RESEARCH ARTICLE

GWAS for main effects and epistatic interactions for grain morphology traits in wheat

Parveen Malik[1] · Jitendra Kumar[1,2] · Shiveta Sharma[1] · Prabina Kumar Meher[3] ·
Harindra Singh Balyan[1] · Pushpendra Kumar Gupta[1] · Shailendra Sharma[1]

Abstract In the present study in wheat, GWAS was conducted for identification of marker trait associations (MTAs) for the following six grain morphology traits: (1) grain cross-sectional area (GCSA), (2) grain perimeter (GP), (3) grain length (GL), (4) grain width (GWid), (5) grain length–width ratio (GLWR) and (6) grain form-density (GFD). The data were recorded on a subset of spring wheat reference set (SWRS) comprising 225 diverse genotypes, which were genotyped using 10,904 SNPs and phenotyped for two consecutive years (2017–2018, 2018–2019). GWAS was conducted using five different models including two single-locus models (CMLM, SUPER), one multi-locus model (FarmCPU), one multi-trait model (mvLMM) and a model for Q x Q epistatic interactions. False discovery rate (FDR) [P value - $\log10(p) \geq 5$] and Bonferroni correction [P value - $\log10(p) \geq 6$] (corrected p value < 0.05) were applied to eliminate false positives due to multiple testing. This exercise gave 88 main effect and 29 epistatic MTAs after FDR and 13 main effect and 6 epistatic MTAs after Bonferroni corrections. MTAs obtained after Bonferroni corrections were further utilized for identification of 55 candidate genes (CGs). In silico expression analysis of CGs
in different tissues at different parts of the seed at different developmental stages was also carried out. MTAs and CGs identified during the present study are useful addition to available resources for MAS to supplement wheat breeding programmes after due validation and also for future strategic basic research.

Introduction

Bread wheat (*Triticum aestivum* L.) is a major staple crop all over the world, which provides food for ∼ 36% of the world's population involving ∼ 20% calories in human diet (http://faostat.fao.org; Maulana et al. 2018). In this crop, as also in other cereals, the grain size is known to be associated with various characteristics of flour (e.g.. hydrolytic enzymes activity), which in turn control baking quality and end-use suitability, the protein content and grain yield (Evers 2000; Breseghello and Sorrells 2007; Gegas et al. 2010). The larger grains also have a positive effect on seedling vigour, market preference, grain yield and flour yield characteristics (Chastain et al. 1995; Gan and Stobbe 1996). The grain size and shape (including length, width, perimeter of the grain, etc.) did not receive the desired attention for the improvement of yield in the current wheat breeding programmes (Kovach et al. 2007). However, significant phenotypic and genetic variation for grain size and grain weight does occur in different *Triticum* species and can be exploited for improvement of grain morphology, indirectly leading to an improvement in grain yield (Gegas et al. 2010; Jing et al. 2007; Rasheed et al. 2014). However, genetics of grain characteristics in wheat

✉ Shailendra Sharma
  Shgjus6@gmail.com

[1] Department of Genetics and Plant Breeding, Ch. Charan Singh University, Meerut, U.P 250 004, India

[2] Department of Biotechnology, National Agri-Food Biotechnology Institute (NABI), Govt. of India, Sector 81 (Knowledge City), S.A.S. Nagar, Mohali, Punjab 140306, India

[3] Centre for Agricultural Bioinformatics, ICAR-Indian Agricultural Statistics Research Institute, New Delhi 110012, India

did not receive as much attention as it did in case of rice, where > 30 genes for grain characteristics have been cloned and characterized (Zheng et al. 2015; Hu et al. 2015; Yu et al. 2017; Jiang et al. 2019; Yang et al. 2019).

The genetic studies in wheat and other cereals already undertaken suggest that grain shape and size are complex quantitative traits (with their component traits), each controlled by a number of major and minor genes. For the study of genetics of such complex traits, interval mapping and GWAS are two important approaches, each with its own merits and demerits. In recent years, GWAS has been preferred, since it utilize a wide range of genetic variation generated through hundreds of cycles of recombination. A beginning of GWAS for the study of genetic architecture of quantitative traits in animals (mainly humans) and plant systems was made in early years of the present century, when relative merits of LD-based association mapping (AM) over the linkage-based interval mapping was documented (for reviews, see Huang et al. 2013; Gupta et al. 2014, 2019; Cortes et al. 2021). The major breakthough, however, took place, when a mixed linear model involving population structure and relatedness was proposed (Yu et al. 2006) and the software TASSEL was made available (Bradbury et al. 2008). This was soon followed by further improvement of GWAS through a series of models proposed during 2010–2020, a process that still continues.

A number of studies involving interval mapping and GWAS are available for genetics of traits related to grain shape and size (Breseghello and Sorrels 2007; Sun et al. 2009; Okamoto et al. 2013; Rasheed et al. 2014; Li et al. 2015; Yan et al. 2017; Kumari et al. 2018; Ma et al. 2019; Wang et al. 2019; Yu et al. 2019; Alemu et al. 2020; Sun et al. 2020; Xin et al. 2020). In most of these genetic studies, a limited number of mapping populations and association panels have been used (Breseghello and Sorrels 2007; Sun et al. 2009; Rasheed et al. 2014). Also, a majority of QTLs that have been reported for grain weight are not suitable for breeding purpose, since these are not stable across environments, explain limited phenotypic variability, and are involved in epistatic interactions (Campbell et al. 2003; Gupta et al. 2007; Prashant et al. 2012; Patil et al. 2013; Cabral et al. 2018). This warrants further studies for detection of additional MTAs using novel germplasm, which has not so far been used for the study of grain morphology traits in wheat.

Most early GWA studies including the above studies, conducted in a number of crops, involved only single locus, single trait analysis and did not involve analysis of epistatic interactions. Although occurrence of epistasis was initially recognized in early years of the last century (Bateson 1909), its importance and methods for detection and estimation, particularly for quantitative traits, have largely been recognized in the present century (Holland 2001; Lu et al. 2011; Ritchie and Van Steen 2018). It is now known that both MTAs/QTLs with main effect and those with no main effect are generally involved in epistatic interactions (Niel et al. 2015; Ritchie and Van Steen 2018; Slim et al. 2020). In some GWA studies also, epistatic interactions have been examined, and significant epistatic interactions have been reported for several traits in different crops including wheat (Mackay 2014; Moellers et al. 2017; Sehgal et al. 2017, 2020).

Keeping the above in view, the present study was planned, which involved the use of the following four improved mixed linear models: (1) Compressed Mixed Linear Model (CMLM;Zhang et al. 2010), (2) Fixed and random model Circulating Probability Unification (Farm-CPU; Liu et al. 2016), (3) Settlement of MLM Under Progressively Exclusive Relationship (SUPER; Wang et al. 2014), and (4) Multi-trait analysis matrix variate linear mixed model (mvLMM; Korte et al. 2012). These models allowed multi-locus and multi-trait analysis and also allowed identification and measurements of epistatic interactions. Hopefully, the results of the present study will prove useful for developing wheat cultivars, through MAS/MARS, with improved grain/flour quality, high market value and grain yield.

Materials and methods

GWAS panel and genotyping details

Originally, 330 spring wheat genotypes belonging to spring wheat reference set (SWRS) were obtained from CIMMYT, Mexico. However, in the present study on GWAS a subset comprising only 225 diverse wheat genotypes was used. The geographical distribution of these genotypes is shown in Fig. 1 and pedigree information for each genotype is available in Table S1. Genotyping data was retrieved from the original data on the entire set of 330 SWRS genotypes that were genotyped using DArT-seq (outsourced by CIMMYT to Diversity Array Technology Pvt. Ltd, Australia, under their "Seed for Discovery" project). The genotypic data for 10,904 SNPs out of 17,937 SNP markers for the whole set of 330 genotypes was available for the subset of 225 genotypes used during the present study, as also in a previous study (Kumar et al. 2018).

The GWAS panel of 225 genotypes was raised during rabi-season in a simple lattice design with two replications at the Research Farm of the Department of Genetics and Plant Breeding, Ch. Charan Singh University, Meerut (location coordinates: 28.984644°N and 77.705956°E) over two consecutive years representing two different environments (E1; 2017–18 and E2; 2018–19). Each genotype was

Fig. 1 Country of origin of the wheat genotypes comprising association mapping panel used during the present study
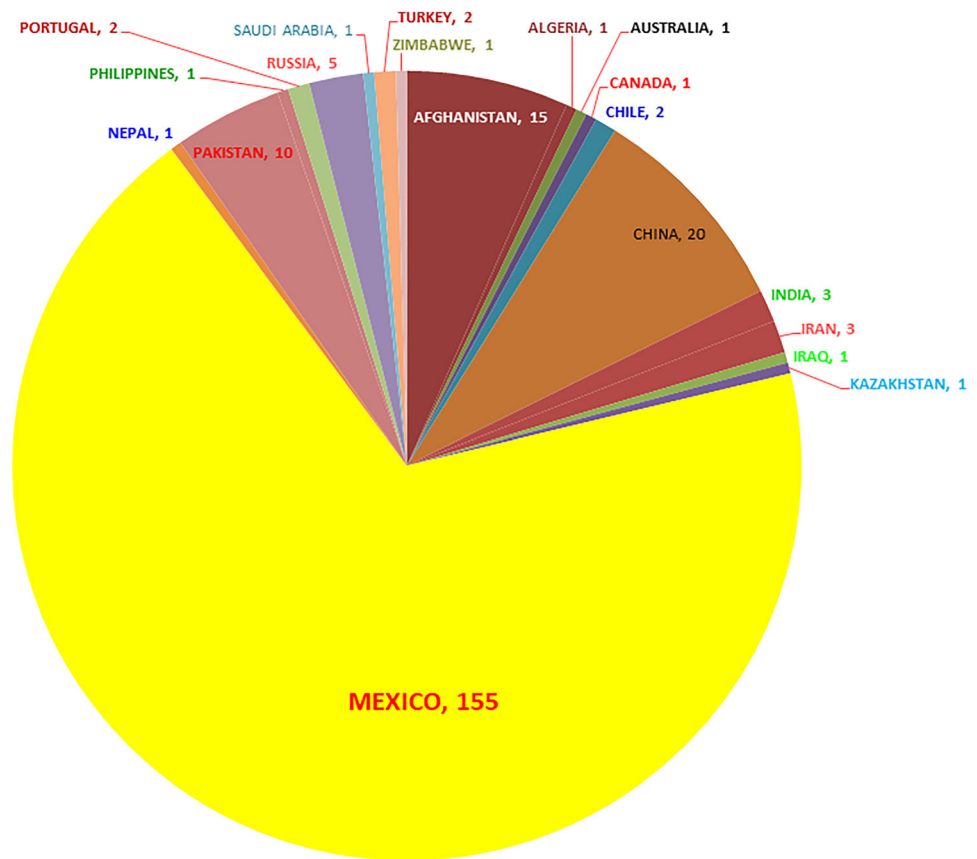
represented by a plot of 3 rows of 1.5 m each, with a row to row distance of 0.25 m. The total number of blocks were 15 with each block containing 45 rows i.e. three rows of each genotype. Normal cultural practices including fertilizer application (i.e., 200 kg/ha fertilizer; N:P:K = 8:8:8) and irrigation were followed.

