

Running Title: GENOTYPIC VARIATIONS OF WHEAT UNDER CADMIUM STRESS

Phenotypic variations of wheat cultivars from the North China plain in response to cadmium stress and associated single nucleotide polymorphisms identified by genome-wide association study

Imdad Ullah ZAID¹, Sajid Hanif MUHAMMAD^{1,2}, Na ZHANG¹, Xin ZHENG^{1,*}, Likun WANG¹ and Xiaofang LI^{1,3,*}

¹Centre for Agricultural Resources Research, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Shijiazhuang 050021, China

²University of Chinese Academy of Sciences, Beijing 100049, China

³CMLR, Sustainable Minerals Institute, The University of Queensland, Brisbane, Queensland 4072, Australia

ABSTRACT

Understanding the genetic mechanisms for cadmium (Cd) uptake and translocation in common wheat is of significance in food Cd pollution control. In this study, a diverse panel of 132 wheat cultivars from the North China Plain was collected and evaluated in terms of their phenotypic variations in response to Cd stress, and subject to a genome-wide association study (GWAS) to identify single nucleotide polymorphisms (SNPs) associated with the phenotypic variations at the seedling stage. Significant phenotypic variations with high heritability were observed among the wheat cultivars under 40 μM Cd for the studied traits, including root length (RL), shoot length (SL), root and shoot dry biomass (RDW and SDW), root and shoot Cd content (RCD and SCD), and Cd translocation factor (TF). Mean values of RCD, SCD and TF ranged from 1.0 to 33.8 mg/g, 0.125 to 2.022 mg/g, and 0.009 to 0.321, respectively. Cluster analysis showed that wheat cultivars with higher RL, SL, RDW and SDW under Cd stress were labile to accumulate more Cd in root, leading to a lower TF. The mixed linear model-based association analysis detected 17 novel significant marker-trait associations (MTAs), four of which were significant at a genome-wide scale. Most of the significant MTAs controlled TF and explained 17.17% to 26.47% of the phenotypic variations. Some of the SNP loci were physically close to a reported Cd-related QTL or gene on wheat chromosomes. Results in this study provided a list of wheat cultivars with the potential of low Cd accumulation and enriched our knowledge on the genetic basis of Cd uptake and translocation in wheat. Pyramiding breeding of superior alleles detected in this study may additionally reduce Cd accumulation of improved wheat cultivars with excellent agronomic traits.

Key Words: cadmium uptake, common wheat, genome-wide association study, phenotypic variation, single nucleotide polymorphism, translocation factor

*Corresponding author. E-mail: xfli@sjziam.ac.cn, zhengxin@sjziam.ac.cn.

INTRODUCTION

Cadmium (Cd) is the most widespread heavy metal in farmland in China as well as in many parts of the world (Ministry of Environmental Protection of China and Ministry of Land and Resources of China, 2014; Zhang *et al.*, 2015). Common wheat ($2n = 6x = 42$, AABBDD) is able to accumulate a considerable amount of Cd in grains and shows a higher Cd bioaccumulation capacity than rice, even in the soils of neutral/alkaline pH where soil Cd bioavailability is thought to be low (Li and Zhou, 2019). It has been reported that around 10% of grains were Cd-polluted in the surveyed market in China (Zhang *et al.*, 2015). Atmospheric deposition, livestock manure, wastewater irrigation, metallo-pesticides or herbicides, phosphate fertilizers and sewage sludge amendments are main sources of the heavy metals in crops (Rai *et al.*, 2019). According to the national food safety standards (GB 2762-2017), the limit for Cd content in wheat grains is 0.1 mg/kg. Excess Cd impacts not only the quality of wheat grains and subsequent yield due to Cd-associated growth retardation but also human health (Khan *et al.*, 2007; Rizwan *et al.*, 2016).

