Local genomic estimates provide a powerful framework for haplotype discovery

Will Shaffer[1], Victor Papin [1], Seema Yadav [1], Kai P. Voss-Fels [1,2], Lee T. Hickey [1], Ben J. Hayes [1], Eric G. Dinglasan[1*]

[1]Queensland Alliance for Agriculture and Food Innovation, The University of Queensland, Brisbane, Queensland, 4072, Australia

[2]Institute of Plant Breeding, Hochschule Geisenheim University, 65366 Geisenheim, Germany

[*]Correspondence: Eric Dinglasan (e.dinglasan@uq.edu.au)

Running title: Local genomic estimates for haplotype discovery

Abstract

Quantitative trait loci (QTL) discovery studies on diversity panels or breeding populations typically use genome-wide association studies (GWAS) to estimate marker effects. For plant and animal breeding applications, researchers increasingly recognize the potential benefits of identifying superior haplotypes (markers in linkage disequilibrium; LD) rather than relying on single markers, as traditional approaches inefficiently account for cumulative signals from incomplete LD with QTL or split effects when multiple markers are in high LD with QTL. Using the genomic prediction framework, the local GEBV (localGEBV) method was developed in animal breeding and has been adopted in crop haplotype mapping studies; however, no study has thoroughly quantified the utility of this method or systematically compared outcomes to traditional GWAS approaches. Here, we characterized a strategy to group markers in chromosomal segments based on LD (haplotype blocks or haploblocks), computed localGEBV as a linear contrast of marker effects within each haploblock, and utilised the variance of localGEBV to enhance QTL discovery compared to traditional GWAS. Marker effects for localGEBV were estimated with ridge-regression best linear unbiased prediction (rrBLUP) and BayesR, with results compared to two common GWAS approaches. Using the barley row-type trait, we demonstrated that localGEBV improved QTL discovery and phenotypic prediction compared to single markers. Furthermore, localGEBV results were robust to the choice of prior marker assumptions and blocking parameters, enabling flexibility in fine or broad-scale QTL mapping. Overall, our findings establish localGEBV as a haplotype-based strategy capable of leveraging localized genomic effects to improve QTL discovery and, potentially, genomic selection.

Key words: local GEBV, linkage disequilibrium, GWAS, genomic selection, haplotype, haploblock

Introduction

Understanding the genetic basis of trait variation is fundamental to advancing crop and livestock improvement programs. Identifying quantitative trait loci (QTL) or genomic regions harbouring mutations that contribute to trait variation provides crucial insights into genetic architecture and enables targeted breeding strategies. Traditional genome-wide association studies (GWAS) detect QTL by testing associations between individual genetic markers and phenotypes using linear regression, while modern approaches utilize linear mixed models, Bayesian hierarchical models, or multi-stage methods. A significant association indicates the marker is in linkage disequilibrium (LD) with a causal mutation.

The single-marker approach proves effective in populations with limited LD, such as human populations with large effective population sizes. However, this approach lacks precision in populations characterized by extensive LD due to small effective population sizes – a common feature of crop and livestock populations. In these populations, markers that are in LD often split the QTL effect when estimated simultaneously, and individual markers may capture only fragments of the total QTL effect through imperfect LD relationships (Meuwissen and Goddard, 2000; Marchini et al., 2004; Goddard, 2009; Zhang et al., 2010). This dilution of signal across multiple correlated markers reduces the power to detect causal variants.

To address these limitations, chromosome segments in strong LD can be leveraged through haplotype-based approaches that capture the collective signal across multiple linked markers (Meuwissen and Goddard, 2000; The Bovine HapMap Consortium et al., 2009; Comadran et al., 2011; Joukhadar et al., 2017). These methods group markers into haplotype blocks, typically defined as LD-based chromosomal segments (LD blocks or haploblocks), and analyse the combined effects within each block. One promising haplotype-based method utilizes the variance of local genomic estimated breeding values (localGEBV), which represents the variance of haplotype effects within a block. This approach has demonstrated the ability to detect QTL where causal mutations are known in livestock populations (Fan et al., 2011; Kemper et al., 2015; Van Den Berg et al., 2019; Xiang et al., 2021). Unlike traditional GWAS that tests markers individually, the localGEBV approach, derived from genomic prediction methods, fits all single nucleotide polymorphism (SNP) simultaneously (Meuwissen et al., 2001), counteracting shrinkage effects and the dilution of marker effects when multiple markers tag the same QTL. Theoretical advantages of the localGEBV approach include reduced susceptibility to false positives from population stratification, less

severe over-estimation of effects, and improved QTL detection power, all stemming from the simultaneous fitting of all SNP rather than individual testing (Yang et al., 2014).

The localGEBV method has also shown promise in crop applications. Voss-Fels et al. (2019) first applied this approach in wheat to identify beneficial haplotypes for yield, though their focus was on breeding program improvements rather than QTL discovery methodology. Others have applied the same method to identify haplotypes important for disease resistance in barley and wheat (Tong et al., 2024; Roy et al., 2025; Vo Van-Zivkovic et al., 2025), stay-green and root traits in barley (Brunner et al., 2024; Alahmad et al., 2025; Aldiss et al., 2025), and chickpea seed weight, flowering time, and plant height (Varshney et al., 2021; Akinlade et al., 2025).

Despite these promising developments, critical knowledge gaps limit our understanding and application of haplotype-based QTL detection. No direct comparison between localGEBV and standard GWAS approaches has been conducted to empirically evaluate their relative performance. Previous applications in QTL discovery in livestock have relied on sliding window approaches, whereas in crop species, haplotype discovery focused on biologically meaningful LD-based block. However, the effect of fundamental methodological aspects, including optimal marker effect priors and blocking parameters, remain poorly characterized for QTL discovery.

Here, we address these gaps through comprehensive evaluation of the localGEBV approach using a barley global diversity panel genotyped with 40K SNP, focusing on spike row morphology, a trait with complex genetic architecture distinguishing 2-row from 6-row varieties. We construct LD-based haploblocks, investigate localGEBV estimation using ridge regression best linear unbiased prediction (rrBLUP) and BayesR to account for different marker effect priors, and directly compare QTL detection capabilities against established multi-locus GWAS methods (FarmCPU and BLINK). Our findings demonstrate that haplotype-based localGEBV offers superior precision and power for QTL discovery in populations with extensive LD by detecting multiple known QTL compared to modern GWAS methods, providing a robust framework applicable across plant and animal breeding programs.

Results

Barley accessions display distinct genetic differentiation for row-type phenotypes

Principal component analysis revealed substantial genetic differentiation among barley accessions, with PC1 and PC2 explaining 30.4% and 14.4% of variation, respectively. *K-means* clustering identified three main genetic clusters (Figure 1A): Cluster 1 (313) comprised predominantly 2-row accessions, while Clusters 2 (297) and 3 (180) were primarily composed of 6-row accessions (Figure 1B and 1D).

LD analysis demonstrated high initial LD ($r^2 = 0.74$) within distances <0.1 Mbp, followed by rapid decay to $r^2 = 0.19$ at 1 Mbp, with subsequent gradual decay thereafter (Figure 1C).

LD-based haploblock construction reveals diverse genomic segmentation patterns

The construction of chromosome segments using varying LD parameters generated distinct haploblock configurations. Relaxed parameters (lower r^2 threshold, higher marker tolerance, tol) generated fewer haploblocks total with more markers per haploblock, while stringent parameters (higher r^2 threshold, lower marker tolerance) produced more numerous haploblocks that were predominantly single-marker blocks (Supplementary Figure 1). For instance, at $r^2 = 0.1$ with zero marker tolerance, 3,738 haploblocks (447,691 possible haplotypes) were constructed, with 60.27% comprising single SNP markers. Increasing the r^2 threshold to 0.5 resulted in 6,781 haploblocks (103,857 possible haplotypes), where 73.72% consisted of single SNP markers.

Increasing marker tolerance substantially reduced haploblock count while expanding block size and possible haplotype numbers. At $r^2 = 0.1$ with marker $tol = 3$, only 1,096 haploblocks were constructed (2,890,043 possible haplotypes). The largest observed haploblock (364 SNP markers spanning 299 Mbp on chromosome 5H) contained 374,192 possible haplotypes.

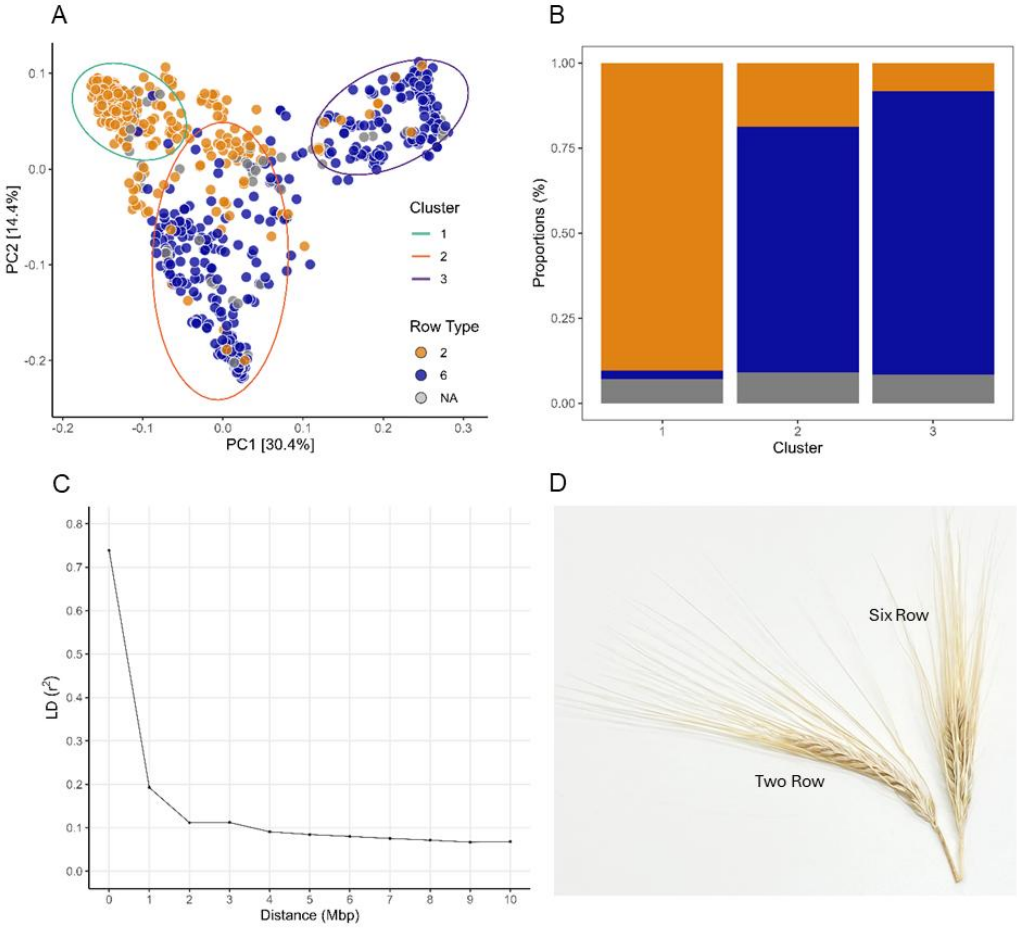


Figure 1. **Population structure and genetic relationships of barley germplasm.** (**A**) Principal component analysis of *N* = 790 barley accessions using 9,050 XT SNP chip markers, revealing three distinct genetic clusters. Shown are PCs 1 and 2 from Roger's genetic distance matrix, accounting for 44.8% cumulative explained variation. The optimum number of clusters were *clusters* = 3: Cluster 1 (313), Cluster 2 (297), and Cluster 3 (180). *Colours* represent the row spike phenotype i.e. 2-row is *orange*, 6-row is *blue*. (**B**) Proportions of barley accessions across the three clusters: Cluster 1 (2-row = 283, 6-row = 8, NA = 22), Cluster 2 (2-row = 56, 6-row = 214, NA = 27), Cluster 3 (2-row = 15, 6-row = 150, NA = 15), where NA represents accessions with ambiguous information and 'irregular' type. (**C**) Genome-wide LD decay as a function of physical distance (Mbp). LD decay is shown at an average distance (Mbp) between adjacent SNP based on the XT SNP chip physical position using cv Morex assembly v2 and visualised up to 10 Mbp. (**D**) An example of the two-row and six-row phenotypes.

Validation of localGEBV approach for mapping *VRS1* gene in chromosome 2H

To determine whether haploblock variance was influenced by segment size, we applied a generalized additive linear regression model, which revealed no meaningful linear relationship between haploblock variance and block size (number of markers; $R^2 = 0.07$, $P <$ 2e-16; Figure 2, Supplementary Figure 3). Instead, a significant and meaningful relationship was observed between haploblock variance and haplotype effects ($R^2 = 0.48$, $P <$ 2e-16).

We validated the localGEBV approach by examining the well-characterized *VRS1* gene on chromosome 2H, which determines row-type phenotype (Komatsuda et al., 2007). Marker effect estimation using rrBLUP showed relatively small effects distributed across the chromosome, with modest increase in the *VRS1* region (Supplementary Figure 2). In contrast, BayesR attributed larger effects to SNP collocated with *VRS1* compared to surrounding markers. Analysis of localGEBV variance across different LD parameters consistently identified haploblocks peaking at the *VRS1* region (Figure 3A). Both rrBLUP and BayesR detected maximum variance at $r^2 = 0.1$ and marker tolerance, $tol = 3$, co-locating with *VRS1* and explaining 0.81% and 1.32% of localGEBV variance, respectively.

Closer examination of the highest variance haploblock in chromosome 2H revealed progressive refinement from 49 SNP spanning 10.23 Mbp at $r^2 = 0.1$ to 4 SNP spanning 0.88 Mbp at $r^2 = 0.5$ (Figure 3B). BayesR consistently produced higher variance estimates at the *VRS1* region compared to rrBLUP, likely due to its heterogeneous variance assumptions allowing larger effects for SNP in high LD with QTL. Additionally, rrBLUP identified dispersed regions of additional high variance haploblocks in chromosome 2H, particularly within 231.42 to 431.25 Mbp when using relaxed LD parameters. These regions diminished in BayesR at higher LD thresholds, consistent with its assumption that most SNP have no effect on the trait (Supplementary Figure 2).

