BMC Plant Biology

RESEARCH

Open Access

Genome-wide association study for agronomic and yield-related traits in spring wheat (*Triticum aestivum* L.) germplasm

Ankita Thakur[1], Raman Dhariwal[2], Arun Kumar Joshi[3,4], Vinod Kumar Mishra[5], Sandeep Sharma[5], Mukesh Kumar Singh[6], Sundeep Kumar[7*] and Neeraj Kumar Vasistha[1,8*]

Abstract

Background Common wheat (*Triticum aestivum* L.) is one of the most widely grown and consumed cereal crops, but its complicated genome makes it difficult to study how genes affect important agronomic and yield-related traits. Genome-wide association study (GWAS) is a useful method for finding specific loci that control complex agronomic and yield-related traits.

Results The present investigation revealed significant phenotypic variability across the genotypes examined for all traits. The broad sense heritability (H^2) for all traits ranged from 0.50 to 0.71 (Env1; 2021–2022) and 0.53 to 0.81 (Env2; 2022–2023). Using two environments' phenotypic data, and high-throughput single-nucleotide polymorphisms (SNPs) genotypic data of 20,996 markers, we discovered 114 grain-yield-related quantitative trait loci (QTLs) and 300 associated SNP markers. Eighty-five of the identified markers were stable, consistently detected across environments (Env1 and Env2) and combined environment (CE) data, and showed a significant association with 32 different QTLs. The trait with the most associated QTLs (28) was the number of fertile tillers (NFT), with 70 markers. This was followed by 20 QTLs for each, spike length (SL) and spikelet number per spike (SPS), with 69 and 82 SNPs, respectively. Conversely, six SNPs that exhibited association with multiple traits were also identified. Twenty-nine of the total 114 identified QTLs were located in chromosomal positions where at least one marker-trait association had been previously identified.

Conclusion This study has found new SNPs, and useful QTLs that may help us to understand the biological processes behind each studied trait. Further validation in various genetic backgrounds and environments is necessary to confirm the potential utility of the significant alleles found in this study for breeding wheat varieties with improved agronomic and yield-related traits.

Keywords Agronomic traits, FarmCPU, GWAS, WAMI panel, Wheat

*Correspondence:
Sundeep Kumar
Sundeep.Kumar@icar.gov.in
Neeraj Kumar Vasistha
neeraj.vasistha@rgu.ac.in

Full list of author information is available at the end of the article

Background

Common wheat is an important source of calories, and nutrients; it is the principal staple food crop for around 35% of the world's population [1]. In the year 2023/2024, the estimated global production of wheat was 788.95 million metric tonnes (https://fas.usda.gov/data/production/commodity/0410000). This showed a −0.01% decrease when compared to the previous year's production. Conversely, wheat yield must be raised by about 60% by the year 2050 to feed the world's expanding population [2], which emphasises continuous efforts to breed wheat with a higher genetic gain for grain yield. Consequently, one of the primary objectives of wheat breeding is the development of cultivars having higher yield potential [3]. Identifying QTLs related to agronomic traits, and characterising their genetic makeup is important for marker-assisted selection (MAS). GWAS, and biparental QTL mapping are methods used to find the specific locations in the genome that control important traits in crops. In which GWAS is more advantageous in some aspects, such as no need to make a biparental mapping population for QTL mapping, and being less time-consuming [4, 5].

For GWAS, a collection of diverse genotypes is required to initiate the program by identifying QTLs/marker-trait associations (MTAs). The new alleles that come from different genetic backgrounds may increase the grain yield significantly in old or newly developed genotypes of wheat to balance supply, and demand globally. Therefore, GWAS presents a viable alternative to biparental QTL mapping [3] for precisely dissecting grain yield-related traits. GWAS is purely based on the correlation between the observable traits, and genotype of the plants. The regions that have QTLs can be identified with the help of the linkage disequilibrium (LD) that formed in a population across generations so that regions, even if the causal mutations are not necessarily involved among the set of available molecular markers [6]. The use of high-density genotyping that covers the whole genome makes GWAS an invaluable strategy for identifying loci linked to desirable traits [6].

Different type of models have been used in GWAS [7, 8]. These model broadly divided into two group's i.e. single-locus model (SLM) models, and multilocus (ML) models. SL methods like: (a) mixed linear model (MLM) [9], (b) general linear model (GLM) [10], (c) compressed mixed linear model (CMLM) [11], (d) enriched compressed mixed linear model (ECMLM) [12], and (e) settlement of MLM under progressively exclusive relationship (SUPER) [13], only utilised ordinary mixed models for genome scan, and these models do not adequately account for large effect loci, and these problem may be mitigated by ML models [14, 15] such as (f) Bayesian information, and linkage-disequilibrium iteratively nested keyway (BLINK) [16], (g) fixed and random model circulating probability unification (FarmCPU) [17] and (h) multilocus mixed model (MLMM) [15]. So that ML methods are considered more effective, and reliable than single locus SL models for the scanning of genomic regions because all marker effects are simultaneously estimated by the ML methods [7, 15, 18–22]. Furthermore, ML methods do not necessitate testing of detected correlations using strict multiple testing corrections, which typically lead to the rejection of significant associations, in contrast to SLM [21].

Wheat breeders emphasise numerous yield-related traits to maximise yields per unit area [23]. However, recent studies show that the key traits for increasing yield are kernels per spike (KPS), and thousand-kernel weight (TKW), along with harvest index (HI), total spikelets per spike (SPS), spike length (SL), and number of fertile tillers (NFT) [24–26]. Plant height (PH) is also an important trait influencing kernel weight [27]. To maximise wheat production, optimising several agronomic traits is essential [28]. Significant efforts have been previously made to identify QTLs influencing these traits by utilising linkage mapping in biparental populations for their potential use in marker-assisted selection [29, 30]. However, these studies usually have reported a few QTLs per population [31].

A large number of QTLs/MTAs linked with the agronomic traits have been identified using GWAS under the normal irrigated field conditions in the different parts of the world. More than 30 studies in wheat related to the GWAS for agronomic traits have been conducted all over the world under different agroclimatic conditions [3, 6, 32–52]. It is also well known that the yield-related agronomic traits are highly affected by the specific environmental conditions. Given the complex and quantitative nature of yield-related traits in wheat, we hypothesize that these traits are controlled by multiple small- to moderate-effect loci distributed across the genome. Our GWAS approach using the FarmCPU model was designed to detect such loci while accounting for confounding factors. Therefore, this programmes may benefit more from environment-specific QTLs/MTAs.

The objective of this study was to identify QTLs for seven important grain-yield-related agronomic traits (PH, NFT, SL, SPS, KPS, TKW, and HI) using multi-locus GWAS. We also compared identified loci with that of previously published genes/QTLs, and identified high-confidence candidate genes in each QTL region.

Results

Phenotypic variation of agronomic and yield related traits

The current study evaluated 286 wheat genotypes at Eternal University, Baru Sahib, India, over two consecutive field seasons/environments (Env1: 2021–2022 and

Table 1 A statistical overview of the phenotypic variation seen in seven agronomic traits

Trait	Environment	Min	Max	Mean	Median	LSD	CV	SD	H^2
PH	Env1	51.0	113.0	84.4	84.7	8.1	5.9	4.9	0.68
	Env2	64.0	116.1	87.2	86.7	7.2	4.8	4.1	0.75
	CE	57.5	114.0	85.4	84.5	5.4	7.9	6.7	0.74
NFT	Env1	16.2	17.1	7.7	7.2	3.3	20.1	1.5	0.71
	Env2	17.4	18.1	10.2	10.2	2.9	19.8	2.0	0.53
	CE	16.8	17.5	9.0	9.2	3.5	19.1	1.7	0.45
SL	Env1	6.2	12.8	10.1	10.1	1.3	8.5	8.5	0.62
	Env2	6.0	12.7	9.8	9.8	1.2	6.9	6.6	0.81
	CE	6.1	12.6	9.9	10.0	1.0	7.5	7.4	0.78
SPS	Env1	11.0	21.5	17.4	17.5	1.8	7.4	1.2	0.50
	Env2	12.7	21.7	17.6	17.7	1.5	5.0	8.8	0.79
	CE	11.8	21.6	17.5	17.6	4.3	16.0	2.8	0.67
KPS	Env1	18.0	69.9	43.0	45.9	14.3	20.1	8.6	0.55
	Env2	23.0	70.0	48.0	50.1	11.7	14.4	6.9	0.71
	CE	20.5	70.9	45.5	48.0	7.5	18.2	8.2	0.51
TKW	Env1	20.0	69.1	47.2	47.8	8.2	10.5	4.9	0.68
	Env2	25.7	73.2	49.7	51.3	6.2	7.1	3.5	0.75
	CE	22.6	71.1	48.4	49.0	4.4	11.8	5.7	0.70
HI	Env1	20.9	85.5	39.5	40.4	7.7	12.6	4.9	0.62
	Env2	25.1	100.0	40.0	39.8	7.0	10.3	4.1	0.73
	CE	20.9	100.0	6.8	40.1	5.1	7.1	4.8	0.76

Min = Range minimum; Max = Range maximum; LSD = Least significant difference; CV = Coefficient of variance; SD = Standard deviation; $H_{2=}$ Heritability

Table 2 ANOVA of seven agronomic traits (PH, NFT, SL, SPS, KPS, TKW, and HI) in response to environment (Env), replicate (Rep), genotype (Gen), and their interaction (Env × Gen)

Source	Df	PH	NFT	SL	SPS	KPS	TKW	HI
Env	1	2229.8**	1755.3	19.9**	16.7**	377.9**	150.8**	77.9
Rep	3	230.0*	0.1	0.3	100.0*	1958.0*	2.6	125.0
Gen	285	134.6**	12.6*	4.1**	5.3**	305**	106.7**	118.5**
Env × Gen	285	17.3	10.5	0.4	2.3**	1.0	1.5	14.05

Signif. codes:0.001'**', 0.01'*'

Env2: 2022–2023). The broad-sense heritability (H^2) ranged from 0.53 to 0.81, with SL having the highest values (0.81) followed by the SPS (0.79), both in Env2, whereas the SPS also had the lowest values (0.50) in Env1, followed by the NFT (0.53) in Env2. The value of H^2 and descriptive statistics of the studied traits in both environments is given in Table 1. Analysis of variance (ANOVA) revealed highly significant variation among the genotypes for all the studied traits (Table 2). For PH, there was a highly significant genotype effect ($p < 0.001$). Similarly, significant genotype effects were observed for NFT ($F = 12.6$, $Df = 285$, $p < 0.01$), SL ($p < 0.001$), SPS) ($p < 0.001$), KPS ($p < 0.001$), TKW ($p < 0.001$), and HI ($p < 0.001$). The relative performance of genotypes for SPS and TKW varies with the environment due to a significant 'Environment × Genotype' (Env × Gen) interaction. However, the performance of different genotypes was the same in all environments for traits (PH, NFT, SL, KPS, and HI) where the interaction was not significant. This highlights how crucial it is to take particular environmental factors into account when choosing genotypes for SPS and TKW (Fig. 1; Table 2). The coefficient of variation (CV) ranged from 5.9 to 20.1% and from 4.8 to 19.8% in Env1 and Env2, respectively. Traits NFT and KPS showed high CV in both environments. The CVs for NFT and KPS were 19.2% and 19.8% in Env1, respectively, and 20.1% and 14.4% at Env2. The frequency distribution pattern was observed using the raincloud plot for evaluated traits showed a normal distribution, which indicates the traits are quantitative and complex in nature (Fig. 2). In each scenario, the NFT, SL, SPS, KPS, TKW, and HI traits appeared to tend towards higher values due to their positive skew. In the CE (data of combined environment) context, most features exhibit a slight positive skew or follow a normal distribution. This scenario implies that certain qualities might make CE more advantageous (Fig. 2A-G). Overall, sufficient phenotypic variation has also been observed in the WAMI panel to support GWAS analysis. The study found strong negative correlations between NFT, SPS, KPS, TKW, and

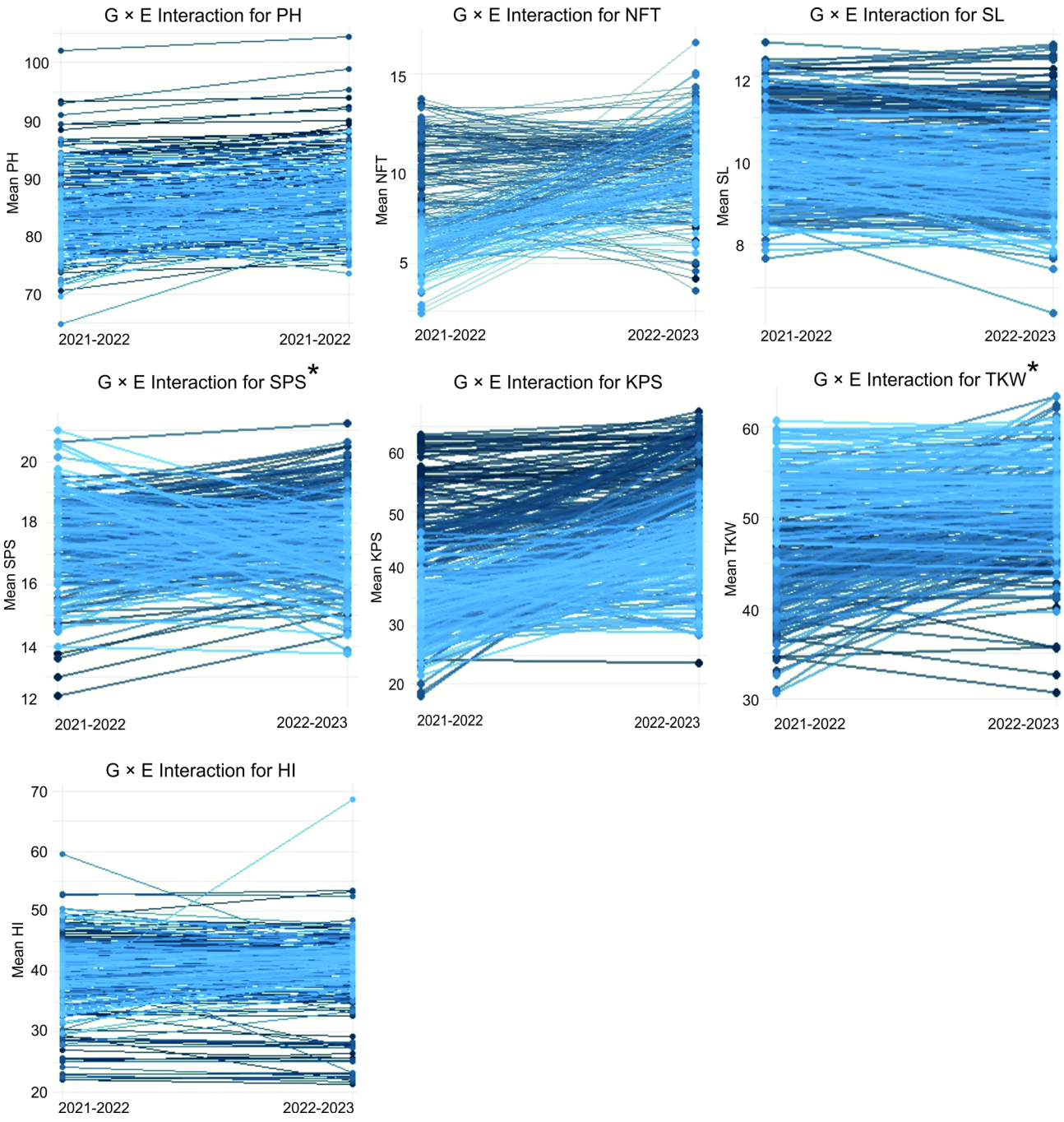

Fig. 1 A visual depiction of the genotype-environment interaction (**G** × **E**) for each of the seven traits in WAMI panel

HI, while SL had a weak positive correlation with KPS. A significant negative correlation has also been found between PH and NFT. SPS, and KPS showed weak positive correlations. TKW had a strong negative correlation with SL, and moderate negative correlations with KPS, and HI. The overall correlations were consistent with individual observations, providing a more robust understanding of relationships across environments (Fig. 2H). Similar results were obtained in both environments (Fig. S1). Principal component 1 (PC1; 40.3%), and principal component (PC2; 18.3%) accounted for most of the data variance, indicating significant factors influencing variability (Fig. 2I).

