Theor Appl Genet (2014) 127:2619–2633
DOI 10.1007/s00122-014-2403-y

ORIGINAL PAPER

Applying association mapping and genomic selection to the dissection of key traits in elite European wheat

Alison R. Bentley · Marco Scutari · Nicolas Gosman · Sebastien Faure · Felicity Bedford · Phil Howell · James Cockram · Gemma A. Rose · Tobias Barber · Jose Irigoyen · Richard Horsnell · Claire Pumfrey · Emma Winnie · Johannes Schacht · Katia Beauchêne · Sebastien Praud · Andy Greenland · David Balding · Ian J. Mackay

Received: 30 May 2014 / Accepted: 20 September 2014 / Published online: 2 October 2014

Abstract
Key message **We show the application of association mapping and genomic selection for key breeding targets using a large panel of elite winter wheat varieties and a large volume of agronomic data**.
Abstract The heightening urgency to increase wheat production in line with the needs of a growing population, and in the face of climatic uncertainty, mean new approaches, including association mapping (AM) and genomic selection (GS) need to be validated and applied in wheat breeding. Key adaptive responses are the cornerstone of regional breeding. There is evidence that new ideotypes for long-standing traits such as flowering time may be required. In order to detect targets for future marker-assisted improvement and validate the practical application of GS for wheat breeding we genotyped 376 elite wheat varieties with 3,046 DArT, single nucleotide polymorphism and gene markers and measured seven traits in replicated yield trials over 2 years in France, Germany and the UK. The scale of the phenotyping exceeds the breadth of previous AM and GS studies in these key economic wheat production regions of Northern Europe. Mixed-linear modelling (MLM) detected significant marker-trait associations across and within regions. Genomic prediction using elastic net gave low to high prediction accuracies depending on the trait, and could be experimentally increased by modifying the constituents of the training population (TP). We also tested the use of differentially penalised regression to integrate candidate gene and genome-wide markers to predict traits, demonstrating the validity and simplicity of this approach. Overall, our results suggest that whilst AM offers potential for application in both research and breeding, GS represents an exciting opportunity to select key traits, and that optimisation of the TP is crucial to its successful implementation.

Communicated by Jochen Reif.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-014-2403-y) contains supplementary material, which is available to authorized users.

A. R. Bentley (✉) · N. Gosman · F. Bedford · P. Howell · J. Cockram · G. A. Rose · T. Barber · J. Irigoyen · R. Horsnell · C. Pumfrey · E. Winnie · A. Greenland · I. J. Mackay
The John Bingham Laboratory, NIAB, Huntingdon Road, Cambridge CB3 0LE, UK
e-mail: alison.bentley@niab.com

M. Scutari · D. Balding
Genetics Institute, University College London, London WC1E 6BT, UK

S. Faure · S. Praud
Biogemma, Site de La Garenne, 63720 Chappes, France

J. Schacht
Limagrain GmbH, Salder Straße 4, 31226 Peine-Rosenthal, Germany

K. Beauchêne
ARVALIS-Institut du végétal, 45 voie Romaine-BP 23, 41240 Ouzouer Le Marché, France

Introduction

A ‘perfect storm’ is building in the necessity to balance food production, land utilisation and challenging production environments to meet future needs (Beddington 2009). Growing environmental concerns will also necessitate reductions in agricultural inputs. In 2012, yields of wheat (*Triticum aestivum*), a central contributor to current and future food security, grown on two million hectares of agricultural land in the United Kingdom (GBR) decreased

by 13 %. This reflected a growing season featuring high disease pressure, persistent late-season rainfall, cool conditions and low sunlight levels (DEFRA 2012). Earlier flowering varieties tended to outperform later varieties, a reversal of normal season expectations (HGCA 2012). Instability for long-standing adaptive response represents a new challenge to the sustained production of high-yielding wheat varieties.

The quantitative inheritance of agronomic traits, including yield, is an ongoing limitation to their successful incorporation into wheat breeding programmes. These traits also display strong environmental interactions, and dissection of their genetic control has been limited to date. In contrast, the major genes controlling adaptative traits including flowering time (FT) and height (HT) have been relatively well characterised in wheat, although additional variation exists. In a seminal European multi-year, multi-environment study, Worland (1996) illustrated the importance of matching appropriate FT to agri-environment. Mutant, photoperiod insensitive (PI) lines (*Ppd-D1a,* genetically characterised by Beales et al. 2007) were early flowering and gave a yield advantage in southern Europe through reduced exposure to late-season temperature extremes and drought stress. The opposite effect was observed in cooler GBR conditions where wild-type photoperiod sensitive (PS) lines were able to exploit the longer growing season to build yield. HT is also central to adaptation, and the reduced HT alleles *Rht-B1b* and *Rht-D1b* each confer a 15–20 % HT reduction without compromising grain yield (reviewed by Hedden 2003), and are used worldwide (Wilhelm et al. 2013).

Linkage disequilibrium or association mapping (AM), originally developed for detecting the genetic control of complex traits in human studies (Bodmer 1987), offers new potential for characterising traits of interest for wheat breeding. This utility is bolstered by the increasing availability of affordable, high density, genome-wide molecular markers. AM exploits high allelic diversity for quantitative trait loci (QTL) detection, with linkage disequilibrium (LD) maintained over generations between linked loci (Neumann et al. 2011). The ability to survey large gene pools that are more representative of the breeding pool within any given country or geographic area lends itself to the detection and mapping of multiple traits in a single panel of genotypes (Neumann et al. 2011). Despite the well-documented limitations of AM, including its inability to detect rare alleles, and the effects of unaccounted (Lewis 2002; Zhao et al. 2007) or overcorrected (Segura et al. 2012) population structure, AM has the advantage of being able to detect robust loci that have an effect across genetic backgrounds (Jannink 2007).

AM studies have been employed to detect the genetic basis of a number of traits in wheat. Crossa et al. (2007) used DArT markers to detect numerous marker-trait associations in historical multi-location field trial data for traits including grain yield (YLD) and foliar disease resistance. Neumann et al. (2011) tested the association of 20 agronomic characters over multiple years in a diverse breeders' collection of 96 winter wheat accessions. More recently, Reif et al. (2011) reported AM results for end-use quality traits including thousand grain weight (TGW), grain protein content (GPC) and sedimentation volume in a collection of 207 soft winter wheat lines genotyped with 115 microsatellite markers. Maccaferri et al. (2011) assessed the association of numerous traits including FT, HT and TGW in 189 durum wheat lines grown in multiple Mediterranean environments over 2 years using 186 microsatellite markers.

In contrast to AM in which markers tagging major effect QTL are detected, genomic selection (GS), proposed by Meuwissen et al. (2001) uses all genetic markers simultaneously to generate a model of phenotypes and genotypes to predict genomic breeding values (GEBVs), thereby aiming to capture the total additive genetic variance for the trait of interest. Selection is based on favourable alleles from both parents for overall population improvement, and offers opportunities for rapidly selecting low heritability traits that are under the control of multiple, additive, small effect QTL. This approach is particularly appealing for quantitative traits, where MAS to pyramid multiple target genes into adapted material is unrealistic, and has already been shown to have potential in plant breeding (Bernardo 2009). Model training is based on a training population (TP) composed of material with both phenotypic and genotypic information, requiring genome-wide marker coverage to ensure there is a marker in LD with each QTL. Optimizing the constituents of the training population is a well-recognised challenge in the application of GS (reviewed by Habier et al. 2010; Clark et al. 2011). The model derived from the TP allows selection in a test set on genotype alone, reducing the quantity of phenotyping required and enabling a reduction in cycle time.

Accuracy of trait prediction may be enhanced by inclusion of markers of known effect, markers tagging candidate genes, or markers tagging previously discovered QTL. Adding these markers as fixed effects (i.e. without shrinkage) may be problematic: some may have no effect within the TP; significant marker selection can result in bias; and there may be colinearity problems (where the marker class is large). Pooling selected markers with a substantially larger number of genome-wide markers may also underestimate their effect. To address these issues, differentially penalised regression (DiPR), in which two or more sets of covariates are shrunk independently, can be employed. To date, DiPR has been used to increase the accuracy of predicting seed yield in wheat from metabolite and marker data (Ward, Shewry, Mackay, *personal communication*).

The overall aim of this study was to identify environmentally stable QTL targets for wheat improvement in Northern Europe via AM. The AM panel consisted of 376 elite winter wheat varieties from France (FRA), Germany (DEU) and GBR and was phenotyped in replicated field trials in each of the countries over 2 years. The panel was genotyped with DArT, SNP and candidate gene markers. Traits assessed include the breeding targets YLD, FT, HT, GPC, specific weight (SW), TGW and winter kill (WK). To complement the AM, GS was also performed on the same dataset with the aim of experimentally validating its application in elite germplasm with significant pedigree overlap, therefore informing future practical application based on known variation at key adaptive loci. We also tested the use of DiPR to integrate candidate adaptation gene information with genome-wide DArT markers to predict six of the seven traits, demonstrating the validity and simplicity of this approach.

Materials and methods

Plant material and phenotyping

A panel of 376 elite winter wheat varieties registered in FRA, DEU and GBR between 1946 and 2007 and assembled as part of the European TriticeaeGenome project (http://www.triticeaegenome.eu) was used in this study (Table S1). Reference samples have been deposited in the Germplasm Resources Unit of the John Innes Centre, Norwich Research Park, and are available on request (E-mail: JIC.geneticresources@jic.ac.uk).