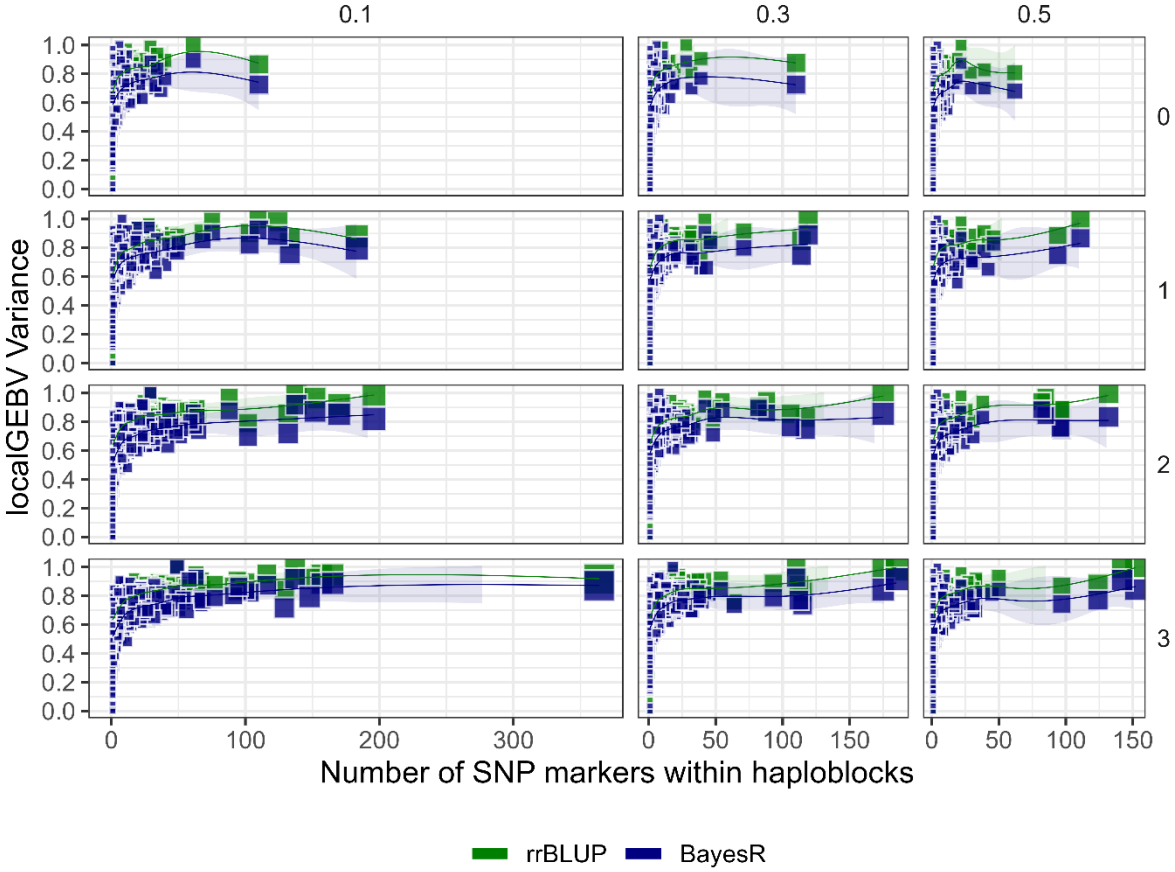


Figure 2. Relationship between haploblock variance and haploblock size. Haploblock variance derived from localGEBV estimates using rrBLUP (*green*) and BayesR (*blue*), plotted on the *y-axis*. Block size is defined by the number of SNP markers within haploblocks, plotted on the *x-axis*. In the figure, different LD parameters are shown as facets: *column facets* represent LD thresholds of $r^2 \in \{0.1, 0.3, 0.5\}$; and *row facet* as marker tolerance, $tol \in \{0, 1, 2, 3\}$. The localGEBV variance was scaled using a min-max scaling of the $\log_{10}$ transformed variance. Blocks were connected using a generalized additive model for non-linear relationships to represent trends between haploblock size and localGEBV variance.

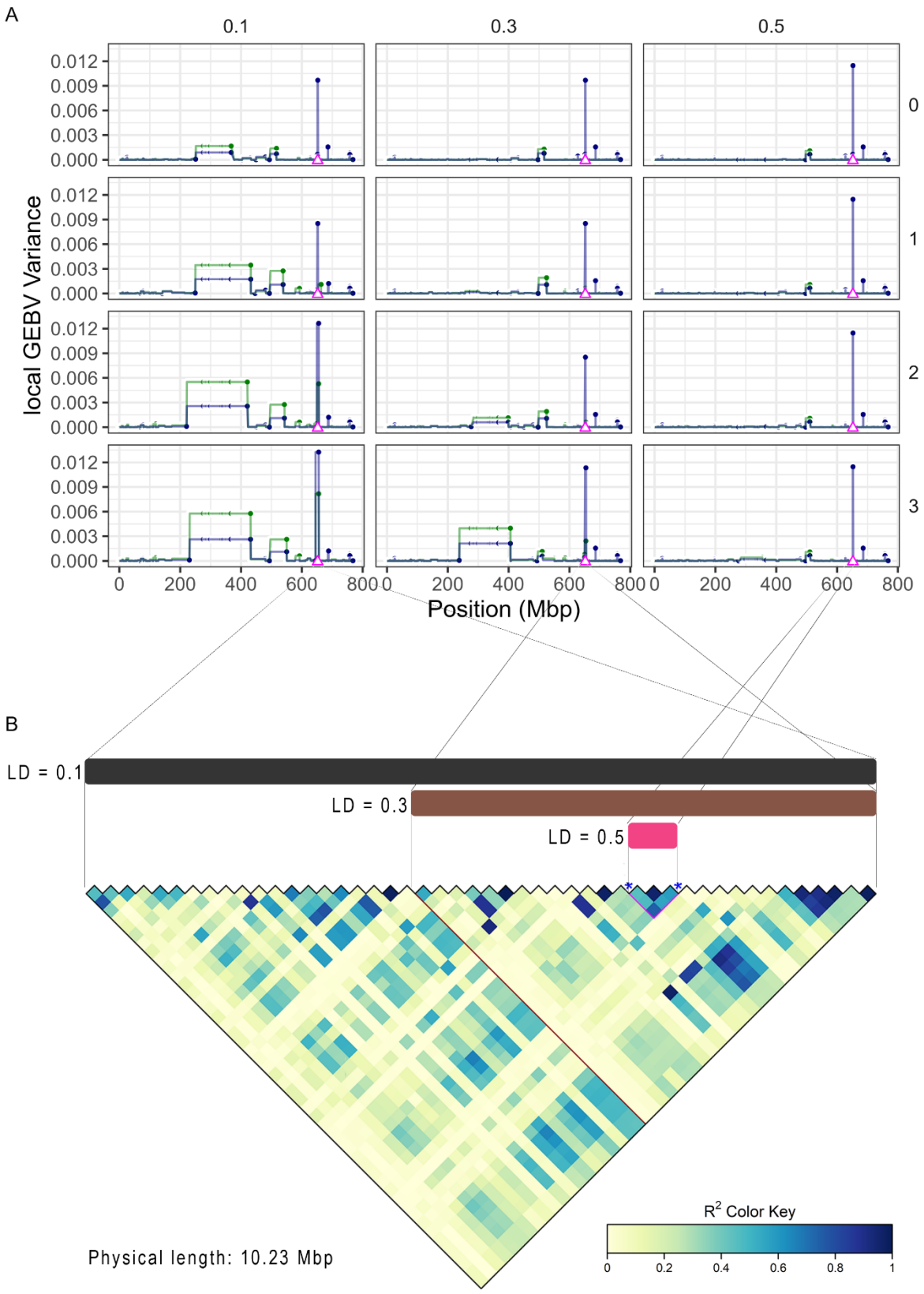


Figure 3. LocalGEBV variance of haploblocks in chromosome 2H. The *VRS1* gene (marked as Δ) is co-located at 652.03 Mbp based on the XT SNP chip physical position using cv Morex assembly v2. (**A**) The localGEBV variance profiles across chromosome 2H

estimated using rrBLUP (*green*) or BayesR (*blue*). Haploblocks are constructed according to different LD parameters: where *column facet* represents LD threshold at $r^2 \in \{0.1, 0.3, 0.5\}$; and *row facet* as $tol \in \{0, 1, 2, 3\}$. (**B**) Varying sizes of the haploblocks based on LD threshold show fine-scale mapping in the *VRS1* region. Haploblocks constructed at $r^2 \in \{0.1, 0.3, 0.5\}$ with $tol \in \{0, 1, 2, 3\}$. Shown together is the corresponding LD heatmap.

Ability of localGEBV to detect QTL

We assessed QTL detection ability using scaled inverse chi-squared probability values ($P_{invchisq}$) calculations for localGEBV variance. Detection ability was generally higher with relaxed LD parameters (lower LD threshold and higher marker tolerance), with comparable performance between rrBLUP and BayesR methods (Figure 4A). However, detection ability decreased with stringent LD parameters.

From the highest detection ability (Figure 4A), we summarised results from haploblocks built on parameters $r^2 = 0.1$ and $tol = 3$. We identified 102 high-variance haploblocks from the total of $J = 1096$ haploblocks that surpassed the mean haploblock variance (Figure 6; Table 1). The corresponding $P_{invchisq}$ from the visually determined scaled inverse chi-squared distribution of these blocks ($\nu = 0.5$ and $\tau^2 = 0.5$) ranged from 4.9e-324 to 0.003 (Table 1). Furthermore, strong correlations were observed between haploblock variance and $P_{invchisq}$ values (rrBLUP $r = 0.70$; BayesR $r = 0.88$). Given we defined significance at $P_{invchisq} < 0.001$, our results indicate that utilising haploblock variance with the mean haploblock variance as a significance threshold is likely a good indicator of whether the variance would likely be considered significant. This indicates that the haploblock variance is a robust alternative to the significance testing paradigm usually found in GWAS methodologies.

Comparative GWAS analysis using FarmCPU and BLINK identified 6 significant markers: 5 on chromosome 2H (2 detected by both methods) and 1 on 4H (FarmCPU only). In contrast, localGEBV identified 20 high variance haploblocks distributed across all chromosomes, with 7 detected by both rrBLUP and BayesR, and 2 and 11 uniquely identified by each method, respectively (Table 1, Figure 4B).

Haploblock 2H:b000235 exhibited the highest localGEBV variance (rrBLUP = 0.0082, $P_{invchisq}$ = 1.52e-05; BayesR = 0.0133, $P_{invchisq}$ = 0.003). This haploblock (spanning 645.10 to655.33 Mbp) contained all 4 significant markers detected by GWAS, including the most

198 significant marker 2H:AVRIG03047 ($-\log_{10} P_{adj}$; FarmCPU = 17.6, BLINK = 10.5) at
199 652.42 Mbp, co-locating with the *VRS1* gene (Komatsuda et al., 2007; Table 1, Figure 4B).

200 Notably, the localGEBV approach identified additional high variance haploblocks ($P_{invchisq}$
201 range = 4.0e-08 to 4.1e-69) that collocated with known genes and candidate genes related to
202 row phenotype. Both rrBLUP and BayesR identified haploblocks 1H:b000049 and
203 3H:b000382 co-locating with *VRS3* (associated with grain uniformity; Van Esse et al., 2017;
204 Bull et al., 2017) and *HORVU3Hr1G036930* (linked to lateral spikelet fertility; Youssef et al.,
205 2020), respectively. BayesR uniquely identified haploblocks 1H:b000091, 4H:b000552, and
206 7H:b001032 co-locating with genes and QTL associated with "third outer glume" structure
207 (TRD) and lateral spikelet fertility (*VRS5* and *HORVU7Hr1G090180*; Ramsay et al., 2011;
208 Youssef et al., 2020), none of which contained significant markers in the FarmCPU and
209 BLINK analyses.

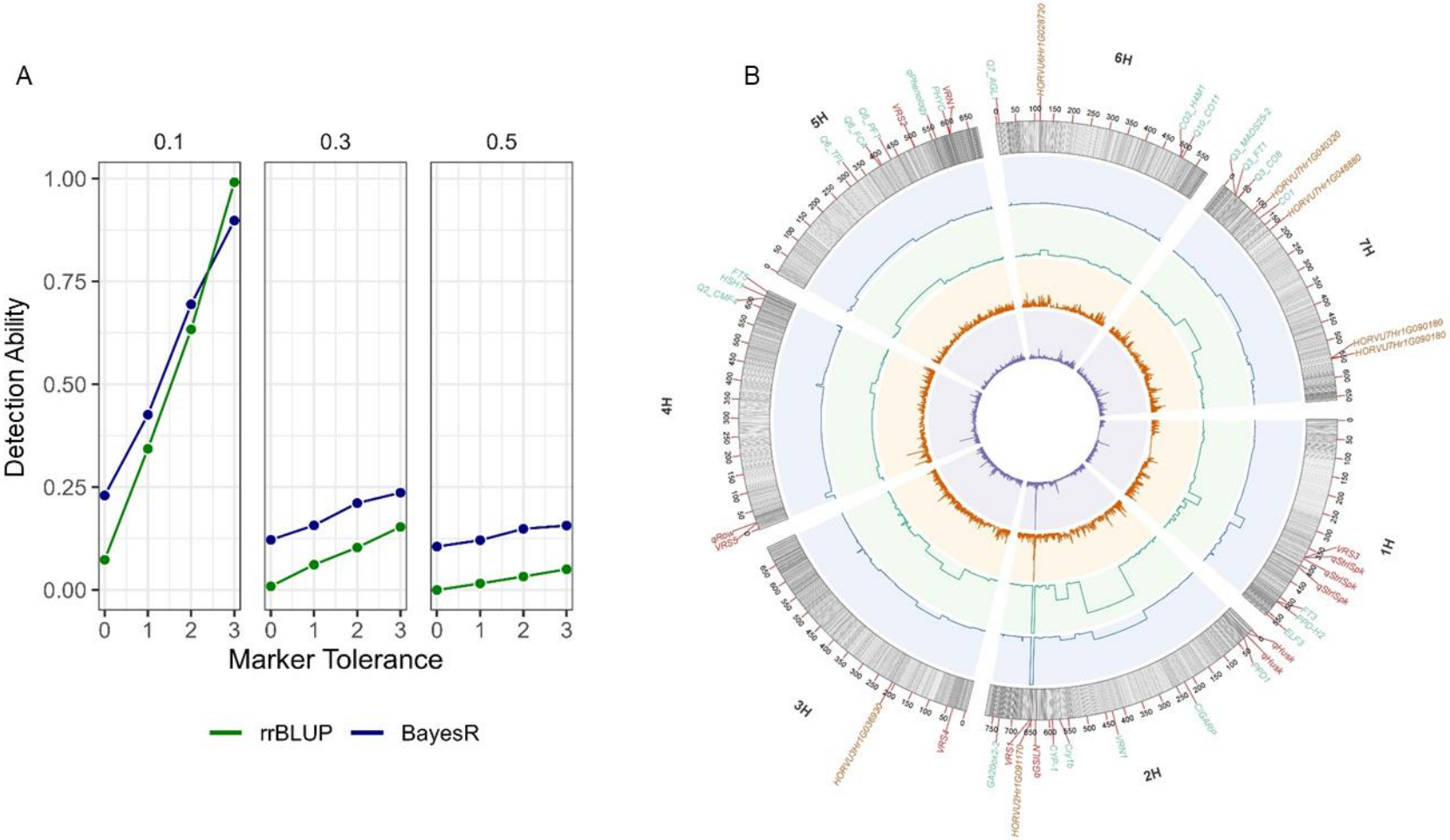


Figure 4. Detection of haploblocks containing QTL. (**A**) Detection ability of localGEBV from estimates using rrBLUP (*green*) or BayesR (*blue*). Haploblocks are constructed according to different LD parameters: where *column facet* represents LD threshold at $r^2 \in \{0.1, 0.3, 0.5\}$;

and *x-axis* as marker $tol \in \{0, 1, 2, 3\}$. (**B**) Physical position in barley chromosome of SNP markers associated with row-type spike architecture detected from localGEBV and GWAS**.** ***Tracks 1 to 5 (outer to innermost):*** **Track 1** Barley chromosome showing physical position (Mbp) based on cv Morex assembly v2 from the 40K XT SNP chip. SNP markers are coloured in *grey*. Colocation of known genes and QTL from the literatures is shown: row-type known genes (*VRS3* in 1H, *VRS1* in 2H, *VRS4* in 3H, *VRS5* in 4H, *VRS2* in 5H; *dark red*), phenology genes (*FT3*, *PPD1*, *CIGARP*, *Cry1b*, *CYP-1*, *GA20ox2*-2, Q2_*CMF4*, *HSH1*, *FT5*, Q6_*TFL*, Q6_*FCA*, *Q6_PFT*, *PHYC*, *Q7_AGL1*, *CO2_H4M1*, *Q10_CO11*, *Q3_MADS25-2*, *Q3_FT1*, *Q3_CO8*, *CO1*; *green*), yield-component candidate genes (*brown*); and QTL for spike sterility (qStrlSpk), grain spiculation of inner lateral nerve (qGSILN), spike row number (qRow), are also shown in *dark red*. **Track 2** LocalGEBV estimated using Bayes R, track colour in *blue*. **Track 3** LocalGEBV estimated using rrBLUP, track colour in *green*. **Track 4** GWAS using BLINK, track colour in *orange*. **Track 5** GWAS using FarmCPU, track colour in *purple*. In both **Track 2** and **Track 3**, localGEBV haploblocks were according to LD blocking parameters: LD threshold $r^2 = 0.1$ and marker $tol = 3$; and shown is the variance localGEBV according to the position of flanking markers within haploblocks. In both **Track 4** and **Track 5**, markers were considered significant according to thresholds after FDR corrections ($-\log_{10} P_{adj} > -\log_{10} 0.008$ for FarmCPU and $-\log_{10} P_{adj} > -\log_{10} 0.004$ for BLINK).

