

Article

Genomic Prediction and Genome-Wide Association Studies of Flour Yield and Alveograph Quality Traits Using Advanced Winter Wheat Breeding Material

Peter S. Kristensen ^{1,*} , Just Jensen ² , Jeppe R. Andersen ¹, Carlos Guzmán ³, Jihad Orabi ¹ and Ahmed Jahoor ^{1,4}

¹ Nordic Seed A/S, 8300 Odder, Denmark

² Department of Molecular Biology and Genetics, Aarhus University, 8830 Tjele, Denmark

³ Departamento de Genética, Escuela Técnica Superior de Ingeniería Agronómica y de Montes, Edificio Gregor Mendel, Campus de Rabanales, Universidad de Córdoba, CeiA3, 14071 Córdoba, Spain

⁴ Department of Plant Breeding, The Swedish University of Agricultural Sciences, 23053 Alnarp, Sweden

* Correspondence: pskr@nordicseed.com

Received: 8 August 2019; Accepted: 29 August 2019; Published: 31 August 2019



Abstract: Use of genetic markers and genomic prediction might improve genetic gain for quality traits in wheat breeding programs. Here, flour yield and Alveograph quality traits were inspected in 635 F₆ winter wheat breeding lines from two breeding cycles. Genome-wide association studies revealed single nucleotide polymorphisms (SNPs) on chromosome 5D significantly associated with flour yield, Alveograph P (dough tenacity), and Alveograph W (dough strength). Additionally, SNPs on chromosome 1D were associated with Alveograph P and W, SNPs on chromosome 1B were associated with Alveograph P, and SNPs on chromosome 4A were associated with Alveograph L (dough extensibility). Predictive abilities based on genomic best linear unbiased prediction (GBLUP) models ranged from 0.50 for flour yield to 0.79 for Alveograph W based on a leave-one-out cross-validation strategy. Predictive abilities were negatively affected by smaller training set sizes, lower genetic relationship between lines in training and validation sets, and by genotype–environment (G×E) interactions. Bayesian Power Lasso models and genomic feature models resulted in similar or slightly improved predictions compared to GBLUP models. SNPs with the largest effects can be used for screening large numbers of lines in early generations in breeding programs to select lines that potentially have good quality traits. In later generations, genomic predictions might be used for a more accurate selection of high quality wheat lines.

Keywords: wheat breeding; baking quality; Alveograph; flour yield; genomic selection; GWAS

1. Introduction

Baking quality of wheat is a complex trait controlled by many genes with minor effects and few genes with larger effects [1,2]. The amount and the composition of gluten proteins have large effects on baking quality of wheat. The major gluten loci are the high molecular weight glutenins (HMWGs) *Glu-A1*, *Glu-B1*, and *Glu-D1* and the low molecular weight glutenins (LMWGs) *Glu-A3*, *Glu-B3*, and *Glu-D3* [3,4]. Milling quality and water absorption are affected by the hardness of the grain. Grain hardness is, to a large extent, controlled by the *Hardness* locus on chromosome 5D, consisting of the genes *Pina-D1*, *Pinb-D1*, and *Gsp-1* [5,6].

Breeding for improved wheat quality is challenging, because phenotyping of most quality traits requires laborious analyses of relatively large amounts of grain using expensive equipment. Baking tests can be used for evaluating the quality of wheat lines by determining bread loaf volume and texture. However, breeding programs typically do not have the resources to perform baking tests with

large numbers of lines [7]. Another way of testing baking quality is rheological measurements with, for example, an Alveograph. Here, a piece of dough is inflated with air into a bubble, and dough tenacity (Alveo P), extensibility (Alveo L), and strength (Alveo W) are determined. Alveo W and the ratio of Alveo P/L are correlated with bread loaf volume and can give good indications of the quality of wheat lines [8]. Together with protein content and grain hardness, the Alveograph traits can be used for predicting the potential end-use for particular wheat lines [9]. Previously, quantitative trait loci (QTL) for the Alveograph traits were identified on many chromosomes, e.g., on chromosome 1B and 5D for both Alveo P and W. Additionally, QTL were identified on chromosomes 3A, 3B, and 5B for Alveo W and on chromosomes 2A and 2B for Alveo P [10–13]. Groos et al. (2004) [10], Tadesse et al. (2015) [14], and Zanetti et al. (2001) [13] identified QTL for Alveo L on chromosomes 2B, 3B, 4A, and 5B. Similarly, QTL for flour yield were identified on most chromosomes. In several studies of spring or winter wheat, QTL for flour yield were identified on chromosomes 1B, 2A, 2B, 3B, 4A, 5A, 5D, or 6A [2,12,15–20].

Those in Denmark are allowed to use more nitrogen for the cultivation of wheat for bread production if certain criteria are fulfilled. The cultivar must be on a list that has been approved by The Danish Agricultural Agency. For a cultivar to be approved, it must have a bread volume and a Zeleny sedimentation value of, at minimum, 90% of the average of the cultivars that have already been approved, and it must not have sticky dough. In addition, the grain used for bread production should, as a minimum, have a test weight of 78 kg/hL, a falling number of 275 s, and a protein content of 11.5% [21,22]. Therefore, it is important for breeders to be able to select lines with high quality.

The use of DNA markers to predict quality traits can reduce costs and enable higher selection intensities compared to having to do the phenotypic measurements. Thereby, higher numbers of wheat lines might be tested in breeding programs. Marker assisted selection based on few DNA markers can be effective for traits that are controlled by few QTL with large effects [23]. However, many traits are more complex and are controlled by many QTL with small effects. Furthermore, major QTL might be fixed in breeding programs and can therefore not be used for selection. In such cases, genomic predictions based on a large number of genome-wide markers could be a good approach. A training set of lines that have been both genotyped and phenotyped is needed to train a model for prediction of genomic estimated breeding values (GEBVs) in other genotyped lines (validation or breeding set) [24]. Different types of models can be used for genomic predictions, and their accuracies depend on the genetic architecture of the trait and on the relatedness of lines in the training and validation set [25,26]. In genomic best linear unbiased prediction (GBLUP) models, breeding values are predicted based on a genomic relationship matrix (G-matrix) [27]. In Bayesian models, marker effects can be assigned prior distributions that might fit better than a normal distribution for certain traits or populations [28]. Genomic prediction was first implemented in cattle breeding and is now widely used in animal breeding programs [28]. Many studies have been focused on the effectiveness and the optimal use of genomic predictions in different plant species and for different traits [29–32]. Generally, heritabilities and predictive abilities are lower for complex traits, such as grain yield, than for more simple traits, such as heading date or resistance against certain diseases, which are controlled by both minor and major genes [33–35]. For wheat quality traits, predictive abilities are often moderate to high [25,36–38]. However, the use of markers with large effects could likely be done in a better way than in the standard GBLUP or Bayesian models, and the optimal implementation strategy might differ between traits and breeding programs [26,34,39,40]. The size and the composition of the training set is crucial for accurate genomic predictions [29,38,41]. A recent study of quality traits and grain yield using wheat hybrids from each of the German quality classes E, A, B, and C reported that genomic predictions within one quality class worked well when the training set contained individuals from the same class. However, for predictions between quality classes, individuals from each of the classes should be included to obtain predictive abilities as high as for the within-class predictions [42].

The aims of the present study were to identify SNPs affecting the baking quality traits flour yield and Alveographs P, L, and W and to evaluate and compare genomic prediction models in order to facilitate implementation of genomic selection for baking quality traits in breeding programs.

2. Materials and Methods

2.1. Plant Material

In total, 635 F₆ winter wheat lines from two breeding cycles in the Danish plant breeding company Nordic Seed A/S (Holeby, Denmark) were used in this study. The 321 lines of the first breeding cycle (set2014) were harvested in 2014, and the 314 lines of the second breeding cycle (set2015) were harvested in 2015. Six out of 96 crossing parents were used for both sets (years), while the remaining 90 crossing parents were used for one of the sets only. Each line was grown in an unreplicated 9.9 m² plot at Lolland, Denmark, following standard Danish agricultural practices. Approximately 180 kg of nitrogen were applied per hectare during the growth season, and no irrigation was used.

2.2. Phenotyping

Phenotyping was done at the Wheat Chemistry and Quality Laboratory at International Maize and Wheat Improvement Center (CIMMYT), Mexico. Grain samples were conditioned to 13.5% moisture content and then milled one time using a Brabender Quadrumat Jr. (Brabender GmbH & Co. KG, Duisburg, Germany). Flour yield was measured as the percentage of refined flour obtained from each grain sample after the bran fraction was sieved away through a 75 µm mesh sieve. A Chopin Alveograph (Tripette and Renaud, Villeneuve-la-Garenne, France) was used to obtain the Alveograph traits P, L, and W (Alveo P is dough tenacity, Alveo L is dough extensibility, and Alveo W is dough strength) using a modified version of the American Association of Cereal Chemists (AACC) method 54-30A [8,43]. Flour was mixed with a saltwater solution to form a dough, which was cut into discs. After resting 20 min at 25 °C, the dough discs were inflated with air, thus the dough expanded as a bubble. During inflation of each disc, a curve was recorded of pressure inside the bubble until it burst. Alveo P was the maximum height of the curve, Alveo L was the length of the curve, and Alveo W was the area under the curve. The coefficient of variation was calculated for each trait by dividing the standard deviation of the raw phenotypes with the mean.

2.3. Genotyping

DNA extraction was performed with a modified cetyl trimethylammonium bromide (CTAB) method [44] using leaves of three bulked, two-week-old seedlings for each line. Genotyping was done by TraitGenetics (Gatersleben, Germany) with the 15K Illumina Infinium iSelect HD Custom Genotyping BeadChip technology. The 13,006 called SNP markers were edited for minor allele frequency (MAF) > 1% and missing values < 10%, and the remaining 10,802 SNPs were used for the analyses. For each line, at least 90% of the SNPs were successfully genotyped.