All 376 varieties were grown in replicate field plots (2 × 4 m harvested area) planted in an incomplete block design (with two blocks) in 2010 and 2011 in GBR (NIAB: Cambridge 2010; Callow 2011), FRA (Arvalis: Villiers le bâcle) and DEU (Limagrain Gmbh: Adenstedt). Three traits were scored in all trials: YLD (t/ha at 15 % moisture), FT [days to 50 % ear emerged from the flag leaf (GS55; Zadoks et al. 1974)] and HT (cm from stem base to top of spike at maturity). YLD, FT and HT data were analysed using a one-stage analysis in GenStat (14th edition) to generate estimates for best linear unbiased estimates (BLUEs) of variety performance over all sites, and at each site (variety effects were fixed, site and interaction effects were random) for use in the AM and GS. The relationship between YLD, FT and HT and both country of origin and variety age (based on year of variety registration and divided into three arbitrary categories: pre 1990, 1990–1999 and 2000–2007; Table S1) were analysed using residual maximum likelihood (REML) in GenStat. In addition, four traits were scored in specific locations: WK (DEU 2010/11 on a 1–9 visual scale from 1 = less damaged to 9 = heavily damaged), SW (FRA 2010/11, hectolitre weight), TGW (FRA 2010/11) and GPC (FRA 2010/11 using an InfraTec system) and BLUEs calculated as above. Trait heritabilities (h^2) were estimated for YLD, FT and HT on a single plot basis as $V_G/(V_G + V_{GS} + V_E)$ with variation V_G for varieties, V_{GS} for sites × varieties and V_E for within sites (between plots) from a random effects ANOVA in GenStat. For traits scored in only a subset of locations (TGW, SW, GPC, WK), h^2 was calculated as $V_G/(V_G + V_E/2)$. The phenotypic data are available via CerealsDB (http://www.cerealsdb.uk.net).

Genotyping, linkage disequilibrium, population structure

Genomic DNA was extracted from 2-week-old seedlings using a modified Tanksley extraction method (Fulton et al. 1995) and analysed by Triticarte Pty Ltd. (Yarralumla, Australia; www.triticarte.com.au) using the Wheat *Pst*I (*Taq*I) v3 high-density DArT array (7,000 markers). A total of 2,712 polymorphic DArT markers were scored. Markers with >10 % missing data were removed (43 markers), as were markers with a minor allele frequency <2 % (134 markers), leaving 2,535 DArT markers, which are also available via CerealsDB (as above). Then 2,012 DArT markers were assigned map positions based on published mapping information (Huang et al. 2012) and the remaining 523 markers were classified as 'unmapped'. Additional published gene marker assays were run on the panel, including *Vrn-A1*, *Vrn-B1*, *Vrn-D1* (Yan et al. 2004; Fu et al. 2005), *Vrn-B3* (Yan et al. 2006), *Ppd-D1* (Beales et al. 2007), *Ppd-B1* copy number variation (Diaz et al. 2012), *Ppd-A1* (Wilhelm et al. 2009), *Rht-B1* (*Rht1*) (Ellis et al. 2002), *Rht-D1* (*Rht2*) (Ellis et al. 2002), *Rht8* (*gwm261* used to detect 175 bp (referred to as *Rht8_175*) and 192 bp (*Rht8_192*) alleles; Korzun et al. 1998) and the 1BL/1RS wheat-rye translocation (de Froidmont 1998) using published protocols. The panel was also screened with 324 commercial-in-confidence single nucleotide polymorphism (SNP) markers developed by Biogemma which were unmapped.

Missing genotype data were imputed into the full dataset using the package 'impute' in R version 3.0.1 for Windows (http://cran.r-project.org) (Hastie et al. 2013). After imputation the data were thinned by removing one marker from each pair of markers with an absolute correlation coefficient of >0.95. The final dataset consisted of 1,804 markers, with their distribution across the genome as summarised in Table 1. LD was estimated between each pair of markers on the same chromosome, and across the genome with the squared allele frequency correlation (r^2) using the 'popgen' package in R (Marchini 2013).

Population structure was assessed on the final dataset using two methods. The first used Bayesian model-based clustering as implemented in STRUCTURE v2.2 (Pritchard

Table 1 Distribution of 1,804 mapped and unmapped DArT and SNP markers across bread wheat genomes (A, B, D) and groups

Chromosome	Genome			Total by group
	A	B	D	
1	97	91	26	214
2	49	113	48	210
3	47	143	10	200
4	63	30	2	95
5	30	73	3	106
6	98	79	15	192
7	83	62	73	218
Total by genome	467	591	177	1,235
Unmapped				569
Total				1,804

et al. 2000). An admixture model was run with default parameters using a burn-in period of 100,000 iterations and 100,000 Markov Chain Monte Carlo (MCMC) iterations to model scenarios for 2–10 populations (*K*) over three runs. Number of populations was selected using the graphical methods of Evanno et al. (2005). The second method used principal coordinate analysis (PCoA), conducted in R with the first two principle coordinates used as a proxy for population structure.

Association mapping

AM was performed on all traits by implementing a MLM in efficient mixed-model association (EMMA) as implemented in the genome association and prediction integrated tool (GAPIT) in R (Lipka et al. 2012). This used a marker-based kinship matrix derived from a subset of 927 mapped markers giving a minimum inter-marker distance of 1 cM and based on an identical-by-state (IBS) allele-sharing matrix to account for genetic relatedness, as described by Kang et al. (2008). The significance of associations between markers and phenotypes was assessed using the false discovery rate (FDR) (Benjamini & Hochberg 1995) with a *q* value cut-off of 0.05. YLD AM was based on BLUEs across experiments, and incorporating *Rht-B1b*, HT and year of variety registration as covariates, and on BLUEs from individual trials. FT AM was done on BLUEs across experiments, and incorporating *Ppd-D1* as a covariate, and for each individual trial. HT AM was done on BLUEs across experiments, and incorporating *Ppd-D1* and *Rht-D1b*. AM for WK, SW, TGW and GPC was done using BLUEs from trials on which they were measured, and with YLD as a covariate (GPC only). Where an association was with an unmapped marker, AM was re-run in GAPIT (marker as a trait) to provide an approximate marker position.

Statistical power

In order to investigate statistical power, traits were simulated using the final marker dataset with h^2 of 0.001, 0.01, 0.05, 0.10, 0.25, 0.50 and 0.75 for two scenarios: (a) a single QTL explaining all the genotypic variation and (b) 50 % of residual phenotypic variation attributed to 100 background markers. For scenario (a), a marker was selected at random as the target trait locus (x) and its two alleles assigned as 0 and 1. Environmental error for each line was simulated by random draws from a normal distribution with variance scaled to fit the target QTL h^2. For scenario (b), a marker was selected at random as the target trait locus and an environmental error simulated as previously. 100 background markers were also selected, with alleles at each assigned as 0 and 1 and summed to provide a residual genetic effect for each line. Residual genetic and environmental variances were scaled to be equal and to fit the target QTL h^2. The 100 background markers were selected to be at least 20 cM from the target QTL but were otherwise located at random. AM for all simulated traits for both scenarios was performed using MLM in EMMA as above. All markers within a 20 cM interval of the target locus, including the target locus itself, were tested for association for both scenario (a) (Table S2a) and (b) (Table S2b). In addition, a single unlinked marker (>200 cM from the target locus) was analysed to test for adequate control of false positives. For each combination of target locus and h^2, 1,000 simulations were performed to assess the power and precision to detect the target locus.

Genomic prediction

We tested three commonly used methods for predicting genetic values [ridge regression (RR), LASSO and elastic net (EN)] on their ability to predict YLD within country of variety registration using the R package 'penalized' (Goeman et al. 2012). All further tests used EN with the exception of the joint GPC/YLD predictions. The varieties in the panel were assigned to groups based on the country of variety registration (FRA: 210 lines; DEU: 90 lines; GBR: 75 lines [one US variety (Table S1) was excluded]. TPs were then assembled using varieties from a single country and to predict GEBVs for varieties from the same country, and from the remaining two countries. The models' parameters were estimated on the TP using five runs of tenfold cross-validation, and predictive power was measured using the average Pearson's correlation across runs. Standard deviations were tested using 100,000 bootstraps. The panel was also split based on variety age [pre-1990 (103 lines); 1990–1999 (120 lines); post-2000 (153 lines)]. TPs were assembled from each time period and used to predict GEBVs from the same and subsequent time periods. In addition,

Table 2 Proportion of varieties from each country of origin with known adaptation genes for vernalization (*Vrn: vrn-A1/vrn-B1/vrn-D1*: winter alleles) and photoperiod (*Ppd: Ppd-A1a/Ppd-B1a/Ppd-D1a*: photoperiod insensitive alleles) response, reduced height (*Rht: Rht-B1b/Rht-D1b/Rht8_193*: reduced height alleles) and the 1BL/1RS wheat-rye translocation

	vrn-A1	*vrn-B1*	*vrn-D1*	*vrn-B3*	*Ppd-A1a*	*Ppd-B1a*	*Ppd-D1a*	*Rht-B1b*	*Rht-D1b*	*Rht8_192*	*1BL/1RS*
FRA	0.92	0.97	0.99	0.70	0.00	0.03	0.40	0.11	0.51	0.05	0.12
DEU	0.97	0.99	1.00	0.87	0.00	0.00	0.01	0.07	0.40	0.01	0.26
GBR	0.84	0.97	0.99	0.94	0.00	0.00	0.03	0.04	0.92	0.00	0.36

for FT and HT, varieties were assigned to TP and test set based on their *Ppd-D1* and *Rht-D1* allele, respectively, with groups consisting exclusively of *Ppd-D1a* or *Rht-B1b* lines, *Ppd-D1b* or *Rht-D1b* lines, or a mixture of the two used for calculating GEBVs within and between groups. We also investigated predictions using a multivariate approach from a joint ridge regression BLUP (RR-BLUP) model of GPC and YLD using the 'synbreed' package in R (Wimmer et al. 2012) to assess the strength of the link between the two traits and to predict YLD as a function of GPC. To do this, both traits (GPC, YLD) were regressed simultaneously against the markers so that the residuals were correlated. The residuals' correlation structure was removed in the prediction, giving the correlation between traits arising from their common genetic basis.