Haploblocks demonstrate superior predictability compared to single markers

We conducted principal component analysis on the marker effects from each method of estimating marker effects and conducting GWAS (FarmCPU, BLINK, rrBLUP, and Bayes R). Principal component analysis of marker effect estimates from all methods showed PC1 and PC2 explaining 73.93% of cumulative explained variance, with marker 2H:AVRIG03047 contributing most substantially (17.49%), followed by 4H:AVRIG07945 (5.64%; Figure 5A). We further evaluated the regression of row-spike status on SNP markers or localGEBV for row-spike phenotype prediction accuracy utilising both a LPM and logistic regression in a generalized linear model (GLM) framework. Correlations between the predicted phenotype and actual phenotype in the test set were consistently higher for both significant markers and high variance haploblocks compared to randomly selected markers or haploblocks (Figure 5B). Notably, both rrBLUP and BayesR achieved higher correlation values when using haploblocks as predictors (LPM = 0.934, GLM = 0.889; Root Mean Squared Error; RMSE: LPM = 0.09, GLM = 0.47) compared to single markers (LPM = 0.716, GLM = 0.751; RMSE: LPM = 0.04, GLM = 3.45). The RMSE for the linear probability model (LPM) utilising single markers was lower. This is likely a downwards biased estimate because of the violation of ordinary least squares assumptions and should not be utilised to justify a better model fit than the logistic regression model.

Haplotype analysis reveals distinctive features of localGEBV

We explored the relationship between haploblock variance and the effect of individual haplotypes (Figure 6). In general, as the magnitude of a localGEBV increased, the variance of the corresponding block was greater (Figure 6, Supplementary Figure 3).

Detailed haplotype profiling of haploblock 2H:b000235 (*VRS1* region) at higher LD resolution ($r^2 = 0.5, tol = 3$) revealed 4 SNP markers spanning 0.88 Mbp (651.54 to 652.42 Mbp) with varying allele frequencies (Figure 7B). From these markers, 30 haplotypes were identified, with 13 occurring at frequencies ≥5 accessions. Two predominant haplotypes emerged as complete allelic reversals: hap1 (TTTT) associated primarily with 2-row phenotype (252) and hap2 (AAAA) with 6-row phenotype (239).

Haplotype network analysis demonstrated that the direction of haplotype effects aligned with phenotypic expression: negative effects for hap1 is associated with 2-row and positive effects for hap2 is linked with 6-row (Figures 6, 7C); this aligns with the regression analysis in accurately predicting 2-row or 6-row phenotypes utilising the same haplotype effects (Figure

257 5B). Hierarchical clustering based on similarity matrix scores differentiated distinct clusters
258 corresponding to haplotype-phenotype associations, wherein Australian Grains Genebank
259 (AGG) accessions carrying hap1 had 2-row spikes while AGG accessions with hap2 had 6-
260 row spikes (Figure 7D).

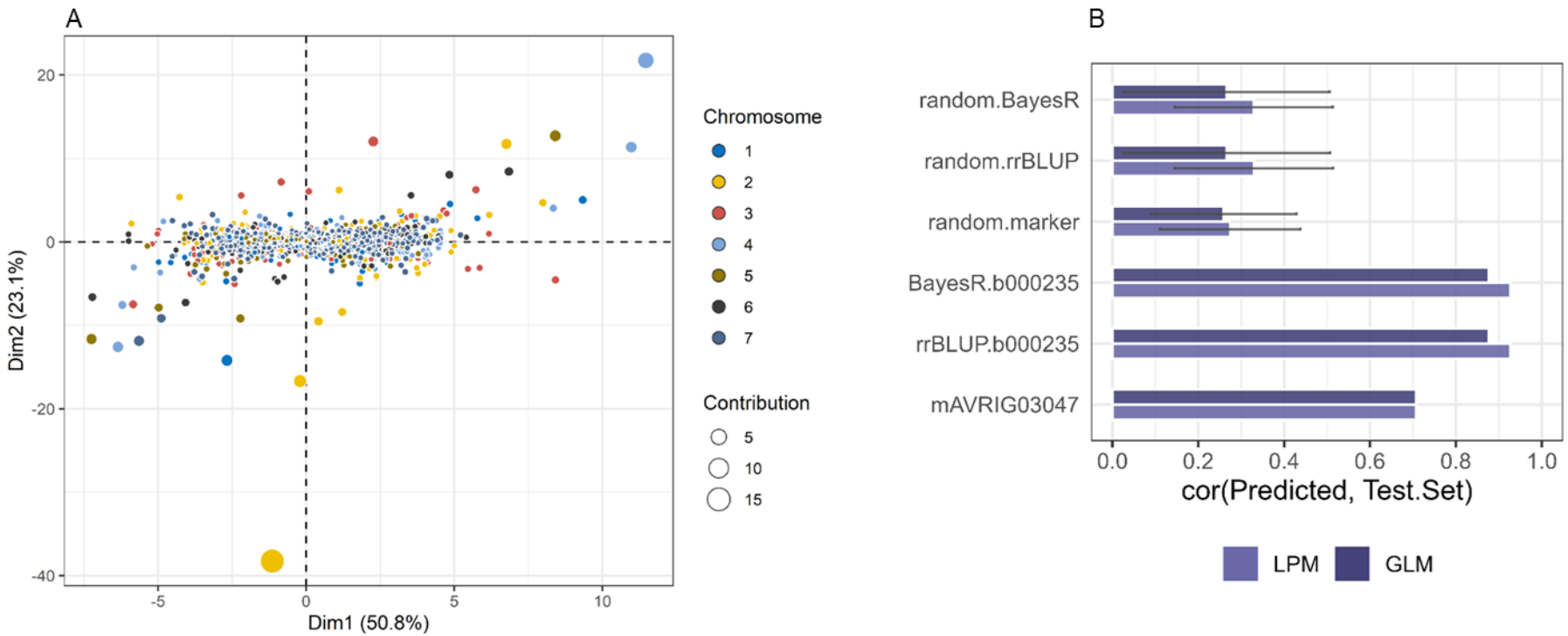


Figure 5. Accuracy of predictors in identifying informative associations with row-spike phenotype. (**A**) Principal component analysis showing the effects contribution of $M = 9{,}050$ XT SNP marker estimates from FarmCPU, BLINK, rrBLUP, and BayesR. *Colours* represent the 7 barley chromosomes; *size* of each dots represents the SNP effect contribution to the principal components relating to row-spike phenotype. (**B**) Regression analysis of markers or haploblocks as predictors. Shown is the accuracy calculated as correlations of the predicted value in the testing set; predicted values calculated using ordinary least squares linear probability model (LPM, *light purple*) and logistic regression in the generalised linear model (GLM, *dark purple*). The most significant marker (2H:AVRIG03047) identified by both FarmCPU and BLINK and the haploblock with the highest variance from rrBLUP and BayesR localGEBV estimates (2H:b000235) were analysed in each model and compared

269 with non-significant random markers/haploblocks. Analyses for random markers/haploblocks were repeated 100 times, and the average
270 correlation and standard deviations of the 100 repetitions represented as error bars were shown.

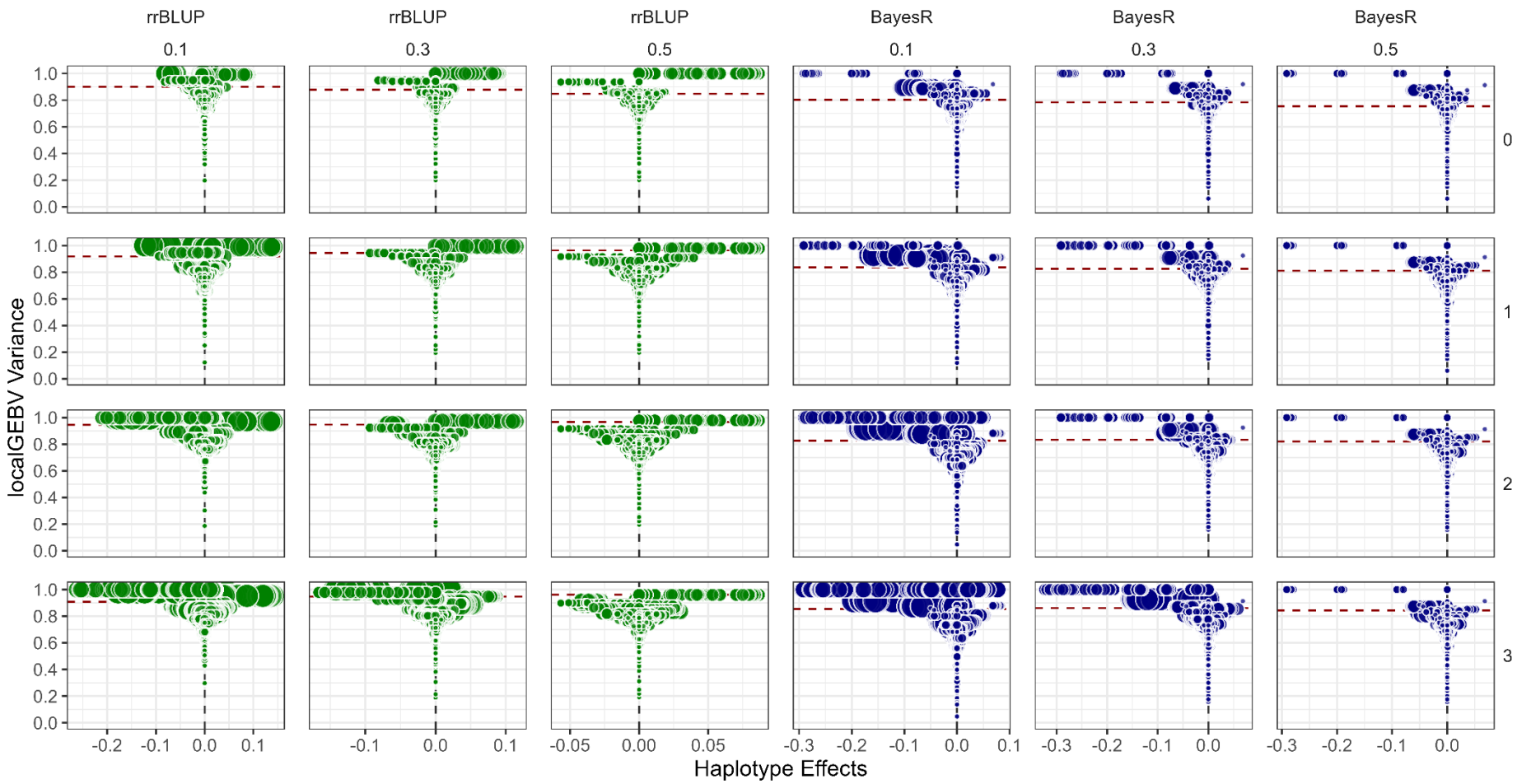


Figure 6. Relationship between haploblock variance and haplotype effects. Haploblock variance was determined from the localGEBV estimates using rrBLUP (*green*) and BayesR (*blue*); plotted as min-max scaled of the log_{10} transformed variance for visual clarity. In the figure, different LD parameters are shown as facets: *column facet* represented by LD threshold at $r^2 \in \{0.1, 0.3, 0.5\}$; and *row facet* as marker tolerance

$tol \in \{0, 1, 2, 3\}$. *Dashed horizontal lines* represent the average haploblock variance at each LD parameter, where haploblocks above the threshold are considered containing the QTL. *Size* of each circle corresponds to the number of markers within the haploblock.

