

Research article

Open Access

Association mapping of partitioning loci in barley

James Cockram^{†1}, Jon White^{†1}, Fiona J Leigh¹, Vincent J Lea¹,
Elena Chiapparino², David A Laurie³, Ian J Mackay¹, Wayne Powell¹ and
Donal M O'Sullivan^{*1}

Address: ¹Centre for Plant Genetics, Breeding and Evaluation, National Institute of Agricultural Botany, Cambridge, UK, ²CRA – Unità di Ricerca, Sant'Angelo Lodigiano (Lodi), Italy and ³Crop Genetics, John Innes Centre, Norwich, UK

Email: James Cockram - james.cockram@niab.com; Jon White - jon.white@niab.com; Fiona J Leigh - fiona.leigh@niab.com; Vincent J Lea - vincelea@btinternet.com; Elena Chiapparino - elenachia@yahoo.com; David A Laurie - david.laurie@bbsrc.ac.uk; Ian J Mackay - ian.mackay@niab.com; Wayne Powell - wayne.powell@niab.com; Donal M O'Sullivan* - donal.osullivan@niab.com

* Corresponding author †Equal contributors

Published: 18 February 2008

Received: 11 September 2007

BMC Genetics 2008, 9:16 doi:10.1186/1471-2156-9-16

Accepted: 18 February 2008

This article is available from: <http://www.biomedcentral.com/1471-2156/9/16>

© 2008 Cockram et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: Association mapping, initially developed in human disease genetics, is now being applied to plant species. The model species *Arabidopsis* provided some of the first examples of association mapping in plants, identifying previously cloned flowering time genes, despite high population sub-structure. More recently, association genetics has been applied to barley, where breeding activity has resulted in a high degree of population sub-structure. A major genotypic division within barley is that between winter- and spring-sown varieties, which differ in their requirement for vernalization to promote subsequent flowering. To date, all attempts to validate association genetics in barley by identifying major flowering time loci that control vernalization requirement (*VRN-H1* and *VRN-H2*) have failed. Here, we validate the use of association genetics in barley by identifying *VRN-H1* and *VRN-H2*, despite their prominent role in determining population sub-structure.

Results: By taking barley as a typical inbreeding crop, and seasonal growth habit as a major partitioning phenotype, we develop an association mapping approach which successfully identifies *VRN-H1* and *VRN-H2*, the underlying loci largely responsible for this agronomic division. We find a combination of Structured Association followed by Genomic Control to correct for population structure and inflation of the test statistic, resolved significant associations only with *VRN-H1* and the *VRN-H2* candidate genes, as well as two genes closely linked to *VRN-H1* (*HvCSFs1* and *HvPHYC*).

Conclusion: We show that, after employing appropriate statistical methods to correct for population sub-structure, the genome-wide partitioning effect of allelic status at *VRN-H1* and *VRN-H2* does not result in the high levels of spurious association expected to occur in highly structured samples. Furthermore, we demonstrate that both *VRN-H1* and the candidate *VRN-H2* genes can be identified using association mapping. Discrimination between intragenic *VRN-H1* markers was achieved, indicating that candidate causative polymorphisms may be discerned and prioritised within a larger set of positive associations. This proof of concept study demonstrates the feasibility of association mapping in barley, even within highly structured populations. A major advantage of this method is that it does not require large numbers of genome-wide markers, and is therefore suitable for fine mapping and candidate gene evaluation, especially in species for which large numbers of genetic markers are either unavailable or too costly.

Background

Determining the genetic basis of economically important complex traits is a major goal of plant breeding, and has largely been accomplished using linkage mapping of quantitative trait loci (QTL). However, focus is now turning towards the use of association mapping (recently reviewed by [1]), initially applied in human disease genetics. Both approaches rely on the strength of associations between genetic markers and phenotype. However, while linkage analysis searches for associations within populations developed from bi-parental crosses, association mapping utilizes historic patterns of recombination that have occurred within a sample of individuals (e.g. a collection of varieties, landraces or breeders' lines). This has the advantage of allowing existing collections to be screened for many different phenotypes, as well as taking advantage of historical phenotype data from lines thoroughly characterized during variety development. Association mapping is based on the principle that over multiple generations of recombination, correlations only with markers tightly linked to the trait of interest will remain. As a predominantly inbreeding species, cultivated barley is an attractive target for association mapping as its genome contains extensive blocks of chromatin in linkage disequilibrium (LD) [2,3], providing a well-defined haplotype structure from which marker-trait associations can be identified. However, spurious associations between genotype and trait may be detected due to the degree of structure or subdivision within the population, necessitating development of various statistical methods to account for population structure (recently reviewed by [4]).

One of the major divisions within barley germplasm is the distinction between vernalization sensitive and insensitive varieties. Plants with a vernalization requirement require a prolonged period of cold (vernalization) in order to promote subsequent flowering. This is characteristic of the wild ancestors of temperate crops such as barley and wheat. During the spread of agriculture, human selection has resulted in the partitioning of the majority of barley germplasm into spring- and winter-sown varieties which lack or retain a vernalization requirement, respectively. The molecular genetics of vernalization requirement in barley is relatively well characterised (recently reviewed by [5]), and is controlled predominantly by two major loci: *VRN-H1* and *VRN-H2* [6]. Spring alleles are thought to be due to deletions spanning putative *cis*-elements in *VRN-H1* intron I [7-10], or by deletion of part or all of the genomic region carrying the *VRN-H2* candidate genes [7,11,12]. These well documented genetic determinants of the phenotype make the genetics of vernalization response an attractive test-bed for association mapping in barley.

Spurious associations between genotype and trait due to population sub-structure is widely recognised as a serious obstacle to association mapping. This is likely to be of particular significance in barley, as diversity studies have shown barley germplasm to be highly partitioned, predominantly due to vernalization requirement (spring or winter growth habit) and ear row-number [3,13,14]. Highly stratified populations result in differences in allelic frequency between sub-populations, which may result in false associations between genotype and phenotype in uncorrected regression analysis. Even if association mapping panels are selected within one of these phenotypic classes, sub-structure is still likely to be present due to other factors such as geographic origin and related pedigree. Some of the first association mapping studies in plants were conducted in the model species *Arabidopsis*, and demonstrated the ability to identify previously characterised flowering time loci involved in local adaptation, and hence correlated with population structure [15,16]. As a first step in the application and validation of association genetics to barley, previous studies attempted to identify *VRN-H1* and *VRN-H2* [3,17]. However, to date, all such association studies have failed to identify these loci, despite their well characterised genetics and robust phenotypes. Here, we apply a combination of Structured Association (SA) [18] to correct for population sub-structure and Genomic Control (GC) [19] to correct for residual inflation of the test statistic due to unidentified population structure effects and residual confounding arising, for example, from close pedigree relationships among varieties. GC alone is very effective at controlling false positive rates but can result in loss of power in the presence of large effects of population structure [20]. SA will account for population substructure, but is less effective at adjusting for genetic relationships within subpopulations. In addition, identifying the appropriate number of subpopulations for SA analysis can be problematic. We reasoned that the combination of the two methods would negate the deficiencies in each. Although other approaches to association genetics exist, in particular the mixed model approach encoded in the software TASSEL [21] and the principal components method EIGENSTRAT [22], these methods require many more background genetic markers than were available for this study. Using SA + GC, we successfully demonstrate association mapping of *VRN-H1* and *VRN-H2* in a collection of European barley germplasm, despite the prominent role of these genes in defining germplasm structure, yet we maintain a false positive frequency close to the expected value elsewhere in the genome.

Results

Genome-wide Marker Profiling

In order to aid subsequent population structure and association analysis, a combination of Sequence-Specific

Amplification Polymorphism (S-SAP), SSR and gene-based markers were screened across the complete varietal collection. S-SAP techniques generated 129 polymorphic markers, giving a mean S-SAP minor allele frequency (M.A.F.) of 0.22, close to the expected value of 0.25 given a uniform distribution. In addition, three *VRN-H1* region SSRs were screened (*EBmatc0003*, *GMS027* and *Bmag0222*). Four alleles were identified for *EBmatc0003*, with a heterozygosity value of 0.504. *GMS027* (12 alleles) and *Bmag0222* (5 alleles) displayed heterozygosity values of 0.62 and 0.64, respectively. The remaining SSRs are distributed five per chromosome, and are described in detail by [14]. One of these, *Bmag0135* (chromosome 7H), was removed from the final analysis due to the high level of missing data. InDels amplified from *HvCSFs1* and *HvPHYC* were also scored across the varietal set. Interestingly, 5 % of barley varieties were heterozygous for the *HvPHYC* assay, although genome wide markers did not indicate heterozygosity to be present elsewhere in the genomes of these lines. Previous studies have shown that two *HvPHYC* pseudogenes (*HvPHYCΨ1a* and *HvPHYCΨ1b*) map to distinct genetic loci relative to the functional *HvPHYC* gene, and represent duplications of the 5' end of the gene [23]. However, the assay employed here is based on primers at the 3' end of the gene, suggesting the presence of both InDel forms may be due to detection of an additional duplication event specific to these varieties.

