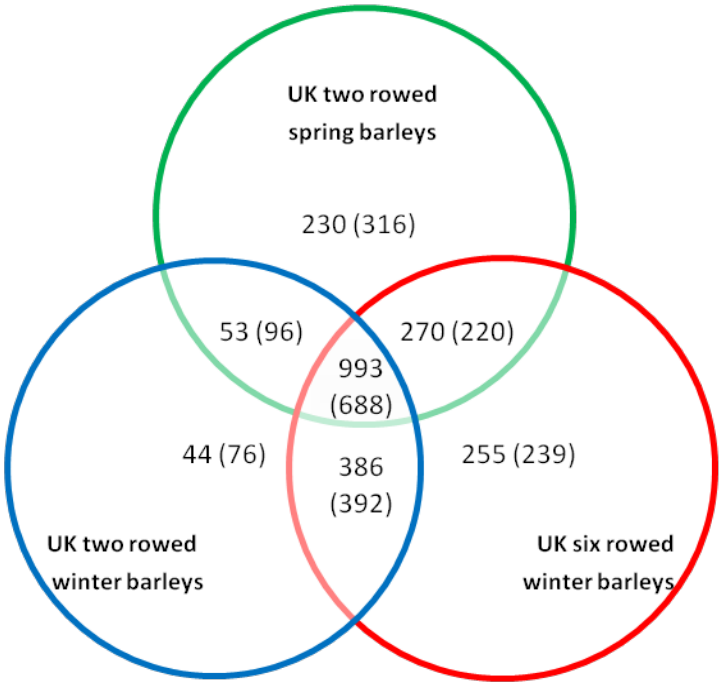


Figure 9. Venn diagram showing polymorphic marker overlap for the major population clusters present in our sample of 728 barley lines genotyped with 2780 SNP markers. The dataset is partitioned into European 2 rowed spring varieties, 2 rowed winter varieties and 6 row winter varieties. A three-way Venn diagram shows SNP markers with minimum allele frequencies (MAF) thresholds of 0.05 and 0.1 (in brackets) respectively.

For a clear picture of spring and winter historical divergence, a threshold of 75 % of estimated ancestry belonging to one of the sub-populations (K = 3) was used to filter the dataset and remove 'recent' cases of inter-population admixture as these cultivars may obscure ancestral sub-population divergence signatures. Exotic cultivars present in our sample (such as Cyrrhus, Corvette, Chevallier Tystofte or Bordeaux KVL 196) used by the barley research community in the past were removed from the sample as they were clearly distinct from the main body of lines and no longer relevant. This strategy did, however, exclude the well known winter malting variety Maris Otter as it was found to have a very high level of admixture with the spring germplasm, as would be expected from its pedigree.

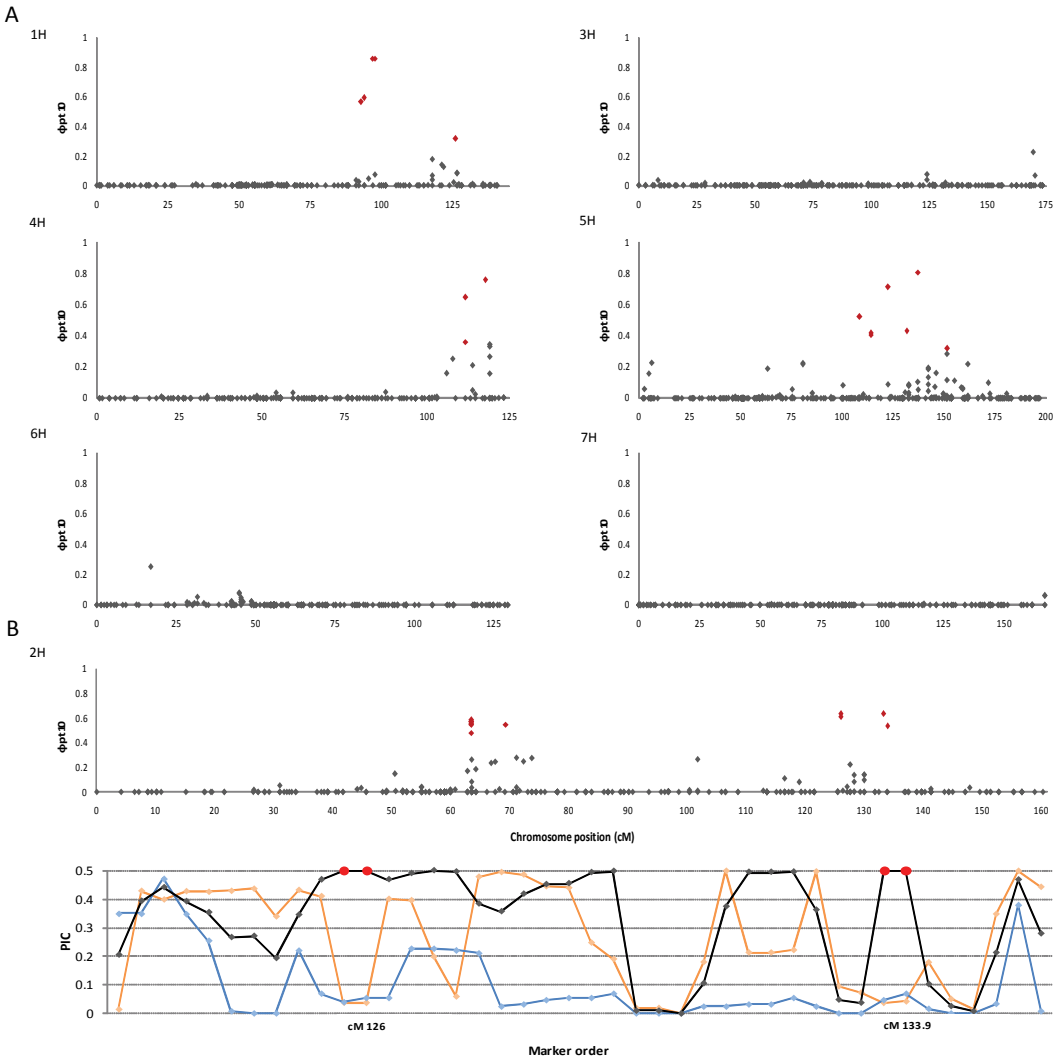


Figure 10. Genome-wide scan for footprints of divergent selection in 681 two-row spring and winter AGOUEB lines genotyped with 2780 SNPs. (A) Chromosomal scans of Φpt10, a measure of population genetic differentiation that maximizes putative traces of divergent selection. Φpt10 values ≥ 0.90 were considered to represent significant divergent selection and are shown in red. (B) Plot of PIC values for each SNP marker in a section of the long arm of chromosome 2H under strong divergent selection. PIC values in black were calculated for the whole data set, in orange for the two row springs and in blue for the two row winters.

The Φpt statistic is a measure of population genetic differentiation that is analogous to Fst and can be obtained via Analysis of Molecular Variance (AMOVA), as implemented in GenAlEX (Peakall & Smouse 2006). We used Φpt to screen the genome for loci displaying higher genetic differentiation than would be expected (high outlier) from neutral evolution from a pool of spring and a pool of winter barleys. The UK six row winter cultivars were not included in the genome scan as their historical relationship with UK two row winter cultivars is unclear. The analysis revealed 28 SNP showing strong evidence for adaptative divergent selection (Φpt values higher 0.9, Figure 10). The 28 SNP loci correspond to 15 genomic regions at least 3 cM apart, and because there are intermediate SNP markers without signs of divergent selection between them we can presume they might refer to different loci (Figure 10).

The analysis revealed 28 SNPs with Φpt10 values higher than 0.9, which indicates fixation of alternate alleles in the winter and spring two row gene pools and thus provides strong evidence for adaptative divergent selection. The 28 SNP loci (summarised in Table 1) correspond to 15 genomic regions, each of which is separated by at least 3 cM with intermediate SNP markers that have low Φpt10 values, so we have assumed that they probably represent selection for different loci.

As expected, signatures of adaptative selection were detected in the vicinity of known genes affecting cold tolerance, vernalization and photoperiod requirement. For the second segment on chromosome 1H, we used the syntenic relationship of barley with *Brachypodium distachyon* and *Oryza sativa* to find that SNP 12_31319 is closely linked to the photoperiod gene Ppd-H2, a candidate for HvFT3 which is a homologue for Flowering Locus T (Kikuchi et al. 2009). We would expect the major vernalisation loci Vrn1 and Vrn2 to be under divergent selection and have indeed identified SNPs in the region of HvVRN2 (11_21210) and HvVRN1 (11_11448) that are virtually fixed for alternative alleles in the spring and winter gene pools. Cold binding factors (cbf) play a major role in the low temperature tolerance of barley and other SNP markers have been developed to target key cbf loci at the frost resistance locus Fr-H2 (Francia et al. 2004; Francia et al. 2007) on chromosome 5H. Some of these SNP markers were included on BOPA2 and we found that SNP 12_30850, a marker for cbf14, was identified by our analysis as being subject to divergent selection in the spring and winter gene pools. We also detected divergent selection in the regions of the large effect flowering time quantitative trait loci (QTL) eam6 and Flowering time-2L (Flt-2L) (Chen et al. 2009a; Chen et al. 2009b). We inferred the position of eam6 on the SNP consensus map from BOPA1 data for the North American barley cultivar Bowman and its isogenic lines BW507 and BW508, which have introgressed segments for the mutants mat-b and mat-c, which correspond to eam6 (Druka et al. 2011). The mutants result in an early heading phenotype (Gustafsson et al. 1960), are photo- and thermo-period sensitive and are classified as long-day

types (Gustafsson & Lundqvist 1976). Again, we inferred the position of Flt-2L on the SNP consensus map from the fine mapping of the region reported in the literature (Chen et al. 2009a). Flt-2L was originally reported as a flowering time QTL closely linked to a frost tolerance QTL (Chen et al. 2009b) and we identified two linked signatures of selection co-segregating with the Flt-2L locus (Figure 10B), which may reflect this linkage.

Table 1. List of 28 SNP markers with signals of divergent selection (Φpt10 ≥ 0.9) found in 15 separate genomic regions (alternate clear and shaded sections) together with their chromosomal positions and possible candidate genes in the region.

Φpt	**snp_name**	**chromosome**	**position (cM)**	**candidate gene**
0.94	11_20475	1H	92.8	
0.95	11_10433	1H	93.9	
0.99	11_10396	1H	96.9	*Ppd-H2**
0.99	12_31319	1H	97.7	
0.9	11_21140	1H	126	
0.95	11_10692	2H	63.5	*eam6***
0.95	11_10191	2H	63.5	
0.94	11_21399	2H	63.5	
0.94	12_30265	2H	63.5	
0.94	12_30323	2H	63.5	
0.93	11_20438	2H	63.5	
0.94	11_21144	2H	69.3	
0.96	11_20480	2H	126	*Flt-2L***
0.95	11_21440	2H	126	
0.96	11_11227	2H	133.2	*Flt-2L***
0.94	12_30106	2H	133.9	
0.96	11_11299	4H	111.7	
0.9	11_10150	4H	111.7	
0.97	11_21210	4H	117.6	*sgh1 (HvVrn2)*
0.9	11_20089	4H	119.1	
0.9	11_20013	4H	119.1	
0.94	12_30850	5H	108.2	(*cbf14*) *cbf* gene cluster
0.92	11_11341	5H	113.8	
0.91	12_30619	5H	113.8	
0.97	11_10094	5H	122.4	
0.92	12_30067	5H	131.6	
0.98	11_11448	5H	137	*Sgh2 (HvVrn1)*
0.9	12_31050	5H	151.4	

* Causal gene identified in the literature but not functionally validated in barley

** Fine-mapped in the literature but causal gene not identified yet

Germplasm Relationships in the Public and Breeders Private Sets

Whilst the breeders private lines remain confidential to the submitting breeder, it was important to check if there were any strong groupings associated with any of the breeders that might render the extension of the association genetics analyses developed within the project problematic. As with the public set, we used Genstat 13 to apply principal coordinate analysis (PCO) based on simple matching of the BOPA1 genotypic data for all the public lines plus the breeders private lines to study whether or not there was any significant clustering of the lines related to the individual breeders.

The clustering of the lines is very similar to that observed within the public set, partitioning the dataset as expected, into two-rowed spring barley varieties, two-rowed winter barley varieties and six-rowed winter barley varieties. As before, PCO separated the winter (low value) from the spring (high value) types and PCO2 separated six row (high value) and two row (low value) winter types (Table 2). The composition of the breeders lines varied considerably between individual breeders (Figure 11 upper triangle) with the 'brown' breeder using a number of lines with potential admixture between six row winter and two row spring groups.

Table 2. Percentage of the genetic variation accounted by the first 4 PCOs in analysis of 950 public and private lines genotyped with 1432 BOPA1 SNPs with <10% missing values.

Percentage variation			
PCO1	*PCO2*	*PCO3*	*PCO4*
20.56	5.97	3.66	3.01

When all the breeders private sets are considered together, it is clear that virtually all their lines are concentrated within the major three clusters detected by the analysis of the Public Set, irrespective of which two of the first four PCOs are plotted together (Figure 11, lower triangle). The conclusion from this is that breeders are still making progress by concentrating their efforts within the current well-established elite gene pools, although this would depend upon the genotypes used by individual breeders to form their own private set.

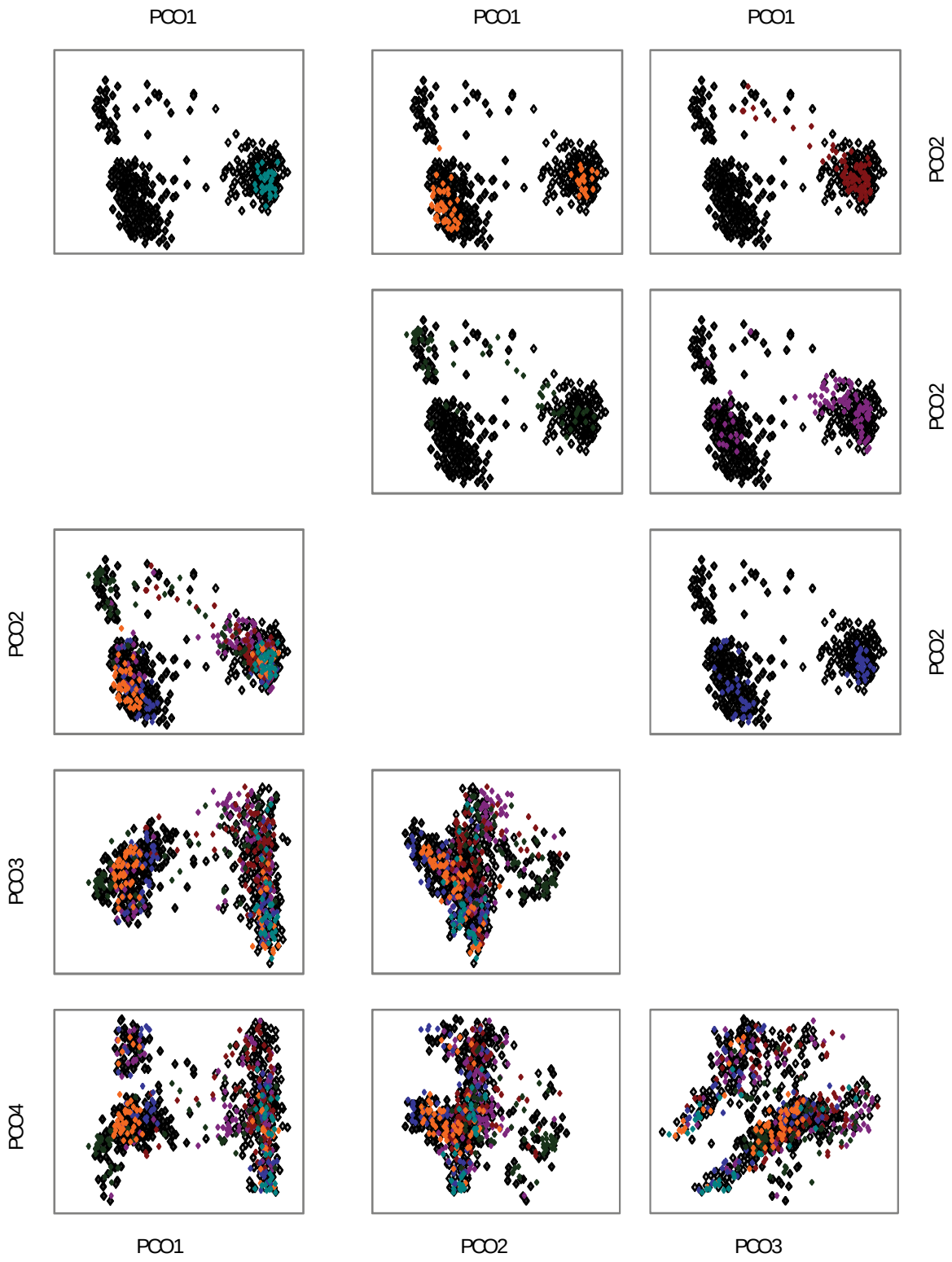


Figure 11. PCO analysis of relationships between AGOUEB Public Set and breeders private lines. Each of 6 breeders private sets was coloured and plotted individually for the first two principal co-ordinates on top of the Public set, where spring lines are coloured grey and winter lines black (upper diagonal). The lower diagonal shows all the breeders lines coloured as in the upper diagonal but plotted together with all the Public set for two of each of the first four principal co-ordinates.

3.3. Phenotyping

3.3.1. 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 two years during the awarding of National Listing (UK) and Plant Breeders' Rights (EU), during which submissions are assigned a unique identifier (AFP number). 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 (Figure 12). 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 (listed in Table 3). 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/).

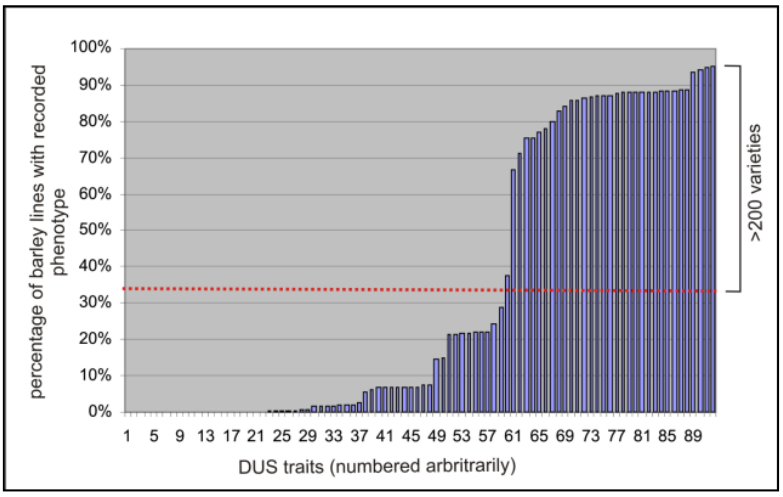


Figure 12. Histogram showing the percentage of varieties with recorded data associated with each of the 94 DUS characters recorded. Thirty-three DUS traits had a fill of ≥ 33 % (≥ 200 barley lines).

Table 3. Traits for which phenotypic data is available for ≥ 200 cultivars within the association panel. The recorded character states and ranges are indicated. *Phenotypes for which significant associations were identified by GWA analysis ($-\log_{10} p \geq 4.35$, ≥ 2 significant markers within a 4 cM window).

Trait (number of varieties with data)	Character states	Character description
Aleurone colour (526) *	1,2,3	none to strong
Auricle anthocyanin coloration (469) *	1,9	present or absent
Auricle anthocyanin intensity (459) *	1,2,3,4,5,6,7,8,9	absent to very strong
Awn anthocyanin coloration (504) *	1,9	present or absent
Awn anthocyanin intensity (459) *	1,2,3,4,5,6,7,8,9	absent to very strong
Awn length cf. glume (537)	3,4,5,6,7	short to long
Awn margin spiculation (434)	1,5,9	absent, reduced, present
Collar type (230)	1,2,3,4,5,6,7	recurrent - platform - cup
Ear attitude (536) *	1,2,3,4,5,6,7,8,9	erect to recurved
Ear emergence (537)	1,2,3,4,5,6,7,8,9	very early to very late
Ear glaucosity (540)	1,2,3,4,5,6,7,8,9	absent to very strong
Ear grain density (536)	1,2,3,4,5,6,7,8,9	very lax to very dense
Ear length (522)	1,2,3,4,5,6,7,8,9	very short to very long
Ear row-number (577) *	1,2	2-row, 6-row
Ear shape (535)	3,4,5,6,7	tapering-parallel-fusiform
Flag leaf-sheath glaucosity (538)	1,2,3,4,5,6,7,8,9	absent to very strong
Grain furrow hair (573) *	1,5,9	absent, sharkskin, present
Grain husk (535)	1,9	absent or present
Grain lateral nerve spiculation (541) *	1,2,3,4,5,6,7,8,9	absent to very strong
Grain rachilla hair type (578) *	1,2	short or long
Growth Habit (533)	1,2,3,4,5,6,7,8,9	erect to prostrate
Hairiness of leaf sheath (569) *	1,9	absent or present
Lemma nerve antho intensity (529) *	1,2,3,4,5,6,7,8,9	absent to very strong
Lodicule disposition	1,2	Frontal; clasping
Plant height (534)	1,2,3,4,5,6,7,8,9	very short to very tall
Rachis first segment curvature (531)	1,2,3,4,5,6,7,8,9	absent to very strong
Rachis first segment length (534)	3,4,5,6,7	short to long
Recurved leaf frequency (512)	1,2,3,4,5,6,7,8,9	absent to very high
Seasonal growth habit (530)*	1,2,3	spring, alternative, winter
Spikelet glume and awn length (528)	3,5,7	short to long
Sterile spikelet attitude (475) *	1,2,3	parallel to divergent
Sterile spikelet development (486) *	1,2	none or full
Sterile spikelet tip shape (406)	1,2,3	pointed, rounded, squared

3.3.2. Historical VCU data

Description of sites

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 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.

There were 936 unique site names in the official barley trial set yet inspection revealed that many of these were effectively the same site. We therefore condensed these sites to a set of 104 unique trial sites located in the same geographic region and assigned alphanumeric codes to them based upon the system applied to NL, RL and BSPB trials in 1988. Nine of these unique sites were used in each of the 19 years surveyed and a further 40 were used in over half the years. Figure 13 shows the distribution of the more frequently used sites.

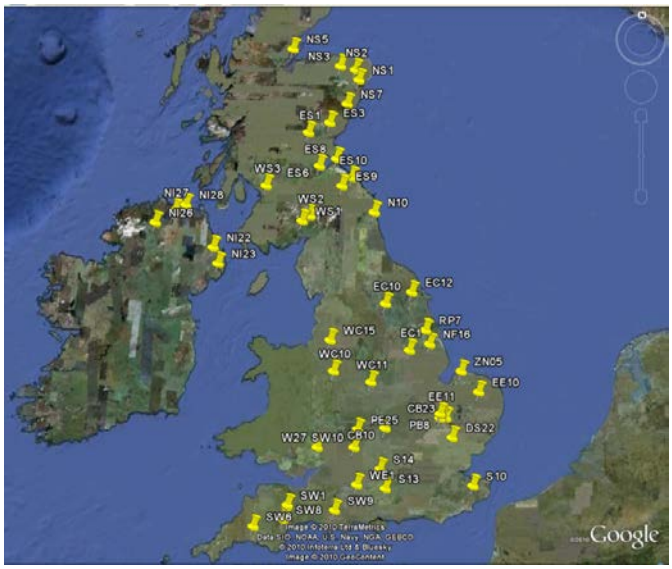


Figure 13. Distribution of the most frequently used sites across Great Britain and Northern Ireland.

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.

Variates scored

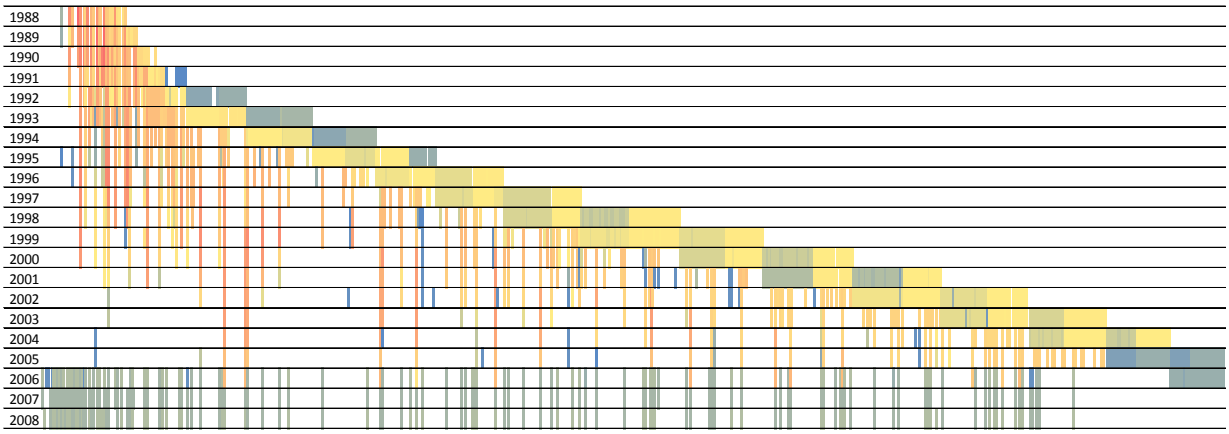


Figure 14. Graphical representation of the coverage of yield data for 630 lines in trials from 1988 to 2008 that we included in our analysis. Lines are ordered by AFP (see section 3.3.1) from left to right with blue and dark colours indicating few trials (<10) in a year and red indicating many (>25). White space means the line was not in trial for that year.