Population structure and LD analysis

Principal component analysis (PCA) showed a largely even distribution of allele frequencies within the panel (Fig. 3A). The first three PCs (PC1, PC2, and PC3)

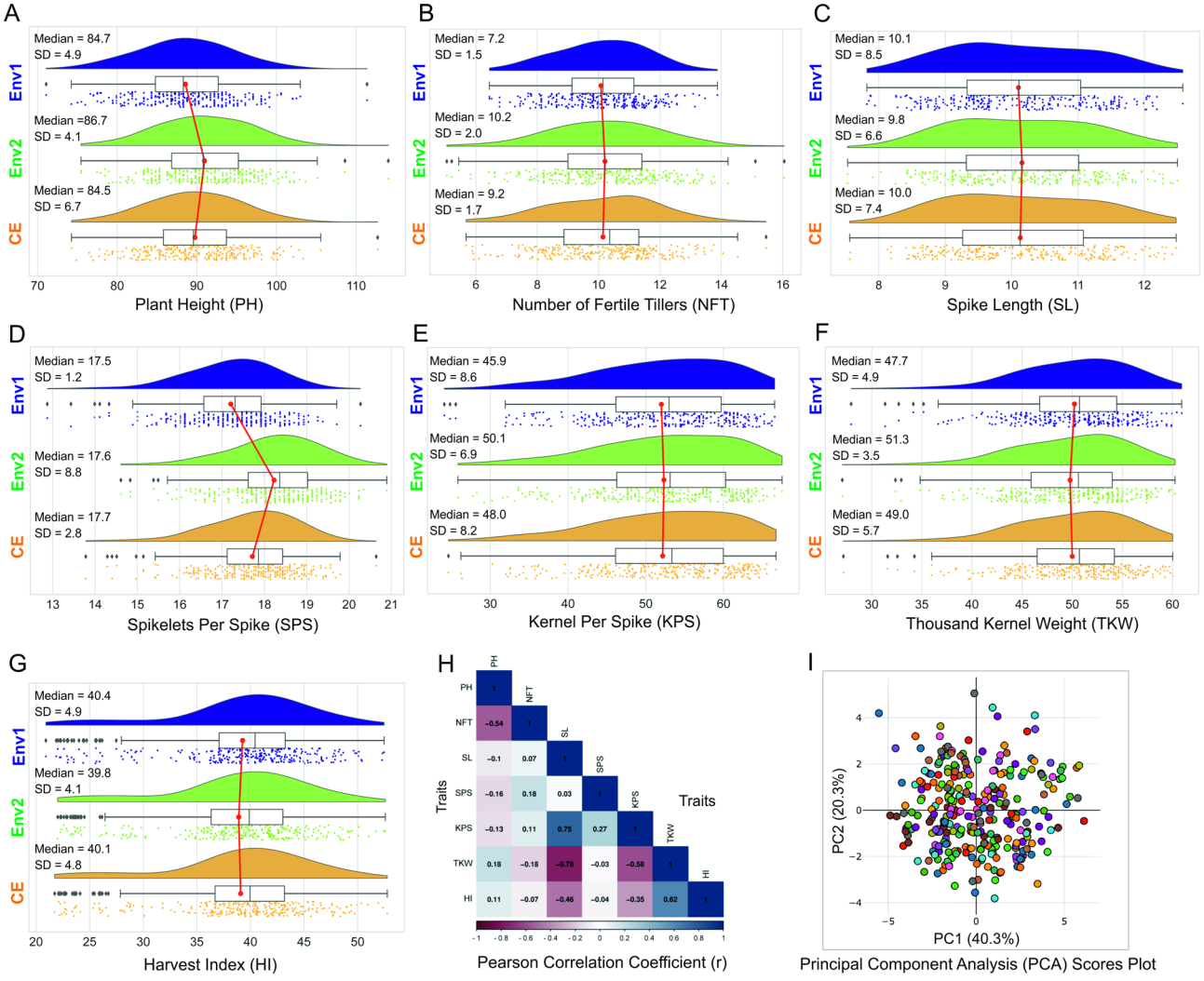

Fig. 2 (**A**-**G**) Figure presents raincloud plots illustrating the distribution of environmental factors across two independent environments: Environment 1 (Env1), Environment 2 (Env2), and their Combined Environment (**CE**) dataset. The individual data points are plotted below each density plot and box plot, giving a granular view of the data. A red line connects the medians of the three groups, illustrating the trend across the years. (**H**) This figure represents Pearson's correlation coefficient among the average data of each of the seven traits. (**I**) The figure presents the results of a PCA analysis performed on the agronomic data of seven traits, demonstrating the high diversity among the investigated panel

accounted for 43.3%, 18.3%, and 15.4% of the total variance, respectively. So, using PC1, PC2, and PC3 together showed a large part of the genetic differences in the wheat association mapping initiative (WAMI) panel, helping to find, and separate different groups. The neighbour-joining dendrogram in circular form, with 286 wheat genotypes, 20,996 SNPs, and 14 clusters, illuminates the varied genetic profiles of the examined wheat varieties, and provides significant insights into their common, and distinct genetic traits (Fig. 3B). The LD analysis was done with TASSEL in a 100 sliding window size for all chromosomes except 3D because no MTA was found on it. The genome had an average squared allele-frequency correlations (r^2) of 0.12, and LD decay began at 0.48, and reached half-decay at 0.23 (average for all the chromosome). The LD decay curve crossed the half-decay point, and the average standard critical point (r^2 D 0.3) at distances of LD, and half LD (Mb), for each chromosome respectively (Fig. 4). The highest LD decay distance of 171.3 Mb was found in the 2 A chromosome, followed by 38.8 Mb in the 4 A chromosome. However, several chromosomes have the lowest LD decay distance of 10.0 Mb (Fig. 4). Half decay for each chromosome was computed independently in this study. This value sets the important distance across the entire genome for finding connections, and quantitative trait nucleotide (QTNs) within this distance are treated as one QTL [53].

Associations between traits and SNPs

The phenotypic traits at physiological maturity, and SNP markers underwent GWAS. In total, 300 MTAs were identified for seven phenotypic traits using the three

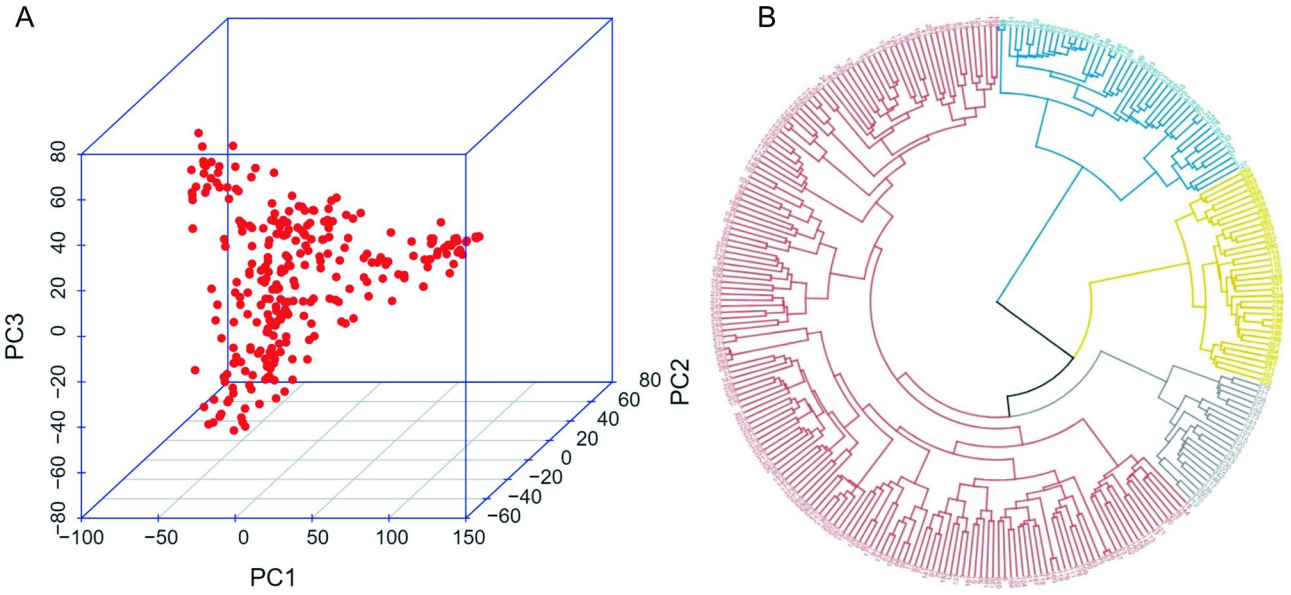

Fig. 3 (**A**) Illustration of a three-dimensional scatter plot representing genotypes, with the plot based on the first three main PCs (PC1, PC2, and PC3). (**B**) A phylogenetic network was created using the ward D2 linkage method. A hierarchical clustering analysis was performed using 20,996 SNPs on 286 genotypes from the WAMI panel and breeding lines. The labels of the breeding lines were color-coded to represent four distinct groupings

datasets (Env1, Env2, and CE) at $P < 0.001$ (Table S1-S7). A total of 187 SNPs demonstrated significant associations with the same trait across two or more datasets, while six SNPs exhibited significant associations with two or more traits (Table S8). Using half LD analysis, we found 114 QTLs, which are linked to different traits: 11 for PH, 28 for NFT, 20 for SL, 19 for SPS, 10 for KPS, 12 for TKW, and 14 for HI (Table S1-S7).

Plant height (PH)

Eleven QTLs linked to PH were found mainly on chromosomes 1D, 2 A, 2B 2D, 3B, 4B, and 6B (Figs. 5 and 6, Fig. S2-S5, Table 3, and Table S1). Eight of these SNPs linked with QTL were regularly seen in at least one of the single-environment, and combined environment (CE) analyses. These SNPs correspond to seven QTLs: *QPh.eu-2 A* (1 SNP), *QPh.eu-2D* (1 SNP), *QPh.eu-2D.1* (2 SNP), *QPh.eu-2D.3* (1 SNPs), *QPh.eu-3B* (1 SNP), *QPh.eu-3B.1* (1 SNP), and *QPh.eu-6B* (1 SNP) (Figs. 5 and 6, Fig. S2-S5, Tables 3 and 3, and Table S1). Among these QTLs, *QPh.eu-3B.1*, located on chromosome 3B at 783.5 Mb, is strongly associated with PH, explaining a substantial 12.1% of the phenotypic variation. QTL *QPh.eu-2D* followed closely, accounting for 11.3% (Table 3). In contrast, one SNP from *QPh.eu-4B* on chromosome 4B positioned at 37.9 Mb demonstrated a minimal contribution, explaining only 2.7% of the phenotypic variation (Table 3). Of the eleven QTLs, six (*QPh.eu-2B*, *QPh.eu-2D.1*, *QPh.eu-2D.2* *QPh.eu-2D.3*, *QPh.eu-3B*, and *QPh.eu-4B*) appeared potentially novel, distinct from previously identified QTLs/MTAs (Figs. 5 and 6, Fig. S2-S5, Table 3 and Table S1).

Number of fertile tillers (NFT)

Seventy important SNPs (associated with 28 QTLs) were found to be linked to the NFT, spread out across different chromosomes. In particular, chromosomes 6B (26 SNPs), 5B (12 SNPs), and 2 A (8 SNPs) had the most SNPs. Notably, chromosomes 6B (26 SNPs), 5B (12 SNPs), and 2 A (8 SNPs) showed prominent numbers (Figs. 5 and 6, Fig. S2-S5, Table 3 and Table S2). The QTLs such as *QNft.eu-6B.2* with 21 SNPs, *QNft.eu-2 A* with 8 SNPs, and *QNft.eu-5B* with 7 SNPs were located on various chromosomes, indicating that many QTLs affect the trait (Table 3, and Table S2). The QTL *QNft.eu-2 A* was especially important, accounting for the largest share of phenotypic variation explained (PVE) at 9%, with *QNft.eu-1B* and *QNft.eu-2B.2* both were close behind at 7.0%. On the other hand, the QTL *QNft.eu-2B*, found at 18.9 on chromosome 2B, had the least impact, explaining only 1.0% of the phenotypic variation (Table 3). Remarkably, among the 28 identified QTLs associated with NFT, 24 are potentially novel discoveries, as no corresponding QTLs or markers were reported in previous studies related to same trait (Figs. 5 and 6, Fig. S2-S5, Table 3 and Table S2).