Data on grain morphology traits

The data on six grain morphology traits were collected using 24 grains (picked up randomly) for each of the 225 genotypes per replication using *SmartGrain* software ver. 1.2 (Tanabata et al. 2012). *SmartGrain* software makes use of grain images to record data on grain morphology, using all the grains within a digital image; it detects outlines, and then estimates all grain size parameters viz, cross-sectional area (CSA), grain perimeter (GP), grain length (GL), grain width (GWid), grain length–width ratio (GLWR) (Fig. 2). Thousand grain weight (TGW) in grams (g) was also measured by weighing 1000-grains of each of the 225 genotypes in each replication. Another parameter (grain form-density; GFD) was calculated with the help of three different parameters (TGW, GL & GWid) using the following formula: GFD $= \frac{TGW}{GL \times GWid}$, which determines variation in grain weight that is not accounted for by the differences in grain length and weight (Giura and Saulescu 1996). For each trait, average values obtained for all 24 grains were utilized for further analysis.

Statistical analysis

The violin plots were prepared to depict the distribution of phenotypic data for all the six traits for each of the two individual environments (E1 and E2) and also using BLUP values (B). The BLUP values were generated using the '*lme4*' package in the R programme (Bates et al. 2015). Pearson's correlation coefficients were estimated using R package *Performance Analytics* (Peterson et al. 2018) and ANOVA was conducted using AMMI (additive main effects and multiplicative interactions) available in *Agricolae* R package (Mendiburu and Yaseen 2020). Broad sense heritability (H^2) was calculated as the ratio between genotypic variance (σ^2_g) and phenotypic variance (σ^2_p) using mathematical formula; $H^2 = \frac{\sigma 2g}{\sigma 2p} \times 100$ by Microsoft Excel 2010 (Allard 1999).

Fig. 2 Images of seeds showing method of recording phenotypic data: (1) 24 seeds (left panel) from one genotype, (2) single whole wheat seed (middle panel), and (3) flowchart for estimation of seed morphology (right panel) using *SmartGrain* software. Grain morphological traits in the middle panel: (1) GCSA = grain cross-sectional area; shaded region of grain; (2) GP = grain perimeter in red line; (3) GL = grain length as yellow line, (4) Gwid = grain width as green line
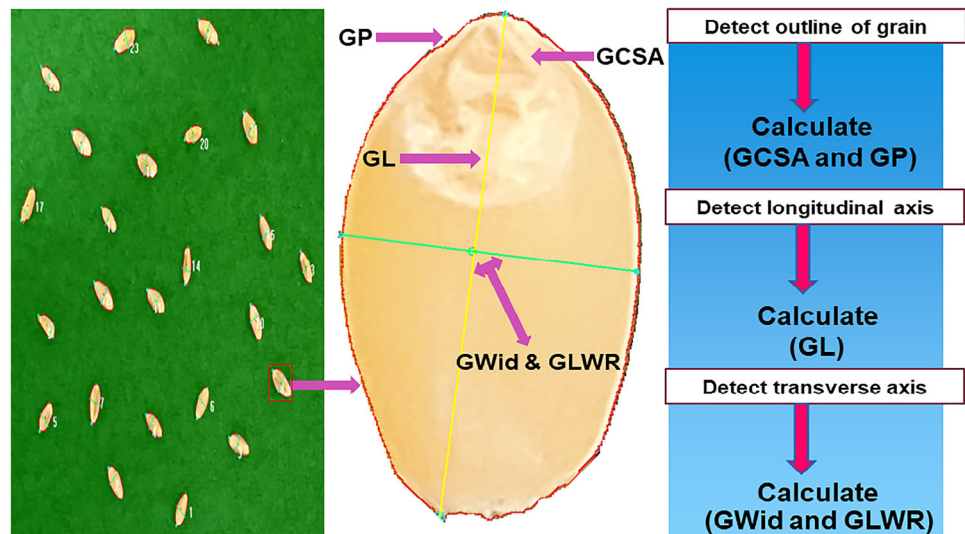
Population structure using principal component (PC) analysis and, kinship matrix

The Q matrix based on PCs and kinship matrix based on relatedness among markers were automatically generated using default set of parameters in GWAS models (Van-Raden 2008; Lipka et al. 2012). These were used to perform GWAS using software for CMLM, FarmCPU and SUPER available in the package 'Genomic Association and Prediction Integrated Tool (GAPIT version 2.0; Tang et al 2016) in R programme; mvLMM for multi-trait analysis was perormed using the software GEMMA (Zhou and Stephens 2012).

Marker-trait association (MTAs)

Main effect MTAs were identified using three different models, namely CMLM, FarmCPU and SUPER using software available in GAPIT version 2.0 using R programme (Tang et al. 2016). Matrix variate linear mixed model (mvLMM) for multi-trait analysis was used for the identification of SNPs associated with more than two traits (Furlotte and Eskin 2015). The epistatic interactions were identified using R package *SNPassoc*, where the function *interactionPval* was executed for computing the statistical significance of paired SNP–SNP interactions (Gonzalez et al. 2007). Main effect MTAs and those involved in epistatic interactions, identified by different methods, were independently subjected to FDR (− log p value = 5.0) and Bonferroni correction to address the problem of multiple testing; very few MTAs were obtained using Bonferroni correction (− log p value = 6.0); therefore, MTAs after applying FDR were used for further analysis. MTAs obtained after Bonferroni correction, were also used for identification of candidate genes (CGs) (see later). The

chromosomal locations and positions of unmapped markers associated with traits were identified using BLASTn from IWGSC RefSeq v1.0 Ensembl Plants using nucleotide sequences (SNP tags) of SNPs with 100% identities (https://plants.ensembl.org/Triticum_aestivum/Info/Index).

Comparison of MTAs with known QTLs

The MTAs obtained in the present study were compared with known QTLs/MTAs of each trait. For this purpose, physical positions of known MTAs were obtained utilizing the data available in Ensemble Plant (version 50; https://plants.ensembl.org/Triticum_aestivum/Info/Index).

Identification of putativeCGs and their in silico expression analysis

For identification of CGs, only the MTAs which qualified Bonferroni correction, were utilized. For this purpose, SNP tags associated with MTAs were aligned with reference genome version IWGSC1.0 (IWGSC 2018), available at Ensembl Plants (http://www.ensembl.org/info/docs/tools/vep/index.html) and 200 kb window (100 kb on either side of SNP) was obtained for identification of CGs. The gene ontology (GO) annotation information of all these CGs was extracted from the IWGSC website (http://www.wheatgenome.org/). The RNA-seq expression data from Wheat Expression Browser was used for *in-silico* gene expression analysis (http://www.wheat-expression.com/). The expression of each CGs in different parts of the seed and at different developmental stages was examined and presented in the form of a heatmap.

Homology-based mining of genes

Protein sequences of known rice genes, for the traits of interest, were retrieved from NCBI database (https://www.ncbi.nlm.nih.gov/), and used against wheat genome sequences (IWGSC RefSeq v1.0) available at Ensembl Plants. Orthologous wheat genome sequences were identified using BLASTx search. From these wheat genome sequences, the sequences which were located within the CGs were used for further study. The rice genes and their functions were identified using oryzabase database of rice (https://shigen.nig.ac.jp/rice/oryzabase/).

Results

### Frequency distribution, ANOVA, heritability and correlations

The frequency distributions of phenotypic data for all the six traits based on BLUP values (Fig. 3) and for each of the two environments are presented in the form of violin plots (Fig. S1). The results of ANOVA for all the six traits are presented in Table 1. Genotypic differences were significant for all the six traits, but G X E interactions were significant for only four (GCSA, GP, GL, and GFD) of the

six traits. Broad sense heritability (H^2) ranged from 41% (GWid) to 89% (GCSA) (Table 1). All 21 pairwise Pearson's correlations involving all six grain morphology traits and TGW for all the three sets of data including BLUP and two environments are provided in Fig S2. Significant correlations included 12 positive and 7 negative correlations in E1 (Fig. S2), 8 positive and 4 negative correlations in E2 (Fig. S3), and 11 positive and 7 negative correlations using BLUP values (Fig. 4). Significant positive correlations were also available between TGW andGCSA in all the three sets of data (Fig. 4, Fig. S2 & S3).

Population structure using principal component analysis (PCA)

Population structure was worked out using principle component analysis (PCA), where the first three PCs produced a 3D scatter plot showing distribution of genotypes into sub-groups. The PCs divided the whole association panel into three sub-groups with variation within sub-groups ranging from 3.44% (PCA3) to 6% (PCA 1) (Malik et al. 2021). These three subgroups represented the population structure and were used for development of Q matrix for GWAS.

Fig. 3 Violin plots showing the frequency distribution of BLUP values for six traits of grain morphology. Shaded regions of the violin plots represent the frequency distribution of data, in each case, the vertical solid bar indicates range of average values, and median is shown as white circle, depicting the lower, medium and upper quartile
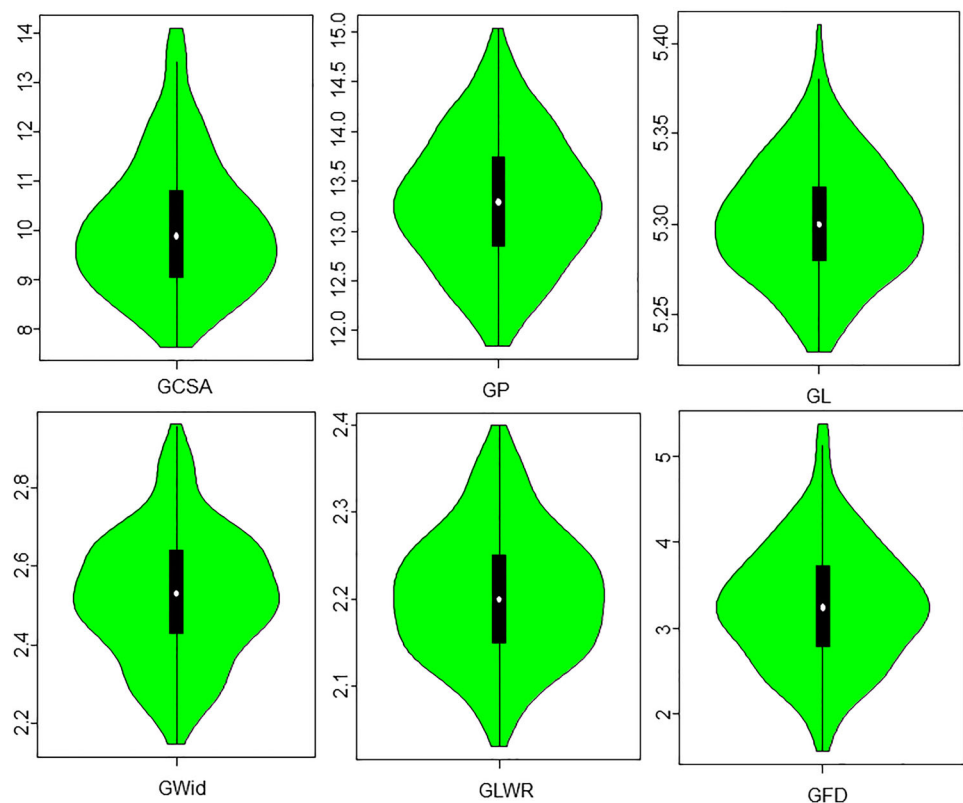
Table 1 Analysis of variance (ANOVA) and broad sense heritability (H^2) of wheat genotypes tested in two different environments (E1:2017–18 and E2:2018–19) with two replications in each environment