Over the time period that we sampled, official barley trials could largely be classified into four series: NL1, NL2, BSPB and RL. 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 the majority were in trials for two or three years (Figure 14). 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 (Figure 14).

Components of variance

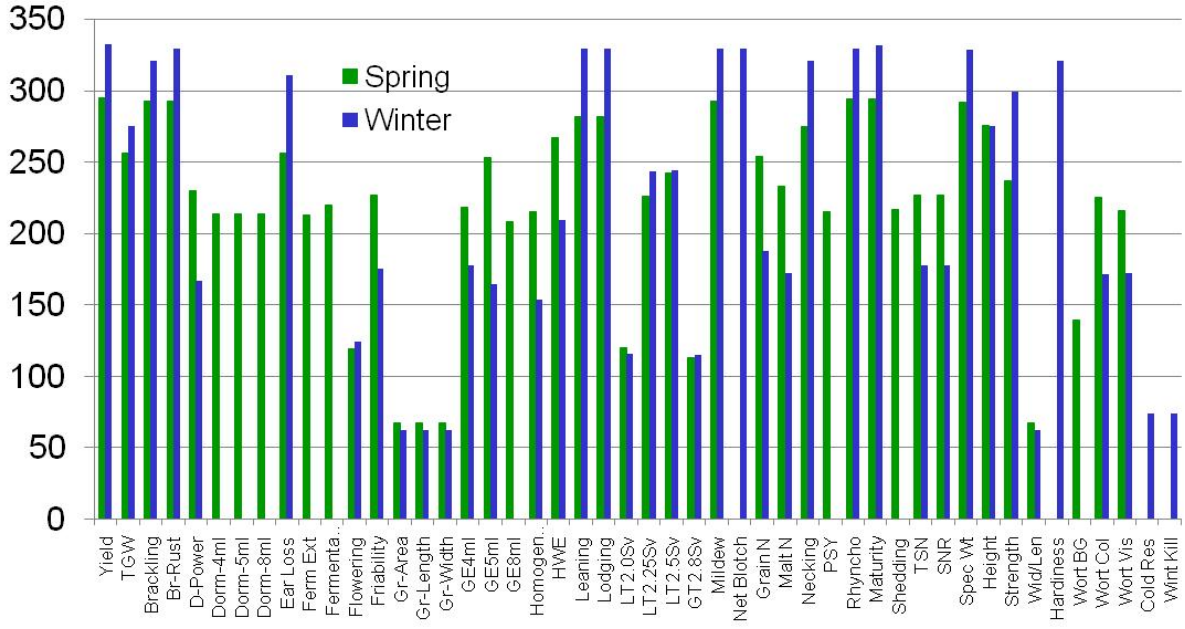


Figure 15. Numbers of spring and winter barley genotypes with BLUPs (predicted means) for each of 48 variates analysed from the combined set of phenotypic data.

In general, we restricted our analyses to the treated trials as these will provide the best estimate of the genetic potential of lines. We therefore used REML analysis to predict the means of the individual lines and the years over which the trials were grown. We fitted all terms as random effects with AFP number used to represent an individual line, Year, Site and Trial Series as main effects and Year x Site, AFP x Year, AFP x Site and Year x Site x Trial Series as interaction effects. The inclusion of extra interactions was not really feasible due to the sparse coverage for most of the variates and we considered those listed above to the most interesting. Some variates, e.g. those relating to grain dimensions, were only scored on the AGOUEB series so all terms relating to Trial Series were removed from the model. For the two cold hardiness variates, we could only fit main effects for genotype and year, with the genotype x year interaction being used as the error term. Some key progenitors did not have AFP numbers so we assigned them a negative number to distinguish them from those that had genuine numbers. This resulted in means for an average of 219 and 216 spring and winter barley lines respectively and the number of lines with means for each variate is shown in Figure 15.

As would be expected, there were considerable differences between years and sites for most of the variates that we analysed and their interactions were also generally high compared to genotypic effects. The differences between the different trial series were, however, rarely significant and generally negligible. Considering the differences between the genotypes, heritabilities were generally high with an average of 36% for spring and 45% for winter barley. For the springs, ear loss had the lowest heritability and grain area and grain length the highest, although the last two were just measured on the AGOUEB trial series and so the progenitors could have a greater influence. For characters measured over all the trials, height had the greatest heritability (Figure 16). GxE effects were generally small relative to genetic and error effects with the Genotype x Year term being 5% and the Genotype x Site slightly less at 4% on average. The combined interactions were noticeably larger than the main genetic effect for all the characters relating to straw strength (Leaning, Lodging and Strength) with the site interaction being slightly larger than the year. It was also noticeable that the interactions were much greater in the germination measures soon after harvest (‘Dorm’ variates) than the measures after appropriate storage (‘GE’ variates).

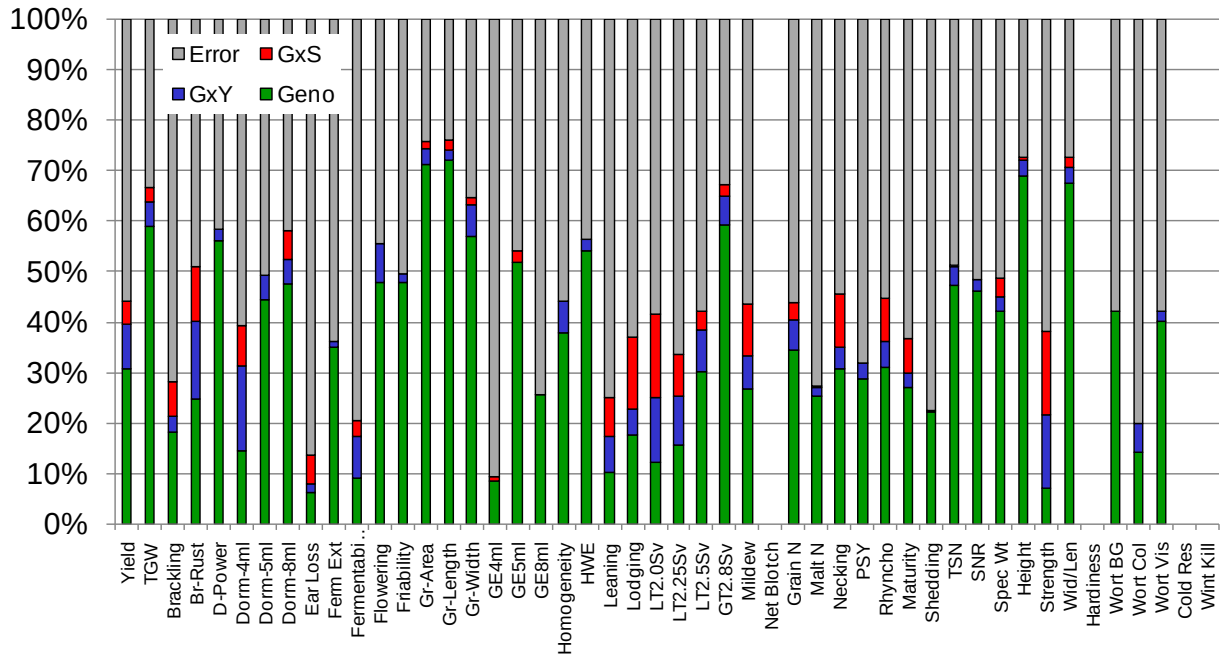


Figure 16. Estimated proportion of Genotype (Geno), Genotype x Year (GxY), Genotype x Site (GxS) and Error variance components for 44 variates analysed from the combined data set for spring barley. The proportion of the Genotypic variation (Green bar) is an estimate of the heritability of a character.

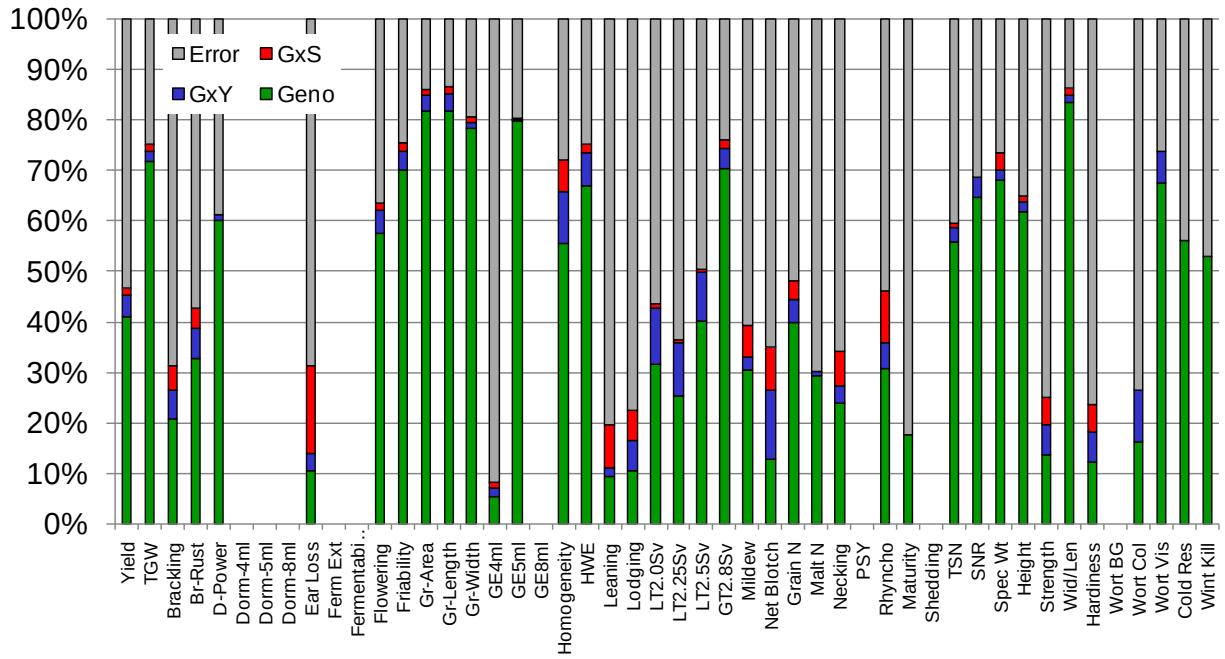


Figure 17. Estimated proportion of Genotype (Geno), Genotype x Year (GxY), Genotype x Site (GxS) and Error variance components for 39 variates analysed from the combined data set for winter barley. The proportion of the Genotypic variation (Green bar) is an estimate of the heritability of a character.

For the winter barleys, the lowest heritability was found for germinative energy in 4ml of water, although ear loss was again low. Like the spring barleys, very high heritabilities were found for the grain dimension measurement characters and these might also be biased by the greater influence

of the progenitors in variates measured in just one trial series over three years. Thousand grain weight had the highest heritability of characters measured over all trial series. GxE effects showed a similar pattern in the winter barley trials as in the spring (Figure 17).

The summary statistics for the Best Linear Unbiased Predictors of the scores for the spring genotypes for the 44 variates analysed are shown in Table 4 where the overall mean plus the range and the two extreme genotypes plus the average standard error of a difference (SED) are presented.

The highest yielding line, Quench, was over 0.8 t/ha greater than the overall mean but the lowest yielding variety was over 1.5 t/ha less. There were also considerable differences in the TGW BLUPs with the highest scoring line being 8g heavier than the mean and the lowest nearly 10g less. Most of the extremes were as would be expected. Tyne is an ari-e.GP semi-dwarf type and the gene is known to produce small grain so it would be expected to score low for the grain dimension characters, whilst Hart was noted for its big bold grain and would be expected to score high. Similarly, Tankard tending to produce rounder grain and thus would have a high grain width to length ratio. Digger was always regarded as having very hot water extract whereas Tartan gained a special recommendation as a malting variety and would have had a high value for this character. Large ranges were found for the disease resistance characters, which may in part reflect the differences in disease incidence in the years that lines were trialled that the analysis could not fully compensate for. For instance, Heron had the lowest incidence of powdery mildew, which would be consistent with its mlo resistance gene and Prisma had the highest incidence, which is consistent with it possessing ineffective resistance genes (see Section 3.3.4) but Golden Promise, whilst significantly greater than the mean, was relatively low scoring for a susceptible line at 6%). The semi-dwarf line Brazil was the shortest lines whereas the non-dwarf line Atem was over 40 cm taller. In terms of straw strength and lodging, the Triumph type semi-dwarves were significantly stronger than the non-dwarf types like Cindy in terms of straw strength but it was surprising that Golden Promise had the highest lodging score, despite possessing a semi-dwarf gene, albeit a different one.

Table 4. Minima, maxima, overall mean, SED and the extreme genotypes for each of 44 variates measured on the AGOUEB public set of spring barley.

	Minimum			Maximum		
Character	**Genotype**	**Value**	**Mean**	**Value**	**Genotype**	**SED**
Yield	Newgrange	5.39	6.99	7.80	Quench	0.091
TGW	Franklin	38.58	48.39	56.71	Velvet	1.154
Brackling	Rakaia	2.09	13.93	43.05	Ingmar	3.646
Br-Rust	Goldie	0.52	8.91	38.35	Attraction	3.194
D-Power	Georgie	65.53	99.10	136.50	Hanka	6.473
Dorm-4ml	Trinity	84.54	95.38	98.51	Ursa	1.696
Dorm-5ml	Neruda	43.37	84.59	97.60	Henni	3.788
Dorm-8ml	Alliot	17.68	41.13	77.94	Penthouse	7.308
Ear Loss	Onyx	7.81	8.51	8.97	Ardila	0.147
Ferm Ext	Henni	66.57	70.88	72.24	Rummy	0.373
Fermentability	Hart	77.48	86.97	87.99	Flick Sejet	0.378
Flowering	Tartan	77.87	86.75	97.74	SW SCANIA	0.592
Friability	Klaxon	65.54	88.76	96.42	Sacha	1.913
Gr-Area	Tyne	21.23	26.10	28.94	Hart	0.130
Gr-Length	Tyne	7.69	8.55	9.19	Hart	0.034
Gr-Width	Tyne	3.53	3.95	4.11	Tocada	0.013
GE4ml	Spotlight	93.79	97.73	99.06	Spey	0.445
GE5ml	Alabama	72.47	94.32	99.60	Ricarda	1.220
GE8ml	Blenheim	38.39	67.35	85.03	Ferment	5.282
Homogeneity	Klaxon	81.23	97.33	99.62	Pitcher	0.988
HWE	Digger	291.39	312.11	316.94	Tartan	0.994
Leaning	Nimbus	1.55	7.96	39.15	Pyla	2.672
Lodging	Mentor	-0.07	5.52	36.07	G. Promise	2.776
LT2.0Sv	Maypole	0.38	1.28	9.04	Crusader	0.309
LT2.25Sv	Patron	0.80	1.89	10.01	Digger	0.690
LT2.5Sv	Akita	2.99	5.77	25.83	Digger	1.583
GT2.8Sv	Livet	31.80	75.62	87.23	Quartet	2.693
Mildew	Heron	0.18	2.58	25.03	Prisma	1.605
Grain N	Talbot	1.42	1.60	1.95	Jersey	0.024
Malt N	Mandolin	1.26	1.51	1.79	Tyne	0.030
Necking	Hydra	1.89	12.89	46.68	Carlota	3.743
PSY	Alexis	398.51	427.30	435.00	Rummy	3.308
Rhyncho	Century	0.72	7.09	23.68	W 8257	1.802
Maturity	Spectre	134.36	148.35	169.54	Franklin	2.023
Shedding	Chad	7.47	8.80	8.98	SW 2808	0.125
TSN	Otira	0.46	0.60	0.76	Goldie	0.018
SNR	Hart	30.81	40.33	45.85	Decanter	1.109
Spec Wt	Otira	63.60	68.80	76.04	Penthouse	0.402
Height	Brazil	52.68	71.64	93.59	Atem	1.838
Strength	Madras	0.97	6.78	40.03	Cindy	3.732
Wid/Len	Cork	0.43	0.46	0.49	Tankard	0.002
Wort B-Glucan	Toby	59.99	130.56	407.64	Klaxon	27.291
Wort Col	Tyne	2.07	2.84	4.36	Georgie	0.161
Wort Vis	Amber	1.43	1.51	1.96	Klaxon	0.020

The same summaries are presented for the winter barley lines in Table 5, where the highest yielding line (Pelican) was over 1.8 t/ha greater than the mean whereas the lowest (Hiberna) was over 1.9 t/ha less. Pelican was a six row cultivar first recommended for growing in 2007 as a high yielding cultivar whereas Hiberna was a hulless two rowed variety and would be expected to be lower yielding. Considering thousand grain weight, the heaviest line (Cypress) was nearly 13g greater than the mean whereas the lowest (Chess) was over 14g less. The high value for Cypress is as expected but Chess was a two rowed line and would have been expected to have had a higher thousand grain weight so must have been adversely affected by some abiotic or biotic stress over the two seasons that it was grown in trial. Whilst there were considerable differences between the maxima and minima for the disease resistance characters, the ranges were noticeably less than for the equivalent characters in the spring genotypes. As for the springs, there was over 40cm difference between the shortest and tallest line, despite the winter lines not being known to possess any of the two major dwarfing genes found in the springs. The weakest lines in terms of lodging and strength were taller two rowed types, whereas the strongest lines were both six row types.

Table 5. Minima, maxima, overall mean, SED and the extreme genotypes for each of 44 variates measured on the AGOUEB public set of winter barley.

	Minimum			Maximum		
Character	**Genotype**	**Value**	**Mean**	**Value**	**Genotype**	**SED**
Yield	Hiberna	6.02	7.98	9.84	Pelican	0.126
TGW	Chess	32.47	46.86	59.48	Cypress	1.202
Brackling	Tudor	2.40	22.33	58.59	Azurel	3.937
Br-Rust	SW Sienna	0.64	4.91	18.80	Angora	1.488
D-Power	Peridot	67.83	96.84	151.78	Plaisant	7.506
Ear Loss	Posaune	4.49	8.25	8.93	Portrait	0.168
Flowering	NORD 96601/6	214.92	223.52	235.24	Blythe	0.670
Friability	Igri	38.01	71.32	92.88	Portrait	3.053
Gr-Area	Esterel	23.08	26.01	29.03	Intro	0.170
Gr-Length	Maris Otter	8.00	8.74	9.80	Siberia	0.045
Gr-Width	Esterel	3.52	3.84	4.10	Haka	0.017
GE4ml	Gypsy	94.29	96.87	98.05	CPBT B66	1.026
GE5ml	Oleron	65.57	95.29	101.69	Jewel	1.531
Homogeneity	Igri	58.09	90.17	98.85	Flagon	1.813
HWE	Majestic	268.45	299.26	309.46	Tiffany	1.742
Leaning	Kite	-0.24	8.82	22.56	Lomerit	2.673
Lodging	Courtois	0.45	5.87	24.21	Symphony	2.764
LT2.0Sv	Artist	0.32	1.67	7.60	Silverstone	0.339
LT2.25Sv	Opal	0.87	2.71	8.60	Pict	0.945
LT2.5Sv	Opal	2.65	9.20	28.00	Ayana	2.359
GT2.8Sv	Tokyo	27.97	63.01	93.93	Opal	3.331
Mildew	Blythe	0.36	4.52	28.74	Vanilla	1.290
Net Blotch	Willow	0.43	3.33	13.34	Anvil	1.513
Grain N	Courtois	1.55	1.74	2.20	Malta	0.043
Malt N	Jessica	1.46	1.71	2.07	Esterel	0.049
Necking	Hurricane	2.17	14.22	44.22	Saigon	3.656
Rhyncho	NORD 96601/6	0.02	4.78	19.90	Gazelle	1.114
Maturity	Esterel	274.78	290.42	303.55	Milano	0.618
TSN	Antonia	0.44	0.56	0.71	Spirit	0.023
SNR	Tallica	24.45	33.65	40.31	Jessica	1.210
Spec Wt	Oleron	60.47	68.70	77.40	Godiva	0.450
Height	Digby	81.86	97.58	128.05	Jerez	3.337
Strength	Kite	-0.03	8.85	32.54	Halcyon	3.489
Wid/Len	Siberia	0.37	0.44	0.48	Opal	0.002
Hardiness	Arctic	6.28	7.64	8.54	Cleopatra	0.180
Wort Col	Firefly	2.11	2.62	3.23	Vertige	0.146
Wort Vis	Chestnut	1.44	1.70	2.64	Igri	0.046
Cold Res	Manitou	2.09	4.94	6.98	Angora	0.621
Wint Kill	Scylla	1.97	3.76	9.98	Gaelic	0.742

3.3.3. AGOUEB trials

Composition

As noted above, the entries in the AGOUEB spring and winter barley trials reflected lines that represented recommended varieties that were commercially successful or were not taken up by the market in terms of seed sales over the period from 1990 to 2005. We added some previous market successes and key progenitors to this list together with a selection of lines that failed to progress beyond the Recommended List Candidate stage. An initial multiplication of seed was carried out at SCRI prior to the first sowing and the seed harvested from the multiplications was used to sow all the trials for harvest year 2006 to eliminate possible confounding effects of seed source with genotype. Seed harvested from the 2006 SCRI trial was used to re-sow trial and multiplication plots at SCRI only for harvest 2007. The 2007 multiplication plots were rogued and used to sow the trials for harvest 2008. The sites used were the main trial sites for KWS (UK), Limagrain, LS Plant Breeding, NIAB, SCRI, Secobra (UK) and Syngenta Seeds for spring and winter barley with an additional site at Lantmannen SW Seed for spring barley. These trials were grown according to the VCU fungicide treated protocol for each of the years in which they were grown. Additionally, small plots were sown without fungicides in disease nurseries at the trial sites for harvest years 2007 and 2008.

Variates scored

During the growing season, the plots were scored for the following variates:

1. Heading - date when 50% plot at DGS53 (ear 25% emerged)
2. Height to the tip of the ear (cm)
3. Leaning - % plot deviating from vertical by <45°
4. Lodging - % plot deviating from vertical by >45°
5. Brackling - % plot with broken stems
6. Necking - % plot with kinked peduncles
7. Maturity – date when 50% plot at DGS91 (hard grain)

When ripe, the plots were harvested by a small plot combine, weighed and moisture content assessed to express the plot weight at 15% moisture content. This was then divided by the plot area, measured as harvested plot length multiplied by the plot width plus half the inter-plot gap on either side, to express the plot yields in t ha^{-1} (variate 9). Harvested seed was retained from each plot and sent to SCRI where a 1.2 kg sub-sample was cleaned and graded. The fractions passing over and through a 2.5mm slot sieve were weighed separately and used to estimate screenings (%) of each plot (variate 10). Sub-samples of the cleaned and graded samples were then used for the following grain quality tests;

8. Grain Length (mm) by Marvin digital analysis of ~200 seeds
9. Grain Width (mm) by Marvin digital analysis of ~200 seeds

10. Grain Area (mm2) by Marvin digital analysis of ~200 seeds
11. Thousand Grain Weight (g) from weight of seed / number counted by Marvin
12. Width:Length (ratio) derived from 11 and 12
13. Grain nitrogen content, measured by Perten DA7200
14. Grain hardness, measured by Perten Single Kernel Characterisation System (SKCS)
15. Grain diameter (mm), measured by Perten SKCS
16. Thousand Grain Weight (g), measured by Perten SKCS

The untreated disease nursery plots were scored during the growing season for the amount of infection by the following diseases, using a 1-9 scale where 1=resistant:

17. Powdery Mildew (Blumeria graminis f.sp. hordei)
18. Leaf Scald (Rhynchosporium secalis
19. Brown (Leaf) Rust (Puccinia hordei)
20. Net Blotch (Pyrenophora teres)

Micro-malting analyses of the cleaned and graded seed from selected plots was carried out by six member companies of the Maltsters Association of Great Britain; Boortmalt (formerly Greencore), Crisp Malting Group, Diageo Scotland, Molson Coors Brewing Company (UK), Muntons, and Simpsons Malt. Each maltster micro-malted 16 samples from each of the harvest years from 2006-8. For 2006, the same 16 spring and 8 winter varieties were selected from 4 sites but for harvest years 2007 and 2008, the same 4 varieties from the 2006 set of 16 winters were again selected from 4 sites each year but a different set of 12 other varieties added. Similarly for the winters, 2 of the 8 from 2006 were retained but supplemented with a different set of 6 varieties from each of 4 sites each year. This strategy ensured that we had a set of common controls over sites and years with which we could estimate the genetic potential of all 64 spring barley and 56 of the winter barley varieties during the course of the project. Each maltster carried out their own malt analysis according to the schedule used for NL trials and one sub-sample of the malt was sent to BRi and another to SWRI for testing. The following variates were scored on the malt samples:

21. Germinative Energy (%) of 100 seeds in 4ml water
22. Germinative Energy (%) of 100 seeds in 8ml water
23. Wort Colour (EBC units)
24. Hot Water Extract Lintner° kg-1
25. Malt Nitrogen content (%)
26. Total Soluble Nitrogen (%)
27. Soluble Nitrogen Ratio (%)
28. Free Amino Nitrogen (mg l-1)
29. Diastatic Power °IOB
30. Alpha Amylase (DU)

31. Wort Viscosity (mPas)
32. Wort Beta glucan (mg l-1)
33. Friability (%)
34. Homogeneity (%)
35. Whole Corns (%)
36. Fermentability (%)
37. Fermentable Extract (%)
38. Predicted Spirit Yield (l t-1)
39. Vmax (g), the theoretical mass of filtrate that can be collected over infinite time.

