



OPEN

Genetic dissection of grain iron and zinc, and thousand kernel weight in wheat (*Triticum aestivum* L.) using genome-wide association study

Gopalareddy Krishnappa^{1,2}, Hanif Khan^{1✉}, Hari Krishna³, Satish Kumar¹, Chandra Nath Mishra¹, Om Parkash¹, Narayana Bhat Devate³, Thirunavukkarasu Nepolean^{3,4}, Nagenahalli Dharmegowda Rathan³, Harohalli Masthigowda Mamrutha¹, Puja Srivastava⁵, Suma Biradar⁶, Govindareddy Uday⁶, Monu Kumar⁷, Gyanendra Singh¹ & Gyanendra Pratap Singh^{1✉}

Genetic biofortification is recognized as a cost-effective and sustainable strategy to reduce micronutrient malnutrition. Genomic regions governing grain iron concentration (GFeC), grain zinc concentration (GZnC), and thousand kernel weight (TKW) were investigated in a set of 280 diverse bread wheat genotypes. The genome-wide association (GWAS) panel was genotyped using 35 K Axiom Array and phenotyped in five environments. The GWAS analysis showed a total of 17 Bonferroni-corrected marker-trait associations (MTAs) in nine chromosomes representing all the three wheat subgenomes. The TKW showed the highest MTAs (7), followed by GZnC (5) and GFeC (5). Furthermore, 14 MTAs were identified with more than 10% phenotypic variation. One stable MTA i.e. *AX-95025823* was identified for TKW in both E4 and E5 environments along with pooled data, which is located at 68.9 Mb on 6A chromosome. In silico analysis revealed that the SNPs were located on important putative candidate genes such as *Multi antimicrobial extrusion protein*, *F-box domain*, *Late embryogenesis abundant protein*, *LEA-18*, *Leucine-rich repeat domain superfamily*, and *C3H4 type zinc finger protein*, involved in iron translocation, iron and zinc homeostasis, and grain size modifications. The identified novel MTAs will be validated to estimate their effects in different genetic backgrounds for subsequent use in marker-assisted selection. The identified SNPs will be valuable in the rapid development of biofortified wheat varieties to ameliorate the malnutrition problems.

Over three billion global population suffers from diseases associated with micronutrient deficiencies including iron and zinc and the problem is more severe in the countries where the food habits are dominated by cereal-based diets¹. Iron, zinc, and vitamin A are the three nutrients that are recognized as limiting factors in the diet by the world health organization². Approximately, one-fourth of the population around the globe is suffering from anemia due to iron deficiency³. The women of reproductive age are more vulnerable as one in every three women is anemic, which lead to 0.12 million deaths and a loss of 48.2 million disability-adjusted life years (DALY) in 2010⁴. Anemia due to acute iron deficiency, particularly in children, pregnant, and lactating women lead to life-threatening health complexes such as chronic heart disease, kidney failure, and inflammatory bowel disease⁵. Zinc is another important micronutrient essential for various immunological and biochemical functions and severe deficiency may lead to impaired growth and development, altered immunity, pregnancy issues, and neuro-behavioral difficulties⁶. Around 17% of the world's population is suffering from diseases related to zinc deficiency⁷, which leads to 97,330 deaths and a loss of 9.1 million DALY's in 2010⁴. Micronutrient deficiency is

¹ICAR-Indian Institute of Wheat and Barley Research, Karnal, India. ²ICAR-Sugarcane Breeding Institute, Coimbatore, India. ³ICAR-Indian Agricultural Research Institute, New Delhi, India. ⁴ICAR-Indian Institute of Millets Research, Hyderabad, India. ⁵Punjab Agricultural University (PAU), Ludhiana, India. ⁶University of Agricultural Sciences, Dharwad, India. ⁷ICAR-Indian Agricultural Research Institute, Gauria Karma, Jharkhand, India. ✉email: Hanif.Khan@icar.gov.in; GP.Singh@icar.gov.in

the major risk factor for health loss in developing countries and the most vulnerable groups are pregnant women and children⁸.

Wheat is one of the most widely cultivated cereals and plays a key role in global food and nutritional security. Although wheat is nutritionally rich as compared to the other two major cereals (rice and maize), still most wheat-based diets fail to meet the required quantity of essential nutrients including iron and zinc. The problem of micronutrient malnutrition can be overcome by food fortification, supplementation, and diet diversification, but were unsustainable in a long run. The affordability and accessibility, particularly for the rural poor in remote areas are the other shortcomings associated with the above-mentioned approaches⁹. Therefore, enhancement of the nutritional value of crops through conventional and molecular approaches, termed as “biofortification”, has been recognized as an economical and sustainable strategy to reduce the problems associated with micronutrient and protein malnutrition.

Genetic dissection of complex quantitative traits through trait mapping approaches is essential for developing better marker-assisted breeding and genomic selection strategies. The identification of linked molecular markers governing complex traits is highly useful and economical for trait improvement, especially in the post-genomics era where the genotyping costs become much cheaper. The quantitative inheritance of wheat quality traits and significant effects of environment and genotype-environment interaction (GEI) on the expression of GFeC GZnC and TKW were documented in several studies^{10–13}. In the past decade, extensive efforts have been made to identify QTLs associated with GFeC and GZnC^{14–28}, and TKW^{25,26,29–33} in wheat through bi-parental populations based QTL mapping. However, QTLs identified in such approaches had a low resolution due to the restricted number of crossovers. In contrast, the mapping resolution could be greatly increased by using linkage disequilibrium (LD)-based association mapping approach where the mapping population represents a more diverse gene pool and considers historical recombination events³⁴. This approach allows the detection of non-random associations of genome-wide markers with the phenotype³⁵ and has been used widely to detect the markers associated with the genomic regions governing complex traits in crop plants³⁶. The QTL resolution in association mapping has been significantly improved by using unrelated diverse genotypes that have accumulated many historical crossover events since their last common progenitors diverged³⁷.

Although many GWAS studies have been performed for various agro-morphological traits, only a limited number of studies were conducted for nutritional quality traits in wheat. Furthermore, hexaploid wheat has a genome size of ~ 17 Gb³⁸, and LD decay has not been well characterized. Alomari et al.³⁹ identified 40 MTAs for GZnC covering all the three wheat subgenomes in a panel of 369 genotypes using a high-density SNP array. Similarly, Bhatta et al.⁴⁰ used a diversity panel of synthetic hexaploid wheat (SHW), being a great reservoir of grain micronutrients, to identify 92 MTAs for 10 micronutrients including GFeC and GZnC. Velu et al.⁴¹ reported 39 MTAs for GZnC in a set of 330 bread wheat genotypes phenotyped in a wide range of environments. Liu et al.⁴² identified 14 significant MTAs for GFeC and GZnC, and manganese in a panel of 161 wild emmer-derived advanced lines. Genetic dissection of micronutrients including GFeC and GZnC has been performed in a diverse HarvestPlus association mapping panel consisting of 330 genotypes from CIMMYT’s biofortification breeding program⁴³. A total of 16 loci were identified which are associated with the GZnC on 11 different chromosomes covering all three wheat subgenomes in a set of 246 wheat varieties⁴⁴. Similarly, Calderini et al.⁴⁵ used a set of 167 *Ae. tauschii* accessions to map nine MTAs governing GFeC and GZnC⁴⁶. A total of 29 unique loci associated with grain GZnC was identified in a diversity panel of 207 bread wheat genotypes⁴⁷.

The TKW has no nutritional value per se in wheat, however, it has a dilution effect on protein and micronutrients. Therefore, TKW is one of the important breeding objectives due to its twin effects on yield and protein. The MTAs have been identified for TKW^{48–51,91} using different compositions of GWAS panels. Therefore, more GWAS studies would be helpful to identify the genomic regions governing nutritional traits in wheat and also to identify the candidate genes to develop biofortified cultivars. The present study aimed to identify the genomic region(s) associated with GFeC, GZnC, and TKW in diverse bread wheat genotypes in a range of environments through the GWAS approach and the putative candidate genes associated with the SNPs.

Materials and methods

Plant material and field experiments. A set of 280 genetically diverse bread wheat genotypes (Supplementary Table S1) consisting of advanced breeding lines and commercial cultivars were used for GWAS analysis. The study material in GWAS panel with 280 genotypes was selected from All India Coordinated Research Project on Wheat and Barley. The GWAS panel was evaluated at five different environments: E1-University of Agricultural Sciences, research farm, Dharwad (15° 29′ 20.71″ N, 74° 59′ 3.35″ E, 750 m AMSL), E2-ICAR-Indian Agricultural Research Institute, New Delhi (28° 38′ 30.5″ N, 77° 09′ 58.2″ E, 228 m AMSL), E3-Indian Agricultural Research Institute, Jharkhand (24° 16′ 58.4″ N, 85° 21′ 16.1″ E, 651 m AMSL), E4-ICAR-Indian Institute of Wheat and Barley, Karnal (29° 41′ 8.2644″ N, 76° 59′ 25.9692″ E, 250 m AMSL), and E5-Punjab Agricultural University, Ludhiana (30° 54′ N, 75° 48′ E, 247 m AMSL). The crop was sown in the first fortnight of november during the 2020–2021 *Rabi* (winter) season under irrigated condition. The genotypes were planted in an augmented block design with only the checks (DBW187, MACS6222, WH1124, and WH1142) repeated in a 2 row of 2 m length with a row spacing of 20 cm.