Source of variation	DF	GCSA	GP	GL	GWid	GLWR	GFD
Environments	1	15.27	87.66*	8.04***	1.35	0.05	46.35**
Replication within environments	3	6.66***	7.69***	0.03	0.59***	0.61***	0.85*
Genotypes	224	12.44***	4.89***	0.50***	0.16***	0.33***	2.47***
Environment x genotype	448	0.45**	0.49*	0.16*	0.07	0.06	1.28***
Broad sense heritability (H^2) %	89	73	42	41	56	63	

***0.1% level of significance; **1% level of significance; *5% level of significance, DF; degrees of freedom; GCSA; grain cross-sectional area, GP; grain perimeter, GL; grain length, GWid; grain width, GLWR; grain length–width ratio and GFD; grain form-density

Fig. 4 Pairwise Pearson's correlation coefficients among the six grain morphology traits and 1000-grain weight estimated using BLUP values of each trait
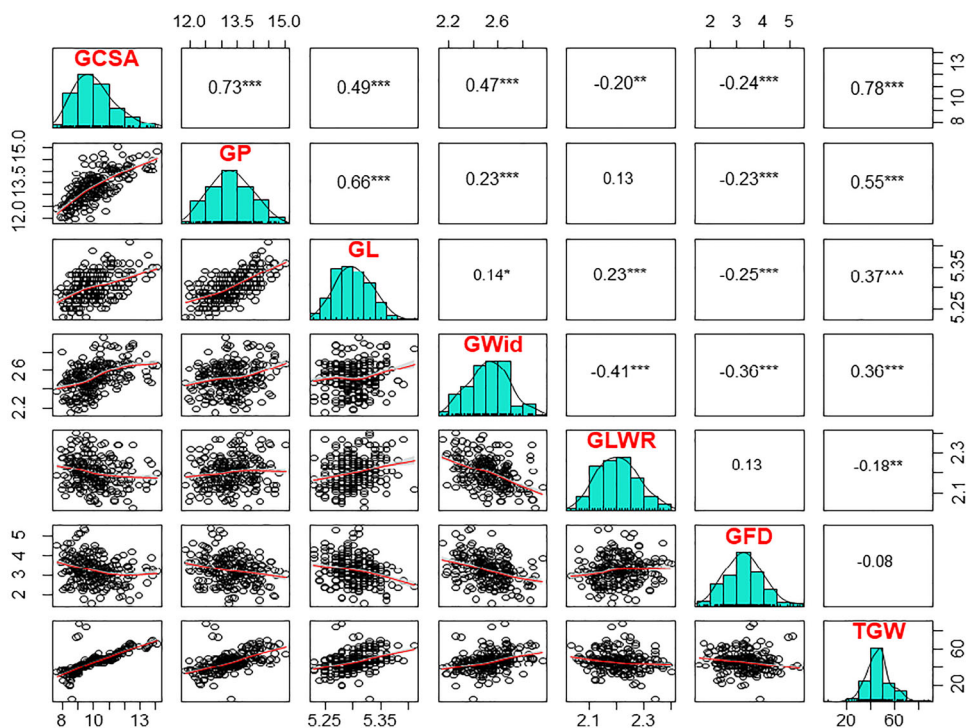
MTAs for two individual environments and BLUP values

The results of GWAS for individual traits involving two environments and BLUP and those for multi-traits will be described separately.

MTAs for two environments

Using data from two environments (E1 and E2), 824 MTAs for individual traits were identified using three models of GWAS (CMLM, FarmCPU; and SUPER; Table S2). After FDR correction, only 21 significant MTAs were identified. The number of trait-wise MTAs identified using each of the three models were as follows; (1) CMLM: 11 MTAs for three trait (Table 2 and Fig. 5), (2) FarmCPU: 9 MTAs for five traits (Table 3 and Fig. 6) and (3) SUPER: only 1 MTA for GL (SNP_3202: 3A). After Bonferroni correction, only one significant MTA (SNP_3202) for GL in E2 was available in each of the following three models: CMLM, FarmCPU, and SUPER methods; this MTA is located on chromosome 3A at 593,990,302–593,990,370 bp. Only MTAs obtained using FDR and Bonferroni correction were used for further study.

MTAs for BLUP data

Using BLUP values, 234 significant MTAs were identified; the distribution of trait-wise MTAs for three models were as follows: (1) CMLM: 65 MTAs for six traits; this number was reduced to 11 MTAs after FDR; only four MTAs qualified Bonferroni multiple correction for three traits (Table 2 & Fig. 5), (2) FarmCPU: 93 MTAs for six traits; only 10 MTAs after FDR for five traits, and only two MTAs after Bonferroni correction for two traits (Table 3 & Fig. 6), (3) SUPER: 76 MTAs were found for six traits, but

Table 2 Summary of significant MTAs detected in the two environments (E1, 2017–18 and E2, 2018–19) and BLUP values after false discovery rate (FDR) correction using CMLM

Trait	SNP	Allele	Chr	Pos. of SNP tag (bp)	R^2	Effect
E1						
GLWR	SNP_13927	G/A#	2B	658,485,687–658,485,737	0.12	−0.09
E2						
GL	SNP_3202*	A/G	3A	593,990,302–593,990,370	0.14	−0.28
	SNP_1476	A/C	6B	19,922,669–19,922,737	0.12	−0.22
	SNP_8572	A/G	5B	4,832,008–4,832,050	0.12	−0.13
	SNP_10532	G/C	4A	603,050,386–603,050,443	0.12	−0.23
	SNP_1716	C/T	2B	41,473,091–41,473,159	0.11	−0.19
	SNP_9406	G/A	3B	36,470,465–36,470,533	0.11	0.11
	SNP_4632	T/C	7A	44,352,216–44,352,284	0.11	0.21
GWid	SNP_2896	G/A	7A	77,195,120–77,195,188	0.1	−0.08
	SNP_330	T/A	2B	615,353,108–615,353,155	0.09	−0.1
	SNP_3275	T/C	1A	555,132,587–555,132,655	0.09	−0.07
BLUP						
GCSA	SNP_4563	A/C	2B	571,070,836–571,070,884	0.12	−0.61
GL	SNP_7441*	A/G	3A	731,844,353–731,844,410	0.12	0.01
	SNP_14066	T/C	2B	12,009,346–12,009,414	0.13	0.01
GP	SNP_12816	A/G	7B	647,426,653–647,426,721	0.13	−0.29
	SNP_13748	C/T	3A	715,576,111–715,576,179	0.13	0.2
	SNP_9771	G/T	7B	684,269,233–684,269,301	0.12	0.21
GWid	SNP_5401*	G/T	2A	79,343,336–79,343,404	0.1	0.05
	SNP_4648*	G/T	7A	94,143,005–94,143,073	0.1	0.12
GLWR	SNP_5518	G/A	1D	225,564,124–225,564,192	0.09	−0.05
	SNP_12288*	T/C	2B	683,768,477–683,768,545	0.08	−0.03
GFD	SNP_15086	G/A	7B	641,807,115–641,807,183	0.11	−0.22

E1; environment 1, E2; environment 2, *; Bonferroni multiple correction passed, #; desirable allele, Chr; chromosome, Pos; position

no MTA was detected after FDR or Bonferroni corrections for any trait. Two MTAs were also common in each of the following pairs involving BLUP, environments and models: BLUP-FarmCPU, E1-FarmCPU and E2-FarmCPU; these MTAs were also associated with GCSA and GFD, respectively (Table 3).

MTAs using multi-trait analysis

Multi-trait analysis, using mvLMM, gave 46 MTAs in 16 trait combinations (2–4 correlated traits) in both environments (E1 and E2) after FDR correction (Table 4), but after Bonferroni correction only four significant MTAs (out of 514 identified MTAs) for one trait combination (GL-GWid-GLWR-FFD) were available in each of the two environments. These four MTAs involved only two SNPs, one each on 2B and 6B (SNP_9744 on 2B at 800,127,644–800,127,712 bp and SNP_912 on 6B at 674,558,317–674,558,364 bp: Table 4).

MTAs involved in epistasis

MTAs involving 360 pairs of epistatic interactions involving all the six traits were also identified; after FDR correction, only MTAs for 29 epistatic interactions involving 47 markers (11 markers were common among the interactions) were available. The trait-wise interactions included a number of interactions for individual traits ranging from a minimum of one for GP to a maximum of 12 interactions for GFD (Table S3). After Bonferroni corrections, only three epistatic interactions remained significant, which involved two of the six traits (Table S3).