Data collected for variates 1-23 were analysed for each trial to identify the most appropriate model to correct for spatial trends and this was then utilised in a mixed model analysis to generate means for each variety in each trial at each site. A meta-analysis was used to incorporate the different spatial effects for each trial in an over sites and years analysis to generate overall means for each variate. A simpler analysis was applied to the micro-malting derived variates as the structure only permitted the fitting of variety and year/site effects with the interaction of variety with year/site being the error term. All analyses were conducted using the REML directive in Genstat.

Breeding Progress

The recommended varieties used in the AGOUEB trials can be used to obtain an unbiased estimate of breeding progress as they were all grown alongside each other in the same trials over three years. We can therefore regress their overall mean for each of the variates that were scored on them against the year when they were first placed on the Recommended List. If there is a significant relationship of the variate with time, then breeding has clearly affected the variate under study and the slope of the regression line can be used to estimate the rate of breeding progress for that variate.

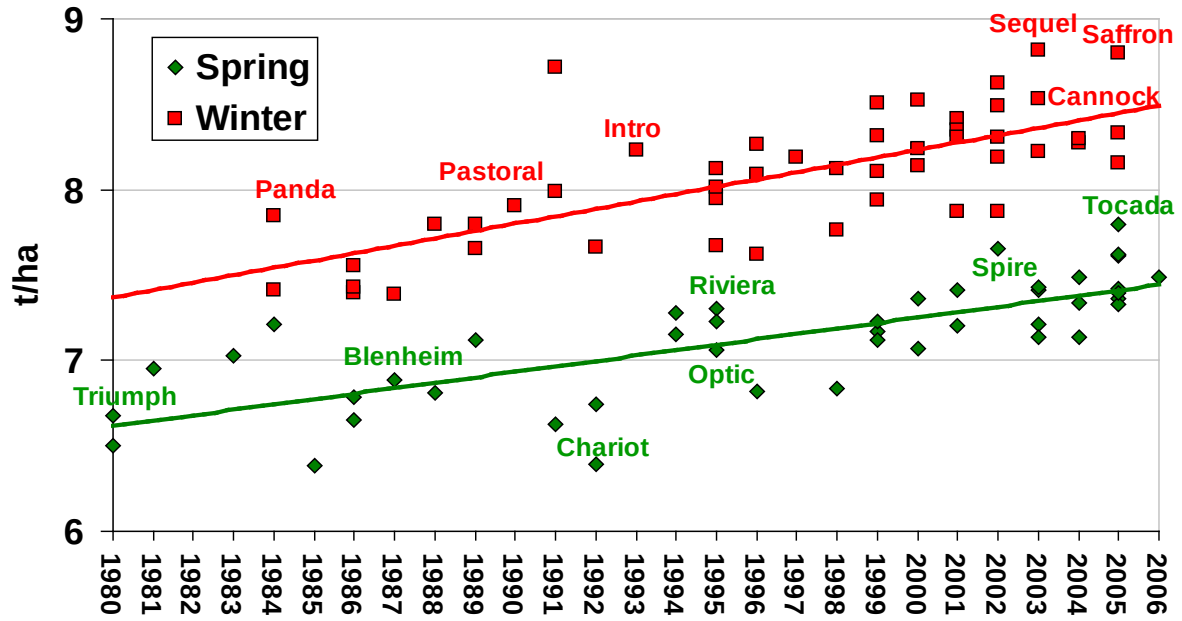


Figure 18. Regression of mean yield against year of first recommendation for all spring (green) and winter (red) recommended varieties that were included in the AGOUEB trials from 2006-8.

We included 44 spring barleys that had been recommended between 1980 and 2006 in the AGOUEB trials and there was a highly significant regression between their year of first recommendation and their mean yield, accounting for over 57% of the variation in the latter. This translates into an improvement of just over 0.8 t ha^{-1} over 26 years, which is approximately 0.5% per annum. Similarly, there was also a highly significant regression between year of first recommendation and mean yield for the 48 recommended winter barley varieties that had been included in the AGOUEB trials. The regression accounted for over 52% of the yield variation and was equivalent to a yield increase of just over 1.1 t ha^{-1} over the period, a rate of progress of 0.6% per annum (Figure 18).

The recommendation of Triumph in 1980 was a major breakthrough in the UK as it was a variety that combined high yield with the best malting quality, although there were some dormancy issues in Northern Britain under poor harvest conditions. All current UK recommended spring barley varieties have Triumph in their ancestry and it is instructive to express variety means for some key variates as a percentage of Triumph to detect breeding progress in other variates. There has been a significant increase in thousand grain weight since 1980 and also a significant decrease in height but no significant change in heading or maturity dates. Triumph possessed the *sdw1* semi-dwarf gene and all current recommended varieties also possess *sdw1* but we had included some of the non-dwarf recommended varieties in our survey set, which could bias the relationship with height as no non-dwarf types have been recommended recently. Even if the non-dwarves are excluded, there has still been a significant decrease in plant height since the introduction of Triumph (Figure 19).

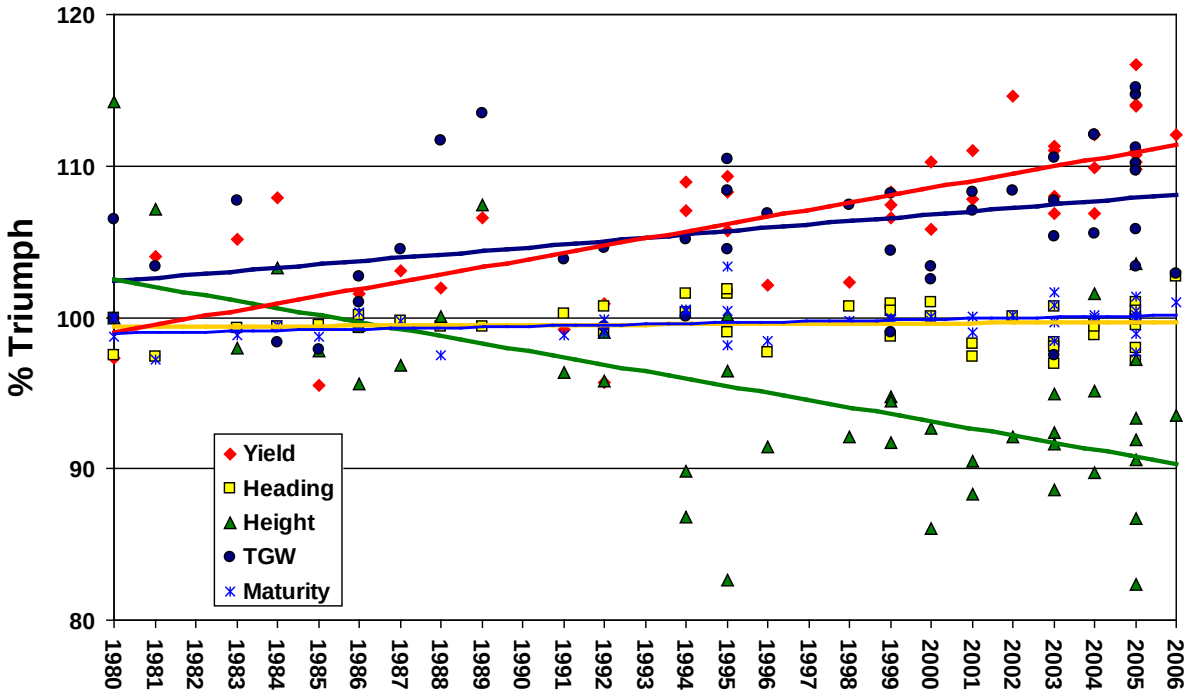


Figure 19. Regression of mean yield, heading, height, thousand grain weight and maturity against year of first recommendation for 44 spring recommended cultivars in the AGOUEB trials from 2006-8. Each point represents an individual variety mean expressed as a percentage of Triumph.

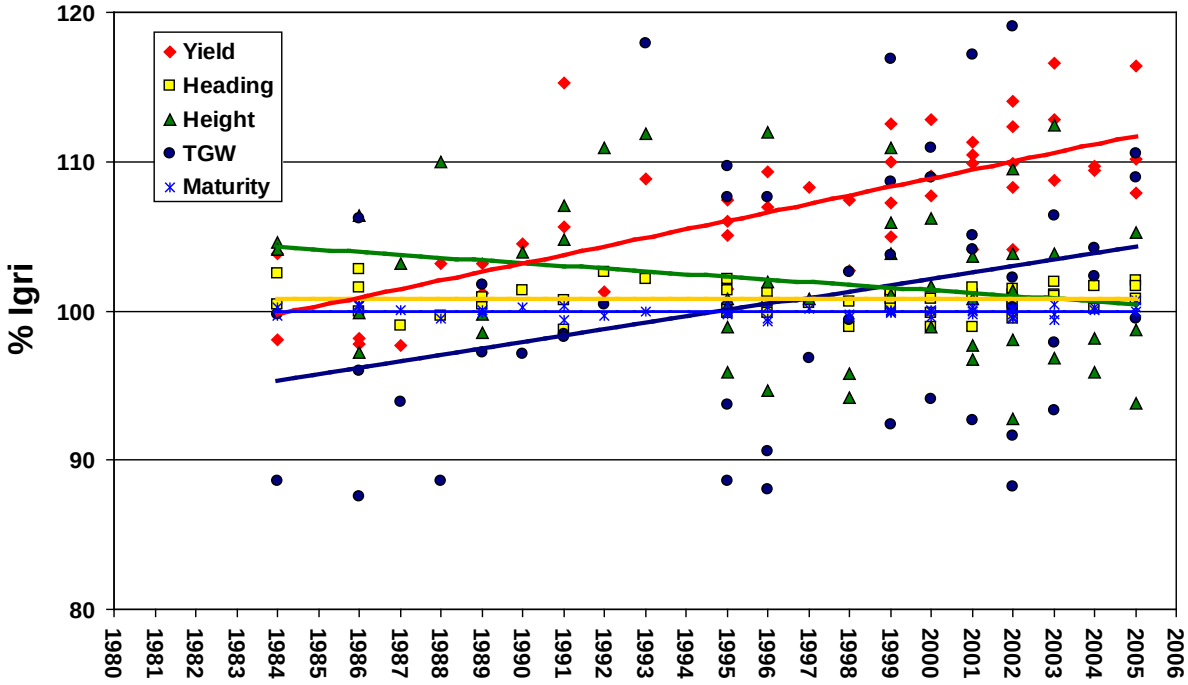


Figure 20. Regression of mean yield, heading, height, thousand grain weight and maturity against year of first recommendation for 48 winter recommended cultivars in the AGOUEB trials from 2006-8. Each point represents an individual variety mean expressed as a percentage of Triumph.

Similarly, the variety Igri had a major impact upon winter barley growing in the UK; although it was first recommended in 1977, no other winter barleys that we surveyed were recommended until 1984 (Halcyon and Panda). When the variety means for heading, height, maturity and thousand grain weight are expressed as a percentage of Igri, we can see that the increase in yield is accompanied by an increase in thousand grain weight. Although this is a significant improvement over time, there was a considerable scatter of the points so the relationship is not as strong as that for yield. No significant change was observed in height, heading date or maturity (Figure 20).

It is also instructive to use the data obtained from the micro-malting analyses to detect if any breeding progress has been made but, in this case, it is more appropriate to restrict the varieties to those which at least gained provisional Institute of Brewing and Distilling approval for use as a malting barley. This results in 26 spring and 16 winter barleys that were included in our study and gained at least provisional approval. For winter barley, there has been significant improvement of 5 Lintner° kg^{-1} in hot water extract over the survey period. Whilst this equates to annual progress of just over 0.05%, it does represent a very real improvement in the amount of alcohol that can be extracted from each tonne of malt and so is also commercially significant. A similar rate of progress can be seen in the winter barley varieties, although the overall potential is some 10 Lintner° kg^{-1} less than that of spring barley. This improvement is not, however, statistically significant as the relationship only accounts for 5% of the variation in hot water extract. This is largely due to exceptionally high values for Melanie and Leonie (Figure 21) and, as these values are derived from just one micro-malt in one year, some caution should be placed upon them. In contrast, each of the spring values is derived from micro-malts carried out over at least two years and their data can be used with greater confidence.

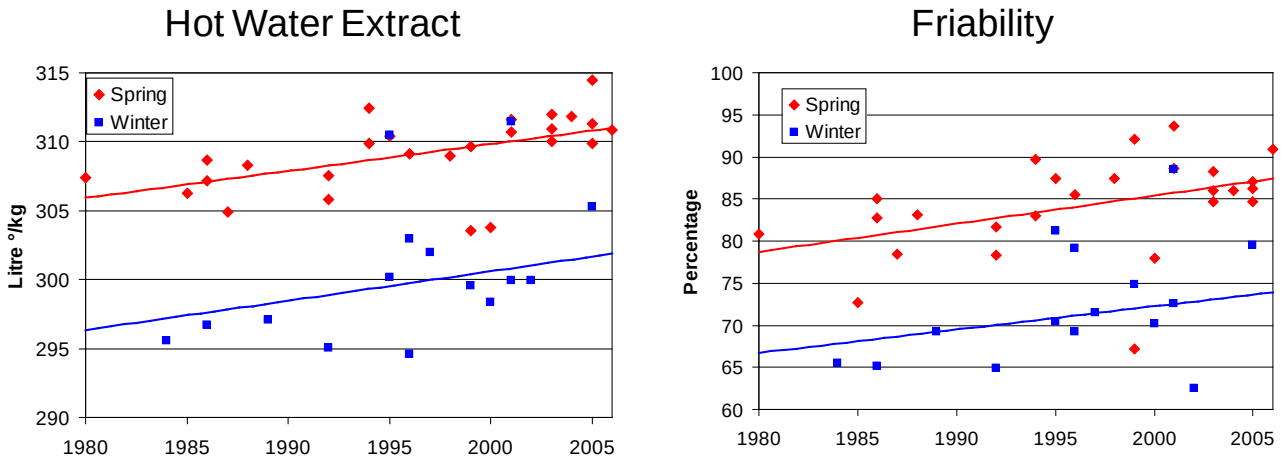


Figure 21. Regression of means for hot water extract and friability against year of first recommendation for spring and winter barleys that gained at least provisional approval by IBD.

There have been apparent improvements in the processability of winter barley varieties over time with a significant decrease in wort viscosity and also a decrease in wort beta-glucan levels. The latter relationship, though similar to the former, is not, however significant, suggesting that the significant reduction in viscosity may well be due to other factors as well as beta-glucan. The winter crop is still, however, significantly poorer than the spring for these two processability parameters (Figure 22) so winter malting barley breeding should target these characteristics for further improvement. There were no significant trends in the spring crop for these two parameters, although some varieties had noticeable wort beta-glucan problems, e.g. Century and Chime both exceeded 380 milli Pascals.

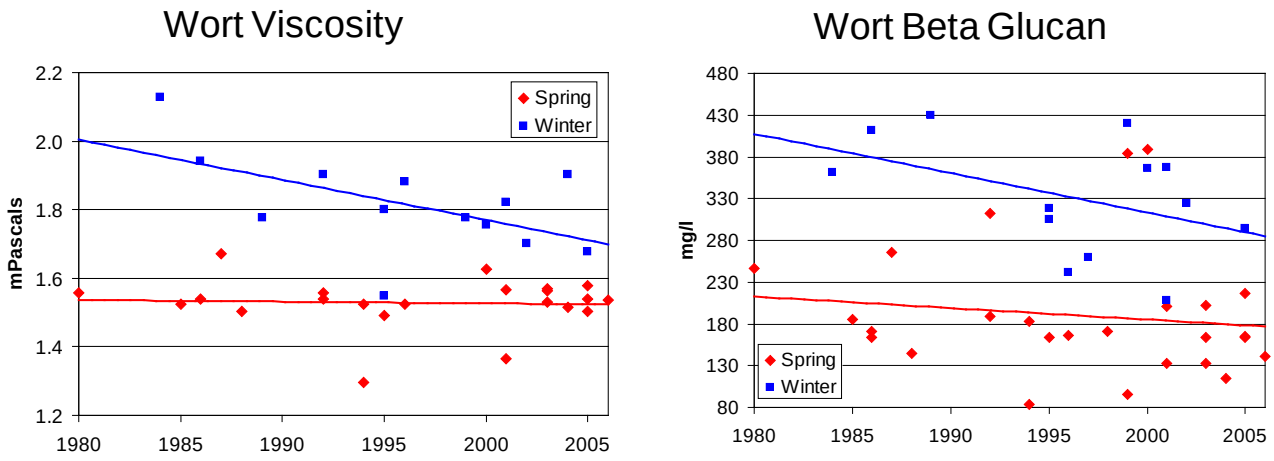


Figure 22. Regression of means for wort viscosity and beta glucan against year of first recommendation for spring and winter barleys that gained at least provisional approval by IBD.

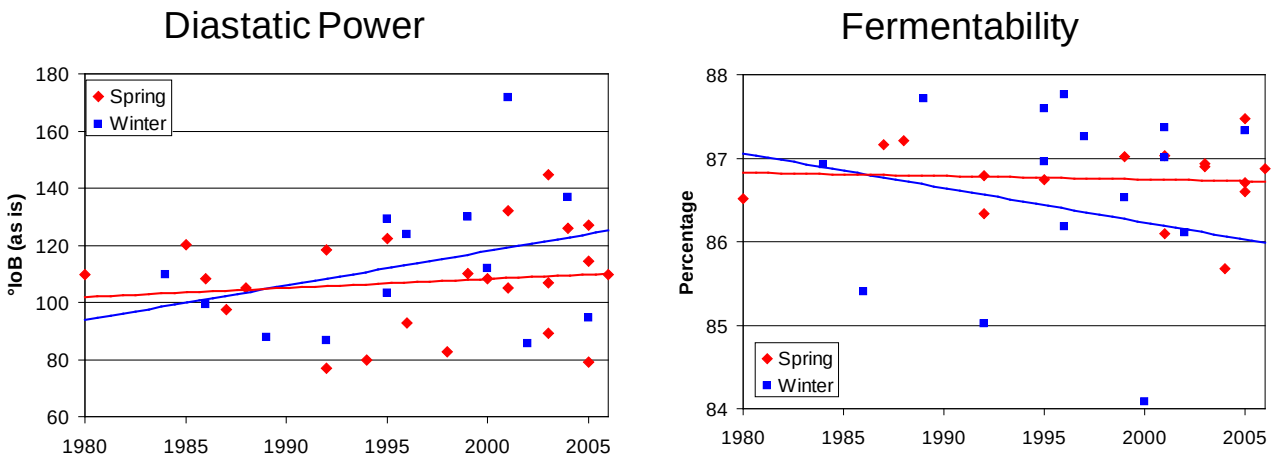


Figure 23. Regression of means for wort diastatic power and fermentability against year of first recommendation for spring and winter barleys that gained at least provisional approval by IBD.

Similarly, there has been no significant improvement in two processing parameters relating to alcohol production; diastatic power and fermentability. Whilst there has been an apparent increase

in the diastatic power of the winter barley varieties, their fermentability has declined this may, in part, be due to changes in grain nitrogen content as Opal had the highest diastatic power but also the second highest grain nitrogen content. There are, however, considerable deviations from the relationship in the winter crop leading to the non-significant trend. Spring barley varieties have hardly changed over time, although Prestige is notable for a high diastatic power and Oxbridge for a high fermentability, although this did not result in the latter having the highest predicted spirit yield (Figure 23).

3.3.4. Brewing trials

Description

Our original intention was to utilise bulks of seed from the breeders trials to provide the required sample of 25kg cleaned and sieved seed for micro-malting and pilot brewing studies. We attempted this after the 2006 harvest but found that there was too much heterogeneity between the various trials to produce samples that would be viewed as acceptable for brewing. We therefore adopted a different strategy during the project of growing strips of 8 and 4 selected varieties of spring and winter barley respectively at two sites for each crop and apply a typical malting barley management regime to them. We then sampled 25kg cleaned and sieved seed from each variety at each site and sent it to the laboratories at Campden BRi for pilot brewing, where they were subjected to a range of tests upon the raw grain, the malted grain, .processability for brewing, the wort, and the resulting beer, including sensory attributes. In addition, data on the performance of each sample during lautering and fermentation was collected.

Whilst the samples were unreplicated at each site, the use of two sites enabled us to conduct a statistical analysis of the data, fitting main effects for variety and site. As there was no replication, we cannot estimate the true error effect but could use the variety x site interaction as an alternative, although it will probably be biased by the inclusion of interaction as well as error effects.

Grain characters

Analysis of the grain samples showed that the spring and winter samples had acceptable germination levels with grain nitrogen contents that were generally well within the limits for malting barley. There were some varieties with water sensitivity problems, notably Triumph amongst the springs and Fanfare amongst the winters, both varieties with known issues in this respect. There were, however, very few significant differences between varieties (Tables 6 and 7)

Table 6. Means of each of 8 spring barley lines for five characters measured on the dry grain samples from >25kg grain produced at two sites from 'Quality Strips'. Significant genotypic variation was found for variates in red.

Variety	Germ_Cap	Germ_Ener	Water Sens	Grain Moist	Grain N
Cellar	97.5	95.5	81	13.2	1.56
Cocktail	98	97.5	87	13.25	1.47
Optic	95.5	96	68.5	13.15	1.46
Prestige	99	98	74.5	13.5	1.595
Prisma	99.5	97.5	79.5	13.45	1.65
Tipple	97.5	96	84	13.4	1.48
Triumph	98	97	25.5	13.45	1.595
Westminster	99	96.5	66	13.65	1.53
Mean	98	96.75	70.75	13.38125	1.5425
SED	1.402	1.414	6.222	0.1532	0.0815

Table 7. Means of each of four winter barley lines for five characters measured on the dry grain samples from >25kg grain produced at two sites from 'Quality Strips'. Significant genotypic variation was found for variates in red.

Variety	Germ_Cap	Germ_Ener	Water_Sens	Grain_Moist	Grain_N
Fanfare	99	98.5	41	14.35	1.675
Flagon	99.5	99.5	59	13.8	1.65
Maris Otter	96	97.5	75.5	14.3	1.74
Pearl	98	99	66.5	14.3	1.775
Mean	98.125	98.625	60.5	14.1875	1.71
SED	1.339	0.677	14.17	0.1208	0.1008

Malt Characters

All the spring barleys had higher levels of extract and fermentability than the best winter barley. In general, the springs had also modified better had lower viscosity and a better mash filtration test (Tables 8 and 9). The springs also had a lower mean nitrogen content and some of the above differences probably reflect, at least in part, the 0.17% difference in nitrogen. Using the Bishop equation to correct for nitrogen content does reduce the difference in extract but the springs do remain noticeably better than the winters. There were more significant differences between varieties for this group of characters than any other group (Tables 8 and 9).

Table 8. Means of each of 8 spring barley lines for 9 characters measured on the malt samples produced from 25kg grain grown at two sites from 'Quality Strips'. Significant genotypic variation was found for variates in red.

Variety	HWE FG	Colour	SNR	FAN	Ferm	Visc	Friab	DP	Vmax
Cellar	318.5	3.1	39.5	0.115	78.5	1.515	97	107	53
Cocktail	319	3.1	40.5	0.11	78	1.54	97	91	27
Optic	319	2.95	37.5	0.095	76.5	1.565	95.5	89.5	31
Prestige	315.5	3.1	38.5	0.11	77	1.51	93	113.5	49.5
Prisma	314	2.9	37.5	0.105	76	1.52	92.5	77.5	19
Tipple	318	2.8	38.5	0.11	77	1.485	96	100.5	34
Triumph	314	2.65	41	0.115	77	1.56	92.5	79	43
Westminster	321.5	3.25	45	0.135	78	1.54	94.5	111	59
Mean	317.4	2.98	39.75	0.1119	77.25	1.529	94.75	96.125	39.44
SED	1.906	0.2789	1.35	0.006478	0.3273	0.0198	2.478	8.016	15.12

Table 9. Means of each of four winter barley lines for 9 characters measured on the malt samples from 25kg grain grown at two sites from 'Quality Strips'. Significant genotypic variation was found for variates in red.