Phenotyping and phenotypic data analysis. Randomly selected 20–25 spikes were harvested and bulk-threshed manually in a clean cloth bag without touching any metal to avoid contamination. Around 20 g of grain sample from each genotype were used for phenotyping GFeC and GZnC through high-throughput Energy Dispersive X-ray Fluorescence (ED-XRF) machine (model X-Supreme 8000; Oxford Instruments plc, Abingdon, United Kingdom) calibrated with glass beads-based values. To record TKW, the Numigral grain counter was used to count the grain number, the reading was set at 1000 grains and the weight of the grains was recorded

in grams with an electronic balance. Phenotypic data were analysed using the R package ‘augmentedRCBD’⁵². Coefficient of variation (CV), broad-sense heritability (h^2_{BS}), genotypic variance (σ^2_G), and environmental variance (σ^2_E) were calculated using the following formula:

$$CV(\%) = SD/\bar{x} \times 100$$

where SD = Standard deviation; \bar{x} = Arithmetic mean.

$$LSD = t_{0.025,DF_w} \sqrt{MS_w(1/n_1 + 1/n_2)}$$

where $t_{0.025,DF_w}$ = The t-critical value from the t-distribution table with $\alpha=0.025$ and DF_w is the degrees of freedom within groups from the ANOVA table. MS_w = The mean squares within groups from the ANOVA table. n_1 and n_2 = The sample sizes for the first and second comparing samples

$$\text{Heritability}(h^2_{BS}) = \frac{\sigma^2_G}{\sigma^2_G + \frac{\sigma^2_E}{n_{\text{Block}}}} \times 100$$

where σ^2_G = Genetic variance was calculated as $(MS_{\text{treatments}} - MS_{\text{residuals}})/n_{\text{Block}}$; σ^2_E = Residual variance = MS_{residual} ; n_{Block} = Number of blocks

$$\sigma^2_G = \frac{MS_{\text{treatments}} - MS_{\text{residuals}}}{b}$$

$$\sigma^2_E = MS_{\text{residuals}}$$

where $MS_{\text{treatments}}$ = Treatment mean sum of square; $MS_{\text{residuals}}$ = Error mean sum of square; b = Number of blocks.

The CV indicates the degree of precision with which the treatments are compared and is a good index of the experimental reliability. It expresses the experimental error as percentage of the mean and if the value is high then the precision of the experiment is low and vice versa. The h^2_{BS} is the proportion of phenotypic variation that is attributable to an overall genetic variation for the genotypes. LSD is the value at a particular level of statistical probability, when exceeded by the difference between two genotypes means, then the two genotypes are said to be distinct for at that or lesser levels of probability. The σ^2_G is the genetic or inherent variation that remains unaltered by environmental changes, this kind of variation responds to the selection during breeding process. In contrast, σ^2_E does not respond to selection as it is non-heritable, which is entirely due to environmental effects.

Genotyping. Genomic DNA of the GWAS panel was extracted from the leaves of 21 days-old seedlings by Cetyl Trimethyl Ammonium Bromide (CTAB) method⁵³. The panel was genotyped using Axiom Wheat Breeder’s Genotyping Array (Affymetrix, Santa Clara, CA, United States) having 35,143 genome-wide SNPs. The monomorphic, markers with minor allele frequency (MAF) of < 5%, missing data of > 20%, and heterozygote frequency > 25% were removed from the analysis. The remaining set of 14,790 high-quality SNPs was used in GWAS analysis (Supplementary Table S3).

Population Statistics and GWAS. The pair-wise LD values (r^2) between the SNPs located in each chromosome were calculated with Trait Analysis by aSSociation Evolution and Linkage (TASSEL) version 5.0⁵⁴. The LD block size in three different subgenomes as well as in the whole genome was calculated by keeping r^2 threshold at half LD decay (Fig. 3). The principal component analysis (PCA) was done through GAPIT⁵⁵ to understand the structure of the population and included in the GWAS model to correct the structure. Furthermore, Kinship relationship was calculated through GAPIT⁵⁵ and presented in Fig. 2C. Additionally, the structure of the population was evaluated through the STRUCURE program by keeping K-value from 1 to 10. For every single K-value, 3 independent runs were used and each run was set with 10,000 burn-in iterations followed by 10,000 Markov Chain Monte Carlo (MCMC) replications after burn-in. The STRUCTURE HARVESTER⁵⁶ was used to detect the optimal K-value based on ad-hoc method described by Pritchard et al. 2010⁹⁰ as well as Evanno’s method⁵⁷. The suitability of the model to account for population structure was assessed using quantile–quantile (Q–Q) plots.

The phenotypic values of GFeC, GZnC, and TKW of 280 diverse genotypes along with corresponding genotyping data were used in GWAS analysis. Significant MTAs were identified using BLINK (Bayesian-information and Linkage-disequilibrium Iteratively Nested Keyway) model⁵⁸ implemented in Genome Association and Prediction Integrated Tool (GAPIT) version 3.0⁸⁰ in R software package. Determining the correct P-value threshold for statistical significance is critical to differentiate true positives from false positives. To determine the statistical significance threshold in GWAS, Bonferroni correction has been employed. To estimate Bonferroni correction, α was set to 0.05 and which is divided by total number of SNPs. The Bonferroni-corrected SNPs were considered for significant association and R^2 was used to describe the percentage variation explained (PVE) by significant MTAs.

In silico analysis. The sequence information of the significant SNPs was used to search for putative candidate genes with Basic Local Alignment Search Tool (BLAST) using default parameters in the Ensemble Plants database (<http://plants.ensembl.org/index.html>) of the bread wheat genome (IWGSC (RefSeq v1.0)). The genes found in the overlapping region and within the region of 10 Kb intervals flanking either side of the associated marker were considered as putative candidate genes and their molecular functions were determined. In addition,

Trait	Env.	Mean \pm SD	Range	CV (%)	LSD	h_{bs}^2	σ_G^2	σ_E^2
GFeC (mg/kg)	E1	39.5 \pm 3.62	31.9–49.0	3.0	3.3	89.7	12.1	1.4
	E2	37.8 \pm 2.85	30.6–47.5	2.9	3.1	83.6	6.2	1.2
	E3	40.7 \pm 3.1	32.2–49.9	4.1	4.7	71.3	6.7	2.7
	E4	35.3 \pm 4.02	26.3–47.9	7.2	7.2	45.4	5.4	6.4
	E5	40.2 \pm 3.49	32.4–49.8	3.1	3.5	87.7	10.9	1.5
GZnC (mg/kg)	E1	39.6 \pm 7.29	22.6–62.5	9.9	11.1	67.8	32.4	15.4
	E2	43.8 \pm 5.26	32.8–57.8	5.4	6.7	77.7	19.5	5.6
	E3	35.2 \pm 5.91	21.3–52.9	13.1	13.0	34.8	11.3	21.2
	E4	36.9 \pm 4.02	28.6–47.6	8.7	9.0	33.3	5.0	10.1
	E5	45.9 \pm 8.44	24.7–64.1	6.9	9.1	84.9	57.9	10.3
TKW (gm)	E1	45.6 \pm 5.17	29.6–58.2	1.3	1.6	98.8	26.9	0.3
	E2	41.3 \pm 3.65	30.3–51.0	2.4	2.8	92.7	12.4	1.0
	E3	46.1 \pm 4.17	34.9–59.3	2.2	2.8	94.2	16.4	1.0
	E4	44.6 \pm 4.35	29.9–57.9	3.1	3.9	89.9	17.1	1.9
	E5	40.5 \pm 4.11	26.0–49.0	1.6	1.9	97.4	16.3	0.4

Table 1. Descriptive statistics, variance and heritability estimates of grain quality traits in GWAS panel evaluated at five environments during 2020–2021. GFeC: grain iron concentration; GZnC: grain zinc concentration; TKW: thousand kernel weight; E1–Dharwad; E2–IARI, New Delhi; E3–IARI, Jharkhand; E4–Karnal; E5–Ludhiana; Env.: Environment; SD: standard deviation; CV: coefficient of variation; h_{bs}^2 : broad sense heritability; σ_G^2 : genotypic variance; σ_E^2 : environmental variance.

their expression patterns were investigated using the Wheat Expression database (<http://www.wheat-expression.com/>) and potential links to phenotypes was determined using Knetminer tool integrated with Wheat Expression database. The role of the identified putative candidate genes in the regulation of GZnC and GFeC, and TKW was also determined with the previous reports.

Results

Variability, heritability, and correlations. The environment-wise heritability and variance components of the GWAS panel for GFeC, GZnC, and TKW are presented in Table 1. The GFeC ranged from 26.3 mg/kg to 49.9 mg/kg, whereas, the GZnC recorded a wider distribution across the environments, as it ranged from 21.3 mg/kg to 64.1 mg/kg. Similarly, TKW ranged from 26.0 gm to 59.3 gm. The trait-wise heritability was recorded highest for TKW followed by GFeC, and GZnC, whereas, the trend for the coefficient of variation (CV) was exactly opposite with the lowest recorded for TKW followed by GFeC, and GZnC. The environment-wise heritability was ranged from 45.4% (E4) to 89.7% (E1), 33.3% (E4) to 84.9% (E5), and 89.9% (E4) to 98.8% (E1) respectively, for GFeC, GZnC, and TKW. For all the three traits, E4 has been recorded as the lowest heritability, which was corroborated with the highest recorded CV for E4. The genotypic variance (σ_G^2) and environmental variance (σ_E^2) are presented in Table 1.

The trait and environment-wise mean values are illustrated graphically through boxplots and presented in Fig. 1. The location means of GFeC were recorded as similar and highest for E3 and E5 followed by E1, E2, and E4, whereas, E5 was recorded highest pooled mean followed by E2, E1, E4, and E3 for GZnC. The E3 and E1 recorded a similar and highest mean for TKW followed by E4, E2, and E5. The frequency distribution of grain quality traits in the GWAS panel evaluated at E1–E5 during 2020–2021 is presented in Fig. 1. The genotypes in the GWAS panel showed continuous frequency distributions for all the studied traits. Partial correlation coefficient (r^2) of GFeC, and GZnC by keeping TKW as a controlling factor was determined. Highly significant and positive correlation was observed between GFeC and GZnC in E1 (0.296**), E2 (0.276**), E3 (0.202**), E4 (0.520**), and E5 (0.35**) and also in pooled data (0.358**).

SNP markers statistics. The quality processing of 35,143 SNPs from 35 K array resulted in a set of 14,790 cured genome-wide SNPs. These high-quality set of SNPs were further used for GWAS analysis. The chromosome and genome-wise marker distribution are presented in Table 2. The highest number of SNPs were mapped on the B genome (5649) followed by the D genome (4590), and the A genome (4551).