Extent of LD on Chromosomes 4H and 5H

Given the haplotype evidence for a local genetic bottleneck caused by a reliance on a single *VRN-H1/VRN-H2* allelic configuration conferring winter GH [9], a selection of markers spanning appropriate regions of chromosomes 4H and 5H were used to assess any constraint in decay of LD between the two germplasm pools. In order to estimate LD decay, markers were integrated into the barley consensus map of [24]. *VRN-H1* and the candidate *VRN-H2* genes *ZCCT-Ha*, *-Hb* and *-Hc* were mapped to chromosomes 5HL and 4HL, respectively (Figure 1), as previously shown [7,11,23,25]. In addition, *HVM67* mapped 3 cM proximal to the *ZCCT-H* gene cluster, while *HvCSF5s1* and the SSR marker *EBmatc0003* both mapped to chromosome 5HL, 1 and 5 cM distal to *VRN-H1*, respectively. Matrices of r^2 were derived separately for spring and winter varieties using bi-allelic markers only (Figure 2). As expected, the levels of LD on chromosome 4H reveal strong associations between the three closely linked *ZCCT-H* genes ($r^2 = 1.0$), in both the spring and winter subpopulations. However, LD between the *ZCCT-H* locus and flanking markers is only found to extend to *Bmy1*, 6.4 cM distal: the *Bmy1* single nucleotide polymorphism (SNP) T+698/C is invariant in the winter population, while C+1,040/T is in LD with the *ZCCT-H* locus in winter ($r^2 = 0.313$) varieties. No other significant levels of LD

were identified on chromosome 4H. The LD matrix for chromosome 5H shows very strong levels of LD within *VRN-H1* (for which ten markers were assayed, see Methods) in winter varieties ($r^2 = 1$), with 3 invariant markers (*VRN-H1-SNP1*, *VRN-H1-SNP3* and *VRN-H1-intronI-St*). However, breakdown of LD within *VRN-H1* can be seen in spring varieties, although high degrees of correlation are still found (eg. $r^2 = 1$ between *VRN-H1-SNP6* and *VRN-H1-42bp-InDel*). Interestingly, both *HvPHYC* (2 cM proximal to *VRN-H1*, invariant in winter cultivars) and *HvCSFs1* (0.7 cM distal to *VRN-H1*, $r^2 = 1$) are in strong LD with *VRN-H1* markers in the winter sub-population, although this breaks down (particularly in the case of *HvCSFs1*), in the spring population. Evidence of LD is also seen on the short arm of chromosome 5H, in spring and winter cultivars (between *HVM30* and *BARE1-5M59q*), with *GBMS0032* also showing LD with both *HVM30* and *BARE1-5M59q* in winter germplasm.

Population Structure and Association Mapping

To help address false association between markers and phenotype, population substructure within the barley varietal collection was initially investigated using Principal Coordinate Analysis (PCoA). Scatterplots showed distinct clustering according to GH (winter/spring) and row-number (2-row/6-row), identifying four main subgroups: (i) spring 2-row (ii) winter 2-row (iii) spring 6-row (iv) winter 6-row (Figure 3). Some noteworthy outliers can be seen: (i) Three 6-row spring varieties ('Albert', 'Kustaa' and 'Vankkuri') are found within the relatively compact 2-row spring group. However, a search of the ECP/GR European Barley Database [26] returns conflicting row-number data for independent entries for each of the above varieties, suggesting that these varieties may contain passport data errors and require phenotypic verification. (ii) The winter varieties form a notably compact group, with the exception of two outliers ('Almunia' and 'Birgit'). These are found within, but towards the edge of, the spring cluster and represent two of just four winter varieties ('Athene', 'Almunia', 'Birgit' and 'Express') which possess the rare winter *VRN-H1* haplotype 5C, with the remaining winter varieties belonging to haplotype 1A [9]. The confirmation of winter GH in lines containing *VRN-H1* haplotype 5C [9] suggests that despite a significant proportion of their genetic background being of spring origin, a winter phenotype has been retained. (iii) Varieties belonging to *VRN-H1* haplotype 1B form a sub-division within 6-row spring barley, clustering towards the winter cloud and may reflect the close pedigree relationships between members of this haplotype (Additional File 1). Phenotypic analysis of two varieties initially classified as facultative ('Candela' and 'Orria') under long-day photoperiod and non-vernalizing temperatures indicate they possess a spring GH (data not shown). Interestingly, both clustered within the winter group, suggesting they have a predomi-

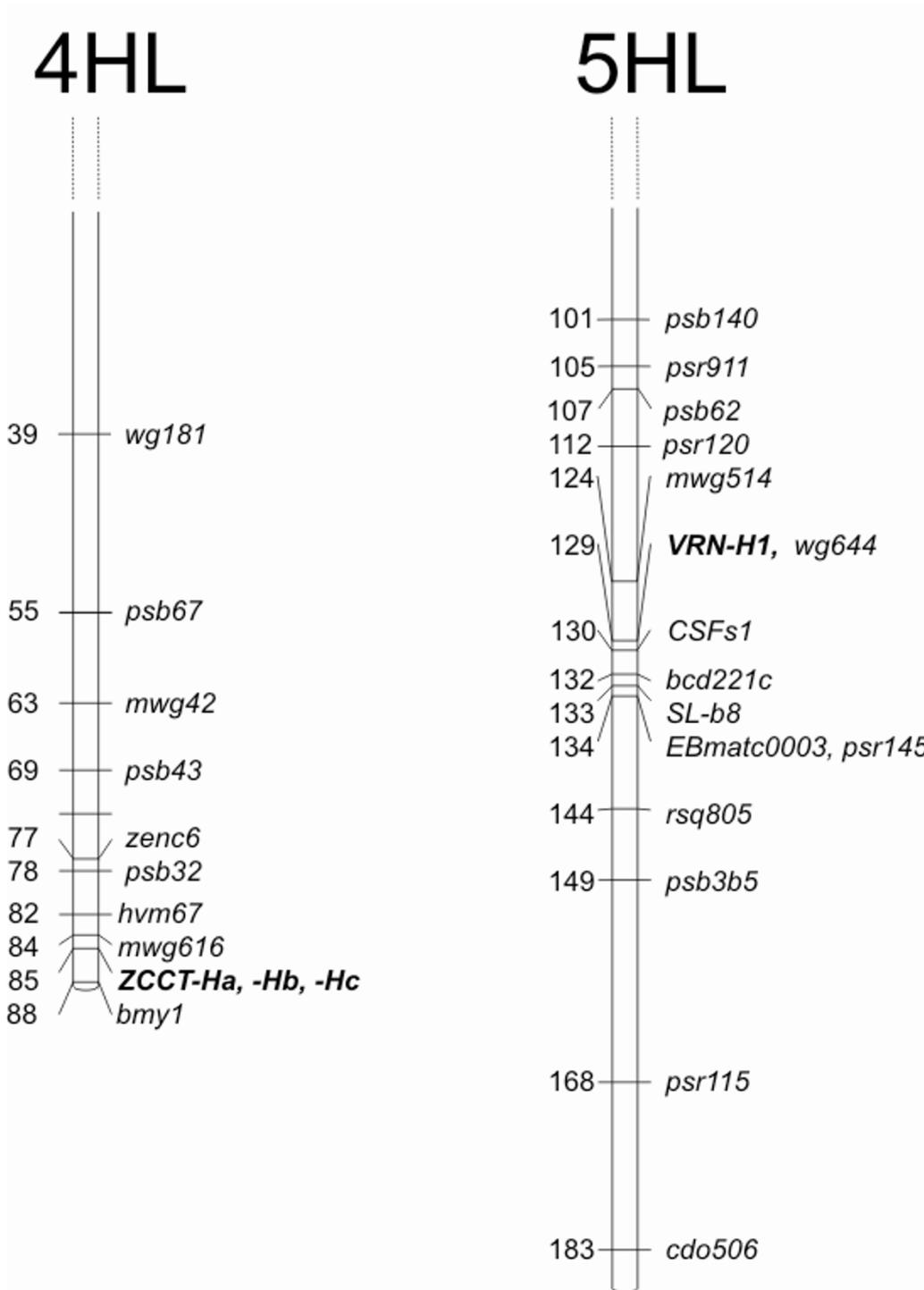


Figure 1
Genetic mapping in the 'Igr1' × 'Triumph' population. Genetic mapping of *ZCCT-Ha, -Hb, -Hc*, *HVM67*, *VRN-H1*, *HvCSFs1*, and *EBmatc0003* in the 'Igr1' × 'Triumph' DH population [6].