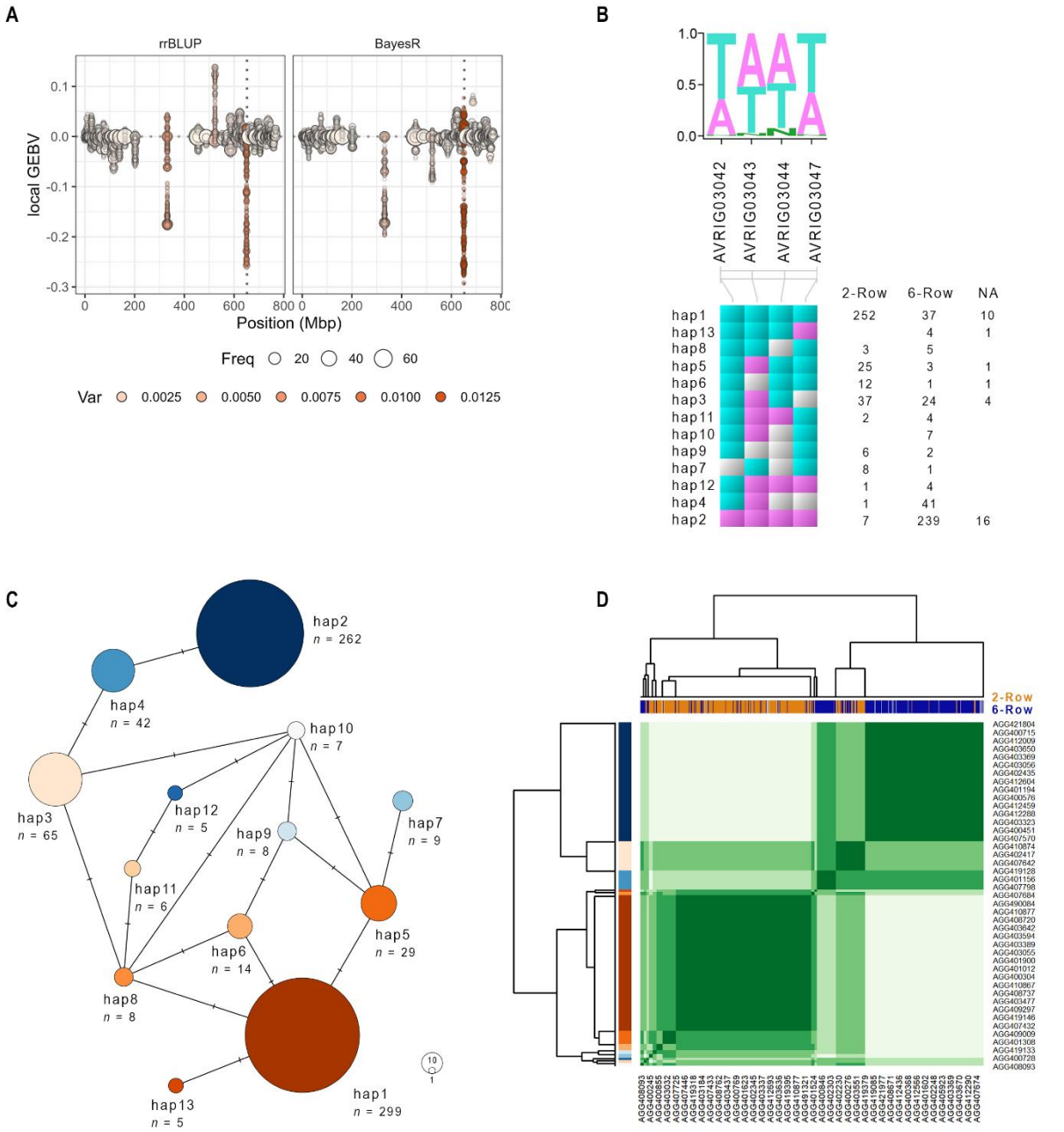


Figure 7. Haplotype analysis. (**A**) Identification of haplotypes from the localGEBV estimates using rrBLUP and BayesR. In the figure, *x-axis* shows the marker physical position (Mbp) of chromosome 2H based on cv Morex assembly v2 from the 40K XT SNP chip; while the *y-axis* corresponds to localGEBV or haplotype effects. The *dotted vertical line* illustrates the position of *VRS1* gene at 652.03 Mbp. Haplotypes associated with row-type phenotypes are determined by the direction of localGEBV estimates, whereby negative localGEBV effects favours 2-row phenotypes while positive localGEBV effects favours 6-row phenotypes. Each *dot* corresponds to a specific haplotype, wherein the *size* of the dots represents the haplotype abundance or the frequency of barley accessions carrying the

287 haplotype; while *orange colour gradient* represents higher haploblock variance. (**B**)
288 Haplotype profile collocating *VRS1* gene. Shown is the 0.88 Mbp finer region within
289 2H:b00023 defined by four markers (i.e. AVRIG03042, AVRIG03043, AVRIG03044,
290 AVRIG03047; spanning 651.54 to 652.42 Mbp); wherein, arbitrarily, *A* corresponds to the
291 reference allele while *T* corresponds to the alternate allele, missing allele is represented by *N*.
292 Marker AVRIG03047 was also the most significant marker identified by FarmCPU
293 ($-\log_{10} P_{adj} = 17.59$) and BLINK ($-\log_{10} P_{adj} = 10.49$). The *top panel* shows the
294 estimated density of alleles at each four marker positions, while the *bottom panel* shows the
295 13 haplotypes identified and their frequency of occurrence across barley accessions relative
296 to each row-type phenotype. Only haplotypes with a frequency *of at least* 5 accessions are
297 shown. **(C**) Haplotype network analysis. Shown is a statistical parsimony network (TCS
298 method; Templeton et al., 1992) of the 13 haplotypes. The *size* of the circle is proportional to
299 the number of barley accessions carrying the respective haplotypes, while the *hash marks*
300 indicate the number of mutational steps. *Colour gradient* illustrates the localGEBV or
301 haplotype effects, with *dark orange* and *dark blue* indicating the lowest (i.e. negative) and
302 highest (i.e., positive) haplotype effects, respectively. (**D**) Heatmap of similarity matrix score
303 of the haplotypes among 759 barley accessions. In the figure, barley accessions are separated
304 into two major clusters according to complete linkage hierarchical clustering, wherein the *top*
305 *colour bar* represents the row-spike phenotypes of the barley accession (i.e., 2-row in *orange*
306 and 6-row in *blue*) while the *left colour bar* represents the 13 haplotypes described in panels
307 B and C.

Discussion

In this study, we demonstrated that localGEBV offers substantial improvements in QTL discovery compared to standard GWAS methods. By integrating LD information with genomic prediction models, the framework effectively partitions the genome into disjoint chromosome segments with estimated haplotype effects (localGEBV) and haploblock variances. This allowed us to leverage local information in the genome enabling more comprehensive detection of genetic loci controlling complex traits. Furthermore, we assessed the impact different marker tolerance and LD threshold blocking parameters had on discovery to provide a general guideline on how to conduct blocking.

Biological complexity of barley row-type phenotype as a model for QTL discovery

The barley row-type phenotype represents an ideal model for testing QTL discovery methods due to its simple phenotype but interconnected genetic control (Bull, 2015). The morphological distinction between 2-row and 6-row phenotypes is determined by the fertility status of lateral spikelets. In 2-row barley, only central spikelets are fertile and develop grains, while lateral spikelets remain sterile. In contrast, 6-row barley exhibits fertility in both central and lateral spikelets, producing more grains per spike and significantly influencing yield component traits (Youssef et al., 2017) such as thousand kernel weights and kernel number.

The genetic architecture underlying this trait involves a network of co-regulated genes, however, there is still little information on the exact mechanism (Zwirek et al., 2019). While *VRS1* on chromosome 2H serves as the most downstream genetic component (Komatsuda et al., 2007), several other characterised genes regulate lateral spikelet fertility and overall spike morphology, including *VRS2* in 5H (Youssef et al., 2017), *VRS3* (syn. INTERMEDIUM-A) in 1H (Van Esse et al., 2017; Bull et al., 2017), *VRS4* (syn. INTERMEDIUM-E) in 3H (Koppolu et al., 2013; Koppolu et al., 2022), and *VRS5* (syn. INTERMEDIUM-C) in 4H (Ramsay et al., 2011). Contemporary 6-row barley cultivars typically contain natural recessive *vrs1* and *vrs5* alleles (Zwirek et al., 2019).

Limitations of standard GWAS approaches

Various GWAS methods remain the predominant method for QTL discovery and genetic architecture analysis but face significant challenges including spurious associations and limited power to explain complex trait variants (Wang and Xu, 2019). Our study utilised a

339 diverse barley panel with pronounced population structure, providing an opportunity to
340 evaluate different mapping approaches with a trait documented to have intricate genetic
341 control.

342 Despite the significant improvement of statistical methods in both FarmCPU and BLINK that
343 employ multi-locus genome scans (Liu et al., 2016; Huang et al., 2019; Wang and Zhang,
344 2021), these approaches only detected the major *VRS1* gene, failing to identify other known
345 genes associated with row-type characteristics (Table 1). This limitation exemplifies a
346 common problem in complex trait analysis where primary morphological characteristics are
347 influenced by secondary traits confounded by genetic background (Zwirek et al., 2019) or
348 masked due to epistatic interaction (Ramsay et al., 2011). The low detection ability also likely
349 stems from uncompounded effects of co-regulated genetic interactions (Tam et al., 2019),
350 even after controlling for population structure. This can occur when the effect of a QTL is
351 split amongst multiple markers because of extensive LD (correlation) structures between the
352 markers and QTL. Alternatively, multiple markers may only capture a part of the QTL effect
353 each if they are only partially correlated with the QTL (Meuwissen and Goddard, 2000;
354 Marchini et al., 2004; Goddard, 2009; Zhang et al., 2010).

355 Notably, a critical constraint of GWAS is the reliance on stringent significance thresholds,
356 which are implemented to decrease false positives at the expense of decreasing power and
357 increasing the number of false negatives for variants with modest effects. The conventional
358 approach of applying genome-wide significance thresholds (FDR adjustments or Bonferroni-
359 corrected values) prioritizes certainty of associations over discovery breadth. This limits
360 statistical power, excluding true causal variants that fall below the significance threshold
361 (Wang et al., 2016; Tam et al., 2019). A recent alternative to Bonferroni and FDR adjustment
362 utilized the LD between markers to calculate the true number of effective tests and control
363 type I error (Gao et al., 2008). This would likely increase the power to detect QTL and is
364 based on a similar philosophy to haploblocking, but it would not necessarily recover any split
365 signal like the localGEBV method would.

366 In a trait where multiple genes contribute through intricate regulatory networks, like barley
367 row-type, this statistical stringency disproportionately penalises small-effect loci that
368 collectively determine phenotypic expression for complex, polygenic traits. This
369 phenomenon has been well-documented across multiple crop species. In maize, numerous
370 small-effect loci controlling numerous phenotypic traits were missed by conventional GWAS

371 but enriched with synonymous SNP because of linkage with nearby causal SNP (Wallace et
372 al., 2014). In rice, chlorophyll content and stay-green traits are governed by networks of
373 interacting genes whose individual effects were too small to reach significance thresholds yet
374 collectively explained substantial phenotypic variation (Zhao et al., 2019). Bustos-Korts et al.
375 (2019) demonstrated that barley adaptation traits, such as developmental and row-type traits,
376 are controlled by complex networks of interacting loci whose effects become diluted when
377 analysed individually. However, these effects can be captured by exomic haplotype states due
378 to the larger number of allelic states, supporting our observation that genetic interactions in
379 the *VRS* gene network may go undetected using conventional approaches.

380 **Methodological advantages of localGEBV**

381 In contrast to standard GWAS, the localGEBV approach demonstrated superior detection
382 power for known genes associated with barley spike row phenotype. That is, we likely
383 controlled the number of false positives while maintaining statistical power. While FarmCPU
384 and BLINK detected only the major *VRS1* gene on chromosome 2H, the localGEBV
385 approach additionally identified haploblocks containing other known row-phenotype
386 regulators, including *VRS3* in 1H (Van Esse et al., 2017; Bull et al., 2017), *VRS5* in 4H
387 (Ramsay et al., 2011), and several annotated QTL (Youssef et al., 2017) that remained
388 undetected by conventional methods (Figure 4B). Notably, the localGEBV methods reduced
389 the background noise compared to the GWAS results from FarmCPU and BLINK, and clearly
390 delineated genomic signals regardless of whether marker effects were estimated using
391 rrBLUP or BayesR. The smoothing of background noise may be indicative of marker effects
392 from non-QTL regions averaging towards small values and reducing the corresponding
393 haploblock variance when combined into localGEBV utilising LD structure. This
394 demonstrates our method's ability to leverage localised genomic information to capture QTL
395 signals that would otherwise be lost to background noise and false-positive control.

396 In addition to greater detection power, the localGEBV method leverages localised
397 information on LD to fine-tune the region of interest. This relies on historic recombination,
398 drift, and selection patterns that influenced the LD development in the genomic region. In
399 contrast, when interpreting GWAS results, there is no defined region of interest and genome-
400 wide average LD patterns are often utilised to investigate potential QTL or genes of interest.
401 Thus, our methodology helps to accurately assess where a QTL may reside (Figure 3B).

402 Finally, a key advantage of the local GEBV approach beyond statistical power and accurately
403 defining QTL regions is its flexibility in accommodating different models to estimate
404 haplotype effects, as exemplified in Figure 4B and 5B. BayesR demonstrated exceptional
405 precision in estimating large-effect QTL, particularly visible in the *VRS1* region. This is not
406 surprising as it aligns with the major role of *VRS1* in the development of the six-rowed spike
407 phenotype controlled by a single recessive allele (*vrs1*), in contrast to the dominant allele for
408 the two-rowed phenotype (Komatsuda et al., 2007). This advantage stems from the model's
409 ability to accommodate various assumptions in estimating marker effects (Erbe et al., 2012;
410 Moser et al., 2015; Wolc and Dekkers, 2022), leading to improved genome-wide QTL
411 discovery (Kemper et al., 2015; Van Den Berg et al., 2019; Xiang et al., 2021). Conversely,
412 rrBLUP showed advantages in detecting regions with multiple small-effect alleles that
413 cumulatively form a significant signal (Figure 4B, Supplementary Figure 2). This model
414 flexibility allows application without prior knowledge of the trait's genetic architecture,
415 making it suitable for diverse breeding scenarios. For instance, this is evident in the nearby
416 *VRS1* region, where rrBLUP captured higher variance blocks (2H:b000210; Figure 3,
417 Supplementary Figure 2). This region was found not to co-locate with known genes or QTL,
418 suggesting it could represent a potentially new QTL associated with row-type and requires
419 further investigation. Thus, our results indicate that the localGEBV method can capture
420 different distributions of marker effects by utilising the infinitesimal assumptions found in
421 rrBLUP.

422 **LD parameters and genetic implications**

423 The precision of QTL discovery using localGEBV is directly influenced by LD structure
424 within the population. Indirectly, it is also based on marker positions. While we had a high-
425 quality reference map, it may be prudent to ignore marker positions and use the LD patterns
426 (genetic map) to form haploblocks if the reference is poorly constructed. It is also worth
427 noting that we utilized genome-wide blocking parameters, but parameters can also be defined
428 at the chromosomal level.

429 Our analysis revealed that while varying LD parameters affected haploblock size, they did
430 not alter the detection of informative segments for large effects (e.g., Figure 3A). This finding
431 aligns with previous studies concluding that no 'universal' or optimal haploblock construction
432 method applied to breeding programs exists (Andrade et al., 2019; Difabachew et al., 2023;
433 Weber et al., 2023). Stringent LD ($r^2 = 0.5$) produced smaller haploblocks (often comprised

of a single marker) enabling fine mapping of specific genes, while relaxed LD ($r^2 = 0.1$) generated larger haploblocks with enhanced detection power (Figure 3) independent of the genomic prediction method. The ability to choose parameters allows for flexibility and balancing trade-offs when conducting QTL discovery studies, a clear advantage not currently found in most, if not all, GWAS methods without utilising different priors on the underlying genetic architecture.

The biological significance of haploblock size warrants consideration. Smaller haploblocks may signify high LD between markers and the QTL (Andrade et al., 2019), facilitating precise candidate gene identification for post-GWAS analysis (Clark, 2004). On the other hand, larger haploblocks likely capture important biological mechanisms including local epistasis, pleiotropy (Kouyos et al., 2007; Zhang et al., 2014; Jiang et al., 2018), and adaptive mechanisms from balancing selection, mutation, and genetic drift (Barton, 2010). We did not observe this in our case, but we utilized a genetically intricate trait rather than fully Mendelian.

Additionally, information on recombination hotspots at haploblock boundaries (Kim et al., 2018; Epstein et al., 2023) can be valuable for breaking unfavourable linkage drag in breeding programs (Bohra et al., 2022) or stacking favourable haplotypes. Utilizing haploblocks provides an interesting advantage when introgressing new QTL (haplotypes) when there is complex population structure. Utilizing localGEBV from haploblocks provides a "genetic insurance policy" by co-selecting target genes with their local genetic background, thereby ensuring more predictable and stable trait expression when introgressed into new cohorts. The latter point is especially important in genetic improvement to maintain diversity and construct the “ultimate genotype” (Cole and VanRaden, 2011; Kemper et al., 2012; Hayes et al., 2024). This suggests our results likely have implications in breeding programs, crossing designs, and genomic improvement. This is comparable to GWAS and genomic selection sharing the same underlying framework and considerations, so it is natural to extend our framework and its implications to genomic improvement.