Spike length (SL)

A total of 69 SNPs were linked to SL, and found on chromosomes 1 A (6 SNPs), 1B (1 SNP), 2 A (6 SNPs), 2B (13 SNPs), 2D (35 SNPs), 5B (3 SNPs), 6D (1 SNP), 7 A (1 SNP), and 7B (3 SNPs) (Figs. 5 and 6, Fig. S2-S5, Table 3

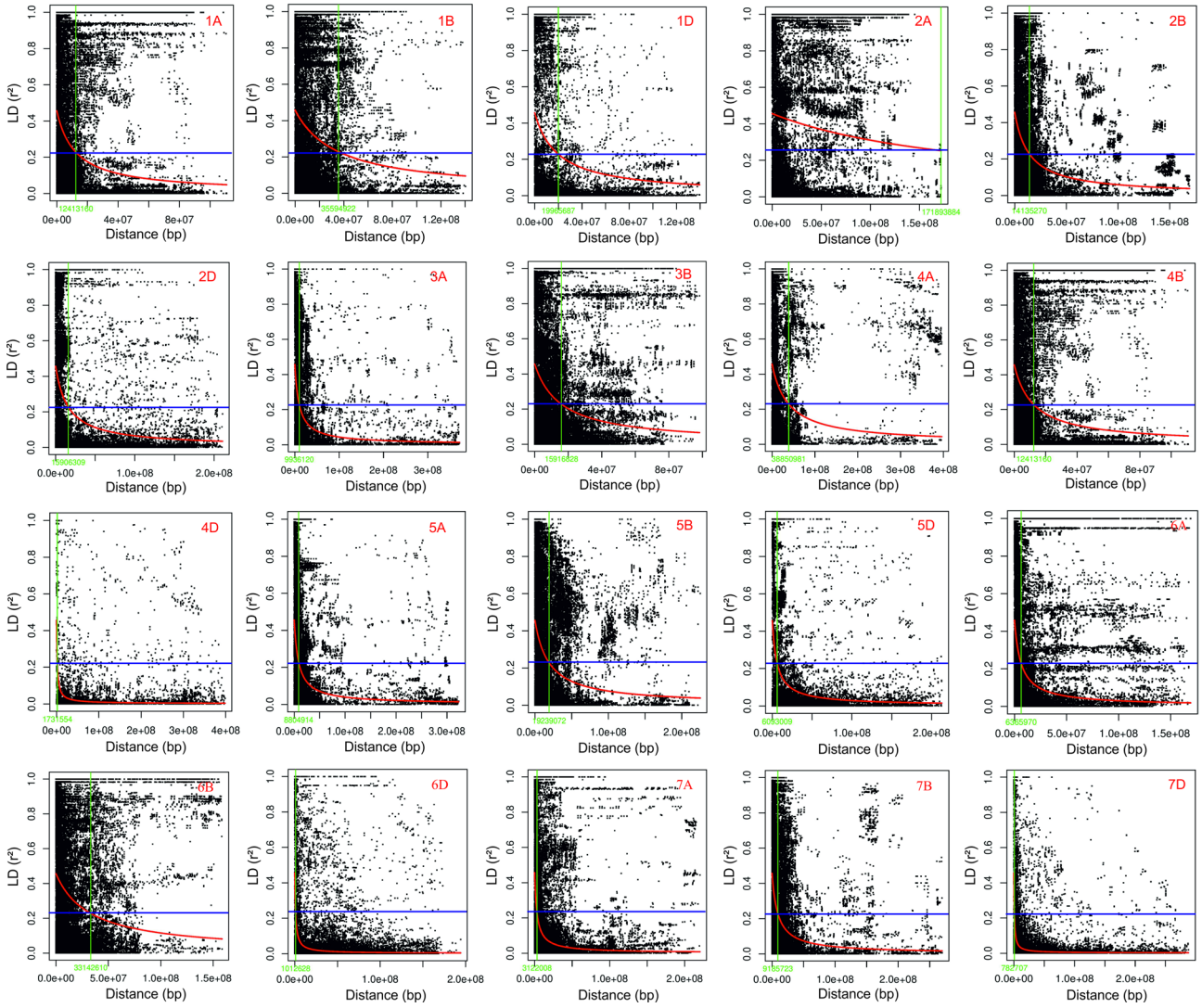

Fig. 4 Scatter plot illustrating linkage disequilibrium (LD) decay on the 20 chromosomes of wheat (except 3D, where no significant MTA was found). The plot displays the physical distance in base pair(bp) against the LD estimate (R^2) for marker pairs

and Table S3). Among these, 19 SNPs were consistently identified in two different environments, and in CE (Figs. 5 and 6, Fig. S2-S5, Table 3 and Table S3). Simultaneously, 32 SNPs were identified in one environment, and in CE (Table S3). These SNPs underlie 20 QTLs, specifically *Qsl.eu-1 A* (1 SNP), *Qsl.eu-1 A.1* (5 SNPs), *Qsl.eu-1B* (1 SNP), *Qsl.eu-2 A* (6 SNPs), *Qsl.eu-2B* (3 SNPs), *Qsl.eu-2B.1* (6 SNPs), *Qsl.eu-2B.2* (2 SNPs), *Qsl.eu-2B.3* (2 SNPs), *Qsl.eu-2B.4* (1 SNP), *Qsl.eu-2D* (1 SNP), *Qsl.eu-2D.1* (12 SNPs), *Qsl.eu-2D.2* (13 SNPs), *Qsl.eu-2D.3* (6 SNPs), *Qsl.eu-2D.4* (3 SNPs), *Qsl.eu-5B* (1 SNP), *Qsl.eu-5B.1* (2 SNPs), *Qsl.eu-6D* (1 SNP), *Qsl.eu-7 A* (1 SNP), *Qsl.eu-7B* (2 SNPs), and *Qsl.eu-7B.1* (1 SNP) (Table S3). The QTL *QSl.eu-1 A* showed the most significant impact on phenotypic variation at 6%, while *QSl.eu-1 A.1*, *QSl.eu-1B*, *QSl.eu-2 A*, *QSl.eu-2D.1*, and *QSl.eu-7B* each accounted for 5.0% of the variation. In contrast, the QTLs *QSl.eu-2B.4*, *QSl.eu-2D.2*, and *QSl.eu-5B* had the least impact, explaining only 1.0% of the variation. (Table 3). Out of the 20 QTLs found for this trait, eight were located in the same or nearby areas as those found in an earlier study for the same trait, while the other 12 QTLs might be new genetic locations.

Spikelets per Spike (SPS)

A total of 82 important SNPs were linked to SPS, mostly found on chromosomes 1 A (54 SNPs), 2 A (1 SNP), 2B (7 SNPs), 4 A (7 SNPs), 4B (1 SNP), 4D (1 SNP), 5B (9 SNPs), 5D (1 SNP), and 7D (1 SNP) (Figs. 5 and 6, Fig. S2-S5, Table 3 and Table S4). These SNPs collectively contributed to the identification of 20 QTLs: *Qsps.eu-1 A* (1 SNP), *Qsps.eu-1 A.1* (5 SNPs), *Qsps.eu-1 A.2* (1 SNP), *Qsps.eu-1 A.3* (8 SNPs), *Qsps.eu-1 A.4* (21 SNPs), *Qsps.eu-1 A.5* (4 SNPs), *Qsps.eu-1 A.6* (12 SNPs), *Qsps.*

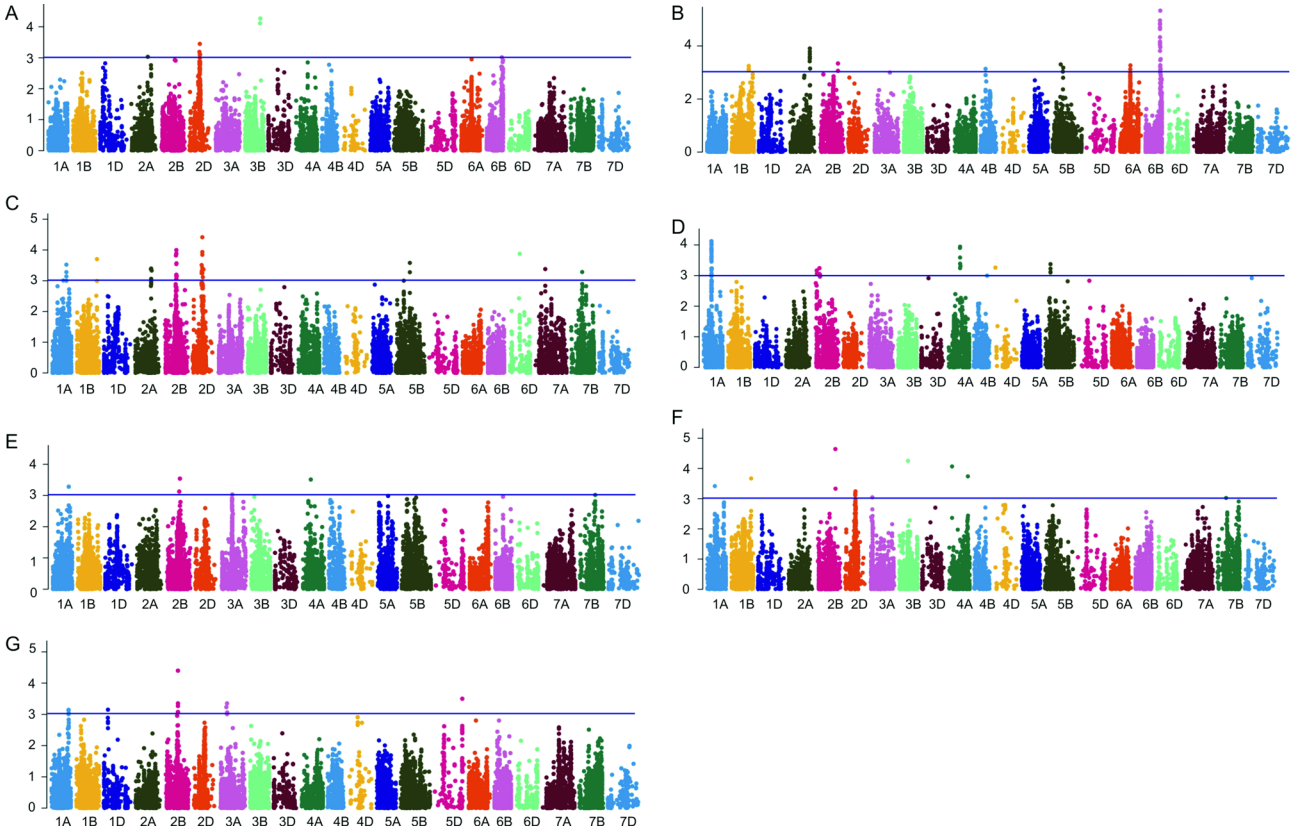

Fig. 5 Manhattan plot for grain yield related seven agronomic traits in 286 wheat accessions using the combined BLUP value of two environment's data (**CE**). (**A**) PH, (**B**) NFT, (**C**) SL, (**D**) SPS, (**E**) KPS, (**F**) TKW and (**G**) HI. The various chromosomes are indicated by different colours. The horizontal line is drawn at a specific y-value to indicate the threshold of statistical significance

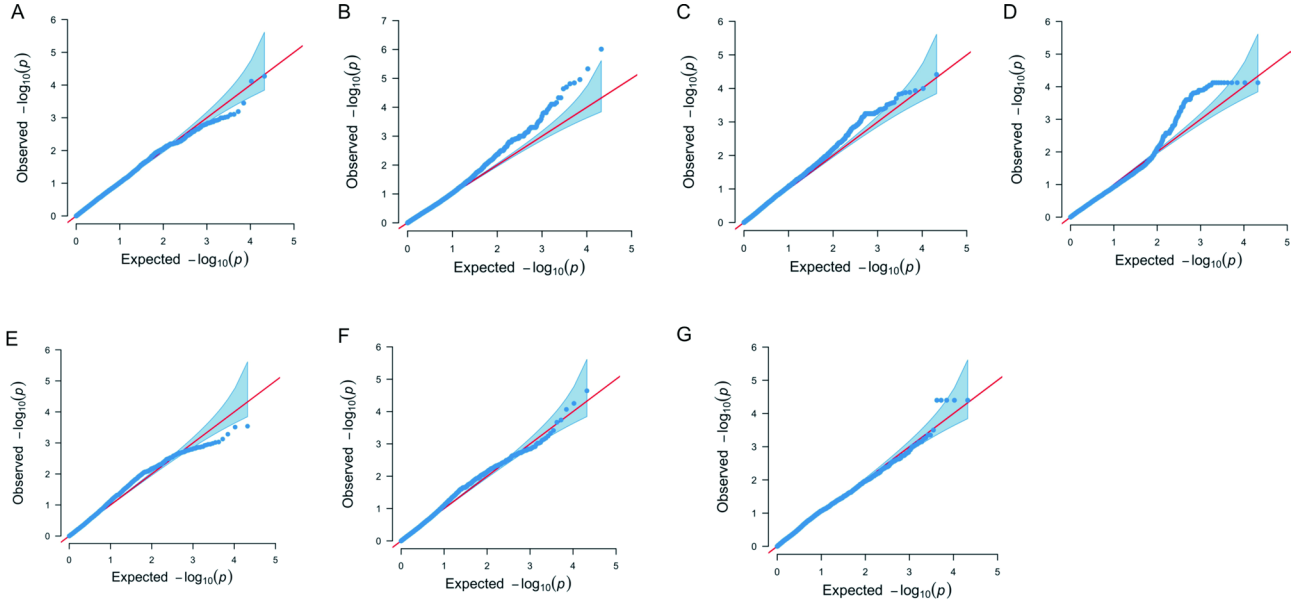

Fig. 6 Q-Q plot for grain yield related seven agronomic traits in 286 wheat accessions the combined BLUP value of two environment's data (**CE**). (**A**) PH, (**B**) NFT, (**C**) SL, (**D**) SPS, (**E**) KPS, (**F**) TKW and (**G**) HI

Table 3 Summary of QTLs with number of SNP(s) per QTL, chromosomal location, genetic position, phenotypic variation explained (PVE%), and environment, along with references to previously identified qtls/mtas in proximity