The use of low-Cd crops has been proposed as a promising approach for food Cd pollution control (Li, 2019). Some studies have reported low-Cd rice cultivars with respect to genetic mapping and engineering of Cd-related traits (Lu *et al.*, 2019; Wang *et al.*, 2020). For wheat, a number of studies have also been dedicated to the screening of low-Cd wheat cultivars. An early research has investigated the phenotypic variations in Cd accumulation of wheat varieties from Australia (Oliver *et al.*, 1995). Recently, some large-scale screenings of wheat cultivars were carried out in search of low-Cd wheat cultivars (Sun *et al.*, 2015; Chen *et al.*, 2017; Xiong *et al.*, 2018; Ai *et al.*, 2019; Li *et al.*, 2019; Li and Zhou, 2019). However, most of the low-Cd cultivars identified so far based on the phenotypic screening may be not sufficient for remediation practice, considering that these cultivars were bred purposefully for other traits like yield, biotic/abiotic stress resistance, and nutrient use efficiency (Zaid *et al.*, 2018). The use of modern breeding strategies is thus necessary to breed wheat varieties dedicated to low Cd accumulation, which necessitates the exploration of genetic regions controlling Cd accumulation in wheat (Ma *et al.*, 2020). Deciphering these regions will be beneficial to identify marker loci for pyramiding breeding (Talukdar *et al.*, 2017).

Genome-wide association study (GWAS) is commonly used to explore marker-trait associations (MTAs) in cereal crops, and has been recently applied for barley (Wu *et al.*, 2015), rice (Oono *et al.*, 2016), oat (Tanhuanpaa *et al.*, 2007) and maize (Zhao *et al.*, 2018) in regards to their Cd stress response. It utilizes the natural populations for uncovering the significantly associated loci (Ueda *et al.*, 2015). In wheat, high-throughput genotyping has been possible with the advent of genotyping-by-sequencing (Rife *et al.*, 2015) and single nucleotide polymorphism (SNP) discovery (Winfield *et al.*, 2012), as well as the subsequent development of SNP-based marker arrays. The advances in wheat genome sequencing have provided a platform to breeders to identify marker regions in wheat. Previous studies have reported several Cd-related marker loci in wheat genome, such as Cd uptake-related SNPs (Guttieri *et al.*, 2015; Oladzad-Abbasabadi *et al.*, 2018; Quraishi, 2020), random amplification polymorphic DNA (RAPD) markers for Cd exclusion (Penner *et al.*, 1995), and quantitative trait loci (QTLs) for Cd tolerance (Ci *et al.*, 2012). AbuHammad mapped a major SNP on chromosome 5BL, which contributed to low Cd uptake explaining about 54.3% of the phenotypic variation (AbuHammad *et al.*, 2016). Maccaferri reported a variation on the *HMA3* gene, which caused a two to three-fold change of Cd uptake in durum wheat (Maccaferri *et al.*, 2019). Overexpression of *HMA3* or introducing the superior allele of *HMA3* via crossbreeding largely reduced grain Cd content (Lei *et al.*, 2020; Zhang *et al.*, 2020).

The primary objective of this study is to evaluate the Cd responding phenotypic variations of wheat cultivars from the North China Plain and identify associated SNPs in wheat. To this end, a diverse panel

of 132 wheat cultivars was collected from the North China Plain, which was subjected to a long-term Cd exposure to evaluate traits variations so as to determine genome-wide superior or inferior SNP alleles controlling Cd uptake. The outcomes of this study may provide wheat cultivars with desired traits for remediation practice, and the SNP loci associated with wheat Cd response is of values for future wheat breeding or bioengineering.

MATERIALS AND METHODS

Wheat cultivars and treatments

In total 132 wheat cultivars from the North China Plain were collected and tested in the present study. These cultivars have been reported in terms of their kernel traits in our previous work (Zhang *et al.*, 2020). Seeds germination was initiated by surface sterilization with 30% hydrogen peroxide solution for 10 minutes followed by thoroughly washing with sterile water. After sterilization, seeds were put on filter paper placed in petri dishes to germinate for 48 hours. Two-day-old uniform seedlings were selected and transferred into holes of specialized polyethylene germination bags containing 40 ml hydroponic solution as a growth medium. The germination bags were kept in a growth chamber under controlled conditions (light/dark regimes of 16/8 h, humidity 55/65% and temperature 20/18°C). The standard nutrient solution consists of 2.5 g/L (K_2SO_4 , KCl, $MgSO_4 \cdot 7H_2O$), 2.5 g/L $Ca(NO_3)_2 \cdot 4H_2O$, 5 g/L $NH_4H_2PO_4$, 0.5 g/L NaFe(III)EDTA, 0.2 g/L $MnSO_4 \cdot H_2O$, $ZnSO_4 \cdot 7H_2O$, $CuSO_4 \cdot 5H_2O$ and 0.5 g/L ($(NH_4)_6Mo_7O_{24} \cdot 4H_2O$, H_3BO_3). The Cd stress of 40 μM , which led to substantial phenotypic variations, was applied as $CdCl_2 \cdot 2.5H_2O$ with a final nutrient solution. The Cd-containing nutrient solution was regularly replaced at a 5-day interval. The experiment was terminated at 30 days post to seedlings exposure to Cd stress.