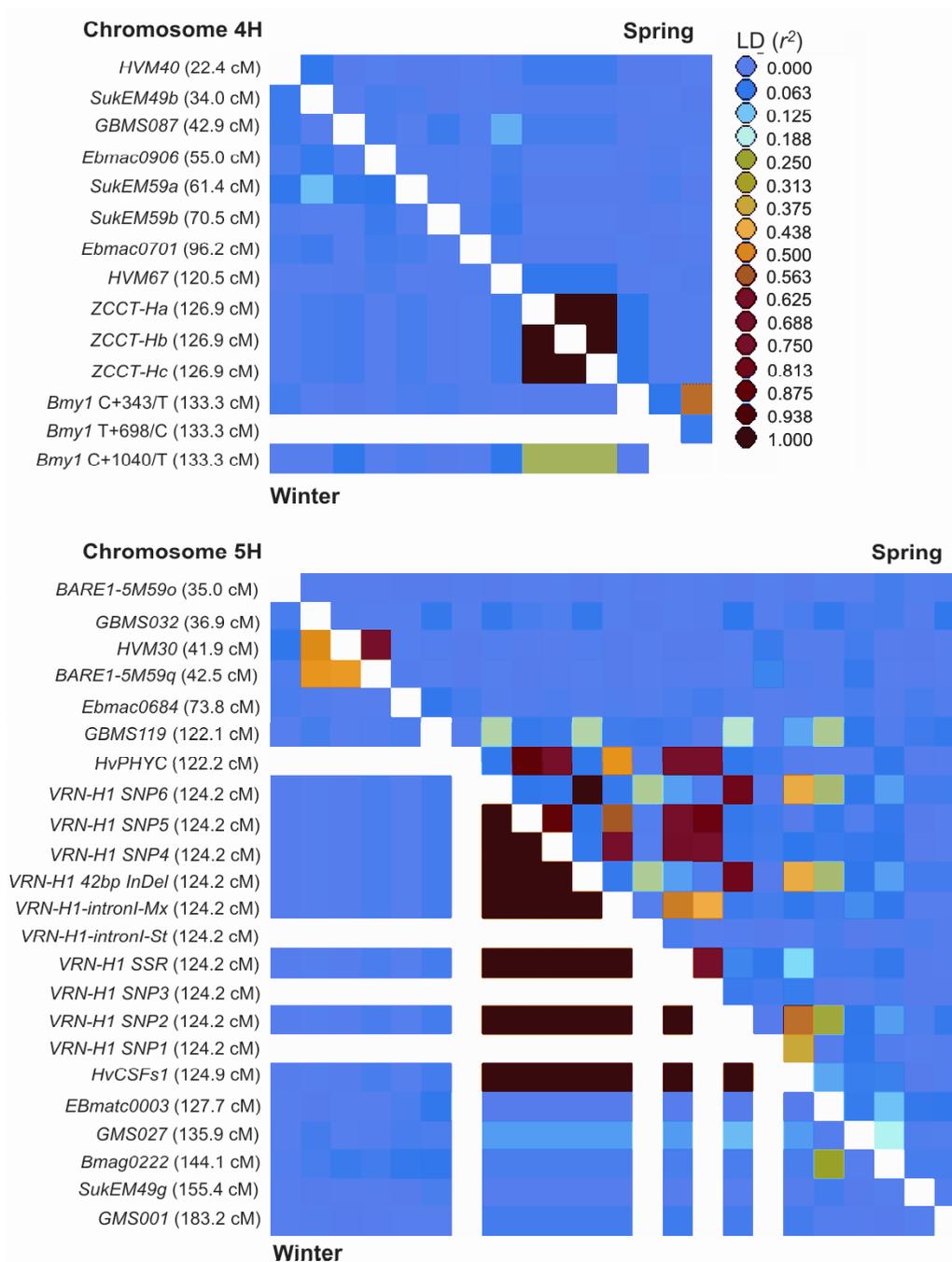


Figure 2

LD on chromosomes 4H and 5H. LD matrix of spring (above the diagonal) and winter (below the diagonal) European barley varieties, as measured by r^2 (colour coded by magnitude). Markers along the x-axis follow the same order as those indicated on the y-axis. Horizontal and vertical white blocks represent uncalculated data, due to invariant genotypes at these loci.

Genetic map positions are from the consensus map of Varshney *et al* [24], after integration of additional markers where necessary. The scarcity of anchor markers close to *HvPHYC* in the 'Dicktoo' × 'Morex' genetic map [23] for integration into the consensus map, resulted in incorrect estimations of map position. Therefore, *HvPHYC* has been positioned 2 cM proximal to *VRN-H1*, as described by Szűcs *et al* [23].

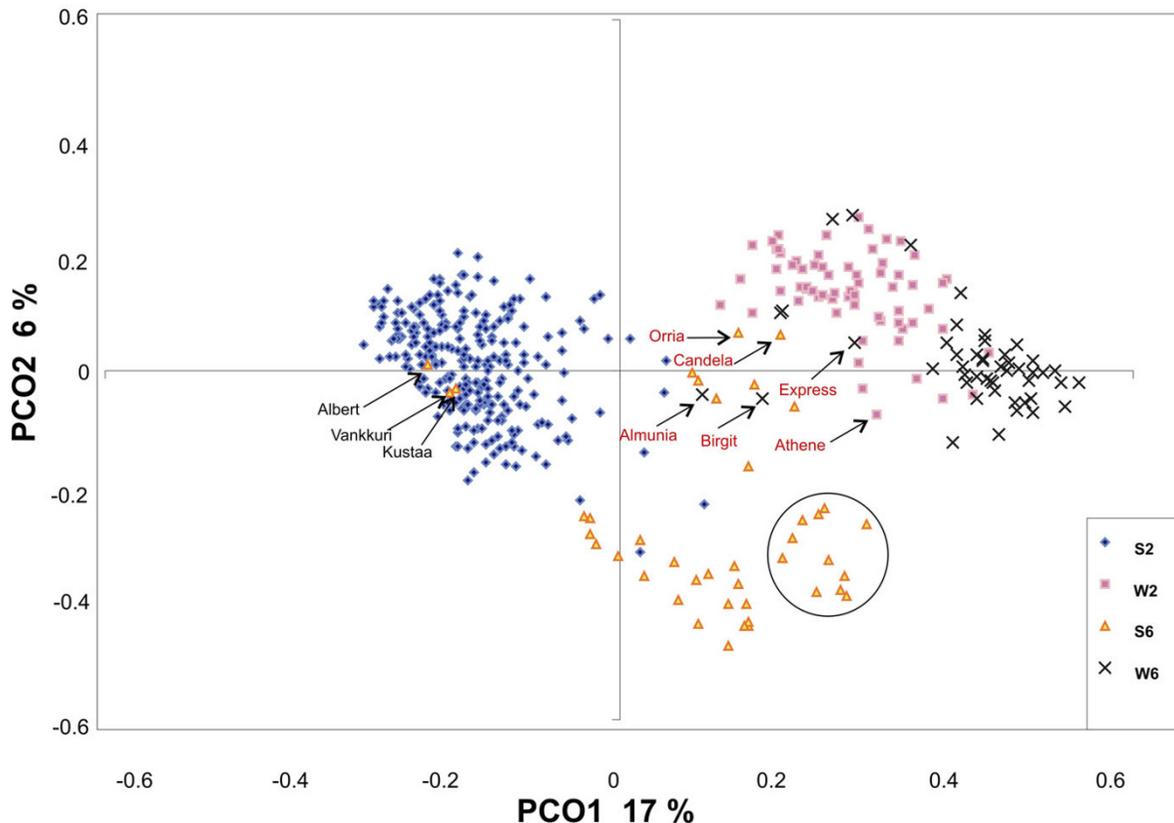


Figure 3

Principal coordinate analysis of 416 barley varieties. Growth habit (S = spring, W = winter) and ear-row number (2 = 2-row, 6 = 6-row) are indicated. *VRN-H1* haplotype designations are previously described [9]. 'Almunia', 'Athene', 'Birgit' and 'Express' (*VRN-H1* haplotype 5C) are the only four winter varieties that do not belong to haplotype 1A. All included members of the spring *VRN-H1* haplotype 1B form a sub-cluster within 6-row spring varieties (circled). The proportion of the total variation accounted for by component 1 (PCO1) and 2 (PCO2) is shown in brackets.

nantly winter genetic background while retaining a spring GH, presumably due to the spring alleles observed at *VRN-H1* [9].

PCoA showed the first and second principal components accounted for only 17 % and 6 % of the genetic variation, respectively, suggesting significant substructure might remain. To help account for this, the Bayesian model-based clustering method implemented by the software STRUCTURE was employed to detect population structure within the varietal collection. The major phenotypic distinctions in cultivated barley suggest $K = 4$ (corresponding to the four possible GH and row-number combinations) might be enough to capture most of the sub-population stratification present, although the PCoA conducted above suggests that further sub-structure exists. In order to explore population stratification, values for K in the range

1 to 20 were evaluated in terms of $\text{Ln}(P|D)$ and AIC (Figure 4). We noted that the value of $\text{Ln}(P|D)$ rose steadily up to $K = 6$, with good duplication of matrices (measured as r) until $K = 8$. However replication of $\text{Ln}(P|D)$ was poor for K in the range 7 to 9, despite multiple replicate analyses. Above $K = 9$ $\text{Ln}(P|D)$ was more stable, but it proved impossible to obtain high correlation between replicate matrices. Thus, analysis of values of $\text{Ln}(P|D)$ did not permit clear interpretation. However, AIC showed that variation in winter/spring phenotype was adequately described at $K = 4$. We reason that since it is spurious association between trait and population structure for which we wish to control then, *for this trait*, the appropriate minimum value of K is 4. Values of K greater than 6 would be difficult to use reliably since they either show poor agreement between replicates in terms of $\text{Ln}(P|D)$ and/or in terms of correlation between replicate matrices.