QTL	No. of SNP within QTL	Chr	Genetic position (cM)	Physical position (Mb)	Average PVE (%)	Environment	Previously identified QTL/MTAs or SNP on the same or nearby location	Reference
Plant height (PH)								
Qph.eu-1D	1	1D	39	12.5	5.00	Env1	PH_AX-94,418,622	[75]
Qph.eu-2 A	1	2 A	124	716.5	4.00	Env1, CE	PH_AX_95085564, qPH2A.2	[3, 45]
Qph.eu-2B	1	2B	109	571.2	10.60	Env2	PNQ	
Qph.eu-2D	3	2D	77–80	571.0-574.3	10.60	Env1, Env2, CE	PH_KukRi_c205_223	[44]
Qph.eu-2D.1	3	2D	80–82	608.7-612.1	9.40	Env1, Env2, CE	PNQ	
Qph.eu-2D.2	1	2D	78	709.8	6.60	Env2	PNQ	
Qph.eu-2D.3	1	2D	80	742.5	7.10	Env2, CE	PNQ	
Qph.eu-3B	1	3B	117	587.0	10.20	Env1, Env2, CE	PNQ	
Qph.eu-3B.1	1	3B	119	783.5	12.10	Env1, Env2, CE	*QPH.sicau-3BL.1*	[76]
Qph.eu-4B	1	4B	56	37.9	2.70	Env2	PNQ	
Qph.eu-6B	1	6B	113	611.6	3.00	Env2, CE	*qPH6B.3*, PH_Kukri_rep_c106092_300	[45, 78]
Number of fertile tiller (NFT)								
Qnft.eu-1 A	1	1 A	51	25.5	3.30	Env1	PNQ	
Qnft.eu-1B	1	1B	134	476.3	7.00	CE	GENE-0035_150	[47]
Qnft.eu-1B.1	2	1B	134	662.5-662.6	6.40	CE	PNQ	
Qnft.eu-2 A	8	2 A	148–151	753.6-759.8	9.00	Env1, CE	BobWhite_c11479_157	[24]
Qnft.eu-2B	1	2B	32	18.9	1.00	Env2	PNQ	
Qnft.eu-2B.1	1	2B	107	637.6	2.00	Env1	PNQ	
Qnft.eu-2B.2	2	2B	142	759.8-765.3	7.00	CE	PNQ	
Qnft.eu-2D	1	2D	11	18.8	6.00	Env2	D_F1BEJMU01DOWJ3_176	[47]
Qnft.eu-2D.1	1	2D	80	607.4	4.00	Env1	PNQ	
Qnft.eu-3 A	1	3 A	153	32.9	2.00	Env1	GENE-0826_51	[47]
Qnft.eu-3B	1	3B	56	60.9	2.00	Env1	PNQ	
Qnft.eu-4B	2	4B	39–44	15.2–20.7	5.00	Env1, CE	PNQ	
Qnft.eu-5B	7	5B	60	482.4-484.7	4.00	Env2, CE	PNQ	
Qnft.eu-5B.1	2	5B	60	510.9	3.10	Env2, CE	PNQ	
Qnft.eu-5B.2	1	5B	76	539.1	4.00	CE	PNQ	
Qnft.eu-5B.3	1	5B	79	559.1	4.00	Env2, CE	PNQ	
Qnft.eu-5B.4	1	5B	71	705.4	4.00	Env1	PNQ	
Qnft.eu-6 A	1	6 A	79	234.0	4.00	CE	PNQ	
Qnft.eu-6 A.1	2	6 A	79	249.1-252.8	3.30	CE	PNQ	
Qnft.eu-6 A.2	1	6 A	79	288.4	3.30	CE	PNQ	
Qnft.eu-6 A.3	2	6 A	79	320.4-322.1	3.30	CE	PNQ	
Qnft.eu-6 A.4	1	6 A	79	367.1	3.30	CE	PNQ	
Qnft.eu-6 A.5	1	6 A	79	381.0	3.30	CE	PNQ	
Qnft.eu-6 A.6	1	6 A	79	399.9	3.30	CE	PNQ	
Qnft.eu-6 A.7	1	6 A	80	454.6	4.00	CE	PNQ	
Qnft.eu-6B	4	6B	108–109	461.9–462.0	6.00	CE	PNQ	
Qnft.eu-6B.1	1	6B	110	609.5	3.00	CE	PNQ	
Qnft.eu-6B.2	21	6B	108–113	703.2–710.0	2.00	Env1, Env2, CE	PNQ	
Spike length (SL)								
Qsl.eu-1 A	1	1 A	95	530.2	6.00	Env2, CE	*qSL1A.1*	[45]
Qsl.eu-1 A.1	6	1 A	118	516.9	5.00	Env1, Env2, CE	AX-94,725,964	[75]
Qsl.eu-1B	1	1B	158	487.2	5.00	Env1, Env2, CE	PNQ	
Qsl.eu-2 A	6	2 A	122–128	716.2-722.8	5.00	Env1, Env2, CE	SL_AX-94,625,204	[75]
Qsl.eu-2B	3	2B	91–93	139.8–152.0	4.00	Env1, Env2, CE	SL_Ra_c27275_627	[79]
Qsl.eu-2B.1	6	2B	93–96	164.0-174.0	3.00	Env1, Env2, CE	wsnp_Ex_c62844_62315607	[79]
Qsl.eu-2B.2	2	2B	96	179.4-184.3	3.00	Env2, CE	PNQ	
Qsl.eu-2B.3	1	2B	97	202.7	3.00	Env1, CE	PNQ	

Table 3 (continued)

QTL	No. of SNP within QTL	Chr	Genetic position (cM)	Physical position (Mb)	Average PVE (%)	Environment	Previously identified QTL/MTAs or SNP on the same or nearby location	Reference
Qsl.eu-2B.4	1	2B	99	513.5	3.00	Env1, Env2, CE	RFL_Contig4718_1269	[38]
Qsl.eu-2D	1	2D	80	122.3	1.00	Env1	PNQ	
Qsl.eu-2D.1	12	2D	71–76	579.5-587.7	5.00	Env1, Env2, CE	PNQ	
Qsl.eu-2D.2	13.0	2D	71–80	590.0-597.0	3.00	Env1, CE	PNQ	
Qsl.eu-2D.3	6	2D	70–76	710.9-720.6	5.00	Env1, Env2, CE	PNQ	
Qsl.eu-2D.4	3.0	2D	71–76	726.0-729.3	5.00	Env1, Env2, CE	PNQ	
Qsl.eu-5B	1	5B	61	479.7	1.00	Env1, CE	wsnp_Ex_rep_c69631_68583363	[38]
Qsl.eu-5B.1	2	5B	107	577.3	4.00	Env1, CE	AX-108,931,712, QSL.ta.ipbb-5B	[52, 80]
Qsl.eu-6D	1	6D	82	95.7	4.00	Env1, Env2, CE	PNQ	
Qsl.eu-7 A	1	7 A	80	48.1	4.00	Env1, Env2, CE	PNQ	
Qsl.eu-7B	2	7B	83	599.4-605.5	5.00	Env1, Env2, CE	PNQ	
Qsl.eu-7B.1	1	7B	113	675.3	2.00	Env1	PNQ	
Spiklets per spike (SPS)								
Qsps.eu-1 A	1	1 A	71	261.7	5	Env1, CE	PNQ	
Qsps.eu-1 A.1	5	1 A	70–71	290.4-296.4	5	Env1, Env2, CE	PNQ	
Qsps.eu-1 A.2	1	1 A	71	325.0	5	Env1, CE	PNQ	
Qsps.eu-1 A.3	8	1 A	71	335.1-341.9	5	Env1, Env2, CE	PNQ	
Qsps.eu-1 A.4	21.0	1 A	70–71	348.9-396.5	5	Env1, Env2, CE	PNQ	
Qsps.eu-1 A.5	4.0	1 A	70–74	357.2-362.9	4	Env1, Env2, CE	PNQ	
Qsps.eu-1 A.6	13.0	1 A	70–71	364.0-370.1	5	Env1, Env2, CE	PNQ	
Qsps.eu-1 A.7	1.0	1 A	70	388.0	4	Env1, Env2, CE	PNQ	
Qsps.eu-2 A	1	2 A	142	608.8	2.00	Env1	PNQ	
Qsps.eu-2B	4	2B	10–19	2.4–6.2	3.00	Env2, CE	PNQ	
Qsps.eu-2B.1	3.0	2B	38–42	25.0-25.2	4.00	Env2, CE	PNQ	
Qsps.eu-4 A	1	4 A	95	97.6	3.00	Env1, CE	PNQ	
Qsps.eu-4 A.1	1	4 A	95	537.7	3.00	Env1, Env2, CE	PNQ	
Qsps.eu-4 A.2	4	4 A	94–95	629.9-631.9	3.00	Env1, Env2, CE	IAAV4112	[56]
Qsps.eu-4 A.3	1	4 A	95	678.3	3.00	Env1, CE	PNQ	
Qsps.eu-4B	1	4B	101	606.6	4.00	Env2	tplb0026o15_1634	[24]
Qsps.eu-4D	1	4D	9	1.9	3.00	Env2, CE	PNQ	
Qsps.eu-5B	9	5B	39	40.7–40.8	2.00	Env2, CE	PNQ	
Qsps.eu-5D	1	5D	80	390.3	4.00	Env2	PNQ	
Qsps.eu-7D	1	7D	29	16.2	6.00	Env2	AX-111,622,533	[73]
Kernels per spike (KPS)								
Qkps.eu-1 A	1	1 A	130	566.5	3.00	Env1, CE	PNQ	
Qkps.eu-2B	1	2B	99	412.7	1.00	Env1, Env2, CE	PNQ	
Qkps.eu-2B.1	1	2B	104	558.3	5.00	Env1, Env2, CE	PNQ	
Qkps.eu-3 A	8	3 A	88–90	556.0-562.3	2.00	Env2, CE	*QKNS.sicau-3AL.1*	[76]
Qkps.eu-3B	1	3B	25	662.0	6.00	Env2	PNQ	[43]
Qkps.eu-4 A	1	4 A	66	605.7	8.00	Env1, Env2, CE	PNQ	
Qkps.eu-5 A	1	5 A	82	564.2	2.00	Env2	*qGN5A.2*	[45]
Qkps.eu-5B	1	5B	39	40.8	1.00	Env2	PNQ	
Qkps.eu-6B	1	6B	56	134.6	3.00	Env1	PNQ	
Qkps.eu-7B	1	7B	113	675.3	3.00	Env1, CE	PNQ	
Thousand kernel weight (TKW)								
Qtkw.eu-1 A	1	1 A	70	254.4	4.00	Env1, Env2, CE	ExcalibuR_c6255_1119	[44]
Qtkw.eu-1B	1	1B	158	673.7	5.00	Env1, Env2, CE	*qTGW1B.1*	[45]
Qtkw.eu-2B	2	2B	144–145	775.8-777.1	5.00	Env1, Env2, CE	RAC875_c10626_2089	[24]
Qtkw.eu-2D	2	2D	80–82	610.2-613.3	9.00	Env1, Env2, CE	Ex_c10068_1509	[56]
Qtkw.eu-2D.1	3	2D	80	610.2-613.3	9.00	Env1, Env2, CE	PNQ	
Qtkw.eu-2D.2	1	2D	82	730.6	8.50	Env1	PNQ	

Table 3 (continued)

QTL	No. of SNP within QTL	Chr	Genetic position (cM)	Physical position (Mb)	Average PVE (%)	Environment	Previously identified QTL/MTAs or SNP on the same or nearby location	Reference
Qtkw.eu-3 A	1	3 A	24	13.5	1.20	Env1, CE	PNQ	
Qtkw.eu-3B	1	3B	70	574.7	10.00	Env1, Env2, CE	PNQ	
Qtkw.eu-4 A	1	4 A	29	12.2	5.00	Env1, Env2, CE	PNQ	
Qtkw.eu-4 A.1	1	4 A	151	737.3	5.00	Env1, Env2, CE	wsnp_Ra_c22775_32274079	[24, 75]
Qtkw.eu-7B	2	7B	67	134.5-136.6	1.00	Env2, CE	AX-94,761,577	
Qtkw.eu-7B.1	3	7B	164	739.3-740.1	1.10	Env1	PNQ	
Harvest index (HI)								
Qhi.eu-1 A	5	1 A	139–140	580.4-581.8	4.00	Env2, CE	BobWhite_c12960_138	[24]
Qhi.eu-1B	1	1B	39	15.3	3.00	Env2	PNQ	
Qhi.eu-1B.1	1	1B	63	255.7	3.00	Env2	PNQ	
Qhi.eu-1D	1	1D	32	10.7	2.00	Env1, CE	PNQ	
Qhi.eu-2B	1	2B	93	155.0	3.00	Env2	wsnp_Ex_c30_66389	[24]
Qhi.eu-2B.1	2	2B	99	244.5-251.8	3.00	Env1, Env2, CE	PNQ	
Qhi.eu-2B.2	1	2B	99	298.8	3.00	Env1, Env2, CE	PNQ	
Qhi.eu-2B.3	5	2B	99	438.9-439.8	3.00	Env1, Env2, CE	PNQ	
Qhi.eu-2D	2	2D	80–86	608.7-612.5	5.00	Env2	RAC875_c5673_1209	[24]
Qhi.eu-3 A	4	3 A	55–61	36.4–38.0	3.00	Env1, Env2, CE	PNQ	
Qhi.eu-4D	1	4D	69	38.3	5.00	Env2	PNQ	
Qhi.eu-4D.1	1	4D	101	659.9	1.00	Env2	PNQ	
Qhi.eu-5D	1	5D	204	694.2	1.00	Env1, Env2, CE	PNQ	
Qhi.eu-7 A	2	7 A	126	132.5	3.00	Env2	PNQ	

Chr = Chromosome; PNQ = potentially novel QTL

eu-1 A.7 (1 SNP), *Qsps.eu-2 A* (1 SNP), *Qsps.eu-2B* (4 SNPs), *Qsps.eu-2B.1* (3 SNPs), *Qsps.eu-4 A* (1 SNP), *Qsps.eu-4 A.1* (1 SNP), *Qsps.eu-4 A.2* (5 SNP), *Qsps.eu-4 A.3* (1 SNP), *Qsps.eu-4B* (1 SNP), *Qsps.eu-4D* (1 SNP), *Qsps.eu-5B* (9 SNPs), *Qsps.eu-5D* (1 SNP), and *Qsps.eu-7D* (1 SNPs) (Figs. 5 and 6, Fig. S2-S5, Table 3 and Table S4). The QTL *QSps.eu-7D* explained the most variation in traits at 6.0%, while the QTL *QSps.eu-2B*, explained the least at 1.0%. Among the 20 identified QTLs for this trait, 17 were likely novel discoveries, potentially representing previously unrecognised genetic loci contributing to spikelet numbers per spike (Figs. 5 and 6, Fig. S2-S5, and Table S4).

