Genome-Wide Association Study Reveals the Genetic Architecture for Calcium Accumulation in Grains of Hexaploid Wheat (*Triticum Aestivum L.*)

Xia Shi
Henan Academy of Agricultural Sciences
Zhengfu Zhou
Henan Academy of Agricultural Sciences
Wenxu Li
Henan Academy of Agricultural Sciences
Maomao Qin
Henan Academy of Agricultural Sciences
Pan Yang
Henan Academy of Agricultural Sciences
Jinna Hou
Henan Academy of Agricultural Sciences
Fangfang Huang
Henan Agricultural University
Zhensheng Lei (zhenshenglei@126.com)
Henan Academy of Agricultural Sciences
Zhengqing Wu
Henan Academy of Agricultural Sciences
Jiansheng Wang
Pingdingshan University

Research Article

Keywords: Hexaploid wheat (Triticum aestivum L.), calcium accumulation, genetic loci, superior and inferior alleles, pyramid breeding, genome-wide association analysis

Posted Date: November 30th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-1104188/v1

Abstract

Background: Hexaploid wheat (*Triticum aestivum L.*) is a leading cereal crop worldwide. Understanding the mechanism of calcium (Ca) accumulation in wheat is important to reduce the risk of human micronutrient deficiencies. However, the mechanisms of Ca accumulation in wheat grain are only partly understood.

Results: Here, we performed a genome-wide association study to identify the genetic basis of Ca accumulation in wheat grain using an association population consisting of 207 varieties, with phenotypic data from three locations and the combined locations. In total, 18 non-redundant quantitative trait loci (QTLs) associated with Ca concentration were identified that explained, on average, 9.61%–26.93% of the phenotypic variation. Cultivars containing more superior alleles and fewer inferior alleles had increased grain Ca concentrations. Notably, six non-redundant loci were identified in at least two environments, indicating their stability across different environments. Searches of public databases revealed six putative candidate genes linked to Ca accumulation. Among them, two subunits of *V-type Proton ATPase* (*TraesCS4A01G428900* and *TraesCS3B01G241000*) are encoded by genes associated with stable genetic loci on chromosomes 4A (AX-108912427) and 3B (AX-110922471), respectively, and they are typical generators of a proton gradient that might be involved in Ca homeostasis in wheat grain.

Conclusion: This study could increases our understanding of the genetic architecture of grain Ca accumulation in wheat, and we plan to develop the identified superior alleles into molecular markers for wheat Ca biofortification pyramid breeding in the future.

Background

Hexaploid wheat (*Triticum aestivum L.*) is one of the world's three major food crops, and its products are main sources of human dietary nutrients [1]. In the past 50 years, with the continuous improvement of breeding and cultivation technology, the yields of major food crops have increased greatly, helping to solve the problem of food security [2]. However, past breeding targets have focused too much on increasing yields and protein contents while ignoring the development of nutritional quality, including that supplied by microelements [3]. However, with changes in consumers' dietary structure and the continuous improvement in living standards, the demand for nutritional high-quality wheat is growing.

Ca is the fifth most required microelement, and its total content forms approximately 2% of the body's weight. It plays important roles in maintaining the body's normal physiological and biochemical functions, making it an important nutrient for human health [4]. Some basic regulatory functions in the human body involve Ca, including hormone secretion, blood clotting, enzyme reaction activation, vascular diastolic, muscle function, nerve impulse delivery, cell proliferation and intracellular metabolism [5]. The appropriate amount of Ca intake has protective effects on colorectal, ovarian, breast and other types of cancer, and it can effectively reduce the risk of cardiovascular disease [6, 7]. Moreover, Ca is also necessary for plant growth and development, playing important roles in plant cell structures and

physiological functions. It is involved in maintaining the stability of cell walls, cell membranes and membrane-binding proteins, and it participates in regulating intracellular homeostasis and plant development [8, 9]. As a second messenger in plant cell signaling, Ca has a sensitive response to various stimuli, such as environmental and endogenous hormone signals [9]. It participates in signal transduction processes that rely on concentration gradient changes to transmit signals [10]. However, depending on the downstream functional protein (such as protein kinases/protein kinase cascade) amplification signals, it also participates in biotic and abiotic stress responses, hormone regulation and other physiological and biochemical functions in plants [11, 12]. Therefore, both plants and humans require optimum intakes of Ca for their normal physiological and biochemical activities.

Quantitative trait loci (QTL) associated with Ca accumulation have been identified in different plant species and crops, such as wheat, rice, sorghum, barley, corn, pearl millet and beans [13–16]. Goel et al. (2011) identified 31 QTLs in rice that regulate Ca accumulation and 28 QTLs in sorghum that affect Ca accumulation [17]. Five QTLs identified in Arabidopsis thaliana account for 36.4% of the Ca content variation [18]. However, QTL identification strategies using bi-parental populations have low resolutions and only relevant sites with residual variation at the parental level can be obtained. In contrast, a genome-wide association analysis (GWAS) relies on more representative and diverse natural population genome information to detect non-random associations between genotypes and phenotypes, greatly improving the resolution of the QTL mapping [19]. This method has been widely used in the QTL mapping of complex quantitative traits in multiple species. However, there are only a few studies in which GWAS has been used to identify QTLs associated with grain Ca accumulation in wheat. Alomari et al. (2017) used a natural population containing 353 wheat varieties to identify the major QTLs associated with the grain Ca accumulation efficiency on chromosomes 2A, 5B and 6A [20]. Bhatta et al. (2018) identified QTLs associated with Ca accumulation in grains in 14 different genomic regions on chromosomes 1B, 2B, 2D, 3A, 3B, 3D, 6A, 6B and 7A, which explained 2.7–21.5% of the phenotypic variation [21].

Previous studies have studied the accumulations of microelements in wheat grains using association and linkage analyses. However, there are limited reports on Ca accumulation in wheat grains; consequently, the genetic basis of Ca accumulation in wheat grains remains unclear. In this study, a natural population was employed to identify QTLs for wheat grain Ca accumulation using a GWAS. This study aimed to 1) dissect the genetic architecture of wheat grain Ca accumulation; 2) identify potential candidate genes associated with Ca accumulation; 3) evaluate the genetic effects of stable non-redundant QTLs; and 4) explore molecular markers that can be used to guide Ca biofortification breeding. This study provides useful information for further elucidating the molecular and genetic mechanisms of Ca accumulation in common wheat.

Results

Genetic diversity analysis

Before the GWAS, the wheat 660K SNP assay was used to perform genotyping of the natural population, which included 207 lines. After filtering using the criteria of minor allele frequency > 0.05 and missing data < 10%, 244,508 single-nucleotide polymorphism (SNP) markers were retained for further analyses (Figure S1). The population structure results were integrated using Admixture software (www.genetics.ucla.edu/software/admixture). As shown in Figure 1, at K = 8, the cross-validation error (CV) was the lowest, indicating that this was the optimal K value; this demonstrated that the natural population has a rich genetic background and could be divided into eight sub-populations. Thus, this population was suitable for a further GWAS of wheat grain Ca accumulations.

Phenotypic variations in Ca concentrations

The Ca levels were investigated for the whole natural wheat population across field studies in Yuanyang (YY), Shangqiu (SQ) and Kaifeng (KF) in 2017 (Table S1). In each environment, the Ca concentration showed a broad range of variation (Figure 2). The highest mean Ca concentration (376.02 mg/kg) was recorded in SQ, where values ranged from 139.18 to 676.04 mg/kg. The second highest mean Ca concentration was 319.05 mg/kg recorded in KF, where values ranged from 121.71 to 552.19 mg/kg. A lower mean Ca concentration value of 297.37 mg/kg was recorded in YY, where the values ranged from 139.18 to 676.04 mg/kg. The resulting of the best linear unbiased predictors (BLUPs) for grain Ca concentrations across all environments ranged from 320.96 to 348.08 mg/kg, with a mean of 330.81 mg/kg (Table 1). Additionally, the frequencies of the Ca concentrations in individual and combined environments exhibited an approximately normal distribution (Figure S2). Thus, this is an ideal population for the association mapping of wheat grain Ca concentrations.

