Mol Breeding (2023) 43:14
https://doi.org/10.1007/s11032-023-01357-5

GWAS and genomic prediction for pre-harvest sprouting tolerance involving sprouting score and two other related traits in spring wheat

Manoj Kumar · Sachin Kumar · Karansher Singh Sandhu · Neeraj Kumar · Gautam Saripalli · Ram Prakash · Akash Nambardar · Hemant Sharma · Tinku Gautam · Harindra Singh Balyan · Pushpendra Kumar Gupta

Received: 15 June 2022 / Accepted: 26 January 2023 / Published online: 20 February 2023

Abstract In wheat, a genome-wide association study (GWAS) and genomic prediction (GP) analysis were conducted for pre-harvest sprouting (PHS) tolerance and two of its related traits. For this purpose, an association panel of 190 accessions was phenotyped for PHS (using sprouting score), falling number, and grain color over two years and genotyped with 9904 DArTseq based SNP markers. GWAS for main-effect quantitative trait nucleotides (M-QTNs) using three different models (CMLM, SUPER, and FarmCPU) and epistatic QTNs (E-QTNs) using PLINK were performed. A total of 171 M-QTNs (CMLM, 47; SUPER, 70; FarmCPU, 54) for all three traits, and 15 E-QTNs involved in 20 first-order epistatic interactions were identified. Some of the above QTNs overlapped the previously reported QTLs, MTAs, and cloned genes, allowing delineating 26 PHS-responsive genomic regions that spread over 16 wheat chromosomes. As many as 20 definitive and stable QTNs were considered important for use in marker-assisted recurrent selection (MARS). The gene, *TaPHS1*, for PHS tolerance (PHST) associated with one of the QTNs was also validated using the KASP assay. Some of the M-QTNs were shown to have a key role in the abscisic acid pathway involved in PHST. Genomic prediction accuracies (based on the cross-validation approach) using three different models ranged from 0.41 to 0.55, which are comparable to the results of previous studies. In summary, the results of the present study improved our understanding of the genetic architecture of PHST and its related traits in wheat and provided novel genomic resources for wheat breeding based on MARS and GP.

Keywords Wheat · Pre-harvest sprouting tolerance · Quantitative trait nucleotides · Genome-wide association study · Genomic perdition · Candidate genes

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11032-023-01357-5.

M. Kumar · S. Kumar (✉) · G. Saripalli · R. Prakash · A. Nambardar · H. Sharma · T. Gautam · H. S. Balyan · P. K. Gupta
Department of Genetics and Plant Breeding, Chaudhary Charan Singh University, Meerut, UP, India
e-mail: sachinkpsingh@gmail.com

K. S. Sandhu
Bayer Crop Sciences, Chesterfield, MO 63017, USA

N. Kumar
Department of Plant and Environmental Sciences, Clemson University, Clemson, SC, USA

G. Saripalli
Department of Plant Science and Landscape Architecture, University of Maryland, College Park, MD, USA

Introduction

Wheat (*Triticum aestivum* L.) is an important cereal crop and is widely grown in almost every continent

around the globe. Approximately 2.5 billion people in ~90 countries rely on wheat grain as their staple food (http://wheat.org). In addition to biotic stresses, a range of abiotic stresses including pre-harvest sprouting (PHS) adversely affect wheat grain yield and quality, thus limiting farmer's income. Globally, the annual losses due to PHS alone have been estimated to be US$ one billion (see review by Ali et al. 2019). As is widely known, PHS involves sprouting of grains in the mature wheat spikes, while still attached to the mother plant (Nyachiro 2012); this occurs, when there is high humidity or prolonged rainfall in fields during physiological maturity of the crop.

In the past, PHS tolerance (PHST) has generally been evaluated using one or more of the following traits: sprouting index and/or sprouting score (SS), grain color (GC), falling number (FN), germination index, seed dormancy and alpha-amylase activity (Rasul et al. 2009; Knox et al. 2012; Kumar et al. 2015; Martinez et al. 2018; for reviews, see Ali et al. 2019 and Gupta et al. 2020). Among these traits, SS and FN have been shown to be more important (DePauw et al. 2012; Martinez et al. 2018). It has also been recognized that the coat color of wheat grain is associated with tolerance to PHS, such that the wheat cultivars with red GC are generally more resistant to PHS relative to white-grained wheats (Himi et al. 2002; Lin et al. 2016; Zhou et al. 2017a, b). However, white grained-wheat genotypes with tolerance to PHS have been reported, both in natural populations as well as in pre-bred lines developed using molecular markers (Warner et al. 2000; Himi et al. 2002; Shao et al. 2018; Gautam et al. 2021).

PHST is an ordinal trait because no continuous scale (required for quantitative traits) is available for recording data; a scale of 1–5 or 1–10 is often used for recording the SS data (DePauw et al. 2012). The genetic studies on PHST have suggested that the trait is polygenic in nature (Moore et al. 2017; Liu et al. 2021; for recent reviews see Gupta et al. 2020; Tai et al. 2021). Many studies involving either linkage-based interval mapping or linkage disequilibrium (LD) based GWAS have already been conducted for PHST in wheat. As a result, at least 250 QTLs through interval mapping and an equal number of MTAs using GWAS have already been reported (Singh et al. 2021) with many more QTLs/MTAs reported during the past two years; these QTLs/MTAs are distributed on all 21 wheat chromosomes. More than half a dozen genes for PHS tolerance have also been cloned and characterized, which include the following: *TaMFT*, *TaMKK3-A*, *TaVp1*, *TaQsdr1*, *TaDOG1*, *TaSdr, TaPHS1, TaMF,* and *Tamyb10* (reviewed in Ali et al. 2019; Gupta et al. 2020; Tai et al. 2021). Meta-QTL analyses for this trait have also been undertaken, which included the latest study conducted by Tai et al. (2021), where 188 valid QTLs from 40 studies and 28 characterized genes were utilized leading to identification of 66 meta-QTL distributed on all 21 wheat chromosomes. The other studies on meta-QTLs for PHST included studies from our own laboratory (Tyagi and Gupta 2012; Tyagi et al. 2015).

For traits like PHST that are each controlled by hundreds of genes, genomic selection (GS), first proposed more than two decades ago, has also been recommended as an alternative approach (Meuwissen et al. 2001). GS makes use of a training population for estimation of genomic estimated breeding values (GEBVs), which are then used for selection of superior individual plants in a segregating population, without phenotyping (Meuwissen et al. 2001; Sandhu et al. 2021b, c). Moore et al. (2017) studied the potential of GS for predicting PHST in wheat by using different values of fixed effects, marker density; the prediction accuracy in their study was 0.49–0.62.

Development of PHS tolerant wheat cultivars is a useful strategy to deal with the problem of PHS. However, marker-assisted selection (MAS) for such polygenic traits each controlled by hundreds of QTLs (each QTL with a small effect), may not be very effective (Bernardo 2016), although examples of some success are available (Kumar et al. 2010; Gautam et al. 2021). Therefore, while still working on the genetics of the trait using interval mapping and GWAS, one should also explore the possibilities of using genomic prediction (GP) for developing PHS tolerant wheat cultivars. GP associated with GWAS has also been attempted for traits such as Septoria tritici blotch (STB), grain protein percent stability, agronomic, physiological, and grain quality traits (Odilbekov et al. 2019; Sandhu et al. 2021a; Alemu et al. 2021).

Despite the availability of an enormous amount of literature on the genetics of PHST, every new study on the genetics of PHST leads to identification of a few novel QTLs/MTAs, suggesting that the genetics of PHST has not been fully worked out and that there

is always a scope for discovery of new loci. Keeping this in view, the present GWAS was planned, where a new association panel comprising 190 spring wheat accessions was used. Three different models (CMLM, SUPER, and FarmCPU) for detection of main-effect QTNs and PLINK for detection of epistatic QTNs, were employed; of these models, FarmCPU also has the merit of providing multi-locus analysis with built in facility for addressing the problem of multiple testing. Apparently, GWAS along with GP based on three traits, (SS, FN, and GC) has not been undertaken so far for PHST. As expected, the present study resolved many novel and/or common genomic regions involved in PHST. Novel, epistatic interactions were also detected.

Materials and methods

Genetic materials

A panel of 190 diverse accessions (with similar phenology) was utilized that was selected from a worldwide collection of 330 spring wheats (a Spring Wheat Reference Set, described as SWRS) procured from CIMMYT, Mexico (Table S1). In our own laboratory, the SWRS or its subsets have earlier been used for GWAS for drought tolerance, grain quality, grain morphology, yield and yield related traits (Kumar et al. 2018; Gahlaut et al. 2019, 2021; Malik et al. 2021, 2022).

A set of 17,937 polymorphic SNP markers generated on the above accessions using DArTseq at Diversity Array Technology Pvt. Ltd., Australia, under the "Seed for Discovery" project of CIMMYT, Mexico was used. Of these as many as 8637 SNPs were physically mapped on all 21 wheat chromosomes (Table S2); the remaining 9300 SNPs could not be assigned to any wheat chromosome and were designated as chromosome unknown (ChrUn). Among the physically mapped SNPs, 2973 SNPs belonged to the A sub-genome, 4505 belonged to the B sub-genome, and 1159 belonged to the D sub-genome. DArTseq SNPs with 20% missing values (3782), ≤ 5% minor allele frequency (2729), and wheat accessions with more than 10% missing calls (1522) were excluded from whole SNP dataset. Finally, a total of 9904 (mapped + unmapped) high quality SNPs were used in GWAS analysis (Table S2). However, after eliminating SNPs with ≤ 5% minor allele frequency, as many as 15,208 polymorphic SNPs were utilized for GP.

Field trials and phenotyping for three PHS-related traits

The field trials were conducted in alpha-lattice design with two replications at Agriculture Research Farm, Chaudhary Charan Singh University, Meerut (28.984644°N and 77.705956°E), India, over two consecutive years (2018 and 2019; referred to as E1 and E2, respectively). Each experimental line was seeded in a 2 m plot comprising three rows 20 cm apart. Space between plots was 50 cm.

The sprouting score (SS) and two other related traits (FN and GC) were used for recording phenotyping data using the following procedures, (*i*) *Sprouting Score* (*SS*): five random spikes of each genotype per replication were hand-harvested from the field at physiological maturity (when 50% glumes in a plot turned yellow) and allowed to dry (after-ripening) at room temperature for 5 days, avoiding direct sunlight or high-temperature. Spikes were then stored at –20 °C to maintain dormancy (Mares 1983) until tests were conducted (within 1–3 months). Once spikes of all the accessions were collected and stored, spike-wetting tests were performed in an artificial rain simulator (mist chamber). The spikes were misted for 30 min every 6 h for seven days. The rare spikes infected with mold, if any, were discarded. Data on SS on each spike of individual accessions were recorded on a scale of 1 to 9 (Fig. S1), where a value of 1 represented no visible sprouting and a value of 9 represented complete sprouting (modified after McMaster and Derera 1976). (*ii*) *Falling Number* (*FN*): ten random spikes of each accession per replication were hand-harvested at physiological maturity and were artificially weathered for 48 h to enhance the activity of alpha-amylase in sampled spikes. The weathered spikes were dried up to 12% moisture and the seeds were threshed gently. Dried grain samples were ground to flour using FOSS Labtech CT410 Sample Mill and mixed with water to form slurry following ICC standard No. 107/1 (1995) and the AACC Method 56–81.03A (1999). The slurry was transferred in a glass test tube and then placed in a water bath with boiling water (100 °C) to gelatinize the starch. The FN apparatus

monitors a plunger (or nozzle) as it falls through the flour slurry in the test tube and calculates the time in seconds (s) that plunger takes to move to the bottom of the test tube. Thick (viscous) slurry takes longer time for the plunger to move down to the bottom of the test tube. Therefore, by evaluating the difference in thickness between flour slurry made from weathered and non-weathered grains, FN indirectly determines alpha-amylase activity. This method quantifies the variation in FN among wheat accessions. (*iii*) *Grain color* (*GC*): seeds from matured spikes were threshed by hand and stored at room temperature till further use to estimate seed coat color. To determine the GC, Petri plates were first washed by liquid detergent, rinsed thoroughly with tap water, and then air-dried. A solution was prepared by dissolving 25 g of NaOH in 500 ml (5% w/v) of distilled water in a volumetric flask and then stored in a polyethylene bottle. Each Petri plate was filled with 10 ml 5% NaOH solution. Approximately 20 seeds for each accession were soaked in a Petri plate containing 5% NaOH solution for 45 min at room temperature. Data on coat color of NaOH soaked seeds was recorded visually on a scale of 1 to 5 (1 = white, 2 = amber, 3 = light red, 4 = medium red and 5 = dark red) (Fig. S2).

Statistical analysis of the phenotypic traits

Phenotypic data were recorded on SS, FN, and GC for two years (E1 and E2). The data of E1 and E2 was separately pooled for three independent traits, and pooled data were denoted as PE. The data of E1, E2, and PE were separately used for all the statistical analyses, including GWAS.

The SPSS v16.0 software was used for calculating mean, median, range, standard deviation (SD), and coefficient of variation (CV) for each of the three traits. Pearson's correlation coefficient, frequency distribution and scatter plots were depicted in a correlation chart generated using Performance Analytics Package in R (https://cran.r-project.org/web/packages/Performance.pdf). Best linear unbiased estimates (BLUEs) of two replicates under each environment and across environments were estimated using META-R (Alvarado et al. 2015) and the values were used for GWAS analysis. Heritability was estimated separately for each environment using the R package lme4 (Bates et al. 2015) treating genotypes as a random effect with the model equation: $H^2 = \sigma^2_g / (\sigma^2_g + \sigma^2_e)$, where H^2 is the broad-sense heritability, σ^2_g is the genotypic variance, and σ^2_e is the residual variance (Bates et al. 2015).

Marker-trait associations (MTAs)

Single locus main-effect MTAs MTAs for each of the three traits were worked out using each of the following three models: (i) Compressed Mixed Linear Model (CMLM; Zhang et al. 2010), (ii) Settlement of MLM Under Progressively Exclusive Relationship (SUPER; Wang et al. 2014), and (iii) Fixed and random model Circulating Probability Unification (FarmCPU; Liu et al. 2016). All three models were implemented in R using GAPIT software package (Lipka et al. 2012). The CMLM and SUPER allowed single-locus analysis while the FarmCPU allowed multi-locus analysis.

The problem of population structure was addressed using principal component analysis (PCA), where the first three components of PCA were used as covariate in FarmCPU. The FarmCPU eliminates confounding issues arising due to population structure (Q matrix), kinship (K matrix), and multiple testing, etc. The Q and K matrices are automatically generated (VanRaden 2008; Lipka et al. 2012) using genotypic data with the help of a default set of parameters. Bonferroni-correction for eliminating false positives due to multiple testing is built-in with FarmCPU and SNPs with $p \leq 0.001$ $\{-\log10(p) \geq 3\}$ were declared significantly associated markers. For the remaining two single-locus models (CMLM and SUPER) also, the above threshold *p*-value was used to declare significant MTAs with the confidence that false positives will not be identified in these two models also. The MTAs identified following the multi-locus model FarmCPU are described as main effect quantitative trait nucleotides (M-QTNs) as is the usual practice; however, for convenience, the MTAs identified by CMLM, and SUPER are also described as M-QTNs during the present study. Circular Manhattan plots were generated using the QQMAN R package in R 3.4.1 (https://www.r-project.org/). The chromosome map showing the physical positions of QTNs associated with PHST for comparing the positions between interval mapping and GWAS was constructed using MapChart v2.2 (Voorrips 2002).