Kernel per Spike (KPS)

In a comprehensive analysis, 17 SNPs displayed statistically significant associations with the KPS. These genetic markers were predominantly situated across various chromosomes, with notable concentrations on chromosome 3 A (8 SNPs) (Figs. 5 and 6, Fig. S2-S5, Table 3 and Table S5). These SNPs were linked to 11 different QTLs, named *Qkps.eu-1 A* (1 SNP), *Qkps.eu-2B* (1 SNP), *Qkps.eu-2B.1* (1 SNP), *Qkps.eu-3 A* (8 SNPs), *Qkps.eu-3B* (1 SNP), *Qkps.eu-4 A* (1 SNP), *Qkps.eu-5 A* (1 SNP), *Qkps.eu-5B* (1 SNP), *Qkps.eu-6B* (1 SNP), and *Qkps.eu-7B* (1 SNP), all of which play a role in the genetics of KPS. One important QTL is *QKps.eu-4 A*, located at 605.7 Mb on chromosome 4 A, which explained the most variation in traits at 8.0%. In contrast, *QKps.eu-1B*, and *QKps.eu-5B* showed the least variation at 1.0%. Conversely, *QKps.eu-2B*, and *QKps.eu-5B* demonstrated the lowest phenotypic variation at 1.0%. Interestingly, Eight of the identified QTLs for KPS appear to be potentially novel, suggesting a novel genetic architecture influencing this trait (Figs. 5 and 6, Fig. S2-S5, Table 3 and Table S5).

Thousand kernels weight (TKW)

Nineteen significant SNPs were significantly associated with TKW. These SNPs were spread out across the genome, with six linked to TGW on chromosome 2D, five on 7B, two each on 2B, and 4 A, and one each on 1 A, 1B, 3 A, and 3B (Figs. 5 and 6, Fig. S2-S5, Table 3 and Table S6). These 19 SNPs collectively contributed to the phenotypic variation in TKW, with the PVE ranging from 1.0 to 10.0% (Table 3). Additionally, these SNPs underlie 12 QTLs, each associated with specific genomic positions. The QTL *QTkw.eu-3B*, found at 574.7 Mb on chromosome 3B, had the highest PVE (10.0%), while *QTkw.eu-7B.1*, located within the 739.4 to 740.1 Mb on chromosome 7B, had the lowest PVE at 1.0% (Table 3). Importantly, seven of the 12 identified QTLs, which are.

Qtkw.eu-2D, *Qtkw.eu-2D.2*, *Qtkw.eu-3 A*, *Qtkw.eu-3B*, *Qtkw.eu-4 A*, and *Qtkw.eu-4 A.1* and *Qtkw.eu-7B.1*) were not found in earlier studies, indicating that they may be new discoveries.

Harvest index (HI)

A total of 28 important SNPs were linked to HI, showing a wide spread across different chromosomes, especially on chromosomes 2B (9 SNPs), 1 A (5 SNPs), and 3 A (4 SNPs) (Table 3, and Table S7). These SNPs collectively contributed to the identification of 14 QTLs. The QTLs *QHi.eu-1 A* and *QHi.eu-2B.3* emerged as particularly noteworthy, being associated with the maximum number of SNPs (each 5), while *QHi.eu-3 A* followed closely with four associated SNPs. The PVE reached its highest at 5.0%, and was linked to the QTLs *QHi.eu-2D*, and *QHi.eu-4D*, reflecting their substantial influence on the trait (Table 3). Conversely, the QTL *QHi.eu-4D.1* exhibited the lowest PVE at 1.0%. Of particular interest is the potential novelty of the identified QTLs for HI, as 11 out of the 14 do not align with any previously reported QTLs or MTAs associated with the same trait in wheat (Figs. 5 and 6, Fig. S2-S5, Table 3 and Table S7). This suggests that these QTLs may represent potentially novel genetic loci governing HI in wheat.

Traits having common associations

The dataset revealed associations between traits, the number of shared SNPs, and the corresponding environments (Table S8). A single SNP (Kukri_c36783_91) was identified for HI in Env2, and CE and for SL in Env1, and Env2. Another SNP (Excalibur_c18353_55) was detected in environments Env2, and CE for HI, and Env1, and Env2 for PH. A common SNP (Kukri_c11141_203) was found in CE, and Env2 for KPS, and Env1 for SL. Similarly, a shared SNP (Excalibur_rep_c108100_149) was identified in Env2 for KPS, and CE, and Env2 for SPS. A common SNP (wsnp_Ex_c59095_60108118) was detected in CE, and Env1 for PH, and CE, and Env2 for SL. Lastly, a single SNP (TA001163-0861) was shared in CE, and Env1 for SL, and CE, and Env1 for TKW (Table S8). These connections, and shared SNPs help us understand how the specific traits are related to each other in different environments, which adds to our knowledge of the genetic structure in the population we studied.

Discussion

Any crop improvement programme requires the evaluation and characterisation of yield-contributing traits. Evaluation and characterisation lead to the identification of superior genotypes, which provide breeders with the most vital input for successful crop breeding. The study revealed that the wheat germplasm exhibits a higher degree of variation for economically significant traits like PH, NFT, SL, SPS, KPS, TKW, and HI. Breeders detect, and use new genes, and QTLs that control these essential traits. However, a limited number of studies were conducted on wheat on this subject [52, 54–56]. The analysis of agronomic traits involved data from two separate years, as well as a combined dataset, to pinpoint QTLs. We recognise the significance of broader environmental representation; however, our GWAS concentrated on replicated phenotypic data from these two specific environments. This method enabled us to focus on and pinpoint stable QTL by taking into account only those loci that were consistently observed in both environments. We agree that including phenotypic data from various environments in future studies would greatly strengthen the robustness and validation of the identified QTL, and we intend to pursue that goal in subsequent research. This methodology is consistent with earlier investigations, including those by Malik et al. [48], which similarly documented noteworthy findings using two years of agronomic data.

The broad-sense heritability of the agronomic trait in this study ranged from 53 to 81% across the environment. The heritability range suggests a significant influence of genetics on these traits, which makes them suitable for a breeding program. Although the environment contributes to a certain extent (4.8–20.1% CV), it is essential to consider both genetic, and environmental factors to maximise crop performance (Table 1) [57–59].During the Env × Gen interaction, we have found a significant environment interaction with SPS and KPS that show a significant interaction between environment and genotypes, meaning that the QTLs influencing these traits are not consistent. To identify broadly adapted QTLs or account for environmental specificity, targeted breeding strategies are required. However, the lack of significant interaction for PH, NFT, SL, TKW, and HI suggests that the QTLs for these traits are more stable, meaning they perform reliably across different conditions.

This study utilised a p value < 0.001 as a stringent criterion. Nevertheless, several other GWAS studies related to agronomic traits in wheats also used the same criterion for identifying significant MTAs to avoid the risk of false negatives [6, 40, 60–62].

In wheat, the majority of genome-wide association studies focusing on agronomic and grain yield traits utilise SLM for analysis [63]. Nonetheless, there is a possibility that SLM may not identify all MTAs [17]. To address this limitation, the sophisticated multilocus GWAS method known as "FarmCPU" (fixed and random model circulating probability unification) has been established [17]. In this study, we utilised FarmCPU due to its superior statistical power and enhanced computational efficiency compared to other existing methods for GWAS analysis, including EMMA, EMMAX, and

GEMMA. FarmCPU addresses confounding issues that emerge from population structure, kinship, and the need for multiple testing correction [17, 42, 48, 64–66]. The analysis employs a fixed-effect model (FEM) alongside a test marker that incorporates pseudo QTNs as covariates, as well as a random effect model (REM). The tests iteratively estimate pseudo QTNs [17]. In the FarmCPU model, PCA is treated as a fixed effect, while kinship is considered a random effect [17, 21]. Model overfitting is mitigated through the estimation of kinship [66]. Furthermore, Q-Q plots validated the appropriateness of the multi-locus association model employed in this investigation.

False positives are a common limitation of GWAS, which require strict statistical thresholds because of the large number of comparisons. Genetic variations within populations can result in erroneous associations and skewed effect size estimates, which is a major concern with regard to population structure [10, 67, 68]. Although techniques MLMs [11] and principal PCA [69] are frequently employed to account for these biases, they might not always fully eliminate the problem, particularly when dealing with intricate population histories [70]. Inaccurate genotype calls can also increase the false-positive rate, especially for rare variants [71]. These elements emphasise the necessity of rigorous analytical methods like false discovery rate (FDR) [72] and cautious interpretation in GWAS.

In this study, we found 214 QTLs, which include 300 SNPs linked to traits that affect crop yield, and these were consistently identified in different environments. These QTLs were distributed across all chromosomes except for chromosome 3D. However, it's noteworthy that QTLs affecting agronomic traits have been reported on chromosome 3D in previous studies [40, 73]. The reason may be that chromosome 3D contains a lower marker density, and there was no separation for trait differences related to that chromosome, or the influence of a locus was too small to get detected [74]. We discuss in detail the stable loci identified across both environments, and combined data.A total of 11 QTLs were identified for PH, with five of them (*QPh.eu-1D*, *QPh.eu-2 A*, *QPh.eu-2D*, *QPh.eu-3B.1*, and *QPh.eu-6B*) having been previously documented in the literature. The QTL *QPh.eu-1D* discovered in our investigation was located at 12.5 Mb, which closely corresponds to a MTA (AX-94418622) reported at 16.4 Mb in a previous genome-wide association analysis conducted by Amalova et al. [75].

In the current study, the position of a QTL (*QPh.eu-2 A*) at 716.4 Mb was very close to the MTA *QSl.sicau-1 A* (715.3–721.6 Mb) [3]. Another QTL (qPH2A.2), which was found at the location of 720.3–720.5 Mb by the Pang et al. [45]. Alemu et al. [44]found a MTA (PH_KukRi_c205_223) at a position of 77 cM that was similar to the *QPh.eu-2D* region, which spans a genetic distance of 77–82 cM. Additionally, Zhang et al. [76] reported a previously found QTL (*QPH.sicau-3BL.1*) at location 787.8 Mb, nearby the *QPh.eu-3B.1* (278.3 Mb). The QTL *QPh.eu-6B*, which we identified at 113 cM and 611.6 Mb in our analysis, appears to be a potentially important genetic location. Other studies have found *qPH6B.3* at 620.9-621.3 [45], and MTA Kukri_rep_c106092_300 at 113.7 cM [77, 78], which are near our QTL. The results show the importance of *QPh.eu-6B* in the genetic control of plant height.

Among the 28 QTLs identified for NFT, four had been previously reported near the same location. The *QNft.eu-1B* was found at 476.2 Mb, close to another MTA (GENE-0035_150) located at 465.5 Mb [47]. The QTL *Qnft.eu-2 A* is found on the chromosome 2 A between 753.6 and 759.8 Mb, just nearby the location of MTA (BobWhite_c11479_157; 764.0 Mb) found in the other study [24]. A QTL (*Qnft.eu-2D*) was located on chromosome 2D at 18.8 Mb, right next to another MTA (D_F1BEJMU01DOWJ3_176; 13.8 Mb) found in an earlier study [47]. Another QTL (*QNft.eu-3 A*; 32.9 Mb) showed similarity with an MTA (GENE-0826_51) present at 32.1 Mb [47].

SL identified a total of 20 QTLs, with eight of them (*QSl.eu-1 A*, *QSl.eu-1 A.1*, *QSl.eu-2 A*, *QSl.eu-2B*, *QSl.eu-2B.1*, *QSl.eu-2B.4*, *QSl.eu-5B*, and *QSl.eu-5B.1*) found close to where some genetic markers were already known, while the other 12 QTLs might be new discoveries. *QSl.eu-1 A* was present at 530.2 Mb and was located nearby the location of QTL (*qSL1A.1*) at 531.6-532.1 Mb [45]. *QSl.eu-1 A.1* has been found at the position of 516.9 Mb, just beside the position of MTA (AX-94725964) found in the previous study at 517.5 Mb [47]. Also, a QTL (*QSl.eu-2 A*) spans the area from 716.5 to 722.8 Mb and is close to another MTA (AX-94625204) that was found earlier at 722.3 Mb [75]. Moreover, *QSl.eu-2B* (91 to 93 cM) and *QSl.eu-2B* (93 to 96 cM) overlap with MTAs (Ra_c27275_627 and wsnp_Ex_c62844_62315607) that were previously identified at 89 cM and 94 cM respectively [79]. Another QTL (*Qsl.eu-2B.4*) from the present study has also been identified at 99 cM on chromosome 2B. This concurrence was also found with the [RFL_Contig4718_1269] identified by Ma et al. [38] at 103 cM. The *QSl.eu-5B* locus, located at 61 cM, was close to a previously identified MTA (wsnp_Ex_rep_c69631_68583363) at 56.6 cM [38]. In addition, there was an overlap between *QSl.eu-5B.1* at 107 cM with previously identified QTLs at 87 cm and 107 cM in two different studies [52, 80].

Our study has shown a total of 20 QTLs linked to SPS, indicating the complex genetic nature of this trait. For SPS, the present investigation identified the three QTLs (*Qsps.eu-4 A.2*, *Qsps.eu-4B*, and *Qsps.eu-7D*), and their positions have been previously reported. *Qsps.eu-4 A.2*

was found in this study at 629.9 to 631.9 Mb which is very close to the location of a previously found MTA (IAAV4112) at 620 Mb [56]. The QTL *Qsps.eu-4B* is located at 606.6 Mb, near the position of the previously identified MTA (tplb0026o15_1634) at 600.0 Mb [24]. Another QTL (*QSps.eu-7D*) is located at 16.2 Mb, and near the location of previously identified QTL at 21.0 Mb [73].