GWAS of wheat grain Ca concentrations

Three models, general linear models (GLM), mixed linear models (MLM) and fixed and random model Circulating Probability Unification (FarmCPU), were used to perform the GWAS for the Ca concentration of wheat grains at the three locations and a combination of the three locations. By comparing the results of the quantile–quantile (QQ) plots of different analysis models, it was determined that the GLM model controlled the population structure with a high risk of type I error (false positive). The FarmCPU model's screening criteria were too strict for significant association sites and resulted in a higher type II error (false negative). The MLM model not only controlled the population structure, but also controlled the relative kinship in the analysis process. The QQ plot indicated that the MLM model was more reliable than the others, reducing the probabilities of type I and II errors (Figure S3). We determined it to be an ideal model for the GWAS analysis of wheat grain Ca concentrations.

Using the MLM model, 111 significant SNPs at the suggested $p < 1.0 \times 10^{-4}$ for wheat grain Ca concentrations were identified in the surveyed environments (including a combination of the three locations) (Table S2). On the basis of the definition of the QTLs, combined with the physical locations of these significant SNPs, the SNPs were categorized into 18 non-redundant QTLs. Briefly, in YY, 14 non-redundant QTLs involving 37 significant SNPs associated with Ca concentration in wheat grains were

identified. They are mainly distributed on chromosomes 1D, 2A, 2B, 3A, 3B, 3D, 4A, 5A, 5B, 6A, 7A and 7B, and each locus could explain phenotypic variation (PVE) ranging from 13.75–26.93%, with a mean of 16.34%. In SQ, 6 non-redundant QTLs involving 10 significant SNPs associated with wheat grain Ca concentrations were identified. They were located on chromosomes 3A, 3D, 4A, 4B and 5B, and the PVE of each locus ranged from 9.33–12.89%, with a mean of 10.29%. Similarly, in KF, two non-redundant QTLs (26 significant SNPs) were identified on chromosomes 2A and 4A, each explaining PVE from 7.32–16.55%, with a mean of 14.36%. Moreover, 5 non-redundant QTLs involving 46 significant SNPs were identified in the combination of the three individual locations. They were mainly distributed on chromosomes 3A, 3B, 3D and 4A, and the PVE ranged from 16.50–19.55%, with a mean of 17.04% (Figure 3, Table 2).

The repeatable SNPs involved in the co-localization of non-redundant QTLs identified across different locations are summarized in Table 2. Overall, 6 non-redundant QTLs involving 6 peak SNPs were detected in at least two environments (including the combined location, BLUP) (Table 2), and they might be stable across different environments. Additionally, the SNP AX-108912427 was identified simultaneously in all three locations and the combination of the three locations. It is located on chromosome 4A and exhibited the highest PVE in each environment, ranging from 12.89–26.93% (Table 2). Furthermore, a QTL hot spot exists at the end of chromosome 4A. This QTL interval could be a key factor in regulating wheat grain Ca accumulation. Detailed analyses of the candidate genes underlying these six non-redundant QTLs will provide useful information concerning Ca accumulation.

Putative candidate gene predictions for QTLs controlling Ca accumulation

On the basis of the GWAS, six stable non-redundant QTLs associated with the regulation of Ca accumulation in wheat grain were identified in this study. Prior to the candidate gene prediction, we compared the QTLs with correlated research on Ca accumulation in wheat and found that these loci co-localize (Table 3). Using the wheat reference genome annotation information from IWGSC RefSeq v2.0, six promising genes were located within these loci, and they were considered as candidate genes related to Ca accumulation in wheat grain. A bioinformatics analysis detected the most significant SNP, AX-108912427, simultaneously in all environments. It is located on chromosome 4A, which harbors the gene *TraesCS4A01G428900* that encodes a *V-type proton ATPase subunit e*, which is relevant to Ca homeostasis and transporters. The SNP AX-94729264, located on the 3D chromosome, was detected in three different locations, and it is associated with the predicted gene *TraesCS3D01G079600* that encodes an *ubiquitin family protein*, which might be involved in Ca homeostasis. On chromosome 3B, the significant SNP AX-110013515 was significantly associated with Ca accumulations in two environments. The gene underlying this marker encodes a *transmembrane protein* that has ion channel activity and is likely involved in Ca transport, whereas the second significant locus linked with SNP AX-110922471 on the same chromosome carries the gene *TraesCS3B01G241000* that encodes *V-type proton ATPase subunit d*, which may be related to Ca accumulation. Additionally, two SNPs, AX-110634514 and AX-109541359, located on chromosomes 2A and 3A, respectively, were linked with the predicted genes

TraesCS2A01G098100 and *TraesCS3A01G424700*, respectively. The first gene encodes a *calmodulin*, and the second gene encodes a *cytochrome P450*, which may be relevant to Ca regulation in grain. Detailed information on the most likely candidate genes are list in Table 3.

Relationship between Ca accumulation and the number of superior alleles

All these SNPs were associated with grain Ca concentrations in at least two environments, and they may be considered as stable loci for controlling Ca accumulations in wheat grains. The repeatable SNP polymorphic effects were investigated using the ANOVA method based on the phenotypic values of the natural population. For each of the six SNPs, cultivars with the superior alleles showed higher Ca concentration in grains than cultivars with the inferior alleles (Figure 4). Statistical analyses showed that the grain Ca concentration differences, accessions with superior alleles compared with the accessions with inferior alleles, have reached significant or highly significant levels in all or some environments (Table 4).

To further understand the pyramiding effects of alleles on the grain Ca concentration, we investigated the numbers of superior and inferior alleles in each cultivar. The number of superior alleles ranged from 0 to 13, compared with 3 to 17 inferior alleles. The Ca concentration-related BLUP value of the natural population was used to examine the relationship between Ca accumulation and the numbers of superior and inferior alleles by linear regression. A linear relationship between grain Ca concentration and the number of superior alleles per genotype was observed. The regression coefficients between the Ca concentration and numbers of superior alleles and inferior alleles were 0.2242, and 0.2444, respectively, implying that cultivars with more superior alleles and fewer inferior alleles contributed to increasing grain Ca concentrations (Figure 5). Thus, pyramiding more superior alleles and fewer inferior alleles should enhance the wheat grain Ca concentration, and this strategy can be used in crop genetic biofortification breeding programs.

Discussion

Genetic biofortification is an effective method for enhancing crop microelement contents. A large number of genetic loci controlling microelement accumulation in wheat, rice and other crops have been identified in previous research [22–25]. However, there are relatively few studies on the genetic mechanisms of Ca ion accumulation in wheat grain; consequently, limited information is available on wheat grain genetic control and molecular physiological mechanisms. GWAS is a powerful tool for dissecting the genetics of complex traits and identifying the chromosomal regions harboring genes suitable for use in breeding programs. In this study, a GWAS was used to dissect the genetic basis of Ca accumulation in wheat grain using a natural population.