Phenotyping and statistical analysis

The 132 wheat cultivars were evaluated in two experimental repeats for measuring agronomic trait values. The experimental design, for both repeats, was a randomized complete block design (Gomez *et al.*, 1984). Each cultivar was grown as three replicates with each consisting of six seedlings. Seedlings of each cultivar without the addition of Cd were set as controls. At 30th day post to Cd treatment, three wheat seedlings per cultivar in each replicate were randomly selected to measure root length (RL) and shoot length (SL), and then separated into roots and shoots parts. The mean phenotypic value of the cultivars collected out of two experimental repeats was used for statistical analysis and further association analysis.

All values for biomass and Cd contents were obtained in three replications by using complete randomized design (Gomez *et al.*, 1984). The root and shoot samples were thoroughly washed with deionized water and all enzymatic activity was halted by oven-drying. The root dry weight (RDW) and shoot dry weight (SDW) were weighed after oven-dried at 65°C for 48 hours and then at 85°C to constant weight. The dried root and shoot of each sample were chopped manually. The fine powder of samples was analyzed for Cd content after wet digestion by microwave digestion system (Multiwave Pro - Anton Paar, GmbH, Austria). The digestion was performed according to a previous study with modification (Fan *et al.*, 2005). In brief, to each sample 6 mL of nitric acid (HNO_3) were added and heated from room temperature to 180°C for 90 minutes, followed by adding 1 mL perchloric acid ($HClO_4$) and heating at 150°C for 120 minutes. Upon cooling, the residual was dissolved using the ultra-pure water and transferred to a 10 mL volumetric flask. The Cd concentration in the solution was determined by Inductively Coupled Plasma Mass Spectrometry (Nexion 300D, PerkinElmer, Shelton,

CT, USA) and the Cd contents in root (RCD, mg/g dry weight) and shoot (SCD, mg/g dry weight) were finally estimated. The values of the Cd translocation factor ($TF = SCD/RCD$) were determined.

Differences in traits among cultivars were determined through the analysis of variance (ANOVA) using SAS PROC GLM (Release 9.1.3; SAS Institute, Cary, NC). The genetic parameters, including the phenotypic coefficient of variation (CV) and genotypic coefficient of variation for mean values, were computed on the basis of ANOVA (Robert, 1999). Broad sense heritability (h_B^2) estimates were calculated from the phenotypic and genotypic variances according to the well-established method (Hanson *et al.*, 1956).

Genotyping and SNP resources

Wheat 55K SNP array containing 53,063 markers (China Golden Marker Biotech, Beijing, China) was utilized to genotype the 132 genomic DNA samples, which has been reported in our previous work (Zhang *et al.*, 2020). Briefly, genomic DNA was extracted and purified from leaf tissues, and DNA samples for each wheat cultivar was hybridized on the Wheat55K SNP array (Liu *et al.*, 2018). The SNPs with high resolution and best quality were captured for further analysis. Finally, a total of 7,948 high confidence SNPs after filtering, with B genome having a maximum number (3,039) of SNPs followed by the A genome (2,993) and the D genome (1,916), were used to carry out the population structure, kinship and marker-trait association analysis. The 132 genotypes were clustered into two sub-populations, by the structure analysis based on the SNP data.