2.4. Statistical Analysis

Genome-wide associations were studied using single marker regression. The following model was run for each of the 10,802 SNPs:

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{w}_i a_i + \mathbf{Z}_1 \mathbf{u} + \mathbf{e} \quad (1)$$

where \mathbf{y} is a vector of observed phenotypes, \mathbf{X} and \mathbf{Z}_1 are design matrices, \mathbf{b} is a vector of fixed effects (mean and year/set), \mathbf{w}_i is the vector of genotypes of the i^{th} SNP coded as 1, 0, -1, a_i is the additive genetic effect of the i^{th} SNP, \mathbf{u} is a vector of additive genetic effects of the lines ($\mathbf{u} \sim N(0, \mathbf{G}_1 \sigma_g^2)$, where \mathbf{G}_1 is a G-matrix (genomic relationship matrix) and σ_g^2 is additive genetic variance), and \mathbf{e} is a vector of random residual effects ($\mathbf{e} \sim N(0, \mathbf{I} \sigma_e^2)$, where \mathbf{I} is an identity matrix and σ_e^2 is the residual variance). Model effects and variance components were estimated by restricted maximum likelihood using the software package DMU [45].

For each chromosome, a G-matrix was calculated based only on the SNPs mapped to the remaining chromosomes and then used to correct for structure when analyzing the SNPs mapped to the excluded

chromosome. This was done so that the SNP effect was not included twice in the model. G-matrices were calculated using the first method proposed by Van Raden [27]:

$$G = \frac{Z_2 Z_2'}{2 \sum p_i (1 - p_i)} \quad (2)$$

where p_i is the MAF of i^{th} marker, $Z_2 = M - P$, M is a matrix with the marker alleles coded as 1, 0, -1, and P is a matrix where the i^{th} column contains the MAF of SNP i calculated as $2(p_i - 0.5)$. Missing genotypes were set to 0 in matrix Z_2 .

Genomic inflation factors, λ_{IF} , were calculated for each trait by dividing the observed median value of the chi-squared statistic for the SNPs with the expected median value [46]. The inflation factor λ_{IF} was used to correct the p -values for inflation by dividing the chi-squared statistic with λ_{IF} and then re-calculating the p -values. The significance threshold was set using a Bonferroni correction: 5% divided by number of SNPs ($0.05/10,802 = 4.6 \times 10^{-6}$).

DNA sequences surrounding the significantly associated SNPs were blasted against the annotated reference genome of the bread wheat variety *Chinese Spring*, IWGSC RefSeq v1.0 [47], using the BLAST tool of EnsemblPlants [48].

A Bayesian Power Lasso model where all SNPs were fitted simultaneously was also used for genome-wide association analyses in the Bayz software [49]:

$$y = Xb + Z_3u + e \quad (3)$$

where y is a vector of observed phenotypes, b is a vector of the mean + year/set effect with design matrix X , Z_3 is a matrix of the alleles of the SNPs coded as 0, 1, 2, u is a vector of additive genetic SNP effects, and e is a vector of residual effects. The prior distribution of SNP effects was specified as an exponential power distribution [50]:

$$p(u) = \prod_{i=1}^m \frac{1}{2} \lambda_{RP} e^{-\lambda_{RP}|u_i|^\beta} \quad (4)$$

where m is the number of markers, and β is shape parameter to control the sparsity, which affects the shrinkage of the SNP effects. Setting β to 1 makes the model equivalent to a standard Bayesian Lasso model. If β is set to less than 1, the difference between large and small marker effects can be increased [50]. Models with β of 0.2, 0.4, 0.8, and 1.0 were run, and the Deviance Information Criterion was used to determine the optimal β for each trait [51]. Residual effects were assigned a normal prior distribution. The residual variance, the mean, the year/set effect, and the rate parameter, λ_{RP} , were assigned flat prior distributions. Model parameters were estimated using Markov Chain Monte Carlo with a length of 100,000 with 30,000 cycles as burn-in. The tool pbayz supplied with Bayz was used to compute posterior means, and the R package CODA was used to check for convergence to the posterior distribution [52].

Genomic predictions based on all 10,802 SNPs were conducted using the Bayesian Power Lasso model (3) and using a GBLUP model:

$$y = Xb + Z_4u + e \quad (5)$$

where y is a vector of observed phenotypes, X and Z_4 are design matrices, b is a vector of fixed effect (mean and year/set), u is a vector of additive genetic effects ($u \sim N(0, G_2 \sigma_g^2)$, where G_2 is a G-matrix computed as above (2) using all SNPs, σ_g^2 is additive genetic variance), and e is a vector of random residual effects ($e \sim N(0, I\sigma_e^2)$).

Model effects and variance components for the GBLUP and Bayesian Power Lasso models were estimated by DMU and Bayz packages, respectively. For the GBLUP models, the narrow-sense genomic heritability corresponding to records of single plots was calculated as:

$$h^2 = \frac{d(G_2)\sigma_g^2}{d(G_2)\sigma_g^2 + \sigma_e^2} \quad (6)$$

where $d(G_2)$ is the average diagonal element of the G-matrix (calculated using all SNPs), σ_g^2 is additive genetic variance, and σ_e^2 is residual variance.

The following cross-validations (CVs) were used to study the effectiveness of possible strategies for implementing genomic selection in breeding programs:

Leave-one-out (LOO): The GEBV of each line was predicted from the rest of the lines. The training set used in the LOO strategy was as large as possible (634 lines), and the genetic relationship between lines in the training and validation set was higher compared to the other CV strategies.

Leave-family-out (LFO): The GEBVs of lines in each half-sib family were predicted from lines of the remaining families. The average size of the half-sib families was 46 lines. Using this strategy, the effect of the genetic relationship between the lines in training and validation sets was studied.

Leave-set-out (LSO): The GEBVs of lines in each set were predicted from lines from the other set. The training set sizes were 314 or 321 lines. The LSO CV strategy was used for studying the predictive ability when GEBVs of lines from one breeding cycle were predicted from lines from another breeding cycle.

k-fold: The lines were randomly divided into k folds (2, 5, or 10) of equal size. The GEBVs of lines in each fold were predicted from lines in the other folds. The training set sizes were approximately 318 lines for the 2-fold, 508 lines for the 5-fold, and 572 lines for the 10-fold. Approximately half of the lines in the training sets were from set2014 and half from set2015. The k-fold CV strategy was used for studying the effect of the training set size.

Furthermore, the effect of the training set size was studied by selecting from 10% to 90% of the 635 lines as a training set for genomic predictions using LOO CV. The lines were randomly selected, and selection and predictions were repeated 100 times for each 10% interval.

Correlations between observed phenotypes corrected for fixed effects and GEBVs were calculated to determine predictive abilities of the models and were compared with the maximum correlation (the square root of the narrow-sense genomic heritability). Biases of the genomic predictions were calculated as the deviation from the expectation of the slope (1.0) of the regression line of the corrected phenotypes on the GEBVs.

Genomic feature models with two G-matrices were tested in order to possibly utilize the most significant SNPs better [53,54]. The lines were randomly divided in two folds. The lines of one fold were used for genome-wide association studies (GWAS), and the remaining lines were used for LOO genomic predictions. The SNPs used for computing the two G-matrices were selected for each trait based on their p -value in the GWAS. The most significant SNPs were used for one G-matrix, and the remaining SNPs were used for the other. The following thresholds for number of SNPs to include in the group of most significant were tested: 5, 10, 50, 100, 500, 1000, 3000, 5000, 7000, 10,000, and all 10,802 SNPs. The following model was used for the genomic predictions:

$$y = Xb + Z_5s + Z_6n + e \quad (7)$$

where y is a vector of observed phenotypes, X , Z_5 , and Z_6 are design matrices, b is a vector of fixed effect (mean and year/set), s and n are vectors of additive genetic effects ($s \sim N(0, G_s\sigma_{g_s}^2)$ and $n \sim N(0, G_n\sigma_{g_n}^2)$), where G_s and G_n are G-matrices computed as above (2) using significant (G_s) or nonsignificant (G_n) SNPs, $\sigma_{g_s}^2$ and $\sigma_{g_n}^2$ are additive genetic variances, and e is a vector of random residual effects ($e \sim N(0, I\sigma_e^2)$).

3. Results

3.1. Phenotyping and Genotyping

A total of 635 F₆ winter wheat lines from two different breeding cycles (set2014 and set2015) were phenotyped for the quality traits flour yield and Alveos P, L, and W (Table S1). The phenotypic distribution for each trait is shown in Figure 1. Phenotypic variation was higher for the Alveograph traits than for flour yield (Table 1). The wheat lines were genotyped for 10,802 SNPs (Table S2). A G-matrix was computed using all SNPs. A dendrogram based on the G-matrix showed that the lines were genetically related both within and between sets (Figure 2). In total, the two sets consisted of 159 full-sib families with an average of four full-sibs per family.

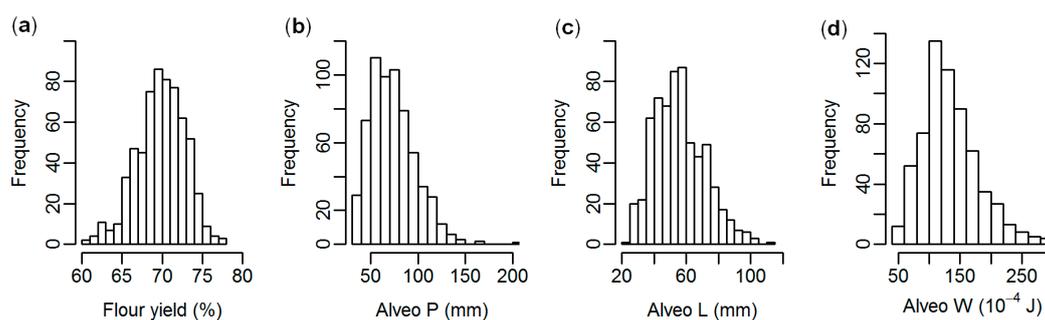


Figure 1. Distribution of phenotypes. (a): Flour yield, (b): Alveo P (dough tenacity), (c): Alveo L (dough extensibility), (d): Alveo W (dough strength).