A total of 11 QTLs were associated with KPS. Out of these, only two QTLs—*Qps.eu-3 A*, and *QKps.eu-5 A*, were found near other QTLs that had been identified before. We have found a QTL (*QKps.eu-3 A*) at the position of 557.4 cM, and it was very close to the position of another QTL (*QKNS.sicau-3AL.1b*) at a position of 551.7 Mb [76]. The QTL *Qkps.eu-5 A* was located at the position of 564.2 Mb in our study, corresponding to a physical location of 564.1 MB. These results share the locations of other QTL (*qGN5A.2*) found by other workers within the interval of 555.5 to 556.6 Mb [45].

We have also identified 12 QTLs for TKW. The locations of five QTLs (*Qtkw.eu-1 A*, *Qtkw.eu-1B*, *Qtkw.eu-2B*, *Qtkw.eu-2D.1*, and *Qtkw.eu-7B*) were identified by other researchers in previous studies. The Qtkw.eu-1 A found in the present study is located at the nearby genetic position (70 cM) of the QTL found in the previous study [44]. The QTL *Qtkw.eu-1B* was located at 673.7 Mb, which was nearby the location of the QTL (*qTGW1B.1*) identified by Pang et al. [45] within the interval of 667.9 to 668.1 Mb. Another QTL (*Qtkw.eu-2B*) was located between the intervals of 775.8 and 777.1 Mb, which is where the QTLs identified by some other workers were at 778.1 [24]. Safadar et al. [24] found an MTA (RAC875_c10626_2089) at 778.6 Mb on chromosome 2B. The present investigation found a QTL (*Qtkw.eu-2B*) that was located in the interval of 775.8 to 777.1 Mb, which coincided with this particular region. The present investigation found *Qtkw.eu-2D* within the interval of 610.2-613.3 Mb, which coincided with the MTA (Ex_c10068_1509; 619 Mb) found in the earlier study [56]. *Qtkw.eu-7B* was located in the interval of 134.5 to 136.6 Mb, whereas Amalova et al. [75] found a comparable MTA (AX-94761577) nearby at 133.4 Mb. These findings provide important information about how TKW is controlled by genetics, showing that they match and align with earlier studies on similar QTLs.

We also found 14 QTLs for HI, of which three were previously reported by other workers. The QTL Qhi.eu-1 A is found at 580.4 Mb, and a previous study also identified an MTA (BobWhite_c12960_138) for the same trait in wheat at 584.7 Mb. Another QTL (*Qhi.eu-2B*) is located at 155.0 Mb on chromosome 2B, close to a previously found one (wsnp_Ex_c30_66389) at 157.7 Mb on the same chromosome [24]. Another QTL (*Qhi.eu-2D*) from this study was found between 608.7 and 612.5 Mb, which is right next to the MTA (RAC875_c5673_1209; 605.8 Mb) discovered in the earlier study. The remaining QTLs identified in this study are potentially novel, and to validate the novelty, fine-mapping, cloning, and transgenic studies are also required.

The dataset al.so shows that some SNPs are found in multiple traits, which could mean that one gene affects many traits (pleiotropy) or that nearby genes are passed down together (LD) [81]. Understanding the genetic underpinnings of trait associations requires being able to distinguish between these two mechanisms. The ^{2} statistic can be used to measure the strength of LD; higher values indicate a stronger association between SNPs [81]. For example, depending on allele frequency matching and genetic distance, ^{2} values between SNPs can vary from low to very high, usually decaying at distances greater than 200 kb [81]. The dataset included a number of shared SNPs, such as Kukri_c36783_91, linked to HI and SL in various contexts. These shared genetic markers help us better understand how traits are related in the population we are studying, either by directly affecting different traits or being near other important genetic variants due to linkage LD. Each shared SNP requires further investigation to distinguish between pleiotropy and LD.

In the present study, we also examined the existence of significant genes affecting essential agronomic traits, including *Rht* genes for reduced height, *Ppd* genes for photoperiod sensitivity, *Vrn* genes for vernalisation requirements, and yield components such as *TaGW2* and *TaGS5*, within the panel. However, when we looked at our 90 K SNP dataset, we found that these important genes were either not included in the available SNPs or showed no genetic differences in this specific group. The lack of representation or polymorphism is likely due to intrinsic limitations in the SNP array design or possibly to little genetic variation at these specific loci within the germplasm of the WAMI panel.

Conclusion

The goal of this study was to identify grain yield increasing loci in wheat. For this purpose, we used the GWAS to analyse seven agronomic traits, utilising a diverse panel of 286 genotypes from International Maize and Wheat Improvement Center (CIMMYT), Mexico. The study indicated that the WAMI panel had an even distribution of allele frequencies, and not much genetic stratification. Finding 300 SNPs at 114 loci, the study mapped genetic loci that affect the agronomic, and yield-related traits. QTLs that demonstrated consistent effects across environments could serve as a suitable target for future fine-mapping, and MAS. The consistency of the identified loci/markers with previously published findings confirms the strength, and reliability of the results in this

investigation. However, it is recognised that additional trials are needed to validate the identified QTLs.

Methods

Plant materials

The Wheat Association Mapping Initiative (WAMI) panel, comprised of 286 genetically diverse, and advanced wheat lines, was used for the GWAS [82]. The seed of the WAMI panel was obtained from the CIMMYT via the International Wheat Improvement Network (IWIN). This panel contains 228 advanced or improved cultivars, 55 breeding lines, and three genetic stocks. All the genotypes were collected by CIMMYT from Mexico (267), India (6), and Nepal (4); two each from Afghanistan, Pakistan, and Bolivia; and one each from Chile, Brazil, and Algeria (https://doi.org/10.18730/BH9PT).

Field trials and trait evaluation

The panel was planted at the Agriculture Research Farm (30.75° N, 77.30° E), Eternal University, Baru Sahib, Himachal Pradesh (India), during the 2021–2022, and 2022–2023 wheat-growing seasons. Individual accessions were placed 5 cm apart in a single 2-metre-long row with a 25 cm row-to-row distance. The experiments were replicated twice using a randomised block design, and each replicate was randomly assigned using the Fisher and Yates technique [83]. The temperature during the sowing of experimental material was ~ 10 °C and during harvesting was ~ 35 °C in both years. The texture of soil was sandy loam with pH 7. The standard rates of the fertilisers were 40 kg K_2O, 60 kg P_2O_5, and 120 kg N. Except for N, which was applied in three doses—50% at sowing, 25% during the first irrigation (21 days after sowing), and the remaining 25% during the second irrigation (45 days after sowing), all fertilisers were applied at sowing time. Normal irrigation was also applied during all the stages of experimental material using the sprinkler method. All seven agronomic traits were measured when the plants were fully mature in both environments, and then the average values were calculated using the data from both places. The PH of ten identical plants, measured from the soil's surface to the tips of their spikes (awns excluded), was averaged. Ten randomly chosen plants were used to determine NFT. SPS was calculated as the average of ten separate plants [84]. The measures of the ten primary tiller spikes were SL, KPS, and TKW. We calculated the HI for each genotype separately by dividing the grain yield by the aboveground biomass, and multiplying the result by 100 after the harvesting. The details of phenotypic data is given in Table S9.

Phenotypic data analysis

The best linear unbiased prediction (BLUP) values were derived using the R's lme4 package [85], from the data of the 2021–2022 (Env1), and 2022–2023 (Env2) wheat-growing seasons along with the combined data of both years (CE). Using the restricted maximum likelihood (REML) method, which is integrated into META-R software v6.0.4, variance components over environments were estimated by considering genotypes, environments, and replication as random effects [86]. The Agricolae (version 1.2–4) package was used to analyse the descriptive statistics. We used the web programme SRPLOT (http://www.bioinformatics.com.cn/srplot; accessed: December 2, 2023) [87] to draw correlation plots. The raincloud diagram was developed with the same program. The phenotypic data of seven agronomic traits was used in this work to calculate PCA based on covariance in the context of an eigenvalue decomposition (EVD) [88, 89], and a PCA graph was developed using the web tool Statistics Kingdom (https://www.statskingdom.com/pca-calculator.html).

SNP genotyping

A dataset consisting of 20,996 high-quality SNPs was utilised for GWAS. We obtained this dataset from the CIMMYT Mexico website (accessible on March 17, 2024; https://data.cimmyt.org/dataset.xhtml?persistentId=hdl:11529/10714). Genotyping of plant material, and data processing were performed as described previously [90]. In simple terms, genotyping was conducted with the Illumina 90 K Infinium iSelect assay at the USDA-ARS Small Grain Genotyping Centre in Fargo. A total of 26,814 bi-allelic SNPs were found using the standard grouping method built into Genome Studio v2011.1. To maintain data quality criteria, Monomorphic SNPs, low-quality, or had a minor allele frequency (MAF) less than 0.05 were eliminated. Following a thorough screening process, 20,996 polymorphic SNPs were selected, and utilised to perform the GWAS analysis.

Population structure, kinship matrix and principal components analyses

PCA was conducted on 20,996 high-quality SNP markers using R software (version 4.0.3) [R Core Team, 2013]. The Bayesian information criterion (BIC) was employed to identify the optimal number of PCs for further analysis [19, 91]. A scatter plot depicting the first three PCs was generated to illustrate the distribution of genotypes across the population. Additionally, a kinship matrix was calculated using R software (version 4.0.3) as done previously [92, 93]. Hierarchical cluster analysis was conducted using dissimilarity values, and the "ward.D2" method within the "hclust," as previously performed [94].

Genome-wide association analyses

Using a sliding window size of 100, the Trait Analysis by Association, Evolution, and Linkage (TASSEL) software

version 5.2.91 [95] was used to find the r^2 between SNP markers. R^2 values were then plotted against physical distance in base pair (bp) to evaluate the linkage disequilibrium (LD) between locations. The LD decay curve was created for the entire genome using a smoothing spline regression line, based on the method explained by Hill and Weir [96], and carried out in the R environment with a script that Marroni et al. [97] had used before. To identify connections between markers, and traits (MTAs), a method called FarmCPU was used [17]. To increase association mapping accuracy, FarmCPU incorporates both random, and fixed effects. In a two-step process, the population structure is first controlled using the fixed effect model, and relatedness between individuals is further accounted for using the random effect model. Statistically significant MTAs were detected using $p < 0.001$ ($-\log10(P) > 3.0$) [98].

Abbreviations

ANOVA	Analysis of variance
BIC	Bayesian information criterion
BLINK	Bayesian information and linkage-disequilibrium iteratively nested keyway
CE	Combined environment
CGs	Candidate genes
cM	centiMorgan
CMLM	Compressed mixed linear model
CV	Coefficient of variation
ECMLM	Enriched compressed mixed linear model
ENV1	Environment 1
ENV2	Environment 2
EVD	Eigenvalue decomposition
FarmCPU	Fixed and random model circulating probability unification
GLM	General linear model
GWAS	Genome-wide association study
H_2	Heritability
HI	Harvest index
IWGSC	International wheat genome sequencing consortium
IWIN	International Wheat Improvement Network
KPS	Kernels per spike
LD	Linkage disequilibrium
MAF	Minor allele frequency
MAS	Marker-assisted selection
MLM	Mixed linear model
MLMM	Multilocus mixed model
MTAs	Marker-trait associations
NFT	The number of fertile tillers
PC1	Principle component 1
PC2	Principle component 2:Principle component 3
PCA	Principal component analysis
PH	Plant height
PVE	Phenotypic variation explained
QTLs	Quantitative trait loci
QTNs	Quantitative trait nucleotide
r^2	Squared allele-frequency correlations
REML	Restricted maximum likelihood
SL	Single-locus models: ML: Multilocus model
SL	Spike length
SNPs	Single nucleotide polymorphisms
SPS	Spikelet number per spike
SUPER	Settlement of MLM under progressively exclusive relationship
TKW	Thousand-kernel weight
WAMI	Wheat Association Mapping Initiative

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12870-025-07263-6.

Supplementary Material 1

Supplementary Material 2

Supplementary Material 3

Supplementary Material 4

Supplementary Material 5

Supplementary Material 6

Acknowledgements

We express our appreciation to Science and Engineering Research Board (SERB), India for supporting NKV with a Start-Up Research Grant (SRG/2020/000091). We also sincerely acknowledge the International Maize and Wheat Improvement Center (CIMMYT), Mexico, for generously providing the WAMI population, and granting access to the molecular data used in this study. Additionally, we extend our gratitude to the Department of Genetics, Plant Breeding, and Biotechnology, Dr. K. S. Gill Akal College of Agriculture, Eternal University, Baru Sahib, HP, India, for their valuable support, and provision of facilities to NKV, and AT throughout the course of this research.

Author contributions

Conceptualization, NKV; methodology, AT, SK and NKV; formal analysis, AT; investigation, AT; resources, NKV, VKM, and AKJ; data curation, AT, RD, and N.K.V.; writing—original draft preparation, AT and NKV; writing—review and editing, NKV, RD, SS, MKS, and AKJ; visualization, RD and NKV; supervision, NKV; project administration, NKV; funding acquisition, NKV. All authors have read and agreed to the published version of the manuscript.

Funding

This research was funded by the Science and Engineering Research Board (SERB), New Delhi, grant number SRG/2020/000091.

Data availability

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Ethics approval and informed consent

All the experiments on plants were carried out in accordance with guidelines of Eternal University.