Two-locus epistatic interactions Pair-wise epistatic interactions were analyzed using PLINK v1.07 (Purcell et al. 2007). Significant interactions were filtered at p-value $< 1 \times 10^{-8}$ (Purcell et al. 2007; Jan et al. 2019). The QTNs involved in epistatic interactions were described as E-QTNs.

Comparison of QTNs with known MTAs/QTLs/cloned genes

Sequence tags of QTNs identified in the present study were subjected to BLASTN against the latest release of IWGSC RefSeq v2.1 (EnsemblPlants database) to obtain their physical coordinates on the chromosomes. Physical coordinates of the previously known MTAs/QTLs/genes related to PHST were also obtained in a similar manner. The sequences of simple sequence repeats (SSRs) were retrieved from GrainGenes website (https://wheat.pw.usda.gov/cgi-bin/GG3/browse.cgi?class=marker) and those for DArT markers were retrieved from Diversity Array Technology website (https://www.diversityarrays.com/technology-and-resources/sequences/). Using the chromosomal coordinates, the QTNs identified in the present study and the previously reported MTAs/QTLs/cloned genes associated with PHS tolerance related traits were projected using MapChart v2.3 (Voorrips 2002). The linkage disequilibrium (LD) decay in wheat genome has been reported at an average distance of 5 mega base pair (Mbp) (Ladejobi et al. 2019). Therefore, any two or more loci (including QTNs detected in the present and/or previously known QTLs/MTAs/cloned genes) were considered co-localized if they were mapped within the 5 Mbp physical region.

Prediction of candidate genes (CGs) underlying the QTNs

Sequence tags of SNP markers associated with the three different traits were subjected to BLASTN analysis against Chinese Spring reference genome sequences version 2.1 (IWGSC RefSeq v2.1) available at EnsemblPlants sequence database (http://www.ensembl.org/info/docs/tools/vep/index.html). The physical coordinates of each SNP tag were used as input in EnsemblPlants database, and the chromosomal regions were each extended to –300 kb and +300 kb around the positions of the QTN. This 600 kb region of a chromosome was searched for potential CGs involved in the processes of seed germination and dormancy. The information for proteins encoded by the genes was obtained from the interpro description using the Biomart tool available at EnsemblPlants database. The annotations of identified CGs were verified from the published reports to determine their putative roles in controlling the targeted PHS traits.

Kompetitive allele-specific PCR (KASP) assay

For KASP assays, 20 prioritized QTNs were selected. For each QTN, two allele-specific forward primers and one common reverse primer were designed using PolyMarker (Ramirez-Gonzalez et al. 2015). The primer sequences are given in Table S3. KASP assays were setup in StepOnePlus™ Real-Time PCR System (Applied BioSystems); KASP reaction mix with a final volume of 10 µl containing 50 ng/µl DNA, and 0.17 µm KASP assay mix (allele-specific primers and common primer) were used for PCR amplification. The profile for KASP assays included the following cycles: 95 °C for 15 min, followed by a 10-cycle touchdown of 94 °C for 20 s and 65–57 °C (dropping 0.8 °C per cycle) for 2 min, and 35 cycles of 94 °C for 20 s and 57 °C for 1 min. The fluorescence detection of the amplified products was analyzed using the allelic discrimination plot function.

Genomic prediction models

Three GP models (rrBLUP, Bayes B and Bayes LASSO) were used to predict the significant values of SS, FN, and GC using a cross-validation approach for each environment. These three models use different assumptions during model training and prediction (Crossa et al. 2017). The BLUE values for the 190 genotypes were used to train the models, which were genotyped using 15,208 DArT-seq SNP markers; prediction accuracies are reported as Pearson's correlations observed between the true observations and the predicted GEBVs. Results were reported separately for each of the two environments (E1 and E2).

Five-fold cross-validation was used for assessing the performance of each model, where 80% of the data was used for model training, and 20% of the data for model testing. Two hundred replications were performed to obtain the average prediction accuracy of the model. Each replication consisted of five iterations; during each

iteration, the dataset was split into five groups, and each group was used as a testing set for each iteration (Sandhu et al. 2021d). Due to the computational burden of these GP models, the whole analysis was performed over cloud computing using the high-speed computing platform available at Washington State University, Pullman, USA (https://hpc.wsu.edu/). Following are the details for each of three GP models used in the present study:

(i) Ridge regression best linear unbiased predictor (rrBLUP): rrBLUP is the most commonly used GP model due to its low computational demand, equivalent to traditional BLUPs in the context of mixed models, and estimation of all marker effects using the shrinkage approach (Rutkoski et al. 2016; Sandhu et al. 2021c). rrBLUP assumes homogenous variances for each marker and shrinks them equally toward the zero. GEBVs were calculated by estimating the genome-wide marker effects using the R package rrBLUP (Endelman 2011; R core team 2020). The following mixed model was used for calculation

$$y = \mu + Zu + e,$$

Where y is the N × 1 for the BLUEs for the associated phenotypic value from the 190 genotypes for each environment, μ is the overall mean, Z is the incidence matrix of dimensions N x M, linking each marker to the associated genotype, μ is the vector of normally distributed random marker effects, as $\mu \sim N(0, I\sigma^2_u)$ where I is the identity matrix, σ^2_u is the variance for marker effects, and e is the residual error with $e \sim N(0, I\sigma^2_e)$. Variance components were estimated using the restricted maximum likelihood function of the rrBLUP package (Kang et al. 2008).

(ii) Bayes B: The assumption of rrBLUP that all the marker effects are evenly distributed across the genome is not satisfactory for some traits. Therefore, Meuwissen et al. (2001) proposed a Bayes B model to relax the assumptions of rrBLUP. This model assumes that some markers have no effect on a particular trait and should not be included for model training (Pérez and de los Campos 2014). The model assumes the marker specific shrinkage for estimating their effects with a unique variance for each marker and can be represented as follows:

$$g(x_i) = \sum_{k=1}^{P} x_{ik}\beta_k\gamma_k$$

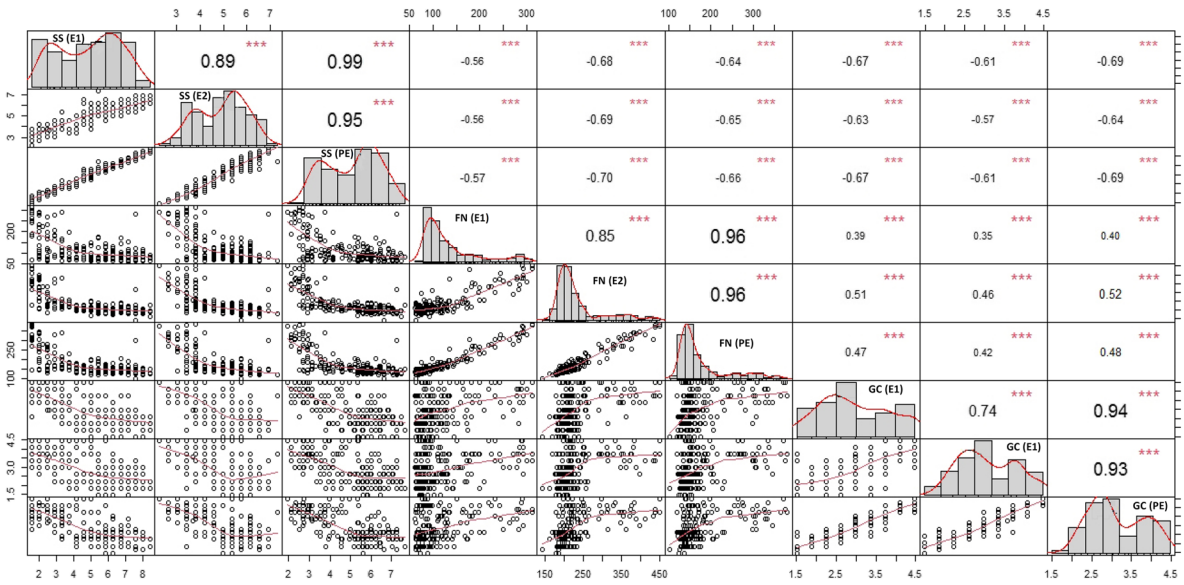


Fig. 1 A diagrammatic representation of frequency distributions (middle diagonal, top left to bottom right), correlations (upper panel), and relationships among three traits [sprouting score (SS), falling number (FN) and grain color (GC)] across environments (E1 = Meerut 2018; E2 = Meerut 2019 and PE = pooled environments) (lower panel). The *X*-axis and *Y*-axis indicate values of each trait. ***, **, and * indicates the significant levels at $p \leq 0.001$, 0.01, and 0.05, respectively

Where $g(x_i)$ is the sum of p marker effects, x_{ik} is the value of kth marker in the ith individual, β_k is the kth marker effect, and γ_k is the indicator variable associating the presence or absence of a particular marker in the model. The model assumes a mixture distribution with mass at zero, and the prior distribution for the variance is represented as

$$\mathrm{Var}(\beta_k) = 0 \text{ (with probability } \pi), \text{ and } \mathrm{Var}(\beta_k) \sim \chi^{-2}(v, s) \text{ (with probability } 1 - \pi)$$

The Bayes B can be reduced to Bayes A by setting $\pi = 0$ in the above model. The prior distribution for the $\chi^{-2}(v, s)$ is the inverse chi-squared distribution χ^{-2}, where v is the degree of freedom and s is the scaling parameter. The complete analysis for Bayes B was performed in the BGLR package implemented in R using cross-validation approaches using prior specified for each model (Pérez and de los Campos 2014). A total of 40,000 iterations were performed, with 10,000 excluded as burn-in.

(iii) Bayes LASSO: The Bayes LASSO (least absolute shrinkage and selection operator) is related to rrBLUP, but it shrinks marker effects strongly relative to rrBLUP with some of the marker coefficients driven to zero, creating a spare model (Tibshirani 1996). The Bayes LASSO assigns a double exponential prior distribution to all the marker effects, with a centre at zero and marker's specific variance (de los Campos et al. 2009). The likelihood function of the model used for Bayes LASSO can be written as follows:

$$f(y|\mu, x, m, \sigma^2) \sim N(\mu + xm, \sigma^2 I),$$

Where y is the vector for phenotype, μ is the overall mean, x is the design matrix for assessing markers to the genotypes, m is the vector of marker effects, $N(\mu + xm, \sigma^2 I)$ denotes the model distribution with mean $\mu + xm$ and variance $\sigma^2 I$. The prior distribution for estimating marker effects were double exponential. The whole analysis was performed using the BGLR package with 40,000 iterations (including 10,000 burn-in iterations) as reported in Sandhu et al. (2021d).

Results

Descriptive statistics

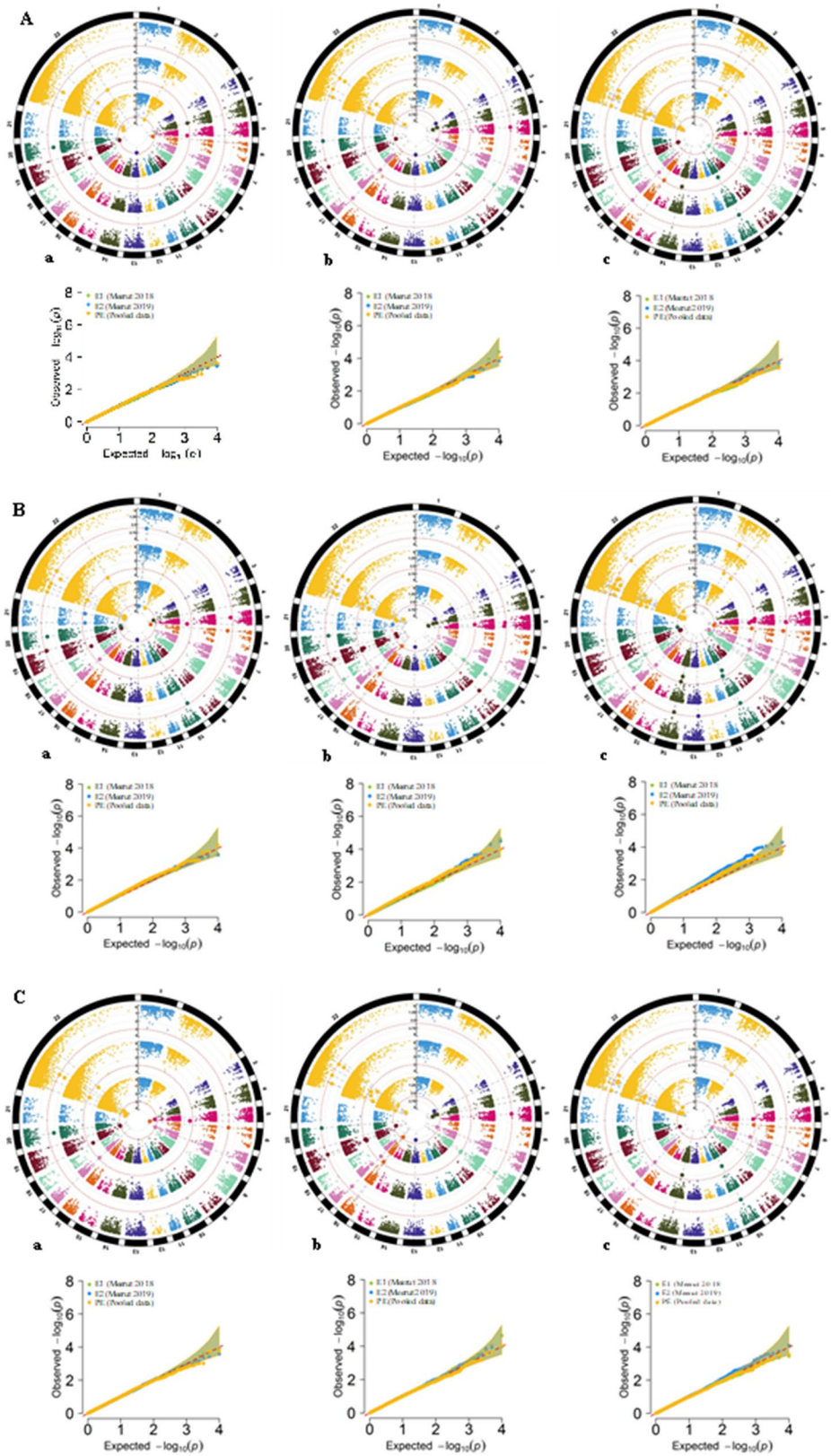
The values of means, range, coefficients of variation (CVs, %) and broad-sense heritability (H^2) for the three traits in E1, E2, and PE (pooled data over environments) are presented in Tables S4. The frequency distribution for the SS, FN, and GC was continuous and was similar in pattern in E1, E2, and PE. However, the distribution patterns for the individual traits differed. For SS and GC the distribution was largely bimodal, whereas for the FN it was skewed toward low FN.