Table 1. Mean, range, coefficient of variation of phenotypic data.

Trait	Mean	Range	Coefficient of Variation (%)
Flour yield (%)	69.9	60.8–77.3	4.3
Alveo P (mm)	73.8	31–201	32.5
Alveo L (mm)	56.1	21–114	28.0
Alveo W (10 ⁻⁴ J)	134.2	40–293	33.4

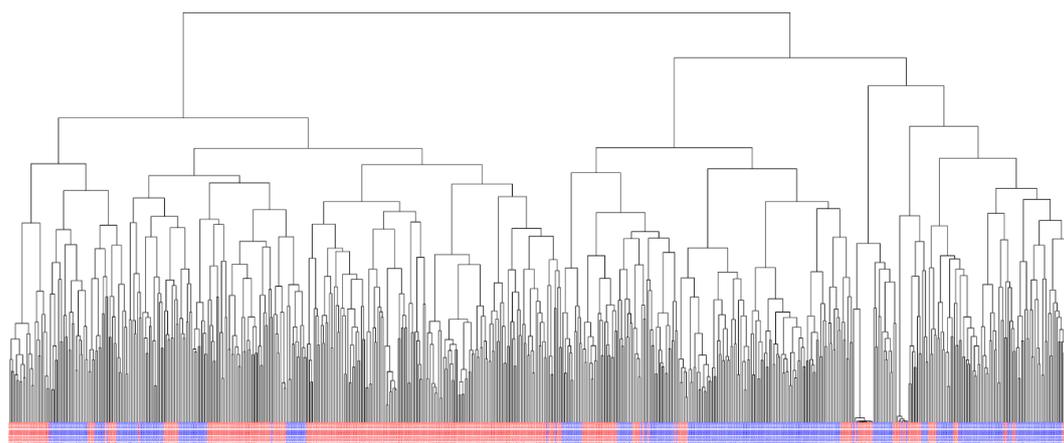


Figure 2. Dendrogram of the 635 wheat lines based on the G-matrix. Lines from set2014 are shown in red and lines from set2015 are shown in blue.

For all traits, additive genetic variance was observed. Variance components were estimated based on GBLUP models and were used for estimation of narrow-sense genomic heritabilities. Heritabilities ranged from 0.38 for flour yield to 0.72 for Alveo W (Table 2).

Table 2. Additive genetic variance components (Var_g), residual variance components (Var_e), and narrow-sense genomic heritabilities (h^2) estimated from genomic best linear unbiased prediction (GBLUP) models.

Trait	Var_g	Var_e	h^2
Flour yield	2.8 ± 0.30	4.6 ± 0.36	0.38 ± 0.046
Alveo P	327.6 ± 23.7	146.8 ± 15.3	0.69 ± 0.053
Alveo L	138.6 ± 11.3	94.1 ± 8.9	0.60 ± 0.054
Alveo W	1108.3 ± 74.4	429.0 ± 45.3	0.72 ± 0.050

3.2. GWAS

Single marker regression was performed for each of the 10,802 SNPs (Figure 3). Two linked SNPs on chromosome 5DS were significantly associated with flour yield, Alveo P, and Alveo W. On chromosome 1DL, significantly associated SNPs were identified for Alveos P and W. Additionally, SNPs associated with Alveo W were identified both on the short arm and on the long arm of chromosome 1B. A region on chromosome 4AL was significantly associated with Alveo L. The frequencies of the SNP alleles that were positively associated with each trait ranged from 13% to 64% (Table 3). A large difference was observed in Alveo P and in Alveo W for lines with the positive alleles of all significant SNPs compared to lines with the negative allele of one or more of the SNPs (Tables 4 and 5).

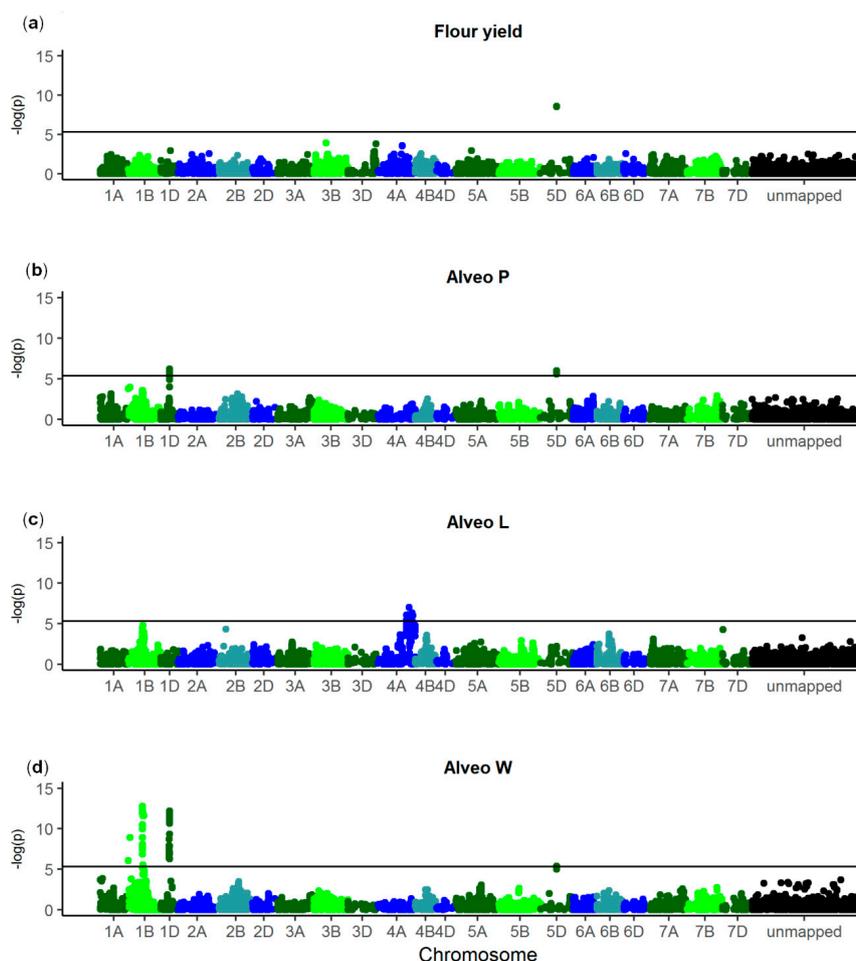


Figure 3. Manhattan plots of $-\log_{10}(p\text{-values})$ after correction with the genomic inflation factors. (a): Flour yield, (b): Alveo P, (c): Alveo L, (d): Alveo W. The horizontal lines show the Bonferroni corrected significance threshold. Total number of single nucleotide polymorphisms (SNPs): 10,802. Last bin is unmapped SNPs.

Table 3. SNPs significantly associated with each of the quality traits based on single marker regressions.

Trait	SNP ¹	Chromo-some	p Value	Allele Frequency (%) ²	Genetic Effect	Explained Genetic Variance (%)	Gene ID	Annotation
Flour yield	NOS_WW_SNP_688	5D	2.44×10^{-9}	28	0.96 ± 0.14	13.3	TraesCS5D02G004300	Nontranslating coding sequence
Alveo P	NOS_WW_SNP_5054	1D	6.22×10^{-7}	15	7.99 ± 1.29	5.0	TraesCS1D02G322300	Uncharacterized protein
	NOS_WW_SNP_688	5D	1.02×10^{-6}	28	6.01 ± 0.99	4.4	TraesCS5D02G004300	Nontranslating coding sequence
Alveo L	NOS_WW_SNP_2731	4A	9.59×10^{-8}	37	4.87 ± 0.82	8.0	TraesCS4A02G447000	Glycosyl-transferase
Alveo W	NOS_WW_SNP_11809	1B	1.77×10^{-13}	64	14.58 ± 1.66	8.8	TraesCS1B02G329600	Uncharacterized protein
	NOS_WW_SNP_6663	1B	1.21×10^{-9}	64	12.04 ± 1.67	6.0	TraesCS1B02G016400	Uncharacterized protein
	NOS_WW_SNP_3056	1D	6.30×10^{-13}	13	20.18 ± 2.35	8.3	TraesCS1D02G317100	Histone deacetylase
	NOS_WW_SNP_688	5D	4.16×10^{-6}	28	9.41 ± 1.74	3.2	TraesCS5D02G004300	Nontranslating coding sequence

¹ Only the most significant SNP for each peak in the Manhattan plots is shown. ² Allele frequencies are for the allele that is positively associated with the trait.

Table 4. Mean values of Alveo P for lines with each combination of alleles of the significant SNPs. Alleles that are positively associated with the trait are marked with green.

NOS_WW_ SNP_5054 (Chr. 1D)	NOS_WW_ SNP_688 (Chr. 5D)	Alveo P (mm)	Number of Lines
G	C	64.0	386
G	T	84.2	125
A	C	100.7	43
A	T	104.7	36

Table 5. Mean values of Alveo W for lines with each combination of alleles of the significant SNPs. Alleles that are positively associated with the trait are marked with green.