TP size impacts the ability of GS models to produce reliable GEBVs, making it difficult to compare the performance of models fitted on population subsets of different sizes (as above). To investigate this effect, stratified bootstrap (Davison & Hinkley 1997) resampling was used to sample new populations of size 50, 100 and 200 from each country. The changes in performance of GS models for cross-country predictions were then studied as a function of the bootstrap samples' size (using 100 such samples for each country). As previously, the models' parameters were estimated on the TP using five runs of tenfold cross-validation.

Differentially penalised regression

Differentially penalised regression was tested on the six traits with significant marker-trait associations (YLD, FT, HT, WK, TGW and GPC) using the 11 candidate adaptation genes screened on the panel (Table 2). Given the interrelationships among traits and the structured nature of the population, each locus was considered to be a potential candidate for each trait. DiPR was employed following the regression model

$$Y = X_1 b_1 + X_2 b_2,$$

where $b_1 = [b_1\ b_2 \ldots b_m]$ is a vector of m fixed, standardised, covariate effects; $b_2 = [b_{m1+1}\ b_{m1+2} \ldots b_k]$ is a vector of $(k\text{-}m)$ separate fixed, standardised, covariate effects; $Y = [y_1\ y_2 \ldots y_n]$ is a vector of n phenotypes; X_1 and X_2 are design matrices for the two sets of covariates and assign values at each to the individual phenotypes in Y.

This model was fitted by minimising

$$\sum_{1}^{n}\left(y_{\mathrm{i}} - \sum_{1}^{m1} b_{1p}x_{ij} - \sum_{m1+1}^{k} b_{2q}x_{ik}\right)^2 + \lambda/w\sum_{1}^{m1} b_{1p}^2 + \lambda/(1-\mathrm{w})\sum_{m1+1}^{k} b_{2q}^2$$

where the b_{1p} and b_{2q} index the elements of b_1 and b_2. The first part of the equation $\left(\sum_1^n \left(y_i - \sum_1^{m1} b_{1p}x_{ij} - \sum_{m1+1}^{k} b_{2q}x_{ik}\right)^2\right)$ is identical to that minimised in ordinary least squares and corresponds to minimising the deviation of observed values from expected. The second two terms penalise the sum of squares of the two sets of regression coefficients, b_1 and b_2, independently, with respective penalties, λ/w and $\lambda/(1-w)$. w and $(1-w)$ are scaling factors for the standard deviation of the two groups of covariates, since a regression on x with an estimated regression coefficient b gives an identical fit to a regression on wx with a regression coefficient b/w. Thus, if the two sets of standardised covariates are additionally scaled by w and $(1-w)$, then λ can be estimated using standard ridge regression software across the pooled set of covariates. Varying w over a 0–1 range will then find optimum values of w and λ. At values of $w = 0$ and $w = 1$, the model is equivalent to ridge regression on, respectively, b_2 or b_1 alone. At $w = 0.5$, both sets of covariates are given equal weight, equivalent to ridge regression on the pooled covariates.

The model was fitted using the R package 'penalized' (Goeman et al. 2012) by searching for values of w, which maximised the cross-validation correlation. Five runs of tenfold cross-validations were run for each trait with w varying in intervals of 0.05.

Results

Variation in key trait phenotypes

Both country of origin ($P = 0.008$; GBR > FRA > DEU) and age ($P = 0.001$; post-2000 > 1990–1999 > pre-1990) of variety had significant YLD effects. Regression of

Table 3 Mean allelic effects across the complete association mapping panel of known genes on YLD, HT and FT

	Ppd-D1a	*Rht-B1b*	*Rht-D1b*	*Rht8_193*
Yield (t/ha)	−0.23*	0.00	0.62*	−0.64*
Height (cm)	−4.68*	−3.39*	−7.53*	0.60
FT (days)	−5.79*	−2.00*	1.20*	−4.11*
FT_FRA (days)	−6.72*	−2.40*	0.80*	−4.40*
FT_DEU (days)	−4.07*	0.00	1.06*	−3.50*
FT_UK (days)	−7.40*	−2.50*	2.20*	−5.05*
FT_2010 (days)	−5.28*	−1.90*	1.40*	−3.75*
FT_2011 (days)	−7.04*	−2.40*	1.40*	−5.00*

* Significant differences conferred by the allele ($P < 0.05$)

yield on variety age showed a significant ($P < 0.001$) YLD increase over time across the dataset ($R^2 = 0.38$), and for each of the countries, with the largest increase recorded for the French varieties ($R^2 = 0.57$). GPC decreased over the dataset with a significant variety origin and age interaction ($P = 0.024$). Pre-1990s varieties from FRA had the highest GPC levels (13.88), with the lowest levels recorded for the most recent (post-2000) GBR varieties (12.59). Country of origin significantly influenced FT ($P < 0.001$), with FRA varieties flowering (GS55) an average of 4 days earlier than DEU and GBR varieties. 40 % of French lines had the early flowering PI *Ppd-D1a* allele (Table 2) with no trend with age, suggesting selection to suit local requirements. Early flowering alleles were virtually absent from the DEU and GBR variety sets. Across experiments, where present, *Ppd-D1a* conferred reduced height (−4.68 cm) and significantly reduced YLD (−0.23 t/ha; $P = 0.003$). The majority of the accessions had wild-type *Vrn* alleles conferring winter growth habit and were semi-dwarf, as conferred by mutant *Rht* alleles (Table 2), most notably *Rht-D1b* (*Rht2*) most frequent in GBR varieties (92 %), followed by the FRA (51 %) and DEU (40 %) material. Both *Rht-D1b* and *Rht-B1b* significantly reduced HT across experiments (−3.39 and −7.53 cm, respectively; Table 3). The 7–8 cm HT reducing *Rht8_192* allele (Korzun et al. 1998) occurred at extremely low frequency, and where present conferred no significant HT reduction, whilst the *Rht8_175* allele gave a slight reduction (−1.6 cm, $P = 0.024$). There was a strong geographic trend for WK score and a variety age and origin interaction ($P = 0.015$). DEU varieties exhibited the highest WK resistance, increasing significantly after 1990, in contrast to FRA resistance which fell significantly after 2000 (Fig. 1). GBR scores were stable over time. Trait heritabilities for single plots were relatively high for FT (0.63) and HT (0.82), but lower for YLD (0.25). Within site heritabilities for traits scored at selected locations ranged from high (WK (0.92); GPC (0.83)) to low (SW (0.13); TGW (0.06)).

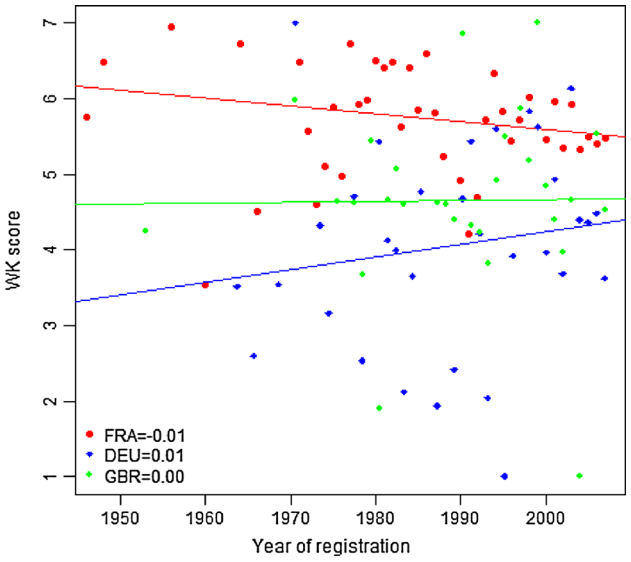

Fig. 1 Opposing trends in WK and country (FRA, DEU, GBR) and year of variety registration for 376 elite Northern European winter wheat varieties

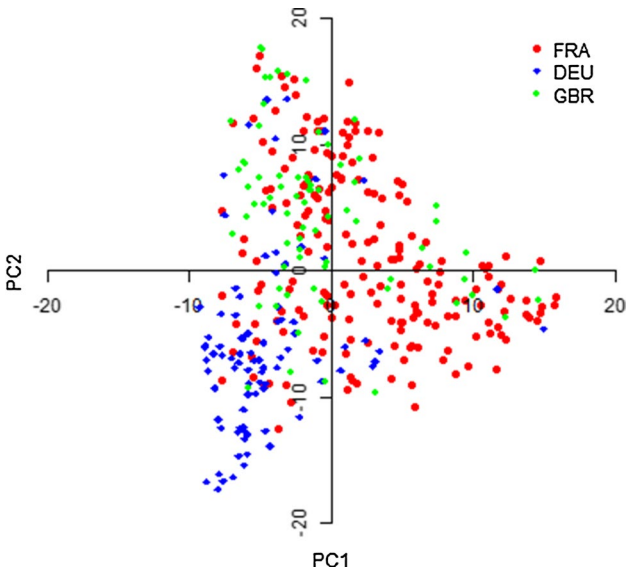

Fig. 2 Population structure of the TriticeaeGenome panel of 376 elite winter wheat varieties as determined by PCoA (coloured by country of origin of each variety), showing PC1 versus PC2 which together explain 12.5 % of the genetic variation in the panel

Linkage disequilibrium, population structure and statistical power

Analysis of STRUCTURE models estimated $K = 2$ based on ΔK (Evanno et al. 2005) (Fig. S1). In the final dataset, PCoA revealed some underlying structure conferred by country of origin (Fig. 2), with PC1 and PC2 describing 8.6 and 3.9 % of the genetic variation. The GBR and

DEU variety sets could be distinguished, whilst the FRA set overlapped both (Fig. 2). Significant LD was estimated using the full dataset of mapped markers to extend to 20 cM (Fig. S2). Simulation results showed good power to detect QTL with modest contribution to trait heritability for the target locus (Table S2; Fig. S3). As expected, on average, the presence of residual genetic variance [Table S2(b); Fig S3(b)] reduces power, though the differences were small for the cases studied here.