Genome-wide association analysis

Genome-wide association analysis of phenotypic data under Cd stress with SNP markers was analyzed using the kinship matrix in a mixed linear model (MLM) by Tassel 5.0 (Zhang *et al.*, 2010). In the MLM analysis, the kinship matrix (K matrix) assessed by SNP markers was combined with population structure (Q matrix) to increase the statistical power of the association analysis (Pritchard *et al.*, 2000). The regression-based coefficient of determination values of all significantly associated SNPs was recorded to determine the variations explained by each SNP locus. Bonferroni correction threshold based on the effective numbers of independent markers was adjusted at $P < 0.05$ ($0.05/7948 = 6.3E-06$), and MTAs with $P < E-04$ was suggested as significant associations in this study.

Bioinformatic analysis

The physical positions of the studied SNP markers were obtained from the website of International Wheat Genome Sequencing Consortium (IWGSC, <http://www.wheatgenome.org/>). The SNPs-associated genes were evaluated for their sequences, domain and subcellular location by using online databases UniProt (<https://www.uniprot.org/>) and InterPro (<https://www.ebi.ac.uk/interpro/>).

RESULTS AND DISCUSSION

Phenotypic variations under Cd stress at seedling stage

The analysis of variance (ANOVA) revealed continuous and broad differences of significance ($P \leq 0.01$) for all wheat traits under 40 μM Cd stress (Table 1 and Fig. S1). These variations indicate a substantial phenotypic variability of the tested cultivars in response to Cd stress. Similar effects of Cd stress on different wheat populations have been observed in previous studies (Ahmad *et al.*, 2012; Ahmad *et al.*, 2013; Quraishi, 2020). RL is one of the most primary parameters to assess the Cd sensitivity of seedlings (Ouzounidou *et al.*, 1997; Greger and Lofstedt, 2003; Vassilev *et al.*, 2004). In

this study, a big difference in RL reduction after Cd stress was observed between the tolerant and sensitive wheat cultivars (Fig. S2). The mean values of RL and SL under Cd stress varied from 13.5 to 32.1 cm, and 12.5 to 23.1 cm, respectively, with the CV value ranging from 11.23 to 18.68% (Table 1). The wheat cultivars, Shi10-4195 and Jimai161, showed the smallest RL reduction rate (4.9%), while Taishan 22 and Jintai 170 showed the smallest SL reduction rate (2.5 and 6.6%, respectively). The mean values for RDW and SDW were also reduced by Cd stress substantially. Taishan 22 showed the highest RDW (22.2 mg) and SDW (38.6 mg) under Cd stress. Higher CVs were observed with the traits of RCD, SCD, and TF, and the mean values for RCD, SCD, and TF ranged from 1.0 to 33.8 mg/g, 0.125 to 2.022 mg/g, and 0.009 to 0.321, respectively (Table 1).

TABLE 1

Statistics and heritability estimation for the traits of the association mapping panel

Trait	Min.	Max.	Mean	SD	CV (%)	H (%)
RL (cm)	13.467	32.089	22.894	4.277	18.683	92.9
SL (cm)	12.461	23.072	15.969	1.794	11.232	97.5
RDW (mg)	5.933	22.200	12.809	3.311	25.848	81.1
SDW (mg)	12.517	38.567	22.923	5.073	22.129	87.7
RCD (mg/g)	1.031	33.836	10.886	5.503	50.550	83.4
SCD (mg/g)	0.125	2.022	0.717	0.378	52.769	82.2
TF	0.009	0.321	0.089	0.061	68.859	45.5

RL, root length; SL, shoot length; RDW, root dry weight; SDW, shoot dry weight; RCD, root cadmium content; SCD, shoot cadmium content; TF, translocation factor; SD, standard deviation; CV, coefficient variation; H, board-sense heritability

Cd content was found higher in roots than in shoots for all cultivars, consistent with the previous findings that root is a major sink for excess Cd in wheat (Ci *et al.*, 2009; Naeem *et al.*, 2016). High h_B^2 was also observed for most of the traits, suggesting that genetic factors play important roles in the formation of these traits (Salsman *et al.*, 2018). The current heritability results are in agreement with earlier studies showing that traits under Cd stress were highly heritable (Clarke *et al.*, 1997; AbuHammad *et al.*, 2016).