Comparison of MTAs with QTLs reported earlier

A total of 116 significant MTAs (21 CMLM + 19 FarmCPU + 1 SUPER + 46 mvLMM + 29 epistasis) were available for all the six traits after applying FDR. These significant MTAs were compared with the previously reported QTLs/MTAs for same traits. Only two MTAs one each for GP and GWid were detected in the flanking region
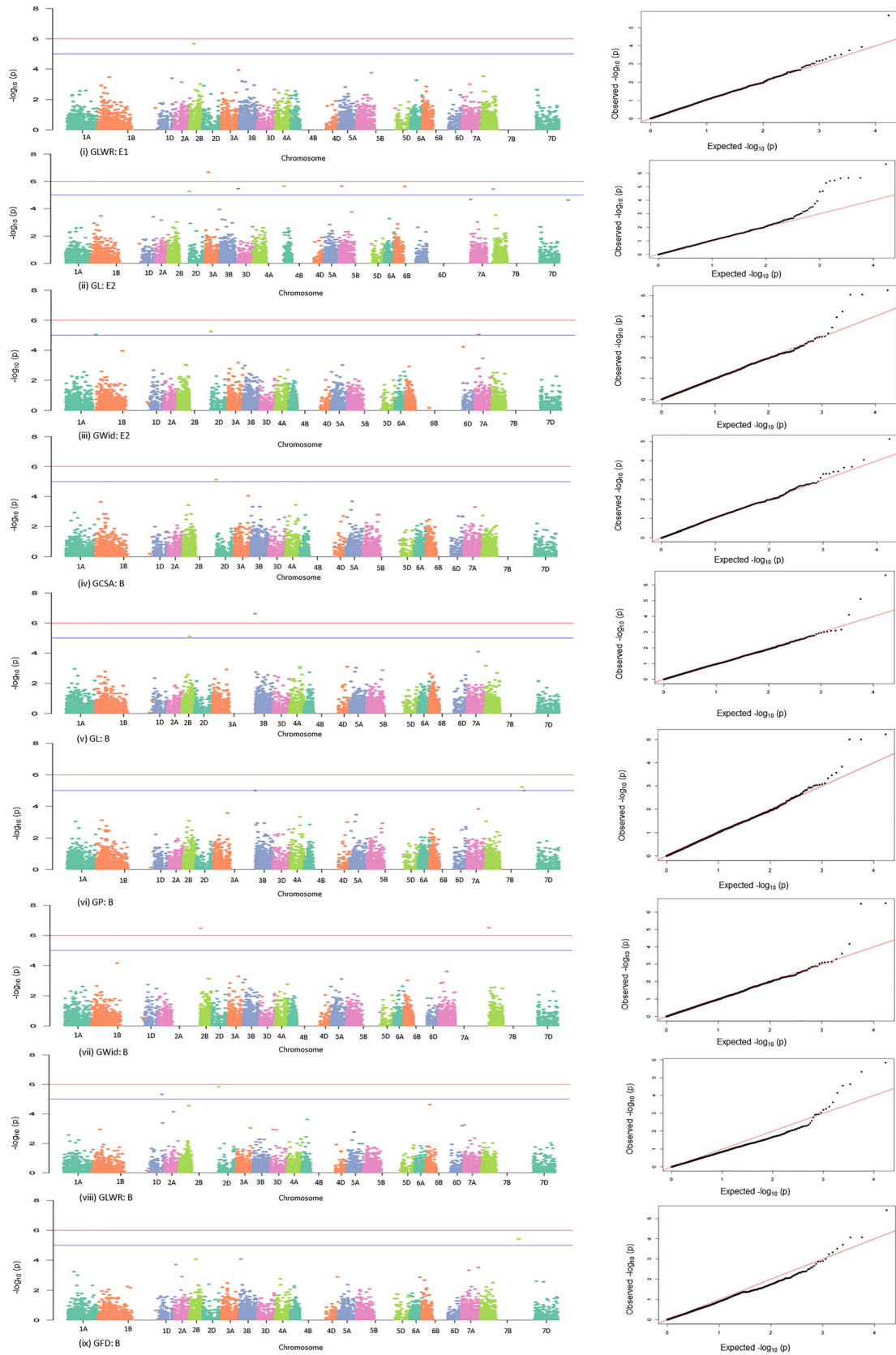
Manhattan plots (left side) showing significant marker-trait associations (MTAs) on 21 chromosomes and ump (represents unmapped MTAs) using CMLM with corrected *P*-value after FDR (blue line) and Bonferroni (red line) corrections in each case. Q-Q plots (right side) represent distribution of observed and expected *P* values for the same traits using environment 1 & 2 (E1 and E2) and BLUP values

of the reported QTL/MTAs. However, three MTAs were identified in the vicinity (< 50 Mb) of the reported QTL regions (Table 5).

Important MTAs for marker assisted selection (MAS)

MTAs identified in the present study, were subjected to further scrutiny in order to identify the most important MTAs, which could be recommended for MAS. MTAs were selected, which fulfilled at least one of the following criteria: (1) highest R^2, (2) lowest *P*-value, (3) availability in more than one models, (4) stability (identified in all the environments), (5) availability in earlier studies (including both interval mapping and GWAS). (6) qualified Bonferroni corrections, including those involved in epistatic interaction. Using these criteria, only 14 MTAs were short-listed, which all qualified Bonferroni and included 4 MTAs from CMLM, 1 MTA from FarmCPU, 2 MTAs from mvLMM, 6 MTAs from epistasis and 1 common MTA form CMLM, FarmCPU and SUPER (Table 6).
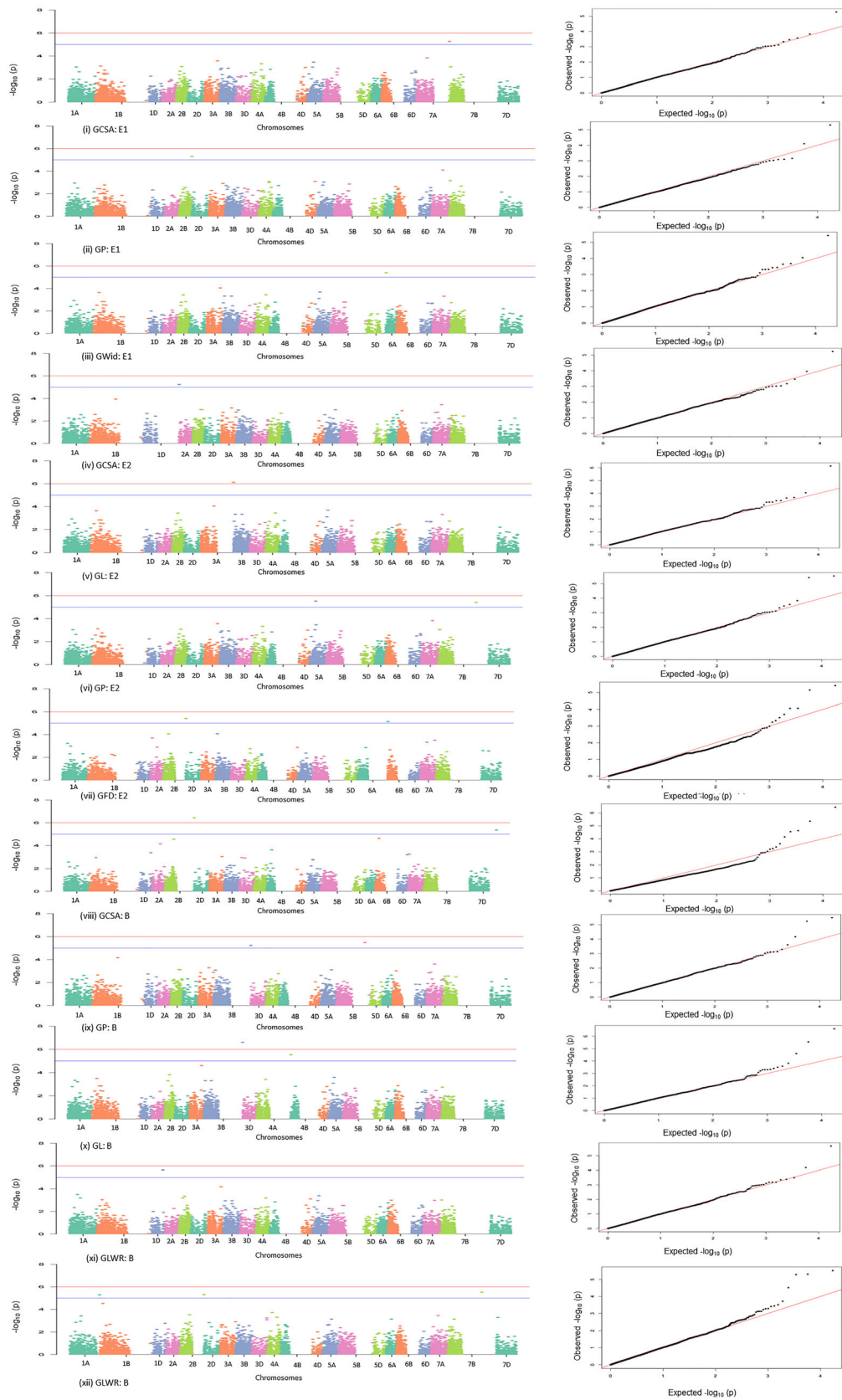
CGs based on GWAS-MTAs

Using GWAS-MTAs, which qualified Bonferroni correction, 25 CGs were available for 9 MTAs (all six traits were represented). Of these 25 CGs, two CGs were available for a solitary common MTA identified using each of the three models (CMLM, FarmCPU and SUPER); six CGs were also available for one MTA detected using mvLMM and 17 CGs were available from 6 MTAs involved in epistatic interactions. However no hit was detected for two MTAs (one each from mvLMM, and epistatic interaction; Table S4). Similarly, 30 CGs were identified using six MTAs obtained using BLUP values for the four traits (Table S4). CGs so detected were found to be involved in different biological (including grain related traits) and molecular functions. Some CGs encoded important proteins such as sucrose synthase (*TraesCS7B02G482200*: Table S4). These protein domains were associated either directly or indirectly with grain traits in wheat.

Table 3 A list of significant MTAs that qualify FDR correction detected in each of the two the environments (E1-2017–18 and E2-2018–19) using FarmCPU

Trait	SNP	Allele	Chr	Pos. of SNP tag (bp)	Effect
E1					
GCSA	SNP_2665	C/T#	7A	611,728,422–611,728,490	0.75
GP	SNP_4263	C/G	2B	215,951,648–215,951,716	− 0.65
GWid	SNP_7518	G/C	5D	368,249,809–368,249,877	− 0.09
E2					
GCSA	SNP_1597	C/T	1D	67,863,683–67,863,742	− 1.12
GL	SNP_3202*	A/G	3A	593,990,302–593,990,370	− 0.28
GP	SNP_4865	T/C	5A	10,724,684–10,724,752	0.76
	SNP_9771	T/G	7B	684,269,233–684,269,301	0.38
GFD	SNP_6775	T/A	2B	436,016,073–436,016,141	− 0.42
	SNP_10672	T/C	6A	602,752,656–602,752,703	0.42
BLUP					
GCSA	SNP_4563*	C/A#	2B	571,070,836–571,070,884	− 0.61
	SNP_2665	T/C	7D	531,439,752–531,439,820	0.57
GP	SNP_4865	T/C	5B	547,290,640–547,290,708	0.01
	SNP_13730	T/C	3B	70,909,819–70,909,887	0.01
GL	SNP_7441*	G/A	3B	809,510,260–809,510,317	0.2
	SNP_8141	C/T	4A	702,149,172–702,149,240	0.05
GLWR	SNP_5518	G/A	1D	225,564,124–225,564,192	0.12
GFD	SNP_15086	A/G	7B	641,807,115–641,807,183	− 0.22
	SNP_1653	G/T	1A	550,948,719–550,948,771	− 0.24
	SNP_6775	A/T	2B	436,016,073–436,016,141	0.51

E1; environment 1, E2; environment 2, *; Bonferroni multiple correction passed, #; desirable allele, Chr; chromosome and Pos; position

(i) GCSA: E1

(ii) GP: E1

(iii) GWid: E1

(iv) GCSA: E2

(v) GL: E2

(vi) GP: E2

(vii) GFD: E2

(viii) GCSA: B

(ix) GP: B

(x) GL: B

(xi) GLWR: B

(xii) GLWR: B

◄**Fig. 6** Manhattan plots (left side) showing significant marker-trait associations (MTAs) on 21 chromosomes and ump (represents unmapped MTAs) using FarmCPU with corrected *P*-value < 0.05 after FDR (blue line) and Bonferroni (red line) corrections in each case. Q-Q plots (right side) represent distribution of observed and expected *P*-values for the same traits using environment 1 & 2 (E1 and E2) and BLUP values

Expression analysis of CGs

In silico expression analysis suggested that CGs differed in their level of expression in different tissues of the seed (e.g. embryo, endosperm, aleurone, seed coat) and at different stages of development (Fig. 7). It is apparent from the figure that most genes have tissue-specific expression and that many more genes have high expression only in embryo proper and seed coat.