Detection of non-perfectly linked markers ranged from 0 to 0.430 ($h^2 = 0.75$; 0 cM < $x \leq 1$ cM from target; 100 background markers). With h^2 of 0.25, 0.50 and 0.75 the target locus was always identified (Table S2), showing good power to detect QTL of modest heritability utilising the experimental marker density of 1,804 markers both with and without background effects. The mean inter-marker interval was 2.18 cM, but with a total map length of 2,851 cM, 35 % of inter-maker intervals were >10 cM indicating a requirement for higher marker density to gain full benefit from the panel.

Association mapping

MLMs for all traits except SW identified significant marker associations. For YLD, two significant marker-trait associations were detected across experiments: BWS2767 and *Rht-D1* (chromosome 4D) (Table 4). Accounting for *Rht-D1* (as a covariate) retained the BWS2767 association and detected an additional association with BWS9255, although this was lost when HT was also considered (as a covariate). The DArT markers wPt_664488 on the distal part of chromosome 3AS and wPt_3451 on chromosome 1BL (respectively) were most highly associated with these unmapped SNP markers. When year of registration was included as a covariate, no significant YLD associations were detected. There was strong evidence that both the 3AS and 1BL loci were under selection, with a significant ($P < 0.001$) interaction between these markers and year, and near fixation in more recent material.

The association between YLD and BWS2767 (wPt_664488, 3AS) was detected in all of the countries in the study, although only in a single year (Table 4). *Rht-D1* had a significant association with YLD in FRA, as did wPt_0472 (tagging *Rht-D1*). Three additional markers were detected in single years in GBR, namely BWS9255 (wPt_3451, 1BL; 2010), wPt_6966 (4AL; 2010) and BWS3158 (wPt_729839, 6AS; 2011). A SNP marker (BWS2641) tagging *VRN-B3* was also detected in DEU in 2011 and GBR in 2010.

AM for FT confirmed the role of *Ppd-D1* (2D), which was significantly associated with FT across experiments (Table 4). The SNP marker BWS2525 and a 2D DArT marker (wPt_733227), tagging *Ppd-D1* were also significantly associated with FT in all but one experiment (DEU 2011). A SNP marker tagging *VRN-B3* (BWS2641) was also associated with FT in DEU in 2010. Accounting for *Ppd-D1* (as a covariate) detected association between *VRN-B3* and wPt_1912 (1BS) and FT.

Six marker-trait associations were detected for HT (Table 4), including *Rht-D1*, wPt_0472 (tagging *Rht-D1*) and *Ppd-D1*. Three other markers, two on 5AL (BWS2245 and BWS5515, both associated with tPt_6495) and one on 6AS (wPt_732355) were also significantly associated with HT. When *Rht-D1* was run as a covariate, the effect of *Ppd-D1* remained, as did the effect of wPt_732355 on 6AS. *Rht-B1* was also detected in the absence of *Rht-D1*. The effects of both wPt_732355 (6AS) and tPt_6495 (5AL) were retained with *Ppd-D1* as a covariate.

WK was associated with *Ppd-D1* and BWS1493 (wPt_4142, 6BS). TGW was significantly associated with a single SNP marker BWS1515 (wPt_4426, 2BL). GPC was also associated with a single SNP marker BWS3158 (wPt_729839, 6AS), although this association was not detectable when YLD was included as a covariate, and was also found to be associated with GBR YLD in 2011.

Genomic selection

The tested GS protocols gave similar predictive correlations for YLD within countries (EN: FRA $r = 0.76$, DEU $r = 0.62$, GBR $r = 0.67$; LASSO: FRA $r = 0.70$, DEU $r = 0.54$, GBR $r = 0.55$; RR: FRA $r = 0.71$, DEU $r = 0.52$, GBR $r = 0.53$). Using EN, YLD was predicted with the highest accuracy in the FRA varieties ($r = 0.75$) followed by GBR ($r = 0.67$) and DEU ($r = 0.62$). Cross-country YLD predictions were the highest from and to FRA (Table 5). Predictions for YLD made on variety age (Table 6) were highest from training on the lines registered from 1990 to 2000 to predict the YLD of the post-2000 lines ($r = 0.58$), and were also relatively high within year groupings (pre-1990 $r = 0.53$; 1990–2000 $r = 0.49$; post-2000 $r = 0.46$).

FT predictions were high within FRA ($r = 0.77$), but reduced in all other comparisons. To investigate whether reduced predictive ability in DEU, GBR and across countries was conferred by distribution of *Ppd-D1* in the TP and test set, correlations were tested using TPs composed only of *Ppd-D1a* or *Ppd-D1b* lines, and mixtures of the two (Table 7). Pooling lines based on *Ppd-D1* led to increased predictive correlations in all cases, allowing differentiation of early and late flowering types in the test set, but not to the point of delineating early or late lines on degrees of earliness or lateness (Fig. S4).

HT predictions were highest in the FRA material ($r = 0.72$), and in the FRA to DEU ($r = 0.61$) cross-country predictions. GBR lines gave the lowest HT

Table 4 Significant marker-trait associations (given as q values) across and within experiments for 376 wheat varieties with the significance of associations determined using $q = 0.05$ false discovery rate (FDR)

Trait[a]	Covariate[b]	Marker	Chr[c]	Across sites	Within sites					
					FRA_2010	FRA_2011	DEU_2010	DEU_2011	GBR_2010	GBR_2011
YLD		BWS2767	3A	0.0005	0.0027			0.0033		0.0006
YLD	*Rht-D1*	BWS2767	3A	0.0001						
YLD	HT	BWS2767	3A	0.0012						
YLD		*Rht-D1*	4D	0.0006	0.0027	0.0001				
YLD	*Rht-D1*	BWS9255	1B	0.0025					0.0048	
YLD	*Rht-D1*	BWS3158	6A	0.0025						0.0063
YLD		wPt_0472	4D			0.0015				
YLD		BWS2641	7B					0.0051	0.0005	
YLD		wPt_6966	4A						0.0044	
FT		*Ppd-D1*	2D	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
FT		BWS2525	2D	0.0000	0.0000	0.0000	0.0001		0.0009	0.0001
FT		wPt_733227	2D	0.0000	0.0003	0.0000	0.0005		0.0033	0.0000
FT	*Ppd-D1*	*VRN-B3*	7B	0.0025						
FT	*Ppd-D1*	wPt_1912	1B	0.0032						
FT		BWS2641	7B				0.0004			
HT		*Rht-D1*	4D	0.0000						
HT	*Ppd-D1*	*Rht-D1*	4D	0.0000						
HT		*Ppd-D1*	2D	0.0001						
HT	*Rht-D1*	*Ppd-D1*	2D	0.0000						
HT		BWS2245	5A	0.0002						
HT	*Ppd-D1*	BWS2245	5A	0.0001						
HT		wPt_0472	4D	0.0002						
HT		wPt_732355	6A	0.0002						
HT	*Rht-D1*	wPt_732355	6A	0.0000						
HT	*Ppd-D1*	wPt_732355	6A	0.0006						
HT		BWS5515	5A	0.0020						
HT	*Ppd-D1*	BWS5515	5A	0.0016						
HT	*Rht-D1*	BWS2952	6A	0.0041						
HT	*Rht-D1*	*Rht-B1*	4B	0.0000						
WK		*Ppd-D1*	2D	0.0002						
WK		BWS1493	6B	0.0021						
TGW		BWS1515	2B	0.0062						
GPC		BWS3158	6A	0.0018						

[a] AM for HT, WK, TGW and GPC was done on line means across the experiments for which the traits were recorded

[b] Term included as a covariate in the AM

[c] Where significant markers were unmapped, chromosome assignment is estimated (via association mapping) to identify the closest mapped marker

predictions ($r = 0.36$). As with FT/*Ppd-D1*, the allelic distribution of *Rht-D1* alleles (Table 2) was tested as the cause of this reduced prediction accuracy. Across the pooled dataset (Table S3) the highest predictions were made between TP and test sets with mixtures of *Rht-D1* alleles ($r = 0.78$), compared to predictions between a TP of only *Rht-D1b* lines to a test set of *Rht-D1a* lines ($r = 0.52$) (Table S3).