NOS_WW_ SNP_3056 (Chr. 1D)	NOS_WW_ SNP_11809 (Chr. 1B)	NOS_WW_ SNP_6663 (Chr. 1B)	NOS_WW_ SNP_688 (Chr. 5D)	Alveo W (10 ⁻⁴ J)	Number of Lines
C	G	G	C	88.9	75
C	G	G	T	83.8	11
C	G	A	C	127.4	86
C	A	G	T	132.9	26
C	A	A	C	137.0	144
C	A	A	T	164.7	74
T	G	G	C	131.8	5
T	G	G	T	127.0	3
T	G	A	C	160.2	5
T	G	A	T	167.3	3
T	A	G	C	145.5	6
T	A	G	T	170.5	2
T	A	A	C	204.7	14
T	A	A	T	214.5	26

GWAS were also performed using Bayesian Power Lasso models to fit all 10,802 SNPs simultaneously [50]. The optimal value for β was 0.4 for flour yield and for Alveo W, 0.6 for Alveo P, and 0.8 for Alveo L, respectively. The SNPs most significantly associated with flour yield, Alveo P, and Alveo W according to the single marker regressions were also the SNPs with the highest genetic effects according to the Bayesian Power Lasso models (Figure 4). For Alveo L, each SNP had very low genetic effect (less than 0.2).

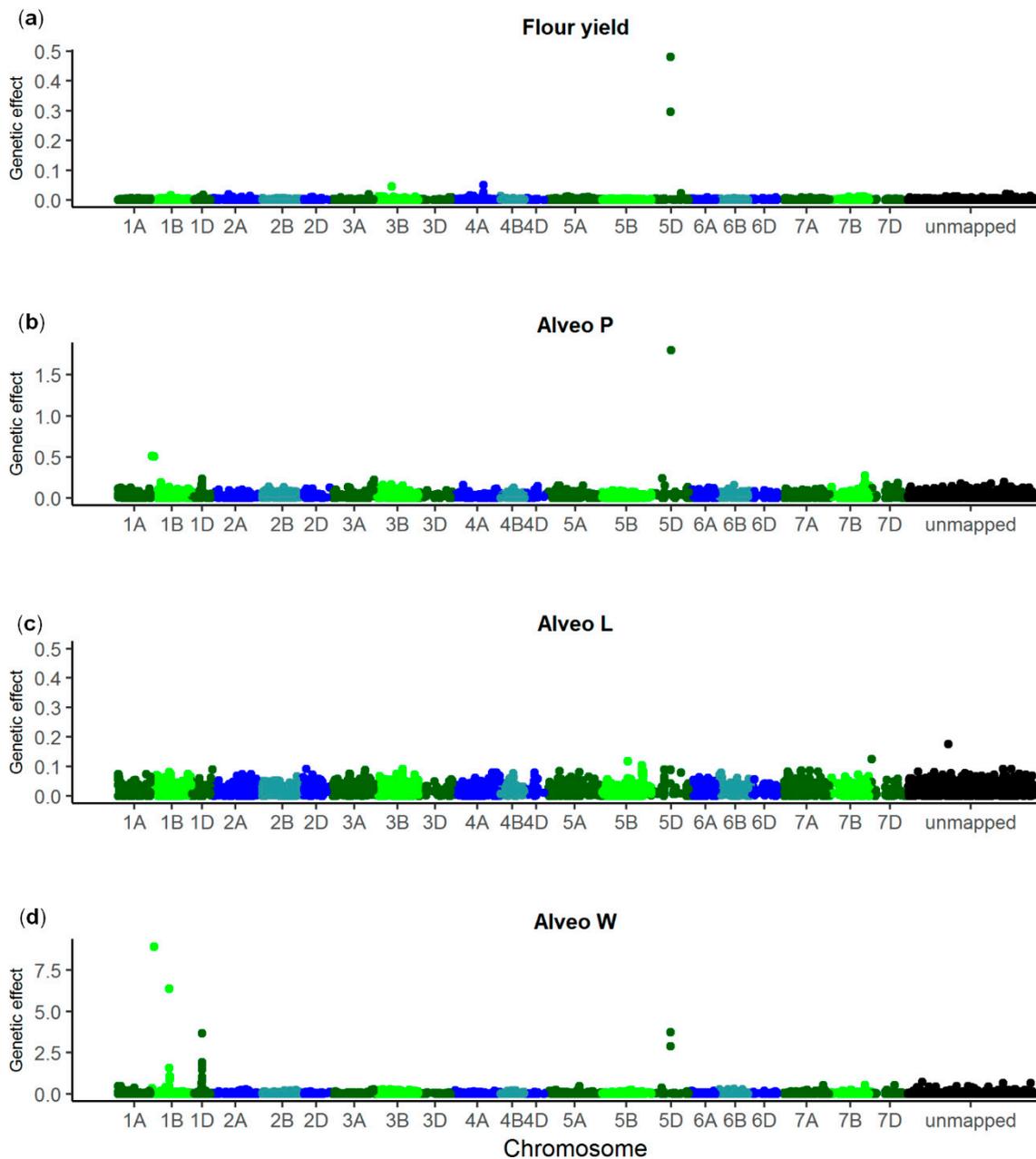


Figure 4. Genetic effect of SNPs estimated by the Bayesian Power Lasso models. (a): Flour yield, (b): Alveo P, (c): Alveo L, (d): Alveo W. Total number of SNPs: 10,802. Last bin is unmapped SNPs.

3.3. Genomic Predictions

Genomic predictions based on all 10,802 SNPs using a GBLUP model were evaluated using different CV strategies. Predictive abilities of the models were determined as the correlations between observed phenotypes corrected for fixed effects and the GEBVs. The predictive abilities were intermediate to high, ranging from 0.50 for flour yield to 0.79 for Alveo W based on the LOO CV (Figure 5). The LFO and the LSO CVs resulted in lower predictive abilities. The lowest predictive ability was 0.3 for flour yield based on the LSO CV. The predictive abilities of the k-fold CVs increased slightly when using a higher number of folds, and they were very close to the LOO when using 10 folds. The predictions were unbiased for the LOO and the k-fold CV and only slightly biased for the LFO and the LSO (Table 6).

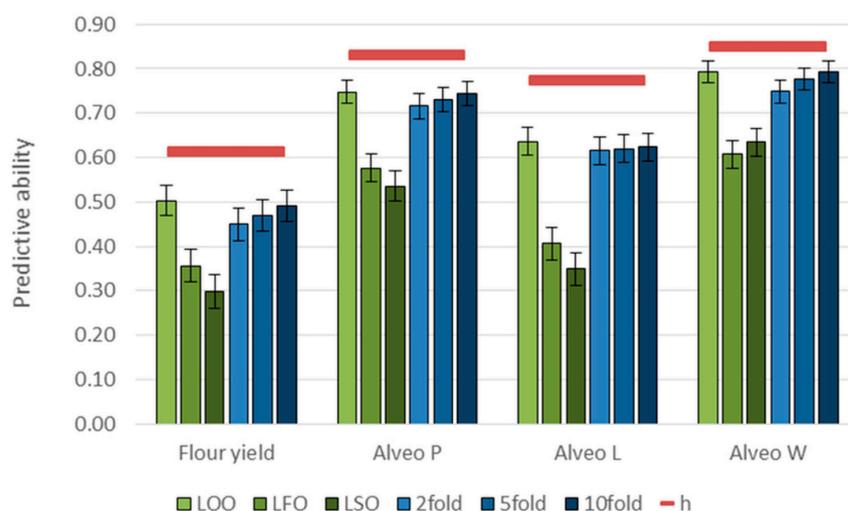


Figure 5. Predictive abilities for each quality trait based on GBLUP models. Red lines are the square root of narrow-sense genomic heritabilities (h) for each trait. Cross-validation (CV) strategies: LOO= leave-one-out, LFO= leave-family-out, LSO= leave-set-out, and k-fold= 2-, 5-, and 10-fold.

Table 6. Regressions of corrected phenotypes on genomic estimated breeding values (GEBVs) based on different cross-validation strategies using the GBLUP models. Bias is the deviation from the expected regression of 1.0.

Trait	LOO ¹	LFO	LSO	2-Fold	5-Fold	10-Fold
Flour yield	1.00 ± 0.07	0.89 ± 0.09	0.71 ± 0.09	0.95 ± 0.08	0.97 ± 0.07	0.99 ± 0.07
Alveo P	1.00 ± 0.04	0.98 ± 0.06	0.90 ± 0.06	1.02 ± 0.04	0.99 ± 0.04	1.00 ± 0.04
Alveo L	0.99 ± 0.05	0.94 ± 0.08	0.81 ± 0.09	1.03 ± 0.05	0.97 ± 0.05	0.98 ± 0.05
Alveo W	1.02 ± 0.03	0.95 ± 0.05	0.96 ± 0.05	1.03 ± 0.04	1.02 ± 0.03	1.03 ± 0.03

¹ Cross-validation strategies: leave-one-out (LOO), leave-family-out (LFO), leave-set-out (LSO), and k-fold: 2-, 5-, and 10-fold.

For every CV strategy, the predictive abilities of the Bayesian Power Lasso model were a little better compared to the GBLUP for flour yield, Alveo P, and Alveo W, but not for Alveo L (Figure 6).

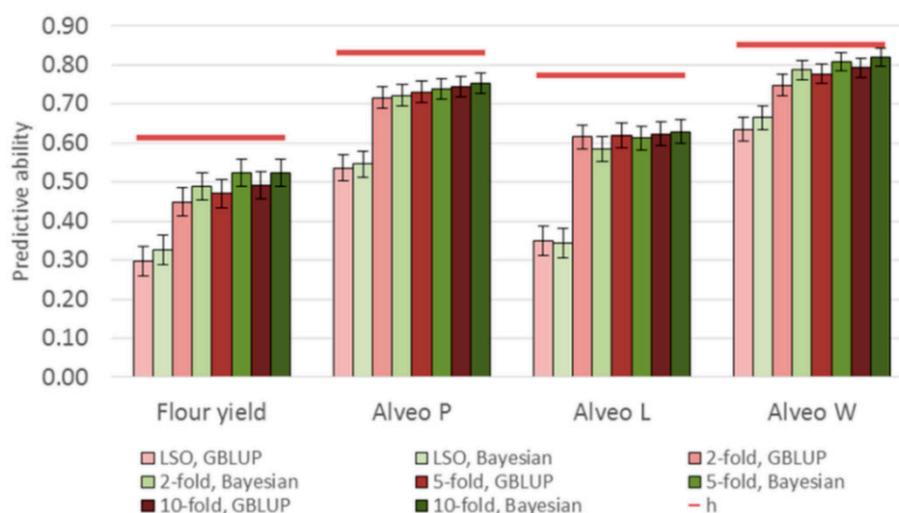


Figure 6. Comparison of predictive abilities based on GBLUP and Bayesian Power Lasso models. Red lines are the square root of narrow-sense genomic heritabilities (h) for each trait based on the GBLUP models.