Wheat orthologues of rice genes for grain morphology

Wheat orthologs of known rice genes for grain morphology traits were also identified (Table S5). For this purpose, eight rice genes were selected on the basis of their functions relevant to the traits of our interest and location within the regions of corresponding wheat CGs. Each of

Table 4 A summary of significant MTAs for 2–4 traits that qualify FDR correction and detected in both the environments (E1-2017–18 and E2-2018–19) using mvLMM

Trait combination	SNP	Allele#	Chr	Pos. of SNP tag (bp)	Trait combination	SNP	Allele#	Chr	Pos. of SNP tag (bp)
E1									
GL-GWid-GLWR-GFD	SNP_912*	C/G#	6B	674,558,317–674,558,364	GP-GWid-GLWR-GFD	SNP_8616	C/A	7A	41,428,595–41,428,650
	SNP_9744*	G/C	2B	800,127,644–800,127,712		SNP_9656	T/A	1A	351,340,404–351,340,472
	SNP_17773	G/C	3B	819,747,424–819,747,492	GCSA-GL-GLWR-GFD	SNP_8616	C/A	7A	41,428,595–41,428,650
GCSA-GWID-GLWR-GFD	SNP_8616	A/A	7A	41,428,595–41,428,650	GP-GL-GWLR-GFD	SNP_12761	T/G	6B	20,093,212–20,093,280
E2									
GCSA-GP-GL	SNP_8221	T/C	1D	408,525,725–408,525,793	GCSA-GWid-GLWR-GFD	SNP_912*	C/G	6B	674,558,317–674,558,364
	SNP_3393	A/T	5A	536,221,003–536,221,071		SNP_9744*	G/C	2B	800,127,644–800,127,712
	SNP_16292	C/T	5B	697,526,371–697,526,439		SNP_7751	T/C	2A	51,261,956–51,262,014
GP-GL-GWid	SNP_8221	T/C	1D	408,525,725–408,525,793		SNP_3393	A/T	5A	536,221,003–536,221,071
	SNP_3393	T/A	5A	536,221,003–536,221,071		SNP_11003	C/A	6A	584,705,543–584,705,611
GL-GWid-GLWR	SNP_7751	T/C	2A	51,261,956–51,262,014		SNP_259	G/A	2B	703,904,164–703,904,221
	SNP_11003	A/C	6A	584,705,543–584,705,611		SNP_6171	G/C	7A	681,786,065–681,786,132
	SNP_3968	C/T	6A	583,269,004–583,269,072		SNP_3968	C/T	6A	583,269,004–583,269,072
	SNP_3393	T/A	5A	536,221,003–536,221,071		SNP_3393	A/T	5A	536,221,003–536,221,071
GWid-GLWR-GFD	SNP_3915	C/T	3A	502,497,785–502,497,853	GP-GWid-GLWR-GFD	SNP_3915	T/C	3A	502,497,785–502,497,853
	SNP_373	G/T	1A	573,920,780–573,920,835		SNP_2735	T/C	6A	69,057,764–69,057,802
	SNP_259	A/G	2B	703,904,164–703,904,221		SNP_373	T/G	1A	573,920,780–573,920,835
GCSA-GP-GL-GWid	SNP_3393	T/A	5A	536,221,003–536,221,071		SNP_3915	T/C	3A	502,497,785–502,497,853
	SNP_16292	C/T	5B	697,526,371–697,526,439		SNP_8221	C/T	1D	408,525,725–408,525,793
GP-GL-GWid-GLWR	SNP_7751	C/T	2A	51,261,956–51,262,014	GCSA-GL-GLWR-GFD	SNP_3393	A/T	5A	536,221,003–536,221,071
	SNP_11003	A/C	6A	584,705,543–584,705,611		SNP_373	T/G	1A	573,920,780–573,920,835
	SNP_3968	C/T	6A	583,269,004–583,269,072	GP-GL-GLWR-FFD	SNP_373	T/G	1A	573,920,780–573,920,835
GL-GWid-GLWR-GFD	SNP_373	T/G	1A	573,920,780–573,920,835		SNP_688	A/G	4A	692,758,028–692,758,090
	SNP_3915	T/C	3A	502,497,785–502,497,853		SNP_3915	T/C	3A	502,497,785–502,497,853

*; Bonferroni multiple correction passed, Chr; chromosome, Pos; position and # desirable allele

Table 5 MTAs for six traits identified during the present study that correspond to known QTLs (either overlapping known QTLs or located in the vicinity of known QTLs) in wheat

MTAs				QTL reported earlier	
Trait	SNP (MTA)	Chromosome	Position (Mb)	Reported region (Mb)	References for QTL in earlier study
Associated SNP identified within flanking region of the reported QTL					
GP	SNP_9771	7B	684.27	136.1–740.72	Kumari et al. (2018)
GWid	SNP_5401	2A	79.34	59.55–112.75	Williams et al. (2013)
Associated SNP identified in the vicinity (< 50 Mb) to reported QTL regions					
GWid	SNP_8883	3A	510.68	528.21	Gao et al. (2021)
GLWR	SNP_12288	2B	683.77	718.96	
GFD	SNP_2816	1A	549.31	511.45	Kumari et al. (2018)

Table 6 Summary of most important MTAs for MAS

SNP	Chr	Pos. of SNP tags	Allele#	Description
SNP_3202	3A	593,990,302–593,990,370	A/G#	Detected using CMLM, FarmCPU and SUPER: associated with GL
SNP_9744	2B	800,127,644–800,127,712	G/C	Detected using mvLMM and identified in the both environments: associated with GL-GWid-GLWR-GFD
SNP_912	6B	674,558,317–674,558,364	C/G	Same as above
SNP_7441	3A	731,844,353–731,844,410	A/G	Detected using CMLM: associated with GL
SNP_5401	2A	79,343,336–79,343,404	G/T	Detected using CMLM: associated with Gwid
SNP_4648	7A	94,143,005–94,143,073	G/T	Same as above
SNP_12288	2B	683,768,477–683,768,545	T/C	Detected using CMLM: associated with GLWR
SNP_4563	2B	571,070,836–571,070,884	C/A	Detected using FarmCPU: associated with GCSA
SNP_11360	1A	6,341,227–6,341,295	T/G	Detected using epistasis: associated with GLWR
SNP_11245	7B	739,999,777–739,999,837	C/G	Same as above
SNP_12855	5B	43,824,419–43,824,487	G/A	Same as above
SNP_2977	7B	678,267,823–678,267,891	T/G	Same as above
SNP_5634	2D	82,867,198–82,867,237	G/A	Detected using epistasis: associated with GFD
SNP_1974	5A	480,099,548–480,099,616	T/G	Same as above

Chr; chromosome, Pos; position (base pair) and #; desirable allele

these rice genes is known to play an important role in seed morphology and reproductive organ development. The results of this exercise are summarised in Table 7. Among these eight rice genes, following two genes deserved special attention. (1) The wheat gene *TraesCS7B02G409000* (associated with GLWR of wheat) corresponded to the rice gene *OsGW2*, which is known to play an important role in grain width/weight in rice. This gene is highly expressed in seed development stages and the encoded protein is localized in the nucleus and cytoplasm. (2) The wheat gene *TraesCS2B02G487800* (associated with GLWR) corresponded rice gene *OsVP1*, which is known to play an important role in seed maturation and dormancy.

Discussion

Grain morphology is an important complex quantitative trait, which includes a number of component traits including six traits that were utilized for GWAS in the present study. It is known that a long tiny primitive grain of the progenitors of wheat was transformed during domestication into the present-day much bigger round grain of wheat (Gegas et al. 2010). Among the three progenitors of wheat sub-genomes, the progenitor of sub-genome A, namely *T. urartu*, is known to be the major determinant of grain shape (Feldman et al. 2012). Among other alien species, the sub-genome D progenitor, namely *Ae. tauschii* is also known to affect grain morphology of wheat (Röder et al. 2008; Okamoto et al. 2013; Rasheed et al. 2014).
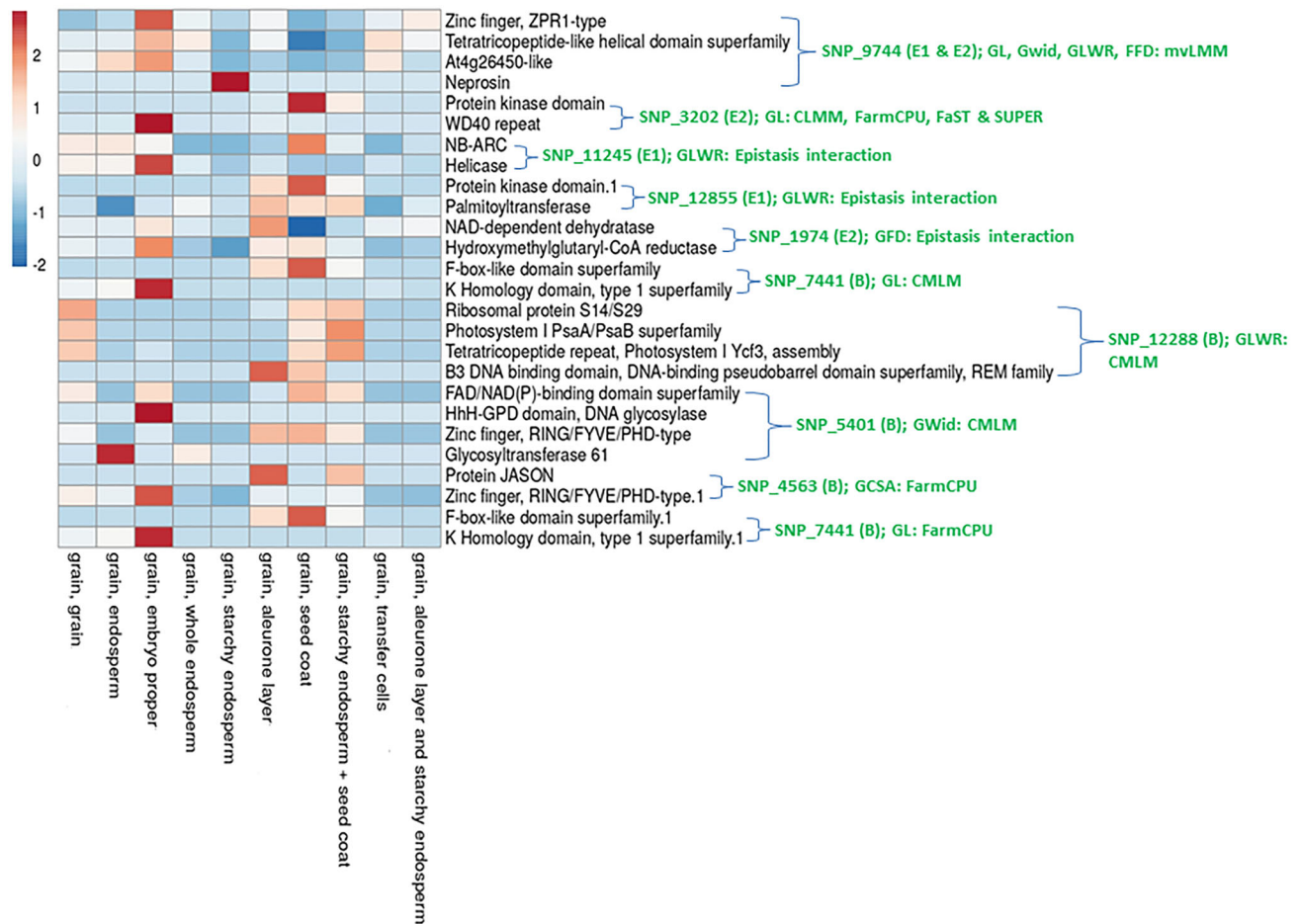
Fig. 7 *In-silico* expression analysis of CGs in different tissues in wheat; columns represent 10 different development stages/tissues of grain and rows represent 26 different proteins encoded by candidate genes (CGs)