Correlation coefficients among the SS, FN, and GC

The details of 36 possible Pearson's correlation coefficients among values of three traits (SS, FN, and GC) each in E1, E2, and PE are presented in Fig. 1; the correlations were generally significant and ranged from 0.74 to 0.99. Among the three traits, significant negative but moderately high correlations were noted between SS and FN (r = –0.56 to –0.70) and between SS and GC (r = –0.57 to –0.69); the correlation between FN and GC were positive but only moderately significant (r = 0.35 to 0.52).

Identification of QTNs for three PHS-related traits using GWAS

PCA revealed first three PCs explaining 6.82, 3.82, and 3.33% variation (total variation of 13.97%; Fig. S3). Utilizing these three PCs, the data is presented in three separate two dimensional plots showing no more than two sub-groups of the genotypes (Fig. S3). Based on the LOD threshold ($-\log10(p) \geq 3$), a total of 171 significant QTNs for three PHS-related traits (SS, FN, and GC) were identified using the phenotypic data of E1, E2, and PE. This included 47 QTNs due to CMLM, 70 QTNs due to SUPER, and 54 QTNs due to FarmCPU (Table S5). Manhattan plots along with Q-Q plots of


A
B
C
a
b
c
E1 (Maruti 2018)
E2 (Maruti 2019)
PE (Pooled data)
Observed $-\log_{10}(p)$
Expected $-\log_{10}(p)$

◂**Fig. 2** Circular Manhattan plots obtained from CMLM (**A**), SUPER (**B**), FarmCPU (**C**) for the significant associations of 9904 SNPs with three pre-harvest sprouting tolerance traits (SS, FN, GC) distributed on 21 wheat chromosomes and chromosome unknown (shown by numbers 1–22). In each circular plot, inner, middle and outer plots are representing E1 (Meerut 2018), E2 (Meerut 2019), and PE (pooled) environments, respectively. The LOD threshold value—log10(p) ≥ 3 is indicated as red-colored dotted circle. For each case, multi-track Q-Q plots are shown below the circular Manhattan plots

the GWAS results for CMLM, SUPER, and FarmCPU are presented in Fig. 2A–C.

A comparison of single locus main-effect QTNs (M-QTNs) for three PHS-related traits and three GWAS models suggest that 86 unique M-QTNs were detected in different combinations (individually and jointly) of PHS-related traits as well as GWAS models (Table 1); 32 M-QTNs were detected by each of the three models (Table 1 and Table S5). The physical positions of only 76 M-QTNs (located on 19 wheat chromosomes; leaving out 4B and 4D) could be determined using IWGSC RefSeq v2.1 (Fig. 5); the remaining 10 M-QTNs could not be assigned to any chromosome and are, therefore, described as ChrUn (chromosome unknown) (Table S5). The A sub-genome harbored maximum number of M-QTNs (34), followed by B (25) and D (17) sub-genomes. A total of 11 M-QTNs were each found to be associated with more than one trait in one or more environments. Among these, nine M-QTNs {M8876 (3B), M2654 (5A), M8514 (2B), M11378 (ChrUn), M10484 (5A), M12953 (ChrUn), M2207 (7A), M11590 (1A), and M9692 (6B)} were found to be associated with both FN and GC. The remaining two M-QTNs, namely M6844 (3B) and M10913 (1B), were associated with both SS and GC (Table 1 and Table S5).

Epistatic interactions among the QTNs

First order (two-locus) epistatic interactions included 20 pairs of significant ($P < 10^{-8}$) QTN x QTN interactions (SNP-SNP interactions), which included three pairs for SS, seven pairs for FN, and ten pairs for GC; five pairs were common for FN and GC, thus resulting in 15 unique SNP pairs (Fig. 3 and Table S6). None of these epistatic QTNs (E-QTN) involved M-QTNs. E-QTNs were distributed on 16 of 21 wheat chromosomes (excluding 2D, 4D, 5A, 5B, and 6D. The chromosome 7B carried two pairs of interacting E-QTNs [SNPs M3515 (626.5 Mb) and M6976 (10.1 Mb) for FN and GC, and interaction between SNPs M5158 (509.5) and M11451 (650.6 Mb) involved only GC]. Another SNP locus M11451 on 7B was involved in an interaction with SNP M5538 (486.1 Mb) on chromosome 7D for only GC (Fig. 3 and Table S6).

QTNs colocalized with previously reported QTLs/MTAs/cloned genes within 5 Mbp genomic regions

The above 76 M-QTNs and 30 E-QTNs were physically mapped along with the 26 known QTLs (based on interval mapping), 51 MTAs (based on GWAS), and 19 known genes for PHST (Table S7). As many as 29 M-QTNs and the two E-QTNs were mapped with the previously reported nine QTLs, 24 MTAs, and five known genes within 26 genomic regions (GRs) distributed on 16 wheat chromosomes (Fig. 4). Since, LD decays at an average distance of ~5 Mbp in wheat (Ladejobi et al. 2019), we assume that PHS-associated loci within each of these 26 GRs are in LD. We like to call these GRs PHS responsive GRs (PRGRs 1–26). These physically mapped and previously reported loci belong to the following eight PHS-related traits: (i) SS/sprouting index (ii) germination index, (iii) germination ratio, (iv) percent germination, (v) FN, (vi) GC, (vii) ABA responsiveness, and (viii) late maturity alpha amylase (LMA). Five of the M-QTNs associated with SS, FN, and GC were co-localized with one PHS tolerance gene each within the five individual PRGRs (Table 2). The remaining 14 reported genes were mapped outside the 26 PRGRs.

Two E-QTNs were also co-located with known QTLs/MTAs. E-QTN M6934 was co-localized with two MTAs (IWB13172 and IWB64868) for SS within the PRGR5 on chromosome 1B. Similarly, E-QTN M3464 was co-localized with one QTL (associated with *Xgwm637*) and an MTA (associated with IWB23723) each for SS within the PRGR15 on chromosome 4A (Fig. 4).

Novel QTNs for three PHS-related traits

As many as 47 M-QTNs and 28 E-QTNs detected during the present study were novel, because these were not co-localized with any of the previously reported QTLs/MTAs/genes. Among these novel QTNs, 12 QTNs (9 M-QTNs and 3 E-QTNs) for FN,

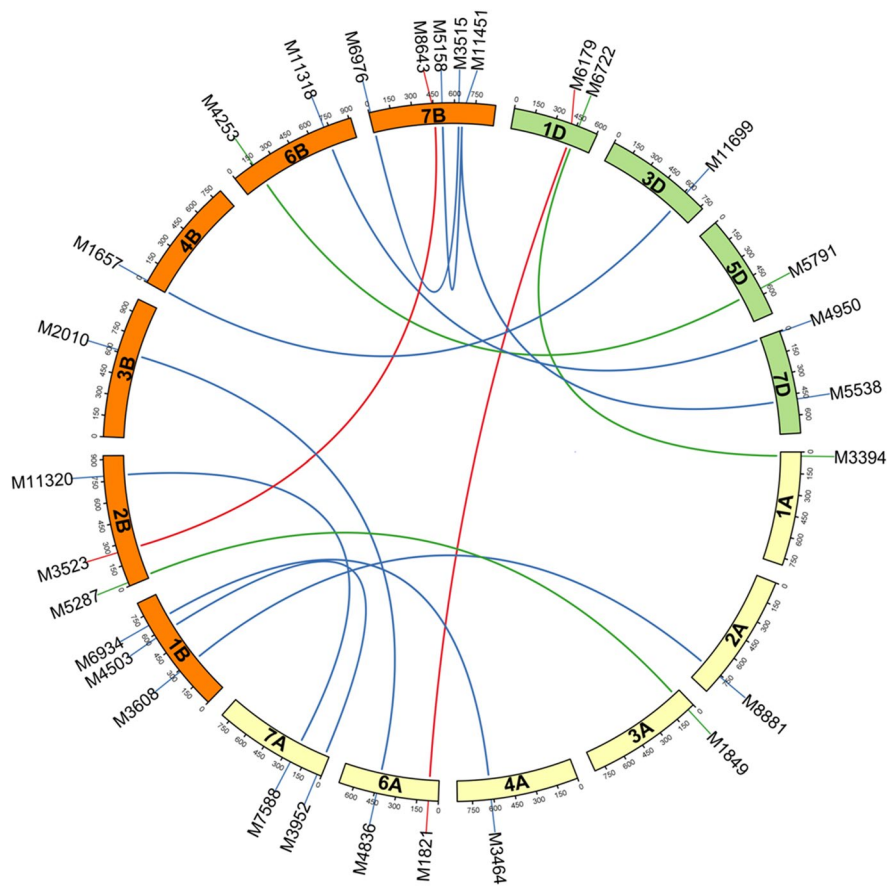


Fig. 3 Epistatic interaction among three traits, namely sprouting score (SS), falling number (FN), and grain color (GC) of pre-harvesting sprouting tolerance. The names of chromosome are shown within the circle. The SNPs involved in epistatic interactions on respective chromosomes are shown outside the circle and their positions (Mb) are given along the chromosomes. Inner curved color lines indicate SNP–SNP interactions. The interactions for SS, FN, and GC are represented by green, red, and blue curved lines, respectively

Table 1 Summary of unique main-effect quantitative trait nucleotides (M-QTNs) identified for three parameters of pre-harvest sprouting tolerance using single- and multi-locus GWAS models

Trait	GWAS method							Total
	CMLM	SUPER	FarmCPU	CMLM + SUPER	CMLM + FarmCPU	SUPER + FarmCPU	CMLM + SUPER + FarmCPU	
SS	1	9	2	4	1	3	6	26*
FN	0	13	4	0	1	5	6	29*
GC	4	0	0	1	3	3	9	20*
SS + GC	0	0	0	0	0	0	2	2
FN + GC	0	0	0	0	0	0	9	9
Total	5	22	6	5	5	11	32	86

GWAS, genome-wide association study; *SS*, sprouting score; *FN*, falling number; *GC*, grain color; *CMLM*, compressed mixed linear model; *SUPER*, settlement of MLM under progressively exclusive relationship; *FarmCPU*, fixed and random model circulating probability unification. *Does not include QTNs detected for more than one trait

SS, GC, and FN + GC were mapped in pairs within six 5 Mbp regions (excluding the PRGRs) on six different chromosomes (1B, 1D, 2A, 2B, 6D, and 7A) of wheat (Fig. 5). For example, M-QTN M5740 was 2.1 Mb away from the E-QTN M6179; each significantly associated with FN on chromosome 1D.

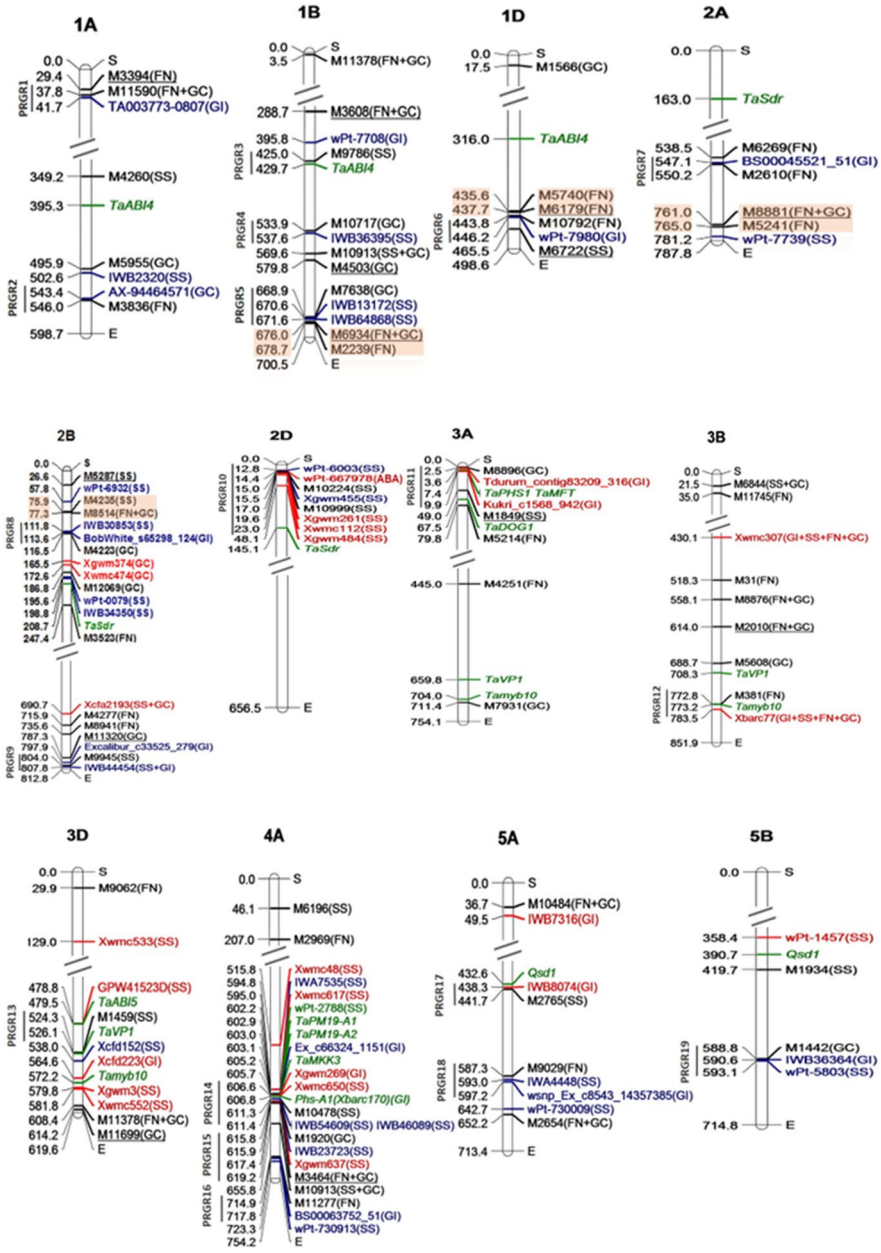


Fig. 4 Distribution of main-effect QTNs (M-QTNs) and epistatic QTNs (E-QTNs) for three pre-harvest sprouting (PHS)-related traits with the previously reported QTLs, MTAs, and cloned genes within 26 PHS responsive genomic regions (PRGR1-26) denoted by the thick black vertical line on the left of each chromosome. The markers name is shown on the right and their physical positions (Mbp; mega base pair) are on the left of each chromosome. For each mapped locus, the associated trait is given in parenthesis as follows: SS for sprouting score, FN for falling number, GC for grain color, GI for germination index, and LMA for late maturity amylase. The QTNs detected during the present study and those reported in the previous studies are shown in different colors as follows: (i) M-QTNs detected in the present study (simple black); (ii) previously published QTLs (red), MTAs (blue), specific known genes (green) and E-QTNs are underlined. The QTNs detected during the present study and mapped in pairs within a 5 Mbp region (excluding the PRGRs) are highlighted in light orange color