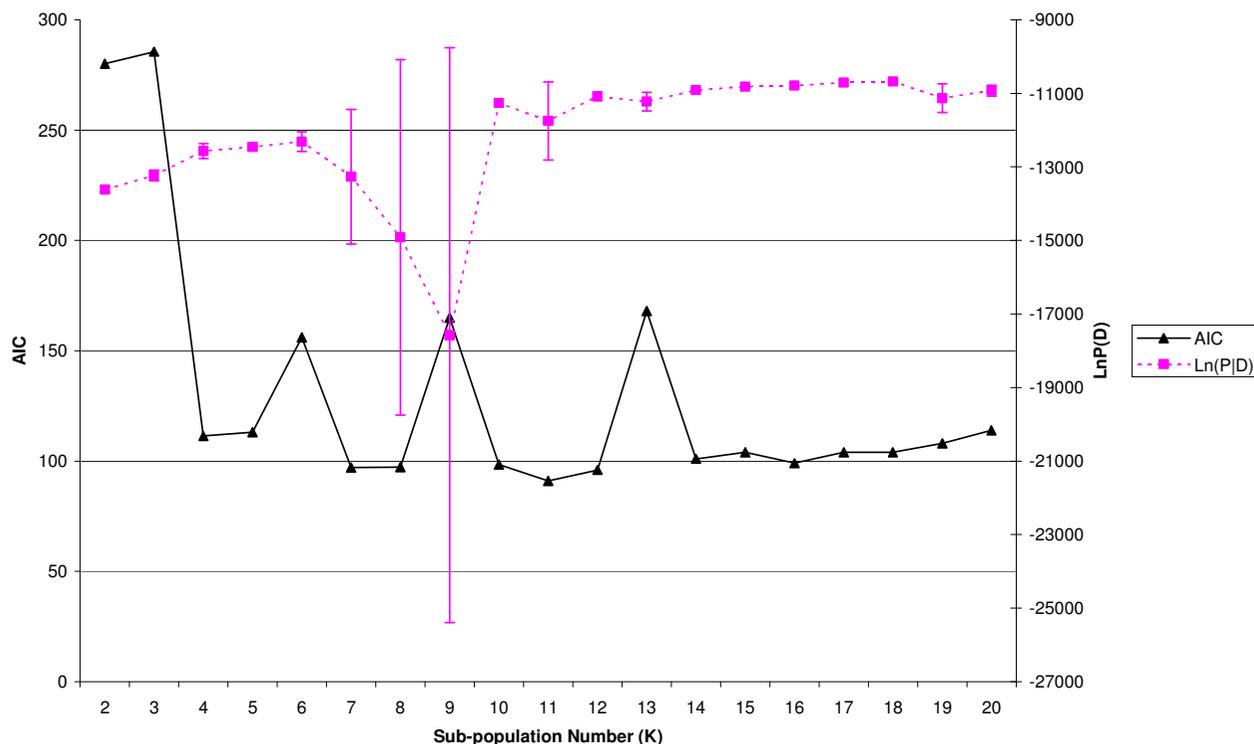


Figure 4
Identifying the appropriate sub-population number (K). Sub-population number (K) against $\text{Ln}(P|D) \pm 1 \text{ SD}$ (100,000 burn-in and 900,000 MCMC iterations, 2 replicates) and Akaike's Information Criterion (AIC).

We used $K = 4$ in SA analysis [18,27] with and without GC, with GH designations used as a two-state categorical phenotype. For comparison we repeated the analysis without implementing SA ($K = 1$). Logistic regression against population membership, and subsequently on marker genotypes (all mapped S-SAP and bi-allelic genome-wide markers), was performed to identify statistically significant marker-trait associations after adjustment for population structure [18,27]. Strength of single marker associations were calculated with and without the inclusion of population structure effects in the model for K equal to 1 and 4. Additional control for the high level of spurious association predicted in highly structured samples was implemented using Genomic Control [19,28], whereby marker chi-squared values (1 d.f.) were divided by the estimate of their robust mean $[(\text{median } \chi^2)/0.455]$ of the S-SAPs. Results of these analyses are summarised in Figures 5 and 6. Comparison of statistical corrections applied to the unmapped S-SAP markers shows that without correction, 77 % of the markers show association at $p = 0.05$, but this falls rapidly with the application of SA (23 %), SA and GC together (9 %) and GC alone (7 %) (Figure 5). Figure 6 shows the effect over mapped markers on chromosomes 4H and 5H. The critical observation is that

SA appears to help distinguish between association which is dependent or independent of sub-population membership. Figure 6B shows significant associations were observed on chromosomes 5H and 4H (on which *VRN-H1* and *VRN-H2* are located, respectively). Amongst the mapped markers, fourteen associations with GH were significant at $p < 0.05$ (Bonferroni adjusted) after adjusting for population structure alone. Subsequent implementation of GC reduced this to nine markers: four within *VRN-H1* (*VRN-H1-intron1-St*, *VRN-H1-SNP2*, *VRN-H1-SNP4*, *VRN-H1-42bp-InDel*), the *VRN-H2* candidate genes *ZCCT-Ha*, *-Hb*, *-Hc*, as well as *HvCSFs1* (0.7 cM distal to *VRN-H1*) and *HvPHYC*. No additional markers were found to be significant.

Discussion

For association mapping to be possible, LD must be present in the collection of individuals under study, with the levels of LD varying according to species and locus investigated. Previous studies in maize and the predominantly self-pollinating model species *Arabidopsis* and rice have demonstrated strong levels of LD surrounding genes controlling flowering time and disease resistance, extending from tens of kb up to 1 cM [14,23,25,27,29-32]. In the

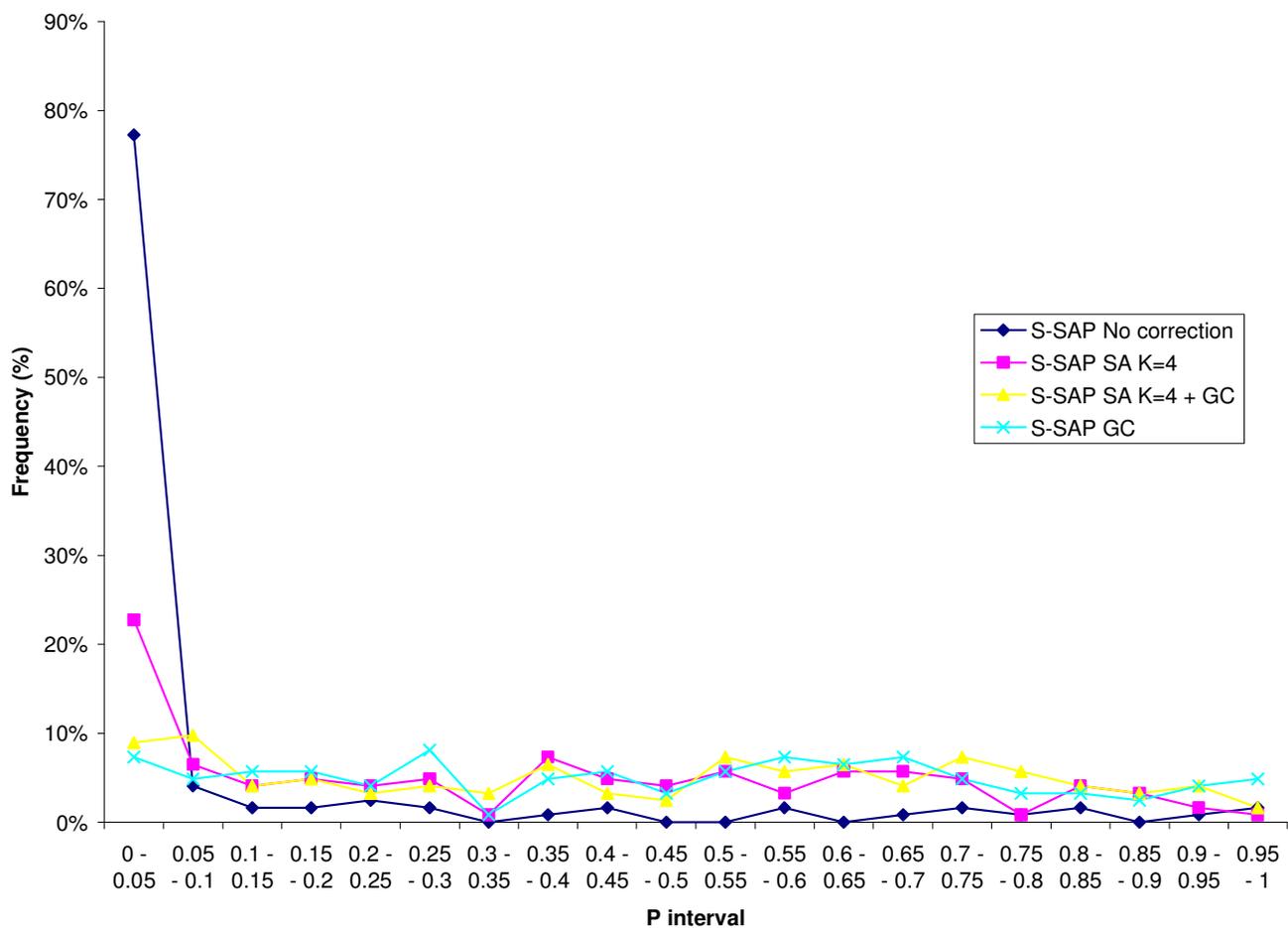


Figure 5
Frequency distribution of p values for unmapped S-SAP markers. p values shown without correction, with Structured Association (SA) with 4 sub-populations (K = 4), SA (K = 4) + Genomic Control (GC), and GC alone.