The objective of the present study was to investigate the genetic architecture of six grain morphology traits utilizing a set of 225 wheat genotypes comprising an association panel. The genotypes in the above panel were also a part of the association panels used in three of our earlier studies (Kumar et al. 2018; Gahlaut et al. 2019, 2021). Several studies involving interval mapping or association mapping have been conducted earlier to study the genetic architecture of the grain weight in wheat (Ain et al. 2015; Sun et al. 2017; Kumar et al. 2018; Singh et al. 2021). Some of these studies also involved grain morphology traits, including the following traits used in the present study: GCSA, GWid, GLWR, GP, GL and GFD (Ramya et al. 2010; Gegas et al. 2010; Prashant et al. 2012; Williams et al. 2013; Okamoto et al. 2013; Williams and Sorrells 2014; Tyagi et al. 2014; Rasheed et al. 2014; Zhang et al. 2015; Wu et al. 2016; Arora et al. 2017; Yan et al. 2017; Kumari et al. 2018; Yoshioka et al. 2019; Alemu et al. 2020; Gao et al. 2021; Schierenbeck et al. 2021). The present study is yet another study, which added a number of novel MTAs to the ever-growing list of markers associated with grain morphology

traits. In the present study, three sets of data (E1, E2 and BLUP) were utilized. Majority of MTAs were unique in these three sets with only 4 MTAs that were common in two of the three sets. This suggested that majority of MTAs were environment specific and that there were strong QT L x Environment interactions, making it difficult to use the same MTAs for MAS under different environment. This is in agreement with the results of some earlier studies (Gahlaut et al. 2020; Alemu et al., 2021; Thudi et al. 2021).

The four models and another model for epistatic inter-actions were used in the present study. Software for these models (except mvLMM) are available in the package GAPIT version 2.0, made available in 2016 (Tang et al. 2016), although now GAPIT version 3 is also available (Wang and Zhang 2021); GAPIT version 1.0 was developed earlier (Lipka et al. 2012). Each of these four models have their own merits over initial MLM proposed by Ed Buckler's group at Cornell University (Yu et al. 2006). For instance, CMLM overcomes the computational problem associated with large datasets by decreasing the effective sample size by clustering individuals into groups.

Table 7 List of selected homology-based mining of candidate genes

Methods	Trait (Env.); Associated SNP: Chr	Wheat candidate genes	Rice gene id	Rice gene	Function
CLMM, FarmCPU, SUPER	GL (E2); SNP_3202: 3A	*TraesCS3A02G344400*	*Os01t0822200-01*	*OsRLCK47*	Panicle and seed development
Epistasis interaction	GLWR (E1); SNP_11245: 7B & GLWR (E1); SNP_2977: 7B	*TraesCS7B02G482200* *TraesCS7B02G482400* *TraesCS7B02G409000* *TraesCS7B02G409200*	*Os04t0309600-01* *Os06t0102800-01* & *Os06t0652550-00 Os03t0325900-00*	*OsSUS5 Osenl1* & *OsGW2 OsEnS-8*	Seed—Morphological traits character as QTL—Yield and productivity Seed—Morphological traits – Endosperm Seed—Morphological traits—Grain shape Seed—Morphological traits—Grain shape
BLUP	GLWR; SNP_12288: 2B	*TraesCS2B02G487800*	*Os03t0212300-01*	*OsVp1*	Seed—Physiological traits—Dormancy
	GWid; SNP_5401: 2A	*TraesCS2A02G132300*	*Os06t0552900-00*	*OsFTL12*	Reproductive organ—Heading date
	GWid; SNP_4648: 7A	*TraesCS7A02G142600*	*Os01t0213300-00*	*OsMADS91*	Reproductive organ—Spikelet, flower, glume, awn, panicle

Env; environment, Chr; chromosome, E1; environment 1 and E2; environment 2

FarmCPU eliminates confounding problems arising due to kinship, population structure, multiple testing, etc. through the use of Fixed Effect Model (FEM) and a Random Effect Model (REM) iteratively. SUPER solves the computational burden problem, and overcomes the limitation, where number of SNPs should be less than the number of individuals through extracting a small subset of SNPs, thus increasing the statistical power without reducing the number of SNPs. The mvLMM allows use of two or more correlated traits for multi-trait analysis.

In the present study, different models gave different results suggesting that different models differ in their superiority over TASSEL. Only a solitary MTA (SNP_3202) was identified using three different models (Table 2). Similar results have been reported in several recent GWAS studies involving multiple models (Alemu et al. 2021; Thudi et al. 2021). This feature alone shows the utility of multiple models for GWAS. The results of the present study also support the general conclusion that till recently, FarmCPU has been the best model for GWAS. However, FarmCPU is a bin method under the unrealistic assumption that quantitative trait nucleotides (QTNs) are evenly distributed throughout the genome. In FarmCPU we use REM which is computationally expensive. In the recently developed tool named BLINK, REM is replaced with FEM in order to eliminate the requirement that QTNs are evenly distributed throughout the genome, which further improved the statistical power of FarmCPU, in addition to reduced computing time, so that a large dataset with one million individuals and one-half million markers can

be analyzed within three hours, instead of one week using FarmCPU (Huang et al. 2018). Therefore, we anticipate that BLINK will be increasingly used in future replacing FarmCPU, particularly, when large association panels with millions of markers are used for GWAS.

Correlations between different traits including SA, GP, GL and GWid also suggested that there may be QTLs which contribute each to more than one trait involved in grain size (Fig. 4, Figs. S2 & S3; Okamoto et al. 2013; Kumari et al. 2018; Yoshioka et al. 2019; Alemu et al. 2020; Mérida-García et al. 2020; Gao et al. 2021). Significant negative correlation of GFD with all the traits except GLWR is also in agreement with earlier reports (Dholakia et al. 2003; Kumari et al. 2018). A significant negative correlation between GLWR and GWid observed in the present study in both the environments (Figs. S2 & S3) may be the result of positive selection for round grains (Okamoto et al. 2013; Kumari et al. 2018; Gao et al. 2021).

The association panel used during the present study exhibited a low level of population structure containing only three sub-populations, as was also reported in two of our earlier GWA studies, where association panels included the genotypes used during the present study (Kumar et al. 2018; Gahlaut et al. 2019, 2021). However, the number of sub-populations in some other previous GWA studies ranged from three (Wang et al. 2017; Rahimi et al. 2019) to five (Qaseem et al. 2018; Jamil et al. 2019). Low level of population structure, as observed in the present study and in our earlier GWA studies is a desirable feature for conducting GWAS.

The availability of a large number of MTAs identified during the present study suggested that multiple loci are involved in controlling each of the six traits (Fig. 3 and Table 1). A comparison of the results of the present study with those reported earlier suggested that 14 of the novel 112 MTAs identified in the present study (Tables 5, 6) may be useful for MAS. The multi-locus analysis (allowed by FarmCPU) and multi-trait analysis also allowed identification of MTAs, each controlling more than one trait. This feature of mvLMM provides an additional advantage over multi-trait mixed model (MTMM) initially developed by Korte et al. (2012), which permits detection of MTAs for not more than two traits at a time (Jaiswal et al. 2016; Kumar et al. 2018). Although multi-trait analysis using MTMM and involving only two traits has been used in earlier GWA studies, more recently mvLMM permitting identification of MTAs involving more than two traits has been used in several studies (Furlotte and Eskin 2015; Kumar et al. 2018; Deng et al. 2018; Gao et al. 2021; Chen et al. 2021; Malik et al. 2021). Availability of 46 MTAs, each involving more than one trait (Table 4), suggest a significant role of pleiotropy or close linkage in controlling the grain morphology traits in wheat. These MTAs included four significant MTAs for GL-GWid-GLWR-GFD that qualified Bonferroni correction (Table 4). These four MTAs were common among both the environments and should prove useful for simultaneous improvement of more than one trait using MAS. Earlier reports on multi-trait association analysis reported 6 to 38 MTAs involving up to five traits for a variety of grain quality and yield traits in wheat (Kumar et al. 2018; Gao et al. 2021; Malik et al. 2021). However, specifically for grain morphology traits, only six MTAs involving two to four traits were reported by Gao et al. (2021), which agrees with the results of the present investigation.

The availability of 29 first order epistatic interactions, which qualified FDR correction (Table S3) also suggested that epistatic interactions were not uncommon. However, epistatic MTAs have been sparingly used in MAS for crop improvement (Reif et al. 2011; Kao et al. 1999; Langer et al. 2014; Jaiswal et al. 2016; Sehgal et al. 2017; Kumar et al. 2018). Therefore, on the basis of the present study, it is recommended that MTAs involved in epistatic interactions, which significantly contribute to the available genetic variation for the traits under study, should also be used for MAS. Epistatic interactions in wheat using GWAS have been reported in several earlier studies involving a number of traits, which include the following: (1) flowering time (Reif et al. 2011; Langer et al. 2014), (2) stem rust resistance (Yu et al. 2011), and (3) agronomic traits (Sehgal et al. 2017). Epistatic interactions were also earlier identified in several studies from our own laboratory, where 63 epistatic interactions for 13 different yield traits were identified in one study and 73 epistatic interactions

for three micronutrients were identified in the other study (Jaiswal et al. 2016); Kumar et al. 2018). Similarly, epistatic interactions were also detected using bi-parental populations through interval mapping (Li et al. 2011; Xu et al. 2012; Rouse et al. 2014; Boeven et al. 2020). Therefore, the results of epistatic interactions from earlier studies and the present study may be useful for MAS to supplement conventional wheat breeding.