GPC predictions were high within (DEU/FRA $r = 0.66$, GBR $r = 0.58$) and across countries (Table 5). Fitting GPC jointly with YLD and predicting the latter from the former across the dataset gave the highest predictive correlations ($r = 0.77$) (Fig. 3). In contrast, SW predictions were low between countries (ranging from 0 to 0.22), with the highest predictions within countries (FRA $r = 0.47$; DEU $r = 0.36$). TGW within country

Table 5 Predictive correlations (r) for all traits within and between countries [FRA (210 lines); DEU (90); GBR (75)] with standard deviations based on 100,000 bootstraps indicated in parentheses

Trait	FRA	DEU	GBR	FRA		DEU		GBR	
	FRA	DEU	GBR	DEU	GBR	FRA	GBR	FRA	DEU
YLD	0.75 (0.03)	0.62 (0.06)	0.67 (0.06)	0.55 (0.07)	0.54 (0.08)	0.52 (0.04)	0.32 (0.10)	0.50 (0.05)	0.22 (0.10)
FT	0.77 (0.03)	0.43 (0.08)	0.36 (0.09)	0.23 (0.14)	0.36 (0.17)	0.11 (0.07)	0.08 (0.13)	0.21 (0.07)	0.11 (0.10)
HT	0.72 (0.04)	0.55 (0.09)	0.36 (0.07)	0.61 (0.07)	0.57 (0.07)	0.45 (0.05)	0.55 (0.08)	0.54 (0.04)	0.55 (0.07)
GPC	0.66 (0.04)	0.66 (0.05)	0.58 (0.08)	0.48 (0.08)	0.54 (0.08)	0.43 (0.05)	0.31 (0.08)	0.48 (0.05)	0.49 (0.07)
SW	0.49 (0.05)	0.24 (0.09)	0.07 (0.10)	0.22 (0.11)	0.12 (0.13)	0.19 (0.07)	0.00 (0.11)	0.22 (0.06)	0.04 (0.09)
TGW	0.36 (0.05)	0.12 (0.11)	0.31 (0.10)	0.15 (0.11)	0.32 (0.10)	0.17 (0.06)	0.21 (0.10)	0.20 (0.06)	0.00 (0.10)
WK	0.48 (0.06)	0.75 (0.04)	0.25 (0.11)	0.44 (0.07)	0.23 (0.09)	0.14 (0.07)	0.00 (0.12)	0.19 (0.06)	0.00 (0.10)

Table 6 Predictive correlations by year of variety registration with standard deviations based on 100,000 bootstraps indicated in parentheses

From/to	<1990	1990–2000	>2000
<1990	0.53 (0.06)	0.40 (0.09)	0.42 (0.07)
1990–2000		0.49 (0.07)	0.58 (0.06)
>2000			0.46 (0.06)

predictions ranged from 0.20 to 0.37, whilst cross-country predictions ranged from 0 to 0.32. WK could not be predicted between DEU and GBR in either direction, but predictions within the DEU material were high ($r = 0.75$).

Cross-country YLD predictions did not show any marked improvement when increasing sample size (bootstrap samples with 50, 100 and 200 observations) (Fig. S5), with the exception of predictions from GBR to DEU ($r = 0.10$ to $r = 0.29$). Similarly, HT, SW and WK were approximately constant (Fig. S5) with the exception of predictions from FRA to DEU (HT: $r = 0.46$ to $r = 0.58$; SW: $r = 0.13$ to $r = 0.21$; WK: $r = 0.09$ to $r = 0.43$). WK predictions also improved from DEU to FRA ($r = 0.09$–0.21). FT predictions improved noticeably ($r = 0.10$ to $r = 0.25$–0.36), except between GBR and DEU ($r = 0.10$–0.15). GPC and TGW predictions from FRA to GBR and from GBR to FRA also improved (GPC: $r = 0.26$–0.27 to $r = 0.45$–0.47; TGW: $r = 0.09$–0.10 to $r = 0.21$–0.26).

Differentially penalised regression

Figure 4 plots average cross-validation correlations for each trait against w, the scaling factor for the standardised covariates where $w = 0$ corresponds to prediction from DArT markers alone, $w = 1$ corresponds to prediction from the ten candidate loci markers alone, and $w = 0.5$ corresponds to prediction from ridge regression equally weighted on the pooled markers. The cross-validation correlations at these values of w and the maximum cross-validation correlation are given in Table 8. There is little or no difference from prediction on DArT markers alone ($w = 0$) and from unweighted prediction ($w = 0.5$) on all markers. Values of w which maximise the cross-validation correlation are all between 0.5 and 1, i.e. where relatively more weight is placed on the candidates than on the DArT markers. The improvement is large for HT and FT because there is substantial information in the candidate markers for these traits. Using these markers alone, cross-validation correlations are high (FT = 0.76, HT = 0.67). For FT, this is higher than from any regression incorporating the DArT markers, except DiPR itself.

Discussion

The genetic gain in elite variety YLD over time in key production regions of Northern Europe (FRA, DEU and GBR) confirms the vital role of plant breeding efforts (Mackay

Table 7 Predictive correlations (r) for FT with different groups of *Ppd-D1* alleles in the training population and test set with standard deviations based on 100,000 bootstraps indicated in parentheses

From/to	*Ppd-D1a* (83)	*Ppd-D1b* (288)	*Ppd-D1a* (40) + *Ppd-D1b* (143)	*Ppd-D1a* (42) + *Ppd-D1b* (144)
Ppd-D1a (83)	0.38 (0.08)	0.22 (0.10)		
Ppd-D1b (288)	0.26 (0.10)	0.57 (0.04)		
Ppd-D1a (40) + *Ppd-D1b* (143)			0.80 (0.03)	0.79 (0.03)
Ppd-D1a (42) + *Ppd-D1b* (144)			0.79 (0.03)	0.80 (0.03)

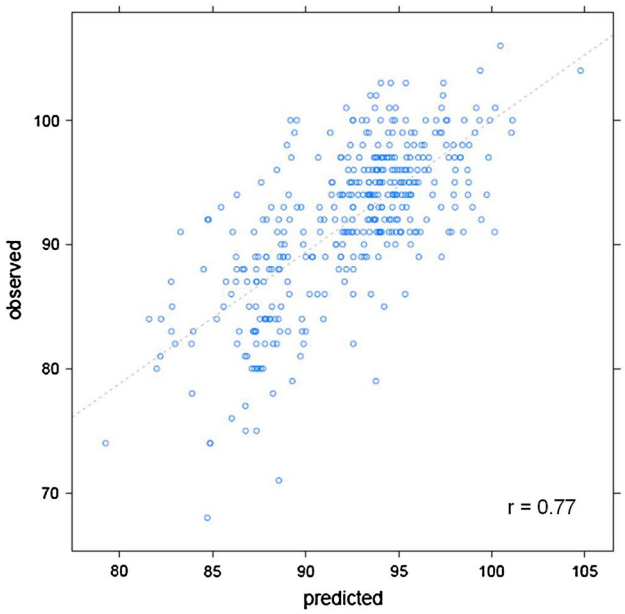

Fig. 3 Fitting GPC jointly with YLD and predicting the latter from the former showed a strong predictive correlation ($r = 0.77$)

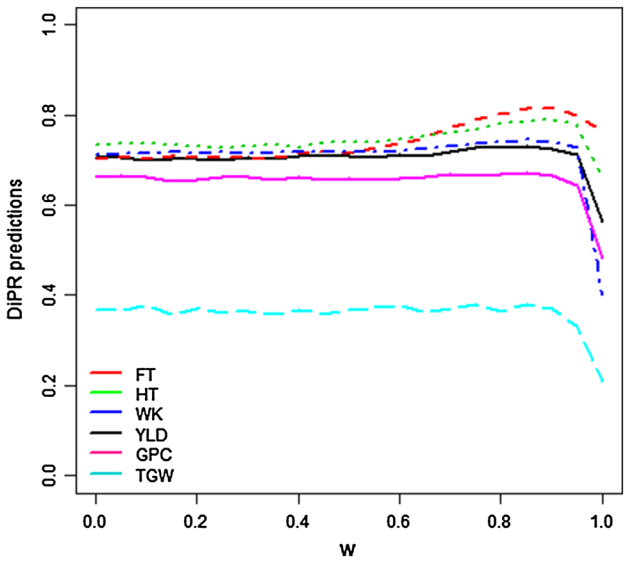

Fig. 4 Average DiPR correlations (based on five times tenfold cross-validation) for six traits against the standardised scaling factor for covariates (w)

et al. 2011). AM for YLD incorporating year of variety registration resulted in no significant marker-trait association, supporting YLD gains through time as the result of breeding, and the introduction of, and selection for, advantageous combinations of alleles. These YLD gains have come at the expense of GPC, a well-documented trade-off (Kibite and Evans 1984; Simmonds 1995). GPC is a breeding target both for health (grain quality) and its proposed economic benefit (reducing nitrogenous fertiliser application) (reviewed by Balyan et al. 2013). Early work suggested no genetic basis for the inverse relationship between YLD and GPC, proposing strong, environmental and physiological interactions (Kibite and Evans 1984; Simmonds 1995). Despite the identification (Joppa et al. 1997; Olmos et al. 2003) and refinement (Distelfeld et al. 2004, 2006) of the *Gpc-B1* locus on chromosome 6B governing GPC its utility for increasing GPC stably across environments independent of genetic background effects without agronomic losses remains unproven (Carter et al. 2012; Balyan et al. 2013). In the current study, a single significant marker association was detected on chromosome 6AS. Two of seven previous QTL studies in bi-parental populations have identified QTL for GPC in a similar location on chromosome 6AS (Sourdille et al. 2003; Huang et al. 2006). The loss of this association when YLD was included as a covariate, confirms an intricate relationship in the genetic control of YLD and GPC effects. Employing GS to fit GPC jointly with YLD and use the former to predict the latter gave higher cross-validated correlations than for predicting YLD in isolation, further supporting the strength of this relationship.