Variety	HWE FG	Colour	SNR	FAN	Ferm	Visc	Friab	DP	Vmax
Fanfare	310.5	2.5	35.5	0.11	75.5	1.555	87	75.5	19
Flagon	311.5	2.4	36	0.1	73.5	1.585	89	69.5	12
Maris Otter	309	2.8	34	0.1	74.5	1.56	92.5	82	21
Pearl	310	2.6	34.5	0.105	74	1.595	89.5	106.5	19.5
Mean	310.3	2.58	35	0.1038	74.38	1.574	89.5	83.375	17.88
SED	1.732	0.08165	1.291	0.00354	0.3536	0.0195	5.244	7.334	13.72

Processability Characters

Significant differences for the processability characters were only detected for extract amongst the spring varieties. It was, however, noticeable again that the springs produced a greater extract, reflecting the results from the malt analyses. Also, the specific gravity at the end of fermentation was higher in the winters, resulting in a lower drop during this phase of the whole process (Tables 10 and 11).

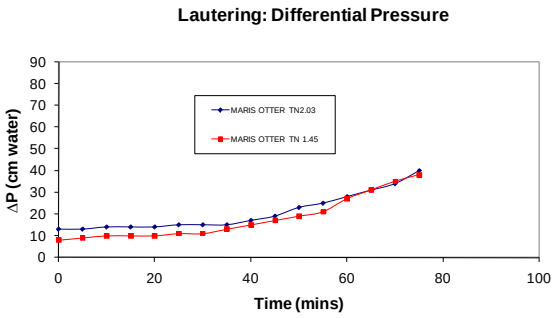
Table 10. Means of each of 8 spring barley lines for 8 characters measured during the mashing and fermentation of the malt samples produced from grain grown at two sites from 'Quality Strips'. Significant genotypic variation was found for variates in red.

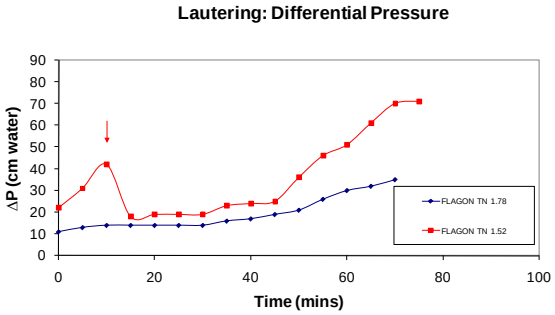
Variety	SG_Last	Extract	SG_Pitch	Volume	Yeast_Viab	SG_End	SG_Drop
Cellar	1006.9	4932.5	1044.3	105	90	1011.4	32.9
Cocktail	1005.0	5100	1043.7	104.75	90.5	1013.8	29.95
Optic	1005.4	5074.5	1043.9	106	93.5	1014.2	29.65
Prestige	1006.4	4975.5	1044	104.5	91	1013.3	30.7
Prisma	1006.5	4839	1043.9	102.25	90.5	1014.4	29.35
Tipple	1005.1	5072	1043.9	104.25	89.5	1010.6	33.4
Triumph	1005.3	4932	1044.2	102.75	90	1010.1	34.1
Westminster	1004.7	5114	1044	107.5	86	1010.9	33.15
Mean	1005.6	5004.9	1044.0	104.63	90.13	1012.3	31.65
SED	1.217	29.02	0.3217	1.452	3.737	2.293	2.291

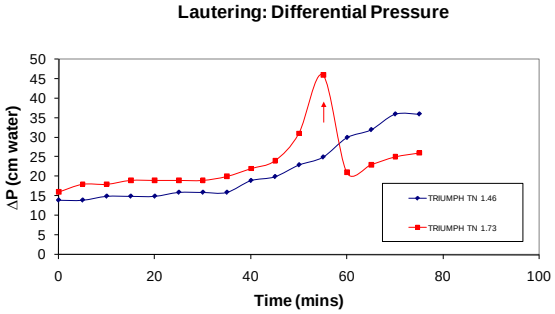
Table 11. Means of each of four winter barley lines for 8 characters measured during the mashing and fermentation of the malt samples produced from grain grown at two sites from 'Quality Strips'. Significant genotypic variation was found for variates in red.

Variety	SG_Last	Extract	SG_Pitch	Volume	Yeast_Viab	SG_End	SG_Drop
Fanfare	1004.5	4888.5	1044.1	102.75	86.5	1015.7	28.4
Flagon	1006.5	4782.5	1044.2	98.5	90	1016.3	27.9
Maris Otter	1007.4	4794.5	1043.9	101.25	87.5	1014.5	29.45
Pearl	1006.1	4966	1043.8	105.5	90.5	1013.5	30.25
Mean	1006.1	4857.9	1044.0	102	88.625	1015.0	29
SED	0.715	66.52	0.2475	1.646	2.336	2.055	2.178

In addition, there were notable differences between varieties in terms of their filtration performance. Amongst the winters, Maris Otter was noticeably better than the other three varieties, running off well even from a high nitrogen sample, confirming its reputation as an excellent malting barley (Figure 24). The springs generally performed better than the winters but both Westminster samples required raking, as did the higher nitrogen samples of Cocktail, Optic and Triumph. It was, however, noticeable that the newer variety Tipple performed much better than the oldest variety tested, Triumph (Figure 24)







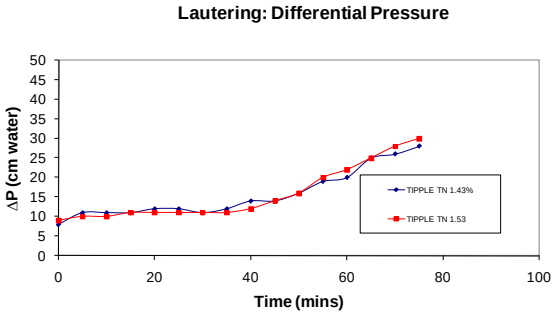


Figure 24. Lauter traces of older winter and spring barley varieties (Maris Otter and Triumph respectively) compared to newer winter and spring barley varieties (Flagon and Tipple respectively). Red lines indicate a higher grain nitrogen sample than the blue line. The red arrows on the traces for Flagon and Triumph indicate that the sample had to be raked for filtration to proceed adequately.

Wort Characters

There were no significant differences between varieties for any of the wort characters and the winters and springs were quite similar for most characters. Whilst the winters produced a slightly greater bitterness and a higher soluble nitrogen, the springs had a greater free amino nitrogen (Tables 12 and 13).

Table 12. Means of each of 8 spring barley lines for seven characters measured on the wort produced from malt samples produced from grain grown at two sites from 'Quality Strips'. Significant genotypic variation was found for variates in red.

Variety	pH	Colour	Gravity	Bitter	FAN	TSN	Ferm
Cellar	5.25	15	43.95	31	215	816	72.5
Cocktail	5.28	14	43.6	29.5	208.5	814	73
Optic	5.235	14	43.45	29	164	720.5	71.5
Prestige	5.255	13.5	43.85	28	195	780	73.5
Prisma	5.285	14	43.3	30	210.5	799.5	71.5
Tipple	5.335	13.5	43.5	30.5	184.5	763	73.5
Triumph	5.265	13.5	43.7	28.5	180	817	72.5
Westminster	5.235	14.5	43.75	28	186.5	872	72.5
Mean	5.2675	14	43.6375	29.3125	193	797.75	72.5625
SED	0.05976	1.254	0.4136	1.306	20.64	55.74	0.7008

Table 13. Means of each of four winter barley lines for seven characters measure on the wort produced from malt samples produced from grain grown at two sites from 'Quality Strips'. Significant genotypic variation was found for variates in red.

Variety	pH	Colour	Gravity	Bitter	FAN	TSN	Ferm
Fanfare	5.305	15	43.8	30.5	173	799.5	73
Flagon	5.33	12.5	44.1	32.5	164.5	791.5	72
Maris Otter	5.27	15	44.05	30.5	168	831	74
Pearl	5.27	14	44	31.5	190	884	73.5
Mean	5.29375	14.125	43.9875	31.25	173.875	826.5	73.125
SED	0.05784	1.947	0.2669	0.7071	11.64	27.98	1.568

Beer Characters

The only significant differences for the beer characters were detected amongst the springs for FAN, where Optic was noticeably low and Westminster high. Differences between the winters and springs were generally slight for this group of characters (Tables 14 and 15).

Table 14. Means of each of 8 spring barley lines for 9 characters measured on pilot brews produced from malt samples of grain grown at two sites from 'Quality Strips'. Significant genotypic variation was found for variates in red.

Variety	pH	Colour	Gravity	Att Limit	Head_Ret	Bitter	FAN	TSN	Ethanol
Cellar	3.98	10	8.95	4.2	9.22E+07	17.5	60.5	564	4.45
Cocktail	4.065	9.5	11.8	4.015	8.62E+07	20.5	52.5	531	4.15
Optic	3.96	10.5	12.6	4.77	1.03E+08	20.5	40	489.5	4
Prestige	3.95	10.5	11.8	4.715	9.12E+07	19	54	562.5	4.2
Prisma	4.025	10.5	13.15	4.68	9.02E+07	21.5	56	580.5	4
Tipple	4.05	9.5	8.75	3.965	8.42E+07	21.5	52	530.5	4.5
Triumph	3.985	11	8.5	4.275	9.47E+07	20.5	56.5	577.5	4.7
Westminster	3.985	10.5	8.3	4.09	7.92E+07	19	71.5	624.5	4.6
Mean	4	10.25	10.48	4.33875	90052488	20	55.38	557.5	4.33
SED	0.0518	0.535	1.857	0.7301	5.95E+06	1.363	6.299	35.97	0.2413

Table 15. Means of each of four winter barley lines for 9 characters measured on pilot brews produced from malt samples of grain grown at two sites from 'Quality Strips'. Significant genotypic variation was found for variates in red.

Variety	pH	Colour	Gravity	Att Limit	Head_Ret	Bitter	FAN	TSN	Ethanol
Fanfare	4.075	10	13.85	5.705	9.77E+07	21	50.9	565.5	3.95
Flagon	4.16	9.9	15.05	6.585	1.06E+08	23	56.35	596	3.9
Maris Otter	4.175	10	13	5.56	9.37E+07	21.5	60	584.5	4
Pearl	4.02	10.5	11.085	4.85	8.32E+07	19	50.55	576.5	4.75
Mean	4.1075	10.1	13.24625	5.675	95062325	21.125	54.45	580.625	4.15
SED	0.1448	0.383	1.904	0.7454	8655486	2.336	8.457	32.21	0.4933

Table 16. Means of each of 8 spring barley lines for 10 beer volatile characters measured on pilot brews produced from malt samples of grain grown at two sites from 'Quality Strips'. Significant genotypic variation was found for variates in red.

Variety	Ethyl hexanoate	Iso amyl alcohol	Iso amyl acetate	Iso butanol	n Propanol	Ethyl acetate	DMS	Acetal dehyde	%2 3 Pent anedione	Diacetyl
Cellar	0.105	43.35	0.97	8.25	10.9	16.8	38	11.4	0.01	0.045
Cocktail	0.075	40.05	0.71	9.05	10.95	12.77	31	11.6	0.01	0.05
Optic	0.07	40.15	0.68	10.3	10.9	11.96	28	11.1	0.02	0.08
Prestige	0.11	43.15	0.93	10.2	12.7	13.7	35.5	12.6	0.01	0.05
Prisma	0.09	36.05	0.67	7.1	10.8	12.4	44.5	12.45	0.015	0.05
Tipple	0.095	39.15	0.825	8.5	12.45	15.5	36.5	10.45	0.01	0.05
Triumph	0.1	44.55	0.925	9.9	12.75	16.7	52.5	8.75	0.02	0.065
Westminster	0.09	47.1	0.97	7.1	11.05	14.85	33	11.35	0.01	0.05
Mean	0.0919	41.7	0.835	8.8	11.56	14.34	37.38	11.21	0.013	0.055
SED	0.0194	3.094	0.185	1.57	1.026	2.07	3.32	1.518	0.008	0.0177

Table 17. Means of each of four winter barley lines for 10 beer volatile characters measured on pilot brews produced from malt samples of grain grown at two sites from 'Quality Strips'. Significant genotypic variation was found for variates in red.

Variety	Ethyl hexanoate	Iso amyl alcohol	Iso amyl acetate	Iso butanol	n Propanol	Ethyl acetate	DMS	Acetal dehyde	%2 3 Pent anedione	Diacetyl
Fanfare	0.05	37	0.45	5.75	10.05	11.6	27	13.15	0.025	0.075
Flagon	0.05	33.65	0.435	4.75	9.95	11.05	25	13.95	0.035	0.11
Maris Otter	0.085	35.35	0.775	6.35	10.75	14.05	35	13.05	0.02	0.07
Pearl	0.085	39.75	0.83	6.85	11.25	13.25	29	11.8	0.015	0.055
Mean	0.0675	36.4375	0.6225	5.925	10.5	12.4875	29	12.9875	0.02375	0.0775
SED	0.009129	4.135	0.08651	1.913	2.518	1.547	5.033	0.6334	0.00677	0.01

Table 18. Means of each of 8 spring barley lines for 11 beer aroma characters measured on pilot brews produced from malt samples of grain grown at two sites from 'Quality Strips'. Significant genotypic variation was found for variates in red.

Variety	Fruity Estery	Alcoholic Solvent	Fruity Citrus	Hoppy	Malty	Cereal Grainy	Toffee	DMS	Other Sulphury	Sweet	Other
Cellar	4.04	3.88	2.665	3.825	3.79	1.745	3.79	0.26	0.925	3.565	0.18
Cocktail	3.83	4.01	2.99	3.71	3.44	1.61	3.44	0.955	1.015	3.635	0.18
Optic	3.455	3.84	2.84	3.515	3.29	1.62	3.29	0.26	0.78	3.255	0
Prestige	3.835	3.52	2.895	3.62	3.785	1.685	3.785	0.63	0.76	3.63	0.575
Prisma	3.58	3.665	2.585	3.66	3.71	1.535	3.71	0.17	0.735	3.28	0.225
Tipple	3.43	3.68	2.47	3.37	3.045	1.585	3.045	0.31	1.285	3.31	0.4
Triumph	3.48	3.73	2.47	3.67	3.845	1.985	3.845	0.795	0.83	3.385	0.64
Westminster	4.1	3.845	2.88	3.62	3.765	1.765	3.765	0.355	0.78	3.655	0.335
Mean	3.719	3.771	2.7245	3.624	3.584	1.691	3.584	0.467	0.889	3.464	0.317
SED	0.4853	0.2119	0.2729	0.2201	0.2816	0.2889	0.2816	0.174	0.3688	0.4285	0.3155

Table 19. Means of each of four winter barley lines for11 beer aroma characters measured on pilot brews produced from malt samples of grain grown at two sites from 'Quality Strips'. Significant genotypic variation was found for variates in red.

Variety	Fruity Estery	Alcoholic Solvent	Fruity Citrus	Hoppy	Malty	Cereal Grainy	Toffee	DMS	Other Sulphury	Sweet	Other
Fanfare	2.9	3.7	2.3	3.45	3.05	1.45	3.05	0.6	1.25	3.15	0.75
Flagon	2.8	3.4	2.35	3.1	3.2	1.4	3.2	0.4	0.9	3.05	0.7
Maris Otter	3.15	3.45	2.45	3.5	3.65	1.7	3.65	1.55	1	3.55	0.9
Pearl	3.25	3.55	2.6	3.35	3.55	1.7	3.55	0.7	0.65	3.5	0.75
Mean	3.025	3.525	2.425	3.35	3.3625	1.5625	3.3625	0.8125	0.95	3.3125	0.775
SED	0.3291	0.3082	0.434	0.3291	0.5983	0.08898	0.5983	0.9476	0.2082	0.4178	0.3873

Table 20. Means of each of 8 spring barley lines for 16 beer aroma characters measured on pilot brews produced from malt samples of grain grown at two sites from 'Quality Strips'. Significant genotypic variation was found for variates in red.

Variety	Fruity Estery	Alcoholic Solvent	Fruity Citrus	Hoppy	Malty	Cereal Grainy	Toffee	DMS	Other Sulphury	Sweet	Sour	Bitter	Astringent	Other	Body	Linger
Cellar	3.79	3.97	2.93	3.92	3.80	1.79	3.80	0.565	1.015	3.56	2.83	4.70	4.22	0.25	3.92	5.04
Cocktail	3.89	4.37	3.18	4.00	3.54	1.67	3.54	0.915	0.880	3.76	2.36	4.91	4.26	0.71	3.87	4.88
Optic	3.62	3.89	3.22	3.99	3.61	1.62	3.61	0.270	0.835	3.36	2.54	4.95	4.02	0.36	4.03	5.16
Prestige	4.05	3.96	2.81	3.67	3.53	1.53	3.53	0.385	0.910	3.71	2.28	4.58	4.18	0.44	3.86	5.09
Prisma	3.77	4.00	2.64	4.00	3.64	1.69	3.64	0.365	0.595	3.59	2.68	4.68	4.23	0.27	3.64	5.05
Tipple	3.64	4.10	3.13	4.00	3.69	1.62	3.69	0.155	1.000	3.26	2.55	5.01	4.28	0.37	4.09	5.11
Triumph	3.69	4.06	3.20	4.00	3.79	1.91	3.79	0.650	1.150	3.67	2.48	4.94	4.12	0.22	4.47	5.39
West-minster	4.03	4.28	2.97	3.83	3.54	1.51	3.54	0.530	0.810	3.51	3.02	5.03	4.63	0.26	4.00	5.54
Mean	3.81	4.08	3.01	3.92	3.64	1.67	3.64	0.479	0.899	3.55	2.59	4.85	4.24	0.36	3.98	5.16
SED	0.28	0.25	0.39	0.20	0.33	0.28	0.33	0.390	0.224	0.29	0.33	0.25	0.25	0.41	0.31	0.21

Table 21. Means of each of four winter barley lines for 16 beer aroma characters measured on pilot brews produced from malt samples of grain grown at two sites from 'Quality Strips'. Significant genotypic variation was found for variates in red.

Variety	Fruity Estery	Alcoholic Solvent	Fruity Citrus	Hoppy	Malty	Cereal Grainy	Toffee	DMS	Other Sulphury	Sweet	Sour	Bitter	Astringent	Other	Body	Linger
Fanfare	3.50	3.70	2.85	3.45	3.10	1.75	3.10	0.600	0.950	3.80	2.25	4.90	4.80	0.85	3.80	5.15
Flagon	3.70	4.30	2.80	3.50	3.10	1.70	3.10	0.500	0.900	3.85	2.50	4.65	4.65	1.25	4.15	5.30
Maris Otter	3.40	3.95	2.80	3.50	3.45	1.90	3.45	1.250	1.000	3.65	2.60	4.60	4.25	0.90	4.00	5.05
Pearl	3.40	3.80	3.05	3.50	3.40	1.90	3.40	0.500	1.100	3.40	2.60	4.80	4.45	0.85	3.90	5.40
Mean	3.50	3.94	2.88	3.49	3.26	1.81	3.26	0.713	0.988	3.68	2.49	4.74	4.54	0.96	3.96	5.23
SED	0.24	0.20	0.22	0.21	0.58	0.54	0.58	0.754	0.234	0.19	0.30	0.33	0.36	0.67	0.11	0.36

Volatiles and Sensory Characters

There were also very few differences between varieties within the spring and winter sets for the beer volatiles (Tables 14 and 15) and the sensory (Tables 16-21) attributes. It was noticeable that the DiMethyl Sulphide (DMS) volatile and aroma did vary significantly amongst the springs with Triumph and Optic having the highest and lowest values respectively for the volatile DMS. The same two also had high and low values for aromatic DMS but the highest value for this character was found for Cocktail and the lowest for Prisma. There were significant differences amongst the winters for three volatiles, where Maris Otter and Pearl had higher values for Ethyl hexanoate and Iso amyl acetate than Fanfare and Flagon. In contrast, Flagon had a significantly higher value for Diacetyl and Pearl a significantly lower value.

3.3.5. Disease data

Mildew genes

Our objective was to assemble a list of hypothesised mildew resistance alleles for the set of barley varieties on the AGOUEB project's winter barley and spring barley sowing lists for 2006, together with a number of varieties found commonly in their pedigrees. We assembled the resistance gene information tabulated in Table 22 for 146 barley varieties and mildew differentials from the following three sources:

1. From mildew resistance allele identifications published by the UK Cereal Pathogen Virulence Survey (UKCPVS) in its Annual Reports from 1990 to 2005
2. From listings on the Cereal Pathogen Resistance Allele Database (CPRAD) at http://cprad.scri.ac.uk/QueryPage.asp
3. From a limited number of supplementary ne novo tests, using definitive isolates of barley mildew (*Blumeria graminis* f.sp *hordei*) of known pathogenicity from the UKCPVS collection.

Table 22. Major gene mildew resistances detected amongst a sample of 66 spring, 67 winter and 14 differential barley genotypes

		Hypothesised Mildew resistance alleles				
Variety	**w/s/d***	**Ml**	**Ml**	**Ml**	**Ml**	**Ml**
Alexis	s	o9				
Amber	s	a13				
Angela	w	h	La	ra		
Angora	w	h	ra			
Antonia	w	h	g			
Apex	d	o				
Appaloosa	s	UN				
Aquarelle	w	a7	ra?			
Aramir	s	a12	g			
Ariel	s	a12				
Aspen	s	o				
Atem	s	La	o11			
Athena	s	o11				
Avenue	w	ra				
Barke	s	o11				
Beatrix	s	g	CP	La	a12	a7
Blenheim	s	Ab	a12			
Boost	w	h	ra	a6		
Brise	s	o11				
Bronze	w	ra				
Camargue	s	Ab	a13			
Camion	w	h	ra	g	CP	
Cannock	w	ra				
Carafe	s	UN				
Carat	w	a12	ra			
Celebrity	w	h	ra	g	CP	
Cellar	s	o				
Centurion	s	a13				
Century	s	o11				
Chalice	s	o11				
Chariot	s	o11				
Chime	s	o				
Clarine	w	0				
Class	s	o				
Cocktail	s	a13				
Colibri	w	h	ra			
Colossus	w	h	ra			
Cooper	s	La	a1			
Cork	s	Ab	a1			
Corniche	s	Ab	a12			
Cypress	w	h	ra			
Decanter	s	o11				
Derkado	s	o11				
Diamond	w	a12	ra			
Digger	d	a13				
Dolphin	w	h	ra			
Doyen	s	a3				
Extract	s	o11				
Fanfare	w	ra	g	CP		
Farenheit	w	ra	12			

		Hypothesised Mildew resistance alleles				
Variety	**w/s/d***	**Ml**	**Ml**	**Ml**	**Ml**	**Ml**
Flagon	w	h	ra			
Flute	w	a1	a12	g		
Forester	s	o				
G Promise	s	0				
Georgie	s	La	g			
Gleam	w	h	ra	a6	g	CP
Goldfoil	d	g				
Golf	s	La	g	CP		
Gypsy	w	h	g			
Haka	w	h	ra			
Halcyon	w	0				
Hanna	w	a8				
Hart	s	o11				
Hassan	d	a12				
Heligan	w	ra				
Henni	s	o11				
Heron	s	o				
Igri	w	ra				
Intro	w	ra				
Jewel	w	a12				
Kestrel	w	h	ra	g	CP	
Kingston	w	h	ra	Ab?	v?	
Kira	w	g				
Kirsty	s	Van				
Klaxon	s	k	La	a7		
Krona	s	11				
Kym	s	La	a9	g		
Leonie	w	Van'				
Livet	s	o11				
Lofa Abed	d	La				
Lotta	d	Ab				
Magie	w	g				
Marinka	w	CP				
Maris Otter	w	0				
Melanie	w	h	ra			
Midas	d	a6				
Monalisa	w	ra	a6			
Muscat	w	ra				
Natasha	s	Ab				
Neruda	s	o				
Novello	s	UN				
Opal	w	h	ra	a6	g	a8
Optic	s	Ab	a12			
Oxbridge	s	Van				
Panda	w	h	g			
Pastoral	w	ra				
Pearl	w	h	ra	g	CP	
Pedigree	w	a12				
Pewter	s	o				
Pict	w	ra				
Pipkin	w	a13				
Platoon	s	a1				
Porter	d	a7				

		Hypothesised Mildew resistance alleles				
Variety	**w/s/d***	**Ml**	**Ml**	**Ml**	**Ml**	**Ml**
Power	s	o				
Prestige	s	o11				
Prisma	s	Ab	a12	g	CP	
Puffin	w	a13				
Putney	s	Van				
Rattle	w	h	ra	g	a12	
Rebecca	s	Van				
Regina	w	a7	ra			
Ricardo	d	a3				
Rifle	w	a12	ra			
Riviera	s	o11				
Riviera	d	o				
Roland	d	a9				
Saffron	w	h	ra	g	CP	a6
Scylla	w	h	ra	g	CP	
Sebastian	s	a12	Ab			
Sequel	w	ra	Ab			
Siberia	w	h	ra			
Sonja	w	ra				
Spectrum	w	h	ra	a7		
Spire	s	a13				
Sprite	w	ra				
Static	s	o11				
Sumo	w	UN				
Sunrise	w	h	ra	g	CP	Ab
Surtees	w	h	ra			
Swift	w	ra				
Tallica	w	h	Ab			
Tankard	s	o11				
Tavern	s	a1	Ab	La		
Tipple	s	UN				
Tocada	s	UN				
Torrent	w	ra	a6	g		
Triumph	s	Ab	a7	Tr3		
Troon	s	o				
Tyra	d	a12				
Vanessa	w	Van				
Vertige	w	h	a7	ra		
Waggon	s	o				
Westminster	s	o				
Wicket	s	a12	Ab			
Willow	w	h	a12			
Zephyr	d	g	CP			

* w = winter variety s = spring variety d = mildew differential cultivar

UN = unidentified resistance

Four of the varieties (Clarine, Golden Promise, Halcyon and Maris Otter) did not possess any resistance gene to the isolates tested, 36 possessed two resistance genes, 12 possessed three, 8 possessed four and five possessed five. There were some notable differences in the distribution of resistance alleles across the winter and spring genotypes. The resistance alleles Mla3, ML(Ab),

Mlk, Ml(La) and alleles at mlo were only found amongst the springs whereas Mla7, Mla8, Ml(CP), Mlh and Mlra were only found amongst the winters. Whilst the sample size is small, it is noticeable that the all but two of the 13 genotypes with four or five resistance genes were winter types, the exceptions being Beatrix (5) and Prisma (4). The winter varieties with five genes were Gleam, Opal, Saffron, and Sunrise.