Identification of CGs associated with QTNs

A set of 176 PHS-related CGs were detected, which were associated with 52 M-QTNs (out of 76 M-QTNs) and included 14 M-QTNs for SS, 20 M-QTNs for FN, 15 M-QTNs for GC, and 3 M-QTNs for FN + GC. No CGs relevant to PHST were identified for the remaining 24 M-QTNs. Detailed information of CGs and their functional annotations are presented in Table S8. These 176 CGs encoded proteins that contained 33 different types of domains (Fig. 5). Some of the important domains include the following: (i) leucine-rich repeat (LRR) superfamily, (ii) NAC domain superfamily, (iii) serine/threonine protein kinase, (iv) F-box domain, (v) WRKY domain, (vi) APETALA2 (AP2)/EREB (ethylene responsive element binding factor) domain, (vii) MADS-box transcription

Fig. 4 (continued)

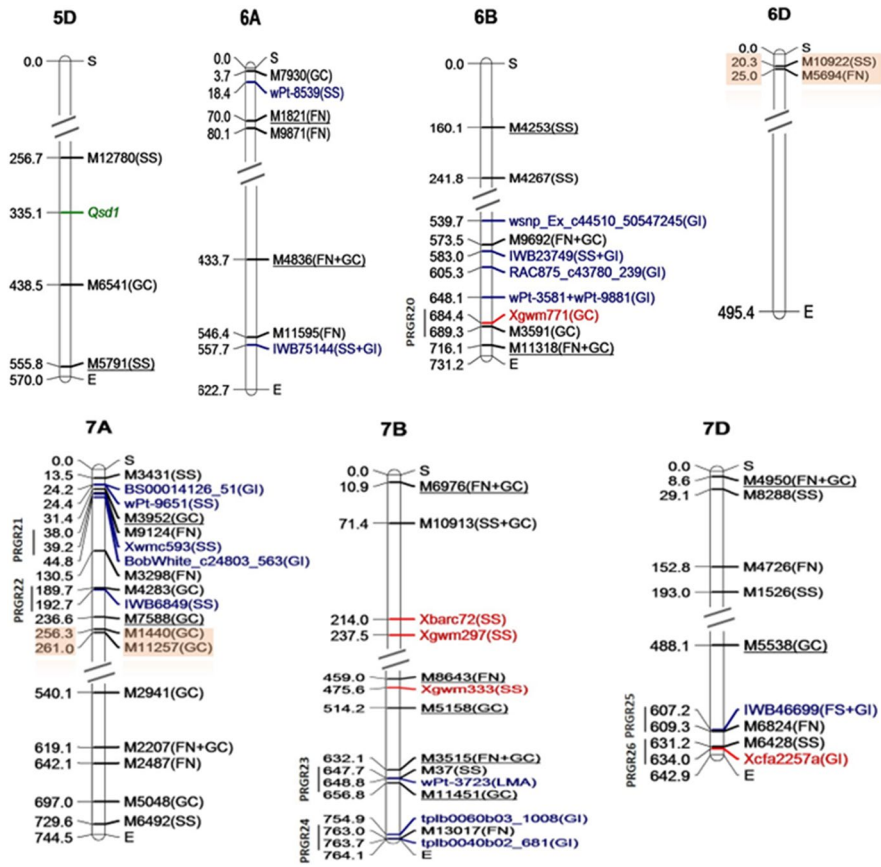


Table 2 Main-effect QTNs (M-QTNs) for the three parameters and five cloned genes for pre-harvest sprouting tolerance co-localized within five PRGRs (Fig. 4)

PRGR number (chromosome); parameter	M-QTN (position, Mbp)	Cloned gene (starting position, Mbp)	Reference
3 (1B); SS	M9786 (425.0)	*TaABI4* (429.7)	Xiao et al. 2021
13 (3D); SS	M1459 (524.3)	*TaVP1* (526.1)	Dale et al. 2017
14 (4A); SS	M10478 (611.3)	*Phs-A1* (606.8)	Shorinola et al. 2017
12 (3B); FN	M381 (772.8)	*Tamyb10* (773.2)	Zhou et al. 2017a, b
11 (3A); GC	M8896 (2.5)	*TaPHS1*/*TaMFT* (7.4)	Nakamura et al. 2011

SS, sprouting score; *FN*, falling number; *GC*, grain color; *M-QTN*, main effect-quantitative trait nucleotide; *PRGRs*, PHS responsive genomic regions

factors, (viii) aspartic peptidase A1 family, etc. A total of 19 of the 33 protein/domain types were related to the ABA signaling pathway. A summary of the M-QTNs for three different traits and associated CGs encoding two major classes of proteins containing two different domains is presented in Table 3. A set of 43 CGs underlying 18 M-QTNs (6 each for SS and FN, 3 for GC, and 3 for FN + GC) located on 11 different wheat chromosomes encoded proteins that contained LRRs (Table 3). Similarly, 27 CGs associated with 13 M-QTN (3 for SS, 7 for FN, 2 for GC, and 1 for FN + GC) on nine different chromosomes encoded proteins that contained serine/threonine-protein kinase domain. A few M-QTN regions contained only one or a few important genes, which encoded proteins with

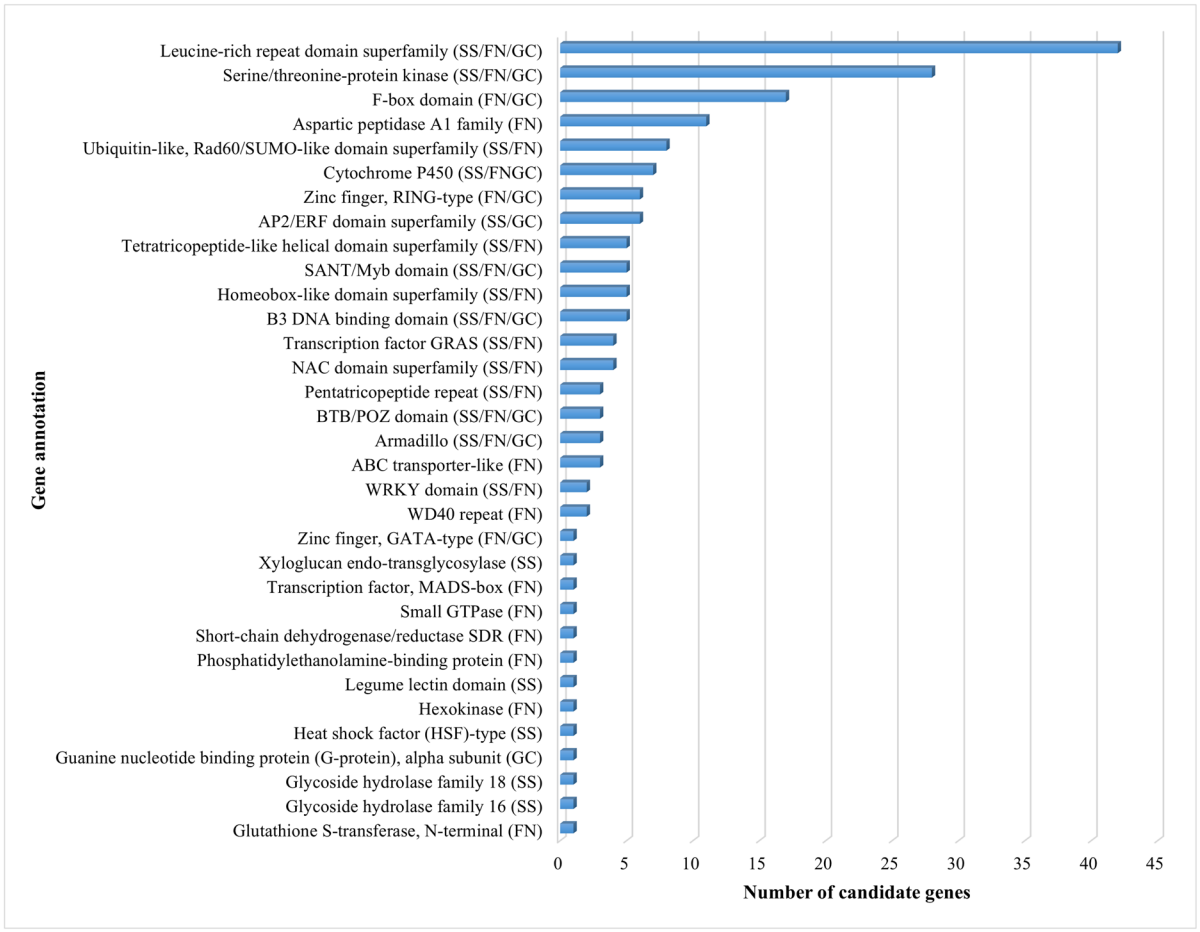


Fig. 5 Frequencies of candidate genes (with functions related to pre-harvest sprouting tolerance) encoding proteins with each of the 33 domains. In parenthesis, SS, FN, and GC refer to sprouting score, falling number, and grain color, respectively

unique domains involved in grain germination and dormancy (Table S8).

KASP assay for 10 prioritized QTNs

KASP assays in 96 wheat accessions were conducted for 10 M-QTNs prioritized for MARS. However, nine of these KASP assays failed to discriminate alternative SNP alleles (Fig. S4 a-i). However, the KASP assay for *TaPHS1* gene known for PHS tolerance, closely associated with an M-QTN (M8896) discriminated two separate SNP alleles, associated with PHS tolerance and susceptibility (Fig. 6).

Genomic prediction for SS, FN, and GC

In the present study, GP for three traits ranged from 0.41 to 0.55 with highest prediction accuracy being 0.55 for GC due to Bayes B, 0.49 for FN due to rrBLUP and 0.50 for SS due to rrBLUP (Fig. 7 and Table S9). Higher prediction accuracy for GC with Bayes B suggested that few QTLs control this trait. On the other hand, FN and SS were more accurately predicted with the help of rrBLUP, but the difference was minor relative to those due to Bayes LASSO and Bayes B (Table S9). The range/distribution of prediction accuracies for each model and trait are shown in Fig. 7.

Table 3 A summary of the main-effect QTNs (M-QTNs) and associated CGs that encode proteins containing two major domains

Parameter: M-QTN	Candidate gene ID	Functional role
(a) Leucine-rich repeat domain superfamily		
(i) SS: M10224, M10999, M1459, M6196, M3431, M8288	TraesCS2D02G041400, TraesCS2D02G041700, TraesCS2D02G042100, TraesCS2D02G042200, TraesCS2D02G044900, TraesCS2D02G045400, TraesCS3D02G409400, TraesCS3D02G409500, TraesCS3D02G409700, TraesCS3D02G410100, TraesCS4A02G055000, TraesCS4A02G055400, TraesCS7A02G031300, TraesCS7D02G053300, TraesCS7D02G053400	Activate expression of ABA responsive TFs leading to delayed germination, inhibited early root growth
(ii) FN: M3836, M3836, M5740, M9062, M9029, M9124	TraesCS1A02G364900, TraesCS1B02G381300, TraesCS1D02G345000, TraesCS1D02G344700, TraesCS3D02G067600, TraesCS3D02G068100, TraesCS3D02G068300, TraesCS5A02G388800, TraesCS7A02G069600	
(iii) GC: M7638, M7931, M3591	TraesCS1B02G437700, TraesCS1B02G437900, TraesCS1B02G438000, TraesCS1B02G438100, TraesCS1B02G438200, TraesCS1B02G438500, TraesCS1B02G438600, TraesCS3A02G478500, TraesCS3A02G478600, TraesCS3A02G478700, TraesCS3A02G479100, TraesCS3A02G479200, TraesCS3A02G479400, TraesCS6B02G405000, TraesCS6B02G405500, TraesCS6B02G405800	
(iv) FN + GC: M2654, M10484, M2207	TraesCS5A02G475800, TraesCS5A02G037500, TraesCS7A02G423400	
(b) Serine/threonine-protein kinase		
(i) SS: M10224, M10999, M8288	TraesCS2D02G039400, TraesCS2D02G040700, TraesCS2D02G045900, TraesCS7D02G052900, TraesCS7D02G053400	Involve in the regulation of auxin signaling and enhance seed dormancy
(ii) FN: M5740, M8941, M11277, M9029, M6824, M2239, M4277	TraesCS1D02G344700, TraesCS2B02G531800, TraesCS4A02G441600, TraesCS5A02G388800, TraesCS5A02G389200, TraesCS7D02G500900, TraesCS1B02G454000, TraesCS1B02G454100, TraesCS1B02G454400, TraesCS1B02G454600, TraesCS1B02G454700, TraesCS1B02G454900, TraesCS2B02G511300, TraesCS2B02G511600, TraesCS2B02G511700, TraesCS2B02G512000, TraesCS2B02G512100, TraesCS2B02G512200	
(iii) GC: M8896, M2941	TraesCS3A02G001500, TraesCS7A02G361500	
(iv) FN + GC: M2654	TraesCS5A02G474400	

SS, sprouting score; *FN*, falling number; *GC*, grain color

Discussion

During the past few decades, the genetic architecture of PHST in wheat has been studied using both linkage-based interval mapping and LD-based association mapping. These extensive studies suggested that PHST is a complex trait controlled by many QTLs/genes distributed on all 21 wheat chromosomes (Kumar et al. 2015; He et al. 2021; Liton et al. 2021; Li et al. 2021; Dhariwal et al. 2021; Khumalo et al. 2022; for reviews see Gupta et al. 2020 and Tai et al. 2021). A recent survey on interval mapping and GWAS for PHST suggested that 575 QTLs/MTAs are already known in wheat (Singh et al. 2021); this number is still growing, suggesting that additional studies are warranted using novel material for the discovery of new QTLs/MTAs for PHST. A large number of loci controlling PHS has made it rather difficult to suggest a package of molecular markers for MAS (or even marker assisted recurrent selection or MARS) for improvement of PHST in wheat, although some efforts have been made in this direction (Kumar et al. 2010; Gautam et al. 2021). Therefore, GS (particularly one based on markers associated with QTLs/MTAs) may eventually prove to be superior to MAS, although its routine effective use in wheat is yet to be demonstrated.

Efforts were also made during the present study to estimate genomic prediction (GP) accuracy for possible use in GS; three different models that were used for this purpose included rrBLUP, Bayes B and Bayes LASSO, although many more models are available. For instance, in a recent study, 11 different genomic prediction models were evaluated for their relative merit in prediction accuracy for two different traits in three different crops, namely maize, rice, and soybean (Kaler et al. 2022). Bayes B was found to be the most suitable model in their study; and therefore Bayes B model was used in the present study. However, in the present study rrBLUP was found to be the best model (see later).