temperate crop barley, LD has been identified over a ~200 kb region encompassing the *Ha* locus controlling grain hardness [33], and up to a distance of 5.5 cM from a gene conferring resistance to the barley yellow mosaic virus (BMYV) complex [34]. Here we estimate LD to extend at least 0.7 cM from *VRN-H1* and 6.4 cM from *VRN-H2*. This is within a similar range of LD previously identified surrounding selected traits in barley, and helps to indicate the scale of marker saturation required to permit future association mapping of loci of interest in barley. High levels of LD were observed within *VRN-H1* over distances of up to 16.7 kb (winter *VRN-H1* haplotype 1A), suggesting LD decay within *VRN-H1* may differ between European germplasm, compared to the low levels suggested to occur in North American germplasm [17].

As an initial inspection of population substructure, PCoA suggested primary division within the sample investigated

is due to GH. Division within GH pools due to row-number was more distinct within the spring germplasm. Furthermore, notable sub-clustering was observed for varieties belonging to the spring *VRN-H1* haplotype 1B [9], found predominantly in Scandinavian varieties. PCoA clearly showed that sub-structure existed within the varietal population and would therefore need to be adequately accounted for during association analysis. It is generally acknowledged that establishing a value for K (the number of sub-populations) prior to association analysis, is not a trivial exercise [35,36]. To identify the appropriate value of K for use in this study, we attempted to find a maximum for $\ln(P|D)$ (a measure of the maximum likelihood of the specific sub-population model as a description of these data) and also to minimise AIC (which finds the value of K at which the phenotypic variation is best accounted for by the sub-population membership matrix). Neither approach gave an unambiguous answer

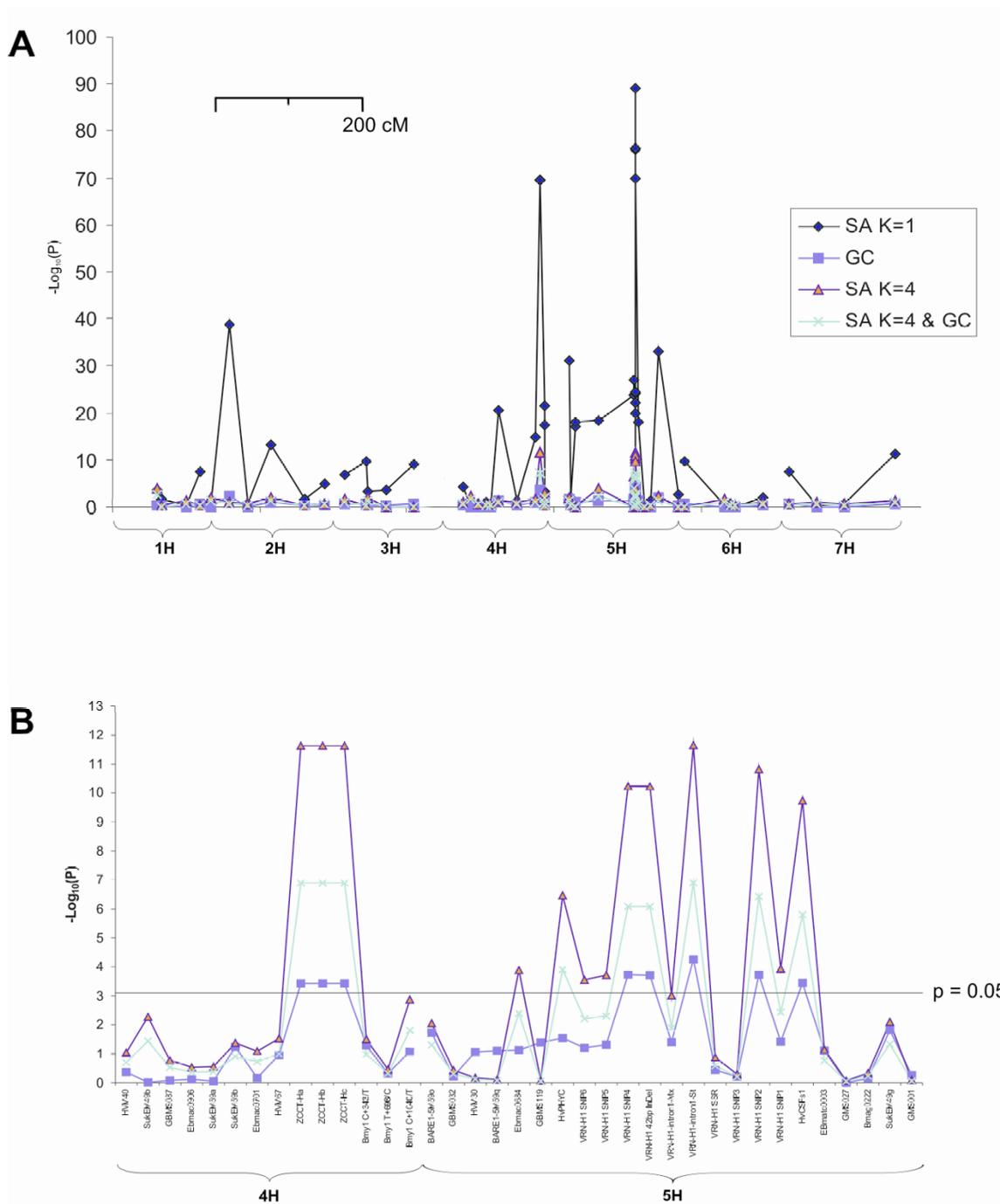


Figure 6
Association mapping of GH in barley. Association of GH with mapped genetic markers using association (SA K = 1), Genomic Control (GC), SA with a population sub-structure = 4 (SA K = 4) and SA with population sub-structure = 4 plus Genomic Control (SA K = 4 & GC). GC was performed using the robust mean of LRT for the unmapped S-SAP markers. **(A)** All mapped markers. Distance between markers is proportional to scale (cM). **(B)** Detailed view of mapped markers on chromosomes 4H and 5H (SA K = 1 not shown). Associations between marker and GH are considered significant above $p = 0.05/\text{marker number}$ (Bonferroni correction). Distance between markers is not to scale.

but instability in determining the Q matrix suggested $K > 6$ was unsafe, while AIC suggested most of the phenotypic variation was described at $K = 4$. Varietal membership to the four groups does not correspond to the expected winter/spring, 6-row/2-row combinations, although 96 % of sub-population 1 are 2-row spring, and sub-population 3 contains 87 % of all winter varieties (data not shown). SA with $K = 4$ appears to have eliminated a high proportion of false positive results: out of 61 markers tested 40 were significant in uncorrected data at $p = 0.05$ (Bonferroni corrected), falling to 14 after SA was applied (Figure 6). However, in many situations, prior knowledge as we have described above will not be available. In these circumstances we believe that exploration of $\text{Ln}(P|D)$ and AIC values could give useful guidance.