Population structure and linkage disequilibrium. The PCA plot (Fig. 2B) indicated that there were no clear distinct sub-populations in the GWAS panel; however, STRUCTURE grouped the GWAS panel into eight sub-populations (Fig. 2A). The LD was estimated by calculating the squared correlation coefficient (r^2) for all the SNPs and plotted against the genetic distance (bp). The LD decay for the whole genome was 4.9 cM and it was found that the decay was rapid in the A subgenome (3.6 cM) followed by the B subgenome (5.7 cM) and the D subgenome (5.2 cM) (Fig. 3).

Genome-wide association studies. A total of 17 Bonferroni-corrected MTAs were identified for GFeC, GZnC, and TKW. The details of the identified MTAs are presented in Table 3 and illustrated in Manhattan plots

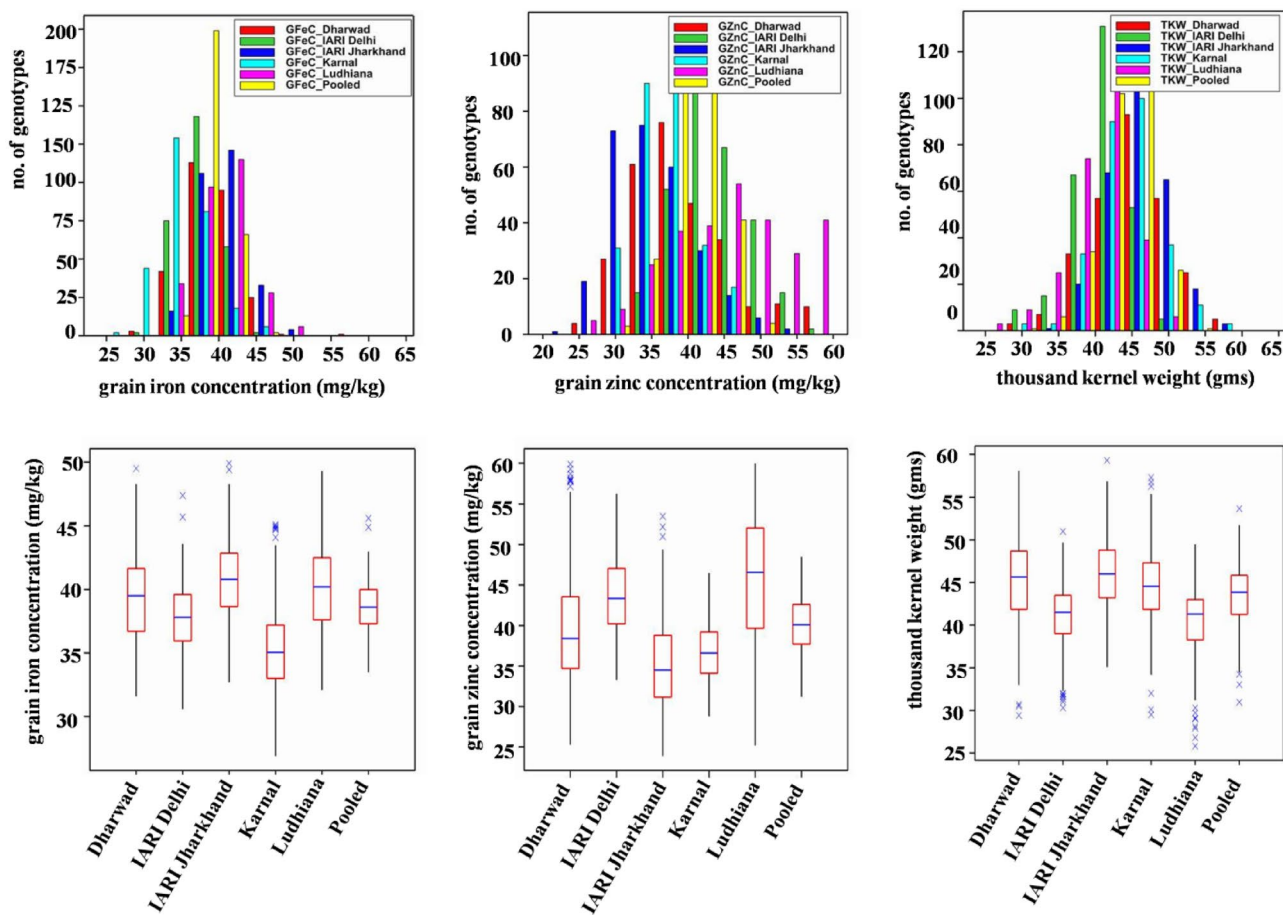


Figure 1. Frequency distribution and boxplots of grain quality traits in GWAS panel evaluated at Dharwad, IARI Delhi, IARI Jharkhand, Karnal, and Ludhiana during 2020–2021.

Genome	Chromosome-wise SNP distribution							Total
	1	2	3	4	5	6	7	
A	751	756	587	493	699	515	750	4551
B	1077	992	726	465	863	766	760	5649
D	986	951	648	264	657	459	625	4590

Table 2. Sub-genome and chromosome-wise distribution of SNP markers in the GWAS panel.

in Fig. 4A,B. The Q-Q plots depicting the observed associations of SNPs and GFeC, GZnC, and TKW compared to the expected associations after accounting for population structure are presented in Fig. 4A,B.

MTAs for GFeC and GZnC. A total of five significant MTAs were identified for GFeC in E2 and E4 environments on chromosomes 6A, 3B, 1A, 7B, and 5A explaining the phenotypic variation ranged from 12.7% to 24.1%. Two major SNPs (*AX-9469986* and *AX-95140213*) on 7B and 5A chromosomes located at 706.0 Mb and 558.3 Mb explained the highest phenotypic variation of 24.1% and 23.1%, respectively, in E4 environment. One SNP each on chromosome 6A (*AX-94423274*), 3B (*AX-94490975*), and 1A (*AX-95195514*) were mapped at 609.1 Mb, 795.8 Mb, and 354.9 Mb, respectively with the phenotypic variation of 15.6%, 12.7%, and 13.0% in E2 environment.

A total of 5 MTAs were identified for GZnC on chromosome 7B, 6A, 2B, 5B, and 7B explaining the phenotypic variation ranged from 5.7% to 10.9%. The B subgenome contributed more MTAs (4) followed by A subgenome (1), whereas, D subgenome didn't contribute for GZnC in the present study. Two major SNPs (*AX-95118780* and *AX-95140213*) on 7B chromosome located at 91.6 Mb and 94.2 Mb explained the highest phenotypic variation of 10.9% and 10.0%, respectively in E1 and E3 environments. Another major SNP (*AX-95113687*) on the A subgenome (6A chromosome) mapped at 595.5 Mb, explained 10.1% phenotypic variation in E1. The remaining two SNPs (*AX-94390652* and *AX-94524014*) on 2B and 5B chromosomes mapped at 201.4 Mb and 440 Mb explained 5.7% and 8.8% phenotypic variation, respectively in E1.

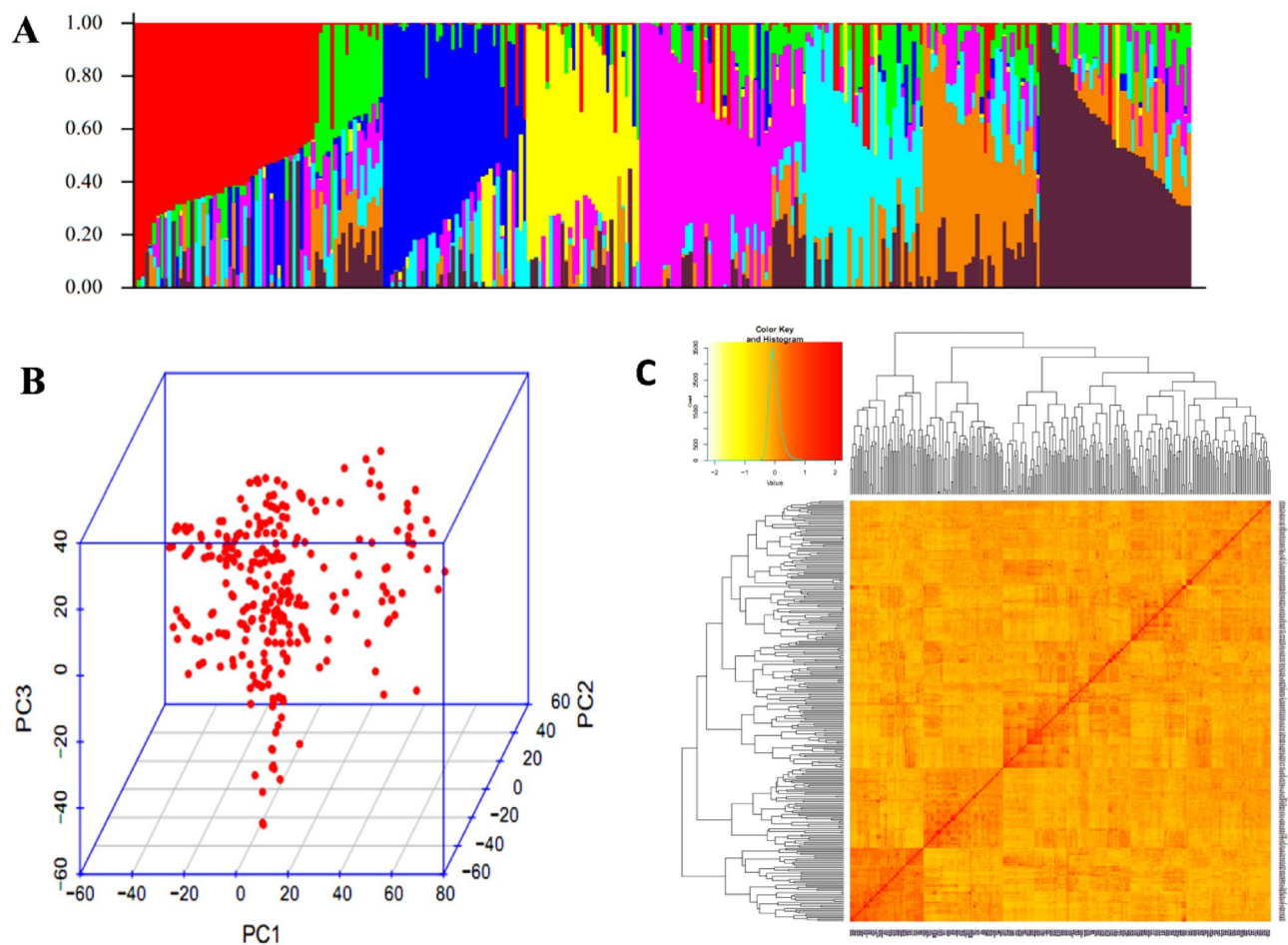


Figure 2. Population groupings in GWAS panel from different models. **(A)** Population structure based on STRUCTURE **(B)** Three-dimensional plot of the first three principal components, and **(C)** heat map of pairwise kinship matrix.