Lastly, it is important to note that constructing haploblocks based on LD is not always possible. When constrained by LD structures from diverse populations (Abdellaoui et al., 2023) like those commonly found in diversity panels, when SNP density is not homogeneous (Li et al., 2021), or when markers are extremely dense with common alleles mixed with rarer alleles (Taliun et al., 2014), alternative haploblock construction methods such as sliding

window approaches (Weber et al., 2023), gene boundary-based collapsing methods (Dering et al., 2011), or incremental haploblock partitioning (Taliun et al., 2014) can be adopted.

Biological insights from localGEBV features

The localGEBV approach provides deeper understanding of trait genetic architecture through two key features:

Haploblock variance. Haploblocks with high variances (greater than the mean haploblock variance) were consistently associated with substantial haplotype effects, representing true biological signals rather than artifacts of segment size (Figure 2 and 6).

This variance metric offers advantages over conventional *P-values* in GWAS, which provides statistical significance but often lack informative biological meaning (Visscher et al., 2017), may suffer from false associations (Yang et al., 2010; Wray et al., 2013) and do not provide a definitive region to explore QTL. They may also only capture part of the QTL signal or share its effect with nearby markers in high LD with the QTL (Meuwissen and Goddard, 2000; Marchini et al., 2004; Goddard, 2009; Zhang et al., 2010); however, this is dependent on the prior assumptions of marker effects, which our approach is more robust towards. By shifting focus from marker significance thresholds to haploblock variance or testing haploblock variance for significance, the localGEBV approach diminishes the problems associated with multiple testing that also makes small-effect loci effectively undetectable. This is because there are often far less haploblocks than markers, so less stringency on Type I error control is needed. Variance patterns can indicate a high degree of polymorphism in regions under selection pressure or fixation (Maynard and Haigh, 2007), absence of selective interactions (Otto and Barton, 1997), or enriched mutations affecting trait heritability (Sullivan et al., 2023).

Haplotype effects. Haplotypes provide superior predictive ability compared to single markers. Our regression analyses demonstrated that the haplotype effects within a particular, high-variance haploblock serve as better predictors of phenotypic traits than individual SNP. This highlights the importance of considering cumulative genetic effects in tightly linked regions (Meuwissen and Goddard, 2000), rather than isolated markers. Lower-frequency alleles with considerable effect can be successfully captured as a single unit within the haploblock (Berdan et al., 2023), enabling exploration of local allelic interactions and epistasis (Charlesworth, 1974).

Lastly, haplotype effects also signify directionality (Han and Pan, 2010) in relation to a reference haplotype, similar to the allele substitution effect of the alternative allele relative to the reference allele. Our analysis confirmed that negative-effect haplotypes were associated with 2-row phenotypes and positive-effect haplotypes with 6-row spikes (Figure 7A and 7D).

Implications for crop breeding

The localGEBV approach offers significant advantages in QTL discovery that could be used in genetic improvement strategies in crops. By exploiting LD blocks, breeders can determine both global and local genetic correlations between multiple traits (Visscher et al., 2017; Olasege et al., 2022), identify sources of genetic control under pleiotropic influence (Watanabe et al., 2019), and select favourable haplotype configurations. The approach effectively identifies compound-effect loci (Bevan et al., 2017) that influences complex traits. This is particularly valuable for traits that are difficult to phenotype; for example, fertility in barley is difficult to phenotype because fertility of lateral spikelets drives row-type morphology. Consequently, fertility can influence other yield components such as grain number and thousand kernel weight. As an example of a useful application, haploblocks that may benefit two unfavourably correlated traits can be identified to guide parental selection and stack haplotypes that may increase the rate of long-term genetic gain compared to utilising genome-wide GEBV. In essence, precise parental combinations and mating pathways can be identified to ensure long-term, consistent genetic gain in unfavourably correlated traits. This would be at odds with multi-trait selection in traditional genomic prediction frameworks, because multi-trait selection is expected to exhaust favourable genetic covariation between the traits, leaving only neutral or unfavourable genetic covariation (Itoh, 1991). This would make genetic progress for both traits increasingly more difficult over time (Itoh, 1991).

The ability to map haploblocks and identify specific haplotypes associated with traits provides a powerful framework for unravelling genetic architecture that can be used to advance crop breeding efforts (e.g., Varshney et al., 2021; Brunner et al., 2024; Tong et al., 2024; Akinlade et al., 2025; Alahmad et al., 2025; Roy et al., 2025; Vo Van-Zivkovic et al., 2025; Aldiss et al., 2025). These insights are essential for developing haplotype-based selection strategies that assemble beneficial allele combinations and optimise genotype configurations (Hayes et al., 2024), complementing gene networks and genotype-to-

528 phenotype systems approaches (Cooper et al., 2009; Kemper et al., 2012; Villiers et al.,
529 2024), and improving genomic predictions (Schopp et al., 2017) in crop breeding programs.

530 To summarize the benefits to QTL discovery and breeding programs, the haploblock
531 approach represents an innovative compromise between GWAS and genomic selection,
532 combining the strengths of both methodologies while addressing their respective limitations.
533 While GWAS provides precise identification of QTL underlying traits of interest, its
534 applications remain largely confined to pre-breeding activities with limited direct utility for
535 practical breeding programs. Conversely, genomic selection effectively estimates genome-
536 wide breeding values for direct use by breeders but offers limited insights into the underlying
537 genetic architecture of target traits and can severely impact diversity with intense selection.
538 The localGEBV method bridges this gap by simultaneously enabling the dissection of genetic
539 control mechanisms and the identification of genomically-informed selection units
540 (haploblocks) directly applicable in breeding programs. The flexibility to adjust block size
541 according to specific breeding objectives further enhances the practical utility of this
542 approach, allowing breeders to optimize the balance between genetic resolution and selection
543 efficiency. This dual functionality positions haploblocks as a powerful tool for translating
544 genomic discoveries into actionable breeding strategies.

545 Future research should explore extended genomic prediction models to capture non-additive
546 or epistatic interactions, leverage advancements in SNP arrays or whole genome sequencing
547 to improve haplotype phasing accurately, and explore the applications of localGEBV in
548 selection under extensive genotype-by-environment interactions and multi-trait selection.

549 In conclusion, we demonstrated the localGEBV methodology is a powerful tool for QTL
550 discovery. The localGEBV method was able to detect more known QTL with clearer signals
551 compared to other GWAS methods and alleviated or eliminated many of the limitations
552 associated with multiple-testing, background noise, dependence on prior marker distribution
553 (genetic architecture) assumptions, and defining regions of interest for candidate gene/QTL
554 discovery based on significant associations. The localGEBV method was able to locate the
555 largest effect QTL of interest with better prediction accuracy in a genomic selection context
556 and was able to identify other known QTL as well as several unknown QTL that FarmCPU
557 and BLINK could not identify. When utilised in a genomic prediction or QTL context, the
558 localGEBV method was robust to *a priori* assumptions of the genetic architecture, with near
559 equivalent phenotypic prediction accuracy from a large effect QTL using an infinitesimal

modelling approach or mixture approach to marker effects. This demonstrates that effects of large effect QTL can be re-captured from other markers in the haploblock under an infinitesimal model. We also demonstrated that utilising the localGEBV method with an infinitesimal model, like rrBLUP was able to capture additional polygenic signals not recognized in any other method. However, we did utilize methods to estimate marker effects that considered all markers jointly and applied shrinkage effects. We would recommend caution when extending this inference to methods that fit individual markers as fixed, because marker effects (and thus localGEBV) may be overestimated. Finally, through the analogy of GWAS to genomic selection, we provide clear examples of how the localGEBV method can be extended to genomic selection and is equally applicable to QTL discovery in livestock and crops.

Methods

Barley phenotypic and genotypic datasets

We analysed $N = 812$ barley accessions with complete genotypic and spike row morphology data acquired from the AGG, representing global diversity originating from 80 countries. Accessions were genotyped using the Illumina Infinium wheat-barley 40K XT SNP chip array (Keeble-Gagnère et al., 2021), yielding $M = 12{,}539$ SNP markers. Genotype calls were in 0/1/2 format (homozygous reference, heterozygous, homozygous alternate) relative to the cv Morex assembly IBSC v2.0.

Data filtering excluded unmapped markers and applied the following criteria using SelectionTools (Hofheinz and Frisch, 2014): >20% missing data, <1% expected heterozygosity, and <1% minor allele frequency. This resulted in $M = 9{,}050$ mapped markers for analysis. Accessions with >20% missing marker data were also excluded, leading to $N = 790$ accessions remaining.

Spike row descriptions obtained from AGG classified accessions as 2-rowed (373) or 6-rowed (384). For simplicity, we excluded 33 accessions that were described as ‘irregular’ row-type (Åberg and Wiebe, 1945), had missing information, or had ‘mixed’ designations. After exclusion of accessions with ambiguous designations, the dataset with row-type for QTL discovery analyses comprised $N = 759$ barley accessions. All analyses were performed using R version 3.6.3.

Population genetic structure and linkage disequilibrium analysis

We performed classical multidimensional scaling analysis of Roger's genetic distance using SelectionTools on the M = 9,050 mapped markers to infer population structure across the N = 790 barley accessions prior to filtering accessions for ambiguous row-type phenotypes. A *k-means* clustering algorithm categorised accessions into clusters, with the optimal number (*clusters* = 3) determined according to the majority of the criterion indices from the NbClust package (Charrad et al., 2014).

For linkage disequilibrium (LD) assessment, we calculated the squared pairwise correlations (r^2) between intrachromosomal SNP pairs. The LD decay was evaluated by analysing average r^2 between adjacent SNP as a function of physical distance, with positions aligned to the reference cv Morex assembly IBSC v2.0 as in the 40K XT SNP chip array.

QTL discovery approach

'Local' GEBV approach

The localGEBV approach consists of three key steps:

Step 1: Haploblock construction. Genome-wide markers were first organised into haploblocks using an LD-based approach, which provides insight into physical recombination events. Due to its reduced dependence on allele frequencies and independence from finite population size, we used the squared correlation coefficient, r^2, as the LD measure following methodologies described by Voss-Fels et al. (2019).

Given the LD characteristics of our barley population, we tested multiple blocking criteria: $r^2 \in \{0.1, 0.3, 0.5\}$ and marker tolerances, $tol \in \{0, 1, 2, 3\}$, resulting in 12 LD blocking parameter combinations. The algorithm scanned each barley chromosome independently to group SNP markers into blocks based on the LD threshold and tolerance parameters. To make blocks, the r^2 measurement between the last marker in a block (or the first marker in a new block) and the adjacent marker was evaluated against the specified threshold to determine whether the marker would be included in the haploblock. If the r^2 value was below the threshold, the block would either be terminated, or the tolerance counter would be incremented by one. The marker tolerance parameter specified how many markers could fail the threshold check before a block was terminated. Specifying a tolerance greater than zero allows the algorithm to handle issues in the data sequencing or genome alignment. A block may be composed of a single marker if it is not in LD with other markers. The algorithm continued until all markers were grouped into blocks.

Step 2. Estimation of SNP effects. We tested two models to estimate SNP effects for comparison: rrBLUP and BayesR. Following the approach described by Voss-Fels et al. (2019), the rrBLUP method was implemented in SelectionTools using:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Zu} + \boldsymbol{e}, \qquad [1]$$

where $\boldsymbol{y}$ is the observation vector of N genotypes for row-type trait, $\boldsymbol{X}$ is the $N \times p$ design matrix allocating N phenotypes to p fixed effects, $\boldsymbol{\beta}$ is the size p vector of fixed effect solutions, $\boldsymbol{Z}$ is the $N \times M$ design matrix allocating phenotypes to the random effects of M SNP markers, $\boldsymbol{u}$ is the vector of M SNP effects, and $\boldsymbol{e}$ is the random residual errors. It is assumed that $\boldsymbol{u} \sim N(\boldsymbol{0_M}, \boldsymbol{I_M}\sigma_u^2)$ and $\boldsymbol{e} \sim N(\boldsymbol{0_N}, \boldsymbol{I_N}\sigma_e^2)$; where $\boldsymbol{0_M}$, σ_u^2, σ_e^2 , and $\boldsymbol{I}$ are a vector of zeros of length M, the additive genetic marker variance under infinitesimal assumptions, the residual variance, and identity matrices of size M and N, respectively. The rrBLUP solution for the M SNP marker effects, $\hat{\boldsymbol{u}}$**,** can be solved as:

$$\hat{\boldsymbol{u}} = (\boldsymbol{Z'Z} + \lambda \boldsymbol{I_M})^{-1}\boldsymbol{Z'y}, \qquad [2]$$

where $\lambda = \frac{\sigma_e^2}{\sigma_u^2}$, estimated in the rrBLUP software package, represents the ratio between the residual and SNP marker variances.

BayesR (Erbe et al., 2012; Moser et al., 2015) was implemented using GCTB v2.0 (Genome-wide Complex Trait Bayesian analysis, https://cnsgenomics.com/software/gctb) following the approach described by Kemper et al. (2015). The model was fit following the model in Equation 1 and used the same $\boldsymbol{Z}$ design matrix. However, BayesR assumes *a priori* that SNP effects, $\boldsymbol{u}$, are drawn from a mixture of K normal distributions with mean 0 and variances σ_k^2, with mixture probabilities defined by vector $\boldsymbol{\gamma} = [\gamma_1, \ldots, \gamma_K]$. At each iteration of the Gibbs sampler, the effect of SNP m is updated by first computing its posterior probabilities of membership in each mixture, conditional on the current effects of all other SNP and mixture probabilities in $\boldsymbol{\gamma}$. A component assignment b_m is then sampled from the posterior probabilities of membership to each mixture, and the SNP estimate, $\hat{u}_m$, is drawn from the corresponding normal distribution. After all SNP have been updated, the mixture proportions, $\boldsymbol{\gamma}$, are resampled from a Dirichlet distribution, with $\boldsymbol{\gamma} \sim \text{Dirichlet}(\boldsymbol{\alpha} + \boldsymbol{\delta})$, where $\boldsymbol{\alpha}$ is the vector of prior counts and $\boldsymbol{\delta}$ are the counts of SNP assigned to each of the K mixtures in an iteration of the Gibb's sampler.

In our implementation, we set $K = 4$ mixture components corresponding to zero-mean normal distributions with variance $\sigma_k^2 = \pi_k \sigma_g^2$, where $\boldsymbol{\pi} = [0, 0.0001, 0.001, 0.01]$ and σ_g^2 is

651 the total additive genetic variance. The prior mixture probabilities were defined $\boldsymbol{\gamma} =$
652 $[0.95, 0.03, 0.01, 0.01]$ and the prior counts were defined $\boldsymbol{\alpha} = \mathbf{1}_{\boldsymbol{K}}$, a vector of ones of length
653 *K*. In other words, while the prior on an individual SNP effect distribution was informative,
654 the prior on the overall genome-wide allocation to each mixture was left uninformative and
655 the posterior was driven completely by the data. We ran 25,000 iterations of the Gibbs chain
656 after 5,000 burn-in iterations.