When using a GWAS, the probability of detecting the causal variant and associated loci for a target trait depends on the marker density, population size and statistical methods [26]. Owing to the rapid development and application of molecular marker assays, Wheat 35K, 90K, 660K and 820K SNP

genotyping arrays have been designed and utilized for GWAS and linkage analyses in common wheat [27–29]. Comparative analyses revealed that the Wheat 660K SNP array is reliable and cost-effective, making it the ideal choice for genotyping a population of individuals [30]. In the present study, a credible number of markers (244,508 SNPs) was identified using the Wheat 660K SNP array, and the population met the requirements of a GWAS for Ca accumulation in wheat grains. Population size is another factor limiting the detection efficiency of a GWAS. The effect of increasing the population size on QTL detection efficiency is greater than that of the marker density [31], and increasing the population size may lead to the identification of more smaller-effect QTLs [32]. Previously, population sizes ranging from 100 to 500 have been used for wheat association analyses [33–35]. In this study, although the natural population of 207 diverse accessions was not sufficiently large, the dramatic phenotypic variations in grain Ca concentration was very large, ranging from 121.71 mg/kg to 685.17 mg/kg, and it showed a normal distribution, which is conducive to GWAS. Dramatic phenotype variation may be associated with high genetic diversity [36]. In this study, a population structure analysis based on the genotype data of the natural population interpreted the diversity from another perspective. Our population could be divided into eight subpopulations (Figure 1), which indicated high genetic diversity and suitability for GWAS. The lack of proper modifications to population structure may lead to spurious associations [37, 38]. To eliminate such associations, the population structure, as shown in the Q matrix, was considered as a fixed-effect factor in the GWAS. Additionally, complex traits are sensitive to different statistical models, which vary in their abilities to control type I or type II errors. Thus, a suitable statistical model can effectively increase the confidence of the associated sites. For example, Chen et al. (2017) analyzed 13 agronomic traits of wheat and found that the MLM model effectively controls false positive and false negative errors, making it the best model for a GWAS [39]. In a maize GWAS for arsenic accumulation, the GLM model (only controls the influence on the population structure) was better than the MLM model [40], although it could not completely control the population structure influence. In the present study, the QQ plots indicated that the influence of type I and type II errors was controlled in the GWAS for wheat grain Ca accumulation by MLM [simultaneously controls the influence of fixed-effect factors (Q matrix) and random-effect factors (K matrix)] better than the GLM and FarmCUP statistical models. These results were in accordance with previous GWAS studies for Ca accumulation in wheat grain, in which the MLM model was used for the GWAS [20].

The manifestations of complex quantitative traits (such as microelement accumulation in wheat grain) are often controlled by multiple genetic loci [21, 41, 42]. Several genetic loci affecting Ca accumulation in wheat grain were identified previously by GWAS and linkage analyses [13, 20, 21], which allowed for a comparison among loci identified in known QTLs and those identified in the present study. Here, 18 non-redundant loci for Ca accumulation were distributed on chromosomes 1D, 2A, 2B, 3A, 3B, 3D, 4A, 4B, 5A, 5B, 6A, 7A and 7B, which suggested the involvement of loci on these chromosomes in the natural population regulating Ca concentration variation. QTLs for Ca accumulation are scattered on chromosomes 4A, 4B, 5B, 6A and 7B [13] in a *durum wheat* × *wild emmer* RIL population, which overlapped with loci identified in the current study, indicating that the linkage mapping results were complementary to those of the GWAS for wheat grain Ca accumulation. Bhatta et al. (2018) identified 15

significant marker-trait associations for wheat grain Ca accumulation distributed in 14 different genomic regions on chromosomes 1B, 2B, 2D, 3A, 3B, 3D, 6A, 6B and 7A that explained 2.7–21.5% phenotypic variation using the Hard Winter Wheat Association Mapping Panel (including 299 varieties) and the GWAS method [21]. We only identified eight loci on chromosomes 2B, 3A (2 loci), 3B (2 loci), 3D, 6A and 7A, and not all of the QTLs detected previously were identified in this study. This may be because of (a) the different origins of the populations, or (b) the use of different genotypic identification platforms. This can make it difficult to align the complete genomes of the population's individuals. It is worth noting that, like other complex quantitative traits, the accumulation of Ca in wheat grain is controlled by multiple genetic sites and is susceptible to environmental influences [43]. Therefore, the ideal target genetic loci should be stably identified under multi-environmental conditions. In this study, we found that all six loci were detected in at least two environments with relatively higher PVE values (9.66–26.93%), which suggests that these were stable QTLs significantly associated with wheat grain Ca accumulation that were critical for target trait phenotypic variation. The peak SNP AX-110634514 has been mapped closely to *Kukri_c40035_258* on chromosome 2A [20], and AX-109541359 is co-located in the vicinity of regions harboring *S3A_593702925* and *CFA2262-wPt_3816* on chromosome 3A [21, 44]. On chromosome 3B, the SNPs AX-110013515 and AX-110922471 were mapped to genomic regions near *gwm389-wPt-8093(C)* and *QGCaUE-3B*, respectively. The former QTL has been identified in a natural wheat population [44], and the latter QTL has been detected in a double-haploid wheat population [42]. The peak SNP AX-94729264 was simultaneously identified in SQ, YY, and BLUP, which indicated it co-localizes with *S3D_45073985* [21]. These findings validate the results of the GWAS and increase the confidence in some loci identified in the present study. The hotspot at the end of chromosome 4A linked with SNP AX-108912427 (identified in all the environments with the highest PVE values, ranging from 12.89–26.93%) was simultaneously mapped in the vicinities of three QTLs, *gwm165a-wmc420*, *wmc106-gwm165a* and *gwm610* [13, 44], which implies that this hotspot is a key factor that harbors a major gene for regulating Ca accumulation in wheat grain.

In this study, six peak SNPs, together with corresponding non-redundant QTLs that are associated Ca accumulation, were identified by the GWAS in at least two environments. Combined with the gene's physical positions and functional annotation information, six genes were identified as the most credible candidate genes for Ca accumulation. Two SNP markers, AX-108912427 on chromosome 4A and AX-110922471 on chromosome 3B, were associated with genes encoding *V-type proton ATPase subunit e* (*TraesCS4A01G428900*) and *V-type proton ATPase subunit d* (*TraesCS3B01G241000*), respectively. Both subunits are important components of the V-type proton ATPase that is the typical generator of a proton gradient involved in Ca ion sequestration in plant cells, and it may influence Ca ion homeostasis in wheat grains [45, 46]. The hot spot QTL linked with 39 significant SNPs at the end of the 4A chromosome. The peak SNP, AX-108912427, has the highest average PVE at 21% for Ca concentration and was identified in all the environments, which implies that V-type proton ATPase is a key factor affecting Ca accumulation in wheat grain. Another marker on chromosome 3B, AX-110013515, corresponds to a *transmembrane protein* (*TraesCS3B01G018800*). This is a novel resident protein on the endoplasmic reticulum (ER) and has an ER luminal domain containing rich acidic residues [47]. This amino-terminal luminal domain has a

high capacity, and moderate affinity, for binding Ca ions, and it plays an essential role in ER Ca ion handling [47], which may be closely associated with Ca accumulation in wheat grain. The SNP marker AX-109541359 on chromosome 3A was associated with the gene *TraesCS3A01G424700*, which encodes *cytochrome P450*. The sequential actions of *cytochrome P450* may result in the synthesis of a pleiotropic hormone that regulates several genes that have actions associated with Ca homeostasis as well as cellular growth [48]. The remaining SNPs, AX-110634514 on chromosome 2A and AX-94729264 on chromosome 3D, were associated with candidate genes *TraesCS2A01G098100* and *TraesCS3D01G079600*, respectively. The former encodes *calmodulin*, the predominant Ca receptor with a flexible conformation that regulates both fluxes of transient Ca-ion influx through Ca-ion channels and Ca-ion efflux by Ca transporters [49]; *calmodulin* also has the capacity to regulate Ca-ion pumps in multiple subcellular locations [50, 51]. The latter encodes a *ubiquitin family protein* involved in the process of ubiquitin conjugation. Mutating ubiquitin proteins may alter cell coupling and the resulting Ca elevation [52], which implies that ubiquitin family proteins may participate in altering Ca homeostasis in wheat. These results reveal the complex nature of Ca accumulation in wheat grain and imply that various mechanisms are involved in controlling Ca accumulation.