Table 23 summarises the frequency of different Ml alleles postulated for the varieties listed in Table 22. Over the period represented by these varieties (predominantly since the early 1990s), the most commonly occurring resistance allele was Mlra (in 29.6% of varieties). Mlh, Mlg, Mla12 and MlAb were each identified in between 10% and 20% of varieties. The least commonly postulated alleles were MlTr3, Mlv and Mlk (each in 0.6% varieties) and Mla3, Mla8 and Mla9 (each in 1.2% varieties). mlo (including mlo11, mlo9 and 'mlo') was postulated in a total of 20.3% of varieties, all springs.

The final column of the table shows the mean frequency of virulence corresponding to each Ml allele identified in the UK pathogen population during the period 1998 – 2009. The population was monitored by the UKCPVS in ten years during this 12 year period, viz 1998 – 2005 inclusive, 2007 and 2009. Virulence for Mla8, Tr3 and Mlv were not monitored. No virulence has yet been detected for mlo, although some evidence has been reported of strains with increased 'aggressiveness'.

Table 23. Frequency of mildew resistance alleles in 162 barley varieties listed in Table 22, with frequency of corresponding virulence in the UK mildew pathogen population (from UKCPVS data; mean of 10 years' surveys over period 1998 – 2009)

Ml allele	Frequency %	Corresponding virulence frequency %
None	2.5	
a1	3.1	68
a3	1.2	10
a6	4.3	28
a7	5.6	88
a9	1.2	18
a12	12.3	91
a13	4.9	19
Ab	10.5	82
CP	8.6	94
g	16.0	98
h	18.5	80
k	0.6	47
La	6.8	91
o	8.6	0
o11	11.1	0
o9	0.6	0
ra	29.6	95
Van	3.7	36*
a8	1.2	n/a
Tr3	0.6	n/a
v	0.6	n/a

* 6 years only

All resistance alleles postulated in 5% or more of the varieties had a very high level of corresponding virulence in the mildew population, 80% or greater. These resistances individually have clearly become ineffective against the pathogen population and even combinations of these will still have a high level of corresponding virulence. Resistance alleles postulated in fewer than 5% of the varieties had substantially lower levels of corresponding virulence, for example Mla9 (18%) and Mla13 (19%). Resistance alleles such as these may continue to offer some protection to varieties for a limited period, particularly if they can be pyramided with other resistances with low levels of virulence.

Rhynchosporium genes

For rhynchosporium genes, like powdery mildew the objective was to assemble a list of hypothesised resistance factors for the set of barley varieties primarily on the AGOUEB project's winter barley and spring barley sowing lists for 2006. The resistance information tabulated below in Table 24 is for 120 barley varieties and comes from:

1. From rhynchosporium resistance factor identifications published by the UK Cereal Pathogen Virulence Survey (UKCPVS) in its Annual Reports from 1973 to 2007
2. From listings on the Cereal Pathogen Resistance Allele Database (CPRAD) at http://cprad.scri.ac.uk/QueryPage.asp
3. From a limited number of supplementary *de novo* tests, using three definitive isolates of barley *R. secalis* of known pathogenicity from the SCRI collection.

Table 24. Rhynchosporium resistances factors (BRR – see Table 25) detected amongst a sample of 56 spring, 64 winter barley genotypes

			Rhyncho BRR factor	
Number	**Name**	**Type**	**Rhyn**	**Rhyn**
1	Angela	w	2	-
2	Antonia	w	-	-
3	Aquarelle	w	2	-
4	Avenue	w	2	-
5	Camion	w	4	-
6	Cannock	w	2	?
7	Carat	w	2	5?
8	Cypress	w	2 or 4	-
9	Diamond	w	2	-
10	Dolphin	w	-	-
11	Fanfare	w	5	-
12	Flagon	w	7	-
13	Flute	w	-	-
14	Gleam	w	2	-
15	Gypsy	w	?	-
16	Haka	w	4	-
17	Halcyon	w	2	-
18	Hanna	w	2	-

			Rhyncho BRR factor	
Number	**Name**	**Type**	**Rhyn**	**Rhyn**
19	Heligan	w	2	-
20	Igri	w	4	-
21	Intro	w	2	-
22	Jewel	w	4	-
23	Kestrel	w	2	-
24	Kingston	w	-	-
25	Leonie	w	5	-
26	Maris Otter	w	-	-
27	Muscat	w	2	-
28	Opal	w	-	-
29	Panda	w	-	-
30	Pastoral	w	4	-
31	Pearl	w	4	-
32	Pedigree	w	-	-
33	Pict	w	2	-
34	Pipkin	w	5	-
35	Rattle	w	-	-
36	Regina	w	2	-
37	Rifle	w	-	-
38	Saffron	w	-	-
39	Scylla	w	-	-
40	Sequel	w	2	-
41	Siberia	w	2	4
42	Spectrum	w	2	-
43	Sprite	w	2	-
44	Sumo	w	4	-
45	Sunrise	w	2	-
46	Tallica	w	-	-
47	Torrent	w	2 or 4	-
48	Vanessa	w	2	-
49	Vertige	w	-	-
50	Carstens 2 row	w	2 or 4	-
51	Charleston	w	2 or 4	-
52	Herfordia	w	-	-
53	Houston	w	-	-
54	Magie	w	-	-
55	Melanie	w	2	-
56	Pioneer	w	2 or 4	-
57	Novello	s	8	-
58	Appaloosa	s	-	-
59	Aramir	s	-	-
60	Aspen	s	-	-
61	Astoria	s	-	-
62	Atem	s	-	-
63	Athena	s	-	-
64	Barke	s	-	-
65	Blenheim	s	-	-
66	Braemar	s	-	-
67	Brise	s	-	-
68	Camargue	s	-	-
69	Carafe	s	-	-
70	Cellar	s	-	-

			Rhyncho BRR factor	
Number	**Name**	**Type**	**Rhyn**	**Rhyn**
71	Century	s	5	-
72	Chalice	s	-	-
73	Chariot	s	-	-
74	Chime	s	-	-
75	Class	s	-	-
76	Cocktail	s	-	-
77	Cooper	s	-	-
78	Cork	s	-	-
79	Corniche	s	-	-
80	Decanter	s	-	-
81	Static	s	-	-
82	Derkado	s	-	-
83	Doyen	s	8	-
84	Forester	s	-	-
85	G Promise	s	-	-
86	Georgie	s	-	-
87	Golf	s	-	-
88	Hanka	s	-	-
89	Hart	s	-	-
90	Henni	s	-	-
91	Heron	s	-	-
92	Kirsty	s	-	-
93	Klaxon	s	-	-
94	Krona	s	-	-
95	Kym	s	-	-
96	Livet	s	8	-
97	Meltan	s	-	-
98	Natasha	s	-	-
99	Neruda	s	-	-
100	Optic	s	-	-
101	Oxbridge	s	-	-
102	Pewter	s	8	-
103	Power	s	-	-
104	Prestige	s	-	-
105	Prisma	s	-	-
106	Rebecca	s	5	-
107	Ria	s	-	-
108	Riviera	s	-	-
109	Scarlett	s	-	-
110	Sebastian	s	-	-
111	Spire	s	-	-
112	Tankard	s	-	-
113	Tavern	s	-	-
114	Tipple	s	-	-
115	Tocada	s	-	-
116	Triumph	s	-	-
117	Troon	s	-	-
118	Waggon	s	-	-
119	Westminster	s	5	-
120	Wicket	s	-	-

* w = winter variety s = spring variety ? = ambiguous test results - = no resistance

The resistance factors have been defined by their reactions on the eight UKCPVS differential cultivars (Table 25). Several of these differentiate *Rrs1* alleles and one *Rrs2* but other genes responsible for the resistance are undefined. Of the winter barley cultivars, 70% had identified resistance factors, mostly single ones but a few probably had more. Only 11% of the spring barley cultivars had resistance factors. The most common factor was BRR2 (Astrix), which was found in 22 of the winter varieties. The next most common was BRR4 (Igri) found in seven of the winters, followed by BRR4 (La Mesita), definitely identified in six winter and spring varieties. This probably traces back to the use of Sergeant as a crossing parent in both spring and winter germplasm. The Digger resistance (BRR8) appears to be exclusive to spring germplasm and the Pirate resistance (BRR7) to winter, although it was only detected in Flagon. For the digenic resistances, the fact that Siberia carries both BRR2 and BRR4 suggests that the latter is a separate locus from Rrs1 as it is not feasible for an inbred line to carry two alleles at the same locus. The same reasoning suggests that it is unlikely that Carat carries both BRR2 and BRR5 if the former is definitely an allele at Rrs1.

Table 25. Differential cultivars and isolates used in *de novo* resistance tests.

				Isolate		
Differential cultivar	*Probable gene*	UKCPVS BRR factor	Octal rank (R-->L)	13-13 octal 136	L2A octal 677	214 octal 324
Maris Mink	-	BRR0	-	**+**	**+**	**+**
Armelle	*Rrs1Brier (3H)*	BRR1	1	-	**+**	-
Astrix	*Rrs1?*	BRR2	2	**+**	**+**	-
Athene	*?*	BRR3	3	**+**	**+**	**+**
Igri	*?*	BRR4	4	**+**	**+**	**-**
La Mesita	*Rrs1La Mesita (3H)*	BRR5	5	**+**	**+**	**+**
Osiris	*Rrs1Jet (3H)*	BRR6	6	-	-	-
Pirate	*?*	BRR7	7	**+**	**+**	**+**
Digger	*Rrs2*	BRR8	8	-	**+**	**+**

In addition to the major resistance factor assays, 295 winter barley cultivars were tested against Rhynchosporium isolates 13-13 and L2A on detached leaves in batches of ca. 60 taking detailed measurements of when lesions appeared and their size (Newton et al., 2001). The basic data values were: 1) Lesion total size (the sum of all the lesion lengths from 6 leaves even if no lesion formed on some); 2) Number of leaves with lesions (maximum = 6); 3) Day on which each lesion appeared. From these data were derived: 4) Lesion mean size; 5) Area under the disease progress curve (AUDPC) (total damage duration), and 6) AUDPC per lesion (mean damage duration for each lesion that formed).

Very few cultivars were resistant to L2A but 13-13 allowed differentiation of some major resistance factors. However, the value of most of these data was for assessment of quantitative resistance. Data were very environmentally sensitive, but preliminary association analyses on part of the dataset identified recognised resistance QTL (e.g. loci on chromosomes 4H and 6H) as well as indicating some other regions. Mean values for some lines are shown in Table 26.

Table 26. Examples of the derived quantitative resistance data from detached leaf assays of 295 winter barley cultivars.

13-13	**Spirit**	**Sprite**	**Sumo**	**Sunbeam**	**Surtees**
Lesion number	0.0	6.0	4.0	5.0	1.0
Lesion total	0.0	13.5	8.4	10.8	2.0
Lesion mean	0.0	13.5	12.7	13.0	12.2
AUDPC total	0.0	53.0	28.0	35.4	12.2
AUDPC/ lesion	0.0	53.0	42.1	42.4	73.2
L2A	**Spirit**	**Sprite**	**Sumo**	**Sunbeam**	**Surtees**
Lesion number	4.0	5.0	5.0	5.0	2.0
Lesion total	10.0	12.7	11.5	11.9	4.2
Lesion mean	15.0	15.3	13.8	14.2	12.7
AUDPC total	78.0	89.1	45.5	65.5	12.7
AUDPC/ lesion	117.0	107.0	54.6	78.6	38.0

3.4. Association analyses

3.4.1. Methodology

Reduction of false positives

Principal component analysis was performed using GenStat v.8 (VSN International) on a similarity matrix created using a simple matching coefficient. We used a Bayesian Markov chain Monte-Carlo approach implemented in the program STRUCTURE v2 (Pritchard *et al* 2000a, 2000b; Falush *et al* 2003) to estimate the membership probability of each cultivar to a number of hypothetical founding sub-populations (K). To avoid overestimation of sub-population divergence (Falush *et al* 2003), a subset of 307 genome-wide genetic markers with ≥ 2cM spacing were selected for this analysis. K was estimated using the admixture model with correlated allele frequencies, modeled with a burnin of 2.5×10^5 cycles followed by 10^6 cycles with duplicate runs between K = 2–20, each returning matrices (Q) of fractional sub-population membership for each cultivar. Agreement between duplicates was assessed in terms of difference in Ln(P|D) between replicates and by calculating the average maximum correlation between all combinations of sub-populations in the two matrices. After comparison of various methodologies to correct for population substructure: a Mixed Linear Model (Yu *et al* 2006), Eigenstrat (Price *et al* 2006), STRUCTURE, as well as Genomic Control (Devlin & Roeder, 1999; Reich & Goldstein, 2001)

implemented on its own and in combination with all other methodologies. We found a mixed linear regression model (Yu *et al* 2006), which accounts for multiple levels of relatedness due to historical population substructure and kinship, to perform best. We used the efficient mixed-model association approach (Kang *et al* 2008), implemented in R v 2.9.0 (http://www.R-project.org/) using previously described software (Astle & Balding, 2009). Relatedness between two individuals was estimated as pair-wise correlation based on standardized (subtract mean, divide by standard deviation) genotypes. While it has been commonplace to set negative entries of $\hat{K}$ to zero, in our definition $\hat{K}$ represents a correlation coefficient measuring excess/lack of between-individual allele sharing compared with that expected by chance; thus, negative kinship values are interpretable and provide valuable information about population structure. Accordingly, we retain negative values of $\hat{K}$ in the analysis. SNPHAP v1.3 (http://www-gene.cimr.cam.ac.uk/clayton/software/) was used to infer missing genotypic data, using sliding windows of 30 adjacent markers with 20 marker overlaps. The significance of GWA scans was estimated using a Bonferroni corrected 0.05 *p*-value ($-\log_{10} p = 4.35$). Power was estimated with significance determined using q-value (http://cran.r-project.org/) which implements the method described by (Storey & Tibshirani, 2003). Genomic Control was implemented by dividing the 1 d.f. test statistic for association by the ratio of the observed median of the test statistic to its expected median under the null. Analyses of association without adjustment for population substructure were by 1 d.f. Chi-squared or t-tests.

Power calculations, heritability and residual phenotypic variation

Power was estimated using our experimental markers and individuals, by simulating phenotypes controlled by n_l loci randomly selected from the marker panel. Phenotypes were allocated (i) a genetic component, for which allele '1' was considered positive (contributing $9/n_l$ units), while allele '0' made no contribution (ii) a structural component dependent on the seasonal growth habit and ear-row number status of each variety (spring/2-row = +5, spring/6-row = +7, winter/2-row = +0, winter/6-row = +2). Values for gene and structure effects were arbitrarily chosen. The simulated genetic components were used to estimate the genetic variation (V_G) for the simulated trait. The value of environmental variation (V_E) necessary to achieve a heritability $\left[h^2 = \frac{V_G}{(V_E + V_G)} \right]$ of 0.5 and 0.9 were obtained by drawing values from a distribution ~N(0, V_E). After removal of simulated causative markers from the analysis, simulated traits were subsequently used in GWA scans (after correction using genomic control or the mixed model). Simulations for $n_l = 1$, 2 and 10 were replicated 100, 100 and 375 times, respectively. Q-values were used to determine significance against a threshold of FDR = 0.1. Discoveries were considered true if they were both significant, and fell within a specified window of genetic distance from the known position of the simulated causative locus (± 1, 2, 4 and 10 cM). Mean proportions of loci identified within each discovery window are reported.

Trait heritabilities (h^2) were estimated as $V_G/(V_G+V_E)$ where V_G and V_E are taken from a fit of the null (no marker association) mixed model. The variation explained SNPs was estimated as the difference in residual phenotypic variation between the null and alternative (with marker association) models. Residual phenotypic variation was taken as the sum of the genetic and environmental variance estimates from the mixed model. The proportion of variation explained (V_P) was therefore estimated as $(V_G + V_E - V'_G - V'_E)/(V_G + V_E)$, where V_G and V_E are estimates from the null model and $V'_G - V'_E$ are estimates from the alternative model.

3.4.2. Development of new analyses

Barley (*Hordeum vulgare* L.) is an economically important model plant for genetics research but cultivated germplasm exhibits extensive Linkage Disequilibrium (LD) due to inbreeding and inter-crossing leading to a relatively narrow genetic base and also the population bottleneck during the domestication of modern barley cultivars (Rostoks et al 2006; Malysheva-Otto et al 2006). Barley cultivars therefore provide extant genetic resources that could permit successful Genome Wide Association (GWA) mapping using a relatively small density of markers but the resolution could be limited. As noted in section 3.2.4, we detected a high degree of population sub-structure and relatedness amongst the AGOUEB public set which could lead to serious confounding in association mapping (Balding, 2006). It is well recognized that barley germplasm is highly partitioned, predominantly due to the number of ear rows (two-row and six-row varieties), and the requirement of vernalization (winter- and spring-sown varieties) (von Zitzewitz et al 2005; Rostoks et al 2006). The present barley cultivars include winter and spring types with the former sub-divided into two and six-row lines. Since the molecular bases of ear row number and vernalization requirement in barley have been relatively well characterised (Cockram et al 2008; Komatsuda et al 2007), the barley sample can be served as a test bed to evaluate the effectiveness of various statistical methods established for the association mapping in structured populations.

We therefore investigated the nature of LD in the AGOUEB Public set and its relationship with the barley sub-population division. Next, we tested the performances of one parametric model and one dimensional reduction technique (principal component analysis, PCA) in the inference of population structure of the barley sample.Then, we undertook a genome-wide scan for significant SNP markers linked to a number of phenotypes using six population structure correction methods favoured in association studies (see Astle and Balding 2009 for a comprehensive review of the recent development of statistical solutions to correct for structure in association studies). Comparison of the association tests revealed that, in terms of false positives and false negatives, a mixed model which makes use of pairwise kinship coefficients among study subjects generally performed the best. By effective control of population structure with the use of a mixed linear model, we had confirmed previously reported associations of seasonal growth habit phenotype and

spikelet morphology (ear row number) with candidate genes, and identified significant associations in a number of complex traits unexplored.

One commonly accepted approach to account for population sub-structure is using the model-based STRUCTURE program to estimate a subpopulation membership matrix Q. However, with the present barley sample, this method failed to infer a reliable parameter, k, the number of subpopulations. In addition, it is important to note that different panels of marker genotypes chosen for STRUCTURE analysis yielded discordant inferences. The failure in achieving an reliable inference is reasonable because STRUCTURE attempts to account for population structure by allocating population groupings in such a way that Hardy-Weinberg equilibrium (HWE) is met within subpopulations, whereas the assumption of HWE within populations is actually invalid given the nature of extensive inbreeding of barley cultivars. While inference of population structure depends critically on the model assumption, violation of model assumption in the present dataset impeded the suitability of such a model-based cluster method.

A one dimensional reduction technique was then applied to predict the hidden structure by extracting features from a covariance matrix using PCA analysis. Although dimensional reduction technique is visually appealing graphically, it is difficult to make an interpretation and statistical assessment of its inference output. We therefore compared the result of PCA to another multidimensional scaling (MDS) method; principal coordinate (PCO) analysis of pairwise Identity By State kinship estimates. Although the MDS plot from PCO analysis presented a similar clustering pattern to that from PCA, there exists difference between the inferences from the two analyses (Figures 25 and 26). Clustering the AGOUEB Public set of barley lines by k-means cluster (Lloyd 1982) using the top features extracted from the two dimensional reduction methods showed similar partition patterns when setting k = 2 but differed when setting k = 3. In fact, employing an x-means algorithm (Pelleg and Moore 2000), an extension of the commonly used k-means cluster algorithm with efficient estimation of the optimized number of clusters based on Bayesian information criterion (BIC), indicated that the best cluster model was with k = 2 when using the top three features extracted from PCA while it was with k = 3 when using top three features extracted from PCO. As, apart from morphology, there was a lack of unequivocal clue in the selection of an appropriate model, caution should be taken in applying dimension reduction techniques to infer population structure.

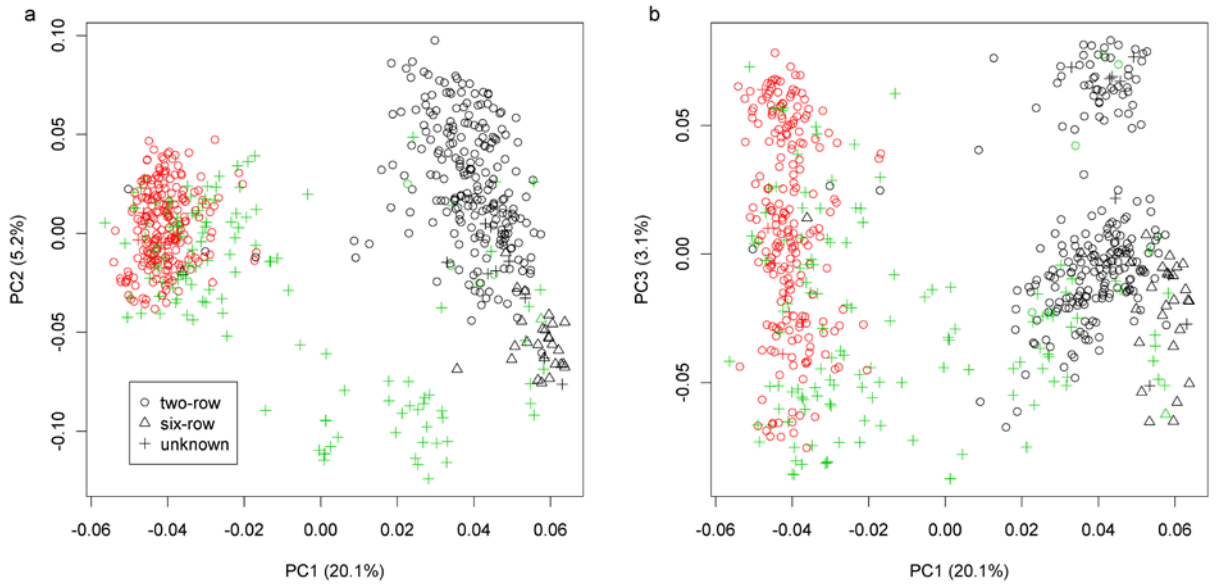


Figure 25. The top three principal components (PCs) in PCA analysis of the variation present in the AGOUEB Public set of barley varieties. Black, red and green colours indicate barley varieties with winter, spring and unknown seasonal growth habit respectively.