657 *Step 3. Estimation of localGEBV and variance*. Linear contrasts of the estimated SNP effects
658 were then specified to aggregate estimated SNP effects within each haploblock, yielding the
659 estimated localGEBV of each individual at each haploblock (Kemper et al., 2015) as follows:

$$\boldsymbol{localGEBV_j} = \boldsymbol{Z_j}\widehat{\boldsymbol{u}}_{\boldsymbol{j}}\,, \qquad [3]$$

660 where $\boldsymbol{Z_j} = \boldsymbol{Z_{m_r:m_R}}$ is the design matrix of the marker genotypes (0/1/2 format for all *N*
661 individuals) comprising haploblock *j* of *J* total haploblocks and $\widehat{\boldsymbol{u}}_{\boldsymbol{j}} = \widehat{\boldsymbol{u}}_{\boldsymbol{m_r:m_R}}$ denotes the
662 vector of estimated SNP effects spanning from the first marker (m_r) to the last marker (m_R) in
663 haploblock *j*. Estimated SNP effects were obtained from Equation 2 in the rrBLUP
664 implementation or from the posteriors of the BayesR implementation of Equation 1 to compare
665 the two methods.

666 Subsequently, the variance of the estimated localGEBV effects (haploblock variance) across
667 individuals in haploblock *j* was defined to be:

$$\sigma^2_{haploblock_j} = var\left(\boldsymbol{localGEBV_j}\right) \qquad [4]$$

668 where $\sigma^2_{haploblock_j}$ is the variance of the localGEBV at haploblock *j*, $\boldsymbol{localGEBV_j}$ is the
669 vector of *N* estimated localGEBV effects for all *N* individuals at block *j*, and the variance
670 represents the population variance calculation on the vector. The haploblock variance serves
671 as a metric for prioritising chromosome segments containing QTL. Haploblocks were
672 considered QTL if the variance surpasses the overall mean haploblock variance (Kemper et
673 al., 2015).

674 To assess whether high variance haploblocks were significantly different from zero, we
675 calculated probability values ($P_{invchisq}$) using a scaled inverse chi-squared distribution:

$$\sigma^2_{haploblock_j} \sim Scaled\ Inv\ \chi^2(\nu, \tau^2) \qquad [5]$$

676 where $\nu \in \{0.1, 0.5, 1, 2\}$ are the potential degrees of freedom and $\tau^2 \in \{0.1, 0.5, 1, 2\}$ are the
677 potential scaling parameter values (Han and Pan, 2010). The distribution of $\sigma^2_{haploblock_j}$ was

678 determined by first plotting the $\boldsymbol{\sigma}^2_{haploblock}$ vector of haploblock variances to determine the
679 best fit, and $\nu = 0.5$ and $\tau^2 = 0.5$ were employed in significance testing. Detection ability
680 was computed relative to number of informative haploblocks (Wei et al., 2009; Hayes, 2013;
681 Wang and Xu, 2019) as the ratio bP_t/B, where bP_t is the number of informative haploblocks
682 from $\boldsymbol{\sigma}^2_{haploblock}$ at a significance threshold of $P_t = 0.001$, and B is the total number of
683 haploblocks. Therefore, considering the 12 LD parameter combinations ($r^2 \in \{0.1, 0.3, 0.5\}$;
684 marker $tol \in \{0, 1, 2, 3\}$) and 2 SNP effect estimation procedures (rrBLUP and BayesR), a
685 total of 24 resulting outputs were analysed.

686 To explore the relationship of haploblock variance with size and haplotype effects, a
687 generalized additive linear regression model, regressing haploblock variance on the number
688 of markers in the haploblock or the haplotype effect was performed using the R package
689 mgcv (Wood, 2001). Each of the 24 constructed LD blocks was analysed independently.

690 *Comparative genome-wide association studies and co-location analysis*

691 For comparative GWAS, we used the same genotype matrix in Equation 1 in two models
692 implemented in Genome Association and Prediction Integrated Tool (GAPIT) v3 in R (Wang
693 and Zhang, 2021): Fixed and random model Circulating Probability Unification (FarmCPU;
694 Liu et al., 2016) and Bayesian-information and Linkage-disequilibrium Iteratively Nested
695 Keyway (BLINK; Huang et al., 2019). FarmCPU iteratively incorporates fixed and random
696 effects, whereas BLINK uses only fixed effects. Unlike the standard single-locus mixed
697 linear model, which tests markers one at a time in a one-dimensional genome scan, both
698 FarmCPU and BLINK perform multi-dimensional genome scans that estimate multiple
699 marker effects simultaneously while controlling for population structure.

700 To account for population stratification, we fitted 15 principal components (explaining ~80%
701 cumulative variance) as covariates. To control for false discovery rate, significant markers
702 were identified following the built-in *P* value thresholds ($-\log_{10} P_{adj} > -\log_{10} 0.008$) for
703 FarmCPU and $-\log_{10} P_{adj} > -\log_{10} 0.004$) for BLINK).

704 Significant SNP markers identified by FarmCPU and BLINK were then compared with
705 haploblocks showing significant haploblock variance from rrBLUP and BayesR with co-
706 location. Co-location was based on the physical position according to the reference cv Morex
707 assembly IBSC v2.0 described in the 40K SNP chip array. We cross-referenced results with
708 known genes and QTL associated with row phenotypes, inflorescence development, fertility

709 of spikelets, overall spike morphology, and agronomic and yield component traits
710 (Komatsuda et al., 2007; Ramsay et al., 2011; Koppolu et al., 2013; Bull, 2015; Youssef et al.,
711 2017; Van Esse et al., 2017; Bull et al., 2017; Koppolu et al., 2022).

712 **Assessment of detecting true haploblocks**

713 First, to summarise the estimates of SNP effects obtained from FarmCPU, BLINK, rrBLUP,
714 and BayesR, we calculated their effect contributions in the principal components using the
715 factoextra (Kassambara and Mundt, 2016) R package.

716 To evaluate the accuracy of detecting a true QTL or haploblocks, we cross-validated the
717 population using an 80:20 split (training set of 589; testing set of 148). Detection accuracy, r,
718 was defined as the correlation between the observed phenotypic values (y) and predicted
719 values ($\hat{y}$) in the testing set. Specifically, we tested the effects of the haploblock with the
720 highest variance identified by rrBLUP and BayesR to predict the row-type phenotype, along
721 with the most significant markers identified by both FarmCPU and BLINK, by fitting their
722 haplotype and marker effects as fixed effects in a LPM and in a logistic regression within a
723 GLM framework in R. In the LPM, estimation was performed by ordinary least squares,
724 assuming the residual was normally, independently, and identically distributed as an
725 approximation because of the binary nature of the row-type trait In the logistic regression in
726 the GLM framework, the same linear predictor from the LPM was linked to the binomial
727 distribution of the row-type trait via the logit link function, thereby mapping the predictor to
728 the probability scale.

729 For comparison, we repeated the analysis with random markers (i.e., non-significant) and
730 random haploblocks (i.e., with localGEBV variance below the threshold). Random selection
731 was repeated 100 times, and the mean correlation, r, and standard deviations (standard error
732 bars) of the 100 iterations were reported.

733 **Haplotype analysis**

734 To explore the two unique features of localGEBV (i.e., haploblock variance and haplotype
735 effects), we examined the specific haplotypes associated with the trait. As an illustrative
736 example, the haploblock with the highest variance was first subjected to statistical parsimony
737 haplotype network analysis (TCS method) implemented in POPART (Leigh et al., 2015).
738 Each haplotype, defined by a series of SNP alleles within the haploblock, was represented as
739 a node in the haplotype network. Using these haplotypes, we calculated pairwise similarity

740 scores based on the series of SNP alleles or haplotypes between accessions using FlapJack
741 (Milne et al., 2010) and subsequently performed complete linkage hierarchical clustering of
742 barley accessions. This enabled us to assess the relationship between haplotype similarity and
743 row-spike phenotypes.

Table 1. Summary statistics of high variance haploblocks for row-spike phenotype from localGEBV estimated using rrBLUP and BayesR, and significant markers identified using FarmCPU and BLINK. Only haploblocks with localGEBV variance exceeding $P_{invchisq} < 0.001$ and average variance from both rrBLUP and BayesR estimates, constructed following LD threshold $r^2 = 0.1$ and marker $tol = 3$ are shown.

Chr	Block ID	Flanking Marker[a]		Pos (Mbp)		Block Size[a]	var(localGEBV)*100 [$P_{invchisq}$]		-log10(*P*)	GAPIT Method	Known or candidate genes
		Start	End	Start	End		rrBLUP	BayesR			
1H	b000049	AVRIG00513	AVRIG00594	352.32	387.55	63	0.264 [7.2e-19]	0.333 [9.6e-15]			*VRS3*
	b000056	AVRIG00697	AVRIG00741	414.06	428.36	34		0.093 [6.4e-56]			
	b000091	AVRIG01071	AVRIG01107	516.80	520.21	32		0.112 [2.9e-46]			*TRD*
2H		AVRIG01687		44.79					7.3	BLINK	
	b000210	AVRIG02156	AVRIG02455	231.42	431.25	137	0.577 [4.0e-08]	0.263 [6.0e-19]			
	b000214	AVRIG02569	AVRIG00273	496.67	550.02	116	0.262 [5.5e-19]	0.111 [1.5e-46]			
	b000235	AVRIG02998	AVRIG03058	645.10	655.33	49	0.817 [1.5e-05]	1.325 [0.00292]			*VRS1*
		AVRIG03038		651.38					6.0	FarmCPU	
		AVRIG03040		651.53					6.9	BLINK	
		AVRIG03047		652.42					17.6; 10.5	FarmCPU, BLINK	
		1_1533		655.11					6.3; 5.8	FarmCPU, BLINK	
	b000251	AVRIG03181	AVRIG03190	685.79	687.83	5		0.121 [1.3e-42]			
	b000324	AVRIG03707		757.38		1		0.084 [2.0e-62]			
3H	b000382	AVRIG04257	AVRIG04489	169.96	269.30	157	0.212 [9.8e-24]	0.136 [6.6e-38]			*HORVU3Hr1G036930*
	b000383	AVRIG04490	AVRIG04708	269.30	371.18	148	0.115 [3.0e-45]				
	b000405	AVRIG05202	AVRIG05226	529.11	533.57	17	0.119 [1.9e-43]				
	b000411	AVRIG05287	AVRIG05299	548.57	550.36	9		0.192 [2.5e-26]			
4H	b000552	AVRIG06251	AVRIG06325	16.16	30.93	58		0.238 [5.0e-21]			*VRS5*
		AVRIG06428		51.58					7.05	FarmCPU	

	b000561	AVRIG06979	AVRIG06987	417.09	422.34	6		0.218 [4.8e-23]			
	b000636	AVRIG07929	AVRIG07949	644.34	645.50	14		0.160 [4.4e-32]			
5H	b000638	AVRIG07962	AVRIG08001	0.40	0.93	22		0.080 [7.4e-66]			
	b000650	AVRIG08140	AVRIG08644	60.21	35.95	364	0.113 [5.3e-46]	0.103 [9.8e-51]			
6H	b000933	AVRIG11948		575.77		1		0.076 [4.1e-69]			
7H	b001020	AVRIG12826	AVRIG13058	233.94	384.44	165	0.223 [1.4e-22]	0.139 [5.5e-37]			
	b001032	AVRIG13334	AVRIG13474	509.15	566.58	95		0.089 [8.5e-59]			*HORVU7Hr1G090180*

[a] Markers within the haplotype blocks are presented as flanking markers; significant markers identified in FarmCPU and BLINK are shown.

[b] Block size corresponds to the number of markers within the defined haploblocks

Funding

We thank the funding from the Grains Research and Development Corporation (UOQ2005-012RTX; Program 3. Minimising the impact of major barley foliar pathogens on yield and profit: Screening of diverse barley germplasm for rapid discovery and utilisation of novel disease resistance in barley using R-HapSelect – A haplotype-based toolkit).

Acknowledgements

We acknowledge Dr Sally Norton and the team at the Australian Grains Genebank for facilitating the acquisition of barley germplasm collection associated accession information; Dr Kerrie Forrest and the team at Agriculture Victoria for genotyping using the 40K Infinium XT SNP Chip; Dr Michael Groszmann at GRDC for discussions on the practical applications of localGEBV method. LTH was supported through an ARC Future Fellowship (FT220100350). WS and VP were supported as Adjunct Postdocs of the Australian Research Council Training Centre in Predictive Breeding for Agricultural Futures (IC230100016). The authors declare no conflict of interest.

Author contributions

Conceptualization, E.D., K.V., L.H., and B.H.; data analysis and writing – review and editing, W.S., E.D., V.P., and S.Y.; All authors have read and approved the final manuscript.

References

Abdellaoui, A., Yengo, L., Verweij, K. J. H., and Visscher, P. M. (2023). 15 years of GWAS discovery: Realizing the promise. *Am. J. Hum. Genet.* **110**:179–194.

Åberg, E., and Wiebe, G. A. (1945). Irregular barley, Hordeum irregulare, sp. nov. *J. Wash. Acad. Sci.* **35**:161–164.

Akinlade, O. J., Robinson, H., Kang, Y., Thudi, M., Samineni, S., Gaur, P., Smith, M. R., Voss-Fels, K. P., Costilla, R., Varshney, R. K., et al. (2025). A chickpea MAGIC population to dissect the genetics of complex traits. *Plant Genome* **18**:e70096.

Alahmad, S., Smith, D., Katsikis, C., Aldiss, Z., Brunner, S. M., Meer, S. V., Meijer, L., Heidariask, B., Chenu, K., and Chapman, S. (2025). Phenotyping the hidden half: Combining UAV phenotyping and machine learning to predict barley root traits in the field. *J. Exp. Bot.* Advance Access published 2025.

Aldiss, Z., Lam, Y., Baraibar, S., Van Der Meer, S., Dinglasan, E., Massel, K., Crisp, P., Godwin, I., Borrell, A., Moody, D., et al. (2025). Haplotype-based insights into seminal root angle in barley. *Plant Genome* **18**:e70088.

Andrade, A. C. B., Viana, J. M. S., Pereira, H. D., Pinto, V. B., and Fonseca E Silva, F. (2019). Linkage disequilibrium and haplotype block patterns in popcorn populations. *PLOS ONE* **14**:e0219417.

Barton, N. H. (2010). Mutation and the evolution of recombination. *Philos. Trans. R. Soc. B Biol. Sci.* **365**:1281–1294.

Berdan, E. L., Barton, N. H., Butlin, R., Charlesworth, B., Faria, R., Fragata, I., Gilbert, K. J., Jay, P., Kapun, M., Lotterhos, K. E., et al. (2023). How chromosomal inversions reorient the evolutionary process. *J. Evol. Biol.* **36**:1761–1782.

Bevan, M. W., Uauy, C., Wulff, B. B. H., Zhou, J., Krasileva, K., and Clark, M. D. (2017). Genomic innovation for crop improvement. *Nature* **543**:346–354.

Bohra, A., Kilian, B., Sivasankar, S., Caccamo, M., Mba, C., McCouch, S. R., and Varshney, R. K. (2022). Reap the crop wild relatives for breeding future crops. *Trends Biotechnol.* **40**:412–431.