MTAs for TKW. A total of seven MTAs were identified covering all the subgenomes. Major phenotypic variation was observed from those MTAs which were ranging from 10.7% to 17.4%. The three subgenomes mapped more or less the same number of MTAs (A subgenome-3, B and D subgenomes-2 each). Six MTAs (AX-94764034, AX-95025823, AX-94452219, AX-94820753, AX-94569403, and AX-95235178) were identified in E4 on chromosome 5A, 6A, 7B, 5B, 2D, and 1A at 444.8 Mb, 68.9 Mb, 131.7 Mb, 689.9 Mb, 461.3 Mb, and 499.8 Mb, respectively with a corresponding phenotypic variation of 16.1%, 16.1%, 14.9%, 13.7%, 17.4%, and 16.7%. A total of 2 MTAs (AX-95117294 and AX-95025823) were mapped in E5 located at 290.3 Mb and 68.9 Mb, which explained 10.7% and 11.7% phenotypic variation on 5D and 6A, respectively.

For pooled TKW data, one MTA (AX-95025823) was mapped on 6A and located at 68.9 Mb, which explained 16.9% phenotypic variation. One stable MTA i.e. AX-95025823 was identified in both E4 and E5 environments along with pooled data, which is located at 68.9 Mb on 6A chromosome.

Identification of putative genes associated with MTAs. The significant SNPs associated with GFeC, GZnC, and TKW were used to identify the putative candidate genes using the annotated wheat reference sequence (RefSeq V1.0) and are presented in Table 4 and Supplementary Table 2. AX-94490975 associated with GFeC found to encode *Multi antimicrobial extrusion protein* (TraesCS3B02G562500). Similarly, another SNP i.e. AX-94699865 associated with GFeC encodes an important *F-box domain* (TraesCS7B02G312400). Two important SNPs i.e. AX-94524014 (TraesCS5B02G257700) and AX-95203413 (TraesCS7B02G083600) associated with GZnC were found to encode *Late embryogenesis abundant protein*, *LEA-18* and *RNA recognition motif domain*. Similarly, AX-95235178 encoding *Leucine-rich repeat domain superfamily* (TraesCS1A02G309000) and AX-95117294 encoding *C3H4 TYPE ZINC FINGER PROTEIN* (TraesCS5D02G188300) identified for TKW.

Discussion

Understanding the genetic basis of complex traits such as GFeC, GZnC, and TKW through GWAS with a diverse panel of genotypes can significantly improve QTL mapping resolution compared to bi-parental populations-based QTL mapping. Using the genome-wide SNPs and multi-environment data, several significant SNPs were identified in this study.

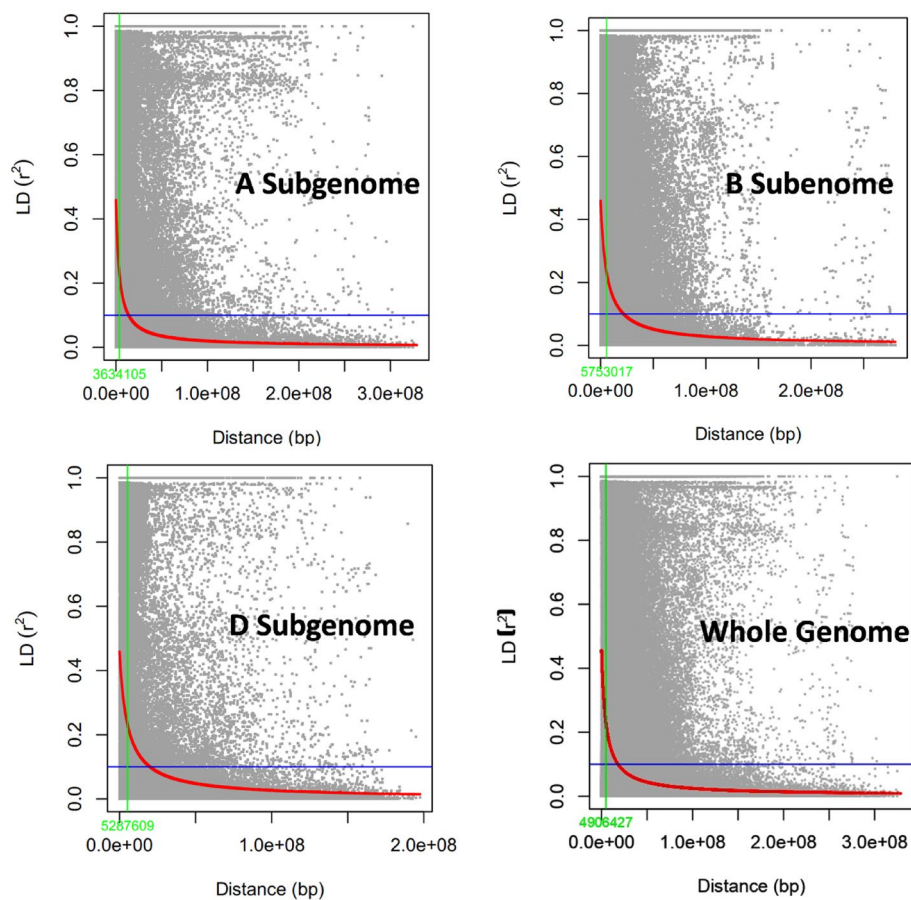


Figure 3. Subgenome and whole genome-wide linkage disequilibrium (LD) decay in GWAS panel of 280 diverse bread wheat genotypes.

Trait	Env	SNPs	Chromosome	Position (bp)	P value	PVE (%)
GFeC	E2	AX-94423274	6A	609111057	1.02E-08	15.6
	E2	AX-94490975	3B	795800318	0.000000747	12.7
	E2	AX-95195514	1A	354950241	0.00000214	13
	E4	AX-94699865	7B	558319306	3.69E-17	24.1
	E4	AX-95140213	5A	706021202	0.00000203	23.1
GZnC	E1	AX-95118780	7B	91660994	1.68E-10	10.9
	E1	AX-95113687	6A	595578858	1.11E-07	10.1
	E1	AX-94390652	2B	201463130	1.10E-06	5.7
	E1	AX-94524014	5B	440179428	2.55E-06	8.8
	E3	AX-95203413	7B	94271868	3.06E-08	10
TKW	E4	AX-94764034	5A	444849916	1.19E-09	16.1
	E4	AX-95025823	6A	68975107	0.00000049	16.1
	E4	AX-94452219	7B	131745573	0.000000747	14.9
	E4	AX-94820753	5B	689950369	0.00000147	13.7
	E4	AX-94569403	2D	461303027	0.00000256	17.4
	E4	AX-95235178	1A	499807792	0.00000257	16.7
	E5	AX-95117294	5D	290389058	0.000000555	10.7
	E5	AX-95025823	6A	68975107	0.000000821	11.7
Across Env	AX-95025823	6A	68975107	5.83E-08	16.9	

Table 3. MTAs for grain quality traits and TKW identified in the GWAS panel from five environments. Env.: Environment; GFeC: grain iron concentration; GZnC: grain zinc concentration; TKW: thousand kernel weight; E1: Dharwad; E2: IARI Delhi; E3: IARI Jharkhand; E4: Karnal; E5: Ludhiana; SNPs: single nucleotide polymorphisms; PVE: phenotypic variation explained.