Genetic variation for SS and two related traits of PHST

Characterization of PHS using multiple traits (as done in the present study) allows detection of many more QTLs relative to those detected using a single trait (Knox et al. 2012; Singh et al. 2014; Kumar et al. 2015). Fairly large variation was observed for all three traits in the present study and the broad-sense heritability for three traits ranged from 0.59 to 0.85. These estimations of the heritability indicate only a marginal role of the environment effects on the expression of the three traits; however, the SS seemed to be relatively more affected by the environment. The pattern of distribution, however, differed for the three traits, although the pattern for individual traits over the environments was similar. The dispersion of phenotypic values for three traits did not follow normal distribution; instead, it had a bimodal distribution for SS and GC and skewed distribution for FN. Similar patterns of distribution for the three traits were also reported in some earlier studies (Martinez et al. 2018; Zuo et al. 2019).

It was also observed in the present study that white/amber grain (light color, GC = 1, 2) as well as low FN are both often associated with high value of PHS (as measured using SS). This distribution pattern is also supported by the inter-se correlations between the three traits, where GC and FN showed significant positive correlation and each of these two traits had negative correlation with SS. These results

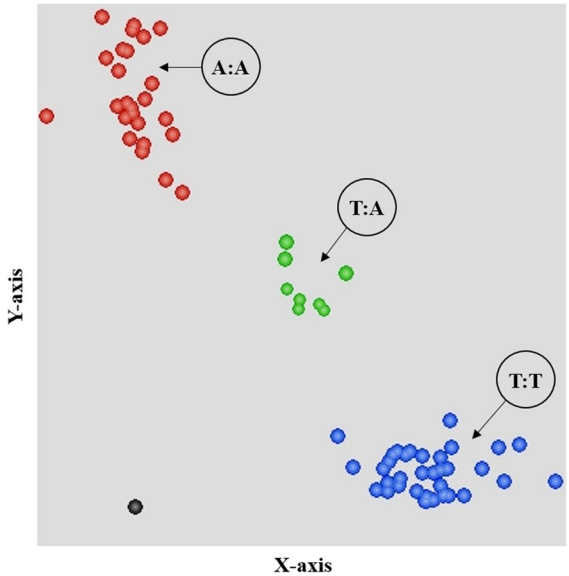


Fig. 6 Allelic discrimination plot, showing clustering of wheat accessions for KASP marker for *TaPHS1* gene for pre-harvest sprouting tolerance (PHST). Y-axis, FAM labeled red dots represent susceptible genotype allele (AA); X-axis, HEX labeled blue dots represent PHS tolerant genotype allele (TT). Green dots stand for heterozygous alleles (A/T). The black dots represent negative controls (no template)

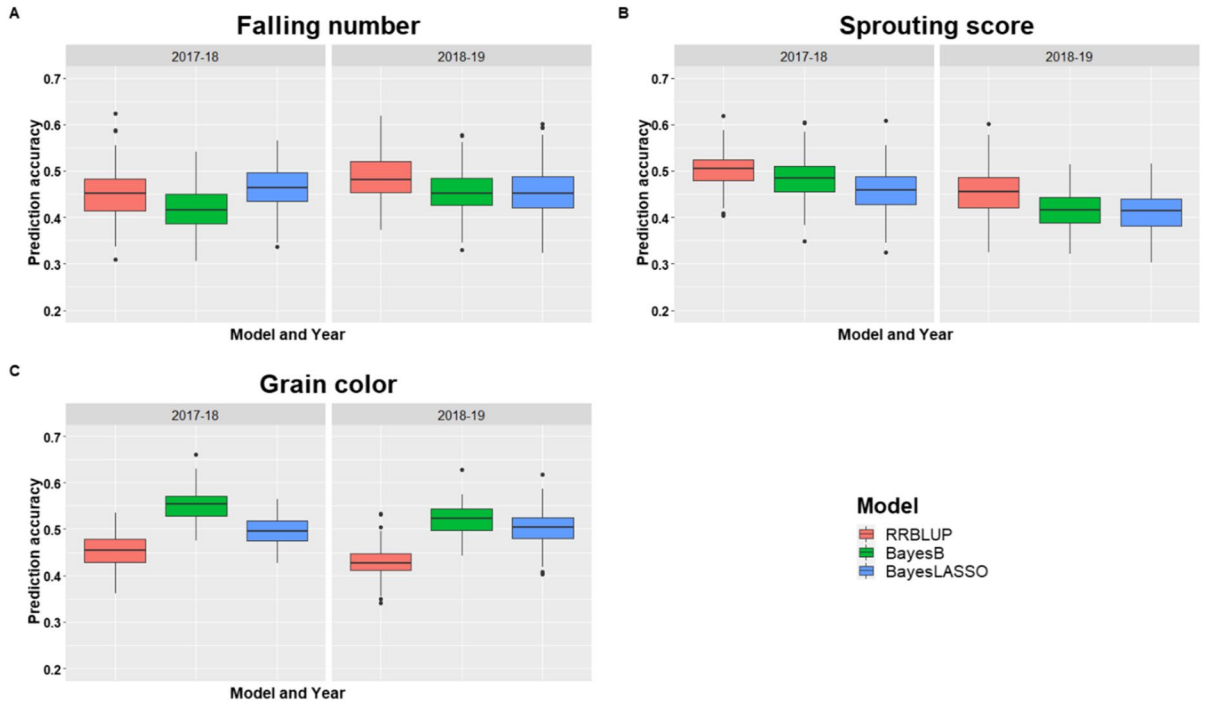


Fig. 7 Genomic prediction (GP) accuracies for three traits for pre-harvest sprouting (PHS) tolerance each using three genomic prediction models. Prediction accuracies were calculated using a cross-validation approach and in two separate environments. X-axis shows the combination of GP models and year, while y-axis depicts the prediction accuracies

also agree with earlier reports, where similar correlations were reported (Rasul et al. 2009; Jimenez et al. 2017; Martinez et al. 2018). The data for the three traits also had high correlations (0.74 to 0.99) among the environments, suggesting a very similar nature of the environments and little role of g x e interaction in the expression of the three traits in different environments. This allowed a fair and comparable analysis of the three traits over the environments.

The GWAS panel of 190 accessions used in the present study also included 19 accessions each with known high level of PHST, associated with low SS and high FN values. This once again confirmed that SS and FN can be used for scoring of PHST and that robust QTNs associated with SS, FN, and GC can be utilized for molecular breeding. The 19 genotypes with high PHST levels would serve as an excellent genetic resource of PHS tolerance to wheat breeders (Table S10). Three accessions (SWRS176, SWRS237, and SWRS325) from the above 19 accessions also had very light red grain color, suggesting that these genotypes can be used as parental genotypes in wheat breeding programs for developing PHS tolerant white (or light red grain color) wheats, preferred in several parts of the world.

The association panel used during the present study showed a low level of population structure comprising only two sub-populations. Similar panels in earlier studies also showed a low level of population structure, with the number of sub-populations ranging from three (see Malik et al. 2021; Liu et al. 2017; Zuo et al. 2019) to five (Zhou et al. 2017a, b). The absence or low level of population structure is a desirable feature for conducting GWAS; therefore, a higher level of confidence can be placed on the QTNs identified during the present study.

QTNs for the three models (CMLM, SUPER, and FarmCPU)

(Relative merits of GWAS models)

It may be recalled that 21 of the 76 (~27%) M-QTNs for the three traits were each detected by at least two of the three GWAS models in at least two of the three datasets (E1, E2, and PE). The remaining 55 M-QTNs were each

recorded in only one of the three sets of data and such M-QTNs were each detected by only one of the three models. Together, these observations once again suggest a low level of QTN x environment interaction and differences in the efficiency of the three models.

Also, nearly one third (~38%; 29/76) of the M-QTNs were co-localized with earlier reported loci for PHS-related traits (see later for more details) and remaining (~62%; 47/76) M-QTNs were novel. Interestingly, 12 novel M-QTNs, each detected by at least two models and in at least two environments explained 5.9 to 7.1% phenotypic variation (R^2); these QTNs are considered definitive and stable (Table 4). The QTLs with PVE values of 5.9% to 7.1%, if calculated using interval mapping, is generally not useful, but R^2 value of this magnitude obtained from GWAS may be useful. The efficiency of three GWAS models and other available models and the importance of definitive and stable M-QTNs in MAS needs further discussion (see next).

With the availability of at least nine different models for GWAS, which became available during the past 10–15 years, only three were used in the present study, which included CMLM, SUPER, and FarmCPU. Their relative merits are known, where FarmCPU has been shown to be more efficient relative to other available models, as also shown in a recent study in soybean and maize involving eight different models (Kaler et al. 2020; for a review, see Gupta 2021). FarmCPU also has in-built Bonferroni correction with a view to minimize the number of false positive QTNs, which appear due to the multiple testing involved in GWAS. It is also widely known that Bonferroni correction also leads to false negatives (an undesirable feature). Therefore, it is emphasized that the Bonferroni correction is a trade-off (Chen et al. 2017). With a view to restrict the false positive QTNs (an undesirable feature) in the remaining two single-locus models (CMLM and SUPER), we used the high threshold p-value of the FarmCPU to identify significant M-QTNs detected by CMLM and SUPER. Therefore, we assume that the M-QTNs detected by all three models in the present study are true and free from false positives.

QTNs co-localized with known MTAs/QTLs/genes

Another important observation of the present study is the availability of 29 M-QNTs for the three traits, which were co-localized with the previously reported loci (QTLs, MTAs, and genes) for PHS tolerance (Kumar et al. 2015; for review see Gupta et al. 2020; Tai et al. 2021; Singh et al. 2021). We are, however, conscious of the fact that co-localization does not necessarily mean identity between loci but can certainly help in identification of some hot spots to be used in molecular breeding. Identification of hot spots for successful implementation of MAS has also been discussed, as reported in a recent study in chickpea (Barmukh et al. 2022).

It may also be recalled that the above 29 M-QTNs are located within the 26 PRGRs identified during the present study; each PRGR is 5 Mbp long (average region in LD; Ladejobi et al. 2019) and thus the M-QTNs detected during the present study may be considered in LD with the previously reported loci and could be used for improvement of PHS tolerance. The majority of these PRGRs are located near the telomeres in the gene rich regions reported by Erayman et al. (2004).

It is also important to note that five cloned genes (out of the 19 identified and characterized genes) for PHS tolerance (Tai et al. 2021) are also located one each within the five different PRGRs on chromosomes 1B, 3A, 3B, 3D, and 4A (Table 2). However, none of these cloned genes carried any M-QTN; it is possible that the use of a much greater number of SNPs for GWAS analysis could have identified causal SNPs within one or more of the five cloned

Table 4 Twenty novel/co-localized, definitive, and stable main-effect QTNs (M-QTNs) for improvement of pre-harvest sprouting (PHS) tolerance using MARS (detailed information is given in Supplementary Table S4)

Parameter: M-QTN (chromosome)	Range of R^2
(a) Novel QTNs	
SS: M4235 (2B), M12780 (5D), M6492 (7B), M1526 (7D); FN: M10792 (1D); GC: M5955 (1A); SS + GC: M10913 (1B); FN + GC: M11378 (1B), M8514 (2B), M10484 (5A), M2654 (5A), M2207 (7A)	6.1–7.9
(b) Co-localized QTNs	
SS: M9945 (2B), M10224 (2D), M37 (7B); FN: M10792 (1D); GC: M1920 (4A), M1442 (5B), M4283 (7A); FN + GC: M11590 (1A)	5.9–8.4

MARS, marker-assisted recurrent selection; *SS*, sprouting score; *FN*, falling number; *GC*, grain color; R^2, Percent phenotypic variation explained

genes. Following are some of the details of these five cloned genes and the associated M-QTNs, which may be useful for further study: (i) *TaABI4* (*ABSCISIC ACID INSENSITIVE 4*) encodes an ABA-responsive transcription factor having a role in seed germination (Xiao et al. 2021); it is co-located with QTN M9786 for SS (PRGR3 on 1B). (ii) *TaPHS1*; a *MOTHER OF FT AND TFL1* (*TaMFT*)-like gene in PRGR11 is co-located with M-QTN M8896 (PRGR11 on 3A). *TaMFT-A1* is known for strong seed dormancy in spring and winter wheats through regulation of ABA and GA signal transduction (Nakamura et al. 2011; Lei et al. 2013; Cao et al. 2016; Ahmad et al. 2019). Interestingly, the QTL *QPHS.wsu-3A.1* for PHS was also reported in the same genomic region by Martinez et al. (2018), emphasizing the importance of the loci in PRGR11. (iii) *Phs-A1* is a gene for PHS tolerance and is mapped close to the M-QTN M10478 for SS (PRGR14 on 4A) (Flintham 2000; Imtiaz et al. 2008; Ogbonnaya et al. 2008; Torada et al. 2008, 2016; Mares and Mrva 2014; Barrero et al. 2015; Kumar et al. 2015; Shorinola et al. 2016, 2017). Barrero et al. (2015) and Torada et al. (2016) independently reported tandem duplicated *Plasma Membrane* genes, namely *PM19-A1* and *PM19-A2* as the leading candidates for *Phs-A1*. Studies further confirmed that the *Phs-A1* is also linked to another gene, *mitogen-activated protein kinase kinase 3* (*TaMKK3-A*) on 4AL. (iv) *TaVP1* gene is associated with M-QTN M1459 and with SS (PRGR13 on 3D) (Bailey et al. 2019; Dale et al. 2017). (v) *Tamyb10* gene is located close to M-QTN (M381) for FN (PRGR12 on 3B) at a distance of 0.4 Mb. Previously, *Tamyb10* transcription factor each, which expresses in developing grains, were reported to be located in the regions of three R loci (Himi and Noda 2005). Later, the *Tamyb10-D* was shown to be associated with GC and PHST in wheat, and thus was recommended for selecting white-grained PHS tolerant wheat cultivars (Lang et al. 2021).

Epistatic QTNs

Epistatic QTNs (E-QTNs) representing 15 pairs of first order digenic epistatic interactions were also identified in the present study. Five of these interactions also had a pleiotropic effect on FN and GC. However, the variation explained by M-QTNs seems to be more important than that due to epistatic interactions for PHS tolerance. Epistatic interactions in wheat using GWAS have also been reported for a number of traits, including flowering time (Reif et al. 2011; Langer et al. 2014), stem rust resistance (Yu et al. 2011), agronomic traits (Sehgal et al. 2017), micronutrients (Kumar et al. 2018), yield related traits (Jaiswal et al. 2016; Malik et al. 2021), nitrogen use efficiency (NUE) (Ranjan et al. 2021; Singh et al. 2023), and seed longevity (Arif et al. 2022). Examples of the application of epistatic interactions in MAS are not available, but the information on epistatic interactions is important for an understanding of the genetic architecture of the traits such as PHS that are relevant for wheat improvement. Optimistically, epistatic QTNs will also be utilized in MAS and GS in the future.