Association mapping by logistic regression without correction for structure or for inflation of λ due to multiple testing showed significant association across chromosomes 4H and 5H, while 77 % of marker/trait associations using the unmapped S-SAP markers had an observed p value ≤ 0.05 (Figure 5 and 6A). Correction using structured association alone ($K = 4$) resulted in a large reduction in associated markers, presumably due to a selective reduction of spurious associations, reflected in the reduction from 77 % to 23 % of markers showing association with $p \leq 0.05$ (Figures 5 and 6B). The addition of genomic control to SA analysis resulted in the elimination of an additional five markers on chromosomes 4H and 5H; applied to the genome wide marker set, the proportion of markers with $p \leq 0.05$ was reduced to 9 %. Interestingly, GC alone reduces the number of markers significant at $p \leq 0.05$ slightly more (7 %) than when used in conjunction with SA (9 %). We attribute this to the discriminating nature of the SA correction: GC alone (based on a robust mean of $\lambda = 24.7$) reduces the test statistic for all markers by an equal proportion; in contrast SA selectively reduces λ at each marker using a correction specifically tailored to the sub-population fractional membership of each individual in the experimental population. SA reduced background association such that the robust mean of λ falls to 1.8. Subsequent GC correction is therefore much less stringent because the general inflation of the test statistic has already been largely removed by SA. In summary, we believe the approach undertaken gives a higher objective threshold for significance and therefore reduces the number of potentially associated markers for consideration. Despite GH representing one of the major divisions within barley germplasm, association mapping using the combined SA and GC approach identified markers within *VRN-H1* and candidate *VRN-H2* genes as significantly associated with GH after applying statistical correction for the population structure they largely define. Of all the *VRN-H1* markers investigated here, the *VRN-H1-intronI-St* assay (which tests for the absence of large intron I dele-

tions within a 'vernalization critical' region associated with the recessive winter *vrn-H1* allele) was previously found to show strong correlation with GH [7-9]. SA found this assay to show the highest association with GH (Figure 6B), demonstrating the statistical approaches undertaken here agree with previous studies aimed at determining functional polymorphism at *VRN-H1*. Five additional *VRN-H1* markers showed highly significant (but lower) p values, although since they are not within the 'vernalization critical' region of *VRN-H1*, their significance is most likely due to the strong LD identified within the gene.

The identification of markers strongly associated with GH, 0.7 cM distal (*HvCSFs1*) and 2 cM proximal (*HvPHYC*) [23] to *VRN-H1* illustrates this locus could potentially have been identified by a genome-wide scan even in the absence of the candidate genes assayed here, given a density of approximately one marker every 1 cM. This level of marker saturation is yet to be routinely achieved in large genome crops, although coverage approaching this density appears feasible in barley [3]. Such practical limitations for the use of association genetics in plants suggests it is better suited to fine-mapping, after localization of the trait of interest by QTL mapping [1]. This approach would have been applicable here, even with no prior knowledge of *VRN-H1* and candidate *VRN-H2* genes. Indeed, the association analysis carried out in this study provided a resolution capable of differentiating between intra-genic *VRN-H1* markers. Previous association mapping studies have failed to identify *VRN-H1*, despite an average marker density of ~ 84 SNPs per chromosome [3]; instead, a marker 27 cM proximal to *VRN-H1* was highly associated with growth habit. Furthermore, the use of a G/C SNP in the 3' UTR of *VRN-H1* in an association study based on a collection of 102 North American barley varieties analysed in conjunction with $\sim 1,100$ genome-wide SNPs, showed no correlation with growth habit [17]. The failure to detect *VRN-H1* and *VRN-H2* in previous association mapping studies may have been due to factors such as sample size, partitioning according to population structure and phenotypic errors. However, even if closely linked markers are available, the power to detect association also depends on marker and trait allele frequencies [37,38]. For example, if the minor allele frequency is very low, or present in a subset of lines not associated with a QTL, the marker is less likely to detect association. This is evident here, where despite the high associations between the majority of *VRN-H1* markers and growth habit, three failed to detect significant association.

A variety of approaches have been successfully used to detect marker-trait associations in cereals. Thornsberry *et al* [27] pioneered the use of Bayesian modelling of population structure (using STRUCTURE) in conjunction with logistic regression in crop plants in an association study of

candidate gene polymorphisms with flowering time variation in maize. In addition to logistic regression, the Buckler group have implemented the General Linear Model (GLM) and Mixed Linear Model (MLM) approaches using TASSEL [39]. GLM has been used by Ravel *et al* [40]; MLM, which incorporates a model of kinship between varieties with complex patterns of shared ancestry, is suitable for applications where larger numbers (some hundreds) of markers are available [41]. The MLM approach with SA was used recently to look for associations with seasonal growth habit in barley [3]. Breseghello *et al* [42] used STRUCTURE to detect population structure and MLM to detect marker trait associations in hexaploid wheat. Kraakman *et al* [43], looking for association with yield and yield stability in spring barley, found no population structure using a Bayesian modelling approach, and instead calculated Pearson correlation coefficients and applied a False Discovery Rate to correct for multiple testing. Although the use of MLM was not possible here (due to the limited marker numbers available), the basic logistic regression with SA reported here is analogous to the logistic regression analysis module available in TASSEL. Our implementation of genomic control follows standards set in human studies [19]. It is, to our best knowledge, the first example of its use to control for residual confounding after adjustment by other means. A similar approach is suggested by Pritchard *et al* [18] in their description of SA, and the use of GC in this manner was made explicit by Price *et al* [22] in a supplementary note to their description of a principal components analysis to correct for stratification.

Conclusion

The following lessons can be drawn from this proof of concept of association mapping in barley: firstly, even with limited marker sets using the right correction models, it is possible to obtain robust associations between genes and major adaptive traits such as vernalization requirement, that themselves define the boundaries of population strata restraining gene flow. Performing association mapping in a population with severe substructure due to the presence of spring and winter varieties could be avoided by sampling within one GH class. However, substructure due to additional factors such as row-number, geographic origin and related pedigrees means population sub-structure is hard to avoid, even when populations are carefully selected. Thus, demonstrating the ability to successfully account for population structure is likely to be important in almost all instances. Secondly, we show that strong LD can be detected between 0.7 and 6 cM from the partitioning loci investigated. Thirdly, discrimination between causative intra-genic markers and those in strong LD versus those presumably more ancient interspersed polymorphisms not in LD with the causative deletions was possible, highlighting the power of associa-

tion mapping in dissecting a locus of interest. Finally, we demonstrate a novel method by which taking population structure into account, and then correcting using genomic control, association analysis is possible using marker numbers too low to be used in alternative approaches. The validation of this model presented here could be of particular significance to secondary crops, for preliminary association mapping screens, or any other instance where marker number may be limited.

Methods

Germplasm, DNA Extraction, Phenotyping and Nomenclature

Four-hundred and twenty-nine spring and winter barley varieties were sampled from thirteen EU countries, collected as part of the GEDIFLUX project to determine the impact of modern breeding on molecular crop diversity [14,44]. DNA extraction protocols, sourcing and phenotyping of growth habit (GH) and further details of plant varieties are previously described [9,45]. A list of all the lines used as well as the *VRN-H1* and *VRN-H2* genotypes, recorded growth habit, country of origin and germplasm source, is previously published [9]. Cereal vernalization locus nomenclature follows that described by [46]; the three closely related members of the *ZCCT-H* gene family (*ZCCT-Ha*, *-Hb* and *-Hc*) follow the nomenclature established by [11].

Genome-Wide Marker Profiling and Candidate Gene Genotyping

The retrotransposon-based marker system, S-SAP, was applied to the complete varietal set using six of the primer combinations described by Leigh *et al* [47] (Sukkula-E0228+Mse-CTA, Sukkula-E0228+Mse-CAG, Sukkula-9900+Mse-CAT, Bare5980+Mse-CTA, Nkita-E2647+MseCTT and Bagy2C0589+MseCAC). The resulting 129 polymorphic markers were used for subsequent population structure and association analysis. Twelve S-SAP markers (1H: BARE1-5M59n, SukEM59c, SukEM59c; 2H: SukEM49k, SukEM59d, NikM62e, BARE1-5M59; 4H: SukEM49b, SukEM59a, SukEM59b; 5H: BARE1-5M59o, BARE1-5M59q, SukEM49g) have previously been mapped [48]. Genotyping of the ten genetic markers distributed throughout the *VRN-H1* gene (*VRN-H1-SNP1* to *-SNP6*, *VRN-H1* SSR, *VRN-H1-intron1-St*, *VRN-H1-intron1-Mx* and *VRN-H1-42bp-InDel*) on chromosome 5HL is previously described [9]. Previously described PCR assays [8] to detect the 5.2 kb intron I deletion characteristic of the spring variety 'Morex' (primer pair Intr1/H/F1 and Intr1/H/R1), and the absence of intron I deletions characteristic of the winter variety 'Strider' (primer pair Intr1/H/F3 and Intr1/H/R3) are termed here *VRN-H1-intron1-Mx* and *VRN-H1-intron1-St*, respectively. InDels within intron XII of *CLEAVAGE STIMULTAION FACTOR subunit 1* (*HvCSFs1*, GenBank accession 635P2.2) and exon IV of

PHYTOCHROME C (*HvPHYC*, DQ238106) were genotyped by agarose gel separation of PCR amplicons generated using the following primer pairs: *HvCSFs1-F1* 5'-GACTTGTGAAGCAATATCCAGG-3'/*HvCSFs1-R1* 5'-AGTAAGGGCGTCCCAGACG-3' and *HvPHYC-F11* 5'-AGTTGTCCACCCAGCGCCAG-3'/*HvPHYC-R4* 5'-TCA-GAAGTTGCTCTTGCTCGT-3'. The presence/absence of the three *VRN-H2* candidate genes *ZCCT-Ha*, *-Hb* and *-Hc*, along with an internal amplification control (*HvSNF2*), were assayed as described by Karsai *et al* [12]. Previously published genotype data obtained using the GEDIFLUX barley varietal set for three polymorphic SNPs in the β -amylase gene, *Bmy1*, closely linked to the *VRN-H2* locus on chromosome 4HL, as well as 35 SSR markers distributed over all seven chromosomes, were also included in the analysis [14,45]. Four additional 5HL SSRs distal to *VRN-H1* (*Ebmatc0003*, *GMS027* and *Bmag0222*) were genotyped over the full varietal collection by detection of PCR products tagged by 5' 6-FAM fluorescent labelling of forward primers. PCR products were separated and sized using the AB3100 sequencing platform (Applied Biosystems), according to the manufacturer's instructions. SSR data was processed using GeneMapper v. 3.7 (Applied Biosystems).