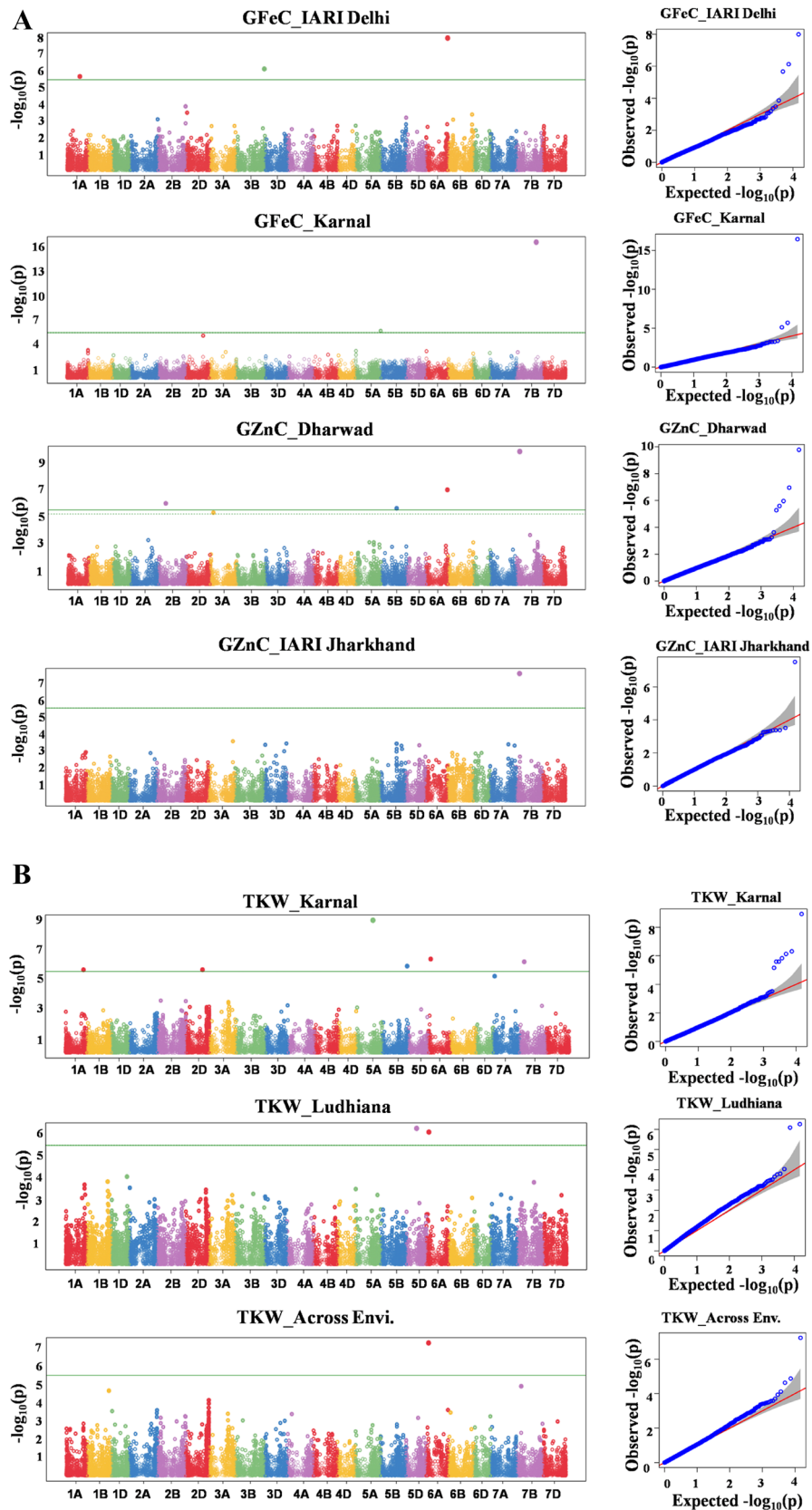


Figure 4. (A) Manhattan and respective-QQ plots for grain iron and zinc concentration in GWAS panel phenotyped at Dharwad, IARI Delhi, IARI Jharkhand, Karnal, and Ludhiana during 2020–2021. (B) Manhattan and respective-QQ plots for thousand kernel weight in GWAS panel phenotyped at Dharwad, IARI Delhi, IARI Jharkhand, Karnal, and Ludhiana during 2020–2021.

Trait	SNP ID	Chromosome	TransID	Position (bp)	Putative candidate genes	Molecular function
GFeC	AX-94490975	3B	TraesCS3B02G562500	795800216–795810270	Multi antimicrobial extrusion protein	Iron translocation in bread wheat ⁵⁹ Efficient translocation of iron from roots to shoots in rice ⁶⁰ Iron homeostasis in Arabidopsis ^{61,81,82} Fe homeostasis and Zn tolerance in Arabidopsis ⁶² Fe influx into aerial parts of the plant and/or for the distribution of intracellular Fe ^{63,64} Aluminum tolerance and iron translocation in <i>Arabidopsis thaliana</i> ⁶⁵ Iron transportation in rice ⁸³ Efficient translocation of Fe under limited Fe conditions in rice ⁶⁰ Efficient translocation of iron in Arabidopsis ⁸⁴ Overexpression of <i>MtMATE69</i> affected Fe and Zn accumulation in <i>M. truncatula</i> and play role in Fe nutrition ⁸⁵ Iron efficiency in soybean ⁸⁶
	AX-94699865	7B	TraesCS7B02G312400	558318164–558320180	F-box domain	F-box domain RAE1 regulates STOP1 in Arabidopsis. STOP1-ALMT1 pathway promote iron accumulation into the apoplast of root tip regions under Pi-deficient conditions ^{66,67}
GZnC	AX-94524014	5B	TraesCS5B02G257700	440178901–440179528	Late embryogenesis abundant protein, LEA-18	Iron transportation in the phloem of castor (<i>Ricinus communis</i> L.) ⁶⁸ Zinc ion binding in cotton ⁶⁹
TKW	AX-95235178	1A	TraesCS1A02G309000	499807479–499809182	Leucine-rich repeat domain superfamily	Regulate grain size in rice ⁷¹ Role in total kernel number and kernel size in maize ⁸⁷ Controls endosperm development and thereby seed size ⁸⁸
	AX-95117294	5D	TraesCS5D02G188300	290386100–290391792	C3H4 TYPE ZINC FINGER PROTEIN	Important agronomic traits in maize yield ⁸⁹

Table 4. Putative candidate genes identified for GFeC, GZnC, and TKW along with their molecular functions.

The expression of GFeC, GZnC, and TKW is significantly affected by the environment and genotype-environment interactions (GEI). Among all traits, GZnC was the most environment-sensitive trait, whereas, TKW was relatively the most stable with minimum environmental influence. The greater magnitude of the environment and GEI have also been reported in previous studies for the expression of GFeC and GZnC^{10,11}, and TKW^{12,13}. The magnitude of environmental interaction decides the identification of environment-specific QTL(s) as well as QTL(s) that can express stably across environments.

The highest heritability was recorded for TKW followed by GFeC, and GZnC, whereas, the trend for the coefficient of variation (CV) was exactly opposite with the lowest CV recorded for TKW and the highest CV for GZnC. The highest and lowest heritability for TKW and GZnC respectively is also concurred with earlier studies^{46,72}. The associations were highly significant positive in all the environments between GFeC and GZnC. Significant and positive correlations found in this study have also been reported in earlier studies^{25,26}. The significant positive correlations between GZnC and GFeC indicated the possibility to map the genomic regions controlling multiple traits. Such co-mapped SNPs will be much useful in marker-assisted selection for simultaneous improvement of correlated traits.

The STRUCTURE model explained 8 sub-groups in the populations. The genotypes in GWAS panel consists of advanced breeding lines suitable for various agro-climatic and production conditions. The first subgroup consists of genotypes mostly selected from international breeding material and suited for North West and North East Plains Zone in India. Similarly, the second group consists of international selections for restricted irrigated or rainfed production conditions. The third subgroup consists of genotypes dominated by 1B.1R translocation with genes for wider adaptation. Subpopulation 4 is mainly dominated by GW322, PASTOR, and OPATA parentage, whereas, 5th subpopulation largely consists of Indian wheat varieties/germplasm in their parentage. High frequency of SOKOLL, KIRITATI, PBW65, and MILAN was present in the 6th subpopulation parentage. Genotypes in 7th subpopulation are dominated by old salinity/alkalinity tolerant varieties. Whereas, 8th subpopulation contains mainly indigenous germplasm, old landraces, and breeding lines. The PC1, PC2 and PC3 of PCA analysis were used as covariates in the GWAS analysis to identify the MTAs. The LD may vary in different populations due to population size, genetic drift, admixtures, selection, mutation, non-random mating, pollination behavior, and recombination frequency^{73,74}. The LD blocks are usually larger in self-pollinated crops such as wheat and hence decay slowly⁷⁵, whereas, in outcrossing crop species like maize⁷⁶, the LD decays rapidly. The presence of high LD across the genome would reduce the QTL mapping resolution and vice versa⁷⁷. In such cases, a better QTL resolution will be achieved by using genome-wide SNPs. The decay of LD was found comparable in the B and D subgenomes (~ 5 cM) compared to the A subgenome, which had a shorter decay distance of around ~ 3 cM. A similar pattern of LD decay was also observed in other GWAS studies in wheat^{49,78,79,91}.

A total of 17 Bonferroni-corrected MTAs were identified for GFeC (5), GZnC (5), and TKW (7). The identified genome-wide MTAs are much higher for B subgenome (8) and A subgenome (7) compared to the D subgenome

(2). A similar trend on MTAs identified in the D subgenome for GFeC and GZnC⁴¹ and yield-contributing traits^{49,50}.

The identified MTAs (5) for GFeC on chromosomes 6A, 3B, 1A, 7B, and 5A in this study were novel, as the earlier reported MTAs on the same chromosomes namely 3B, 7B, and 5A^{43,44}, 1A, 3B, and 5A^{23,26}, 1A⁴⁰ and 6A, 3B⁹¹ were identified at different positions. A total of five novel MTAs was identified for GZnC on chromosome 7B, 6A, 2B, 5B, and 7B. MTAs in the same chromosomes were also identified in different GWAS panels in previous experiments on 6A⁴⁷, 2B^{26,41,43,47}, 5B^{23,43,47}, and 7B^{23,41,91}. Zhou et al.⁴⁷ identified an MTA on 5B chromosome located in a interval of 407.0 Mb – 412.1 Mb, which was similar to that of *AX-94524014* located on 5B chromosome and mapped at 440.1 Mb explained 8.8% phenotypic variation.

A total of seven MTAs in different environments were identified covering the three subgenomes and all were major MTAs as they explained more than 10.0% phenotypic variation. The TKW was relatively the most stable trait compared to the rest of the other two traits, as TKW recorded the highest heritability and lowest coefficient of variation which reflected in detecting the highest number of MTAs as well. All the identified MTAs were mapped on 5A, 6A, 7B, 5B, 2D, 7A, and 5D located at 444.8 Mb, 68.9 Mb, 131.7 Mb, 689.9 Mb, and 461.3 Mb, 499.8 Mb, and 290.3 Mb respectively. Previous reports were also identified MTAs on 6A and 7B^{29,48,50}, 5B^{26,49}, 5A²⁶, 1D²⁹, 1A⁴⁸ and 2D, 5D, 7A and 7B⁹¹.

The various putative candidate genes underlying MTAs with high phenotypic variation for GZnC, GFeC, and TKW were identified through BLAST search (Table 4 and Supplementary Table 2). The MTAs identified in various chromosomes were located in gene coding regions related to transcription factors, transporters, transmembrane protein and kinase-like superfamilies. For example, *Multi antimicrobial extrusion protein* (TraesCS3B02G562500) has a role in the translocation of iron during iron deficiency stress in bread wheat⁵⁹ and multi antimicrobial extrusion protein (MATE) family proteins were observed under iron excess in rice. Few protein members of MATE family were known to be involved in efficient iron translocation from roots to shoots in rice⁶⁰. Also, MATE transporter mediates iron homeostasis under osmotic stress in Arabidopsis⁶¹. Subfamily III of the MATE gene members plays an important role in plant aluminum tolerance and iron translocation in Arabidopsis⁶⁵. *FRD3* MATE transporter locus reveals cross-talk between Fe homeostasis and Zn tolerance in Arabidopsis by loading Zn into xylem⁶². MATE is also a candidate for the mechanism of Fe influx into aerial parts of the plant and the distribution of intracellular Fe^{63,64}.