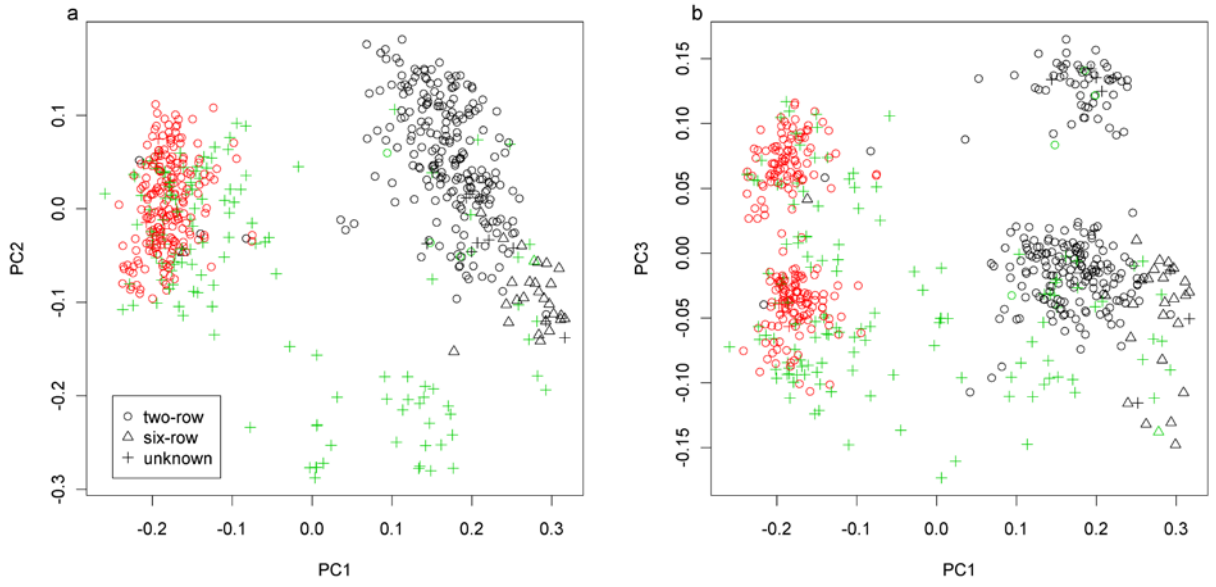


Figure 26. Multidimensional scaling (MDS) plots for pairwise Identity By State coefficients matrix showing the population structure in the AGOUEB Public set of barley varieties. Colours indicate winter (black), spring (red) and unknown (green) seasonal growth habits. PC1-3 are the first three principal coordinates from MDS analysis.

Seven association test approaches were applied in this paper, including Single Marker Association (SMA), StepWise Regression (SWR), Structured Annealing (SA), EIGENSTRAT, Mixed Linear Model (MLM) with Population and Kinship (P+K), MLM (K) and Genomic Control (GC), which

presented highly varying levels of conservatism. Though the more liberal methods resulted in higher mutual predictability, we observed low mutual predictabilities between two liberal methods SWR and SA, suggesting that the added predictions in a liberal method might contain a higher risk of false positives. Through comparison of empirical quantile-quantile plots and Quantitative Trait Nucleotide (QTN) simulations, the MLM (K) model outperforms its rivals in terms of controlling false positives and prediction power. The MLM (P+K) showed a smaller prediction power than MLM (K), which is probably because the kinship matrix K is sufficient in capturing the features of the data while a combination of sub-Populations and K led to over-correction.

Analysis of the structure of LD revealed that the division by seasonal growth habit explained the major source of population structure exhibited in the AGOUEB Public set of barley varieties, which is concordant with a previous study (Rostoks et al 2006). Taking seasonal growth habit as a complex trait, we identified one SNP 11_21210 which was only 1.49 cM away from VRN-H2, one of the major loci previously reported to control seasonal growth habit (Yan et al. 2006; Szűcs et al 2007), on the barley consensus genetic map. Not all of the present seven methods detected SNP 11_21210, however, as SA, EIGENSTRAT and GC failed to reveal a significant association. We detected some other significant associations, including two on chromosome 5 flanking one candidate gene VRN-H1, and a highly significant region on chromosome 1 between 92.8 and 96.9 cM on the consensus map, but these will need to be validated in a future study.

3.4.3. DUS data

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 (Figure 27a); 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 (Figure 27a). The extent of genetic stratification was further investigated using the program STRUCTURE. Analysis of the likelihood of various models (LnP[D] within the range K = 2–20 (burnin = 250,000, MCMC iterations = 1,000,000) indicates the optimum number of sub-populations is ~10 (Figure 27a). 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.

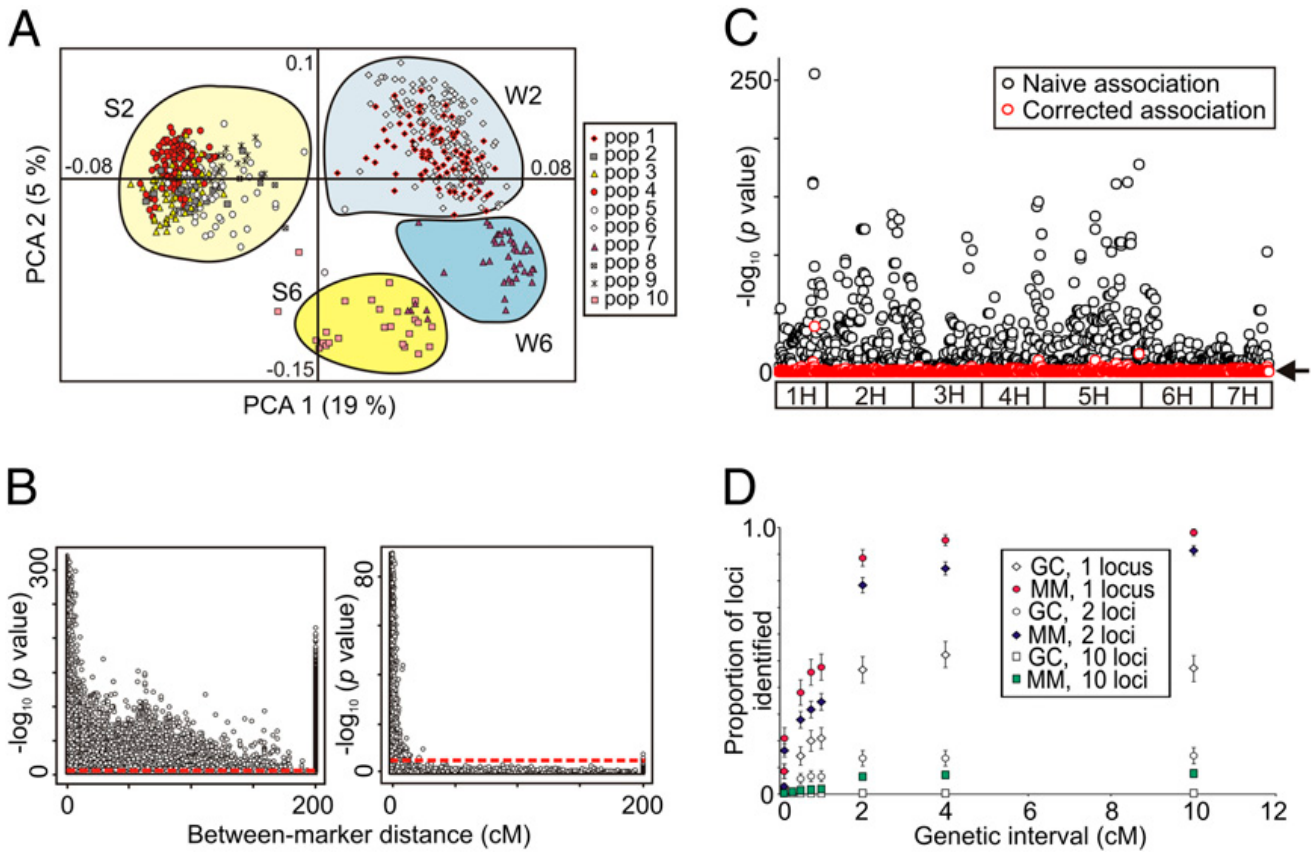


Figure 27a. Principle component analysis. Phenotypic combinations for 'row-number' (2 or 6-row) and 'seasonal growth habit' (S = spring, W = winter) are indicated. Varietal membership to the ten sub-populations identified using STRUCTURE are overlaid. **b.** Decay of pair-wise marker LD over increasing genetic distance (cM) before (left), and after (right) correction using the mixed linear model. Off-chromosome comparisons are shown at 200 cM. The Bonferroni corrected $p = 0.05$ significance threshold is indicated. **c.** Naïve and corrected GWA analysis of 'seasonal growth habit', illustrating the extent of confounding present. Arrow indicates the significance threshold. **d.** Predicted experimental power to detect a trait controlled by 1, 2 and 10 loci ($h^2 = 0.9$) over genetic distance (± cM). Power is measured as the proportion of simulations in which at least one causative locus was detected (q-value ≤0.1). Error bars denote ± 1 standard error.

As identification of marker-trait associations relies on detection of significant LD after correction for spurious signal due to population genealogy, we investigated the extent of pair-wise marker associations, with and without statistical correction for confounding (Figure 27b). Strikingly, we find that for uncorrected analysis, 35 % of inter-chromosomal associations between marker-pairs are significant ($-\log_{10} p \geq 4.35$). Furthermore, significant intra-chromosomal LD is evident across the full length of chromosomes (mean distance between significant marker-pairs = 40.2 cM, median =

30.7 cM). After adjustment using the mixed model, this is reduced to < 10 cM (mean = 1.2 cM, median = 0.6 cM), with the proportion of significant inter-chromosomal associations controlled to just 0.1 %. The specificity of the control achieved using the mixed model is demonstrated by the suppression of off-chromosome and long-distance association, while retaining signal at shorter genetic distances. The extreme levels of confounding encountered are illustrated by naïve analysis of a selected trait ('seasonal growth habit'), where 72 % of all markers return significant associations. This is reduced to 1 % after correction using the mixed model (Figure 27c).

Statistical power

We modeled the power to detect 1, 2 and 10 independent loci distributed randomly across the genome (100, 100 and 375 permutations, respectively) with a heritability (h^2) of 0.5 and 0.9 (Figure 27d). Using the mixed model to correct for genetic substructure, simulations based on a trait controlled by one locus predict our experimental design has a high probability (≥ 0.92 for both values of h^2) of detecting significant (q-value ≤ 0.1) associations within windows of ≤ 8 cM. This compares favorably with correction using Genomic Control, where the power to detect a single locus within the same genetic interval is 0.46 (h^2 = 0.5) and 0.55 (h^2 = 0.9), and falls dramatically in the two locus model (0.18, h^2 = 0.5; 0.23, h^2 = 0.9). For a ten locus trait, even when considering a genetic interval of 20 cM, power to detect one or more loci after correction with the mixed model is low (0.25, h^2 = 0.5; 0.58, h^2 = 0.9), while power after correction with Genomic Control is effectively zero (≤ 0.01 for all scenarios investigated). To report only the most robust marker-trait associations in the experimental dataset, we used a Bonferroni corrected $p \leq 0.05$ threshold ($-\log_{10} p \geq 4.35$) to account for multiple testing. To simplify the results reported, we set our discovery criterion to be ≥ 2 significant markers within a 4 cM interval.

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 (Figure 28). Subsequent GWA analysis using the mixed model to correct for population substructure identified eighteen genomic locations associated (5 ≤ -log10 p ≤ 113) with fifteen traits (Figure 29a; Table 27). 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 $V_P = 0.41$). Overall, seventeen traits (mean $h^2 = 0.18$) failed to return significant associations satisfying our discovery criteria (Figure 29b). To investigate the potential for improved mapping resolution, we exploited colinearity between cereal genomes to develop additional genetic markers within target genomic regions. Using sequence homology between barley ESTs and the rice (Oryza sativa) genome, 49 new gene-specific markers were developed and their genetic map locations determined (Table 28), representing the addition of 3–18 genes within the majority of significant chromosomal regions identified. Re-sequencing these amplicons in a sub-set of 90 varieties identified 195 polymorphisms over ~22 kb of genic sequence (176 SNPs, 13 InDels, 6 SSRs). The inclusion of additional polymorphisms in association analyses within this sub-set identified fourteen genes with associations at least as significant as the peak marker identified in the original GWA scan (Table 28).

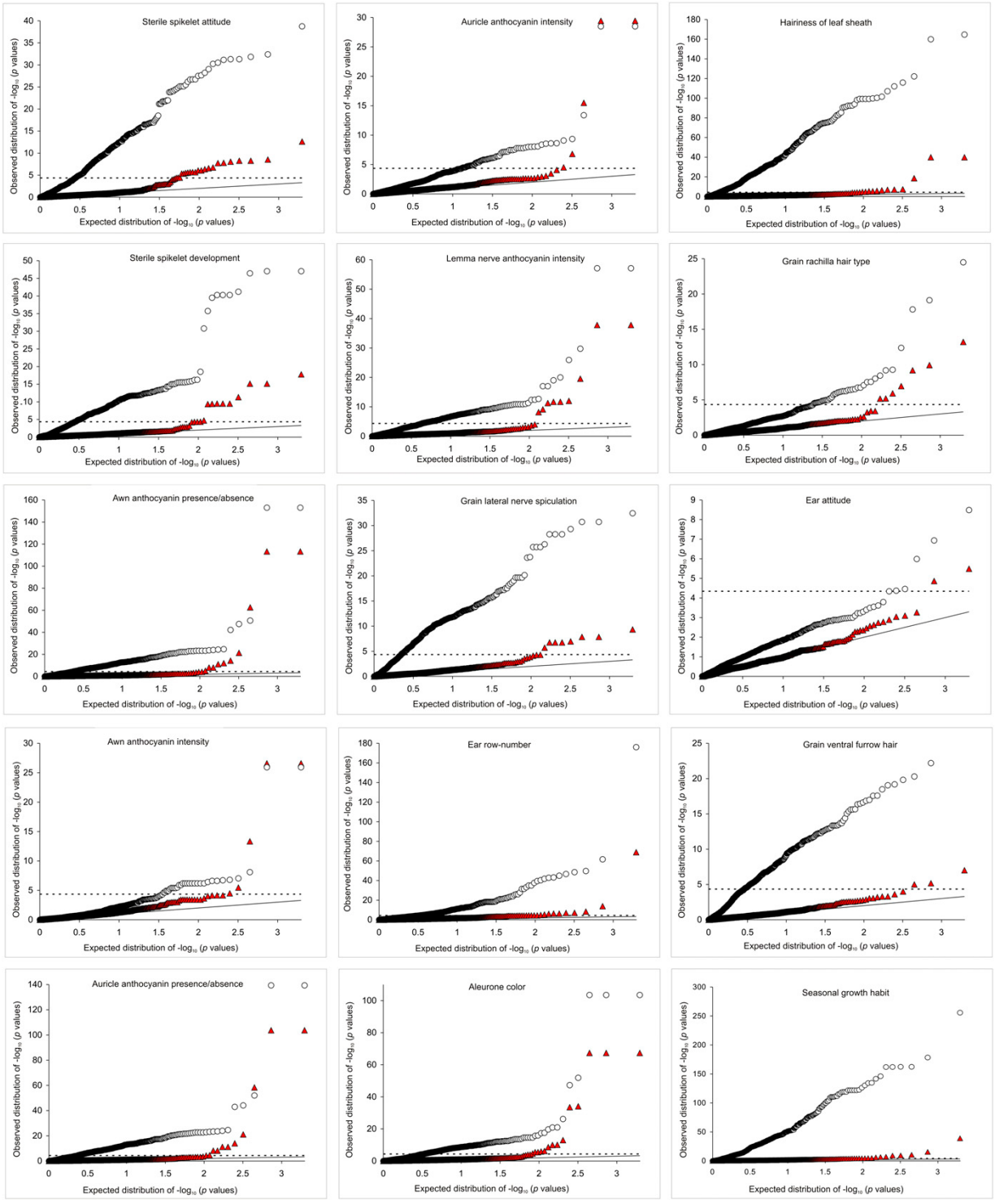
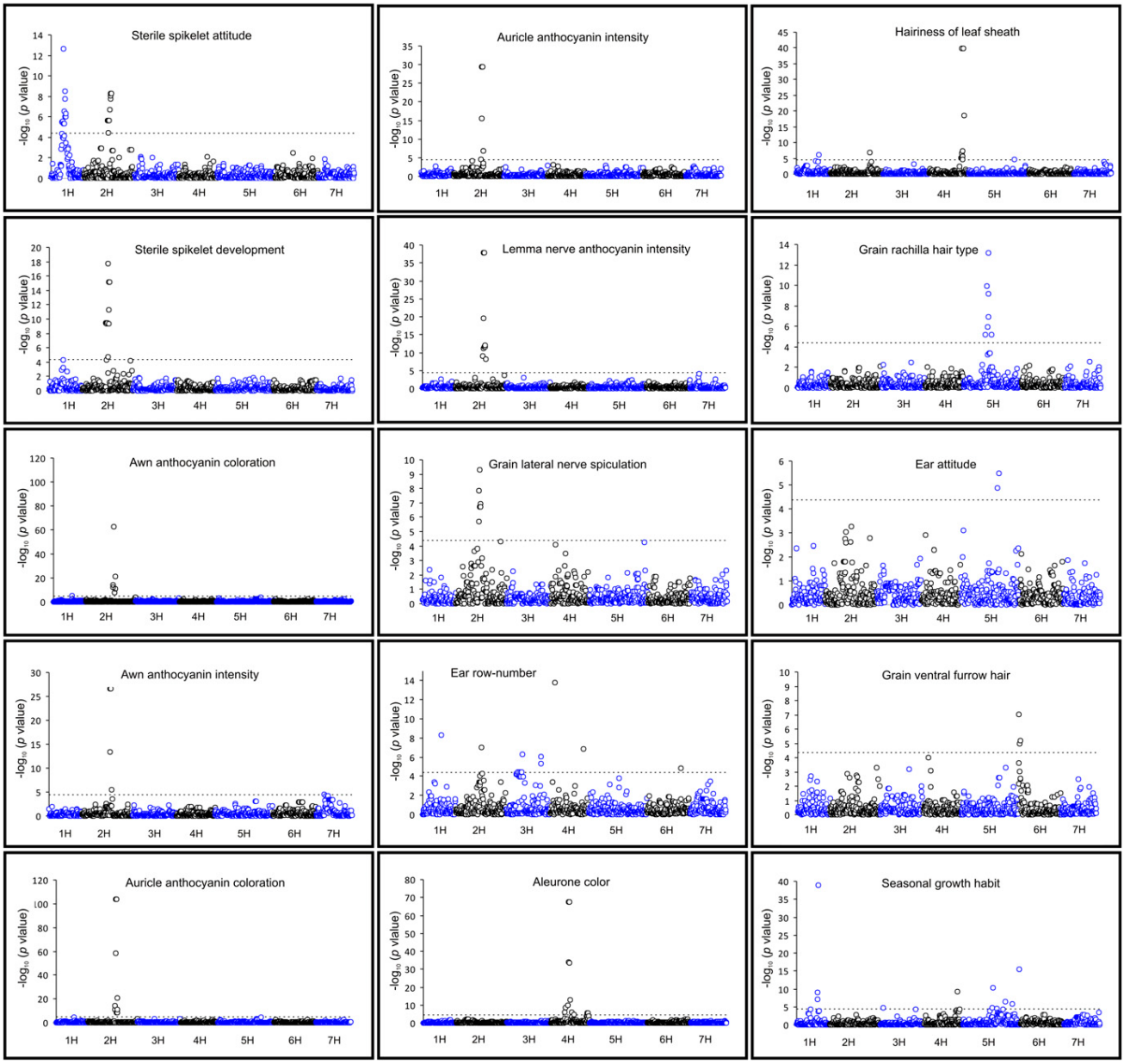


Figure 28. Quantile-quantile plots for traits returning significant associations after GWA analysis. Expected *versus* observed *p*-values are plotted for naive (circles) and mixed model corrected (triangles) analyses. The x = y line (solid) and Bonferroni corrected $p = 0.05$ significance thresholds (dashed line) are indicated.

Table 27. Summary of significant ($-\log_{10} p \geq 4.35$) marker-trait associations identified by genome-wide association scans where a. ≥2 or significant markers were identified within a ± 4 cM window. b. genomic location identified by just one significant marker within a ± 4 cM window. Homology between barley and rice genes was determined by BLASTn analysis of cDNA sequences. S = short arm, L = long arm, C = centromeric. * Mapped to chromosome arm only.

Trait	Chromosome arm (interval, cM)	Peak marker name	$-\log_{10} p$ value	marker position (cM)	Putative rice homologue (e value)
a					
Sterile spikelet attitude	1HC (55.49-60.19)	11_10933	12.61	55.49	LOC_Os04g14150.1 ($1e^{-169}$)
	2HL (82.75-90.1)	11_21245	8.26	89.32	LOC_Os04g49370.1 ($2e^{-90}$)
Seasonal growth habit	1HL (92.80-96.92)	11_10396	38.87	96.92	LOC_Os05g44760.1 (0.0)
Sterile spikelet development	2HL (82.75-90.10)	11_10287	17.76	85.92	LOC_Os04g45490.1 (0.0)
Awn anthocyanin coloration	2HL (93.5-101.78)	11_21175	113.22	96.82	LOC_Os04g46820.1 ($4e^{-26}$)
Awn anthocyanin intensity	2HL (93.5-101.78)	11_21175	26.59	96.82	LOC_Os04g46820.1 ($4e^{-26}$)
Auricle anthocyanin coloration	2HL (93.5-101.78)	11_21175	103.67	96.82	LOC_Os04g46820.1 ($4e^{-26}$)
Auricle anthocyanin intensity	2HL (93.5-101.78)	11_21175	29.43	96.82	LOC_Os04g46820.1 ($4e^{-26}$)
Lemma nerve anthocyanin intensity	2HL (93.5-101.78)	11_21175	37.79	96.82	LOC_Os04g46820.1 ($4e^{-26}$)
Grain lateral nerve spiculation	2HL (78.03-82.75)	11_10786	9.32	82.75	LOC_Os04g44530.1 (0.0)
Ear row-number	3HC (56.40)	11_10349	6.25	56.40	LOC_Os01g44210.1 ($1e^{-148}$)
	3HL (130.19–130.82)	11_10280	6.02	130.19	LOC_Os02g19510.1 (0.0)
	4HS (26.19)	11_20606	68.88	26.19	LOC_Os03g50040.1 (0.0)
Aleurone color	4HL (55.63-65.05)	11_21296	67.32	62.83	LOC_Os03g14040.1 ($1e^{-118}$)
Hairiness of leaf sheath	4HL (106.03-119.09)	11_21210	39.85	117.60	LOC_Os03g01750.1 (0.0)
Grain rachilla hair type	5HL (99.56-119.09)	11_20449	13.21	100.28	LOC_Os09g32526.1 ($4e^{-151}$)
Ear attitude	5HL (135.72-137.16)	11_11080	5.49	137.16	LOC_Os03g54130.1 (0.0)
Grain ventral furrow hair	6HS (1.34-6.07)	11_20881	7.01	1.34	LOC_Os02g01280.1 (0.0)
b					
Auricle anthocyanin intensity	1HS (75.45)	11_20121	4.54	75.45	LOC_Os05g41010.1 (0.0)
Awn anthocyanin coloration	1HL (75.45)	11_20121	5.11	75.45	LOC_Os05g41010.1 (0.0)
Ear-row number	1HL (75.45)	11_20990	8.27	75.45	LOC_Os05g40990.1 (0.0)
	2HL (85.92)	11_10287	7.05	85.92	LOC_Os04g45490.1 (0.0)
	3HC (56.40)	11_10349	6.25	56.40	LOC_Os01g44210.1 ($1e^{-148}$)
	3HL (130.19)	11_10280	6.02	130.19	LOC_Os02g19510.1 (0.0)
	4HL (117.60)	11_21210	6.85	117.60	LOC_Os03g01750.1 (0.0)
	6HL (97.39)	11_10015	4.85	97.39	LOC_Os08g10450.1 (0.0)
Growth habit	1HL (96.92)	11_10396	5.92	96.92	LOC_Os05g44760.1 (0.0)
Hairiness of leaf sheath	1HL (96.92)	11_10396	6.17	96.92	LOC_Os05g44760.1 (0.0)
	2HL (126.03)	11_21440	6.93	126.03	LOC_Os03g61640.1 (0.0)
	5HL (153.51)	11_21355	4.60	153.51	LOC_Os03g58340.1 (0.0)
Seasonal growth habit	3HS (19.15)	11_10565	4.83	19.15	LOC_Os01g04260.1 (0.0)
	3HS (131.59)	11_20612	4.40	131.59	LOC_Os01g68860.1 (0.0)
	4HL (117.60)	11_21210	9.27	117.60	LOC_Os03g01750.1 (0.0)
	5HL (113.83)	11_11341	4.77	113.83	LOC_Os05g25450.1 (0.0)
	5HL (122.38)	11_10094	10.35	122.38	LOC_Os09g38030.1 (0.0)
	5HL (130.84)	11_11375	4.49	130.84	LOC_Os03g40020.1 (0.0)
	5HL (146.00)	11_20568	4.37	146.00	LOC_Os03g57690.1 (0.0)
	5HL (151.36)	11_20100	6.54	151.36	LOC_Os03g57220.1 (0.0)
	5HL (161.58)	11_10336	5.90	161.58	LOC_Os01g36950.1 (0.0)
Grain rachilla hair type	5HL (87.35)	11_20645	5.17	87.35	LOC_Os09g28510.1 ($8e^{-110}$)
Ear glaucosity	6HS (42.36)	11_10129	4.78	42.36	LOC_Os02g03330.1 (0.0)
Flag leaf-sheath glaucosity	7HS*	11_10956	4.40	7HS*	LOC_Os06g05740.1 (0.0)
Awn anthocyanin intensity	7HS (62.88)	11_10721	4.49	62.88	LOC_Os07g12080.1 ($9e^{-18}$)

a.