Organismal Ca requirements must be met through dietary uptake. The Ca intake in the adult population of Asia has been reported to be less than 500 mg/day, and in Africa and South America the Ca intake of the adult population is between 400 and 700 mg/day [15]. These values are far below the recommended standards of the Food and Agriculture Organization, which are 1,300 mg/day for children over 9 years of age and 800–1,300 mg/day for adults. Biofortification is an effective strategy to increase the microelement content of wheat and improve the human intake of Ca. However, early breeding pro-grams mainly focused on yield and ignored the microelement levels. Because of the so-called " dilution effect " in which high yields are negatively correlated with micro-element levels [3], it is difficult to select wheat varieties with high Ca contents at high yield levels using traditional breeding methods. Identifying superior allele loci and developing corresponding molecular markers have been beneficial to pyramid breeding, and this strategy could significantly enhance the microelement, including Ca, levels in wheat grain [53–55]. In this study, 6 of 18 loci were identified as harboring superior alleles and exhibited significantly higher Ca accumulations in two or more environments. By comparing the abilities of lines with both superior and inferior alleles for Ca accumulation in wheat grain at the six loci, we found that phenotypic differences reached significant levels between individuals with either superior or inferior alleles (Figure 4, Table 4). Additionally, a linear regression showed that with an increase in the number of favorable alleles, the Ca concentrations of accessions gradually increased, revealing a significant additive effect, which provides guidance for pyramid breeding (Figure 5). Markers identified by the GWAS that were significantly linked to these loci and associated with wheat grain Ca concentration may be converted into competitive allele-specific PCR markers for molecular marker-assisted selection-based breeding programs [56]. Using marker-assisted selection, the superior alleles for Ca accumulation could be integrated for multi-loci pyramid breeding, which will provide guidance for biofortification breeding. In the future, our studies will focus on validating the effects of these QTLs and developing molecular markers for wheat Ca biofortification pyramid breeding.

Conclusion

In the present study, 18 non-redundant QTLs associated with Ca concentration were identified in the surveyed environments and 6 non-redundant QTLs were stability across different environments. Among them, a QTL hot spot exists at the end of chromosome 4A and exhibited the highest PVE in each environment, ranging from 12.89–26.93%. It is imply that this locus may be embraced a key factor in regulating wheat grain Ca accumulation. Haplotype analysis results showed that cultivars containing more superior alleles and fewer inferior alleles had increased grain Ca concentrations, which was beneficial to marker-assisted selection for varieties with high Ca concentrations in wheat grain at the early developmental stages without needing to phenotype mature plants. This study not only increases our understanding of the genetic architecture of grain Ca accumulation in wheat, but also provide a guidance for wheat Ca biofortification pyramid breeding in the future.

Materials And Methods

Plant material

A natural wheat population, consisting of 207 representative varieties that collected from the Henan Province Crop Germplasm Bank and The International Maize and Wheat Improvement Center (CIMMYT), was used in the GWAS for Ca accumulation in wheat grain and the name of individual varieties were listed in Table S1. The natural population was grown under three environments at Yuanyang (YY; E113°97', N35°5'), Kaifeng (KF; E114°30', N34°80') and Shangqiu (SQ; E115°65', N34°45') in northern China during the 2016–2017 cropping season (October 2016 to June 2017). The average annual temperatures were 14.3°C, 15.0°C and 14.2°C, and the average annual rainfalls were 556, 656 and 623 mm, in YY, KF and SQ, respectively. The natural population was separately sown in different plots of 1.0 m × 1.5 m containing four rows, with 10-cm spaces between individuals and 23cm spaces between rows. The natural population at each location was subjected to standard agronomic practices.

Phenotypic analysis of wheat grains with different Ca concentrations

The whole natural population, including 207 cultivars, was harvested after reaching physiological maturity (8–10% moisture content) at the three different locations. Grain samples (approximately 50g/cultivar) were threshed by hand and carefully cleaned, and broken grains and sundries were removed. The samples were stored in paper sacks for the micronutrient analysis. The samples were milled using a Retch mill (MM301, Germany) and were dried overnight at 40°C. Then, 0.5 g dried powder samples from each cultivar were digested with 5 mL nitric acid (HNO_3, 69%, analytical reagent grade, Merck, Darmstadt, Germany) using a microwave reactor (UltraCLAVE, Milestone, Germany). After cooling, digested samples were adjusted to 25 mL with de-ionized distilled water (Milli-Q Reference System, Merck, Germany). The Ca concentration was measured using inductively coupled plasma–mass

spectrometry (Agilent 7800, Agilent Technologies Inc., USA). Each sample was tested three times to generate technical replicates, and the average values were used for further analyses (Table S1).

Genotyping and quality control

For each accession, total genomic DNA was extracted from young leaf tissue using the CTAB (cetyltrimethylammonium bromide) procedure [57]. The 207 accessions were genotyped using the Wheat Breeders 660 K Axiom® array with the Axiom 2.0 Assay Manual Workflow protocol [28]. The accuracy of SNPs was determined using Plink version 1.9 software (http://www.cog-genomics.org/plink2/) with the criteria of minor allele frequency > 0.05 and missing genotype data < 10%. In total, 244,508 SNPs were considered as creditable markers for further analyses (Figure S1).

GWAS mapping

Before performing the GWAS, the population structure (Q) and kinship (K) matrices were analyzed using ADMIXTURE version 1.3.0 (www.genetics.ucla.edu/software/admixture) and the GAPIT package [58] in R software (R Core Team, 2020), respectively. An admixture model with 10 replicates for each genetic group (K = 1–10) was implemented. A burn-in of 1,000 iterations followed by 1,000 Markov chain Monte Carlo replicates was used to estimate the number of subpopulations. After the operation, the different CV values were produced, and the optimal K was considered as having the minimal CV value. The natural population K matrix was calculated using the VanRaden method to determine relative kinship among the sampled individuals [59]. The values for BLUPs of Ca concentration, according to variety and location (YY, KF and SQ), were calculated using the mixed linear model in R package “lem4”, and they were computed using the following formula: Y = (1|Line) + (1|Loc) + (1|Rep%in%Line: Loc) + (1|Line: Loc).

The GWAS analysis for wheat grain Ca accumulations was performed using the R package GAPIT [58]. To select the optimal statistical model, three models, GLM (only accounts for population structure), MLM (accounts for population structure and relative kinship) and FarmCPU (accounts for fixed and random effects and improves calculation speed and accuracy), were used to analyze the association between phenotypic and genotypic datasets. The suggested p-value for significance was 1.0×10^{-4} to control the genetic false positive error rate for this population [60].

Putative candidate gene predictions

Based on the reported common wheat variety Chinese Spring’s reference genome sequence “IWGSC RefSeq v2.0”, a high confidence gene list was downloaded from IWGSC (http://www.wheatgenome.org/) and used to identify putative candidate genes in each QTL. Candidate genes were annotated using Ensemble Plants (http://plants.ensembl.org/Triticum_aestivum/Info/Index). The physical position of each SNP from the 660K arrays was obtained from the IWGSC website (http://www.wheatgenome.org/Tools-and-Resources/IWGSC-RefSeq-v2.0). For each locus without appropriate candidates, the gene nearest to the peak SNP was assigned.

Abbreviations

BLUPs
Best linear unbiased predictors
Ca
Calcium
CV
Cross-validation error
CTAB
Cetyltrimethylammonium bromide
FarmCPU
Fixed and random model Circulating Probability Unification
GLM
General linear models
GWAS
Genome-wide association analysis
KF
Kaifeng
MLM
mixed linear models
PVE
Phenotypic variation explain
QQ
quantile–quantile
QTL
quantitative trait locus
SQ
Shangqiu
SNP
Single-nucleotide polymorphism
YY
Yuanyang

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

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

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

Not applicable.

Funding

This research was funded by China Postdoctoral Science Foundation (2020M682300), Henan Provincial Science and Technology Research Project (212102110065), Outstanding Youth Foundation of Henan Academy of Agricultural Sciences (2022YQ13, 2020YQ02), Joint Fund of National Natural Science Foundation of China (U1804102).

Authors' contributions

Conceptualization, Z.Z., J.W. and Z.W.; Investigation, X.S., M.Q., W.L. and P.Y.; Methodology, J.H; Project administration, P.Y.; Resources, Z.L. and Z.W.; Software, X.S. and F.H.; Supervision, Z.L., Z.W. and Z.Z.; Validation, J.W.; Visualization, X.S.; Writing – original draft, X.S. and Z.Z. All authors reviewed and approved the final manuscript.