Significant correlations were observed among several traits. In general, significant positive correlations were found between the RL and RDW as well as between SL and SDW, while the dry weight of shoot or root did not correlate well with Cd contents of shoot or root (Table S1). This result indicates that inherent genetic factors but not biomass effects control the Cd accumulation in wheat organs (Li and Zhou, 2019). The correlation coefficient between RCD and TF ($R^2 = -0.579$, $P < 0.01$) was higher than that between SCD and TF ($R^2 = 0.310$, $P < 0.01$) (Table S1), implying that a wheat cultivar of lower root Cd has a great potential to harbor a higher TF, which is consistent with previous reports showing the sink feature of root in response to excess Cd (Ci *et al.*, 2009; Naeem *et al.*, 2016). In remediation practice, it may be meaningful to search of a wheat cultivar with a relatively higher root Cd to be used in the safe utilization of Cd-contaminated farmland.

Genetic diversity and population structure

Our previous study reported that the Wheat 55K SNP Array employed 7,948 SNPs after filtering in the current association mapping panel (Zhang *et al.*, 2020). The population was identified to have a high genetic diversity at the SNP level, based on the polymorphism information analysis (Zhang *et al.*, 2020).

Population structure always has a great impact on the results of marker-trait association analysis, whereas it can lead to spurious associations if not properly controlled. Effective control of population structure directly ensured the accuracy of association analysis (Ma *et al.*, 2018). Here, the association mapping panel was divided into two subgroups based on the SNP data (Zhang *et al.*, 2020). Thus, the current population can be used for association analysis. In other studies, two distant lineages of genotype populations were also detected, i.e., with 161 wheat accessions and 177 *A. tauschii* accessions (Pariyar *et al.*, 2016; Arora *et al.*, 2017). Higher marker diversity and low population structure were obtained with a significant impact on GWAS for identifying factual associations between tested traits and makers alleles (Myles *et al.*, 2009). It also suggested that the addition of population structure and relative kinship will mark the association analysis more accurately (Yu *et al.*, 2006).

To further explore the relationships among all traits in the mapping panel, cluster analysis based on the Cd-related phenotypic data was performed. The current population was separated into two subgroups: one consists of 42 genotypes with higher RL, SL, RDW, SDW, and RCD, and lower SCD and TF; the other one includes 90 genotypes with relatively opposite traits (Table 2). Given the fact that higher RL, SL, RDW, or SDW under Cd stress generally represents greater Cd tolerance, these results indicate that the wheat cultivars with higher Cd tolerance are labile to accumulate more Cd in root, leading to a lower TF. In addition, only 83 (62.9%) of the wheat genotypes shared the same genotypic and phenotypic classifications (Table S2).

TABLE 2
Phenotypic classification and statistics of the association mapping panel

Subgroup		RL	SL	RDW	SDW	RCD	SCD	TF
1 (n=42)	Mean	25.643	17.059	16.045	27.953	13.518	0.690	0.063
	SD	4.083	2.164	2.713	4.286	6.517	0.428	0.038
2 (n=90)	Mean	21.612	15.461	11.299	20.576	9.658	0.731	0.107
	SD	3.744	1.325	2.344	3.450	4.496	0.355	0.074

Significant MTA and potential associated genes

Association analysis of 7,948 SNP markers was performed with all traits investigated under Cd stress. The well-developed MLM (Q+K) method (Zhao *et al.*, 2007) allowed the discovery of 17 significant MTAs with a probability level range of 8.61E-05 – 5.80E-07 (Table S3; Figs 1, S3 and S4), and these MTAs were scattered over 12 chromosomes (1A, 1B, 1D, 2D, 3A, 3D, 4A, 4B, 5A, 5B, 6D, and 7B) of wheat (Table S3). Only four MTAs (two for TF, one for RCD, and one for SCD) were detected with genome-wide significance (P value < 6.29E-06) (Table 3). Given that stringent criteria, such as Bonferroni correction, may cause dismissing of the possible genetic variants (Gordon *et al.*, 2016), all markers with an acceptable probability level (P value < E-03) were listed in Table S3 for further evaluation.