There are relatively few QTL studies reporting wheat YLD QTL across environments, a reflection of its complex nature and environmental dependency (Quarrie et al. 2005). In the current study, several significant markers were identified. A SNP on chromosome 3AS that was independent of HT effects was detected across experiments, and in single trials (FRA 2010; DEU 2011; GBR 2011). Bennett et al. (2012a, b) identified overlapping YLD QTLs in very similar locations on 3AS across southern Australian (2012a) and Mexican (2012b) environments. Snape et al. (2007) also identified a grain filling QTL on 3A that co-located with QTL for YLD in some of the environments tested. The genetic basis of this YLD association merits further investigation.

Rht-D1b previously shown to increase assimilate partitioning to the ear and ears/plant (Hedden 2003; Borrell et al. 1990) was significantly associated with YLD and had a significant allelic effect across the panel (+0.62t/ha). A SNP marker on 1BL was associated with YLD in the absence of the *Rht-D1* effect across experiments, and in GBR 2010, indicating it confers YLD effect independent of this source of reduced HT. Two additional associations were only identified in single years in GBR (4AL 2010; 6AS 2011) with the large GxE component of variation suggesting they may be effective only in specific environments. The vernalization gene *VRN-B3* (7B) was associated with increased YLD in DEU 2011 and GBR 2010, supporting the role of lifecycle duration on YLD, particularly where winter temperatures affect development via vernalization.

AM for FT identified five significant marker-trait associations, three of which could be ascribed to *Ppd-D1*,

Table 8 Cross-validated correlations for each of six traits where *w* (the scaling factor for the standardised covariates) is predicted from DArTs alone ($w = 0$), pooled, equally weighted markers ($w = 0.5$) and the 11 candidate gene markers alone ($w = 1$)

Trait	YLD	FT	HT	WK	TGW	GPC
DArT ($w = 0$)	0.71	0.70	0.73	0.71	0.37	0.66
Pooled ($w = 0.5$)	0.71	0.72	0.74	0.72	0.37	0.66
Candidates ($w = 1$)	0.56	0.76	0.67	0.40	0.21	0.48
DiPR	0.73	0.82	0.79	0.75	0.38	0.67
w at maximum	0.80	0.85	0.90	0.85	0.75	0.70

The DiPR and maximum *w* cross-validation correlations are also given

confirming its well-documented effect on FT under both field (Worland 1996; Bentley et al. 2013) and glasshouse (Bentley et al. 2011) conditions. Only when *Ppd-D1* was accounted for (as a covariate) were other markers associated with FT, namely *VRN-B3* and a 1BS DArT identified, despite previous authors documenting fine-tuning of FT via numerous *earliness per se* QTL (e.g. Griffiths et al. 2009) independent of *Ppd*. *VRN-B3,* a promoter retroelement insertion conferring early flowering, is known to influence FT in winter wheat (Yan et al. 2006). Griffiths et al. (2009) identified two independent FT QTL on 1B in a meta-QTL study, suggesting one was a possible homoeoallele of the stronger, more environmentally stable 1D FT QTL they also identified. The 1BS marker identified in association with FT in this study represents a potentially useful additional controller of FT that warrants further investigation and validation.

Application of GS to predict FT was largely unsuccessful with the exception of the FRA to FRA predictions. Given the dominance of *Ppd-D1* in the genetic control of FT, and its presence in the FRA variety set (40 % *Ppd-D1a*) GS was repeated to test whether *Ppd-D1* was also a primary determinant of prediction efficiency. This was done in two ways: the first used a simple method to account for the strong, single gene effect, namely using it to stratify the TP and test set. The second employed DiPR. Using the first approach, poor predictions from using only *Ppd-D1a* or *Ppd-D1b* variety TPs could be markedly improved when a mixture of *Ppd-D1* alleles were used in the TP and test set, and the GS model could distinguish, and predict, early and late types. Similarly, mixing *Rht-D1b* and *Rht-D1a* alleles in the TP and test sets improved HT prediction accuracy. This approach is similar to that proposed by Hayes et al. (2009) for improving GEBV accuracies in cattle test sets comprised of multiple breeds by ensuring representation of all breeds in the TP. Although degrees of earliness/lateness or short/tall could not be distinguished, these results have practical implications in the design of TP and test sets to predict these traits in elite material differing at key adaptation loci, and for other traits that are strongly influenced by single genes.

DiPR gave the largest increase in accuracy for FT, with a cross-validation correlation of 0.82, compared to 0.72 when the candidate and DArT markers were pooled. FT was predicted with greater accuracy from the candidates alone than from the DArT markers (0.76 compared to 0.72). Given the inclusion of *Ppd-D1* within the candidate marker set, which has a well established large effect on flowering time (e.g. Bentley et al. 2013), this is not surprising. In this dataset *Ppd-D1* was associated with population structure: its allele frequencies differed between countries and therefore, a proportion of the increased prediction accuracy from DiPR may come from more effective tagging of earlier and later flowering groups of lines.

HT was significantly associated with *Rht-D1b*, *Rht-B1b* and *Ppd-D1a*, all of which have been previously shown to reduce height (Wilhelm et al. 2013), and six other markers. The 5AL association was lost when *Rht-D1* was accounted for, supporting previous reports of major QTLs affecting adaptability on 5A (Kato et al. 2000; Huang et al. 2006; Marza et al. 2006). In contrast, the 6AS association remained significant when both *Rht-D1* and *Ppd-D1* were included as covariates, indicating it as an additional regulator of HT warranting further investigation. Marza et al. (2006) reported a QTL on 6A that consistently increased HT across environments. In general, applying GS to predict HT gave relatively good predictions. The lowest predictive power was observed in the GBR set, most likely due to the dominance of *Rht-D1b*, reducing the efficiency of predictions, as for *Ppd-D1*. DiPR also substantially improved the accuracy of prediction for HT where again, the candidate markers (particularly *Rht-D1*) are known to have major effects on the trait, and also tag population subdivision. This effect will occur for any QTL for which allele frequencies differ among populations. The hardiness trait WK was significantly associated with *Ppd-D1* and a SNP marker on 6BS, with early flowering *Ppd-D1a* lines being less hardy. Tolerance to low temperatures is a trait of economic importance, especially in northerly latitudes (such as DEU) making it an interesting regionally specific adaptive trait. It is acquired through the biochemical, structural and physiological processes that contribute to cold

acclimation (Båga et al. 2007). The effect of FT (conferred by *Ppd-D1*) on WK is possibly a reflection of its effect on lifecycle duration, although *Ppd-D1* has not been shown to effect seedling vigour, or early stages in wheat development (Bentley et al. 2013). Its effect to reduce hardiness in the current study probably reflects stabilising selection for this trait in the warmer FRA and GBR climates. Most of the genetic control of low temperature tolerance is reported to come from the group 5 chromosomes, primarily through the effects of vernalization (*Vrn-1*) loci, frost resistance (*Fr)* loci, and additional QTL proximal to *vrn-1* genes (Båga et al. 2007; Zhao et al. 2013). There have been no previous reports of genetic control of WK or other winter hardiness traits on 6BS, making this a key target for further investigation for increased WK tolerance. In addition, DEU to DEU predictions for WK were high, suggesting GS could also be successfully be used for breeding for this trait in DEU.

AM for TGW, a component of YLD, detected a single significant SNP marker on 2B. Groos et al. (2003) reported a TGW QTL on 2BL at six locations during a single year in FRA, indicating there is potential that this QTL is relevant to TGW in Northern Europe. GS for TGW resulted in low predictive correlations which was either a result of the low heritability, or the extremely small effect additive control of this trait. AM for SW detected no significant marker-trait associations, probably due to its low heritability. Early investigations on this trait found poor correlations between YLD and SW, attributing this to genotype and environmental variability and its interaction with grain characters (Hook 1984). The GS gave reasonable predictions within FRA, and to a lesser extent DEU, suggesting that in the absence of suitable genetic markers to apply in MAS, GS could be applied to improve SW. This would have the added advantage of allowing selections to be made without the need to phenotye adult plants. This is also the case for GPC, which had high predictive correlations.

The improvement from DiPR for YLD, GPC, TGW and WK was slight. The lack of substantial improvement is in spite of some of the candidate markers having a statistically significant effect on the trait: namely *Rht-D1* for YLD and *Ppd-D1* for WK. The increase in prediction accuracy from including the candidate markers ($w = 0.5$) compared to the DArT markers alone ($w = 0$) was very slight, confirming that treating markers for known QTL in the same manner as any other marker does not reveal their full benefit. It is important to note that we have not used the AM results directly to select markers for inclusion in trait prediction. Unless significant markers were detected through a process of cross-validation, such use would amount to "double dipping" and could give spurious importance to the significant markers in the trait prediction. DArT markers, used in the current study to obtain genome-wide marker coverage have now been superseded by high-density SNP arrays. The low frequency of markers, particularly on the D-genome (specifically 4D, 5D) reduce the efficiency of both AM and GS, and due to a rapid decay of LD mean significant marker-trait associations were unlikely to be detected on these chromosomes using our dataset. A higher marker density would allow the full benefits of the association mapping panel to be expressed and should increase the precision of the AM. Given that population sizes were relatively low, it would also be interesting to investigate effect on GS. However, even with advanced technology, low D-genome diversity still remains a barrier to marker development (Allen et al. 2012).