Prioritization of QTNs and validation using KASP assay for MAS

Based on the results of the present study, we prioritized the following two sets of M-QTNs (Table 4): (i) Twelve novel M-QTNs, which were not co-localized with known M-QTNs, and (ii) Eight M-QTNs, which were detected by all three models of GWAS and in at least two environments, explaining 5.9% to 8.4% phenotypic variation (R^2) (Table 4). Both these sets of M-QTNs deserve further discussion. Twelve novel and eight co-localized M-QTNs (see above) that were treated as definitive and stable are good candidates for MAS (Tables 4a, b). Seven of these 20 M-QTNs were multi-trait QTNs and can be used for simultaneous improvement of two of the three PHS-related traits.

The validation of the above prioritized M-QTNs using KASP assay before their use in MARS was largely unsuccessful, indicating the challenges associated with successful KASP genotyping in allo-hexaploid wheat (Makhoul et al. 2020). However, in the present study, the KASP marker for the gene *TaPHS1* for PHS tolerance (Wang et al. 2020a), which is closely associated with a M-QTN (M8896) for GC on chromosome 3A, showed clear separation of two clusters belonging to the alternative alleles for PHS tolerance and susceptibility when used with accessions of the GWAS panel (Fig. 6). KASP assays for a variety of individual major genes have also been successful in wheat (Yu et al. 2017; Singh et al. 2019). Therefore, we propose that in future studies, the KASP

assays may be developed for CGs closely associated with MTAs/QTNs rather than for MTAs/QTNs per se and the same may be validated in multiple bi-parental populations because no single mapping population is expected to segregate for all the MTAs/QTNs identified through GWAS.

CGs for PHS-related traits

All 176 CGs reported during the present study have a functional role in controlling various molecular mechanisms involved in determining seed dormancy/PHS tolerance in wheat and other crop plants. These CGs are involved in the following functions, as revealed through gene ontology studies (i) ABA signaling, (ii) auxin signaling, (iii) GA biosynthesis, (iv) dormancy and germination, and (v) starch/carbohydrate degradation. However, a high proportion of CGs (119/176) participate in ABA biosynthesis and signaling network and thus deserve some discussion to understand the molecular basis of PHS tolerance in wheat (Table 3). These identified CGs that participate in encoding ABA have functions related to glucose signaling, metabolism, root growth, defective embryo development, catalytic activity, degradation (or de-activation) of ABA-signaling, etc. (Zhou et al. 2017a, b; Park et al. 2018; Rikiishi et al. 2021; Wang et al. 2020b). For example, the TraesCS2D02G041400 gene underlying M-QTN M10224 (PRGR10) for SS belong to the LRR domain superfamily on chromosome 2D. This gene regulates ABA signal transduction in *Arabidopsis thaliana*, which has been confirmed by knock-out, resulting in ABA insensitivity during seed germination (Osakabe et al. 2005). In general, seed dormancy and the germination (i.e., PHS tolerance) are believed to be antagonistically controlled by two major endogenous hormones, namely ABA and GA in many plant species (Chen et al. 2020). ABA affects dormancy formation, while GA enhances germination (Hilhorst and Karssen 1992, see Sohn et al. 2021 for review). Although, over the past decades, ABA and GA have been studied in some details, the regulatory mechanisms of the ratio of ABA and GA in controlling seed dormancy and germination remain to be fully resolved.

Surprisingly, the CGs underlying M-QTNs for FN have been shown to contain ubiquitin-like, Rad60/SUMO-like domain and WD40 domain both of which are involved in positive/negative regulation of ABA signaling during seed germination (Zheng et al. 2012; Wang et al. 2020b). It is uncommon to find ubiquitin-related genes in the control of dormancy or PHS in wheat. Recently, He et al. (2021) reported TraesCS3D01G466100 as a novel PHS tolerance CG important for seed dormancy. It is proposed that the gene may have a crucial role in the action of ubiquitin ligase enzymes 3 (E3), which could contribute to distinct PHS tolerance phenotypes. These findings indicate that control of PHS or seed dormancy is genetically complex and additional detailed studies are needed to confirm the real function of these CGs in PHS tolerance.

Genomic prediction for SS, FN, and GC

The GP accuracies using three models were found to range from 0.41 to 0.55. These results are comparable to the results of several other studies on PHS in wheat (Heffner et al. 2011; Moore et al. 2017). The highest prediction accuracy ranging from 0.49 to 0.62 was reported by Moore et al. (2017). The GP accuracies for PHST reported in this study and earlier studies were within the range for those reported for grain yield, which ranged from 0.37 to 0.45 in a recent study (Juliana et al. 2021) and from 0.62 to 0.65 in an earlier study (Basnet et al. 2018). In contrast to the present study, multivariate models have also been used and shown promise over the single trait prediction accuracy models (for reviews see Montesinos-López et al. 2021; Gill et al. 2021; Shahi et al. 2022). Notwithstanding the above, differences in the assumptions and algorithms concerning the variances of complex traits, the marker density, population size, trait architecture, heritability, relatedness between testing and training population, and selection intensity together affect the accuracies of GP (Wang et al. 2018; Merrick et al. 2022a, b; for a review see Budhlakoti et al. 2022) in both a variety of parametric models (rrBLUP, GBLUP, LASSO, Bayes A, Bayes B, and Bayesian LASSO) and nonparametric models (RKHS, neural networks and random forests) that have been used for GP (for a review see Shahi et al. 2022). However, it has been argued that with GP accuracies > 0.50, the GS allows maximum annual genetic gains compared to other breeding methods (Longin et al. 2015), although GS is still in a transition phase and is being tested to find out its utility in wheat breeding.

GWAS and GP have been sparingly used together for the same trait in a crop (Odilbekov et al. 2019; Kibe et al. 2020; Sandhu et al. 2021a; Alemu et al. 2021). It needs to be emphasized, however, that a joint study of GWAS and GP will be useful, only when the future GS program will utilize a training population selected from the GWAS panel used for calculation of GP. The breeding population should also be relevant as recommended for GS.

We may also like to examine the merits of the three models that were used for GP. In this study, the model rrBLUP has been shown to be superior in predicting SS and FN while Bayes B performed better for GC; this is partly due to different genetic architecture of each component trait (Merrick et al. 2022a) and suggests that different models may need to be used, depending upon the crop, the trait, and the heritability of the trait involved, as done in a recent study (Kibe et al. 2020).

The model rrBLUP assumes that all markers contribute to the trait, have a random effect with common variance and thus reduces the effect of all markers equally toward zero, which is true for the polygenic traits such as SS and FN in our study (Endelman 2011; Sandhu et al. 2022b). Furthermore, Bayes B uses a selection operator based on an assumption that few loci contribute to the traits and remaining markers have a zero effect. Bayes B also uses a Markov Chain Monte Carlo to estimate the effect of markers and thus are computationally expensive, whereas rrBLUP uses a ridge regression coefficient to estimate the markers effect and is faster. The rrBLUP is mostly the preferred model in wheat breeding programs in addition to GBLUP (Pérez and de los Campos 2014). More information about the working of these models is given in Sandhu et al. (2022a) and Merrick et al. (2022a).

Conclusions

A large number of M-QTNs and E-QTNs for SS, FN, and GC were identified during the present study, suggesting the polygenic nature and role of epistasis in controlling PHS tolerance in wheat. The number of QTNs reported by the three GWAS models also differed, suggesting differences in their efficiencies. Some of the QTNs detected during the present study were mapped together with earlier reported loci and cloned genes in 26 PRGRs, each genomic region being 5Mbp long. These observations validated our results. A sizable frequency of QTNs were novel and make a good source for future studies. A set of definitive and stable QTNs has been proposed for use in MARS for the improvement of PHS tolerance in wheat; however, validation of these QTNs using KASP assay require further efforts in future studies. Validated KASP marker for *TaPHS1* gene in the present study may be utilized in breeding programs aimed at improving PHS tolerance in wheat. GP accuracies (0.41–0.55) for the three traits estimated using three different models suggested that GP can be used for planning a GS program for improvement of PHS tolerance. Such a GS program will have to be based on a training population derived from the GWAS panel used in the present study, and a related breeding population. Further refinement in GP models may also allow improvement in GP accuracy so that GS will be routinely used in due course of time. A set of CG underlying M-QTNs were also identified, which broadly suggested the role of ABA in PHS tolerance. These CGs may be examined in detail in the future studies. In recent years, emphasis is also being laid on post-GWAS, where available statistics from GWAS is used for extracting further information (Gupta et al. 2019).

Acknowledgements The present study was supported by the Science and Engineering Research Board (SERB), Department of Science and Technology (DST), Government of India, New Delhi. HSB was awarded Honorary Scientist position by the Indian National Science Academy (INSA), New Delhi. We are thankful to Professor Parveen Chhuneja, Director, School of Agricultural Biotechnology, Punjab Agricultural University (PAU), Ludhiana, Punjab for allowing us to use KASP assay facilities. The Head, Department of Genetics and Plant Breeding, Chaudhary Charan Singh University, Meerut provided necessary facilities for this study. CIMMYT, Mexico provided the genetic materials for the present research work.

Author's contribution SK and HSB conceived the idea and designed the experiment. MK, RP, AN, and HS conducted the field experiments and phenotyping. MK, GS, and TG performed GWAS analysis. KSS and NK assisted with genomic prediction analysis, interpretation, and writing. MK and SK wrote the first draft of the manuscript. SK, HSB, and PKG critically revised and edited the manuscript. All authors read and approved the final manuscript.

Funding This work was financially supported by the Science and Engineering Research Board (SERB), Department of Science and Technology (DST), Government of India, New Delhi under the Early Career Research Award (Grant No. ECR/2016/001774).

Data availability Relevant data are included in this paper and its associated Supplementary Information (SI).

Declarations

Ethics declaration Not applicable.

Ethics approval and consent to participate Not applicable.

Consent for publication Not applicable.

Competing interests The authors of this manuscript declare no conflicts of interest.

References

Ahmad I, Kamran M, Meng X, Ali S, Bilegjargal B, Cai T, Liu T, Han Q (2019) Effects of plant growth regulators on seed filling, endogenous hormone contents and maize production in semiarid regions. J Plant Growth Regul 38(4):1467–1480. https://doi.org/10.1007/s00344-019-09949-2

Alemu A, El Baouchi A, El Hanafi S, Kehel Z, Eddakhir K, Tadesse W (2021) Genetic analysis of grain protein content and dough quality traits in elite spring bread wheat (*Triticum aestivum*) lines through association study. J Cereal Sci 100:103214. https://doi.org/10.1016/j.jcs.2021.103214

Alvarado M, López M, Vargas A, Vargas M, Pacheco Á, Rodríguez F, Burgueño J, Crossa J (2015) META-R (multi environment trail analysis with R for windows) Version 4.1. CIMMYT Research Data & Software Repository Network, V23, pp 6–8. http://hdl.handle.net/11529/10201. Accessed 6 Jan 2021

Ali A, Cao J, Jiang H, Chang C, Zhang HP, Sheikh SW, Shah L, Ma C (2019) Unravelling molecular and genetic studies of wheat (*Triticum aestivum* L.) resistance against factors causing pre-harvest sprouting. Agronomy 9(3):117. https://doi.org/10.3390/agronomy9030117

Arif MA, Agacka-Mołdoch M, Qualset CO, Börner A (2022) Mapping of additive and epistatic QTLs linked to seed longevity in bread wheat (*Triticum aestivum* L.). Cereal Res Commun 11:1–7. https://doi.org/10.1007/s42976-021-00240-3

Bailey PC, McKibbin RS, Lenton JR, Holdsworth MJ, Flintham JE, Gale MD (2019) Genetic map locations for orthologous Vp1 genes in wheat and rice. Theor Appl Genet 98(2):281–284. https://doi.org/10.1007/s001220051069

Barrero JM, Cavanagh C, Verbyla KL, Tibbits JF, Verbyla AP, Huang BE, Rosewarne GM, Stephen S, Wang P, Whan A, Rigault P (2015) Transcriptomic analysis of wheat near-isogenic lines identifies PM19-A1 and A2 as candidates for a major dormancy QTL. Genome Biol 16(1):1–8. https://doi.org/10.1186/s13059-015-0665-6

Barmukh R, Roorkiwal M, Garg V, Khan AW, German L, Jaganathan D, Chitikineni A, Kholova J, Kudapa H, Kaliamoorthy S, Samineni S, Kale SM, Gaur PM, Sagurthi SR, Benitez-Alfonso Y, Varshney RK (2022) Genetic variation in *CaTIFY4b* contributes to drought adaptation in chickpea. Plant Biotechnol J 1–15. https://doi.org/10.1111/pbi.13840

Basnet BR, Crossa J, Dreisigacker S, Pérez-Rodríguez P, Manes Y, Singh RP, Rosyara U, Camarillo-Castillo F, Murua M (2018) Hybrid wheat prediction using genomic, pedigree, and environmental covariables interaction models. Plant Genome 12:9. https://doi.org/10.3835/plantgenome2018.07.0051

Bates D, Mächler M, Bolker BM, Walker SC (2015) Fitting linear mixed-effects models using lme4. J Stat Softw 67(1):1–48. https://doi.org/10.18637/jss.v067.i01

Bernardo R (2016) Bandwagons I, too, have known. Theor Appl Genet 129:2323–2332. https://doi.org/10.1007/s00122-016-2772-5

Budhlakoti N, Kushwaha AK, Rai A, Chaturvedi KK, Kumar A, Pradhan AK, Kumar U, Kumar RR, Juliana P, Mishra DC, Kumar S (2022) Genomic selection: a tool for accelerating the efficiency of molecular breeding for development of climate-resilient crops. Front Genet 13:832153. https://doi.org/10.3389/fgene.2022.832153

Cao L, Hayashi K, Tokui M, Mori M, Miura H, Onishi K (2016) Detection of QTLs for traits associated with preharvest sprouting resistance in bread wheat (*Triticum aestivum* L.). Breed Sci 66(3):462. https://doi.org/10.1270/jsbbs.66.260

Chen SY, Feng Z, Yi X (2017) A general introduction to adjustment for multiple comparisons. J Thorac Dis 9(6):1725. https://doi.org/10.21037/jtd.2017.05.34

Chen K, Li GJ, Bressan RA, Song CP, Zhu JK, Zhao Y (2020) Abscisic acid dynamics, signaling, and functions in plants. J Integr Plant Biol 62(1):25–54. https://doi.org/10.1111/jipb.12899

Crossa J, Pérez-Rodríguez P, Cuevas J, Montesinos-López O, Jarquín D, de los Campos G, Burgueño J, González-Camacho JM, Pérez-Elizalde S, Beyene Y, Dreisigacker S (2017) Genomic selection in plant breeding: methods, models, and perspectives. Trends Plant Sci 22:961–975. https://doi.org/10.1016/j.tplants.2017.08.011

Dale Z, Jie H, Luyu H, Cancan Z, Yun Z, Yarui S, Suoping L (2017) An advanced backcross population through synthetic octaploid wheat as a "bridge": development and QTL detection for seed dormancy. Front Plant Sci 8:2123. https://doi.org/10.3389/fpls.2017.02123

de los Campos G, Naya H, Gianola D, Crossa J, Legarra A, Manfredi E, Weigel K, Cotes JM (2009) Predicting quantitative traits with regression models for dense molecular markers and pedigree. Genetics 182(1):375–385.https://doi.org/10.1534/genetics.109.101501