Sterile spikelet attitude
Auricle anthocyanin intensity
Hairiness of leaf sheath
Sterile spikelet development
Lemma nerve anthocyanin intensity
Grain rachilla hair type
Awn anthocyanin coloration
Grain lateral nerve spiculation
Ear attitude
Awn anthocyanin intensity
Ear row-number
Grain ventral furrow hair
Auricle anthocyanin coloration
Aleurone color
Seasonal growth habit
-log10 (p value)
1H 2H 3H 4H 5H 6H 7H

b.

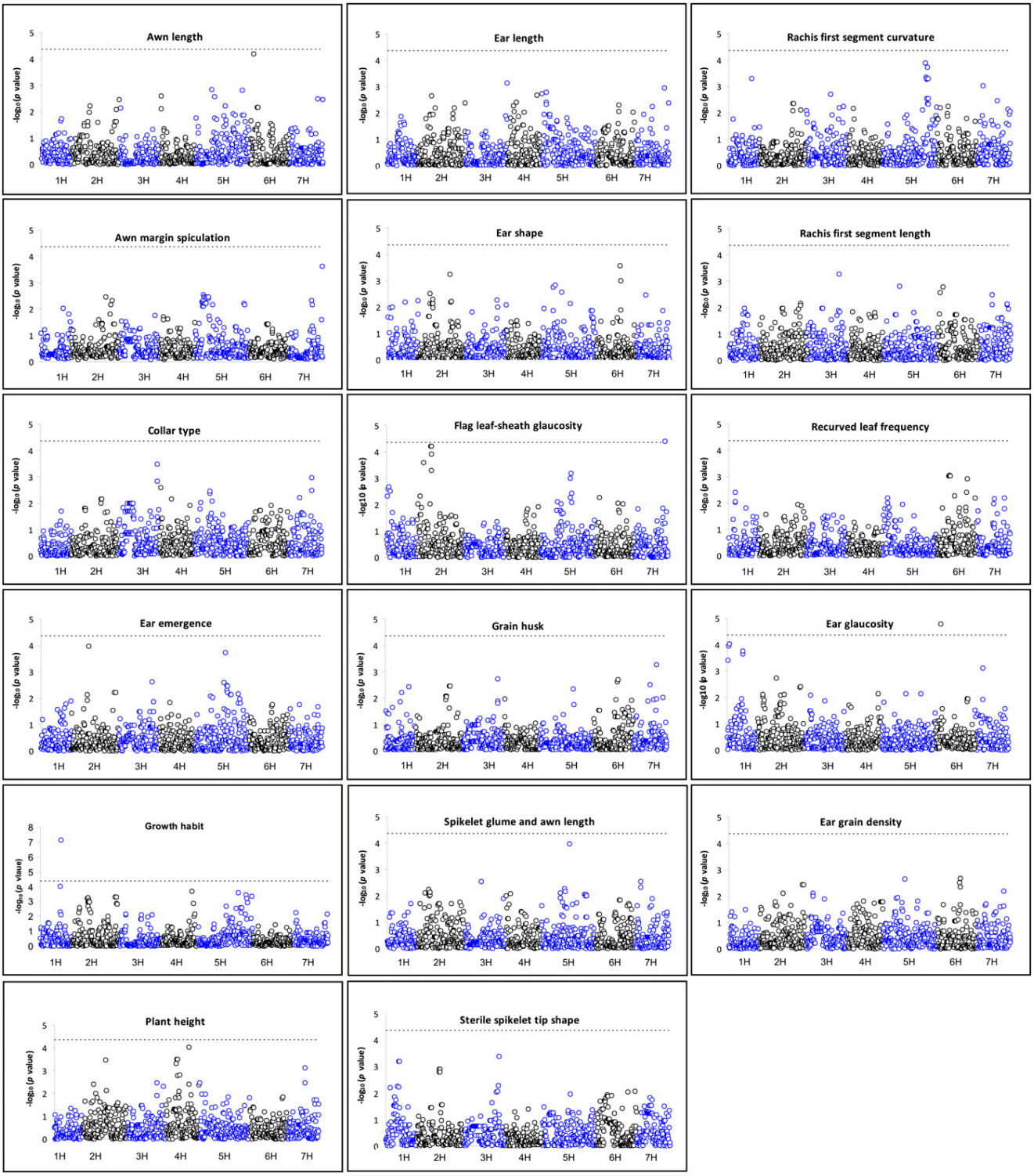


Figure 29. GWA scans. Barley chromosomes 1H to7H are shown. The p <0.05, Bonferroni corrected significance threshold is indicated by a dashed line. a The fifteen traits with significant associations. b. The eighteen traits which did not detect significant associations, according to the discovery criteria employed in this study (-log10 $p \geq 4.35$, ≥ 2 significant markers within a 4 cM window).

Table 28. Current UPOV DUS characters, indicating which of the peak markers identified during GWA scans have been converted to the KASPar platform. As KASPar is not generally suitable for assaying insertion/deletions, we also indicate multiplex PCR/agarose-gel based assays designed for genotyping *VRN-H1.*

Trait	UPOV No.	Marker	Chr
Growth habit	1	HvFT3_FC816A	1H
Lower leaves: hairiness of leaf sheaths	2G	HvOs03g03180_A447G	4H
Lower leaves: hairiness of leaf sheaths	2G	HvOs03g03034_G93A	4H
Lower leaves: hairiness of leaf sheaths	2G	Hv11_11299_GC	4H
Lower leaves: hairiness of leaf sheaths	2G	Hv11_20007_GA	4H
Flag leaf: intensity of anthocyanin colouration of auricles [2]	3	HvANT2_C4289T	2H
Awns: intensity of anthocyanin colouration of awn tips [2]	7	HvANT2_C4289T	2H
Grain: anthocyanin colouration of lemma nerves [2]	23	HvANT2_C4289T	2H
Ear: number of rows [3]	11G	HvVRS1_C349G	2H
Ear: number of rows [3]	11G	HvVRS1_GINS681	2H
Ear: number of rows [4]	11G	HvVRS1_C349G & HvVRS1_GINS681	2H
Ear: number of rows [5]	11G	Hv11_20606_GC	4H
Sterile spikelet: attitude (mid 1/3 of ear) [6]	19	Hv11_10933_GC	1H
Sterile spikelet: attitude (mid 1/3 of ear) [6]	19	Hv11_11359_GC	1H
Sterile spikelet: attitude (mid 1/3 of ear) [6]	19	Hv11_21333_CG	1H
Grain: rachilla hair type	21	Hv11_20449_TA	5H
Grain: rachilla hair type	21	Hv11_10622_GA	5H
Grain: rachilla hair type	21	Hv11_20850_AG	5H
Grain: spiculation of inner lateral nerves [7]	24	Hv11_10818_CA	2H
Grain: spiculation of inner lateral nerves [8]	24	Hv11_11435_AG	2H
Grain: ventral furrow - presence of hairs	25G	HvOs02g01490_G607A	6H
Grain: ventral furrow - presence of hairs	25G	Hv11_21204_GA	6H
Grain: disposition of lodicules [9]	26	HvCly1_A2604G	2H
Grain: disposition of lodicules [10]	26	HvCly1_A2664C	2H
Kernel: colour of aleurone layer [11]	27	HvOs03g14250_C82T	4H
Kernel: colour of aleurone layer [12]	27	HvOs03g14380_G125A	4H
Kernel: colour of aleurone layer [13]	27	Hv11_21296_CA	4H
Seasonal growth habit [14]	28G	VRN-H1 Multiplex PCR	5H
Seasonal growth habit [15]	28G	HvVRNH1_SNP2	5H
Seasonal growth habit [16]	28G	HvVRNH1_SNP2 & HvVRNH1_hap2_InDel	5H

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 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 (Figure 30a). 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 (Figure 30b), 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) (Figure 30b-c). 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 (Figure 30b). 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 (Figure 30b-e), 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 (Figure 31a), 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 (Figure 31b). 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 (Figures 30e-f). 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 (Figure 30d). 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 (Figure 31d).

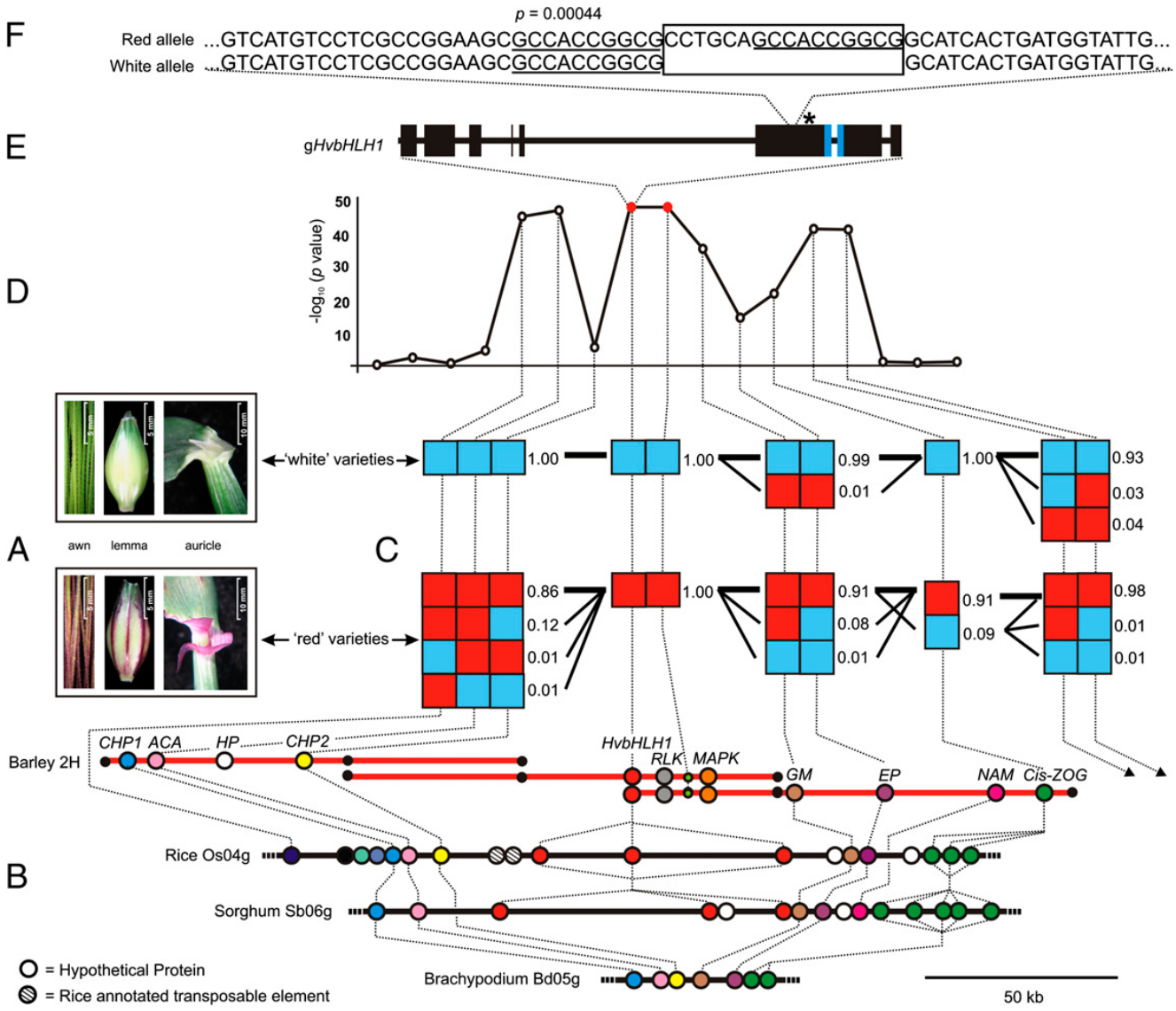


Figure 30. Fine-mapping of *ANT2*. **a.** Anthocyanin pigmentation in target tissues of cvs. Saffron ('white') and 'Retriever' ('red'). **b.** The sequenced physical map of the barley *ANT2* region, aligned to colinear regions of rice, sorghum and Brachypodium. **c.** Haplotype frequencies in 'red' and 'white' varieties. **d.** GWA scan for the trait, 'anthocyanin expression *per se*'. The two peak markers are highlighted in red. **e.** Structure of the *ANT2* candidate gene, *HvbHLH1*. The region encoding the predicted bHLH domain is highlighted in blue. The position of the premature stop codon in the 'white' allele is indicated by an asterisk. **f.** Partial DNA sequence of *HvbHLH1* exon 6, illustrating the position of the 16 bp InDel. The 10 bp repeat sequences flanking the deletion are underlined; the probability (*p*) that these locations are by chance is indicated.

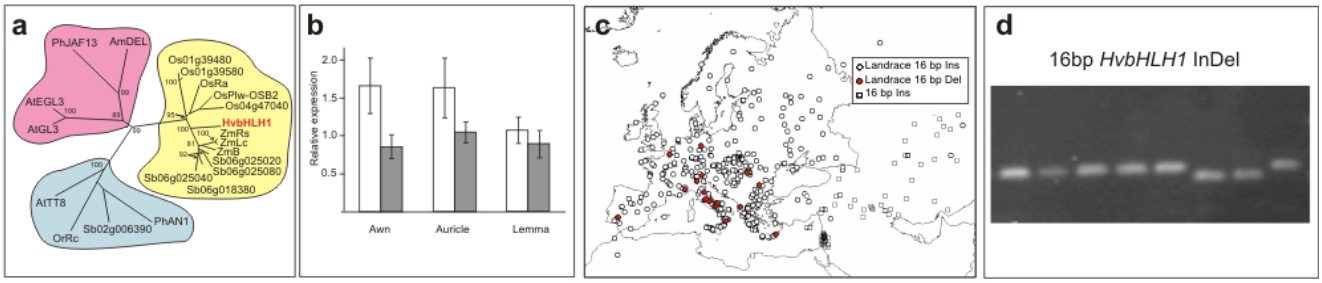


Figure 31a. Unrooted phylogenetic analysis of HvbHLH1 and additional bHLH proteins from rice, sorghum and maize, as well as proteins known to act within the anthocyanin pathways of dicotenous species. Bootstrap frequencies (1,000 replicates) > 80 % are indicated. **b.** Semi-quantitative RT-PCR analysis of *HvbHLH1* expression in tissues from the 'white' variety 'Saffron' and the 'red' variety 'Retriever', normalized against *ACTIN.* Error bars indicate ± 1 standard deviation. **c.** Geographic distribution of the *HvbHLH1* 16 bp InDel in 647 landrace and *H. vulgare* ssp. *spontaneum* accessions. **d.** PCR/agarose gel-based assay for the 16bp *HvbHLH1* InDel.

Origins of the 'white' ant2 allele

Sequence analysis suggests that the 'white' *ant2* allele is a mutated form of the wild-type 'red' allele, due to a 16 bp exonic deletion in *HvbHLH1* that results in a truncated predicted protein. This model is supported by the presence of identical 10 bp sequence motifs that flank the deletion (*p* repeat sequences are located by chance = 0.00044), a hallmark of illegitimate recombination following double-stranded DNA break repair (Figure 30f). To investigate the geographic origin of the 'white' allele, we screened 117 wild barley accessions and 471 predominantly European landraces (distinct locally adapted populations that predate formal crop improvement) for the 16 bp InDel in HvbHLH1. Although not found in wild barley, twenty landraces possessed the deletion, of which thirteen are located within Italy (Figure 31c).

Further analysis of Aleurone colour

As was the case for the majority of GWAs identified, we developed additional barley genetic markers were developed within the GWA peak for the DUS trait 'Aleurone colour' (Figure 29a), which were genotyped on a subset of 96 lines. Of these (HvOs03g14890, HvOs03g14380, HvOs03g14250, HvOs03g13590 and HvOs03g13450), SNPs within the HvOs03g14380 amplicon were found to show an association with the trait at least as good as the peak BOPA1 markers identified in the GWA scan (11_20453, 11_21087, 11_21296). We genotyped HvOs03g14380 across the complete association panel, finding it to be more predictive for phenotypic state than the peak BOPA1 markers. However, HvOs03g14380 did not provide an unambiguous result, as not only did it 'correct' 7 of the incorrect BOPA1 marker character state predictions (CSPs), it also resulted in 4 additional 'incorrect' CSPs. Analysis of Colinearity between rice, brachypodium and barley near the 3 peak BOAP1 markers (which all map at 62.1 cM on chromosome 4H) identifies an important ambiguity in synteny between the species, based on the three peak BOPA1 markers: the rice orthologue of barley marker 11_21296 is located 50 gene-models from peak marker

11_21087 and 63 gene-models from 11_20453. However, the brachypodium orthologue of 11_21296 appears to have undergone a translocation, placing it just 4 gene models and 2 gene models away from peak marks 11_21087 and 11_20453, respectively. Analysis of the gene content of the collinear regions of the sequenced genomes of rice and brachypodium identifies a number of candidate genes are located in the vicinity of all three barley markers (when ordered according to rice); therefore, resolving the physical location of 11_21296 in barley is important for the prioritization of barley candidate genes for further investigation.

Genotyping of known flowering time loci

While GWA scans for flowering time (based on historical DUS data) did not return significant associations using the Bonferronni corrected $P = 0.05$ significance threshold, peaks of association were nevertheless evident (Figure 29b), principally on chromosomes 2HS and 5HL. Four major flowering time loci have been identified in barley, controlling response to vernalization (*VRN-H1, VRN-H2, VRN-H3*) and photoperiod (*PPD-H1, PPD-H2*). Of these, *PPD-H1* is predicted to map close to the (insignificant) GWA peak on chromosome 2HS, while *VRN-H1* is close to the 5HL peak. As a basis from which future GWA scans could be performed using *de novo* flowering time data, we genotyped the five major flowering time loci across the complete public varietal association panel (as well as performing partial sequencing of an amplicon of the ear row-number locus, *VRS1*, spanning all three of the known mutations that result in conversion of a 2-row allele into a 6-row allele). *VRN-H1, VRN-H2 and PPD-H2* were genotyped by PCR/agarose-gel assays; the remaining genes were genotyped by sequencing. This data should prove valuable for future studies on flowering time, where the ability to standardize multi-locus haplotypes at the major flowering time loci should help the identification of novel loci in GWA scans for this highly heritable trait.

Conversion of selected BOPA1 and de novo SNPs to the KASPAR platform

It is of use to researchers and breeders to be able to genotype selected BOPA1 SNPs (as well as *de novo* SNPs identified during the marker enrichment process described above) in a flexible and easily scalable way. Towards this goal, we selected the KASPar platform, provided as a service by KBiosciences (http://www.kbioscience.co.uk/) as a suitable genotyping platform. It is based on their single-plex technology which dispenses with the need for the individually labeled fluorescent probes used in the Taqman system. Once validated, genotyping costs range from 17p or under per data point. A list of the 31 markers validated on KASPar using a panel of 90 barley varieties is shown in Table 28. For each of the selected gene-based markers, one or more polymorphic DNA features were assayed, with the following minimum details recorded for each: (a) DNA sequence, with the targeted polymorphism identified using standard nomenclature (b) The GenBank accession number for the DNA sequence of the reference allele (c) a PMID number, linking to the relevant scientific publication describing the allelic variants (d) genetic map position of the gene

assayed (e) Information describing the SNPs and their associated phenotypes. The calling rate was high for these assays with less than 1% missing data.Information for validated KASPar assays (Figure 32) are to be made accessible via web-pages (under construction) hosted at the NIAB website (http://cropmarkers.niab.com/).

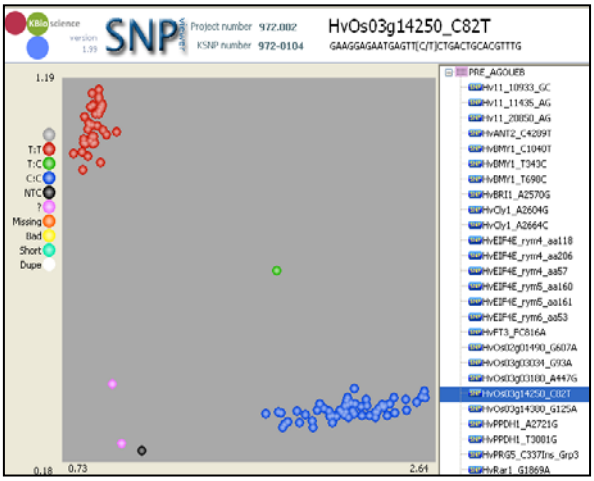


Figure 32. Visualisation of SNP data generated from 'aleurone colour' assay HvOs03g14250_C82T, using SNP Viewer (KBiosciences). The alternative SNP genotypes are clearly distinguishable (T:T = red, C:C = blue), with heterozygous individuals (T:C = green) unambiguously clustered in a separate cloud. The water negative control is shown in black, while unknown calls (predicted to represent wells which lack DNA) are shown in pink.

3.4.4. VCU data (Jordi Comadran)

Population structure:

Population structure and the need to correct and account for its effects when using highly stratified germplasm for association mapping have guided most of the research on GWAS for the last few years (Mackay & Powell, 2007). Several statistical approaches have been proposed in the literature (genomic control, false discovery thresholds, a long list of mixed models accounting for multiple levels of relatedness,…), the biggest problem being the inconsistency amongst the number and identity of the markers that remain significant after population structure correction.

Alongside with statistical development, research focused on population development searching for tools combining high statistical power, high resolution and low population stratification. Heterogenic stocks, known as well as multi-parental advanced generation intercross populations (MAGIC) have been proposed to overcome the handicaps imposed by the strong stratification of the germplasm. Genetic mapping of traits using MAGIC "should" be more complete due to greater genetic diversity and more precise than classical bi-parental populations. Short history of recombination in the *de novo* population development gives high statistical power to detect QTL, while ancestral recombination and diversity accumulated between the parental lines would provide the basis for fine-mapping. The few rounds of inter-crossing will be sufficient to remove long range linkage

disequilibrium present between the parental lines. However, in the creation of MAGIC populations, in order to avoid the creation of small clusters of highly related progenies, thus introducing *de novo* germplasm stratification, complex and time-consuming crossing schemes have to be implemented.

Artificial out-crossing imposed by the breeders coupled with the long recombination history of the germplasm can create a highly diverse germplasm stock without major population sub-divisions. Here we propose the use the elite two-rowed spring barley cluster as an heterogenic stock without strong population stratification. Thus, avoiding complex statistical approaches to deal with population structure effects, avoiding complex and time-consuming crossing schemes and most important perform QTL analysis and discovery in a germplasm set directly related to the breeding pool. Objectives are (1) to check diversity and possible stratification of the germplasm and as an example (2) to perform and present some GWAS of a highly heritable trait related to yield such as thousand kernel weight or height.

The phylogenetic relationship between genotypes and pairwise marker to marker relationships were used as complementary approaches to explore the possible presence of genetically distinct homogeneous clusters, which may raise population structure issues during the analysis.

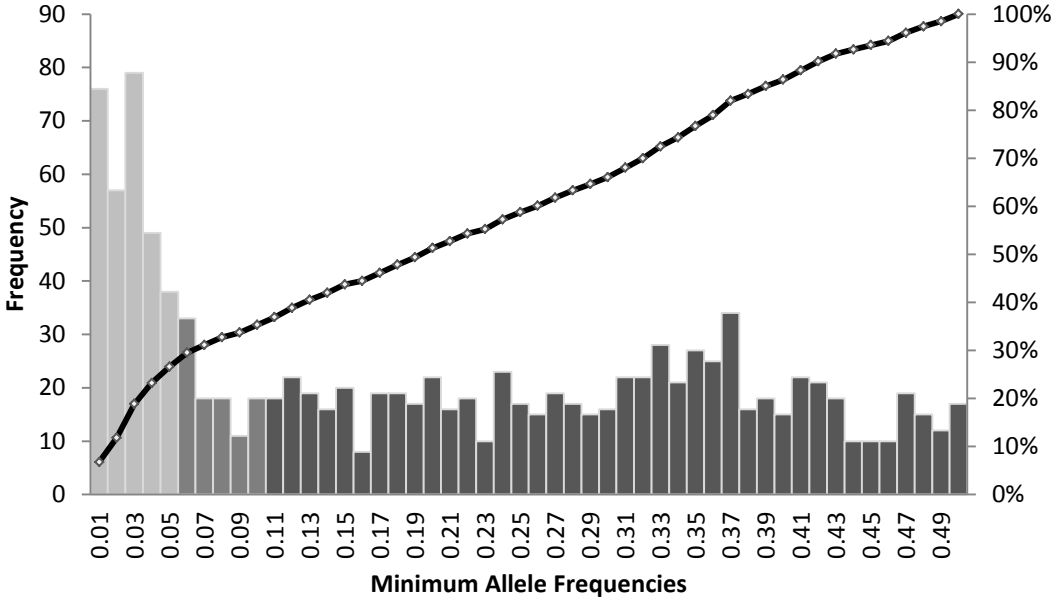


Figure 33. Genetic diversity of 1126 bOPA1 SNPs across the elite spring barley dataset.