Author details

[1]Department of Genetics-Plant Breeding and Biotechnology, Dr. Khem Sigh Gill Akal College of Agriculture, Eternal University, Baru Sahib, HP, India
[2]Lethbridge Research and Development Centre, Agriculture and Agri-Food Canada, Lethbridge, AB, Canada
[3]Borlaug Institute for South Asia (BISA), NASC Complex, DPS Marg, New Delhi, India
[4]CIMMYT Regional Office, NASC complex, DPS Marg, New Delhi, India
[5]Department of Genetics and Plant Breeding, Institute of Agricultural Sciences, Banaras Hindu University, Varanasi, India
[6]Department of Genetics and Plant Breeding, Tilka Manjhi Agriculture College, Birsa Agricultural University, Ranchi, India
[7]Indian Council of Agricultural Research (ICAR) - National Bureau of Plant Genetic Resources, New Delhi, India

[8]Department of Genetics and Plant Breeding, Rajiv Gandhi University, Rono Hills, Itanagar, India

Received: 19 April 2025 / Accepted: 19 August 2025
Published online: 03 November 2025

References

1. Facts IJCIC. International Development Research Centre. Figures on Food and Biodiversity. 2010.
2. Shiferaw B, Smale M, Braun H-J, Duveiller E, Reynolds M, Muricho G. Crops that feed the world 10. Past successes and future challenges to the role played by wheat in global food security. Food Secur. 2013;5:291–317. https://doi.org/10.1007/s12571-013-0263-y.
3. Li F, Wen W, Liu J, Zhang Y, Cao S, He Z, Rasheed A, Jin H, Zhang C, Yan J, Zhang P, Wan Y, Xia X. Genetic architecture of grain yield in bread wheat based on genome-wide association studies. BMC Plant Biol.19:168. https://doi.org/10.1186/s12870-019-1781-3
4. Neumann K, Kobiljski B, Denčić S, Varshney R, Börner A. Genome-wide association mapping: a case study in bread wheat (*Triticum aestivum* L). Mol Breed. 2011;27:37–58. https://doi.org/10.1007/s11032-010-9411-7.
5. Saïdou A-A, Thuillet A-C, Couderc M, Mariac C, Vigouroux Y. Association studies including genotype by environment interactions: prospects and limits. BMC Genet. 2014;15:3. https://doi.org/10.1186/1471-2156-15-3.
6. Rahimi Y, Bihamta MR, Taleei A, Alipour H, Ingvarsson PK. Genome-wide association study of agronomic traits in bread wheat reveals novel putative alleles for future breeding programs. BMC Plant Biol. 2019;19:541. https://doi.org/10.1186/s12870-019-2165-4.
7. Wang S-B, Feng J-Y, Ren W-L, Huang B, Zhou L, Wen Y-J, Zhang J, Dunwell JM, Xu S, Zhang Y-M. Improving power and accuracy of genome-wide association studies via a multi-locus mixed linear model methodology. Sci Rep. 2016;6(1):19444. https://doi.org/10.1038/srep19444.
8. Kaler AS, Gillman JD, Beissinger T, Purcell LC. Comparing different statistical models and multiple testing corrections for association mapping in soybean and maize. Front Plant Sci. 2020;10:1794. https://doi.org/10.3389/fpls.2019.01794.
9. Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, Kresovich S, Buckler ES. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet. 2006;38(2):203–8. https://doi.org/10.1038/ng1702.
10. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 2006;38(8):904–9. https://doi.org/10.1038/ng1847.
11. Zhang Z, Ersoz E, Lai C-Q, Todhunter RJ, Tiwari HK, Gore MA, Bradbury PJ, Yu J, Arnett DK, Ordovas JM, Buckler ES. Mixed linear model approach adapted for genome-wide association studies. Nat Genet. 2010;42(4):355–60. https://doi.org/10.1038/ng.546.
12. Li M, Liu X, Bradbury P, Yu J, Zhang Y-M, Todhunter RJ, Buckler ES, Zhang Z. Enrichment of statistical power for genome-wide association studies. BMC Biol. 2014;12:73. https://doi.org/10.1186/s12915-014-0073-5.
13. Wang Q, Tian F, Pan Y, Buckler ES, Zhang Z. A super powerful method for genome wide association study. PLoS One. 2014;9(9):e107684. https://doi.org/10.1371/journal.pone.0107684.
14. Rakitsch B, Lippert C, Stegle O, Borgwardt K. A lasso multi-marker mixed model for association mapping with population structure correction. Bioinformatics. 2013;29(2):206–14. https://doi.org/10.1093/bioinformatics/bts669.
15. Segura V, Vilhjálmsson BJ, Platt A, Korte A, Seren Ü, Long Q, Nordborg M. An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. Nat Genet. 2012;44(7):825–30. https://doi.org/10.1038/ng.2314.
16. Huang M, Liu X, Zhou Y, Summers RM, Zhang Z. BLINK: a package for the next level of genome-wide association studies with both individuals and markers in the millions. Gigascience. 2019;8(2):giy154. https://doi.org/10.1093/gigascience/giy154.
17. Liu X, Huang M, Fan B, Buckler ES, Zhang Z. Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. PLoS Genet. 2016;12(2):e1005767. https://doi.org/10.1371/journal.pgen.1005767.
18. Lipka AE, Gore MA, Magallanes-Lundback M, Mesberg A, Lin H, Tiede T, Chen C, Buell CR, Buckler ES, Rocheford T, DellaPenna D. Genome-wide association study and pathway-level analysis of tocochromanol levels in maize grain. G3 Genes|Genomes|Genetics. 2013;3(8):1287–99. https://doi.org/10.1534/g3.113.006148.
19. Sauvage C, Segura V, Bauchet G, Stevens R, Do PT, Nikoloski Z, Fernie AR, Causse M. Genome-wide association in tomato reveals 44 candidate loci for fruit metabolic traits. Plant Physiol. 2014;165(3):1120–32. https://doi.org/10.1104/pp.114.241521.
20. Vaughn JN, Nelson RL, Song Q, Cregan PB, Li Z. The genetic architecture of seed composition in soybean is refined by genome-wide association scans across multiple populations. G3 Genes|Genomes|Genetics. 2014;4(11):2283–94. https://doi.org/10.1534/g3.114.013433.
21. Zhang Y-M, Jia Z, Dunwell JM. The applications of new multi-locus GWAS methodologies in the genetic dissection of complex traits. Front Plant Sci. 2019;10:100. https://doi.org/10.3389/fpls.2019.00100.
22. Danakumara T, Kumari J, Singh AK, Sinha SK, Pradhan AK, Sharma S, Jha SK, Bansal R, Kumar S, Jha GK, Yadav MC. Genetic dissection of seedling root system architectural traits in a diverse panel of hexaploid wheat through multi-locus genome-wide association mapping for improving drought tolerance. Int J Mol Sci. 2021;22(13): 7188. https://doi.org/10.3390/ijms22137188.
23. Yadav R, Gupta S, Gaikwad KB, Bainsla NK, Kumar M, Babu P, Ansari R, Dhar N, Dharmateja P, Prasad R. Genetic gain in yield and associated changes in agronomic traits in wheat cultivars developed between 1900 and 2016 for irrigated ecosystems of Northwestern plain zone of India. Front Plant Sci. 2021;12:719394. https://doi.org/10.3389/fpls.2021.719394.
24. Safdar LB, Andleeb T, Latif S, Umer MJ, Tang M, Li X, Liu S, Quraishi UM. Genome-wide association study and QTL meta-analysis identified novel genomic loci controlling potassium use efficiency and agronomic traits in bread wheat. Front Plant Sci. 2020;11:70. https://doi.org/10.3389/fpls.2020.00070.
25. Geneti GS, Kebede SA, Mekonnen TB. Genetic variability and association of traits in bread wheat (*Triticum aestivum* L.) genotypes in Gechi district, South West Ethiopia. Adv Agric. 2022;2022(1):7132424. https://doi.org/10.1155/2022/7132424.
26. Li Z, Luo Q, Deng Y, Du K, Li X, Ren T. Identification and validation of a stable major-effect quantitative trait locus for kernel number per spike on chromosome 2D in wheat (*Triticum aestivum* L). Int J Mol Sci. 2023;24(18): 14289. https://doi.org/10.3390/ijms241814289.
27. Liu H, Shi Z, Ma F, Xu Y, Han G, Zhang J, Liu D, An D. Identification and validation of plant height, spike length and spike compactness loci in common wheat (*Triticum aestivum* L.). BMC Plant Biol. 2022;22(1): 568. https://doi.org/10.1186/s12870-022-03968-0.
28. Fernandes RC, Busanello C, Viana VE, Venske E, de Oliveira VF, Lopes JL, da Maia LC, Costa de Oliveira A, Pegoraro C. Genetic variability and heritability of agronomic traits in a wheat collection used in Southern Brazil. J Crop Sci Biotechnol. 2022;25:337–48. https://doi.org/10.1007/s12892-021-00135-z.
29. Gupta PK, Balyan HS, Gahlaut V. QTL analysis for drought tolerance in wheat: present status and future possibilities. Agronomy. 2017;7(1): 5. https://doi.org/10.3390/agronomy7010005.
30. Lephuthing MC, Khumalo TP, Tolmay VL, Dube E, Tsilo T. Genetic mapping of quantitative trait loci associated with plant height and yield component traits in a wheat (*Triticum aestivum* L.) doubled haploid population derived from Tugela-DN× Elands. Agronomy. 2022;12(10): 2283. https://doi.org/10.3390/agronomy12102283.
31. Korte A, Farlow A. The advantages and limitations of trait analysis with GWAS: a review. Plant Methods. 2013;9:29. https://doi.org/10.1186/1746-4811-9-29.
32. Sukumaran S, Dreisigacker S, Lopes M, Chavez P, Reynolds MP. Genome-wide association study for grain yield and related traits in an elite spring wheat population grown in temperate irrigated environments. Theor Appl Genet. 2015;128:353–63. https://doi.org/10.1007/s00122-014-2435-3.
33. Sukumaran S, Reynolds MP, Sansaloni C. Genome-wide association analyses identify QTL hotspots for yield and component traits in durum wheat grown under yield potential, drought, and heat stress environments. Front Plant Sci. 2018;9:81. https://doi.org/10.3389/fpls.2018.00081.
34. Sun C, Zhang F, Yan X, Zhang X, Dong Z, Cui D, Chen F. 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. 2017;15(8):953–69. https://doi.org/10.1111/pbi.12690.
35. Godoy J, Gizaw S, Chao S, Blake N, Carter A, Cuthbert R, Dubcovsky J, Hucl P, Kephart K, Pozniak C, Prasad PVV, Pumphrey M, Talbert L. Genome-wide association study of agronomic traits in a spring-planted North American elite hard red spring wheat panel. Crop Sci. 2018;58(5):1838–52. https://doi.org/10.2135/cropsci2017.07.0423.