Brunner, S. M., Dinglasan, E., Baraibar, S., Alahmad, S., Katsikis, C., Van Der Meer, S., Godoy, J., Moody, D., Smith, M., Hickey, L., et al. (2024). Characterizing stay-

green in barley across diverse environments: unveiling novel haplotypes. *Theor. Appl. Genet.* **137**:120.

Bull, H. J. (2015). Identification and Characterisation of the Barley Row-Type Gene VRS3 Advance Access published 2015.

Bull, H., Casao, M. C., Zwirek, M., Flavell, A. J., Thomas, W. T. B., Guo, W., Zhang, R., Rapazote-Flores, P., Kyriakidis, S., Russell, J., et al. (2017). Barley SIX-ROWED SPIKE3 encodes a putative Jumonji C-type H3K9me2/me3 demethylase that represses lateral spikelet fertility. *Nat. Commun.* **8**:936.

Bustos-Korts, D., Dawson, I. K., Russell, J., Tondelli, A., Guerra, D., Ferrandi, C., Strozzi, F., Nicolazzi, E. L., Molnar-Lang, M., Ozkan, H., et al. (2019). Exome sequences and multi-environment field trials elucidate the genetic basis of adaptation in barley. *Plant J.* **99**:1172–1191.

Charlesworth, B. (1974). Inversion polymorphism in a two-locus genetic system. *Genet. Res.* **23**:259–280.

Charrad, M., Ghazzali, N., Boiteau, V., and Niknafs, A. (2014). **NbClust** : An *R* Package for Determining the Relevant Number of Clusters in a Data Set. *J. Stat. Softw.* **61**.

Clark, A. G. (2004). The role of haplotypes in candidate gene studies. *Genet. Epidemiol.* **27**:321–333.

Cole, J. B., and VanRaden, P. M. (2011). Use of haplotypes to estimate Mendelian sampling effects and selection limits. *J. Anim. Breed. Genet.* **128**:446–455.

Comadran, J., Ramsay, L., MacKenzie, K., Hayes, P., Close, T. J., Muehlbauer, G., Stein, N., and Waugh, R. (2011). Patterns of polymorphism and linkage disequilibrium in cultivated barley. *Theor. Appl. Genet.* **122**:523–531.

Cooper, M., Van Eeuwijk, F. A., Hammer, G. L., Podlich, D. W., and Messina, C. (2009). Modeling QTL for complex traits: detection and context for plant breeding. *Curr. Opin. Plant Biol.* **12**:231–240.

Dering, C., Hemmelmann, C., Pugh, E., and Ziegler, A. (2011). Statistical analysis of rare sequence variants: an overview of collapsing methods. *Genet. Epidemiol.* **35**.

Difabachew, Y. F., Frisch, M., Langstroff, A. L., Stahl, A., Wittkop, B., Snowdon, R. J., Koch, M., Kirchhoff, M., Cselényi, L., Wolf, M., et al. (2023). Genomic prediction with haplotype blocks in wheat. *Front. Plant Sci.* **14**:1168547.

Epstein, R., Sajai, N., Zelkowski, M., Zhou, A., Robbins, K. R., and Pawlowski, W. P. (2023). Exploring impact of recombination landscapes on breeding outcomes. *Proc. Natl. Acad. Sci.* **120**:e2205785119.

Erbe, M., Hayes, B. J., Matukumalli, L. K., Goswami, S., Bowman, P. J., Reich, C. M., Mason, B. A., and Goddard, M. E. (2012). Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels. *J. Dairy Sci.* **95**:4114–4129.

Fan, B., Onteru, S. Garrick, D. J., Stalder, K. J., and Rothschild, M. F. (2011). Genome-Wide Association Study Identifies Loci for Body Composition and Structural Soundness Traits in Pigs. *PLoS ONE* **6**:e14726.

Gao, X., Starmer, J., and Martin, E. R. (2008). A multiple testing correction method for genetic association studies using correlated single nucleotide polymorphisms. *Genet. Epidemiol.* **32**:361–369.

Goddard, M. (2009). Genomic selection: prediction of accuracy and maximisation of long term response. *Genetica* **136**:245–257.

Han, F., and Pan, W. (2010). A Data-Adaptive Sum Test for Disease Association with Multiple Common or Rare Variants. *Hum. Hered.* **70**:42–54.

Hayes, B. (2013). Overview of Statistical Methods for Genome-Wide Association Studies (GWAS). In *Genome-Wide Association Studies and Genomic Prediction* (ed. Gondro, C.), Van Der Werf, J.), and Hayes, B.), pp. 149–169. Totowa, NJ: Humana Press.

Hayes, B. J., Mahony, T. J., Villiers, K., Warburton, C., Kemper, K. E., Dinglasan, E., Robinson, H., Powell, O., Voss-Fels, K., and Godwin, I. D. (2024). Potential approaches to create ultimate genotypes in crops and livestock. *Nat. Genet.* **56**:2310–2317.

Hofheinz, N., and Frisch, M. (2014). Heteroscedastic Ridge Regression Approaches for Genome-Wide Prediction With a Focus on Computational Efficiency and Accurate Effect Estimation. *G3 GenesGenomesGenetics* **4**:539–546.

Huang, M., Liu, X., Zhou, Y., Summers, R. M., and Zhang, Z. (2019). BLINK: a package for the next level of genome-wide association studies with both individuals and markers in the millions. *GigaScience* **8**:giy154.

Itoh, Y. (1991). Changes in genetic correlations by index selection. *Genet. Sel. Evol.* **23**:301–308.

Jiang, Y., Schmidt, R. H., and Reif, J. C. (2018). Haplotype-Based Genome-Wide Prediction Models Exploit Local Epistatic Interactions Among Markers. *G3 GenesGenomesGenetics* **8**:1687–1699.

Joukhadar, R., Daetwyler, H. D., Bansal, U. K., Gendall, A. R., and Hayden, M. J. (2017). Genetic Diversity, Population Structure and Ancestral Origin of Australian Wheat. *Front. Plant Sci.* **8**:2115.

Kassambara, A., and Mundt, F. (2016). Factoextra: extract and visualize the results of multivariate data analyses. *CRAN Contrib. Packag.* Advance Access published 2016.

Keeble-Gagnère, G., Pasam, R., Forrest, K. L., Wong, D., Robinson, H., Godoy, J., Rattey, A., Moody, D., Mullan, D., Walmsley, T., et al. (2021). Novel Design of Imputation-Enabled SNP Arrays for Breeding and Research Applications Supporting Multi-Species Hybridization. *Front. Plant Sci.* **12**:756877.

Kemper, K. E., Bowman, P. J., Pryce, J. E., Hayes, B. J., and Goddard, M. E. (2012). Long-term selection strategies for complex traits using high-density genetic markers. *J. Dairy Sci.* **95**:4646–4656.

Kemper, K. E., Reich, C. M., Bowman, P. J., Vander Jagt, C. J., Chamberlain, A. J., Mason, B. A., Hayes, B. J., and Goddard, M. E. (2015). Improved precision of QTL mapping using a nonlinear Bayesian method in a multi-breed population leads to greater accuracy of across-breed genomic predictions. *Genet. Sel. Evol.* **47**:29.

Kim, S. A., Cho, C.-S., Kim, S.-R., Bull, S. B., and Yoo, Y. J. (2018). A new haplotype block detection method for dense genome sequencing data based on interval graph modeling of clusters of highly correlated SNPs. *Bioinformatics* **34**:388–397.

Komatsuda, T., Pourkheirandish, M., He, C., Azhaguvel, P., Kanamori, H., Perovic, D., Stein, N., Graner, A., Wicker, T., Tagiri, A., et al. (2007). Six-rowed barley originated from a mutation in a homeodomain-leucine zipper I-class homeobox gene. *Proc. Natl. Acad. Sci.* **104**:1424–1429.

Koppolu, R., Anwar, N., Sakuma, S., Tagiri, A., Lundqvist, U., Pourkheirandish, M., Rutten, T., Seiler, C., Himmelbach, A., Ariyadasa, R., et al. (2013). *Six-rowed spike4* (*Vrs4*) controls spikelet determinacy and row-type in barley. *Proc. Natl. Acad. Sci.* **110**:13198–13203.

Koppolu, R., Jiang, G., Milner, S. G., Muqaddasi, Q. H., Rutten, T., Himmelbach, A., Guo, Y., Stein, N., Mascher, M., and Schnurbusch, T. (2022). The barley mutant multiflorus2.b reveals quantitative genetic variation for new spikelet architecture. *Theor. Appl. Genet.* **135**:571–590.

Kouyos, R. D., Silander, O. K., and Bonhoeffer, S. (2007). Epistasis between deleterious mutations and the evolution of recombination. *Trends Ecol. Evol.* **22**:308–315.

Leigh, J. W., Bryant, D., and Nakagawa, S. (2015). POPART: full-feature software for haplotype network construction. *Methods Ecol. Evol.* **6**.

Li, J., Wang, Z., Fernando, R., and Cheng, H. (2021). Tests of association based on genomic windows can lead to spurious associations when using genotype panels with heterogeneous SNP densities. *Genet. Sel. Evol.* **53**:45.

Liu, X., Huang, M., Fan, B., Buckler, E. S., and Zhang, Z. (2016). Iterative Usage of Fixed and Random Effect Models for Powerful and Efficient Genome-Wide Association Studies. *PLOS Genet.* **12**:e1005767.

Marchini, J., Cardon, L. R., Phillips, M. S., and Donnelly, P. (2004). The effects of human population structure on large genetic association studies. *Nat. Genet.* **36**:512–517.

Maynard, J., and Haigh, J. (2007). The hitch-hiking effect of a favourable gene. *Genet. Res.* **89**:391–403.

Meuwissen, T. H. E., and Goddard, M. E. (2000). Fine Mapping of Quantitative Trait Loci Using Linkage Disequilibria With Closely Linked Marker Loci. *Genetics* **155**:421–430.

Meuwissen, T. H. E., Hayes, B. J., and Goddard, M. E. (2001). Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. *Genetics* **157**:1819–1829.

Milne, I., Shaw, P., Stephen, G., Bayer, M., Cardle, L., Thomas, W. T. B., Flavell, A. J., and Marshall, D. (2010). Flapjack—graphical genotype visualization. *Bioinformatics* **26**:3133–3134.

Moser, G., Lee, S. H., Hayes, B. J., Goddard, M. E., Wray, N. R., and Visscher, P. M. (2015). Simultaneous Discovery, Estimation and Prediction Analysis of Complex Traits Using a Bayesian Mixture Model. *PLOS Genet.* **11**:e1004969.

Olasege, B. S., Porto-Neto, L. R., Tahir, M. S., Gouveia, G. C., Cánovas, A., Hayes, B. J., and Fortes, M. R. S. (2022). Correlation scan: identifying genomic regions that affect genetic correlations applied to fertility traits. *BMC Genomics* **23**:684.

Otto, S. P., and Barton, N. H. (1997). The Evolution of Recombination: Removing the Limits to Natural Selection. *Genetics* **147**:879–906.

Ramsay, L., Comadran, J., Druka, A., Marshall, D. F., Thomas, W. T. B., Macaulay, M., MacKenzie, K., Simpson, C., Fuller, J., Bonar, N., et al. (2011). INTERMEDIUM-C, a modifier of lateral spikelet fertility in barley, is an ortholog of the maize domestication gene TEOSINTE BRANCHED 1. *Nat. Genet.* **43**:169–172.

Roy, D., Dinglasan, E., Fowler, R., Platz, G., Lance, R., Synman, L., Franckowiak, J., Hickey, L. T., Voss-Fels, K., and Robinson, H. (2025). Genomic regions associated with spot blotch resistance in elite barley breeding populations. *Mol. Breed.* **45**:16.

Schopp, P., Müller, D., Wientjes, Y. C., and Melchinger, A. E. (2017). Genomic prediction within and across biparental families: means and variances of prediction accuracy and usefulness of deterministic equations. *G3 Genes Genomes Genet.* **7**:3571–3586.

Sullivan, P. F., Meadows, J. R. S., Gazal, S., Phan, B. N., Li, X., Genereux, D. P., Dong, M. X., Bianchi, M., Andrews, G., Sakthikumar, S., et al. (2023). Leveraging base-pair mammalian constraint to understand genetic variation and human disease. *Science* **380**:eabn2937.

Taliun, D., Gamper, J., and Pattaro, C. (2014). Efficient haplotype block recognition of very long and dense genetic sequences. *BMC Bioinformatics* **15**:10.

Tam, V., Patel, N., Turcotte, M., Bossé, Y., Paré, G., and Meyre, D. (2019). Benefits and limitations of genome-wide association studies. *Nat. Rev. Genet.* **20**:467–484.

Templeton, A. R., Crandall, K. A., and Sing, C. F. (1992). A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics* **132**:619–633.

946 **The Bovine HapMap Consortium, Gibbs, R. A., Taylor, J. F., Van Tassell, C. P.,**
947 **Barendse, W., Eversole, K. A., Gill, C. A., Green, R. D., Hamernik, D. L., Kappes,**
948 **S. M., et al.** (2009). Genome-Wide Survey of SNP Variation Uncovers the Genetic
949 Structure of Cattle Breeds. *Science* **324**:528–532.

950 **Tong, J., Tarekegn, Z. T., Jambuthenne, D., Alahmad, S., Periyannan, S., Hickey, L.,**
951 **Dinglasan, E., and Hayes, B.** (2024). Stacking beneficial haplotypes from the
952 Vavilov wheat collection to accelerate breeding for multiple disease resistance.
953 *Theor. Appl. Genet.* **137**:274.

954 **Van Den Berg, I., Hayes, B. J., Chamberlain, A. J., and Goddard, M. E.** (2019). Overlap
955 between eQTL and QTL associated with production traits and fertility in dairy cattle.
956 *BMC Genomics* **20**:291.

957 **Van Esse, G. W., Walla, A., Finke, A., Koornneef, M., Pecinka, A., and Von Korff, M.**
958 (2017). Six-Rowed Spike3 (VRS3) Is a Histone Demethylase That Controls Lateral
959 Spikelet Development in Barley. *Plant Physiol.* **174**:2397–2408.

960 **Varshney, R. K., Roorkiwal, M., Sun, S., Bajaj, P., Chitikineni, A., Thudi, M., Singh, N. P.,**
961 **Du, X., Upadhyaya, H. D., and Khan, A. W.** (2021). A chickpea genetic variation
962 map based on the sequencing of 3,366 genomes. *Nature* **599**:622–627.

963 **Villiers, K., Voss-Fels, K. P., Dinglasan, E., Jacobs, B., Hickey, L., and Hayes, B. J.**
964 (2024). Evolutionary computing to assemble standing genetic diversity and achieve
965 long-term genetic gain. *Plant Genome* **17**:e20467.

966 **Visscher, P. M., Wray, N. R., Zhang, Q., Sklar, P., McCarthy, M. I., Brown, M. A., and**
967 **Yang, J.** (2017). 10 Years of GWAS Discovery: Biology, Function, and Translation.
968 *Am. J. Hum. Genet.* **101**:5–22.

969 **Vo Van-Zivkovic, N., Dinglasan, E., Tong, J., Watt, C., Goody, J., Mullan, D., Hickey, L.,**
970 **and Robinson, H.** (2025). A large-scale multi-environment study dissecting adult-
971 plant resistance haplotypes for stripe rust resistance in Australian wheat breeding
972 populations. *Theor. Appl. Genet.* **138**:72.