MTAs were also used for identification of CGs for each individual trait; this gave 55 CGs (Table S4). The in silico expression analysis for 26 of the above CGs showed variable expression in grain and related tissues. The CGs which showed very high expression in one or more of the different tissues of grain encode the following different proteins: (i) Zinc finger, ZPR1-type, (ii) Neprosin, (iii) Protein kinase domain, (iv) WD40 repeat, (v) Helicase, (vi) Protein kinase domain.1, (vii) F-box like domain superfamily, (viii) K homology domain, type,1 superfamily, (ix) B3DNA binding domain, (x) HhH GPD domain, (xi) DNA glycolase, (xii) Glycosyltransferase 61, (xiii) Protein JASON, (xiv) Zinc finger, RINGFYV/PHD-type1, and (xv) F-box like domain superfamily.1 (Fig. 7). The CGs encoding these proteins are the targets of future functional genomics research aimed at determining the role of these CGs in controlling the grain morphology traits examined during the present study.

Some other important CGs includes the following; (1) The gene *TaFBA1* encoding F-box protein and its function was confirmed through overexpression in tobacco, suggesting its involvement in plant growth and development including seed germination (An et al. 2019). (2) A sucrose synthase gene, which is a key enzyme of starch biosynthesis, thus affecting contents of starch affecting grain morphology and weight (Dai et al. 2009). (3) Glycoside hydrolase (a superfamily protein) is involved in different processes such as seed development and endosperm cell wall degradation during germination (Dong et al. 2019).

Another exercise in the present study involved identification of eight important CG-based wheat genomic regions carrying orthologues of rice genes for grain morphology. These CGs, were associated with reproduction and regulation of seed morphology and should be the target for future studies in wheat (Table 7). For example, the gene *TraesCS3A02G344400*, which is orthologous to the rice gene *OsRLCK47*, is associated with GL of wheat. Shubha et al. (2008), functionally characterized *RLCKs* from several crops including rice, where it was shown to play roles in development (panicle and seed development) and stress responses. Similarly, *TraesCS7B02G482200* (associated with GLWR of wheat) is orthologue of the gene *OsSUS5*, which plays an important role in seed development and was shown to express in sink tissues like root, flower and immature seed (Cho et al. 2011).

Conclusions

The MTAs for six grain morphology traits, identified in present study, may be useful for MAS for development of wheat cultivars with high grain yield associated with high quality. These MTAs may be validated and may also be used for post-GWAS or joint linkage and association mapping (JLAM; Gupta et al. 2019; Gahlaut et al. 2019). The information of the CGs may also be useful for the development of CG-based functional markers. These markers may also be useful for MAS to facilitate breeding for improvement of grain morphology traits. The CGs identified in the present study may also be used for CG-based association mapping and functional genomics in future research.

Author contributions Shailendra Sharma conceived and conducted or directed the experiments. Parveen Malik, Jitendra Kumar and Shiveta Sharma performed the experiments. Prabina Kumar Meher, Jitendra Kumar, Parveen Malik analysed the data. Harindra Singh Balyan, Pushpendra Kumar Gupta and Shailendra Sharma participated in manuscript writing. Shailendra Sharma finalized the manuscript.

Declarations

Conflict of interest There is no conflict of interest.

References

Ain QU, Rasheed A, Anwar A et al (2015) Genome-wide association for grain yield under rainfed conditions in historical wheat cultivars from Pakistan. Front Plant Sci 6:1–15

Alemu A, Feyissa T, Tuberosa R et al (2020) Genome-wide association mapping for grain shape and color traits in Ethiopian durum wheat (Triticum turgidum ssp durum). Crop J 8:757

Alemu A, Suliman S, Hagras A et al (2021) Multi-model genome-wide association and genomic prediction analysis of 16 agronomic, physiological and quality related traits in ICARDA spring wheat. Euphytica 217:1–22

Allard RW (1999) Principles of plant breeding, 2nd edn. John Wiley & Sons, Hoboken

An J, Li Q, Yang J et al (2019) Wheat F-box protein TaFBA1 positively regulates plant drought tolerance but negatively regulates stomatal closure. Front Plant Sci 10:1–20

Arora S, Singh N, Kaur S et al (2017) Genome-wide association study of grain architecture in wild wheat aegilops tauschii. Front Plant Sci 8:1–13

Bates D, Machler M, Bolker B, Walker S (2015) Fitting linear mixed-effects models using lme4. J Stat Softw 67:1–48

Bateson W (1909) Mendel's Principles of Heredity: Cambridge University Press. März 1909; 2nd Impr, 3:1913.

Boeven PHG, Zhao Y, Thorwarth P et al (2020) Negative dominance and dominance-by-dominance epistatic effects reduce grain-yield heterosis in wide crosses in wheat. Sci Adv 6:1–12

Bradbury PJ, Zhang Z, Kroon DE, et al (2008) Genetics and population analysis TASSEL: software for association mapping of complex traits in diverse samples

Breseghello F, Sorrells ME (2007) QTL analysis of kernel size and shape in two hexaploid wheat mapping populations. Field Crops Res 101:172–179

Cabral AL, Jordan MC, Larson G et al (2018) Relationship between QTL for grain shape, grain weight, test weight, milling yield, and plant height in the spring wheat cross RL4452/ 'AC Domain.' PLoS ONE 13:1–32

Campbell BT, Baenziger PS, Gill KS et al (2003) Identification of QTLs and environmental interactions associated with agronomic traits on chromosome 3A of wheat. Crop Sci 43:1493–1505

Chastain TG, Ward KJ, Wysocki DJ (1995) Stand establishment responses of soft white winter wheat to seedbed residue and seed size. Crop Sci 35:213–218

Chen Y, Wu H, Yang W, Zhao W, Tong C (2021) Multivariate linear mixed model enhanced the power of identifying genome-wide association to poplar tree heights in a randomized complete block design. G3 11:1–32

Cho J II, Kim HB, Kim CY, Hahn TR, Jeon JS (2011) Identification and characterization of the duplicate rice sucrose synthase genes OsSUS5 and OsSUS7 which are associated with the plasma membrane. Mol Cells 31:553–561

Cortes LT, Zhang Z, Yu J (2021) Status and prospects of genome-wide association studies in plants. Plant Genome 14:1–38

Dai Z, Yin Y, Wang Z (2009) Comparison of starch accumulation and enzyme activity in grains of wheat cultivars differing in kernel type. Plant Growth Regul 57:153–162

Deng X, Wang B, Fisher V, Peloso G, Cupples A, Liu CT (2018) Genome-wide association study for multiple phenotype analysis. BMC Proc 12:139–144

Dholakia BB, Ammiraju JSS, Singh H et al (2003) Molecular marker analysis of kernel size and shape in bread wheat. Plant Breed 122:392–395

Dong X, Jiang Y, Hur Y (2019) Genome-wide analysis of glycoside hydrolase family 1 β-glucosidase genes in Brassica rapa and their potential role in Pollen development. Int J Mol Sci 20:1–17

Evers AD (2000) Grain size and morphology: implications for quality. Spec Publ-Royal Soc Chem 212:19–24

Feldman M, Levy AA, Fahima T, Korol A (2012) Genomic asymmetry in allopolyploid plants: wheat as a model. J Exp Bot 63:5045–5059

Furlotte NA, Eskin E (2015) Efficient multiple-trait association and estimation of genetic correlation using the matrix-variate linear mixed model. Genetics 200:59–68

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 9:1–15

Gahlaut V, Kumari P, Jaiswal V, Kumar S (2021) Genetics, genomics and breeding in Rosa species. J Hortic Sci Biotechnol 96:1–16

Gahlaut V, Zinta G, Jaiswal V, Kumar S (2020) Quantitative epigenetics: a new avenue for crop improvement. Epigenomes 4:1–17

Gan Y, Stobbe EH (1996) Seedling vigor and grain yield of "Roblin" wheat affected by seed size. Agron J 88:456–460

Gao Y, Xu X, Jin J et al (2021) Dissecting the genetic basis of grain morphology traits in Chinese wheat by genome wide association study. Euphytica 217:1–12

Gegas VC, Nazari A, Griffiths S et al (2010) A genetic framework for grain size and shape variation in wheat. Plant Cell 22:1046–1056

Giura A, Saulescu NN (1996) Chromosomal location of genes controlling grain size in a large grained selection of wheat (Triticum aestivum L.). Euphytica 89:77–80

Gonzalez JR, Armengol L, Solé X et al (2007) SNPassoc: an R package to perform whole genome association studies. Bioinformatics 23:644–645

Gupta PK, Balyan HS, Kulwal PL et al (2007) QTL analysis for some quantitative traits in bread wheat. J Zhejiang Univ Sci B 8:807–814

Gupta PK, Kulwal PL, Jaiswal V (2014) Association mapping in crop plants: opportunities and challenges. Adv Genet 85:109–147

Gupta PK, Kulwal PL, Jaiswal V (2019) Association mapping in plants in the post-GWAS genomics era. Adv Genet 104:75–154

Holland JB (2001) Epistasis and plant breeding. Plant Breed Rev 21:27–92

Hu J, Wang Y, Fang Y et al (2015) A rare allele of GS2 enhances grain size and grain yield in rice. Mol Plant 8:1455–1465

Huang R, Jiang L, Zheng J et al (2013) Genetic bases of rice grain shape: So many genes, so little known. Trends Plant Sci 18:218–226

Huang M, Liu X, Zhou Y et al (2018) BLINK: a package for the next level of genome-wide association studies with both individuals and markers in the millions. GigaScience 8:1–12

IWGSC (2018) Shifting the limits in wheat research and breeding using a fully annotated reference genome. Science 345:1251788

Jaiswal V, Gahlaut V, Meher PK et al (2016) Genome wide single locus single trait, multi-locus and multi-trait association mapping for some important agronomic traits in common wheat (T. aestivum L.). PLoS ONE 11:1–25

Jamil M, Ali A, Gul A et al (2019) Genome-wide association studies of seven agronomic traits under two sowing conditions in bread wheat. BMC Plant Biol 19:1–18

Jiang C, Xiao S, Li D et al (2019) Identification and expression pattern analysis of bacterial blight resistance genes in Oryza officinalis wall ex watt under xanthomonas oryzae Pv. oryzae Stress. Plant Mol Biol Report 37:436–449

Jing HC, Kornyukhin D, Kanyuka K et al (2007) Identification of variation in adaptively important traits and genome-wide analysis of trait-marker associations in Triticum monococcum. J Exp Bot 58:3749–3764

Kao CH, Zeng ZB, Teasdale RD (1999) Multiple interval mapping for quantitative trait loci. Genetics 152:1203–1216

Korte A, Vilhjálmsson BJ, Segura V, Platt A, Long Q, Nordborg M (2012) A mixed-model approach for genome-wide association studies of correlated traits in structured populations. Nat Genet 44:1066–1071

Kovach MJ, Sweeney MT, McCouch SR (2007) New insights into the history of rice domestication. Trends Genet 23:578–587

Kumar J, Saripalli G, Gahlaut V et al (2018) Genetics of Fe, Zn, β-carotene, GPC and yield traits in bread wheat (Triticum aestivum L.) using multi-locus and multi-traits GWAS. Euphytica 214:1–17