36. Sheoran S, Jaiswal S, Kumar D, Raghav N, Sharma R, Pawar S, Paul S, Iquebal MA, Jaiswar A, Sharma P, Singh R, Singh CP, Gupta A, Kumar N, Angadi UB, Rai A, Singh GP, Kumar D, Tiwari R. Uncovering genomic regions associated with 36 agro-morphological traits in Indian spring wheat using GWAS. Front Plant Sci. 2019;10:527. https://doi.org/10.3389/fpls.2019.00527.
37. Liu J, Xu Z, Fan X, Zhou Q, Cao J, Wang F, Ji G, Yang L, Feng B, Wang T. A genome-wide association study of wheat spike related traits in China. Front Plant Sci. 2018;9:1584. https://doi.org/10.3389/fpls.2018.01584.
38. Ma F, Xu Y, Ma Z, Li L, An D. Genome-wide association and validation of key loci for yield-related traits in wheat founder parent Xiaoyan 6. Mol Breed. 2018;38:91. https://doi.org/10.1007/s11032-018-0837-7.
39. Bajgain P, Zhang X, Anderson JA. Genome-wide association study of yield component traits in intermediate wheatgrass and implications in genomic selection and breeding. G3 (Bethasda). 2019;9(8):2429–39. https://doi.org/10.1534/g3.119.40007
40. Jamil M, Ali A, Gul A, Ghafoor A, Napar AA, Ibrahim AM, Naveed NH, Yasin NA, Mujeeb-Kazi A. Genome-wide association studies of seven agronomic traits under two sowing conditions in bread wheat. BMC Plant Biol. 2019;19:149–218. https://doi.org/10.1186/s12870-019-1754-6.
41. Ward BP, Brown-Guedira G, Kolb FL, Van Sanford DA, Tyagi P, Sneller CH, Griffey CA. Genome-wide association studies for yield-related traits in soft red winter wheat grown in Virginia. PLoS One. 2019;14(2):e0208217. https://doi.org/10.1371/journal.pone.0208217.
42. Ali M, Zhang Y, Rasheed A, Wang J, Zhang L. Genomic prediction for grain yield and yield-related traits in Chinese winter wheat. Int J Mol Sci. 2020;21(4):1342. https://doi.org/10.3390/ijms21041342.
43. Alqudah AM, Haile JK, Alomari DZ, Pozniak CJ, Kobiljski B, Börner A. Genome-wide and SNP network analyses reveal genetic control of spikelet sterility and yield-related traits in wheat. Sci Rep. 2020;10(1):2098. https://doi.org/10.1038/s41598-020-59004-4.
44. Alemu A, Suliman S, Hagras A, Thabet S, Al-Abdallat A, Abdelmula AA, Tadesse W. Multi-model genome-wide association and genomic prediction analysis of 16 agronomic, physiological and quality related traits in ICARDA spring wheat. Euphytica. 2021;217:1–22. https://doi.org/10.1007/s10681-021-02933-6.
45. Pang Y, Liu C, Wang D, St Amand P, Bernardo A, Li W, He F, Li L, Wang L, Yuan X, Dong L, Su Y, Zhang H, Zhao M, Liang Y, Jia H, Shen X, Lu Y, Jiang H, Wu Y, Li A, Wang H, Kong L, Bai G, Liu S. High-resolution genome-wide association study identifies genomic regions and candidate genes for important agronomic traits in wheat. Mol Plant. 2020;13(9):1311–27. https://doi.org/10.1016/j.molp.2020.07.008.
46. Eltaher S, Baenziger PS, Belamkar V, Emara HA, Nower AA, Salem KF, Alqudah AM, Sallam A. GWAS revealed effect of genotype × environment interactions for grain yield of Nebraska winter wheat. BMC Genomics. 2021;22: 2. https://doi.org/10.1186/s12864-020-07308-0.
47. Gao L, Meng C, Yi T, Xu K, Cao H, Zhang S, Yang X, Zhao Y. Genome-wide association study reveals the genetic basis of yield-and quality-related traits in wheat. BMC Plant Biol. 2021;21:144. https://doi.org/10.1186/s12870-021-02925-7.
48. Malik P, Kumar J, Singh S, Sharma S, Meher PK, Sharma MK, Roy JK, Sharma PK, Balyan HS, Gupta PK, Sharma S. Single-trait, multi-locus and multi-trait GWAS using four different models for yield traits in bread wheat. Mol Breed. 2021;41:46–21. https://doi.org/10.1007/s11032-021-01240-1.
49. Saini DK, Chopra Y, Singh J, Sandhu KS, Kumar A, Bazzer S, Srivastava P. Comprehensive evaluation of mapping complex traits in wheat using genome-wide association studies. Mol Breed. 2022;42: 1. https://doi.org/10.1007/s11032-021-01272-7.
50. Zhang J, Yao Q, Li R, Lu Y, Zhou S, Han H, Liu W, Li X, Yang X, Li L. Identification of genetic loci on chromosome 4B for improving the grain number per spike in pre-breeding lines of wheat. Agronomy. 2022;12(1):171. https://doi.org/10.3390/agronomy12010171.
51. Vishwakarma MK, Bhati PK, Kumar U, Singh RP, Kumar S, Govindan V, Mavi GS, Thiyagarajan K, Dhar N, Joshi AK. Genetic dissection of value-added quality traits and agronomic parameters through genome-wide association mapping in bread wheat (*T. aestivum* L). Front Plant Sci. 2024;15:1419227. https://doi.org/10.3389/fpls.2024.1419227.
52. Hu P, Zheng Q, Luo Q, Teng W, Li H, Li B, Li Z. Genome-wide association study of yield and related traits in common wheat under salt-stress conditions. BMC Plant Biol. 2021;21:27. https://doi.org/10.3389/fpls.2022.1098560.
53. Alemu A, Brazauskas G, Gaikpa DS, Henriksson T, Islamov B, Jørgensen LN, Koppel M, Koppel R, Liatukas Ž, Svensson J, Chawade A. Genome-wide association analysis and genomic prediction for adult-plant resistance to septoria tritici blotch and powdery mildew in winter wheat. Front Genet. 2021;12:661742. https://doi.org/10.3389/fgene.2021.661742.
54. Said AA, MacQueen AH, Shawky H, Reynolds M, Juenger TE, El-Soda M. Genome-wide association mapping of genotype-environment interactions affecting yield-related traits of spring wheat grown in three watering regimes. Environ Exp Bot. 2022;194: 104740. https://doi.org/10.1016/j.envexpbot.2021.104740.
55. Tian Y, Liu P, Cui F, Xu H, Han X, Nie Y, Kong D, Sang W, Li W. Genome-wide association study for yield and yield-related traits in Chinese spring wheat. Agronomy. 2023;13(11):2784. https://doi.org/10.3390/agronomy13112784.
56. Guo J, Guo J, Li L, Bai X, Huo X, Shi W, Gao L, Dai K, Jing R, Hao C. Combined linkage analysis and association mapping identifies genomic regions associated with yield-related and drought-tolerance traits in wheat (*Triticum aestivum* L). Theor Appl Genet. 2023;136(12):250. https://doi.org/10.1007/s00122-023-04494-9.
57. Dawson JC, Huggins DR, Jones SS. Characterizing nitrogen use efficiency in natural and agricultural ecosystems to improve the performance of cereal crops in low-input and organic agricultural systems. Field Crops Res. 2008;107(2):89–101. https://doi.org/10.1016/j.fcr.2008.01.001.
58. Lozada DN, Mason RE, Babar MA, Carver BF, Guedira G-B, Merrill K, Arguello MN, Acuna A, Vieira L, Holder A, Addison C, Moon DE, Miller RD, Dreisigacker S. Association mapping reveals loci associated with multiple traits that affect grain yield and adaptation in soft winter wheat. Euphytica. 2017;213:222. https://doi.org/10.1007/s10681-017-2005-2.
59. Nigro D, Gadaleta A, Mangini G, Colasuonno P, Marcotuli I, Giancaspro A, Giove S, Simeone R, Blanco A. Candidate genes and genome-wide association study of grain protein content and protein deviation in durum wheat. Planta. 2019;249:1157–75. https://doi.org/10.1007/s00425-018-03075-1.
60. Dodig D, Zoric M, Kobiljski B, Savic J, Kandic V, Quarrie S, Barnes J. Genetic and association mapping study of wheat agronomic traits under contrasting water regimes. Int J Mol Sci. 2012;13(5):6167–88. https://doi.org/10.3390/ijms13056167.
61. Jung WJ, Lee YJ, Kang CS, Seo YW. Identification of genetic loci associated with major agronomic traits of wheat (*Triticum aestivum* L.) based on genome-wide association analysis. BMC Plant Biol. 2021;21:418. https://doi.org/10.1186/s12870-021-03180-6.
62. Mwadzingeni L, Shimelis H, Rees DJG, Tsilo TJ. Genome-wide association analysis of agronomic traits in wheat under droughtstressed and non-stressed conditions. PLoS ONE. 2017;12(2):e0171692. https://doi.org/10.1371/journal.pone.0211730.
63. Garcia M, Eckermann P, Haefele S, Satija S, Sznajder B, Timmins A, Baumann U, Wolters P, Mather DE, Fleury D. Genome-wide association mapping of grain yield in a diverse collection of spring wheat (*Triticum aestivum* L.) evaluated in southern Australia. PLoS One. 2019;14(2):e0211730. https://doi.org/10.1371/journal.pone.0211730.
64. Kaur S, Zhang X, Mohan A, Dong H, Vikram P, Singh S, Zhang Z, Gill KS, Dhugga KS, Singh J. Genome-wide association study reveals novel genes associated with culm cellulose content in bread wheat (*Triticum aestivum*, L). Front Plant Sci. 2017;8:1913. https://doi.org/10.3389/fpls.2017.01913.
65. Gyawali A, Shrestha V, Guill KE, Flint-Garcia S, Beissinger TM. Single-plant GWAS coupled with bulk segregant analysis allows rapid identification and corroboration of plant-height candidate SNPs. BMC Plant Biol. 2019;19(1):412. https://doi.org/10.1186/s12870-019-2000-y.
66. Muhammad A, Hu W, Li Z, Li J, Xie G, Wang J, Wang L. Appraising the genetic architecture of kernel traits in hexaploid wheat using GWAS. Int J Mol Sci. 2020;21(16): 5649. https://doi.org/10.3390/ijms21165649.
67. Devlin B, Roeder K. Genomic control for association studies. Biometrics. 1999;55(4):997–1004. https://doi.org/10.1111/j.0006-341x.1999.00997.x.
68. Marchini J, Cardon LR, Phillips MS, Donnelly P. The effects of human population structure on large genetic association studies. Nat Genet. 2004;36:512–7. https://doi.org/10.1038/ng1337.
69. Grinde KE, Browning BL, Reiner AP, Thornton TA, Browning SR. Adjusting for principal components can induce collider bias in genome-wide association studies. PLoS Genet. 2024;20(12):e1011242. https://doi.org/10.1371/journal.pgen.1011242.
70. Zaidi AA, Mathieson I. Demographic history mediates the effect of stratification on polygenic scores. Elife. 2020;9:e61548. https://doi.org/10.7554/eLife.61548.
71. Walters RG, Coin LJ, Ruokonen A, de Smith AJ, El-Sayed Moustafa JS, Jacquemont S, Elliott P, Esko T, Hartikainen AL, Laitinen J, Männik K, Martinet D, Meyre D, Nauck M, Schurmann C, Sladek R, Thorleifsson G, Thorsteinsdóttir U, Valsesia A, Waeber G, Zufferey F, Balkau B, Pattou F, Metspalu A, Völzke H,

Vollenweider P, Stefansson K, Järvelin MR, Beckmann JS, Froguel P, Blakemore AI. Rare genomic structural variants in complex disease: lessons from the replication of associations with obesity. PLoS ONE. 2013;8(3):e58048. https://doi.org/10.1371/journal.pone.0058048.
72. Benjamini Y, Drai D, Elmer G, Kafkafi N, Golani I. Controlling the false discovery rate in behavior genetics research. Behav Brain Res. 2001;125(1–2):279–84. https://doi.org/10.1016/s0166-4328(01)00297-2.
73. Li Y, Tang J, Liu W, Yan W, Sun Y, Che J, Tian C, Zhang H, Yu L. The genetic architecture of grain yield in spring wheat based on genome-wide association study. Front Genet. 2021;12: 728472. https://doi.org/10.3389/fgene.2021.728472.
74. Hussain W, Baenziger PS, Belamkar V, Guttieri MJ, Venegas JP, Easterly A, Sallam A, Poland J. Genotyping-by-sequencing derived high-density linkage map and its application to QTL mapping of flag leaf traits in bread wheat. Sci Rep. 2017;(1): 16394. https://doi.org/10.1038/s41598-017-16006-z.
75. Amalova A, Yessimbekova M, Ortaev A, Rsaliyev S, Griffiths S, Burakhoja A, Turuspekov Y, Abugalieva S. Association mapping of quantitative trait loci for agronomic traits in a winter wheat collection grown in Kazakhstan. Agronomy. 2023;13(8):2054. https://doi.org/10.3390/agronomy13082054.
76. Zhang ZP, Zhen LI, Fang HE, Lue JJ, Bin XI, Jing LI, Jian MA, Zheng YL, Wei LI. Genome-wide association and linkage mapping strategies reveal the genetic loci and candidate genes of important agronomic traits in Sichuan wheat. J Integr Agric. 2023;22(11):3380–93. https://doi.org/10.1016/j.jia.2023.02.030.
77. Kumar N, Kulwal P, Balyan H, Gupta PK. QTL mapping for yield and yield contributing traits in two mapping populations of bread wheat. Mol Breed. 2007;19:163–77. https://doi.org/10.1007/s11032-006-9056-8.
78. Wang SX, Zhu YL, Zhang DX, Shao H, Liu P, Hu JB, Zhang H, Zhang HP, Chang C, Lu J, Xia XC, Sun GL, Ma CX. Genome-wide association study for grain yield and related traits in elite wheat varieties and advanced lines using SNP markers. PLoS One. 2017;12(11):e0188662. https://doi.org/10.1371/journal.pone.0188662.
79. Alotaibi FS, Al-Qthanin RN, Aljabri M, Shehzad T, Albaqami M, Abou-Elwafa SF. Identification of genomic regions associated with agronomical traits of bread wheat under two levels of salinity using GWAS. Plant Mol Biol Rep. 2022;40:595–609. https://doi.org/10.1007/s11105-022-01341-x.
80. Amalova A, Abugalieva S, Babkenov A, Babkenova S, Turuspekov Y. Genome-wide association study of yield components in spring wheat collection harvested under two water regimes in Northern Kazakhstan. PeerJ. 2021;9:e11857. https://doi.org/10.7717/peerj.11857.
81. Eberle MA, Rieder MJ, Kruglyak L, Nickerson DA. Allele frequency matching between SNPs reveals an excess of linkage disequilibrium in genic regions of the human genome. PLoS Genet. 2006;2(9):e142. https://doi.org/10.1371/journal.pgen.0020142.
82. Lopes MS, Dreisigacker S, Peña RJ, Sukumaran S, Reynolds MP. Genetic characterization of the wheat association mapping initiative (WAMI) panel for dissection of complex traits in spring wheat. Theor Appl Genet. 2015;128(3):453–64. https://doi.org/10.1007/s00122-014-2444-2.
83. Panse VG, Sukhatme PV. Statistical methods for agricultural workers. 1954.
84. Briggle LW, Reitz LP. Classification of *Triticum* species and of wheat varieties grown in the united States. US Department of Agriculture; 1963.
85. Bates D, Mächler M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. J Stat Softw. 2015;67:1–48. https://doi.org/10.18637/jss.v067.i01.
86. Alvarado G, Rodríguez FM, Pacheco A, Burgueño J, Crossa J, Vargas M, Pérez-Rodríguez P, Lopez-Cruz MA. META-R: a software to analyze data from multi-environment plant breeding trials. Crop J. 2020;8(5):745–56. https://doi.org/10.1016/j.cj.2020.03.010.
87. Tang D, Chen M, Huang X, Zhang G, Zeng L, Zhang G, Wu S, Wang Y. SRplot: a free online platform for data visualization and graphing. PLoS One. 2023;18(11):e0294236. https://doi.org/10.1371/journal.pone.0294236.
88. Flury B. Common principal components & related multivariate models. John Wiley & Sons, Inc.; 1988.
89. Jolliffe IT, Cadima J. Principal component analysis: a review and recent developments. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences. 2016;374(2065):20150202. https://doi.org/10.1098/rsta.2015.0202.
90. Sukumaran S, Crossa J, Jarquin D, Lopes M, Reynolds MP. Genomic prediction with pedigree and genotypex environment interaction in spring wheat grown in South and West Asia, North Africa, and Mexico. G3 Genes Genomes Genet. 2017;7(2):481–95. https://doi.org/10.1534/g3.116.036251.
91. Schwarz G. Estimating the dimension of a model. Ann Statist. 1978;6(2):461–4. https://doi.org/10.1214/aos/1176344136.
92. Van Raden PM. Efficient methods to compute genomic predictions. Ann Stat. 2008;91(11):4414–23. https://doi.org/10.3168/jds.2007-0980.
93. Yin L, Zhang H, Tang Z, Xu J, Yin D, Zhang Z, Yuan X, Zhu M, Zhao S, Li X, Liu X. rMVP: a memory-efficient, visualization-enhanced, and parallel-accelerated tool for genome-wide association study. Genomics Proteomics Bioinf. 2021;19(4):619–28. https://doi.org/10.1016/j.gpb.2020.10.007.
94. Murtagh F, Legendre P. Ward's hierarchical agglomerative clustering method: which algorithms implement ward's criterion? J Classif. 2014;31:274–95. https://doi.org/10.1007/s00357-014-9161-z.
95. Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics. 2007;23(19):2633–5. https://doi.org/10.1093/bioinformatics/btm308.
96. Hill W, Weir BS. Variances and covariances of squared linkage disequilibria in finite populations. Theor Popul Biol. 1988;33(1):54–78. https://doi.org/10.1016/0040-5809(88)90004-4.
97. Marroni F, Pinosio S, Zaina G, Fogolari F, Felice N, Cattonaro F, Morgante M. Nucleotide diversity and linkage disequilibrium in *Populus nigra* cinnamyl alcohol dehydrogenase (CAD4) gene. Tree Genet Genomes. 2011;7:1011–23. https://doi.org/10.1007/s11295-011-0391-5.
98. Kaur R, Vasistha NK, Ravat VK, Mishra VK, Sharma S, Joshi AK, Dhariwal R. Genome-wide association study reveals novel powdery mildew resistance loci in bread wheat. Plants. 2023;12(22): 3864. https://doi.org/10.3390/plants12223864.

Publisher's note