References

1. Shewry PR: **Wheat**. *J Exp Bot* 2009, **60**(6):1537-1553.
2. Bushuk W: **Wheat breeding for end-product use**. *Euphytica* 1998, **100**(1-3):137–145.
3. Garvin DF, Welch RM, Finley JW: **Historical shifts in the seed mineral micronutrient concentration of US hard red winter wheat germplasm**. *Journal of the Science of Food & Agriculture* 2010, **86**(13):2213–2220.
4. Brini M, Ottolini D, Cali T, Carafoli E: **Calcium in health and disease**. *Met Ions Life Sci* 2013, **13**(1):81–137.
5. Pravina P, Sayaji D, Avinash M: **Calcium and its role in human body**. 2012.
6. Gao X, LaValley MP, Tucker KL: **Prospective studies of dairy product and calcium intakes and prostate cancer risk: a meta-analysis**. *J Natl Cancer Inst* 2005, **97**(23):1768–1777.
7. Lin J, Manson JE, Lee IM, Cook NR, Buring JE, Zhang SM: **Intakes of calcium and vitamin D and breast cancer risk in women**. *Arch Intern Med* 2007, **167**(10):1050–1059.
8. White PJ, Bowen HC, Demidchik V, Nichols C, Davies JM: **Genes for calcium-permeable channels in the plasma membrane of plant root cells**. *Biochim Biophys Acta* 2002, **1564**(2):299–309.

9. Dodd AN, Kudla J, Sanders D: **The language of calcium signaling**. *Annu Rev Plant Biol* 2010, **61**:593–620.
10. Allen GJ, Schroeder JI: **Combining genetics and cell biology to crack the code of plant cell calcium signaling**. *Sci STKE* 2001, **2001**(102):re13.
11. Knight H: **Calcium signaling during abiotic stress in plants**. *Int Rev Cytol* 2000, **195**:269–324.
12. S.H. C: **Calcium signaling through protein kinases. The Arabidopsis calcium-dependent protein kinase gene family**. *Plant Physiology* 2002, **129**(2):469.
13. Peleg Z, Cakmak I, Ozturk L, Yazici A, Jun Y, Budak H, Korol AB, Fahima T, Saranga Y: **Quantitative trait loci conferring grain mineral nutrient concentrations in durum wheat x wild emmer wheat RIL population**. *Theor Appl Genet* 2009, **119**(2):353–369.
14. Velu G, Ortiz-Monasterio I, Cakmak I, Hao Y, Singh RP: **Biofortification strategies to increase grain zinc and iron concentrations in wheat**. *Journal of Cereal Science* 2014, **59**(3):365–372.
15. Sharma D, Jamra G, Singh UM, Sood S, Kumar A: **Calcium biofortification: three pronged molecular approaches for dissecting complex trait of calcium nutrition in Finger Millet (*Eleusine coracana L.*) for devising strategies of enrichment of food crops**. *Front Plant Sci* 2016, **7**:2028.
16. Fedorowicz-Stronska O, Koczyk G, Kaczmarek M, Krajewski P, Sadowski J: **Genome-wide identification, characterisation and expression profiles of calcium-dependent protein kinase genes in barley (*Hordeum vulgare L.*)**. *J Appl Genet* 2017, **58**(1):11–22.
17. Goel A, Taj G, Pandey D, Gupta S, Kumar A: **Genome-wide comparative in silico analysis of calcium transporters of rice and sorghum**. *Genomics Proteomics Bioinformatics* 2011, **9**(4-5):138–150.
18. Vreugdenhil D, Aarts MGM, Koornneef M, Nelissen H, Ernst WHO: **Natural variation and QTL analysis for cationic mineral content in seeds of Arabidopsis thaliana**. *Plant Cell & Environment* 2010, **27**(7):828–839.
19. Korte A, Farlow A: **The advantages and limitations of trait analysis with GWAS: a review**. *Plant Methods* 2013, **9**(1):29–38.
20. Alomari DZ, Eggert K, von Wiren N, Pillen K, Roder MS: **Genome-wide association study of calcium accumulation in grains of European wheat cultivars**. *Front Plant Sci* 2017, **8**:1797.
21. Bhatta M, Baenziger PS, Waters BM, Poudel R, Belamkar V, Poland J, Morgounov A: **Genome-wide association study reveals novel genomic regions associated with 10 grain minerals in Synthetic Hexaploid Wheat**. *Int J Mol Sci* 2018, **19**(10):3237.
22. Bálint A, Rder MS, Hell R, Galiba G, Brner A: **Mapping of QTLs affecting copper tolerance and the Cu, Fe, Mn and Zn contents in the shoots of wheat seedlings**. *Biologia Plantarum* 2007, **51**(1):129–134.
23. Crespo-Herrera LA, Govindan V, Stangoulis J, Hao Y, Singh RP: **QTL mapping of grain Zn and Fe concentrations in two hexaploid wheat RIL populations with ample transgressive segregation**. *Front Plant Sci* 2017, **8**:1800.
24. Ganeva G, Landjeva S, Merakchijska M: **Effects of chromosome substitutions on copper toxicity tolerance in wheat seedlings**. *Biologia Plantarum* 2003, **47**(4):621–623.

25. Y.F. X, D.G. A, H.J. L, H.X. X: **Breeding wheat for enhanced micronutrients**. *Canadian Journal of Plant Science* 2011, **91**(2):231–237.
26. Liu S, Fan C, Li J, Cai G, Yang Q, Wu J, Yi X, Zhang C, Zhou Y: **A genome-wide association study reveals novel elite allelic variations in seed oil content of Brassica napus**. *Theor Appl Genet* 2016, **129**(6):1203–1215.
27. Wang S, Wong D, Forrest K, Allen A, Chao S, Huang BE, Maccaferri M, Salvi S, Milner SG, Cattivelli L *et al*: **Characterization of polyploid wheat genomic diversity using a high-density 90,000 single nucleotide polymorphism array**. *Plant Biotechnol J* 2014, **12**(6):787–796.
28. Winfield MO, Allen AM, Burridge AJ, Barker GL, Benbow HR, Wilkinson PA, Coghill J, Waterfall C, Davassi A, Scopes G *et al*: **High-density SNP genotyping array for hexaploid wheat and its secondary and tertiary gene pool**. *Plant Biotechnol J* 2016, **14**(5):1195–1206.
29. Allen AM, Winfield MO, Burridge AJ, Downie RC, Benbow HR, Barker GL, Wilkinson PA, Coghill J, Waterfall C, Davassi A *et al*: **Characterization of a wheat breeders' array suitable for high-throughput SNP genotyping of global accessions of hexaploid bread wheat (***Triticum aestivum L.*). *Plant Biotechnol J* 2017, **15**(3):390–401.
30. Sun C, Dong Z, Zhao L, Ren Y, Zhang N, Chen F: **The Wheat 660K SNP array demonstrates great potential for marker-assisted selection in polyploid wheat**. *Plant Biotechnol J* 2020, **18**(6):1354–1360.
31. Long AD, Langley CH: **The power of association studies to detect the contribution of candidate genetic loci to variation in complex traits**. *Genome Res* 1999, **9**(8):720–731.
32. Zhu C, Gore M, Buckler ES, Yu J: **Status and prospects of association mapping in plants**. *Plant Genome* 2008, **1**(1):5–20.
33. Cormier F, Gouis JL, Dubreuil P, Lafarge S, Praud S: **A genome-wide identification of chromosomal regions determining nitrogen use efficiency components in wheat (***Triticum aestivum L.*). *Tagtheoretical & Applied Geneticstheoretische Und Angewandte Genetik* 2014, **127**(12):2679–2693.
34. Yang X, Pan Y, Singh PK, He X, Ren Y, Zhao L, Zhang N, Cheng S, Chen F: **Investigation and genome-wide association study for Fusarium crown rot resistance in Chinese common wheat**. *BMC Plant Biology* 2019, **19**(1):153.
35. Hitz K, Clark AJ, Sanford DV: **Identifying nitrogen-use efficient soft red winter wheat lines in high and low nitrogen environments**. *Field Crops Research* 2017, **200**:1–9.
36. Flint-Garcia SA: **Genetics and consequences of crop domestication**. *J Agric Food Chem* 2013, **61**(35):8267–8276.
37. Pritchard JK, Stephens M, Rosenberg NA, Donnelly P: **Association mapping in structured populations**. *Am J Hum Genet* 2000, **67**(1):170–181.
38. Yu J, Buckler ES: **Genetic association mapping and genome organization of maize**. *Curr Opin Biotechnol* 2006, **17**(2):155–160.
39. 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):953–969.