The effects of training population sizes ranging from 10% to 90% of the 635 lines were studied (Figure 7). Predictive abilities increased from 0.33 at 10% (64 lines) to 0.50 at 90% (572 lines) for flour yield. For all traits, the predictive abilities increased when increasing the size of the training set, but the increases were smaller at larger sizes. Additionally, the variation around the mean of the predictive abilities decreased when increasing the size of the training set.

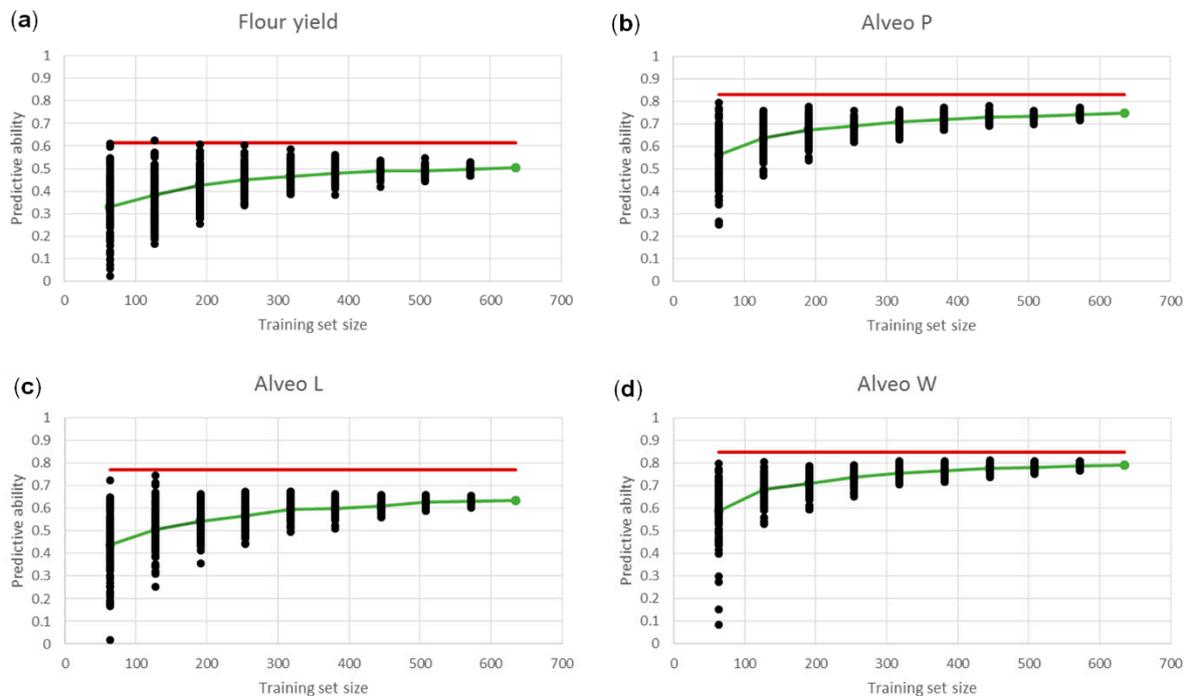


Figure 7. Predictive abilities based on leave-one-out cross-validations using reduced training set sizes. For each training set size, lines were selected randomly 100 times, and the average predictive abilities are shown with the green lines. The red lines are the square root of the narrow-sense genomic heritabilities based on all lines. (a): Flour yield, (b): Alveo P, (c): Alveo L, (d): Alveo W.

Genomic features models with two G-matrices were also tested (Figure 8). One G-matrix was calculated from significant SNPs, and the other G-matrix was calculated from nonsignificant SNPs. The number of SNPs to include as significant ranged from five to all 10,802 SNPs. For flour yield, Alveo P, and Alveo W, predictive abilities were highest (0.52, 0.73, and 0.79, respectively) when fewer than 1000 SNPs were considered significant (10, 500, and 100 SNPs, respectively). For Alveo L, predictive abilities were highest (0.61) when 3000 SNPs were considered significant. For all traits, predictive abilities were higher when using the optimal number of significant SNPs for one G-matrix and the remaining SNPs for another G-matrix compared to using all SNPs for one G-matrix: 0.52 vs. 0.47 for flour yield, 0.73 vs. 0.70 for Alveo P, 0.61 vs. 0.58 for Alveo L, and 0.79 vs. 0.76 for Alveo W.

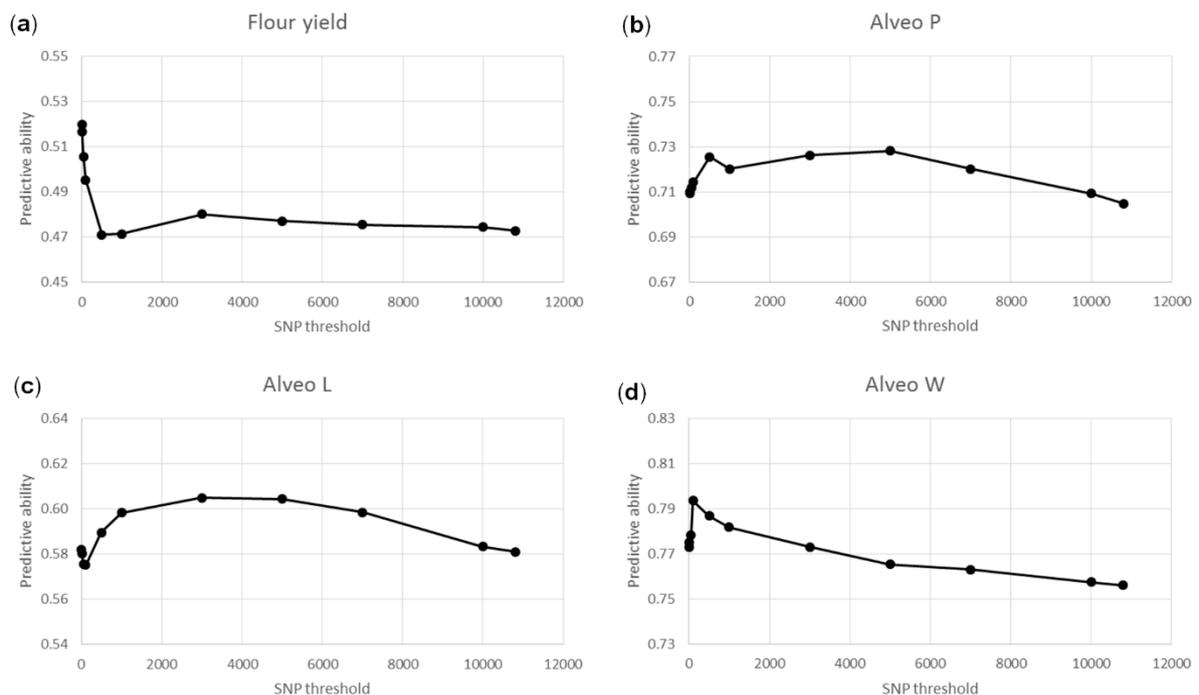


Figure 8. Predictive abilities based on genomic feature models with two G-matrices. The two G-matrices were computed from significant and nonsignificant SNPs, respectively. Half of the lines were used for GWAS to select significant SNPs, and the other half were used for LOO cross-validations. The threshold for number of significant SNPs was from five to 10,802. (a): Flour yield, (b): Alveo P, (c): Alveo L, (d): Alveo W.

4. Discussion

Wheat quality traits typically have intermediate or high heritabilities, although the traits can be considerably affected by environmental effects and genotype–environment (G×E) interactions [36–38]. Thus, additive genetic variation across environments also affects the traits. Here, narrow-sense genomic heritabilities ranged from 0.38 for flour yield to 0.72 for Alveo W (Table 2). Therefore, breeding for improved wheat quality traits should be possible.

Advanced breeding material was used in the present study. The breeding program has, until now, focused more on increasing yield rather than improving baking quality due to restrictions in application of nitrogen fertilization to the fields in Denmark, which have made it challenging to grow high quality bread wheat lines. Nevertheless, variation was observed for the quality traits, indicating that both high and low quality wheat lines were used as crossing parents for the studied lines, and that genetic variation was maintained throughout the breeding program (Figure 1, Table 2). Six of the 96 crossing parents were used in both breeding sets. The dendrogram of the wheat lines indicated genetic relationships within each of the two sets (Figure 2). However, the lines from each set were not clearly divided in the dendrogram, indicating that the lines were also related between the two breeding sets.

QTL for wheat quality traits were identified across the genome [1,2,17]. However, many QTL were only identified in certain environments or populations [55,56]. In the present study, two closely linked SNPs on chromosome 5D were significantly associated with flour yield, Alveo P, and Alveo W (Figure 3). The puroindoline genes *Pina-D1* and *Pinb-D1* have a large effect on grain hardness. These genes are located on chromosome 5D and can affect several quality traits [1,12,57]. Several *Pinb-D1* alleles with a positive effect on wheat quality were identified [58,59]. However, relatively few of the markers in the present study were located on chromosome 5D [37], thus additional markers or sequencing would be needed to distinguish between each of the alleles.