In accordance with the original objectives of the study, significant marker associations were detected for key traits across years and environments in our study and they represent targets for use in MAS. QTL mapping and AM facilitate locating genetic regions underlying complex traits, although the ability of subsequent MAS to combine them in realistically sized breeding programmes is somewhat limited. Therefore, additional, complementary tools are required to optimise selection. To this end, GS has vast potential, as supported by the results of this study in which GS gave good predictions for the majority of traits investigated. The TPs used (based on country or year of registration) were employed as convenient vehicles for grouping varieties. As population size is central to prediction accuracy, bootstrap resampling was also employed, finding that in most cases, a marked increase in the predictive correlation with increased sample size was achieved. These results suggest that the population sizes used for the GS were too small, providing insufficient genetic information. In this dataset low correlations were not necessarily indicative of low heritability or distant relatedness between the varieties in different countries, but rather, of a lack of statistical power in the GS model. Conversely, constant correlations over different sample sizes suggest that the performance of GS, whether good or bad, is primarily determined by other factors.

The desire to include additional sources of information in trait prediction is increasing. Zhao et al. (2014) described weighted best linear unbiased prediction (W-BLUP) which extends RR-BLUP by shrinking functional markers independently of a genome-wide set of markers, with shrinkage based on the ratios of residual variance to marker effect variances for each class. They report improved accuracy in predicting the FT of wheat hybrids from a TP of inbred parents by fitting separate random additive and dominance variance components to both sets based on a functional marker set that comprised *Ppd-D1*, *Rht-B1* and *Rht-D1*. Speed and Balding (2014) have described MultiBLUP, a computationally efficient procedure in which multiple random effects can be included, each corresponding to a separate subset of SNPs. When tested extensively on multiple traits in human and mouse datasets they demonstrate substantial increases

in accuracy compared to BLUP. They also developed a procedure: adaptive MultiBLUP in which SNP classes with different effect sizes are detected within the dataset, though as this process involves significance testing its use in TPs for subsequent genomic selection requires further testing.

Although, as here, the same dataset can be used for both AM and GS, it should be emphasised that their objectives differ. AM aims for unambiguous detection of QTL of major effect and for this purpose uses methods to eliminate marker-trait associations arising from population structure and pedigree relationships together with stringent significance thresholds. In contrast, GS incorporates population structure and kinship effects with no significance testing. For this reason great attention is being paid to training population design (e.g. Hickey et al. 2014) to maintain relevance to the population of candidate individuals for GS. In contrast, functional polymorphisms and markers in high LD with major QTL are likely to have a more consistent effect across populations; in the age of genomic selection their detection is still worthwhile. The panel used consists of elite, registered varieties. Across the key wheat regions of Northern Europe, kinship relationships are high, as seen in the population structure and as a result, predictive abilities were generally high. Where there was less genetic overlap (i.e. between DEU and GBR), predictions fell away. The ability to predict key traits forward in time in elite varieties derived from a common pool of ancestors and accounting for commonly used adaptive variation is of utility to the implementation of GS within and across commercial breeding programmes.

Author contributions IM, AG, NG, KB, SP, JS and SF designed the experiments. PH, GR, TB, JI, CP, JS, SF and KB coordinated the field experiments. FB, EW, RH, SF and JC conducted the genotyping. AB, MS, DB and IM analysed the data. AB wrote the paper.

Acknowledgments This work was supported by a Grant from the European Commission under the 7th Framework Programme for Research and Technological Development (FP7-212019). AB was funded by BB/I002561/1 from the UK Biotechnology and Biological Sciences Research Council whilst working on the manuscript. We thank the Editor and three anonymous reviewers for constructive comments for improving the manuscript.

Conflict of interest The authors declare that they have no conflict of interest.

References

Allen AM, Barker GLA, Wilkinson P, Burridge A, Winfield M, Coghill J, Uauy C, Griffiths S, Jack P, Berry S et al (2012) Discovery and development of exome-based, co-dominant single nucleotide polymorphism markers in hexaploid wheat (*Triticum aestivum* L.). Plant Biotech J 11:279–295

Båga M, Chodaparambi SV, Limin AE, Pecar M, Fowler DB, Chibbar RN (2007) Identification of quantitative trait loci and associated candidate genes for low-temperature tolerance in cold-hardy winter wheat. Funct Integr Genomics 7:53–68

Balyan HS, Gupta PK, Kumar S, Dhariwal R, Jaiswal V, Tyagi S, Agarwal P, Gahlaut V, Kumari S (2013) Genetic improvement of grain protein content and other health-related constituents of wheat grain. Plant Breed. doi:10.1111/pbr.12047

Beales J, Turner A, Griffiths S, Snape JW, Laurie DA (2007) A *Pseudo-Response Regulator* is misexpressed in the photoperiod insensitive *Ppd-D1a* mutant of wheat (*Triticum aestivum* L.). Theor Appl Genet 115:721–733

Beddington J (2009) Food, energy, water and the climate: a perfect storm of global events? http://www.dius.gov.uk/assets/goscience/docs/p/perfect-storm-paper.pdf. Accessed 9 Jan 2014

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Stat Soc B57:289–300

Bennett D, Izanloo A, Reynolds M, Kuchel H, Langridge P, Schnurbusch T (2012a) Genetic dissection of grain yield and physical grain quality in bread wheat (*Triticum aestivum* L.) under water-limited environments. Theor Appl Genet 125:255–271

Bennett D, Reynolds M, Mullan D, Izanloo A, Kuchel H, Langridge P, Schnurbusch T (2012b) Detection of two major grain yield QTL in bread wheat (*Triticum aestivum* L.) under heat, drought and high yield potential environments. Theor Appl Genet 125:1473–1485

Bentley AR, Turner AS, Gosman N, Leigh FJ, Maccaferri M, Dreisigacker S, Greenland A, Laurie DA (2011) Frequency of the photoperiod-insensitive *Ppd-A1a* alleles in tetraploid, hexaploid and synthetic hexaploid wheat germplasm. Plant Breed 130:10–15

Bentley AR, Horsnell R, Werner CP, Turner AS, Rose GA, Bedard C, Howell P, Wilhelm EP, Mackay IJ, Howells RM et al (2013) Short, natural, and extended photoperiod response in BC_2F_4 lines of bread wheat with different *Photoperiod-1* (*Ppd-1*) alleles. Exp Bot 64:1783–1793

Bernardo R (2009) Genomewide selection for rapid introgression of exotic germplasm in maize. Crop Sci 49:419–425

Bodmer WF (1987) Human genetics: the molecular challenge. BioEssays 7:41–45

Borrell A, Incoll LD, Dalling MJ (1990) The influence of the Rht_1 and Rht_2 alleles on the growth of wheat stems and ears. Ann Bot 67:103–110

Carter AH, Santra DK, Kidwell KK (2012) Assessment of the effects of the *Gpc-B1* allele on senescence rate, grain protein concentration and mineral content in hard red spring wheat (*Triticum aestivum* L.) from the Pacific Northwest Region of the USA. Plant Breed 131:62–68

Clark SA, Hickey JM, van der Werf JHJ (2011) Different models of genetic variation and their effect on genomic evaluation. Genet Sel Evol 43:18

Crossa J, Burgueno J, Dreisigacker S, Vargas M, Herrera-Foessel SA, Lillemo M, Singh RP, Trethowan R, Warburton M, Franco J et al (2007) Association analysis of historical bread wheat germplasm using additive genetic covariance of relatives and population structure. Genetics 177:1889–1913

Davison CA, Hinkley DV (1997) Bootstrap methods and their applications. Cambridge University Press, Cambridge

de Froidmont D (1998) A co-dominant marker for the 1BL/1RS wheat-rye translocation via multiplex PCR. J Cereal Sci 27:229–232

DEFRA (2012) http://webarchive.nationalarchives.gov.uk/20130123162956/http:/www.defra.gov.uk/statistics/files/defra-stats-food-farm-landuselivestock-farmingstats-june-statsrelease-jun2012uk-provcrops-1210151.pdf. Accessed 9 Jan 2014

Diaz A, Zikhali M, Turner AS, Isaac P, Laurie DA (2012) Copy number variation affecting the *Photoperiod-B1* and *Vernalisation-A1*

genes is associated with altered flowering time in wheat (*Triticum aestivum*). PLoS ONE 7:e33234

Distelfeld A, Uauy C, Olmos S, Schlatter AR, Dubcovsky J, Fahima T (2004) Microcolinearity between a 2-cM region encompassing the grain protein content locus *Gpc-6B1* on wheat chromosome 6B and a 350-kb region on rice chromosome 2. Funct Integr Genomics 4:59–66

Distelfeld A, Uauy C, Fahima T, Dubcovsky J (2006) Physical map of the wheat high-grain protein content gene *Gpc-B1* and development of a high-throughput molecular marker. New Phytol 169:753–763

Ellis MH, Spielmeyer W, Gale KR, Rebetzke GJ, Richards RA (2002) "Perfect" markers for the *Rht-B1b* and *Rht-D1b* dwarfing genes in wheat. Theor Appl Genet 105:1038–1042