The assembled 648 two-rowed elite spring barley dataset for the study consisted of AGOUEB UK public elite spring lines complemented with AGOUEB “progenitors” and ExBARDIV older North European spring barley landraces featuring strong in the pedigrees of UK germplasm, all of them genotyped with bOPA1 1536 SNP markers. 410 SNPs with more than 10 % of missing allele calls

were removed from the dataset and omitted from further analyses. Figure 33 summarizes the genetic diversity of bOPA1 in the spring set. 27 % of the SNP markers have MAF lower than 0.5 %.

Phylogenetic Relationships

Low recombination in heterochromatic regions can bias the identification of germplasm sub-populations. As an example, a large haplotype block concerning the centromeric region of chromosome 3H, where >30 co-segregating SNP markers in the 56.4 cM delimitate a region of at least 542 and 982 gene models in *Brachipodium* and *Oryza* respectively. In practical terms that means that it would be possible to detect germplasm groupings that reflect centromeric haplotypes rather than true germplasm differentiation. Thus, for the study of population structure and sub-structure, 'redundant' SNPs mapping to the same chromosome position were removed from the dataset to avoid bias of population partitioning of the germplasm towards low recombination genomic regions such as centromers. SNP markers with minimum allele frequencies lower than 10 % were also omitted from further analyses.

The final dataset for the study of the phylogenetic relationship between genotypes encompassed 486 informative non map position-redundant SNPs. It is important to note, though, that the removal of map 'position-redundant' SNPs is based on the SNP consensus map and does not reflect the recombination and diversity patterns observed for those SNPs in the association panel.

The phylogenetic relationships among the 648 selected barley cultivars was evaluated by generating a neighbour-joining population tree based on simple matching allelic distances as implemented in DARwin (Perrier & Jacquemoud-Collet 2006) (Figure 34A). The resulting neighbour-joining tree did not show any signs of clustering. The pair-wise simple matching allelic distances between genotypes (Figure 34B) showed an average genetic distance of 0.63. It is also important to note that less than 1 % of the pair-wise genetic distances had values higher than 0.8 and those lines did not cluster together. Following, principal coordinate analysis, PCO, based on simple matching of SNP alleles was performed with Genstat 13 (Figure 34C). The PCO analysis supported the same conclusions as the phylogenetic analysis with no observable population structure present in the dataset.

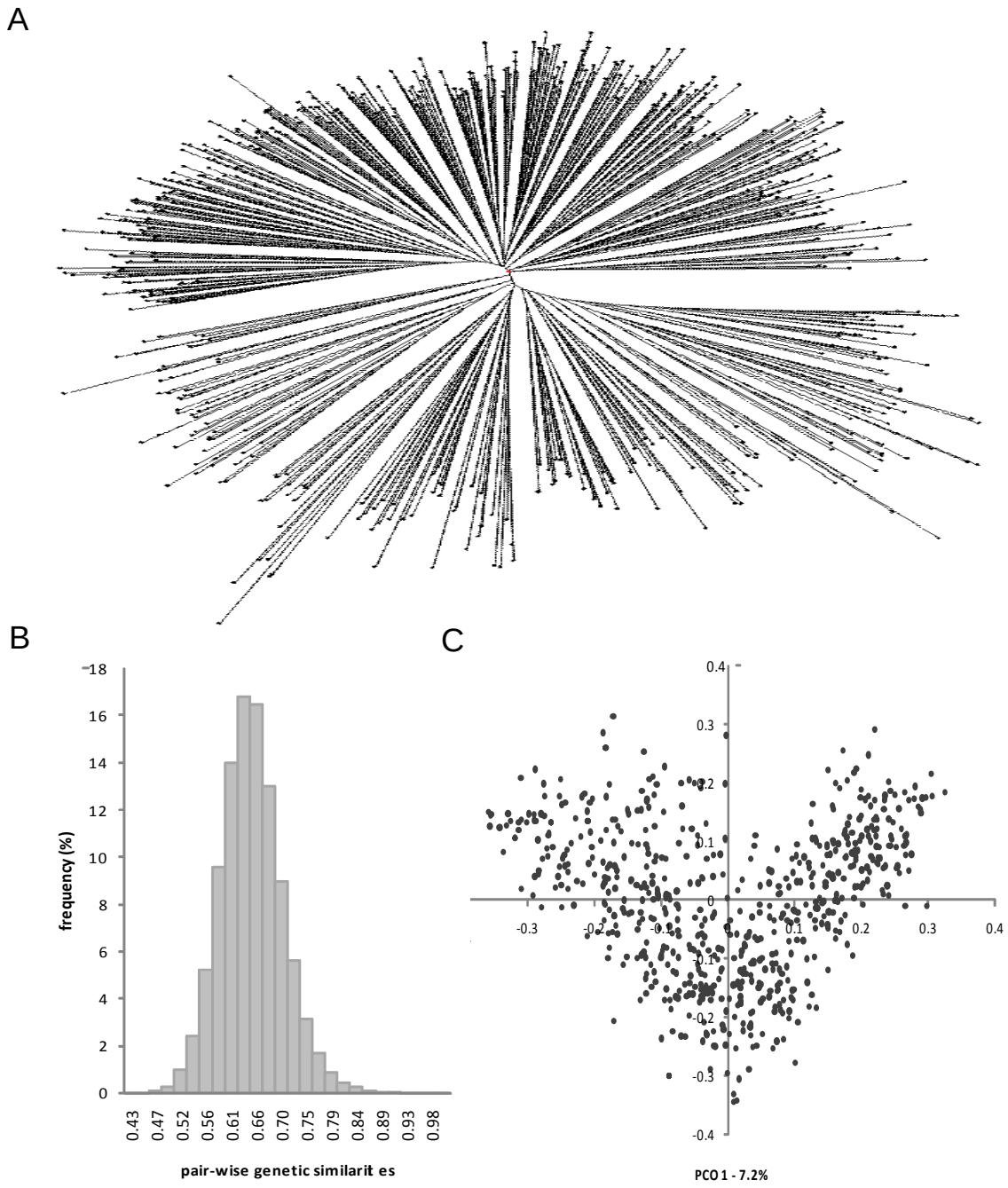


Figure 34. Population structure within the elite two-row spring barley gene-pool (A) Neighbour joining tree of 648 barley cultivars constructed from simple matching distance of 486 non position-redundant SNPs. (B) Pair-wise similarities between barley cultivars. (C) PCO plots of the first two components of 648 barley cultivars.

Following, we used the Bayesian approach implemented in the program STRUCTURE (Pritchard *et al.* 2000a; Pritchard *et al.* 2000b) (*K* = 2-5, 3 replicates, burn-in period = 10,000, MCMC iterations = 10,000) to obtain fractional memberships of the 648 UK barley cultivars to varying numbers of *K* subpopulations. The objective was to "force" STRUCTURE software to allocate the accessions to a varying number of putative "cryptic" sub-populations, so that we could compare the

Fst values of the putative sub-populations and assess the level of genetic differentiation amongst them. Group assignment for each *K* was consistent across STRUCTURE run replicates, thus only Fst value per run is shown in Table 29. The low Fst values observed amongst the several putative *K* clusters is indicative of frequent gene flow and low sub-population genetic differentiation and little population stratification within the elite spring gene-pool and suggests that there are no cryptic populations.

Table 29. Population-differentiation statistic (Fst).

K	Fst_1	Fst_2	Fst_3	Fst_4	Fst_5
2	0.1215	0.3426	-	-	-
3	0.2864	0.418	0.2106	-	-
4	0.2528	0.4282	0.3552	0.3223	-
5	0.3359	0.3692	0.3948	0.4306	0.4319

Pairwise marker to marker relationships:

The main source of false positives in association mapping panels is supposed to be strong linkage disequilibrium (LD) between markers across the genome, largely caused by the coupled effect of sub-population genetic differentiation and isolation. Thus, when a QTL which is segregating within a population matches the sample stratification it is not possible to distinguish between true and false marker:trait associations.

Although phylogenetic, PCO and STRUCTURE analysis of the two row spring panel did not show any signs of population stratification we decided to check whether there was significant long range LD between SNP markers (Figure 35).

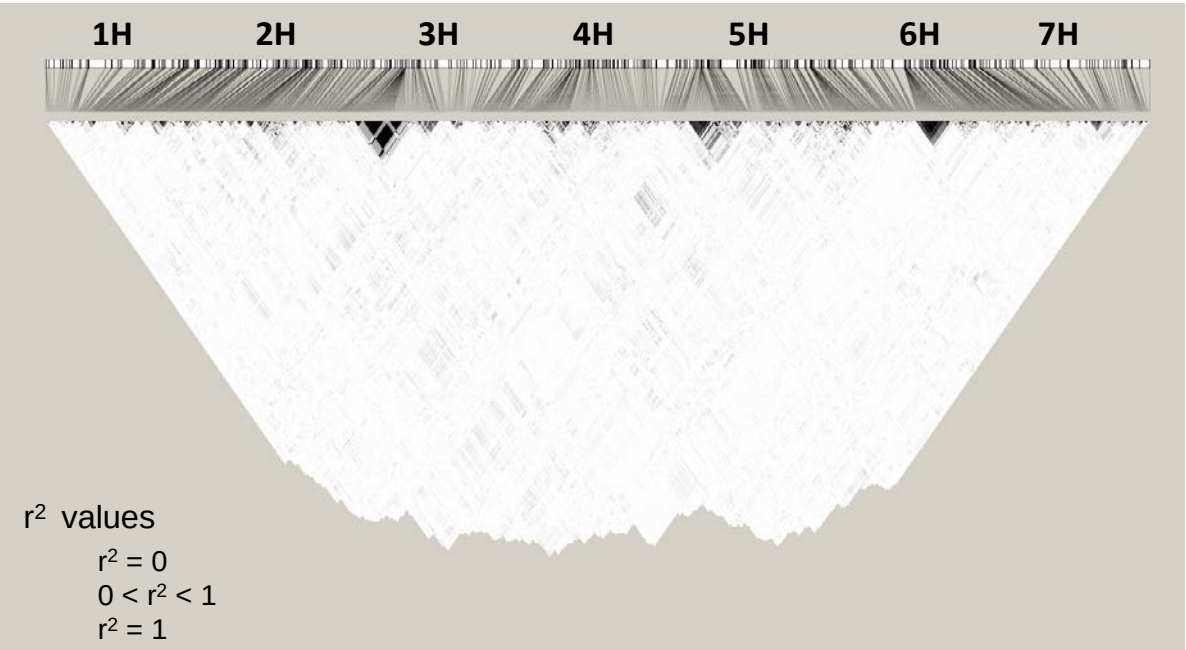


Figure 35. Whole genome r^2 Linkage Disequilibrium SNP pair-wise measures.

There was no reason to exclude the map position redundant SNPs and the analysis was performed with 827 SNP markers with minimum allele frequencies > 10 % and missing data < 10 %. We explored whole genome patterns of LD by using the classic LD algorithm (r^2) as implemented by HAPLOVIEW.

Heatmap charts of the distribution of intra and inter-chromosomal r^2 values across the barley genome chromosome highlight the extended LD values across the genetic centromeres (Figure 35). High LD extends outwards from these regions along the spine of each chromosome forming an axis of blocks of short-range LD, which we believe is caused by genetic linkage. We did not observe signs of long range intra and inter-chromosomal LD that commonly results from population sub-structure and admixture within a germplasm set.

This analysis supported the same findings as the phylogenetic and PCO analyses. Absence of population structure facilitates genome-wide association studies.

Genome-wide association. Preliminary analyses with BOPA1.

For this example we used over 200 lines from the AGOUEB Public Set with historical yield, height and thousand kernel weight data respectively. 811 mapped BOPA1 SNP loci with minimum allele frequencies > 10 % and missing data < 10 %were used for a genome wide association (GWA) scan for yield, height and thousand kernel weight using a mixed model were SNP data were fitted as fixed effects. Figures 36 and 37 as examples (below).

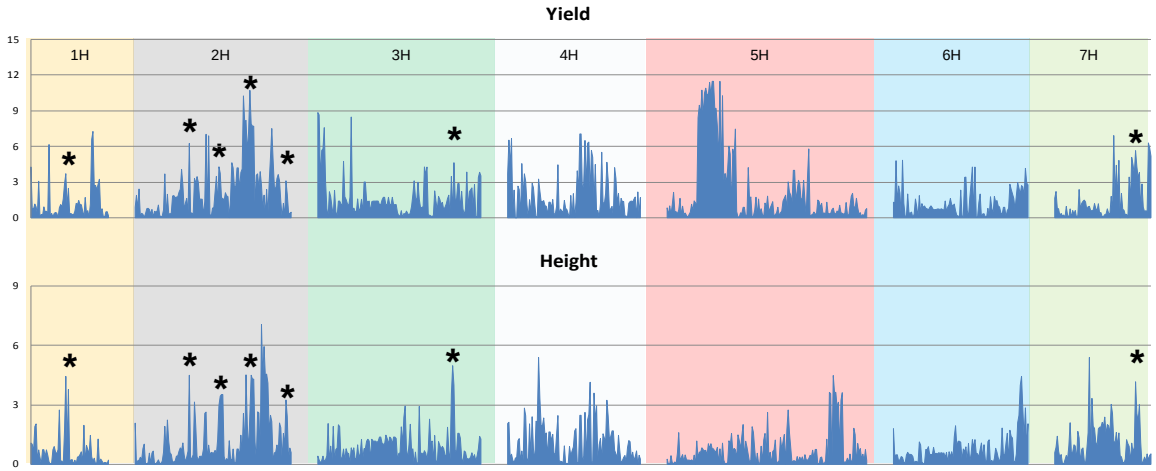


Figure 36. Genome-wide association scans for yield and height. –log10 [fp values] are plotted following chromosomal order and may not reflect genetic distances. Asterisks (*) indicate SNPs both significant for yield and height. We observe opposite signs of the effects for yield and height related to those SNPs suggesting that effects on those loci linked to shorter plants are related to bigger yields.

As an example of a known gene under one of the association peaks, significant height QTL on chromosome 3H at 126 cM is closely linked to the green revolution gene sdw1.

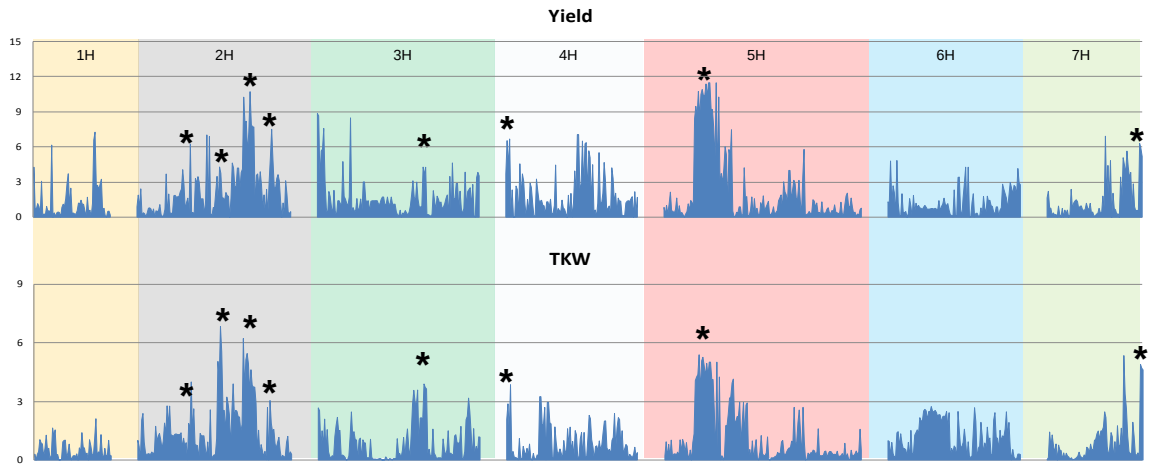


Figure 37. Genome-wide association scans for yield and thousand kernel weight (TKW). –log10 [fp values] are plotted following chromosomal order and may not reflect genetic distances. Asterisks (*) indicate SNPs both significant for yield and height. Same effect direction was observed for all the QTL indicating that bigger grains are correlated with bigger yields.

Genetic nature of the QTL

If one thing is clear is that complex traits are very complex. We were interested to check whether the identified QTL were related to additive or epistatic interactions. A mixed model was fitted to check epistatic interactions of 811 SNP markers with each other. Each SNP combinations was fitted in the random term as SNP1 + SNP2 + SNP1*SNP2, thus removing first the additive effects of both SNP markers to assess the true significancy of the interaction.

Checking the top 10 TKW QTL from the single SNP model, none of them showed significant epistatic interactions with each other suggesting that their effects are mainly additive. In addition to that, we observed significant epistatic SNP*SNP interactions being of special interest (-log10[fp values] > 6) those between the distal end of chromosome 5H and short arm of chromosome 2H, the long arm peri-centromeric region of chromosome 3H and the long arm of chromosome 4H. SNP*SNP epistatic interactions for TKW are summarized in Figure 38.

Figure 38. Most significant SNP markers showing SNP.SNP interactions for thousand kernel weight. Legend: red, green and blue for –log10[fp values] over 4, 5 and 6 respectively.

These results are of special interest considering future implementations of genomic selection tools.

3.5. References

Astle W, Balding DJ (2009) Population structure and cryptic relatedness in genetic association studies. *Stat Sci* **24**: 451–471.

Balding DJ (2006) A tutorial on statistical methods for population association studies. *Nat Rev Genet* **7**: 781-791.

Chen, A., U. Baumann, G. B. Fincher, and N. C. Collins, 2009a Flt-2L, a locus in barley controlling flowering time, spike density, and plant height. *Funct.Integr.Genomics* **9**: 243-254.

Chen, A., J. L. Reinheimer, A. Brûlé-Babel, U. Baumann, G. B. Fincher et al. 2009b Genes and traits associated with barley 2H and 5H chromosome regions controlling sensitivity of reproductive tissues to frost. *Theor Appl Genet* **118**: 1465-1476.

Close et al., (2009

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, Diana L. Zuluaga DL, David Smith D, Comadran J, Malcolm Macaulay M, Zewei Luo Z, Kearsey MJ, Werner P, Harrap D, Tapsell C, Liu H, Hedley PE, Stein N, Schulte D, Steuernagel B, Marshall DF, Thomas WTB, Ramsay L, Mackay I, Balding DJ, The AGOUEB Consortium, Waugh R, O'Sullivan DM (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.

Comadran J, Thomas WTB, van Eeuwijk FA, Ceccarelli S, Grando S, Stanca AM, Pecchioni N, Akar T, Al-Yassin A, Benbelkacem A, Ouabbou H, Bort J, Romagosa I, Hackett CA, Russell JR (2009). Patterns of genetic diversity and linkage disequilibrium in a highly structured Hordeum vulgare association-mapping population for the Mediterranean basin. *Theoret Appl Genet* **119**: 175-187.

Devlin B, Roeder K (1999) Genomic control for association studies. *Biometrics* **55**: 997-1004.

Druka A, Franckowiak J, Lundqvist U, Bonar N, Alexander J, Houston K, Radovic S, Shahinnia F, Vendramin V, Morgante M, Stein N, Waugh R (2011). Genetic Dissection of Barley Morphology and Development. *Plant Physiol* **155**: 617-627

Falush D, Stephens M, Pritchard JK (2003) Inference of Population Structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* **164**: 1567–1587.

Francia, E., D. Barabaschi, A. Tondelli, G. Laido, F. Rizza 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.

Francia, E., F. Rizza, L. Cattivelli, A. M. Stanca, G. Galiba et al. 2004 Two loci on chromosome 5H determine low-temperature tolerance in a 'Nure' (winter) x 'Tremois' (spring) barley map. *Theor.Appl.Genet*. **108**: 670-680.

Gustafsson, Å., A. Hagberg, and U. Lundqvist, 1960 The induction of early mutants in Bonus barley. *Hereditas* **46**: 675-699.

Gustafsson, Å., and U. Lundqvist, 1976 Controlled environment and short-day tolerance in barley mutants., pp. 45-53 in Induced Mutants in Cross-breeding. Proc. Advisory Group., edited by International Atomic Energy Agency. Vienna.

Hanemann, A, Schweizer, GF, Cossu, R, Wicker, T, Roeder, MS (2009) Fine mapping, physical mapping and development of diagnostic markersfor the Rrs2 scald resistance gene in barley. *Theoret Appl Genet* **119**: 1507-1522

Jing RC; Vershinin, A.; Grzebyta, J.; Shaw, P.; Smykal, P.; Marshall, D.; Ambrose, M. J.; Ellis, T. H. N.; Flavell, A. J. (2010) The genetic diversity and evolution of field pea (Pisum) studied by high throughput retrotransposon based insertion polymorphism (RBIP) marker analysis. *BMC Evol Biol* **10**: 44.

Kang HM, Zaitlen NA, Wade CM, Kirby A, et al. (2008) Efficient control of population structure in model organism association mapping. *Genetics* **178**: 1709–1723.

Kikuchi, R., H. Kawahigashi, T. Ando, T. Tonooka, and H. Handa, 2009 Molecular and functional characterization of PEBP genes in barley reveal the diversification of their roles in flowering. Plant Physiol **149**: 1341-1353.

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.

Kopecky, D.; Bartos, J.; Lukaszewski, A. J.; Baird, J. H.; Cernoch, V.; Kolliker, R.; Rognli, O. A.; Blois, H.; Caig, V.; Lubberstedt, T.; Studer, B.; Shaw, P.; Dolezel, J.; Kilian, A. (2009). Development and mapping of DArT markers within the Festuca-Lolium complex. *BMC Genomics* **10**: 473

Lloyd M E (1982). Selecting systems to measure client outcome in human service agencies. *Behavioral Assessment* **5**: 55-70.

Mackay, I; Powell, W (2007) Methods for linkage disequilibrium mapping in crops. *Trends Plant Sci* **12**: 57-63.

Malysheva-Otto, LV; Ganal, MW; Roder, MS (2006) Analysis of molecular diversity, population structure and linkage disequilibrium in a worldwide survey of cultivated barley germplasm (Hordeum vulgare L.). *BMC Genetics* **7**: 6.

Newton AC, Searle J, Guy DC, Hackett CA, Cooke DEL, 2001. Variability in pathotype, aggressiveness, RAPD profile, and rDNA ITS1 sequences of UK isolates of Rhynchosporium secalis. Journal of Plant Diseases and Protection **108**, 446-58.

Peakall, R., and P. E. Smouse, 2006 GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Mol. Eco. Notes* **6**: 288-295.

Pelleg D, Moore A (2000) X-means: extending k-means with efficient estimation of the number of clusters. In: Proceedings of the seventeenth international conference on machine learning. Morgan Kaufmann, San Francisco, pp 727–734.

Perrier X. & Jacquemoud-Collet J.P. DARwin software http://darwin.cirad.fr/darwin. 2006.

Piffanelli, P, Ramsay, L, Waugh, R, Benabdelmouna, A, D'Hont, A, Hollricher, K, Jorgensen, JH, Schulze-Lefert, P, Panstruga, R (2004) A barley cultivation-associated polymorphism conveys resistance to powdery mildew. *Nature* **430**: 887-891

Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, et al. (2006) Principle component analysis corrects for stratification in genome-wide association studies. *Nat Genet* **38**: 904–909.

Pritchard, J. K., M. Stephens, and P. Donnelly, 2000a Inference of population structure using multilocus genotype data. Genetics **155**: 945-959.

Pritchard, J. K., M. Stephens, N. A. Rosenberg, and P. Donnelly, 2000b Association mapping in structured populations. Am.J.Hum.Genet. **67**: 170-181.

Reich DA, Goldstein DB (2001) Detecting association in a case-control study while correcting for population stratification. *Genet Epid* **2**: 4–16.

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., Grainer, 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

Storey JD, Tibshirani R (2003) Statistical significance for genome-wide studies. *Proc Natl Acad Sci USA* **100**: 9440–9445.

Szucs P, Skinner JS, Karsai I, Cuesta-Marcos A, Haggard KG, Corey AE, Chen THH, Hayes PM (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 RP, Laurie DA (2005) The pseudo-response regulator Ppd-H1 provides adaptation to photoperiod in barley. *Science* **310**: 1031-1034.

Van Ooijen JW (2006) Joinmap®4, Software for the calculation of genetic linkage maps in experimental populations. Kyazma B.V., Wageningen, The Netherlands.

von Zitzewitz J, Szucs P, Dubcovsky J, Yan LL, Francia E, Pecchioni N, Casas A, Chen THH, Hayes PM, Skinner JS (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.

Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M, et al. (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet 38: 203–208.