973 **Voss-Fels, K. P., Stahl, A., Wittkop, B., Lichthardt, C., Nagler, S., Rose, T., Chen, T.-W.,**
974 **Zetzsche, H., Seddig, S., and Majid Baig, M.** (2019). Breeding improves wheat
975 productivity under contrasting agrochemical input levels. *Nat. Plants* **5**:706–714.

Wallace, J. G., Bradbury, P. J., Zhang, N., Gibon, Y., Stitt, M., and Buckler, E. S. (2014). Association Mapping across Numerous Traits Reveals Patterns of Functional Variation in Maize. *PLoS Genet.* **10**:e1004845.

Wang, M., and Xu, S. (2019). Statistical power in genome-wide association studies and quantitative trait locus mapping. *Heredity* **123**:287–306.

Wang, J., and Zhang, Z. (2021). GAPIT version 3: boosting power and accuracy for genomic association and prediction. *Genomics Proteomics Bioinformatics* **19**:629–640.

Wang, T., Chen, Y.-P. P., Bowman, P. J., Goddard, M. E., and Hayes, B. J. (2016). A hybrid expectation maximisation and MCMC sampling algorithm to implement Bayesian mixture model based genomic prediction and QTL mapping. *BMC Genomics* **17**:744.

Watanabe, K., Stringer, S., Frei, O., Umićević Mirkov, M., De Leeuw, C., Polderman, T. J. C., Van Der Sluis, S., Andreassen, O. A., Neale, B. M., and Posthuma, D. (2019). A global overview of pleiotropy and genetic architecture in complex traits. *Nat. Genet.* **51**:1339–1348.

Weber, S. E., Frisch, M., Snowdon, R. J., and Voss-Fels, K. P. (2023). Haplotype blocks for genomic prediction: a comparative evaluation in multiple crop datasets. *Front. Plant Sci.* **14**:1217589.

Wei, Z., Sun, W., Wang, K., and Hakonarson, H. (2009). Multiple testing in genome-wide association studies via hidden Markov models. *Bioinformatics* **25**:2802–2808.

Wolc, A., and Dekkers, J. C. M. (2022). Application of Bayesian genomic prediction methods to genome-wide association analyses. *Genet. Sel. Evol.* **54**:31.

Wood, S. N. (2001). mgcv: GAMs and generalized ridge regression for R. *R News* **1**:20–25.

Wray, N. R., Yang, J., Hayes, B. J., Price, A. L., Goddard, M. E., and Visscher, P. M. (2013). Pitfalls of predicting complex traits from SNPs. *Nat. Rev. Genet.* **14**:507–515.

Xiang, R., MacLeod, I. M., Daetwyler, H. D., De Jong, G., O'Connor, E., Schrooten, C., Chamberlain, A. J., and Goddard, M. E. (2021). Genome-wide fine-mapping identifies pleiotropic and functional variants that predict many traits across global cattle populations. *Nat. Commun.* **12**:860.

1005 **Yang, J., Benyamin, B., McEvoy, B. P., Gordon, S., Henders, A. K., Nyholt, D. R.,**
1006 **Madden, P. A., Heath, A. C., Martin, N. G., Montgomery, G. W., et al.** (2010).
1007 Common SNPs explain a large proportion of the heritability for human height. *Nat.*
1008 *Genet.* **42**:565–569.

1009 **Yang, J., Zaitlen, N. A., Goddard, M. E., Visscher, P. M., and Price, A. L.** (2014).
1010 Advantages and pitfalls in the application of mixed-model association methods. *Nat.*
1011 *Genet.* **46**:100–106.

1012 **Youssef, H. M., Eggert, K., Koppolu, R., Alqudah, A. M., Poursarebani, N., Fazeli, A.,**
1013 **Sakuma, S., Tagiri, A., Rutten, T., Govind, G., et al.** (2017). VRS2 regulates
1014 hormone-mediated inflorescence patterning in barley. *Nat. Genet.* **49**:157–161.

1015 **Zhang, Z., Ersoz, E., Lai, C.-Q., Todhunter, R. J., Tiwari, H. K., Gore, M. A., Bradbury, P.**
1016 **J., Yu, J., Arnett, D. K., and Ordovas, J. M.** (2010). Mixed linear model approach
1017 adapted for genome-wide association studies. *Nat. Genet.* **42**:355–360.

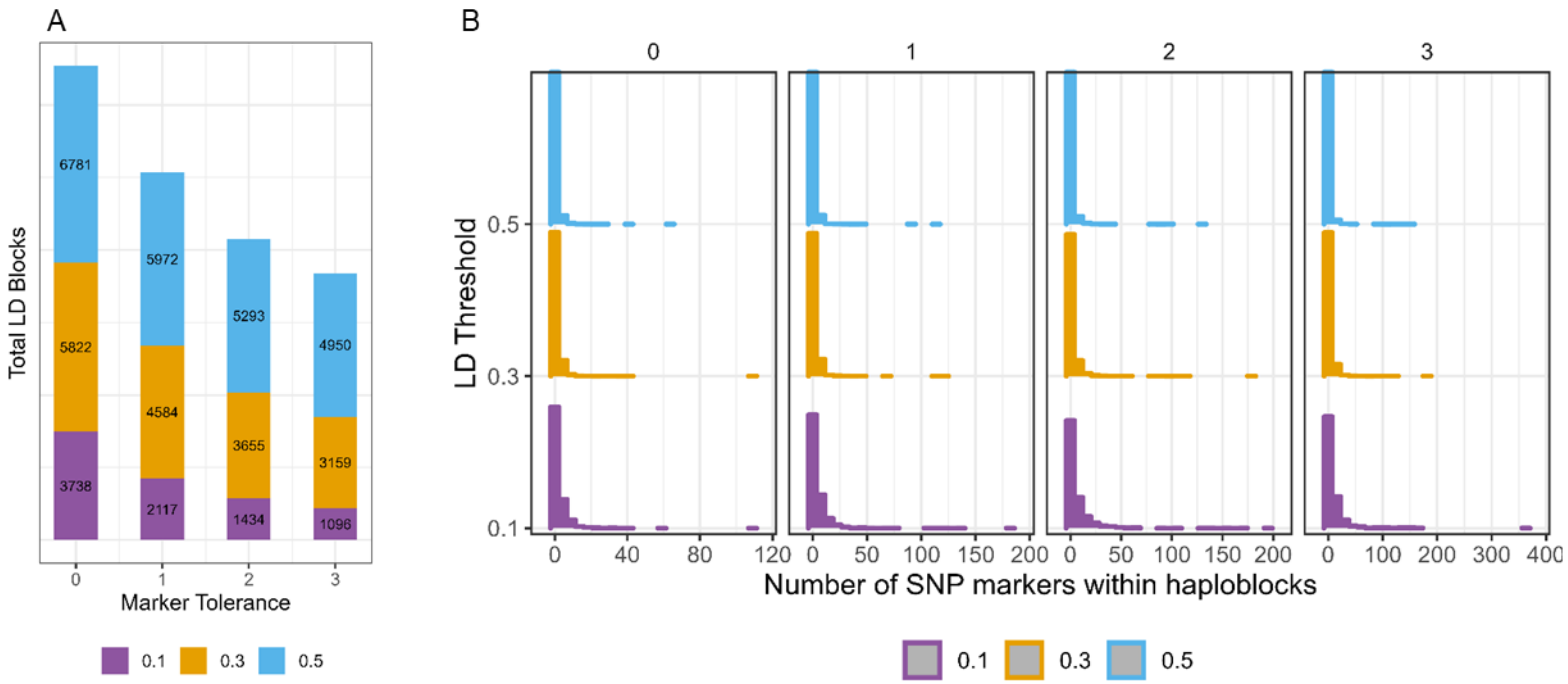
1018 **Zhang, Z., Wang, W., and Valdar, W.** (2014). Bayesian Modeling of Haplotype Effects in
1019 Multiparent Populations. *Genetics* **198**:139–156.

1020 **Zhao, Y., Qiang, C., Wang, X., Chen, Y., Deng, J., Jiang, C., Sun, X., Chen, H., Li, J.,**
1021 **Piao, W., et al.** (2019). New alleles for chlorophyll content and stay-green traits
1022 revealed by a genome wide association study in rice (Oryza sativa). *Sci. Rep.*
1023 **9**:2541.

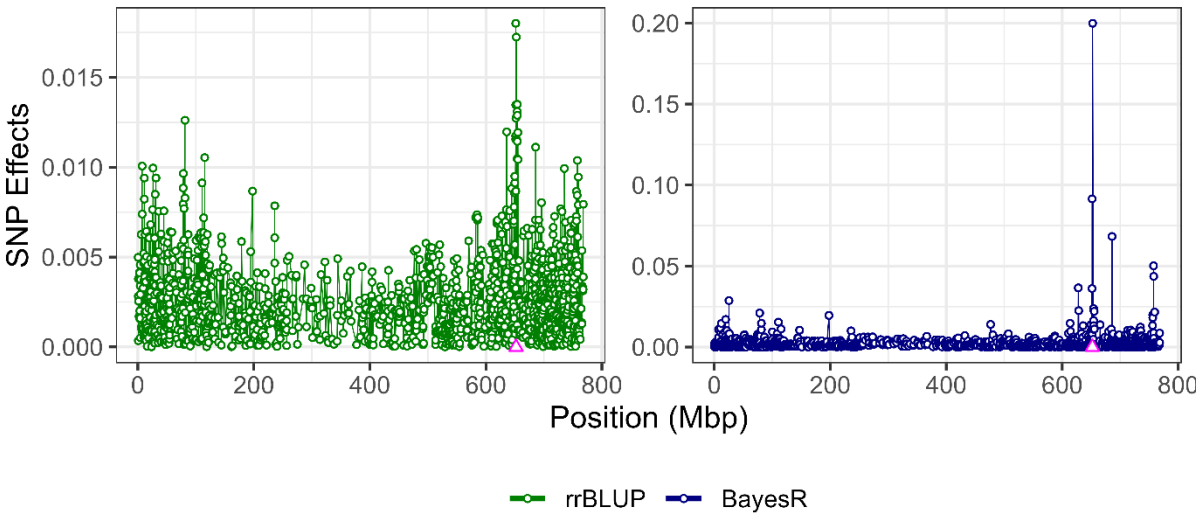
1024 **Zwirek, M., Waugh, R., and McKim, S. M.** (2019). Interaction between row-type genes in
1025 barley controls meristem determinacy and reveals novel routes to improved grain.
1026 *New Phytol.* **221**:1950–1965.

1027

Supplementary Files



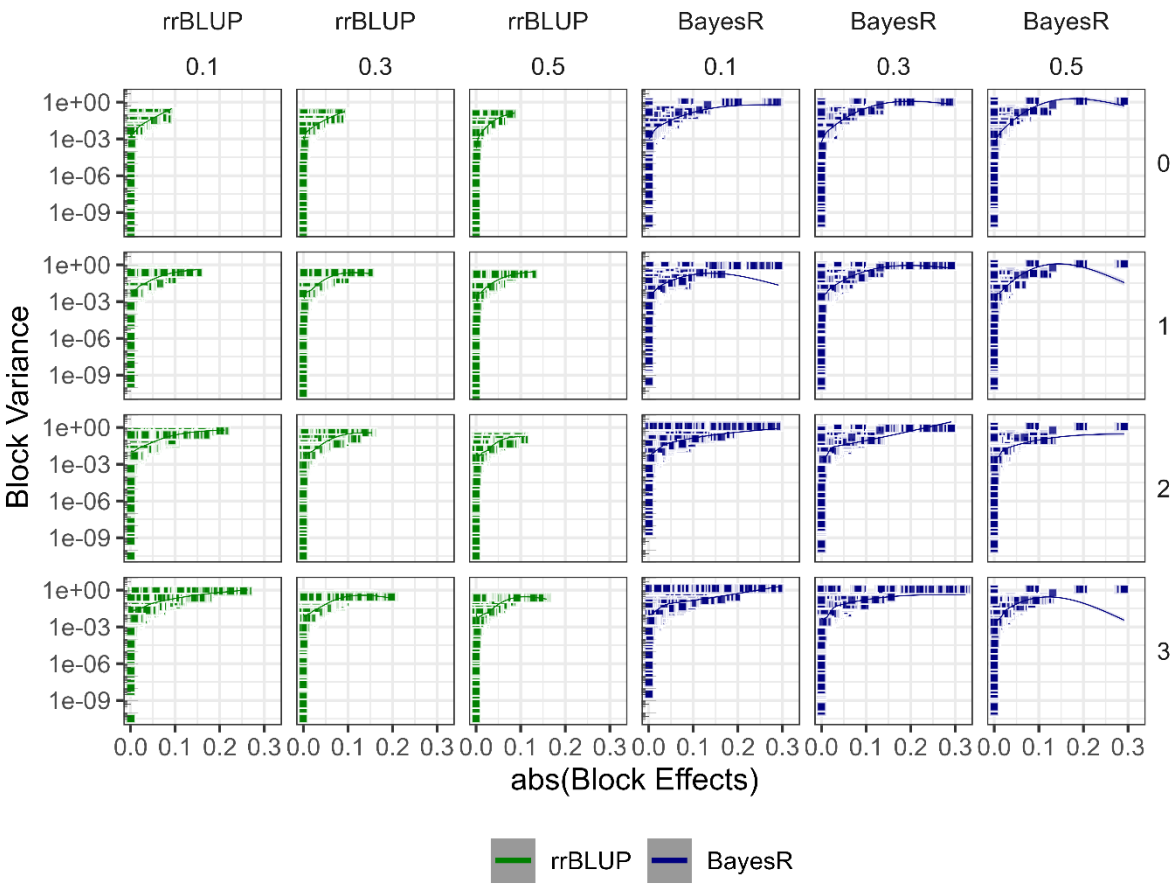
Supplementary Figure 1. LD block construction and its relationship with block size. SNP markers are grouped according to intrachromosomal pairwise LD based on the different LD-blocking parameters (LD threshold $r^2 \in \{0.1,0.3,0.5\}$; and marker $tol \in \{0,1,2,3\}$). (**A**) The number of total blocks for each combination of marker tolerance (x-axis) and LD threshold (colours and y-axis) are presented. (**B**) The distribution of the number of markers in each block for each combination of marker tolerance (x-axis) and LD threshold (colours and y-axis) are presented.



1035

1036 **Supplementary Figure 2. SNP effects estimated by rrBLUP and BayesR for row-type**
1037 **spike architecture in chromosome 2H.** Shown is the absolute value of SNP effect estimates,
1038 note the change of *y-axis* scale for rrBLUP and BayesR. The *VRS1* gene (marked as Δ) is
1039 located at 652.03 Mbp based on the XT SNP chip physical position using cv Morex assembly
1040 v2

1041



Supplementary Figure 3. Relationship between haploblock variance and haplotype effects using generalised additive model. Haploblock variance was determined from the localGEBV estimates using rrBLUP (*green*) and BayesR (*blue*); and plotted as min-max scaled of the $\log_{10}$ transformed variance for visual clarity. Haplotype effect is equivalent to localGEBV, which is the sum of SNP marker effects within haploblocks. In the figure, different LD parameters are shown as facets: *column facet* represented by LD threshold at $r^2 \in \{0.1,0.3,0.5\}$; and *row facet* as marker $tol \in \{0,1,2,3\}$.