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14:2611–2620

Fu D, Szuchs P, Yan L, Helguera M, Skinner JS, von Zitzewitz J, Hayes PM, Dubcovsky J (2005) Large deletions within the first intron in *VRN-1* are associated with spring growth habit in barley and wheat. Mol Genet Genomics 273:54–65

Fulton TM, Chunwongse J, Tanksley SD (1995) Microprep protocol for extraction of DNA from tomato and other herbaceous plants. Plant Mol Biol Rep 13:207–209

Goeman JJ, Meijer R, Chaturvedi N (2012) Penalized. R package version 0.9-41. http://cran.r-project.org/web/packages/penalized/penalized.pdf

Griffiths S, Simmonds J, Leverington M, Wang Y, Fish L, Sayers L, Alibert L, Orford S, Wingen L, Herry L et al (2009) Meta-QTL analysis of the genetic control of ear emergence in elite European winter wheat germplasm. Theor Appl Genet 119:383–395

Groos C, Robert N, Bervas E, Charmet G (2003) Genetic analysis of grain protein-content, grain yield and thousand-kernel weight in bread wheat. Theor Appl Genet 106:1032–1040

Habier D, Tetens J, Seefried F-R, Lichtner P, Thaller G (2010) The impact of genetic relationship information on genomic breeding values in German Holstein cattle. Genet Sel Evol 42:5

Hastie T, Tibshirani R, Narasimhan B and Chu G (2013) Impute: Imputation for microarray data. R package version 1.34.0. http://master.bioconductor.org/packages/release/bioc/manuals/impute/man/impute.pdf

Hayes BJ, Bowman PJ, Chamberlain AJ, Goddard ME (2009) Genomic selection in dairy cattle: progress and challenges. J Dairy Sci 92:433–444

Hedden P (2003) The genes of the green revolution. Trends Genet 19:5–9

HGCA (2012). http://www.hgca.com/document.aspx?fn=load&media_id=8124&publicationId=5624. Accessed 9 Jan 2014

Hickey JM, Dreisigacker S, Crossa J, Hearne S, Babu R, Prasanna BM, Grondona M, Zambelli A, Windhausen VS, Mathews K, Gorjanc G (2014) Evaluation of genomic selection training population designs and genotyping strategies in plant breeding programs using simulation. Crop Sci 54:1476–1488

Hook SCW (1984) Specific weight and wheat quality. J Sci Food Agric 35:1136–1141

Huang XQ, Cloutier S, Lycar L, Radovanovic N, Humphreys DG, Noll JS, Somers DJ, Brown PD (2006) Molecular detection of QTLs for agronomic and quality traits in a doubled haploid population derived from two Canadian wheats (*Triticum aestivum* L.). Theor Appl Genet 113:753–766

Huang BE, George AW, Forrest KL, Kilian A, Hayden MJ, Morell MK, Cavanagh CR (2012) A multiparent advanced generation inter-cross population for genetic analysis in wheat. Plant Biotech J 10:826–839

Jannink JL (2007) Identifying quantitative trait locus by genetic background interactions in association studies. Genetics 176:553–561

Joppa LR, Du C, Hart GE, Hareland GA (1997) Mapping gene(s) for grain protein in tetraploid wheat (*Triticum turgidum* L.) using a population of recombinant inbred chromosome lines. Crop Sci 37:1586–1589

Kang HM, Zaitlen NA, Wade CM, Kirby A, Heckerman D, Daly MJ, Eskin E (2008) Efficient control of population structure in model organism association mapping. Genetics 178:1709–1723

Kato K, Miura H, Sawada S (2000) Mapping QTLs controlling grain yield and its components on chromosome 5A of wheat. Theor Appl Genet 101:1114–1121

Kibite S, Evans LE (1984) Causes of negative correlations between grain yield and grain protein concentration in common wheat. Euphytica 33:801–810

Korzun V, Roder MS, Ganal MW, Worland AJ, Law CN (1998) Genetic analysis of the dwarfing gene (*Rht8*) in wheat. Part I. Molecular mapping of *Rht8* on the short arm of chromosome 2D of bread wheat. (*Triticum aestivum* L.). Theor Appl Genet 96:1104–1109

Lewis CM (2002) Genetic association studies: design, analysis and interpretation. Brief Bioinform 3:146–153

Lipka AE, Tian F, Wang Q, Peiffer J, Li M, Bradbury PJ, Gore M, Buckler ES, Zhang Z (2012) GAPIT: genome association and prediction integrated tool. Bioinformatics 28:2397–2399

Maccaferri M, Sanguineti MC, Demontis A, El-Ahmed A, Garcia del Moral L, Maalouf F, Nachit M, Nserallah N, Ouabbou H, Rhouma S et al (2011) Association mapping in durum wheat grown across a broad range of water regimes. J Exp Bot 62:409–438

Mackay I, Horwell A, Garner J, White J, McKee J, Philpott H (2011) Reanalyses of the historical series of UK variety trials to quantify the contributions of genetic and environmental factors to trends and variability in yield over time. Theor Appl Genet 122:225–238

Marchini JL (2013). Popgen: statistical and population genetics. R package version 1.0-3. http://cran.r-project.org/web/packages/popgen/popgen.pdf

Marza F, Bai G-H, Carver BF, Zhou W-C (2006) Quantitative trait loci for yield and related traits in the wheat population Ning7840 x Clark. Theor Appl Genet 112:688–698

Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–1829

Neumann K, Kobiljski B, Dencic S, Varshney RK, Börner A (2011) Genome-wide association mapping: a case study in bread wheat (*Triticum aestivum* L.). Mol Breed 27:37–58

Olmos S, Distelfeld A, Chicaiza O, Schlatter AR, Fahima T, Echenique V, Dubcovsky J (2003) Precise mapping of a locus affecting grain protein content in durum wheat. Theor Appl Genet 107:1243–1251

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959

Quarrie SA, Steed A, Calestani C, Semikhodskii A, Lebreton C, Chinoy C, Steele N, Pljevljakusic D, Waterman E, Weyen J et al (2005) A high-density genetic map of hexaploid wheat (*Triticum aestivum* L.) from the cross Chinese Spring x SQ1 and its use to compare QTLs for grain yield across a range of environments. Theor Appl Genet 110:865–880

Reif JC, Gowda M, Maurer HP, Longin CFH, Korzun V, Edmeyer E, Bothe R, Pietsch C, Würschum T (2011) Association mapping for quality traits in soft winter wheat. Theor Appl Genet 122:961–970

Segura V, Vilhjálmsson BJ, Platt A, Kortel A, Seren Ü, Long Q, Nordborg M (2012) An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. Nat Genet 44:825–832

Simmonds NW (1995) The relation between yield and protein in cereal grain. J Sci Food Agric 67:309–315

Snape JW, Foulkes MJ, Simmonds J, Leverington M, Fish LJ, Wang Y, Ciavarrella M (2007) Dissecting gene x environmental effects on wheat yields via QTL and physiological analysis. Euphytica 154:401–408

Sourdille P, Cadalen T, Guyomarch H, Snape JW, Perretant MR, Charmet G, Boeuf C, Bernard S, Bernard M (2003) An update of the Courtot x Chinese Spring intervarietal molecular marker linkage map for the QTL detection of agronomic traits in wheat. Theor Appl Genet 106:530–538

Speed D, Balding DJ (2014) MultiBLUP: improved SNP-based prediction for complex traits. Genome Res. doi:10.1101/gr.169375.113

Wilhelm EP, Turner AS, Laurie DA (2009) Photoperiod insensitive *Ppd-A1a* mutations in tetraploid wheat (*Triticum durum* Desf.). Theor Appl Genet 118:285–294

Wilhelm EP, Boulton MI, Al-Kaff N, Balfourier F, Bordes J, Greenland A, Powell W, Mackay IJ (2013) *Rht-1* and *Ppd-D1* associations with height, GA sensitivity, and days to heading in a worldwide bread wheat collection. Theor Appl Genet. doi:10.1007/s00122-013-2130-9

Wimmer V, Albrecht T, Auinger HJ, Schön CC (2012) synbreed: a framework for the analysis of genomic prediction data using R. Bioinformatics 18:2086–2087

Worland AJ (1996) The influence of flowering time genes on environmental adaptability in European wheat. Euphytica 89:49–57

Yan L, Helguera M, Kato K, Fukuyama S, Sherman J, Dubcovsky J (2004) Allelic variation at the *VRN-1* promoter region in polyploidy wheat. Theor Appl Genet 118:1677–1686

Yan L, Fu C, 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*. PNAS 103:19581–19586

Zadoks JC, Chang TT, Konzak CF (1974) A decimal code for the growth stages of cereals. Weed Res 14:415–421

Zhao K, Aranzana MJ, Kim S, Lister C, Shindo C, Tang C, Toomaijan C, Zheng H, Dean C, Marjoram P, Nordborg M (2007) An Arabidopsis example of association mapping in structured samples. PLoS Genet 3:e4

Zhao Y, Gowda M, Wurschum T, Longin CFH, Korzun V, Kollers S, Schachschneider R, Zeng J, Fernando R, Dubcovsky J et al (2013) Dissecting the genetic architecture of frost tolerance in Central European winter wheat. J Exp Bot 64:4453–4460

Zhao Y, Mette MF, Gowda M, Longin CFH, Reif JC (2014) Bridging the gap between marker-assisted and genomic selection of heading time and plant height in hybrid wheat. Heredity 112:638–645