Fig. 1 The SNP loci in Manhattan plot associated with the Cd translocation factor. The lines indicate different levels of significance: red line (P < 6.29E-06); blue line (P < E-04).

TABLE 3
The marker-trait associations with genome-wide significance identified in this study

Trait	Marker	Chromosome	Position (Mb)	P value	PVE (%)
RCD	AX-111054146	1D	492.8	4.98E-06	23.2

TF	AX-108795510	2D	106.3	5.80E-07	26.5
TF	AX-109324739	3D	334.6	3.10E-06	23.3
SCD	AX-108775210	5B	696.1	3.31E-06	24.1

Most of the significant SNP loci (13 of 17) were found to be associated with TF, explaining an average of 18.9% phenotypic variation ranging from 17.2% to 26.5% (Table S3). To our knowledge, all these identified MTAs were novel; this may be due to the complexity of studied traits, growth conditions and sampling time (Xue *et al.*, 2009). The SNP marker AX-108795510 showed the highest peak in this study with a *P* value of 5.8E-07 (Fig. 1). It is noticed that this SNP on chromosome 2D is located inside a protein kinase gene (TraesCS2D02G161300). *AtVIK*, the kinase gene in *Arabidopsis* homologous to TraesCS2D02G161300, has been reported to be involved in auxin-activated and brassinosteroid-mediated signalling pathway (Ceserani *et al.*, 2009). A number of TF-related SNP loci with relatively higher *P* values were also found to be associated with some potential Cd-related genes. The SNP marker AX-110389721, for instance, is located in the exon region of an uncharacterized gene (TraesCS1B02G357100). Bioinformatic prediction showed that TraesCS1B02G357100 encodes a trans-membrane protein containing a cytoplasmic armadillo-type fold domain. Some SNPs may be from one QTL due to their physically close location on the same chromosome. For example, the SNP markers AX-109301565, AX-109980293, AX-110067443 and AX-108904194 are located in the region of 576.2 Mb - 577.0 Mb on chromosome 1B (Table S3). This region contains five annotated genes, which encode a serine/threonine kinase (TraesCS1B02G347100), a Tubby-like F-box protein (TraesCS1B02G347200), a Dof-type Zinc finger protein (TraesCS1B02G347400), a GA2ox-B4 protein (TraesCS1B02G347600), and a monooxygenase (TraesCS1B02G347700), respectively. These genes may be involved in the cellular process of signalling transduction, abiotic stress response, transcription regulation, gibberellin catabolism and biological metabolism, respectively. However, no significant linkage with these clustered SNPs was detected in this region, which may due to the low density of SNP markers (Quraishi, 2020). Considering the importance of this region in determining wheat's Cd TF, the potential function of the embedded genes associated with Cd stress needs further experimental validation.

A few significant MTAs were identified for root/shoot Cd content traits in the current study. For RCD, only one significant MTA were identified (Tables 3 and S3). The associated SNP AX-111054146, which was also associated with SDW (*P* value = 4.43E-04), is located on chromosome 1D. This SNP explained 23.2% of the RCD variation and is physically close to another SNP AX-110960120 (within 0.5 Mb). AX-110960120 has been identified to be significantly associated with kernel traits in the current population (Zhang *et al.*, 2020). An SNP AX-110520028 was also detected to be associated with RCD (*P* value = 3.94E-04). Although the association was not significant based on the analysis of the present study, AX-110520028 is physically close to the SNP IWB51724, which has been reported to be significantly associated with seed Cd content in wheat (Quraishi, 2020). For SCD, two significant SNP loci were found on chromosome 5B (Tables 3 and S3). Actually, a number of previous studies have reported Cd accumulation-related variations on chromosome 5B (Penner *et al.*, 1995; Knox *et al.*, 2009; Wiebe *et al.*, 2010; AbuHammad *et al.*, 2016; Oladzad-Abbasabadi *et al.*, 2018; Salsman *et al.*, 2018; Maccaferri *et al.*, 2019). In the present study, a significant MTA (SNP marker AX-108775210) was detected in the region of *Cdu-B1*, explaining a 24.1% phenotypic variation of SCD. As reported *Cdu-B1* is a major QTL with a physical interval of 4.27 Mb controlling Cd uptake in wheat (Knox *et al.*, 2009; Wiebe *et al.*, 2010). Another MTA (SNP marker AX-108933521, *P* value = 1.30E-04) associated with SCD is close to the gene TraesCS7B02G320100 which is one copy of *TaHMA2* on chromosome 7B. It has been reported that *TaHMA2* encodes a P1B-ATPase and is involved in long-distance transport of Zn/ Cd in transgenic rice materials (Tan *et al.*, 2013). We therefore speculated that the SNPs