Three up-regulated genes i.e. *Os01g0684900*, *Os10g0345100*, and *Os06g0495500* of citrate transporters family (MATE family protein) were observed under excess iron conditions and involved in iron transportation in rice⁸³. Similarly, a MATE gene (*OsFRDL1*), the closest homolog of barley *HvAACT1* (aluminum-activated citrate transporter 1) is involved in the efficient translocation of Fe under limited Fe conditions⁶⁰. *FRD3* is a member of the multidrug and toxin efflux (MATE) family, which is involved in the efficient translocation of iron in Arabidopsis⁸⁴. *FRD3* is mainly expressed in root vascular tissues and is necessary to solubilize Fe and Zn in the extracellular space. Similarly, overexpression of *MtMATE69* affected Fe and Zn accumulation in *Medicago truncatula* hairy roots, further suggesting a function for *MtMATE69* in Fe nutrition⁸⁵. Also, two MATE proteins namely, *GmFRD3a* and *GmFRD3b* play a significant role in iron efficiency in soybean⁸⁶. Cloning and characterization of an Arabidopsis gene i.e. *FRD3*, a member of the multidrug and toxin efflux family is involved in iron homeostasis⁸¹. The *FRD3*, which is an efflux transporter of the efficient Fe chelator citrate is involved in Fe homeostasis maintenance throughout plant growth and development. Additionally to its well-known root expression, *FRD3* is also strongly expressed in seeds and flowers⁸².

One SNP i.e. *AX-94699865* associated with GFeC encodes an important *F-box domain* (TraesCS7B02G312400) regulates STOP1 in Arabidopsis. STOP1-ALMT1 pathway promotes iron accumulation into the apoplast of root tip regions under Pi-deficient conditions^{66,67}. Another SNP i.e. *AX-94524014* (TraesCS5B02G257700) associated with GZnC was found to encode *LEA protein*, where *LEA-18* was involved in the transportation of iron in the phloem of castor⁶⁸. The binding of LEA proteins to different molecules like Zn ion, DNA and ATP binding, were the major activities for the action of upland LEA proteins⁶⁹.

The SNP i.e. *AX-95235178* encoding *Leucine-rich repeat domain superfamily* (TraesCS1A02G309000) was associated with TKW. A total of 32 barley orthologs were identified as potential candidate genes that determine barley grain size or weight. The barley ortholog of the rice *OsBDG1* gene is mapped on 3H chromosome at 666.35 Mb (*HORVU3Hr1G104350*), which encodes the leucine-rich repeat receptor-like protein kinase family⁷⁰. The rice *OsBDG1* gene encoding a small protein with short leucine-rich-repeats possessing cell elongation activity, has previously been proven to positively regulate grain size in rice⁷¹. Therefore, *HORVU3Hr1G104350* could be a reliable candidate gene affecting grain size as the function of the *OsBDG1* gene.

Another grain weight controlling gene i.e. *FASCIATED EAR2 (FEA2)* encodes the maize ortholog of *CLAVATA2 (CLV2)*, encoding a leucine-rich repeat receptor-like protein that regulates meristem size by transmitting signals from *CLAVATA3 (CLV3)* peptide ligand to the *WUSCHEL (WUS)* homeodomain transcription factor. The *FEA2* has a role in total kernel number and kernel size in maize⁸⁷. Similarly, IKU pathway represents one of the well-studied genetic networks involves four major genes including *HAIKU2 (IKU2)*, which encodes a leucine-rich repeat kinase, mutational analyses of these genes in Arabidopsis revealed their physiological significance in controlling endosperm development and thereby seed size through regulating endosperm proliferation and cellularization⁸⁸, and loss of function mutations in IKU pathway genes cause a decrease in seed size. Another SNP (*AX-95117294*) encoding C3H4 type zinc finger protein (TraesCS5D02G188300) was associated with the expression of TKW. Functional prediction of maize C2H2—zinc finger gene revealed its involvement mainly in the formation of important agronomic traits in maize yield⁸⁹.

The study with 280 diverse set of bread wheat GWAS panel has shown that GFeC, GZnC, and TKW were quantitatively inherited traits. The strong positive correlation between the GFeC and GZnC suggested the possibility of improving both the traits simultaneously. A total of 17 MTAs including 5 for GFeC, 5 for GZnC, and 7 for TKW were identified from the GWAS approach. The environment-specific and pooled-data MTAs identified

in the present investigation represented novel genomic regions associated with trait expression. Several putative candidate genes encoding important molecular functions such as iron translocation, iron and zinc homeostasis, and grain size modifications were associated with the identified MTAs. Further validation and functional characterization of the candidate genes to elucidate the role of these genes in wheat is envisaged. The identified SNPs could be useful in marker-assisted selection programs to develop biofortified varieties to reduce micronutrient malnutrition.

Declaration. The set of 280 genotypes used in the present experiment were selected from All India Coordinated Research Project on Wheat and Barley and the imported genotypes have been obtained through the nodal agency for germplasm exchange i.e. National Bureau of Plant Genetic Resources, New Delhi following the prescribed guidelines. Also, the authors have all the required permissions and rights to collect and use the genotypes for research purpose. The experimental research and field experiments in the present study are duly approved by the institute research council of ICAR-Indian Institute of Wheat and Barley Research, Karnal.

Data availability

All the phenotypic and genotypic data used in the study is given as Supplementary Table 3 and also deposited in DRYAD (an international open-access repository of research data). The international open-access data repository link is as follows: https://datadryad.org/stash/share/FaLrCt8Hd4sKQzWyGSNYQOQ_tGtdSKIE3a9rBYAVTM.

Received: 21 February 2022; Accepted: 4 July 2022

Published online: 20 July 2022

References

- Black, R. E. *et al.* Maternal and child undernutrition and overweight in low-income and middle-income countries. *The Lancet* **382**, 427–451. [https://doi.org/10.1016/S0140-6736\(13\)60937-X](https://doi.org/10.1016/S0140-6736(13)60937-X) (2013).
- Ortiz-Monasterio, J. I. *et al.* Enhancing the mineral and vitamin content of wheat and maize through plant breeding. *J. Cereal Sci.* **46**, 293–307. <https://doi.org/10.1016/j.jcs.2007.06.005> (2007).
- World Health Organization. Worldwide prevalence of anaemia 1993–2005 : WHO global database on anaemia. (ed. de Benoist, B., McLean, E., Egli, V., Cogswell, M.) <https://apps.who.int/iris/handle/10665/43894> (2008).
- Lim, S. S. *et al.* A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990–2010: A systematic analysis for the global burden of disease study 2010. *The Lancet* **380**, 2224–2260. [https://doi.org/10.1016/S0140-6736\(12\)61766-8](https://doi.org/10.1016/S0140-6736(12)61766-8) (2013).
- Lopez, A., Cacoub, P., Macdougall, I. C. & Peyrin-Biroulet, L. Iron deficiency anaemia. *The Lancet* **387**(10021), 907–916. [https://doi.org/10.1016/s0140-6736\(15\)60865-0](https://doi.org/10.1016/s0140-6736(15)60865-0) (2016).
- Holtz, C. & Brown, K. H. Assessment of the risk of zinc deficiency in populations and options for its control. *Food Nutr. Bull.* **25**, 94–204 (2004).
- Wessells, K. R. & Brown, K. H. Estimating the global prevalence of zinc deficiency: results based on zinc availability in national food supplies and the prevalence of stunting. *PLoS ONE* **7**(11), e50568. <https://doi.org/10.1371/journal.pone.0050568> (2012).
- Muller, O. & Krawinkel, M. Malnutrition and health in developing countries. *CMAJ* **173**(3), 279–286. <https://doi.org/10.1503/cmaj.050342> (2005).
- Pfeiffer, W. H. & McClafferty, B. HarvestPlus: Breeding crops for better nutrition. *Crop Sci.* **47**, 88–105. <https://doi.org/10.2135/cropsci2007.09.0020IPBS> (2007).
- Velu, G. *et al.* Performance of biofortified spring wheat genotypes in target environments for grain zinc and iron concentrations. *Field Crops Res.* **137**, 261–267. <https://doi.org/10.1016/j.fcr.2012.07.018> (2012).
- Gopalareddy, K., Singh, A. M., Ahlawat, A. K., Singh, G. P. & Jaiswal, J. P. Genotype-environment interaction for grain iron and zinc concentration in recombinant inbred lines of a bread wheat (*Triticum aestivum* L.) cross. *Indian J. Genet. Plant. Breed.* **75**(3), 307–313 (2015).
- Hernandez-Espinos, N. *et al.* Milling, processing and end-use quality traits of CIMMYT spring bread wheat germplasm under drought and heat stress. *Field Crops Res.* **215**, 104–112. <https://doi.org/10.1016/j.fcr.2017.10.003> (2018).
- Krishnappa, G. *et al.* Multi-environment analysis of grain quality traits in recombinant inbred lines of a biparental cross in bread wheat (*Triticum aestivum* L.). *Cereal Res. Commun.* **1**, 1. <https://doi.org/10.1556/0806.47.2019.002> (2019).
- Shi, R. *et al.* Identification of quantitative trait locus of zinc and phosphorus density in wheat (*Triticum aestivum* L.) grain. *Plant Soil* **306**, 95–104. <https://doi.org/10.1007/s11104-007-9483-2> (2008).
- Peleg, Z. *et al.* Quantitative trait loci conferring grain mineral nutrient concentrations in durum wheat × wild emmer wheat RIL population. *Theor. Appl. Genet.* **119**, 353–369. <https://doi.org/10.1007/s00122-009-1044-z> (2009).
- Tiwari, V. K. *et al.* Mapping of quantitative trait loci for grain iron and zinc concentration in diploid A genome wheat. *J. Hered.* **100**, 771–776. <https://doi.org/10.1093/jhered/esp030> (2009).
- Xu, Y. *et al.* Molecular mapping of QTLs for grain zinc, iron and protein concentration of wheat across two environments. *Field Crops Res.* **38**, 57–62. <https://doi.org/10.1016/j.fcr.2012.09.017> (2012).
- Roshanzamir, H., Kordenaej, A. & Bostani, A. Mapping QTLs related to Zn and Fe concentrations in bread wheat (*Triticum aestivum*) grain using microsatellite markers. *Iran. J. Genet. Plant Breed.* **2**, 551–556 (2013).
- Hao, Y., Velu, G., Roberto, J., Pen Sukhwinder, S. & Ravi, P. S. Genetic loci associated with high grain zinc concentration and pleiotropic effect on kernel weight in wheat (*Triticum aestivum* L.). *Mol. Breed.* **34**, 1893–1902. <https://doi.org/10.1007/s11032-014-0147-7> (2014).
- Srinivasa, J. *et al.* Zinc and iron concentration QTL mapped in a *Triticum spelta* × *T. aestivum* cross. *Theor. Appl. Genet.* **127**, 1643–1651. <https://doi.org/10.1007/s00122-014-2327-6> (2014).
- Pu, Z.-E. *et al.* Quantitative trait loci associated with micronutrient concentrations in two recombinant inbred wheat lines. *J. Integr. Agric.* **13**, 2322–2329. [https://doi.org/10.1016/S2095-3119\(13\)60640-1](https://doi.org/10.1016/S2095-3119(13)60640-1) (2014).
- Tiwari, C. *et al.* Molecular mapping of quantitative trait loci for zinc, iron and protein content in the grains of hexaploid wheat. *Euphytica* **207**, 563–570. <https://doi.org/10.1007/s10681-015-1544-7> (2016).
- Crespo-Herrera, L. A., Velu, G., Stangoulis, J., Hao, Y. & Singh, R. P. QTL mapping of grain Zn and Fe concentrations in two hexaploid wheat RIL populations with ample transgressive segregation. *Front. Plant Sci.* **8**, 1800. <https://doi.org/10.3389/fpls.2017.01800> (2017).
- Krishnappa, G. *et al.* Molecular mapping of the grain iron and zinc concentration, protein content and thousand kernel weight in wheat (*Triticum aestivum* L.). *PLoS ONE* **12**(4), 0174972. <https://doi.org/10.1371/journal.pone.0174972> (2017).