Kumari S, Jaiswal V, Mishra VK, Paliwal R, Balyan HS, Gupta PK (2018) QTL mapping for some grain traits in bread wheat (Triticum aestivum L.). Physiol Mol Biol Plants 24:909–920

Langer SM, Longin CFH, Würschum T (2014) Flowering time control in European winter wheat. Front Plant Sci 5:1–12

Li J, Horstman B, Chen Y (2011) Detecting epistatic effects in association studies at a genomic level based on an ensemble approach. Bioinformatics 27:222–229

Li Q, Zhang Y, Liu T et al (2015) Genetic analysis of kernel weight and kernel size in wheat (Triticum aestivum L.) using unconditional and conditional QTL mapping. Mol Breed 35:1–15

Lipka AE, Tian F, Wang Q et al (2012) GAPIT: Genome association and prediction integrated tool. Bioinformatics 28:2397–2399

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 12:1–24

Lu HY, Liu XF, Wei SP, Zhang YM (2011) Epistatic association mapping in homozygous crop cultivars. PLoS One 6:e17773

Ma J, Zhang H, Li S et al (2019) Identification of quantitative trait loci for kernel traits in a wheat cultivar Chuannong16. BMC Genet 20:1–12

Mackay TF (2014) Epistasis and quantitative traits: using model organisms to study gene-gene interactions. Nat Rev Genet 15:22–33

Malik P, Kumar J, Singh S et al (2021) Single-trait, multi-locus and multi-trait GWAS using four different models for yield traits in bread wheat. Mol Breed 41:1–21

Maulana F, Ayalew H, Anderson JD, Kumssa TT, Huang W, Ma XF (2018) Genome-wide association mapping of seedling heat tolerance in winter wheat. Front Plant Sci 9:1–16

Mendiburu D, Yaseen M (2020) Agricolae: s tatistical procedures for agricultural research, R package version 1.4.0.

Merida-García R, Bentley AR, Gálvez S et al (2020) Mapping agronomic and quality traits in elite durum wheat lines under differing water regimes. Agronomy 10:1–23

Moellers TC, Singh A, Zhang J et al (2017) Main and epistatic loci studies in soybean for Sclerotinia sclerotiorum resistance reveal multiple modes of resistance in multi-environments. Sci Rep 7:1–13

Niel C, Sinoquet C, Dina C, Rocheleau G (2015) A survey about methods dedicated to epistasis detection. Front Genet 6:285

Okamoto Y, Nguyen AT, Yoshioka M, Iehisa JCM, Takumi S (2013) Identification of quantitative trait loci controlling grain size and shape in the D genome of synthetic hexaploid wheat lines. Breed Sci 63:423–429

Patil RM, Tamhankar SA, Oak MD et al (2013) Mapping of QTL for agronomic traits and kernel characters in durum wheat (Triticum durum Desf.). Euphytica 190:117–129

Peterson BG, Carl P, Boudt K et al (2018) Package 'PerformanceAnalytics.' R Team Cooperation 3:13–14

Prashant R, Kadoo N, Desale C et al (2012) Kernel morphometric traits in hexaploid wheat (Triticum aestivum L.) are modulated by intricate QTL × QTL and genotype × environment interactions. J Cereal Sci 56:432–439

Qaseem MF, Qureshi R, Muqaddasi QH, Shaheen H, Kousar R, Roder MS (2018) Genome-wide association mapping in bread wheat subjected to independent and combined high temperature and drought stress. PLoS ONE 13:1–22

Rahimi Y, Bihamta MR, Taleei A, Alipour H, Ingvarsson PK (2019) Genome-wide association study of agronomic traits in bread wheat reveals novel putative alleles for future breeding programs. BMC Plant Biol 19:1–19

Ramya P, Chaubal A, Kulkarni K et al (2010) QTL mapping of 1000-kernel weight, kernel length, and kernel width in bread wheat (Triticum aestivum L.). J Appl Genet 51:421–429

Rasheed A, Xia X, Ogbonnaya F et al (2014) Genome-wide association for grain morphology in synthetic hexaploid wheats using digital imaging analysis. BMC Plant Biol 14:1–21

Reif JC, Maurer HP, Korzun V, Ebmeyer E, Miedaner T, Wurschum T (2011) Mapping QTLs with main and epistatic effects

underlying grain yield and heading time in soft winter wheat. Theor Appl Genet 123:283–292

Ritchie MD, Van Steen K (2018) The search for gene-gene interactions in genome-wide association studies: challenges in abundance of methods, practical considerations, and biological interpretation. Ann Transl Med 6:1–14

Rouse MN, Talbert LE, Singh D, Sherman JD (2014) Complementary epistasis involving Sr12 explains adult plant resistance to stem rust in Thatcher wheat (*Triticum aestivum* L.). Theor Appl Genet 127:1549–1559

Röder MS, Huang XQ, Börner A (2008) Fine mapping of the region on wheat chromosome 7D controlling grain weight. Funct Integr Genom 8:79–86

Schierenbeck M, Alqudah AM, Lohwasser U, Tarawneh RA, Simon MR (2021) Genetic dissection of grain architecture-related traits in a winter wheat population. BMC Plant Biol 21:1–14

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 7:1–12

Sehgal D, Rosyara U, Mondal S et al (2020) Incorporating genome-wide association mapping results into genomic prediction models for grain yield and yield stability in CIMMYT spring bread wheat. Front Plant Sci 11:197

Shubha V, Giri J, Dansana PK, Kapoor S, Tyagi AK (2008) The receptor-like cytoplasmic kinase (OsRLCK) gene family in rice: Organization, phylogenetic relationship, and expression during development and stress. Mol Plant 1:732–750

Singh K, Batra R, Sharma S, et al (2021) WheatQTLdb: a QTL database for wheat. Mol Genet Genomics 1–7.

Slim L, Chatelain C, Azencott CA, Vert JP (2020) Novel methods for epistasis detection in genome-wide association studies. PLoS One 15:e0242927

Sun C, Dong Z, Zhao L, Ren Y, Zhang N, Chen F (2020) The Wheat 660K SNP array demonstrates great potential for marker-assisted selection in polyploid wheat. Plant Biotechnol J 18:1354–1360

Sun XY, Wu K, Zhao Y et al (2009) QTL analysis of kernel shape and weight using recombinant inbred lines in wheat. Euphytica 165:615–624

Sun C, Zhang F, Yan X et al (2017) Genome-wide association study for 13 agronomic traits reveals distribution of superior alleles in bread wheat from the Yellow and Huai Valley of China. Plant Biotechnol J 15:953–969

Tanabata T, Shibaya T, Hori K, Ebana K, Yano M (2012) SmartGrain: high-throughput phenotyping software for measuring seed shape through image analysis. Plant Physiol 160:1871–1880

Tang Y, Liu X, Wang J et al (2016) GAPIT version 2: an enhanced integrated tool for genomic association and prediction. Plant Genome 9:1–9

Thudi M, Chen Y, Pang J et al (2021) Novel genes and genetic loci associated with root morphological traits, phosphorus-acquisition efficiency and phosphorus-use efficiency in chickpea. Front Plant Sci. https://doi.org/10.3389/fpls.2021.636973

Tyagi S, Mir RR, Kaur H et al (2014) Marker-assisted pyramiding of eight QTLs/genes for seven different traits in common wheat (*Triticum aestivum* L.). Mol Breed 34:167–175

VanRaden PM (2008) Efficient methods to compute genomic predictions. J Dairy Sci 91:4414–4423

Wang X, Dong L, Hu J et al (2019) Dissecting genetic loci affecting grain morphological traits to improve grain weight via nested association mapping. Theor Appl Genet 132:3115–3128

Wang Q, Tian F, Pan Y, Buckler ES, Zhang Z (2014) A SUPER powerful method for genome wide association study. PLoS ONE 9:1–9

Wang SX, Zhu YL, Zhang DX et al (2017) Genome-wide association study for grain yield and related traits in elite wheat varieties and advanced lines using SNP markers. PLoS ONE 12:1–14

Wang J, Zhang Z (2021) GAPIT Version 3: boosting power and accuracy for genomic association and prediction. Genomics Proteomics Bioinformatics (Published online Sep 4, 2021).

Williams K, Munkvold J, Sorrells M (2013) Comparison of digital image analysis using elliptic Fourier descriptors and major dimensions to phenotype seed shape in hexaploid wheat (*Triticum aestivum* L.). Euphytica 190:99–116

Williams K, Sorrells ME (2014) Three-dimensional seed size and shape QTL in hexaploid wheat (*Triticum aestivum* L.) populations. Crop Sci 54:98–110

Wu J, Zhang J, Wang S, Kong F (2016) Assessment of food security in China: a new perspective based on production-consumption coordination. Sustain 8:1–14

Xin F, Zhu T, Wei S et al (2020) OPEN QTL mapping of kernel traits and validation of a major QTL for kernel length-width ratio using SNP and bulked segregant analysis in wheat. Sci Rep 10:1–12

Xu Y, An D, Liu D, Zhang A, Xu H, Li B (2012) Molecular mapping of QTLs for grain zinc, iron and protein concentration of wheat across two environments. F Crop Res 138:57–62

Yan L, Liang F, Xu H et al (2017) Identification of QTL for grain size and shape on the D genome of natural and synthetic allohexaploid wheats with near-identical AABB genomes. Front Plant Sci 8:1–14

Yang Y, Guo M, Sun S et al (2019) Natural variation of OsGluA2 is involved in grain protein content regulation in rice. Nat Commun 10:1–12

Yoshioka M, Takenaka S, Nitta M, Li J, Mizuno N, Nasuda S (2019) Genetic dissection of grain morphology in hexaploid wheat by analysis of the NBRP-Wheat core collection. Genes Genet Syst 94:35

Yu K, Liu D, Chen Y et al (2019) Unraveling the genetic architecture of grain size in einkorn wheat through linkage and homology mapping and transcriptomic profiling. J Exp Bot 70:4671–4687

Yu K, Liu D, Wu W et al (2017) Development of an integrated linkage map of einkorn wheat and its application for QTL mapping and genome sequence anchoring. Theor Appl Genet 130:53–70

Yu LX, Lorenz A, Rutkoski J et al (2011) Association mapping and gene-gene interaction for stem rust resistance in CIMMYT spring wheat germplasm. Theor Appl Genet 123:1257–1268

Yu J, Pressoir G, Briggs WH et al (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet 38:203–208

Zhang Z, Ersoz E, Lai CQ et al (2010) Mixed linear model approach adapted for genome-wide association studies. Nat Genet 42:355–360

Zhang J, Zhang J, Liu W et al (2015) Introgression of *Agropyron cristatum* 6P chromosome segment into common wheat for enhanced thousand-grain weight and spike length. Theor Appl Genet 128:1827–1837

Zheng J, Zhang Y, Wang C (2015) Molecular functions of genes related to grain shape in rice. Breed Sci 65:120–126

Zhou X, Stephens M (2012) Genome-wide efficient mixed-model analysis for association studies. Nat Genet 44:821–824