40. Zhao Z, Zhang H, Fu Z, Chen H, Lin YN, Yan PS, Li WH, Xie HL, Guo ZY, Zhang XH: **Genetic-based dissection of arsenic accumulation in maize using a genome-wide association analysis method**. *Plant Biotechnology Journal* 2017, **16**(5):1085–1093.
41. Shi RL, Tong YP, Jing RL, Zhang FS, Zou CQ: **Characterization of Quantitative Trait Loci for Grain Minerals in Hexaploid Wheat (***Triticum aestivum L.*). *Journal of Integrative Agriculture* 2013, **12**(9):1512–1521.
42. Shen X, Yuan Y, Zhang H, Guo Y, Zhao Y, Li S, Kong F: **The hot QTL locations for potassium, calcium, and magnesium nutrition and agronomic traits at seedling and maturity stages of wheat under different potassium treatments**. *Genes* 2019, **10**(8):607.
43. Mackay TFC, Stone EA, Ayroles JF: **The genetics of quantitative traits: challenges and prospects**. *Nat Rev Genet* 2009, **10**(8):565–577.
44. Genc Y, Taylor J, Rongala J, Oldach K: **A major locus for chloride accumulation on chromosome 5A in bread wheat**. *PLoS One* 2014, **9**(6):e98845.
45. Bickel-Sandkötter S, Wagner V, Schumann D: **ATP-synthesis in archaea: Structure-function relations of the halobacterial A-ATPase**. *Photosynthesis Research* 1998, **57**(3):335–345.
46. Samuelson JC, Azikiwe N, Shen PS: **Proton-transporting and calcium ion-transporting ATPases of Entamoeba histolytica**. *Parasitology Today* 1995, **11**(11):417–420.
47. Miao Z, Yamazaki T, Yazawa M, Treves S, Nishi M, Murai M, Shibata E, Zorzato F, Takeshima H: **Calumin, a novel Ca^{2+}-binding transmembrane protein on the endoplasmic reticulum**. *Cell Calcium* 2007, **42**(1):83–90.
48. Omdahl JL, Morris HA, May BK: **Hydroxylase enzymes of the vitamin D pathway: expression, function, and regulation**. *Annu Rev Nutr* 2002, **22**:139–166.
49. Sze H, Feng L, Hwang I, Curran AC, Harper JF: **Diversity and regulation of plant Ca^{2+} pumps: Insights from expression in yeast**. *Annual Review of Plant Physiology & Plant Molecular Biology* 2000, **51**(1):433–462.
50. Chung WS, Lee SH, Kim JC, Heo WD, Kim MC, Park CY, Park HC, Lim CO, Kim WB, Harper JF *et al*: **Identification of a calmodulin-regulated soybean Ca^{2+}-ATPase (SCA1) that is located in the plasma membrane**. *Plant Cell* 2000, **12**(8):1393–1407.
51. Harper JF: **Dissecting calcium oscillators in plant cells**. *Trends in Plant Science* 2001.
52. Liu K, Lyu L, Chin D, Gao J, Sun X, Shang F, Caceres A, Chang ML, Rowan S, Peng J *et al*: **Altered ubiquitin causes perturbed calcium homeostasis, hyperactivation of calpain, dysregulated differentiation, and cataract**. *Proc Natl Acad Sci U S A* 2015, **112**(4):1071–1076.
53. Cheema SA, Rehman HU, Kiran A, Bashir K, Wakeel A: **Plant micronutrient use efficiency**. *Elsevier: Amsterdam, Netherlands* 2018, **1st ed**:261–278.
54. Xu Y, An D, Liu D, Zhang A, Xu H, Li B: **Molecular mapping of QTLs for grain zinc, iron and protein concentration of wheat across two environments**. *Field Crops Res* 2012, **138**:57–62.

55. Borrill P, Connorton JM, Balk J, Miller AJ, Sanders D, Uauy C: **Biofortification of wheat grain with iron and zinc: integrating novel genomic resources and knowledge from model crops**. *Front Plant Sci* 2014, **5**:53.
56. Chen J, Zhang F, Zhao C, Lv G, Sun C, Pan Y, Guo X, Chen F: **Genome-wide association study of six quality traits reveals the association of the *TaRPP13L1* gene with flour colour in Chinese bread wheat**. *Plant Biotechnol J* 2019, **17**(11):2106–2122.
57. Doyle JJ, Doyle JL: **A rapid DNA isolation procedure from small quantities of fresh leaf tissues**. *Phytochemistry* 1986, **19**(1).
58. Lipka AE, Tian F, Wang Q, Peiffer J, Li M, Bradbury PJ, Gore MA, Buckler ES, Zhang Z: **GAPIT: genome association and prediction integrated tool**. *Bioinformatics* 2012, **28**(18):2397–2399.
59. Vanraden PM: **Efficient methods to compute genomic predictions**. *Journal of Dairy Science* 2008, **91**(11):4414–4423.
60. Zhou Z, Shi X, Zhao G, Qin M, Wang J: **Identification of novel genomic regions and superior alleles associated with Zn accumulation in wheat using a genome-wide association analysis method**. *International Journal of Molecular Sciences* 2020, **21**(6):1928.

Tables

Due to technical limitations, tables 1 to 4 are only available as a download in the Supplemental Files section.