25. Velu, G. *et al.* QTL mapping for grain zinc and iron concentrations and zinc efficiency in a tetraploid and hexaploid wheat mapping populations. *Plant Soil* **411**, 81–99. <https://doi.org/10.1007/s11104-016-3025-8> (2017).
26. Liu, J., Wu, B., Singh, R. P. & Velu, G. QTL mapping for micronutrients concentration and yield component traits in a hexaploid wheat mapping population. *J. Cereal Sci.* **88**, 57–64. <https://doi.org/10.1016/j.jcs.2019.05.008> (2019).
27. Krishnappa, G. *et al.* Identification of novel genomic regions for biofortification traits using an snp marker-enriched linkage map in wheat (*Triticum aestivum* L.). *Front. Nutr.* **8**, 669444. <https://doi.org/10.3389/fnut.2021.669444> (2021).
28. Rathan, N. D. *et al.* Identification of genetic loci and candidate genes related to grain zinc and iron concentration using a zinc-enriched wheat ‘Zinc-Shakti’. *Front. Genet.* **12**, 652653. <https://doi.org/10.3389/fgene.2021.652653> (2021).
29. Goel, S. *et al.* Analysis of genetic control and QTL mapping of essential wheat grain quality traits in a recombinant inbred population. *PLoS ONE* **14**, e0200669. <https://doi.org/10.1371/journal.pone.0200669> (2019).
30. Nezhad, K. Z. *et al.* QTL analysis for thousand-grain weight under terminal drought stress in bread wheat (*Triticum aestivum* L.). *Euphytica* **186**, 127–138. <https://doi.org/10.1007/s10681-011-0559-y> (2012).
31. Mergoum, M. *et al.* Agronomic and quality QTL mapping in spring wheat. *J. Plant Breed Genet.* **1**, 19–33 (2013).
32. Wei, L. *et al.* QTL positioning of thousand wheat grain weight in qaidam basin. *Open J. Genet.* **4**, 239–244. <https://doi.org/10.4236/ojgen.2014.43024> (2014).
33. Zhang, H. *et al.* Conditional QTL mapping of three yield components in common wheat (*Triticum aestivum* L.). *Crop J.* **4**, 220–228. <https://doi.org/10.1016/j.cj.2016.01.007> (2016).
34. Flintgarcia, S. A., Thornsberry, J. M., And, E. S. & Buckler, I. V. Structure of linkage disequilibrium in plants. *Annu. Rev. Plant Biol.* **54**, 357–374. <https://doi.org/10.1146/annurev.arplant.54.031902.134907> (2003).
35. Zondervan, K. T. & Cardon, L. R. The complex interplay among factors that influence allelic association. *Nat. Rev. Genet.* **5**, 89–100. <https://doi.org/10.1038/nrg1270> (2004).
36. Korte, A. & Farlow, A. The advantages and limitations of trait analysis with GWAS: A review. *Plant Methods* **9**, 29–38. <https://doi.org/10.1186/1746-4811-9-2> (2013).
37. Pang, Y. *et al.* High-resolution genome-wide association study identifies genomic regions and candidate genes for important agronomic traits in wheat. *Mol. Plant.* **13**, 1311–1327 (2020).
38. Zimin, A. V. *et al.* (2017) The first near-complete assembly of the hexaploid bread wheat genome, *Triticum aestivum*. *Gigascience* **6**, 1–7. <https://doi.org/10.1093/gigascience/gix097> (2017).
39. Alomari, D. Z. *et al.* Identifying Candidate Genes for Enhancing Grain Zn Concentration in Wheat. *Front. Plant Sci.* **9**, 1313. <https://doi.org/10.3389/fpls.2018.01313> (2018).
40. Bhatta, M. *et al.* Genome-wide association study reveals novel genomic regions associated with 10 grain minerals in synthetic hexaploid wheat. *Int. J. Mol. Sci.* **19**, 3237. <https://doi.org/10.3390/ijms19103237> (2018).
41. Velu, G. *et al.* Genetic dissection of grain zinc concentration in spring wheat for mainstreaming biofortification in CIMMYT wheat breeding. *Sci. Rep.* **8**, 13526. <https://doi.org/10.1038/s41598-018-31951-z> (2018).
42. Liu, J. *et al.* Genome-wide association study for grain micronutrient concentrations in wheat advanced lines derived from wild Emmer. *Front. Plant Sci.* **12**, 651283. <https://doi.org/10.3389/fpls.2021.651283> (2021).
43. Cu, S. T. *et al.* Genetic dissection of zinc, iron, copper, manganese and phosphorus in wheat (*Triticum aestivum* L.) grain and rachis at two developmental stages. *Plant Sci.* **291**, 1138. <https://doi.org/10.1016/j.plantsci.2019.110338> (2020).
44. Liu, Y. *et al.* A thorough screening based on QTLs controlling zinc and copper accumulation in the grain of different wheat genotypes. *Environ. Scib. Pollut. Res. Int.* **28**(12), 15043–15054. <https://doi.org/10.1007/s11356-020-11690-3> (2020).
45. Calderini, D. F. & Ortiz-Monasterio, I. Are synthetic hexaploids a means of increasing grain element concentrations in wheat. *Euphytica* **134**, 169–178. <https://doi.org/10.1023/B:EUPH.0000003849.10595.ac> (2003).
46. Arora, S., Cheema, J., Poland, J., Uauy, C. & Chhuneja, P. Genome-wide association mapping of grain micronutrients concentration in *Aegilops tauschii*. *Front. Plant Sci.* **10**, 54. <https://doi.org/10.3389/fpls.2019.00054> (2019).
47. Zhou, Z. *et al.* Identification of novel genomic regions and superior alleles associated with Zn accumulation in wheat using a genome-wide association analysis method. *Int. J. Mol. Sci.* **21**, 1928. <https://doi.org/10.3390/ijms21061928> (2020).
48. Godoy, J. *et al.* Genome-wide association study of agronomic traits in a spring-planted North American elite hard red spring wheat panel. *Crop Sci.* **58**, 1838–1852. <https://doi.org/10.2135/cropsci2017.07.0423> (2018).
49. Rahimi, Y., Bihamta, M. R., Taleei, A., Alipour, H. & Ingvarsson, P. K. Genome-wide association study of agronomic traits in bread wheat reveals novel putative alleles for future breeding programs. *BMC Plant. Biol.* **19**, 541. <https://doi.org/10.1186/s12870-019-2165-4> (2019).
50. Ward, B. P. *et al.* Genome-wide association studies for yield-related traits in soft red winter wheat grown in Virginia. *PLoS ONE* **14**(2), e0208217. <https://doi.org/10.1371/journal.pone.0208217> (2019).
51. Gahlaut, V., Jaiswal, V., Balyan, H. S., Joshi, A. K. & Gupta, P. K. Multi-locus GWAS for grain weight-related traits under rain-fed conditions in common wheat (*Triticum aestivum* L.). *Front. Plant Sci.* **12**, 7531. <https://doi.org/10.3389/fpls.2021.758631> (2021).
52. Aravind, J., MukeshSankar, S., Wankhede, D. P. & Kaur, V. AugmentedRCBD: Analysis of Augmented Randomised Complete Block Designs. R package version 0.1.5.9000, <https://aravind-j.github.io/augmentedRCBD/><https://cran.rproject.org/package=augmentedRCBD>(2021).
53. Murray, M. G. & Thompson, W. F. Rapid isolation of high molecular weight plant DNA. *Nucl. Acids Res.* **8**(19), 4321–4325 (1980).
54. Bradbury, P. J. *et al.* TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinfo* **23**, 2633–2635. <https://doi.org/10.1093/bioinformatics/btm308> (2007).
55. Lipka, A. E. *et al.* GAPIT: genome association and prediction integrated tool. *Bioinfo* **28**(18), 2397–2399. <https://doi.org/10.1093/bioinformatics/bts444> (2012).
56. Earl, D. A. & von Holdt, B. M. STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conserv. Genet. Resour* **4**, 359–361. <https://doi.org/10.1007/s12686-011-9548-7> (2012).
57. Evanno, G., Regnaut, S. & Goudet, J. Detecting the number of clusters of individuals using the software structure: A simulation study. *Mol. Ecol.* **14**, 2611–2620. <https://doi.org/10.1111/j.1365-294X.2005.02553.x> (2005).
58. Huang, M., Liu, X., Zhou, Y., Summers, R. M. & Zhang, Z. "BLINK: A package for the next level of genome-wide association studies with both individuals and markers in the millions. *Gigascience* **8**(2), 154. <https://doi.org/10.1093/gigascience/giy154> (2019).
59. Wang, M., Gong, J. & Bhullar, N. K. Iron deficiency triggered transcriptome changes in bread wheat. *Comput. Struct. Biotechnol. J.* **18**, 2709–2722. <https://doi.org/10.1016/j.csbj.2020.09.009> (2020).
60. Yokosho, K., Yamaji, N., Ueno, D., Mitani, N. & Ma, J. F. OsFRDL1 Is a citrate transporter required for efficient translocation of iron in rice. *Plant Physiol.* **149**, 297–305. <https://doi.org/10.1104/pp.108.128132> (2009).
61. Seo, P. J. *et al.* A Golgi-localized MATE transporter mediates iron homeostasis under osmotic stress in Arabidopsis. *Biochem. J.* **442**, 551–561. <https://doi.org/10.1042/BJ20111311> (2012).
62. Pineau, C. *et al.* Natural variation at the FRD3 MATE transporter locus reveals cross-talk between Fe homeostasis and Zn tolerance in Arabidopsis thaliana. *PLoS Genet.* **8**(12), e1003120. <https://doi.org/10.1371/journal.pgen.1003120> (2012).
63. Przybyla-Toscano, J., Boussardon, C., Law, S. R., Rouhier, N. & Keech, O. Gene atlas of iron-containing proteins in Arabidopsis thaliana. *Plant J.* **106**, 258–274. <https://doi.org/10.1111/tpj.15154> (2021).
64. Finazzi, G. *et al.* Ions channels/transporters and chloroplast regulation. *Cell Calcium* **58**, 86–97. <https://doi.org/10.1016/j.ceca.2014.10.002> (2015).