DePauw RM, Knox RE, Singh AK, Fox SL, Humphreys DG, Hucl P (2012) Developing standardized methods for breeding preharvest sprouting resistant wheat, challenges and successes in Canadian wheat. Euphytica 188(1):7–14. https://doi.org/10.1007/s10681-011-0611-y

Dhariwal R, Hiebert CW, Sorrells ME, Spaner D, Graf RJ, Singh J, Randhawa HS (2021) Mapping pre-harvest sprouting resistance loci in AAC Innova × AAC Tenacious spring wheat population. BMC Genom 22(1):1–20. https://doi.org/10.1186/s12864-021-08209-6

Endelman JB (2011) Ridge regression and other kernels for genomic selection with R package rrBLUP. Plant Genome 4:250–255. https://doi.org/10.3835/plantgenome2011.08.0024

Erayman M, Sandhu D, Sidhu D, Dilbirligi M, Baenziger PS, Gill KS (2004) Demarcating the gene-rich regions of the wheat genome. Nucleic Acids Res 32(12):3546–3565. https://doi.org/10.1093/nar/gkh639

Flintham JE (2000) Different genetic components control coat-imposed and embryo-imposed dormancy in wheat. Seed Sci Res 10(1):43–50. https://doi.org/10.1017/S0960258500000052

Gahlaut V, Jaiswal V, Singh S, Balyan HS, Gupta PK (2019) Multi-locus genome wide association mapping for yield and its contributing traits in hexaploid wheat under different water regimes. Sci Rep 9:19486. https://doi.org/10.1038/s41598-019-55520-0

Gahlaut V, Jaiswal V, Balyan HS, Gupta PK (2021) Multi-locus GWAS for grain weight related traits under rain-fed conditions in common wheat (*Triticum aestivum* L.). Front Plant Sci 21:2266. https://doi.org/10.3389/fpls.2021.758631

Gautam T, Kumar K, Agarwal P, Tyagi S, Jaiswal V, Gahlaut V, Kumar S, Prasad P, Chhuneja P, Balyan HS, Gupta PK (2021) Development of white-grained PHS-tolerant wheats with high grain protein and leaf rust resistance. Mol Breed 41(6):1–4. https://doi.org/10.1007/s11032-021-01234-z

Gill HS, Halder J, Zhang J, Brar NK, Rai TS, Hall C, Bernardo A, Amand PS, Bai G, Olson E, Ali S (2021) Multi-trait multi-environment genomic prediction of agronomic traits in advanced breeding lines of winter wheat. Front Plant Sci 12:709545. https://doi.org/10.3389/fpls.2021.709545

Gupta PK, Kulwal PL, Jaiswal V (2019) Association mapping in plants in the post-GWAS genomics era. Adv Genet 104:75–154. https://doi.org/10.1016/bs.adgen.2018.12.001

Gupta PK, Balyan HS, Sharma S, Kumar R (2020) Genetics of yield, abiotic stress tolerance and biofortification in wheat (*Triticum aestivum* L.). Theor Appl Genet 133(5):1569–1602. https://doi.org/10.1007/s00122-020-03583-3

Gupta PK (2021) GWAS for genetics of complex quantitative traits: Genome to pangenome and SNPs to SVs and k-mers. BioEssays 43(11):2100109. https://doi.org/10.1002/bies.202100109

He J, Zhang D, Chen X, Li Y, Hu M, Sun S, Su Q, Su Y, Li S (2021) Identification of QTLs and a candidate gene for reducing pre-harvest sprouting in *Aegilops tauschii–Triticum aestivum* chromosome segment substitution lines. Int J Mol Sci 22(7):3729. https://doi.org/10.3390/ijms22073729

Heffner EL, Jannink JL, Iwata H, Souza E, Sorrells ME (2011) Genomic selection accuracy for grain quality traits in biparental wheat populations. Crop Sci 51:2597–2606. https://doi.org/10.2135/cropsci2011.05.0253

Hilhorst HW, Karssen CM (1992) Seed dormancy and germination: the role of abscisic acid and gibberellins and the importance of hormone mutants. J Plant Growth Regul 11(3):225–238. https://doi.org/10.1007/BF00024561

Himi E, Mares DJ, Yanagisawa A, Noda K (2002) Effect of grain colour gene (R) on grain dormancy and sensitivity of the embryo to abscisic acid (ABA) in wheat. J Exp Bot 53(374):1569–1574. https://doi.org/10.1093/jxb/erf005

Himi E, Noda K (2005) Red grain colour gene (R) of wheat is a Myb-type transcription factor. Euphytica 143(3):239–242

Imtiaz M, Ogbonnaya FC, Oman J, van Ginkel M (2008) Characterization of quantitative trait loci controlling genetic variation for preharvest sprouting in synthetic backcross-derived wheat lines. Genetics 178(3):1725–1736. https://doi.org/10.1534/genetics.107.084939

Jaiswal V, Gahlaut V, Meher PK, Mir RR, Jaiswal JP, Rao AR, Balyan HS, Gupta PK (2016) Genome wide single locus single trait, multi-locus and multi-trait association mapping for some important agronomic traits in common wheat (*Triticum aestivum* L.). PLoS One 11(7):0159343. https://doi.org/10.1371/journal.pone.0159343

Jan HU, Guan M, Yao M, Liu W, Wei D, Abbadi A, Zheng M, He X, Chen H, Guan C, Nichols RA (2019) Genome-wide haplotype analysis improves trait predictions in *Brassica napus* hybrids. Plant Sci 283:157–164. https://doi.org/10.1016/j.plantsci.2019.02.007

Jimenez N, Mares D, Mrva K, Lizana C, Contreras S, Schwember AR (2017) Susceptibility to preharvest sprouting of Chilean and Australian elite cultivars of common wheat. Crop Sci 57(1):462–474. https://doi.org/10.2135/cropsci2016.02.0138

Juliana P, He X, Marza F, Islam R, Anwar B, Poland J, Shrestha S, Singh GP, Chawade A, Joshi AK, Singh RP (2021) Genomic selection for wheat blast in a diversity panel, breeding panel and full-sibs panel. Front Plant Sci 12. https://doi.org/10.3389/fpls.2021.745379

Kaler AS, Gillman JD, Beissinger T, Purcell LC (2020) Comparing different statistical models and multiple testing corrections for association mapping in soybean and maize. Front Plant Sci 10:1794. https://doi.org/10.3389/fpls.2019.01794

Kaler AS, Purcell LC, Beissinger T, Gillman JD (2022) Genomic prediction models for traits differing in heritability for soybean, rice, and maize. BMC Plant Biol 22:87. https://doi.org/10.1186/s12870-022-03479-y

Kang HM, Noah AZ, Claire MW, Andrew K, David H, Mark JD, Eleazar E (2008) Efficient control of population structure in model organism association mapping. Genetics 178(3):1709–1723. https://doi.org/10.1534/genetics.107.080101

Khumalo TP, Hlongoane T, Barnard A, Tsilo TJ (2022) Genomic regions influencing preharvest sprouting tolerance in two doubled-haploid wheat populations (*Triticum aestivum* L.). Agronomy 12(4):832. https://doi.org/10.3390/agronomy12040832

Kibe M, Nair SK, Das B, Bright JM, Makumbi D, Kinyua J, Suresh LM, Beyene Y, Olsen MS, Prasanna BM, Gowda M (2020) Genetic dissection of resistance to gray leaf spot by combining genome-wide association, linkage mapping, and genomic prediction in tropical maize germplasm. Front Plant Sci 11:572027. https://doi.org/10.3389/fpls.2020.572027

Knox RE, Clarke FR, Clarke JM, Fox SL, DePauw RM, Singh AK (2012) Enhancing the identification of genetic loci and transgressive segregants for preharvest sprouting resistance in a durum wheat population. Euphytica 186(1):193–206. https://doi.org/10.1007/s10681-011-0557-0

Kumar J, Mir RR, Kumar N, Kumar A, Mohan A, Prabhu KV, Balyan HS, Gupta PK (2010) Marker-assisted selection for pre-harvest sprouting tolerance and leaf rust resistance in bread wheat. Plant Breed 129(6):617–621. https://doi.org/10.1111/j.1439-0523.2009.01758.x

Kumar J, Saripalli G, Gahlaut V, Goel N, Meher PK, Mishra KK, Mishra PC, Sehgal D, Vikram P, Sansaloni C, Singh S (2018) Genetics of Fe, Zn, β-carotene, GPC and yield traits in bread wheat (*Triticum aestivum* L.) using multilocus and multi-traits GWAS. Euphytica 214(11):1–7. https://doi.org/10.1007/s10681-018-2284-2

Kumar S, Knox RE, Clarke FR, Pozniak CJ, DePauw RM, Cuthbert RD, Fox S (2015) Maximizing the identification of QTL for pre-harvest sprouting resistance using seed dormancy measures in a white-grained hexaploid wheat population. Euphytica 205(1):287–309. https://doi.org/10.1007/s10681-015-1460-x

Ladejobi O, Mackay IJ, Poland J, Praud S, Hibberd JM, Bentley AR (2019) Reference genome anchoring of high-density markers for association mapping and genomic prediction in European winter wheat. Front Plant Sci 10:1278. https://doi.org/10.3389/fpls.2019.01278

Lang J, Fu Y, Zhou Y, Cheng M, Deng M, Li M, Zhu T, Yang J, Guo X, Gui L, Li L (2021) Myb10-D confers PHS-3D resistance to pre-harvest sprouting by regulating NCED in ABA biosynthesis pathway of wheat. New Phytol 230(5):1940–1952

Langer SM, Longin CF, Würschum T (2014) Flowering time control in European winter wheat. Front Plant Sci 5:537. https://doi.org/10.3389/fpls.2014.00537

Lei L, Zhu X, Wang S, Zhu M, Carver BF, Yan L (2013) *TaMFT-A1* is associated with seed germination sensitive to temperature in winter wheat. PLoS ONE 8(9):73330. https://doi.org/10.1371/journal.pone.0073330

Li L, Zhang Y, Zhang Y, Li M, Tian X, Song J, Luo X, Xie L, Wang D, He Z, Xia X (2021) Genome-wide linkage mapping for preharvest sprouting resistance in wheat using 15K single nucleotide polymorphism arrays. Front Plant Sci 12. https://doi.org/10.3389/fpls.2021.749206

Lin M, Zhang D, Liu S, Zhang G, Yu J, Fritz AK, Bai G (2016) Genome-wide association analysis on pre-harvest sprouting resistance and grain color in US winter wheat. BMC Genomics 17(1):1–6. https://doi.org/10.1186/s12864-016-3148-6

Lipka AE, Tian F, Wang Q, Peiffer J, Li M, Bradbury PJ, Gore MA, Buckler ES, Zhang Z (2012) GAPIT: genome association and prediction integrated tool. Bioinformatics 28(18):2397–2399. https://doi.org/10.1093/bioinformatics/bts444

Liton MM, McCartney CA, Hiebert CW, Kumar S, Jordan MC, Ayele BT (2021) Identification of loci for pre-harvest sprouting resistance in the highly dormant spring wheat RL4137. Theor Appl Genet 134(1):113–124. https://doi.org/10.1007/s00122-020-03685-y

Liu C, Parveen RS, Revolinski SR, Campbell KA, Pumphrey MO, Steber CM (2021) The genetics of late maturity alpha-amylase (LMA) in North American spring wheat (*Triticum aestivum* L.). Seed Sci Res 31(3):159–168. https://doi.org/10.1017/S0960258521000064

Liu X, Huang M, Fan B, Buckler ES, Zhang Z (2016) Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. PLoS Genet 12(2):1005767. https://doi.org/10.1371/journal.pgen.1005767

Liu Y, Liu Y, Zhou Y, Wight C, Pu Z, Qi P, Jiang Q, Deng M, Wang Z, Wei Y, Cao W (2017) Conferring resistance to pre-harvest sprouting in durum wheat by a QTL identified in *Triticum spelta*. Euphytica 213:19. https://doi.org/10.1007/s10681-016-1796-x

Longin CF, Mi X, Würschum T (2015) Genomic selection in wheat: optimum allocation of test resources and comparison of breeding strategies for line and hybrid breeding. Theor Appl Genet 128(7):1297–1306. https://doi.org/10.1007/s00122-015-2505-1

Makhoul M, Rambla C, Voss-Fels KP, Hickey LT, Snowdon RJ, Obermeier C (2020) Overcoming polyploidy pitfalls: a user guide for effective SNP conversion into KASP markers in wheat. Theor Appl Genet 133:2413–2430. https://doi.org/10.1007/s00122-020-03608-x

Malik P, Kumar J, Singh S, Sharma S, Meher PK, Sharma MK, Roy JK, Sharma PK, Balyan HS, Gupta PK, Sharma S (2021) Single-trait, multi-locus and multi-trait GWAS using four different models for yield traits in bread wheat. Mol Breed 41:1–21. https://doi.org/10.1007/s11032-021-01240-1

Malik P, Kumar J, Sharma S, Meher PK, Balyan HS, Gupta PK, Sharma S (2022) GWAS for main effects and epistatic interactions for grain morphology traits in wheat. Physiol Mol Biol Plants 28(3):651–668. https://doi.org/10.1007/s12298-022-01164-w

Mares DJ (1983) Preservation of dormancy in freshly harvested wheat grain. Aust J Agric Res 34(1):33–38. https://doi.org/10.1071/AR9830033

Mares DJ, Mrva K (2014) Wheat grain preharvest sprouting and late maturity alpha-amylase. Planta 240:1167–1178. https://doi.org/10.1007/s00425-014-2172-5

Martinez SA, Godoy J, Huang M, Zhang Z, Carter AH, Garland Campbell KA, Steber CM (2018) Genome-wide association mapping for tolerance to preharvest sprouting and low falling numbers in wheat. Front Plant Sci 9:141. https://doi.org/10.3389/fpls.2018.00141

McMaster GJ, Derera NF (1976) Methodology and sample preparation when screening for sprouting damage in cereals. Cereal Res Commun 1:251–254. https://www.jstor.org/stable/23777431

Merrick LF, Herr AW, Sandhu KS, Lozada DN, Carter AH (2022a) Optimizing plant breeding programs for genomic selection. Agronomy 12:714. https://doi.org/10.3390/agronomy12030714

Merrick LF, Herr AW, Sandhu KS, Lozada DN, Carter AH (2022b) Utilizing genomic selection for wheat population development and improvement. Agronomy 12:522. https://doi.org/10.3390/agronomy12020522

Meuwissen TH, Hayes BJ, Goddard M (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–1829. https://doi.org/10.1534/genetics.116.189795

Montesinos-López A, Runcie DE, Ibba MI, Pérez-Rodríguez P, Montesinos-López OA, Crespo LA, Bentley AR, Crossa J (2021) Multi-trait genomic-enabled prediction enhances accuracy in multi-year wheat breeding trials. Gene Genom Genet 11(10):270. https://doi.org/10.1093/g3journal/jkab270