Figures

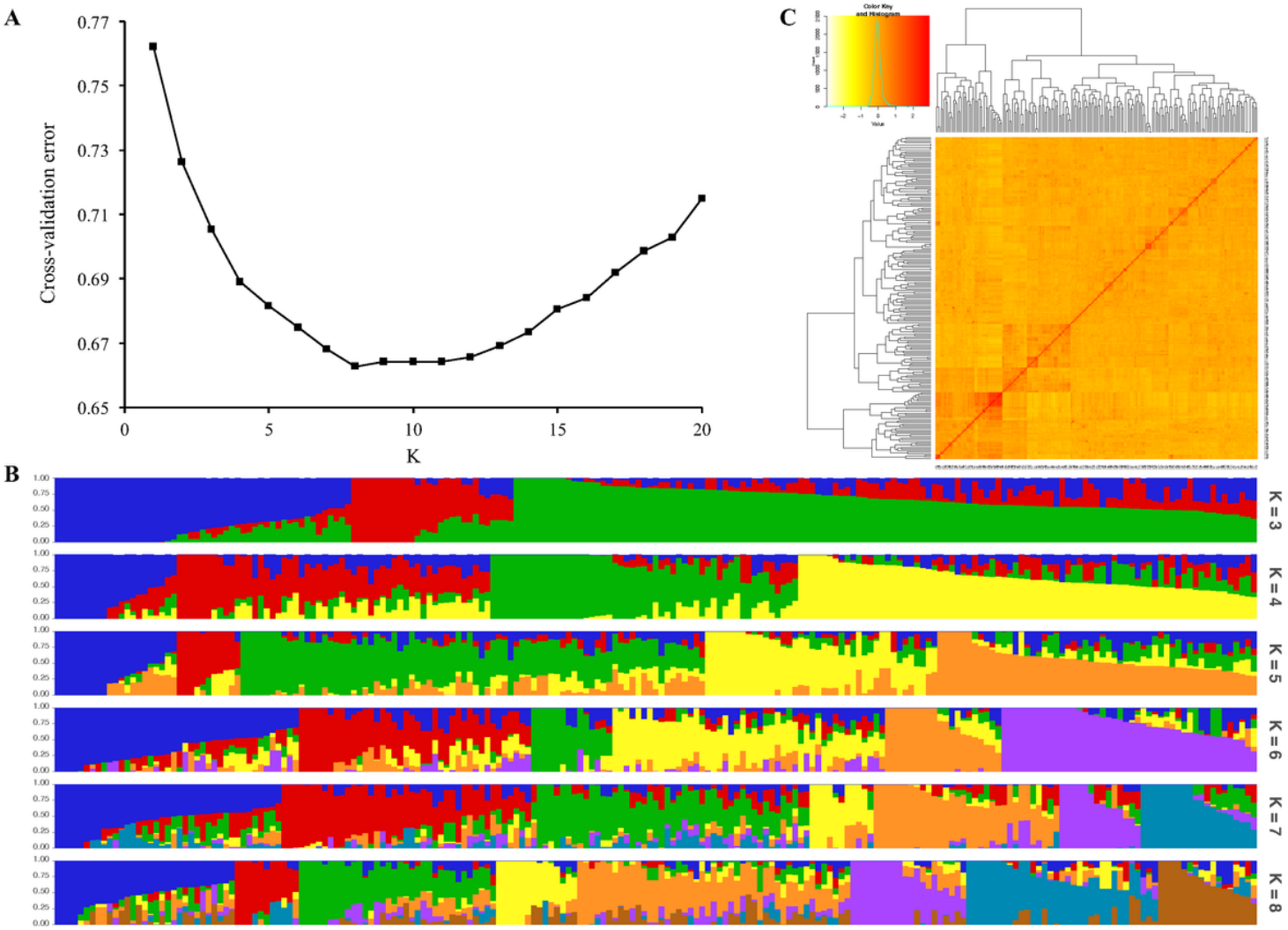

Figure 1

Population structure and kinship estimations in 207 wheat cultivars. (A) Plot of cross-validation errors. (B) Population structure of wheat cultivar panel from K = 3 to K = 8. (C) Kinship matrix, showing pairwise genetic relatedness among individuals.

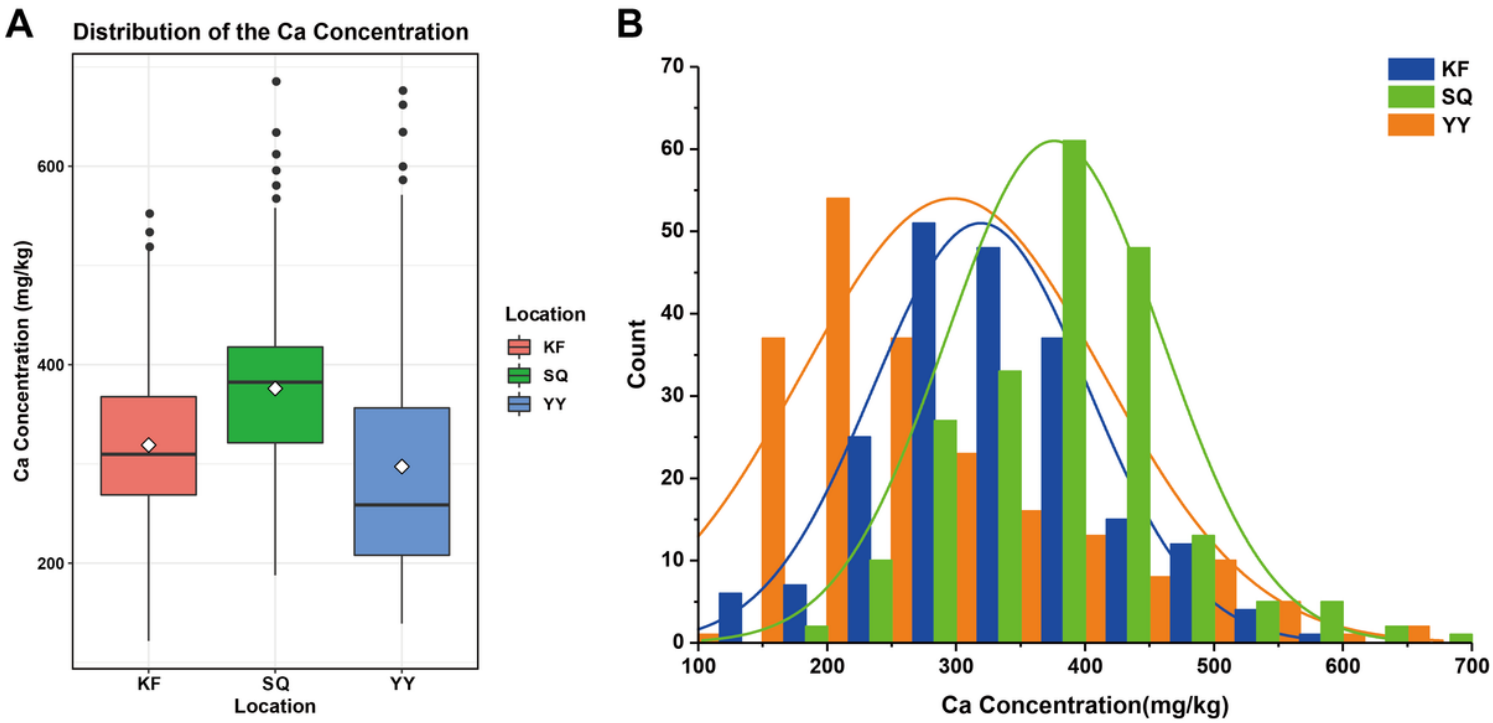

Figure 2

Distribution of the wheat grain calcium concentrations measured in the natural population. (A) Boxplot showing the calcium concentrations among different locations. Data from three locations, Yuanyang (YY), Kaifeng (KF) and Shangqiu (SQ), are shown. (B) Distribution of the calcium con-centrations in the association populations from YY, KF and SQ.