For Alveo P and Alveo W, significant SNPs were also identified on chromosome 1D. The HMWG loci *Glu-D1* is located on chromosome 1D and has a large influence on wheat baking quality [4].

Similarly, the glutenin loci *Glu-B1* and *Glu-B3*, which are located on chromosome 1B, can affect quality. No SNPs on chromosome 1A were significantly associated with the Alveograph traits. The loci *Glu-A1* and *Glu-A3* are also known to affect wheat quality traits [4]. Thus, these loci might not be polymorphic in the studied material, their effects might be too low to detect, or the genotyped SNPs might not be located in or close enough to the loci. Several markers may be required in order to distinguish between the alleles of each of the *Glu* loci [60]. Therefore, the SNP chip array used to genotype the studied wheat lines can perhaps not capture all the genetic variants. Characterization of *Glu*- and *Pin* loci in the breeding material could be useful for more accurate estimation of their effects and consequently more accurate predictions of the baking quality traits [4,38,58].

The GWAS indicated that only few QTL with large or intermediate effects controlled the quality traits. The identified QTL only explained a relatively small proportion of the total genetic variance. This indicates that the traits were also controlled by many QTL with small effects. Identification of such minor QTL is challenging, especially if they have a low MAF or if they are located near major QTL. Lines with the positive alleles of the four largest QTL for Alveo W had considerably higher dough strength than lines with negative alleles of any of the QTL (Table 5). However, the four QTL together explained only 26.3% of the additive genetic variance based on the effects estimated from the single marker regressions, and these effects were possibly overestimated due to the Beavis effect [61,62]. The effects were lower when estimated using the Bayesian Power Lasso (Figure 4). Here, all SNP effects were estimated simultaneously, thus the effects were shrunk towards zero, and each QTL effect might have been distributed across several SNPs.

Since only a relatively small proportion of the genetic variance could be explained by the identified QTL, genomic predictions based on a large number of genome-wide markers could be useful. The different CV strategies showed that the predictive abilities were affected by size of the training set, by the genetic relationship between lines in training and validation sets, and by the G×E interactions (Figure 5). The LSO strategy represents one way of implementing genomic predictions in breeding programs. Predicting GEBVs of new lines based on lines from previous years could possibly enable selection before any phenotypic information is available for the new lines. However, the LSO strategy resulted in lower and more biased predictive abilities than the other CV strategies (Figure 5, Table 6). Possible reasons for the lower predictive abilities could be the size of the training set, the genetic relationship between lines, and the G×E interactions. Including lines from the same year and the same families in the training set improved the predictive abilities considerably. Reducing the size of training sets had a negative impact on the predictive abilities (Figure 7). However, the decrease for very small training sets seen in the present study was not as drastic as in other studies [29,32], possibly due to the high heritabilities of the traits included in this study. Thus, a few hundred lines might be enough for the training set, if they are highly related to the validation set. The G×E interactions and the genetic relationship between lines might be partly confounded, since the lines were only tested from one location. These effects had a larger impact on predictive abilities than the size of the training set and the model used for the predictions. G×E interactions can be accounted for in, for example, reaction norm models if phenotypic data are available for lines replicated across several locations or years. Additionally, data about climatic conditions and soil types might be included to obtain higher predictive abilities [63]. Such models could also be used for selecting lines for target environments or lines that are performing well over many environments. Predictive abilities can also be affected by the number of markers used for the predictions [29,32]. If the markers are selected based on GWAS, the number can possibly be reduced to a few hundred markers without affecting the predictive abilities. However, the markers that are selected would not be the same for each trait, and the markers could change after each new breeding cycle [29].

The predictive abilities of the Bayesian Power Lasso models were slightly higher compared to the GBLUP models for flour yield, Alveo P, and Alveo W (Figure 6). In the Bayesian Power Lasso model, the difference between small and large QTL effects can be bigger than in GBLUP [50]. Thus, the largest QTL effects might be shrunk too much in the GBLUP models. For Alveo L, no improvements were

observed when using the Bayesian models, indicating that no major QTL were present for this trait (Figure 6). Previous studies have also shown that different types of Bayesian models can, in some cases, give slightly more accurate predictions compared to GBLUP models, especially for traits influenced by major QTL and for populations with low genetic relationships between training and validation sets [26,37,50].

Other studies of genomic selection for wheat quality traits using breeding material have reported predictive abilities in similar ranges as in the present study [36,37,64]. Thus, genomic prediction appears to be a promising strategy for improving wheat quality in breeding programs. Nevertheless, the use of information from GWAS or from known major QTL in genomic predictions might be useful for modeling marker effects more accurately. Markers for major QTL can be specified as having fixed effects, while remaining markers have random effects. Bernardo (2014) [39] recommended to specify fixed effects for markers that explain more than 10% genetic variance for traits controlled by fewer than 10 major genes. Zhao et al. (2014) [26] used W-BLUP (weighted BLUP) to give few functional markers a larger weight than other markers. This improved prediction accuracies of heading time and plant height in hybrid wheat compared to marker-assisted selection (MAS), ridge regression BLUP (equivalent to GBLUP) and BayesC π . Improved prediction accuracies were also reported by Arruda et al. (2016) [34] when using fixed effects for QTL associated with Fusarium head blight resistance traits compared to ridge regression BLUP, and similarly for pre-harvest sprouting tolerance by Moore et al. (2017) [40]. However, including QTL identified using the same lines that were also used for the genomic predictions could lead to inflation of the accuracies [34]. In a recent study by Michel et al. (2018) [38], prediction accuracies for several wheat quality traits were higher based on ridge regression BLUP compared to specifying fixed effects for associated markers identified in independent populations. However, including one or more of the three *Glu-1* loci as fixed effects resulted in improved accuracies. For each trait, *Glu-1* loci were included if the locus explained more than 5% genetic variance [38].

In the present study, an alternative approach using one G-matrix computed from the most significant SNPs and another G-matrix computed from the remaining SNPs could slightly improve predictive abilities (Figure 8). However, the optimal significance threshold depended on the trait. For the traits controlled by few QTL with large effects, a conservative significance threshold seemed to be best, while a loose threshold seemed to be better for traits controlled only by QTL with small effects. To avoid inflation of the predictive abilities, half of the lines were used for the GWAS to select significant SNPs, and the other half were used for the genomic predictions. Thus, larger datasets might be necessary to study the predictive abilities of the genomic feature models and the optimal number of SNPs more thoroughly.

Implementation of genetic markers and genomic predictions in breeding programs could likely lead to increased genetic gains for the wheat quality traits. Resources needed for phenotyping could be reduced, and the selection intensity could be increased. The identified SNPs with large effects might be used for screening large numbers of lines early in the breeding program and selecting lines that potentially have good quality traits. Genomic predictions could be used for a more accurate selection of lines with good quality in later generations.

5. Conclusions

SNPs significantly associated with flour yield, Alveo P, and Alveo W were identified on chromosome 5D. For Alveos P and W, associated SNPs were also identified on chromosome 1D. Likely candidate genes could be the *Pina-D1* or the *Pinb-D1* on chromosome 5D and the *Glu-D1* loci on chromosome 1D. Furthermore, SNPs associated with Alveo W were identified on chromosome 1B, and SNPs associated with Alveo L were identified on chromosome 4A. Additive genetic variance explained by a single SNP was up to 13.3% (SNP on 5D for flour yield). The identified SNPs can be used in early generations of breeding programs to screen large numbers of lines. In later generations, it would be advantageous to use a large number of SNPs to ensure accurate prediction of breeding values. Predictive abilities of GBLUP models were 0.50 for flour yield, 0.75 for Alveo P, 0.79 for Alveo W, and

0.64 for Alveo L based on the LOO CV. Predictive abilities were lower when using smaller training sets but were still moderate when using only 10% of the 635 lines as the training set. Furthermore, predictive abilities were significantly lower when using LSO and LFO CV strategies because of reduced genetic relationship between lines in training and validation sets and because of G×E interactions. Predictive abilities were similar or slightly higher based on Bayesian Power Lasso and genomic feature models. Thus, GBLUP models could be used for genomic prediction of wheat quality traits with moderate to high predictive ability. Other models might give slightly higher predictive abilities for traits where major QTL are present in the breeding material.

Supplementary Materials: The following are available online at <http://www.mdpi.com/2073-4425/10/9/669/s1>, Table S1: Phenotypic data, Table S2: Genotypic data.

Author Contributions: Conceptualization, P.S.K., J.J., J.R.A. and A.J.; Data curation, P.S.K. and J.O.; Formal analysis, P.S.K.; Funding acquisition, P.S.K., J.J., J.R.A. and A.J.; Investigation, P.S.K., J.R.A. and C.G.; Methodology, P.S.K., J.J., J.R.A., C.G., J.O. and A.J.; Project administration, J.J. and A.J.; Resources, P.S.K., J.R.A., C.G. and J.O.; Software, P.S.K. and J.J.; Supervision, J.J., J.R.A. and A.J.; Validation, P.S.K. and J.R.A.; Visualization, P.S.K.; Writing—original draft, P.S.K.; Writing—review & editing, P.S.K., J.J., J.R.A., C.G., J.O. and A.J.

Funding: This research was funded by Innovation Fund Denmark, grant number 5139-00018B, Erstatningsfonden for Sædekorn, and Ministry of Environment and Food of Denmark under the Program for Green Development and Demonstrations, grant number 34009-13-0607.

Acknowledgments: We would like to thank Hanne Svenstrup from Nordic Seed A/S, Denmark, and the Wheat Chemistry and Quality Laboratory at CIMMYT, Mexico, for supporting laboratory work and phenotyping. Carlos Guzman gratefully acknowledges the European Social Fund and the Spanish Ministry of Science, Innovation and Universities for financial funding through the Ramon y Cajal Program (RYC-2017-21891).