Moore JK, Manmathan HK, Anderson VA, Poland JA, Morris CF, Haley SD (2017) Improving genomic prediction for pre-harvest sprouting tolerance in wheat by weighting large-effect quantitative trait loci. Crop Sci 57(3):1315–1324. https://doi.org/10.2135/cropsci2016.06.0453

Nakamura S, Abe F, Kawahigashi H, Nakazono K, Tagiri A, Matsumoto T, Utsugi S, Ogawa T, Handa H, Ishida H, Mori M (2011) A wheat homolog of MOTHER OF FT AND TFL1 acts in the regulation of germination. Plant Cell 23(9):3215–3229. https://doi.org/10.1105/tpc.111.088492

Nyachiro JM (2012) Pre-harvest sprouting in cereals. Euphytica 188(1):1–5. https://doi.org/10.1007/s10681-012-0779-9

Odilbekov F, Armonienė R, Koc A, Svensson J, Chawade A (2019) GWAS-assisted genomic prediction to predict resistance to Septoria tritici blotch in Nordic winter wheat at seedling stage. Front Plant Sci 10:1224. https://doi.org/10.3389/fgene.2019.01224

Ogbonnaya FC, Imtiaz M, Ye G, Hearnden PR, Hernandez E, Eastwood RF, Van Ginkel M, Shorter SC, Winchester JM (2008) Genetic and QTL analyses of seed dormancy and preharvest sprouting resistance in the wheat germplasm CN10955. Theor Appl Genet 116(7):891–902. https://doi.org/10.1007/s00122-008-0712-8

Osakabe Y, Maruyama K, Seki M, Satou M, Shinozaki K, Yamaguchi-Shinozaki K (2005) Leucine-rich repeat receptor-like kinase1 is a key membrane-bound regulator of abscisic acid early signaling in Arabidopsis. Plant Cell 17(4):1105–1119. https://doi.org/10.1105/tpc.104.027474

Park J, Lim CJ, Shen M, Park HJ, Cha JY, Iniesto E, Rubio V, Mengiste T, Zhu JK, Bressan RA, Lee SY (2018) Epigenetic switch from repressive to permissive chromatin in response to cold stress. Proc Natl Acad Sci 5:115(23):E5400–9. https://doi.org/10.1073/pnas.1721241115

Pérez P, de los Campos G (2014) Genome-wide regression and prediction with the BGLR statistical package. Genetics 1:198(2):483–495. https://doi.org/10.1534/genetics.114.164442

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, De Bakker PI, Daly MJ, Sham PC (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 1:81(3):559–575. https://doi.org/10.1086/519795

R Core Team (2020) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.r-project.org/. Accessed 12 Mar 2021

Rasul G, Humphreys DG, Brûlé-Babel A, McCartney CA, Knox RE, DePauw RM, Somers DJ (2009) Mapping QTLs for pre-harvest sprouting traits in the spring wheat cross 'RL4452/AC Domain'. Euphytica 168(3):363–378. https://doi.org/10.1007/s10681-009-9934-3

Ramirez-Gonzalez RH, Uauy C, Caccamo M (2015) PolyMarker: a fast polyploid primer design pipeline. Bioinformatics 31(12):2038–2039. https://doi.org/10.1093/bioinformatics/btv069

Ranjan R, Yadav R, Jain N, Sinha N, Bainsla NK, Gaikwad KB, Kumar M (2021) Epistatic QTLs play a major role in nitrogen use efficiency and its component traits in indian spring wheat. Agriculture 11(11):1149. https://doi.org/10.3390/agriculture11111149

Reif JC, Maurer HP, Korzun V, Ebmeyer E, Miedaner T, Würschum T (2011) Mapping QTLs with main and epistatic effects underlying grain yield and heading time in soft winter wheat. Theoret Appl Genet 123(2):283–292. https://doi.org/10.1007/s00122-011-1583-y

Rikiishi K, Sugimoto M, Maekawa M (2021) Transcriptomic analysis of developing seeds in a wheat (*Triticum aestivum* L.) mutant RSD32 with reduced seed dormancy. Breed Sci 71(2):155–166. https://doi.org/10.1270/jsbbs.20016

Rutkoski J, Poland J, Mondal S, Autrique E, Pérez LG, Crossa J, Reynolds M, Singh R (2016) Canopy temperature and vegetation indices from high-throughput phenotyping improve accuracy of pedigree and genomic selection for grain yield in wheat. Gene Genom Genet 1:6(9):2799–808. https://doi.org/10.1534/g3.116.032888

Sandhu KS, Patil SS, Pumphrey M, Carter A (2021a) Multitrait machine- and deep-learning models for genomic selection using spectral information in a wheat breeding program. Plant Genome 14(3):e20119. https://doi.org/10.1002/TPG2.20119

Sandhu KS, Aoun M, Morris CF, Carter AH (2021b) Genomic selection for end-use quality and processing traits in soft white winter wheat breeding program with machine and deep learning models. Biology 10(7):689. https://doi.org/10.3390/biology10070689

Sandhu KS, Lozada DN, Zhang Z, Pumphrey MO, Carter AH (2021c) Deep learning for predicting complex traits in spring wheat breeding program. Front Plant Sci 11:2084. https://doi.org/10.3389/fpls.2020.613325

Sandhu KS, Mihalyov PD, Lewien MJ, Pumphrey MO, Carter AH (2021d) Combining genomic and phenomic information for predicting grain protein content and grain yield in spring wheat. Front Plant Sci 12:170. https://doi.org/10.3389/fpls.2021.613300

Sandhu KS, Merrick LF, Sankaran S, Zhang Z, Carter AH (2022) Prospectus of genomic selection and phenomics in cereal, legume and oilseed breeding programs. Front Genet 12:829131. https://doi.org/10.3389/fgene.2021.829131

Sandhu KS, Patil SS, Aoun M, Carter AH (2022b) Multi-trait multi-environment genomic prediction for end-use quality traits in winter wheat. Front Genet 13:831020. https://doi.org/10.3389/fgene.2022.831020

Sehgal D, Autrique E, Singh R, Ellis M, Singh S, Dreisigacker S (2017) Identification of genomic regions for grain yield and yield stability and their epistatic interactions. Sci Rep 7:41578. https://doi.org/10.1038/srep41578

Shahi D, Guo J, Pradhan S, Khan J, Avci M, Khan N, McBreen J, Bai G, Reynolds M, Foulkes J, Babar MA (2022) multi-trait genomic prediction using in-season physiological parameters increases prediction accuracy of complex traits in US wheat. BMC Genom 23(1):1–3. https://doi.org/10.1186/s12864-022-08487-8

Shao M, Bai G, Rife TW, Poland J, Lin M, Liu S, Chen H, Kumssa T, Fritz A, Trick H, Li Y (2018) QTL mapping of pre-harvest sprouting resistance in a white wheat cultivar Danby. Theor Appl Genet 131(8):1683–1697. https://doi.org/10.1007/s00122-018-3107-5

Shorinola O, Balcárková B, Hyles J, Tibbits JF, Hayden MJ, Holušova K, Valárik M, Distelfeld A, Torada A, Barrero JM, Uauy C (2017) Haplotype analysis of the pre-harvest sprouting resistance locus Phs-A1 reveals a causal role of TaMKK3-A in global germplasm. Front Plant Sci 13(8):1555. https://doi.org/10.3389/fpls.2017.01555

Shorinola O, Bird N, Simmonds J, Berry S, Henriksson T, Jack P, Werner P, Gerjets T, Scholefield D, Balcárková B, Valárik M (2016) The wheat Phs-A1 pre-harvest sprouting resistance locus delays the rate of seed dormancy loss and maps 0.3 cM distal to the PM19genes in UK germplasm. J Exp Bot 67(14):4169–4178. https://doi.org/10.1093/jxb/erw194

Singh AK, Knox RE, Clarke JM, Clarke FR, Singh A, DePauw RM, Cuthbert R (2014) Genetics of pre-harvest sprouting resistance in a cross of Canadian adapted durum wheat genotypes. Mol Breed 33(4):919–929. https://doi.org/10.1007/s11032-013-0006-y

Singh K, Batra R, Sharma S, Saripalli G, Gautam T, Singh R, Pal S, Malik P, Kumar M, Jan I, Singh S (2021) WheatQTLdb: a QTL database for wheat. Mol Genet Genom 296(5):1051–1056. https://doi.org/10.1007/s00438-021-01796-9

Singh L, Anderson JA, Chen J, Gill BS, Tiwari VK, Rawat N (2019) Development and validation of a perfect KASP marker for Fusarium Head Blight resistance gene *Fhb1* in wheat. Plant Pathol J 35(3):200–207

Singh R, Saripalli G, Kumar A, Gautam T, Singh SK, Gahlaut V, Kumar S, Meher PK, Mishra RP, Singh VK, Sharma PK, Balyan HS, Gupta PK (2023) QTL analysis for nitrogen use efficiency in wheat (*Triticum aestivum* L.). Euphytica. 219(1):1–22

Sohn SI, Pandian S, Kumar TS, Zoclanclounon YA, Muthuramalingam P, Shilpha J, Satish L, Ramesh M (2021) Seed dormancy and pre-harvest sprouting in Rice—An updated overview. Int J Mol Sci 22(21):118040. https://doi.org/10.3390/ijms222111804

Tai L, Wang HJ, Xu XJ, Sun WH, Ju L, Liu WT, Li WQ, Sun J, Chen KM (2021) Pre-harvest sprouting in cereals: Genetic and biochemical mechanisms. J Exp Bot 2;72(8):2857–2876. https://doi.org/10.1093/jxb/erab024

Tibshirani R (1996) Regression shrinkage and selection via the lasso. J R Stat Soc Ser B (Methodol) 58(1):267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

Torada A, Koike M, Ikeguchi S, Tsutsui I (2008) Mapping of a major locus controlling seed dormancy using backcrossed progenies in wheat (*Triticum aestivum* L.). Genome 51(6):426–432. https://doi.org/10.1139/G08-007

Torada A, Koike M, Ogawa T, Takenouchi Y, Tadamura K, Wu J, Matsumoto T, Kawaura K, Ogihara Y (2016) A causal gene for seed dormancy on wheat chromosome 4A encodes a MAP kinase kinase. Curr Biol 21;26(6):782–787. https://doi.org/10.1016/j.cub.2016.01.063

Tyagi S, Gupta PK (2012) Meta-analysis of QTLs involved in pre-harvest sprouting tolerance and dormancy in bread wheat. Triticeae Genomics and Genetics 5;3(1). https://doi.org/10.5376/tgg.2012.03.0002

Tyagi S, Mir RR, Balyan HS, Gupta PK (2015) Interval mapping and meta-QTL analysis of grain traits in common wheat (*Triticum aestivum* L.). Euphytica 201(3):367–380. https://doi.org/10.1007/s10681-014-1217-y

VanRaden PM (2008) Efficient methods to compute genomic predictions. Int J Dairy Sci 1;91(11):4414–23. https://doi.org/10.3168/jds.2007-0980

Voorrips R (2002) MapChart: software for the graphical presentation of linkage maps and QTLs. J Hered 1;93(1):77–78. https://doi.org/10.1093/jhered/93.1.77

Wang Q, Tian F, Pan Y, Buckler ES, Zhang Z (2014) A SUPER powerful method for genome wide association study. PloS One 23;9(9):e107684. https://doi.org/10.1371/journal.pone.0107684

Wang X, Xu Y, Hu Z, Xu C (2018) Genomic selection methods for crop improvement: Current status and prospects. Crop J 1;6(4):330–340. https://doi.org/10.1016/j.cj.2018.03.001

Wang D, Pang Y, Dong L, Li A, Kong L, Liu S (2020a) Allelic impacts on pre-harvest sprouting resistance and favorable haplotypes in *TaPHS1* of Chinese wheat accessions. Crop J 8(4):515–521

Wang Z, Ren Z, Cheng C, Wang T, Ji H, Zhao Y, Deng Z, Zhi L, Lu J, Wu X, Xu S (2020b) Counteraction of ABA-mediated inhibition of seed germination and seedling establishment by ABA signaling terminator in Arabidopsis. Mol Plant 13(9):1284–1297. https://doi.org/10.1016/j.molp.2020.06.011

Warner RL, Kudrna DA, Spaeth SC, Jones SS (2000) Dormancy in white-grain mutants of Chinese Spring wheat (*Triticum aestivum* L.). Seed Sci Res 10(1):51–60. https://doi.org/10.1017/S0960258500000064

Xiao C, Liu Y, Chen W, Yang J, Cheng M, Watt C, Cheng J, Wang Z, Tan Z, Li M, Wang J (2021) Characterization and expression quantitative trait loci analysis of *TaABI4*, a pre-harvest sprouting related gene in wheat. Seed Sci Res 31(3):188–198. https://doi.org/10.1017/S0960258521000015

Yu LX, Chao S, Singh RP, Sorrells ME (2017) Identification and validation of single nucleotide polymorphic markers linked to Ug99 stem rust resistance in spring wheat. PLoS One 27;12(2):e0171963

Yu LX, Lorenz A, Rutkoski J, Singh RP, Bhavani S, Huerta-Espino J, Sorrells ME (2011) Association mapping and gene–gene interaction for stem rust resistance in CIMMYT spring wheat germplasm. Theor Appl Genet 123(8):1257–1268. https://doi.org/10.1007/s00122-011-1664-y

Zhang Z, Ersoz E, Lai CQ, Todhunter RJ, Tiwari HK, Gore MA, Bradbury PJ, Yu J, Arnett DK, Ordovas JM, Buckler ES (2010) Mixed linear model approach adapted for genome-wide association studies. Nat Genet 42(4):355–360. https://doi.org/10.1038/ng.546

Zheng Y, Schumaker KS, Guo Y (2012) Sumoylation of transcription factor MYB30 by the small ubiquitin-like modifier E3 ligase SIZ1 mediates abscisic acid response in *Arabidopsis thaliana*. Proc Natl Acad Sci 31;109(31):12822–12827. https://doi.org/10.1073/pnas.120263010

Zhou K, Yang J, Wang ZX, Wang JR (2017a) Sequence analysis and expression profiles of *TaABI5*, a pre-harvest sprouting resistance gene in wheat. Gene Genom 39(2):161–171. https://doi.org/10.1007/s13258-016-0483-6

Zhou Y, Tang H, Cheng MP, Dankwa KO, Chen ZX, Li ZY, Gao S, Liu YX, Jiang QT, Lan XJ, Pu ZE (2017b) Genome-wide association study for pre-harvest sprouting resistance in a large germplasm collection of Chinese wheat landraces. Front Plant Sci 6(8):401. https://doi.org/10.3389/fpls.2017.00401

Zuo J, Lin CT, Cao H, Chen F, Liu Y, Liu J (2019) Genome-wide association study and quantitative trait loci mapping of seed dormancy in common wheat (*Triticum aestivum* L.). Planta 250(1):187–198. https://doi.org/10.1007/s00425-019-03164-9