65. Wang, L. *et al.* The similar and different evolutionary trends of MATE family occurred between rice and *Arabidopsis thaliana*. *BMC Plant Biol.* **16**, 207. <https://doi.org/10.1186/s12870-016-0895-0> (2016).
66. Mora-Macias, J. *et al.* Malate-dependent Fe accumulation is a critical checkpoint in the root developmental response to low phosphate. *Proc. Natl. Acad. Sci. USA* **114**, E3563–E3572 (2017).
67. Balzergue, C. *et al.* Low phosphate activates STOP1-ALMT1 to rapidly inhibit root cell elongation. *Nat. Commun.* **8**, 15300 (2017).
68. Kruger, C., Berkowitz, O., Stephan, U. W. & Hell, R. A metal-binding member of the late embryogenesis abundant protein family transports iron in the phloem of *Ricinus communis* L. *J. Biol. Chem.* **277**(28), 25062–25069. <https://doi.org/10.1074/jbc.M201896200> (2002).
69. Magwanga, R. O. *et al.* Characterization of the late embryogenesis abundant (LEA) proteins family and their role in drought stress tolerance in upland cotton. *BMC Genet.* **19**, 6. <https://doi.org/10.1186/s12863-017-0596-1> (2018).
70. Wang, Q. *et al.* Dissecting the genetic basis of grain size and weight in barley (*Hordeum vulgare* L.) by QTL and Comparative Genetic Analyses. *Front. Plant Sci.* **10**, 469. <https://doi.org/10.3389/fpls.2019.00469> (2019).
71. Jang, S. & Li, H. Y. *Oryza sativa* BRASSINOSTEROIDUPREGULATED1 LIKE1 induces the expression of a gene encoding a small leucine-rich-repeat protein to positively regulate lamina inclination and grain size in rice. *Front. Plant Sci.* **8**, 1253. <https://doi.org/10.3389/fpls.2017.01253> (2017).
72. Giancaspro, A., Giove, S. L., Zacheo, S. A., Blanco, A. & Gadaleta, A. Genetic variation for protein content and yield-related traits in a durum population derived from an inter-specific cross between hexaploid and tetraploid wheat cultivars. *Front. Plant Sci.* **10**, 1509. <https://doi.org/10.3389/fpls.2019.01509> (2019).
73. Gupta, P. K., Rustgi, S. & Kulwal, P. L. Linkage disequilibrium and association studies in higher plants: present status and future prospects. *Plant Mol. Biol.* **57**(4), 461–485. <https://doi.org/10.1007/s11103-005-0257-z> (2005).
74. Vos, P. J. *et al.* Evaluation of LD decay and various ld-decay estimators in simulated and snp-array data of tetraploid potato. *Theor. Appl. Genet.* **130**(1), 123–135. <https://doi.org/10.1007/s00122-016-2798-8> (2017).
75. Yu, H., Deng, Z., Xiang, C. & Tian, J. Analysis of diversity and linkage disequilibrium mapping of agronomic traits on B-genome of wheat. *J. Genom.* **2**, 20–30. <https://doi.org/10.7150/jgen.4089> (2014).
76. Dinesh, A. *et al.* Genetic diversity, linkage disequilibrium and population structure among CIMMYT maize inbred lines, selected for heat tolerance study. *Maydica* **61**(3), 1–7 (2016).
77. Dadshani, S., Mathew, B., Ballvora, A., Mason, A. S. & Leon, J. Detection of breeding signatures in wheat using a linkage disequilibrium-corrected mapping approach. *Sci. Rep.* **11**(1), 5527. <https://doi.org/10.1038/s41598-021-85226-1> (2021).
78. Sukumaran, S., Dreisigacker, S., Lopes, M., Chavez, P. & Reynolds, M. P. Genome-wide association study for grain yield and related traits in an elite spring wheat population grown in temperate irrigated environments. *Theor. Appl. Genet.* **128**(2), 353–363. <https://doi.org/10.1007/s00122-014-2435-3> (2015).
79. Sheoran, S. *et al.* Uncovering genomic regions associated with 36 agro-morphological traits in Indian spring wheat using GWAS. *Front. Plant Sci.* **10**, 527. <https://doi.org/10.3389/fpls.2019.00527> (2019).
80. Wang, J., Zhang, Z. (2021). GAPIT Version 3: Boosting power and accuracy for genomic association and prediction. *Genom. Proteom. Bioinf.*
81. Rogers, E. E. & Guerinot, M. L. FRD3, a member of the multidrug and toxin efflux family, controls iron deficiency responses in *Arabidopsis*. *Plant Cell* **14**, 1787–1799. <https://doi.org/10.1105/tpc.001495> (2002).
82. Roschzttardtz, H., Seguela-Arnaud, M., Briat, J.-F., Vert, G. & Curie, C. The FRD3 citrate effluxer promotes iron nutrition between symplastically disconnected tissues throughout *Arabidopsis* Development. *Plant Cell* **23**, 2725–2737. <https://doi.org/10.1105/tpc.111.088088> (2011).
83. Finatto, T. *et al.* Abiotic stress and genome dynamics: Specific genes and transposable elements response to iron excess in rice. *Rice* **8**, 13. <https://doi.org/10.1186/s12284-015-0045-6> (2015).
84. Durrett, T. P., Gassmann, W. & Rogers, E. E. The FRD3-mediated efflux of citrate into the root vasculature is necessary for efficient iron translocation. *Plant Physiol.* **144**, 197–205. <https://doi.org/10.1104/pp.107.097162> (2007).
85. Wang, J. *et al.* Diverse functions of multidrug and toxin extrusion (MATE) transporters in citric acid efflux and metal homeostasis in *Medicago truncatula*. *Plant J.* **90**, 79–95. <https://doi.org/10.1111/tpj.13471> (2017).
86. Rogers, E. E., Wud, X., Stacey, G. & Nguyen, H. T. Two MATE proteins play a role in iron deficiency in soybean. *J. Plant Physiol.* **166**, 1453–1459. <https://doi.org/10.1016/j.jplph.2009.02.009> (2009).
87. Bommert, P., Nagasawa, N. S. & Jackson, D. Quantitative variation in maize kernel row number is controlled by the FASCIATED EAR2 locus. *Nat. Genet.* **45**(3), 1. <https://doi.org/10.1038/ng.2534> (2013).
88. Wang, A. *et al.* The VQ motif protein IKU1 regulates endosperm growth and seed size in *Arabidopsis*. *Plant J.* **63**, 670–679. <https://doi.org/10.1111/j.1365-313X.2010.04271.x> (2010).
89. Li, J. *et al.* In silico functional prediction and expression analysis of C2H2 Zinc-finger family transcription factor revealed regulatory role of ZmZFP126 in Maize growth. *Front. Genet.* **12**, 770427. <https://doi.org/10.3389/fgene.2021.770427> (2021).
90. Pritchard, J. K., Wen, X. & Falush, D. Documentation for structure software: Version 2.3. *University of Chicago, Chicago, IL*. <http://pritch.bsd.uchicago.edu/structure.html> (2010).
91. Rathan, N. D. *et al.* Genome-wide association study identifies loci and candidate genes for grain micronutrients and quality traits in wheat (*Triticum aestivum* L.). *Sci Rep.* **12**, 7037. <https://doi.org/10.1038/s41598-022-10618-w> (2022).

Acknowledgements

The authors acknowledge the work and the assistance of all the cooperators who have evaluated the trial and provided valuable data. This study was supported with funding provided by the Indian Council of Agricultural Research (ICAR) and Bill & Melinda Gate foundation (BMGF) under the project ICAR-BMGF (Project Code:1011099).

Author contributions

H.K., G.P.S., G.K. conceptualized the investigation and drafted the original manuscript; H.K., G.K., H.Kr., S.K., C.N.M., O.P., P.S., S.B., U.G., M.K. conducted the field experiments; H.Kr., H.K. generation of genotypic data; N.B.D., T.N., M.H.M., G.S., N.D.R. did the phenotypic and GWAS. analysis. All authors contributed to the article editing and approved the submitted version.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-022-15992-z>.

Correspondence and requests for materials should be addressed to H.K. or G.P.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022