Genetic Mapping

The *VRN-H1* SNP, T+14,567/C was mapped using the ABI PRISM® SNaPshot® Multiplex Kit (Applied Biosystems) using primers described by [9]. Gene-specific products were amplified from a doubled haploid (DH) population derived from an 'Igri' (winter) \times 'Triumph' (spring) cross [6]. *HvCSFs1* was mapped in the same population using the amplicon size polymorphism described above, while the SSRs *EBmatc0003* and *HVM67* were mapped following protocols described by [49] to allow integration of *VRN-H1* and *VRN-H2* into a barley consensus map [24]. Genotype data for the S-SAP [49] and GBMS SSR [50] markers previously mapped in 'Steptoe' \times 'Morex' were downloaded from the GrainGenes 2.0 database [51] and combined using JoinMap 3.0 [52]. The map position for *HvPHYC* is as previously published [23], 2 cM proximal to *VRN-H1* in a 'Dicktoo' \times 'Morex' cross. Where necessary, genetic markers were positioned within the consensus map [24], following the protocols they describe.

Estimation of Population Structure

PCoA was carried out as implemented in GenStat v.8 (VSN International, Herts, UK) on a similarity matrix created from all genotype data using a simple matching coefficient. To account for genetic sub-structure in the varietal collection, individuals were allocated to populations, and ancestry coefficients estimated, using the program, STRUCTURE version 2.2 [35,53,54]. Genotypes were presented in haploid format. Results were summarised in matrices of fractional sub-population membership (Q

matrices), modeled with a burnin of 100,000 and a run length of 900,000. SSR genotype data was used to define Q matrices, allowing the unmapped S-SAP data to be used as an independent set of markers for genomic control. 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. Where average maximum correlation (r) was less than 0.99, further replicates with burnin of 200,000 and run length of 1,250,000 were performed to attempt to achieve good duplication. All models were fitted without prior population information provided to the algorithm. Two approaches were taken to estimate the appropriate value for K: (i) from the maximum $\ln(P|D)$, as described by [35]. (ii) Using logistic regression to predict the winter/spring phenotype from the Q matrix obtained at each value of K, and finding the point where change in value of Akaike's Information Criterion (AIC) [55] between subsequent values of K was minimised. Logistic regression was performed using R statistical software v2.5.0 [56]. When a final value of K to be fully tested in association analyses had been determined, new Q matrices were calculated using a burnin of 200,000 and run length of 1,250,000.

Association Testing

SA analyses [18,27] were carried out by logistic regression of GH on estimated coefficients of ancestry (taken from STRUCTURE analysis and on all mapped markers) using R v.2.5.0. For comparison, logistic regression was also carried out directly on markers without first correcting for population structure. Genomic control [19,28] using the 123 unmapped S-SAP markers was used to account for residual confounding remaining after the application of structured association. Strength of single marker associations were measured as $-2 \ln(\lambda)$ (where λ is the likelihood ratio comparing presence and absence of a marker in the analysis) which is asymptotically distributed as a chi-squared distribution with 1 d.f. under the null hypothesis of no association. Significant associations were recorded above the nominal 5 % significance level divided by the number of markers employed, following the standard Bonferroni procedure.

Authors' contributions

JC performed genotyping experiments, analysed and interpreted data, JC and JW wrote the manuscript; IJM carried out PCoA and devised the association testing strategy, JW and IM performed the LD, population structure and association analyses. FL performed S-SAP genotyping, EC contributed genotype and mapping data; DAL, VJL, IJM, WP contributed ideas, supervisory input and commented on the manuscript. DOS devised the study, co-ordinated the research and co-wrote the manuscript.

Additional material

Additional File 1

Pedigree relationships between varieties belonging to VRN-H1 haplotype 1B. Dashed lines indicate crosses involving more than two parental lines. Cultivars displaying VRN-H1 haplotype 1B [9] are highlighted in red and underlined; varieties not included in this study are highlighted in italicised grey. Country and decade of release are indicated, where known.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1471-2156-9-16-S1.doc>]

Acknowledgements

This work was supported in part by Home Grown Cereals Authority and Crop Evaluation Ltd and the GEDIFLUX project (contract No. QLK5-CT-2001-00934), funded by the European Commission. The authors are grateful to members of the GEDIFLUX barley team particularly to Ludmilla Malysheva-Otto and Marion Roeder (IPK) for pre-publication access to SSR genotype data and Paolo Donini and James Reeves for initiating the GEDIFLUX project and encouraging the development of new lines of enquiry arising from it. We would also like to thank Professor D.J. Balding, Imperial College, London for helpful discussion and guidance.