Conflicts of Interest: This research was performed in a collaboration between Aarhus University and the plant breeding company Nordic Seed A/S. Authors P.S.K., J.R.A., J.O. and A.J. were employed by the company Nordic Seed A/S. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

- Bordes, J.; Ravel, C.; Le Gouis, J.; Lapierre, A.; Charmet, G.; Balfourier, F. Use of a global wheat core collection for association analysis of flour and dough quality traits. *J. Cereal Sci.* **2011**, *54*, 137–147. [[CrossRef](#)]
- Smith, N.; Guttieri, M.; Souza, E.; Shoots, J.; Sorrells, M.; Sneller, C. Identification and validation of QTL for grain quality traits in a cross of soft wheat cultivars pioneer brand 25R26 and Foster. *Crop. Sci.* **2011**, *51*, 1424–1436. [[CrossRef](#)]
- He, Z.H.; Liu, L.; Xia, X.C.; Liu, J.J.; Pena, R.J. Composition of HMW and LMW Glutenin Subunits and Their Effects on Dough Properties, Pan Bread, and Noodle Quality of Chinese Bread Wheats. *Cereal Chem.* **2005**, *82*, 345–350. [[CrossRef](#)]
- Payne, P.I.; Nightingale, M.A.; Krattiger, A.F.; Holt, L.M. The relationship between HMW glutenin subunit composition and the bread making quality of British grown wheat varieties. *J. Sci. Food Agric.* **1987**, *40*, 51–65. [[CrossRef](#)]
- Bhave, M.; Morris, C.F. Molecular genetics of puroindolines and related genes: Allelic diversity in wheat and other grasses. *Plant. Mol. Biol.* **2008**, *66*, 205–219. [[CrossRef](#)] [[PubMed](#)]
- Giroux, M.J.; Morris, C.F. A glycine to serine change in puroindoline b is associated with wheat grain hardness and low levels of starch-surface friabilin. *Theor. Appl. Genet.* **1997**, *95*, 857–864. [[CrossRef](#)]
- Graybosch, R.A.; Peterson, C.J.; Hareland, G.A.; Shelton, D.R.; Olewnik, M.C.; He, H.; Stearns, M.M. Relationships between small-scale wheat quality assays and commercial test bakes. *Cereal Chem.* **1999**, *76*, 428–433. [[CrossRef](#)]
- Guzmán, C.; Posadas-romano, G.; Hernández-Espinosa, N.; Morales-Dorantes, A.; Peña, R.J. A new standard water absorption criteria based on solvent retention capacity (SRC) to determine dough mixing properties, viscoelasticity, and bread-making quality. *J. Cereal Sci. J.* **2015**, *66*, 59–65. [[CrossRef](#)]
- Guzman, C.; Peña, R.J.; Singh, R.; Autrique, E.; Dreisigacker, S.; Crossa, J.; Rutkoski, J.; Poland, J.; Battenfield, S. Wheat quality improvement at CIMMYT and the use of genomic selection on it. *Appl. Transl. Genomics.* **2016**, *11*, 3–8. [[CrossRef](#)]

10. Groos, C.; Bervas, E.; Charmet, G. Genetic analysis of grain protein content, grain hardness and dough rheology in a hard X hard bread wheat progeny. *J. Cereal Sci.* **2004**, *40*, 93–100. [[CrossRef](#)]
11. Kerfal, S.; Giraldo, P.; Rodríguez-Quijano, M.; Vázquez, J.F.; Adams, K.; Lukow, O.M.; Röder, M.S.; Somers, D.J.; Carrillo, J.M. Mapping quantitative trait loci (QTLs) associated with dough quality in a soft×hard bread wheat progeny. *J. Cereal Sci.* **2010**, *52*, 46–52. [[CrossRef](#)]
12. Nelson, J.C.; Andreescu, C.; Breseghello, F.; Finney, P.L.; Gualberto, D.G.; Bergman, C.J.; Peña, R.J.; Perretant, M.R.; Leroy, P.; Qualset, C.O.; et al. Quantitative trait locus analysis of wheat quality traits. *Euphytica* **2006**, *149*, 145–159. [[CrossRef](#)]
13. Zanetti, S.; Winzeler, M.; Feuillet, C.; Keller, B.; Messmer, M. Genetic analysis of bread-making quality in wheat and spelt. *Plant. Breed.* **2001**, *120*, 13–19. [[CrossRef](#)]
14. Tadesse, W.; Ogonnaya, F.C.; Jighly, A.; Sohail, Q.; Rajaram, S. Genome-Wide Association Mapping of Yield and Grain Quality Traits in Winter Wheat Genotypes. *PLoS ONE* **2015**, *10*, e0141339. [[CrossRef](#)] [[PubMed](#)]
15. Cabrera, A.; Guttieri, M.; Smith, N.; Souza, E.; Sturbaum, A.; Hua, D.; Griffey, C.; Barnett, M.; Murphy, P.; Ohm, H.; et al. Identification of milling and baking quality QTL in multiple soft wheat mapping populations. *Theor. Appl. Genet.* **2015**, *128*, 2227–2242. [[CrossRef](#)] [[PubMed](#)]
16. Fox, G.P.; Martin, A.; Kelly, A.M.; Sutherland, M.W.; Martin, D.; Banks, P.M.; Sheppard, J. QTLs for water absorption and flour yield identified in the doubled haploid wheat population Lang/QT8766. *Euphytica* **2013**, *192*, 453–462. [[CrossRef](#)]
17. Ishikawa, G.; Nakamura, K.; Ito, H.; Saito, M.; Sato, M.; Jinno, H.; Yoshimura, Y.; Nishimura, T.; Maejima, H.; Uehara, Y.; et al. Association mapping and validation of QTLs for flour yield in the soft winter wheat variety Kitahonami. *PLoS ONE* **2014**, *9*, e111337. [[CrossRef](#)] [[PubMed](#)]
18. McCartney, C.A.; Somers, D.J.; Lukow, O.; Ames, N.; Noll, J.; Cloutier, S.; Humphreys, D.G.; McCallum, B.D. QTL analysis of quality traits in the spring wheat cross RL4452 × “AC domain”. *Plant. Breed.* **2006**, *125*, 565–575. [[CrossRef](#)]
19. Sherman, J.D.; Nash, D.; Lanning, S.P.; Martin, J.M.; Blake, N.K.; Morris, C.F.; Talbert, L.E. Genetics of end-use quality differences between a modern and historical spring wheat. *Crop. Sci.* **2014**, *54*, 1972–1980. [[CrossRef](#)]
20. Tsilo, T.J.; Hareland, G.A.; Chao, S.; Anderson, J.A. Genetic mapping and QTL analysis of flour color and milling yield related traits using recombinant inbred lines in hard red spring wheat. *Crop. Sci.* **2011**, *51*, 237–246. [[CrossRef](#)]
21. Bagningsanalyser af sorter af Vinterhvede ved TystofteFonden. Available online: <https://www.tystofte.dk/wp-content/uploads/2018/06/broedhvedekriterier-til-www.pdf> (accessed on 10 October 2018).
22. Brødhvede. Available online: <https://lbst.dk/landbrug/goedning/kvaelstofregulering/broedhvede/#c51695> (accessed on 10 October 2018).
23. Bernardo, R. Molecular markers and selection for complex traits in plants: Learning from the last 20 years. *Crop. Sci.* **2008**, *48*, 1649–1664. [[CrossRef](#)]
24. Meuwissen, T.H.E.; Hayes, B.J.; Goddard, M.E. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **2001**, *157*, 1819–1829. [[PubMed](#)]
25. Heffner, E.L.; Jannink, J.L.; Iwata, H.; Souza, E.; Sorrells, M.E. Genomic selection accuracy for grain quality traits in biparental wheat populations. *Crop. Sci.* **2011**, *51*, 2597–2606. [[CrossRef](#)]
26. Zhao, Y.; Mette, M.F.; Gowda, M.; Longin, C.F.H.; Reif, J.C. Bridging the gap between marker-assisted and genomic selection of heading time and plant height in hybrid wheat. *Heredity (Edinb)* **2014**, *112*, 638–645. [[CrossRef](#)] [[PubMed](#)]
27. VanRaden, P.M. Efficient methods to compute genomic predictions. *J. Dairy Sci.* **2008**, *91*, 4414–4423. [[CrossRef](#)]
28. Meuwissen, T.H.E.; Hayes, B.; Goddard, M. Genomic selection: A paradigm shift in animal breeding. *Anim. Front.* **2016**, *6*, 6. [[CrossRef](#)]
29. Cericola, F.; Jahoor, A.; Orabi, J.; Andersen, J.R.; Janss, L.L.; Jensen, J. Optimizing Training Population Size and Genotyping Strategy for Genomic Prediction Using Association Study Results and Pedigree Information. A Case of Study in Advanced Wheat Breeding Lines. *PLoS ONE* **2017**, *12*, e0169606. [[CrossRef](#)] [[PubMed](#)]
30. Crossa, J.; Pérez-Rodríguez, P.; Cuevas, J.; Montesinos-López, O.; Jarquín, D.; de los Campos, G.; Burgueño, J.; González-Camacho, J.M.; Pérez-Elizalde, S.; Beyene, Y.; et al. Genomic Selection in Plant Breeding: Methods, Models, and Perspectives. *Trends Plant. Sci.* **2017**, *22*, 961–975. [[CrossRef](#)]