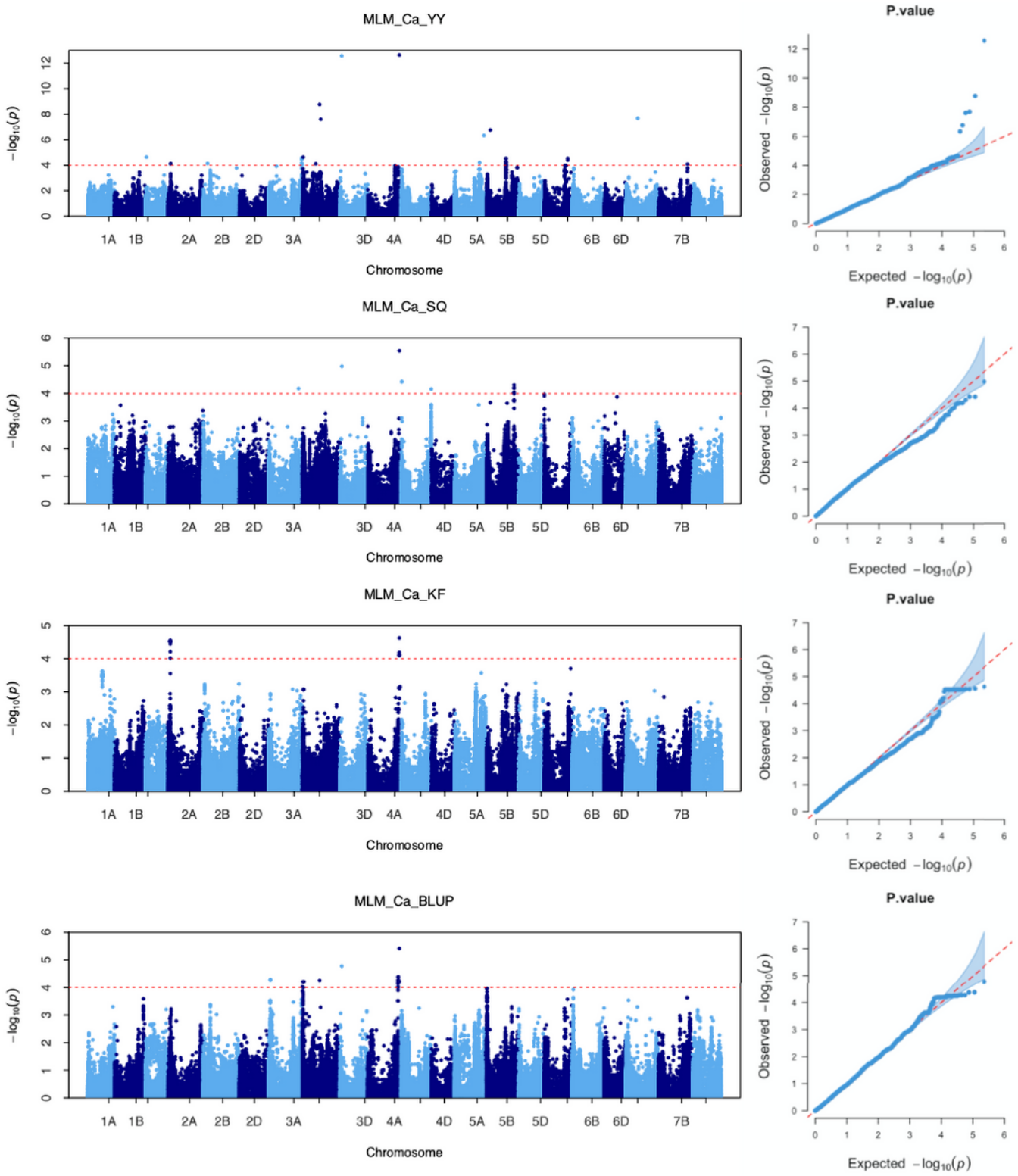

Figure 3

Manhattan and quantile-quantile plots for calcium concentrations using the mixed linear model for wheat grains across different environments (including the combined location, BLUP). The dashed horizontal line represents the significant threshold of −log10(P) = 4.0. The SNPs above the red dotted line are significantly associated with calcium variation.

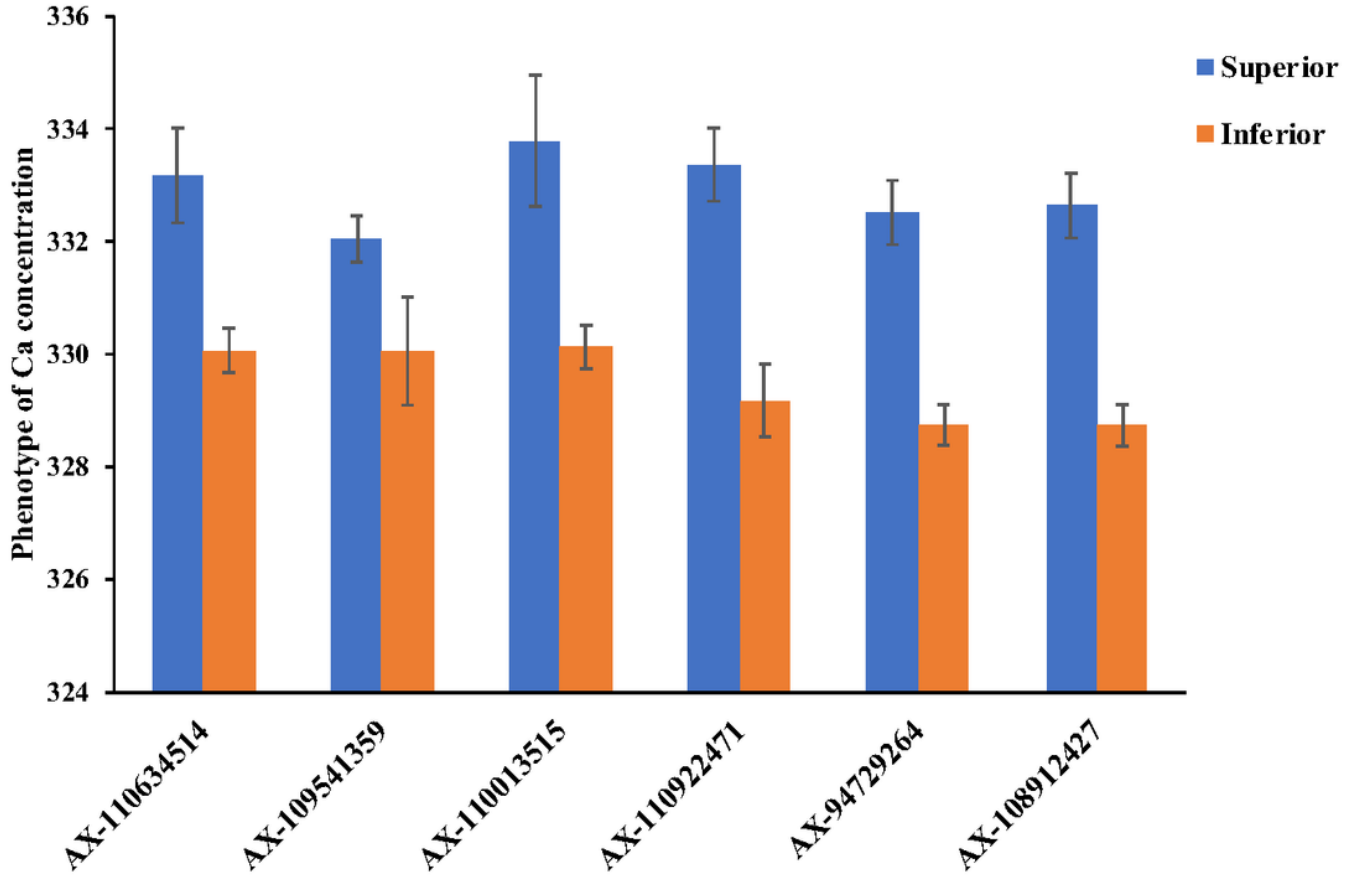

Figure 4

The phenotype values of accessions with superior alleles (blue bar) and inferior alleles (orange bar) of repetitive significant SNPs for wheat grain calcium concentration.

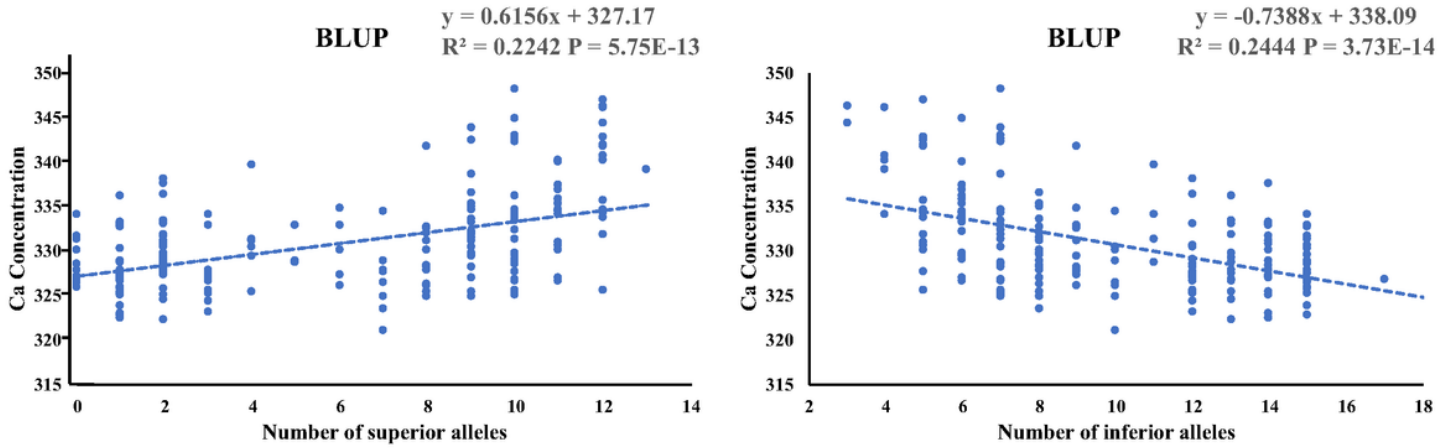

Figure 5

Linear regression between the number of (A) superior alleles and (B) inferior alleles.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryBMCsubmitCa20211128bySX.docx
- Table1.xlsx
- Table2.xlsx
- Table3.xlsx
- Table4.xlsx