References

- Mackay I, Powell W: **Methods for linkage disequilibrium mapping in crops.** *Trends Plant Sci* 2006, **12**:57-63.
- Malysheva-Otto LV, Ganal MW, Röder MS: **Analysis of molecular diversity, population structure and linkage disequilibrium in a worldwide survey of cultivated barley germplasm (*Hordeum vulgare* L.).** *BMC Genet* 2006, **7**:6.
- Rostocks N, Ramsay L, MacKenzie K, Cardle L, Bhat PR, Roose ML, Svensson JT, Stein N, Varshney RK, Marshall DF, Graner A, Close TJ, Waugh R: **Recent history of artificial outcrossing facilitates whole genome association mapping in elite crop varieties.** *Proc Natl Acad Sci USA* 2006, **103**:18656-18661.
- Balding DJ: **A tutorial on statistical methods for population association studies.** *Nat Genet* 2006, **7**:781-791.
- Cockram J, Jones H, Leigh FJ, O'Sullivan D, Powell W, Laurie DA, Greenland AJ: **Control of flowering time in temperate cereals: genes, domestication and sustainable productivity.** *J Exp Bot* 2006, **58**:1231-1244.
- Laurie DA, Pratchett N, Bezant JH, Snape JW: **RFLP mapping of five major genes and eight quantitative trait loci controlling flowering time in a winter x spring barley (*Hordeum vulgare* L.) cross.** *Genome* 1995, **38**:575-585.
- von Zitzewitz J, Szűcs P, Dubcovsky J, Yan L, Francia E, Pecchioni N, Casas A, Chen THH, Hayes P, Skinner J: **Molecular and structural characterization of barley vernalization genes.** *Plant Mol Biol* 2005, **59**:449-467.
- Fu D, Szűcs P, Yan L, Helguera M, Skinner JS, von Zitzewitz J, Hayes PM, Dubcovsky J: **Large deletions within the first intron in *VRN-1* are associated with spring growth habit in barley and wheat.** *Mol Genet Genomics* 2005, **273**:54-65.
- Cockram J, Chiapparino E, Taylor SA, Stamati K, Donini P, Laurie DA, O'Sullivan D: **Haplotype analysis of vernalization loci in European barley germplasm reveals novel *VRN-H1* alleles and a predominant *VRN-H1/VRN-H2* multi-locus haplotype.** *Theor Appl Genet* 2007, **115**:993-1001.
- Szűcs P, Skinner JS, Karsai I, Cuesta-Marcos A, Haggard KG, Corey AE, Chen THH, Hayes PM: **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 Genomics* 2007, **277**:249-261.
- Dubcovsky J, Chen C, Yan L: **Molecular characterization of allelic variation at the *VRN-H2* vernalization locus in barley.** *Mol Breed* 2005, **15**:395-407.
- Karsai I, Szűcs P, Mészáros K, Filichkina T, Hayes PM, Skinner JS, Láng L, Bedó Z: **The *Vrn-H2* locus is a major determinant of flowering time in a facultative x winter growth habit barley (*Hordeum vulgare* L.) mapping population.** *Theor Appl Genet* 2005, **110**:1458-1466.
- Koebner RMD, Donini P, Reeves JC, Cooke RJ, Law JR: **Temporal flux in the morphological and molecular diversity of UK barley.** *Theor Appl Genet* 2003, **106**:550-558.
- Malysheva-Otto LV, Ganal MW, Law JR, Reeves JC, Röder MS: **Temporal trends of genetic diversity in European barley cultivars (*Hordeum vulgare* L.).** *Mol Breed* 2007, **20**:309-322.
- Aranzana MJ, Kim S, Zhao K, Bakker E, Horton M, Jakob K, et al.: **Genome-wide association mapping in *Arabidopsis thaliana* identifies previously known genes responsible for variation in flowering time and pathogen resistance.** *PLoS Genet* 2007, **1**:e60.
- Zhao K, Aranzana MJ, Jim S, Lister C, Shindo C, Tang C, Toomajam C, Zheng H, Dean C, Majoram P, Nordborg M: **An *Arabidopsis* example of association mapping in structured samples.** *PLoS Genet* 2007, **3**:e4.
- Hayes P, Szűcs P: **Disequilibrium and association in barley: Thinking outside the glass.** *Proc Natl Acad Sci USA* 2006, **103**:18385-18386.
- Pritchard JK, Stephens M, Rosenberg NA, Donnelly P: **Association mapping in structured populations.** *Amer J Hum Genet* 2000, **67**:170-181.
- Devlin B, Roeder K: **Genomic control for association studies.** *Biometrics* 1999, **55**:997-1004.
- Setakis E, Stirnadel H, Balding DJ: **Logistic regression protects against population structure in genetic association studies.** *Genome Res* 2006, **16**:290-296.
- Zhang Z, Btadbury PJ, Kroon DE, Casstevens TM, Buckler ES: **TASSEL 2.0: a software package for association and diversity analysis in plants and animals.** In *Plant & Animal Genomes Conference* San Diego, California, USA. 14th-18th January 2006
- Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D: **Principal components analysis corrects for stratification in genome-wide association studies.** *Nature Genet* 2006, **38**:904-909.
- Szűcs P, Karsai I, von Zitzewitz J, Mészáros K, Cooper LLD, Gu YQ, Chen THH, Hayes PM, Skinner JS: **Positional relationships between photoperiod response QTL and photoreceptor and vernalization genes in barley.** *Theor Appl Genet* 2006, **112**:1277-1285.
- Varshney RK, Marcel TC, Ramsay L, Russell J, Röder MS, Stein N, Waugh R, Langridge P, Niks RE, Graner A: **A high density barley microsatellite consensus map with 755 SSR loci.** *Theor Appl Genet* 2007, **114**:1091-1103.
- Yan L, von Zitzewitz J, Skinner JS, Hayes PM, Dubcovsky J: **Molecular characterization of the duplicated meristem identity genes *HvAPIa* and *HvAPIb* in barley.** *Genome* 2005, **48**:905-912.
- IPK Gatersleben: the ECP/GR barley database [<http://barley.ipk-gatersleben.de/ebdb.php3>]
- Thornsberry JM, Goodman MM, Doebley J, Kresovich S, Nielsen D, Buckler ES: ***Dwarf8* polymorphisms associate with variation in flowering time.** *Nat Genet* 2001, **28**:286-289.
- Reich DA, Goldstein DB: **Detecting association in a case-control study while correcting for population stratification.** *Genet Epidemiol* 2001, **20**:4-16.
- Hagenblad J, Nordborg M: **Sequence variation and haplotype structure in the flowering time locus *FRI* in *Arabidopsis thaliana*.** *Genetics* 2002, **161**:289-298.
- Nordborg M, Borevitz JO, Bergelson J, Berry CC, Chory J, Hagenblad J, Kreitman M, Maloof JN, Noyes T, Oefner PJ, Stahl EA, Weigel D: **The extent of linkage disequilibrium in *Arabidopsis thaliana*.** *Nat Genet* 2002, **30**:190-193.
- Garris AJS, McCouch SR, Kresovich S: **Population structure and its effect on haplotype diversity and linkage disequilibrium surrounding the *xa5* locus of rice (*Oryza sativa* L.).** *Genetics* 2003, **165**:759-769.
- Olsen KM, Halldorsdottir SS, Stinchcombe JR, Weining C, Schmitt J, Purugganan MD: **Linkage disequilibrium mapping of *Arabidopsis* *CRY2* flowering time alleles.** *Genetics* 2004, **167**:1361-1369.
- Caldwell KS, Russell J, Langridge P, Powell W: **Extreme population-dependent linkage disequilibrium detected in an**

- inbreeding plant species, *Hordeum vulgare*. *Genetics* 2006, **172**:557-567.
34. Stracke S, Presterl T, Stein N, Perovic D, Ordon F, Graner A: **Effects of introgression and recombination on haplotype structure and linkage disequilibrium surrounding a locus encoding Bymovirus resistance in barley**. *Genetics* 2007, **175**:805-817.
 35. Pritchard J, Wen X, Falush D: *Document for Structure Software Version 2.1* 2007.
 36. Huelsenbeck JP, Andolfatto P: **Inference of population structure under a Dirichlet process model**. *Genetics* 2007, **175**:1787-1802.
 37. Long AD, Langley CH: **The power of association studies to detect the contribution of candidate genetic loci to variation in complex traits**. *Genome Res* 1999, **9**:720-731.
 38. Zondervan KT, Cardon LR: **The complex interplay among factors that influence allelic association**. *Nat Rev Genet* 2004, **5**:89-100.
 39. **Maize Genetics** [<http://www.maizegenetics.net/>]
 40. Ravel C, Praud S, Murigneux A, Linossier L, Dardevet M, Balfourier F, Dufour P, Brunel D, Charmet G: **Identification of *Glu-B1-1* as a candidate gene for the quantity of high-molecular-weight glutenin in bread wheat (*Triticum aestivum* L.) by means of an association study**. *Theor Appl Genet* 2006, **112**:738-743.
 41. Yu J, Pressoir G, Briggs WH, Vroh BI, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, Kresovich S, Buckler ES: **A unified mixed-model method for association mapping that accounts for multiple levels of relatedness**. *Nat Genet* 2006, **38**:203-208.
 42. Brescghello F, Sorrells ME: **Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars**. *Genetics* 2006, **172**:1165-1177.
 43. Kraakman ATW, Nijs RE, Van den Berg P, Stam P, Van Eeuwijk FA: **Linkage disequilibrium mapping of yield and yield stability in modern spring barley cultivars**. *Genetics* 2004, **168**:435-446.
 44. Reeves JC, Chiapparino E, Donini P, Ganai M, Guiard J, Hamrit S, Heckenberger M, Huan XQ, Van Kaauwen M, Kochieva E, Koebner R, Law JR, Lea V, LeClerc V, Van der Lee T, Leigh F, Van der Linden G, Malysheva L, Melchinger AE, Orford S, Reif JC, Röder M, Schulman A, Vosman B, Van der Wiel C, Wolf M, Zhang D: **Changes over time in the genetic diversity of four major European crops: a report from the GEDIFLUX Framework 5 project**. In *Proceedings of the seventeenth EUCARPIA General Congress: 8-11th September 2004; Tulln, Austria* Johann Vollmann; Heinrich Grausgruber; Peter Ruckebauer; 2004:3-7.
 45. Chiapparino E, Donini P, Reeves J, Tuberosa R, O'Sullivan DM: **Distribution of β -amylase I haplotypes among European cultivated barleys**. *Mol Breed* 2006, **18**:341-354.
 46. Dubcovsky J, Lijavetzky D, Appendino L, Tranquilli G: **Comparative RFLP mapping of *Triticum monococcum* genes controlling vernalization requirement**. *Theor Appl Genet* 1998, **97**:968-975.
 47. Leigh F, Kalendar R, Lea V, Lee D, Donini P, Schulman AH: **Comparison of the utility of barley retrotransposon families for genetic analysis by molecular marker techniques**. *Mol Genet Genomics* 2003, **269**:464-474.
 48. Rodriguez M, O'Sullivan D, Donini P, Papa R, Chiapparino E, Leigh F, Attene G: **Integration of retrotransposon-based markers in a linkage map of barley**. *Mol Breed* 2006, **17**:173-184.
 49. Kobiljski B, Quarie S, Dencic S, Kirby J, Ives M: **Genetic diversity of the Novi Sad Wheat Core Collection revealed by microsatellites**. *Cell Mol Biol Lett* 2002, **7**:685-694.
 50. Li JZ, Sjakste TG, Röder MS, Ganai MV: **Development and genetic mapping of 127 new microsatellite markers in barley**. *Theor Appl Genet* 2003, **107**:1021-1027.
 51. **GrainGenes: maps and map data** [<http://wheat.pw.usda.gov/GG2/maps.shtml>]
 52. Van Ooijen JW, Voorrips RE: **JoinMap 3.0: software for the calculation of genetic maps in experimental populations**. CPRO-DLO, Wageningen; 2004.
 53. Pritchard JK, Stephens M, Donnelly P: **Inference of population structure using multilocus genotype data**. *Genetics* 2000, **155**:945-959.
 54. Falush D, Stephens M, Pritchard JK: **Inference of Population Structure using multilocus genotype data: linked loci and correlated allele frequencies**. *Genetics* 2003, **164**:1567-1587.
 55. Akaike H: **Information theory and an extension of the maximum likelihood principle**. In *Proceedings of the second International Symposium of Information Theory 1973; Budapest* Edited by: Petrov BN, Cs'aki F. Akad'emiai Kiad'; 1973:267-281.
 56. Venables WN, Smith DM, the R development Core Team: **An Introduction to R, Version 2.5.0**. 2006 [<http://www.r-project.org/>].

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