31. Marulanda, J.J.; Mi, X.; Melchinger, A.E.; Xu, J.L.; Würschum, T.; Longin, C.F.H. Optimum breeding strategies using genomic selection for hybrid breeding in wheat, maize, rye, barley, rice and triticale. *Theor. Appl. Genet.* **2016**, *129*, 1901–1913. [[CrossRef](#)]
32. Nielsen, N.H.; Jahoor, A.; Jensen, J.D.; Orabi, J.; Cericola, F.; Edriss, V.; Jensen, J. Genomic prediction of seed quality traits using advanced barley breeding lines. *PLoS ONE* **2016**, *11*, e0164494. [[CrossRef](#)]
33. Heffner, E.L.; Jannink, J.; Sorrells, M.E. Genomic Selection Accuracy using Multifamily Prediction Models in a Wheat Breeding Program. *Plant Genome* **2011**, *4*, 65–75. [[CrossRef](#)]
34. Arruda, M.P.; Lipka, A.E.; Brown, P.J.; Krill, A.M.; Thurber, C.; Brown-Guedira, G.; Dong, Y.; Foresman, B.J.; Kolb, F.L. Comparing genomic selection and marker-assisted selection for Fusarium head blight resistance in wheat (*Triticum aestivum* L.). *Mol. Breed.* **2016**, *36*, 1–11. [[CrossRef](#)]
35. Charmet, G.; Storlie, E.; Oury, F.X.; Laurent, V.; Beghin, D.; Chevarin, L.; Lapierre, A.; Perretant, M.R.; Rolland, B.; Heumez, E.; et al. Genome-wide prediction of three important traits in bread wheat. *Mol. Breed.* **2014**, *34*, 1843–1852. [[CrossRef](#)] [[PubMed](#)]
36. Battenfield, S.D.; Guzmán, C.; Gaynor, R.C.; Singh, R.P.; Peña, R.J.; Dreisigacker, S.; Fritz, A.K.; Poland, J.A. Genomic Selection for Processing and End-Use Quality Traits in the CIMMYT Spring Bread Wheat Breeding Program. *Plant Genome* **2016**, *9*. [[CrossRef](#)] [[PubMed](#)]
37. Kristensen, P.S.; Jahoor, A.; Andersen, J.R.; Cericola, F.; Orabi, J.; Janss, L.; Jensen, J. Genome-Wide Association Studies and Comparison of Models and Cross-Validation Strategies for Genomic Prediction of Quality Traits in Advanced Winter Wheat Breeding Lines. *Front. Plant. Sci.* **2018**, *9*, 69. [[CrossRef](#)] [[PubMed](#)]
38. Michel, S.; Kummer, C.; Gallee, M.; Hellinger, J.; Ametz, C.; Akgöl, B.; Epure, D.; Löschenberger, F.; Buerstmayr, H. Improving the baking quality of bread wheat by genomic selection in early generations. *Theor. Appl. Genet.* **2018**, *131*, 477–493. [[CrossRef](#)] [[PubMed](#)]
39. Bernardo, R. Genomewide selection when major genes are known. *Crop. Sci.* **2014**, *54*, 68–75. [[CrossRef](#)]
40. Moore, J.K.; Manmathan, H.K.; Anderson, V.A.; Poland, J.A.; Morris, C.F.; Haley, S.D. Improving genomic prediction for pre-harvest sprouting tolerance in wheat by weighting large-effect quantitative trait loci. *Crop. Sci.* **2017**, *57*, 1315–1324. [[CrossRef](#)]
41. Norman, A.; Taylor, J.; Edwards, J.; Kuchel, H. Optimising Genomic Selection in Wheat: Effect of Marker Density, Population Size and Population Structure on Prediction Accuracy. *G3-Genes Genomes Genetics* **2018**, *8*, 2889–2899. [[CrossRef](#)]
42. Thorwarth, P.; Liu, G.; Ebmeyer, E.; Schacht, J.; Schachschneider, R.; Kazman, E.; Reif, J.C.; Würschum, T.; Longin, C.F.H. Dissecting the genetics underlying the relationship between protein content and grain yield in a large hybrid wheat population. *Theor. Appl. Genet.* **2019**, *132*, 489–500. [[CrossRef](#)]
43. American Association of Cereal Chemists. *Approved Methods of Analysis*, 11th ed.; AACCI International: St. Paul, MN, USA, 2000.
44. Rogers, S.O.; Bendich, A.J. Extraction of DNA from milligram amounts of fresh, herbarium and mummified plant tissues. *Plant. Mol. Biol.* **1985**, *5*, 69–76. [[CrossRef](#)] [[PubMed](#)]
45. Madsen, P.; Jensen, J. DMU: A User's Guide. A Package for Analysing Multivariate Mixed Models. 2013. Available online: <http://dmu.agrsci.dk> (accessed on 8 August 2019).
46. Hinrichs, A.L.; Larkin, E.K.; Suarez, B.K. Population stratification and patterns of linkage disequilibrium. *Genet. Epidemiol.* **2009**, *33*, S88–S92. [[CrossRef](#)] [[PubMed](#)]
47. IWGSC (The International Wheat Genome Sequencing Consortium). Shifting the limits in wheat research and breeding using a fully annotated reference genome. *Science* **2018**, *361*, 7191. [[CrossRef](#)] [[PubMed](#)]
48. Kersey, P.J.; Allen, J.E.; Allot, A.; Barba, M.; Boddu, S.; Bolt, B.J.; Carvalho-Silva, D.; Christensen, M.; Davis, P.; Grabmueller, C.; et al. Ensembl Genomes 2018: An integrated omics infrastructure for non-vertebrate species. *Nucleic Acids Res.* **2018**, *46*, D802–D808. [[CrossRef](#)] [[PubMed](#)]
49. Janss, L.L. Bayz Manual. Available online: <http://bayz.biz> (accessed on 8 August 2019).
50. Gao, H.; Su, G.; Janss, L.; Zhang, Y.; Lund, M.S. Model comparison on genomic predictions using high-density markers for different groups of bulls in the Nordic Holstein population. *J. Dairy Sci.* **2013**, *96*, 4678–4687. [[CrossRef](#)] [[PubMed](#)]
51. Spiegelhalter, D.J.; Best, N.G.; Carlin, B.P.; Van Der Linde, A. Bayesian measures of model complexity and fit. *J. R. Statist. Soc. B* **2002**, *64*, 583–616. [[CrossRef](#)]

52. Plummer, M.; Best, N.; Cowles, K.; Vines, K. CODA: Convergence Diagnosis and Output Analysis for MCMC. *R News*. **2006**, *6*, 7–11. Available online: https://www.r-project.org/doc/Rnews/Rnews_2006-1.pdf (accessed on 8 August 2019).
53. Rohde, P.D.; Gaertner, B.; Ward, K.; Sørensen, P.; Mackay, T.F.C. Genomic analysis of genotype by social environment interaction for *Drosophila* aggressive behavior. *Genetics* **2017**, *206*, 1969–1984. [[CrossRef](#)]
54. Sarup, P.; Jensen, J.; Ostersen, T.; Henryon, M.; Sørensen, P. Increased prediction accuracy using a genomic feature model including prior information on quantitative trait locus regions in purebred Danish Duroc pigs. *BMC Genet.* **2016**, *17*, 1–17. [[CrossRef](#)]
55. Deng, Z.; Tian, J.; Chen, F.; Li, W.; Zheng, F.; Chen, J.; Shi, C.; Sun, C.; Wang, S.; Zhang, Y. Genetic dissection on wheat flour quality traits in two related populations. *Euphytica* **2015**, *203*, 221–235. [[CrossRef](#)]
56. Jin, H.; Wen, W.; Liu, J.; Zhai, S.; Zhang, Y.; Yan, J.; Liu, Z.; Xia, X.; He, Z. Genome-Wide QTL Mapping for Wheat Processing Quality Parameters in a Gaocheng 8901/Zhoumai 16 Recombinant Inbred Line Population. *Front. Plant. Sci.* **2016**, *7*. [[CrossRef](#)] [[PubMed](#)]
57. Maphosa, L.; Langridge, P.; Taylor, H.; Emebiri, L.C.; Mather, D.E. Genetic control of grain protein, dough rheology traits and loaf traits in a bread wheat population grown in three environments. *J. Cereal Sci.* **2015**, *64*, 147–152. [[CrossRef](#)]
58. Mohler, V.; Schmolke, M.; Paladey, E.; Seling, S.; Hartl, L. Association analysis of Puroindoline-D1 and Puroindoline b-2 loci with 13 quality traits in European winter wheat (*Triticum aestivum* L.). *J. Cereal Sci.* **2012**, *56*, 623–628. [[CrossRef](#)]
59. Morris, C.F. Puroindolines: The molecular basis of wheat grain hardness. *Plant Mol Biol Puroindolines: The molecular genetic basis of wheat grain hardness.* *Plant. Mol. Biol.* **2002**, *48*, 633–647. [[CrossRef](#)] [[PubMed](#)]
60. Vagndorf, N.; Kristensen, P.S.; Andersen, J.R.; Jahoor, A.; Orabi, J. Marker-assisted breeding in eheat. In *Next Generation Plant Breeding*; IntechOpen: London, UK, 2018; pp. 3–22. [[CrossRef](#)]
61. Beavis, W.D. QTL analyses: power, precision, and accuracy. In *Molecular Dissection of Complex Traits*; Paterson, A.H., Ed.; CRC Press: New York, NY, USA, 1998; pp. 145–162.
62. Xu, S. Theoretical Basis of the Beavis Effect. *Genetics* **2003**, *165*, 2259–2268. [[PubMed](#)]
63. Jarquín, D.; Crossa, J.; Lacaze, X.; Du Cheyron, P.; Daucourt, J.; Lorgeou, J.; Piroux, F.; Guerreiro, L.; Pérez, P.; Calus, M.; et al. A reaction norm model for genomic selection using high-dimensional genomic and environmental data. *Theor. Appl. Genet.* **2014**, *127*, 595–607. [[CrossRef](#)]
64. Michel, S.; Ametz, C.; Gungor, H.; Epure, D.; Grausgruber, H.; Löschenberger, F.; Buerstmayr, H. Genomic selection across multiple breeding cycles in applied bread wheat breeding. *Theor. Appl. Genet.* **2016**, *129*, 1179–1189. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).