AX-110520028 and AX-108933521 may be of interest in the future as controlling loci associated with wheat shoot Cd uptake.

For agronomic traits like RL, SL, RDW and SDW, only one significant MTA (SNP marker AX-95008359) was identified in this study. The SNP marker AX-95008359 was associated with SL and explained 17.7% of the phenotypic variation. This SNP is located inside the gene TraesCS3D02G424500, which encodes an exostosin domain-containing protein. To our knowledge, the function of TraesCS3D02G424500 under Cd stress remains uncharacterized, and this locus is not located in any reported QTLs controlling Cd accumulation or tolerance in wheat. Further studies may be required to examine the function of this locus and the adjacent genetic regions of wheat variety of practical importance.

Allelic frequency and additive effect of superior alleles

Distribution of superior and inferior alleles for TF was analyzed to identify the relationship between these alleles on genotype averaged (Fig. 2A). Here, alleles associated with lower TF, SCD, and RCD were defined as superior ones, which are key traits of elite low-Cd wheat. In general, 68.2% of the wheat cultivars (90 of 132) carry three to five superior alleles, while 73.5% of the wheat cultivars (97 of 132) carry three or four inferior alleles, at the identified significant loci (Fig. 2A). Most of the wheat cultivars carry a number of both superior and inferior alleles. In conventional wheat breeding aiming for traits like yield and nutrient use efficiency, there have been cultivars with a cluster of superior alleles and fewer inferior alleles after long-term pyramiding (Zhang *et al.*, 2020). Therefore, low-Cd wheat breeding using the population of this study may require a longer term than expected for the selection and pyramiding of superior alleles, particular for Cd TF.

Fig. 2 Distribution and allelic effect of superior and inferior alleles at significant loci associated with TFs.

Pyramiding of superior alleles is a natural process for plant evolution to enhance the biotic/ abiotic stress tolerance and is also an effective strategy for breeding improvement (Zhou, 2011; Mirabella *et al.*, 2016; Zheng *et al.*, 2017). Here, a change of average TF values against the number of superior/ inferior alleles harboring by the 132 wheat cultivars was calculated to estimate the additive effect of superior/ inferior alleles (Fig. 2B). In general, both superior and inferior alleles revealed distinct additive effects. Superior alleles assembling gradually decreased the TF value ($R^2 = 0.67$). When two to four superior alleles were pyramided, the average value of TF was halved relative to that of one superior allele (Fig. 2B). It is noted that five cultivars (Yannong 836, Yannong 0428, Jimai 21, Chang 4640 and Zhengmai 004), with an average TF value of 0.04 ± 0.01 , were found to carry eight superior alleles for Cd TF (maximum pyramiding, data not shown). Conversely, the pyramiding of inferior alleles increased TF value additively ($R^2 = 0.86$). The pyramiding effect tended to get substantially significant when the number of inferior alleles reached four or more (Fig. 2B). Taken together, additive effect analysis further verified the robustness of the GWAS identification of SNPs associated with the wheat Cd. The results here may be of importance in low-Cd wheat breeding in terms of Cd TF.

Positive effects on TF were observed for all superior alleles at significant loci (Fig. 3). It is worth noting that at the most significant locus (SNP marker AX-108795510), the superior allele AA resulted in a 72.8% reduction on TF, compared with the allele AT (Fig. 3). Further exploration indicated that 13 out of the 132 wheat cultivars carry this allele, which can be considered for the low-Cd breeding in the future.

Fig. 3 Significant effects of favorable alleles on the Cd translocation factor in the studied association mapping panel

Wheat cultivars of practical importance in farmland Cd pollution control

Based on the above analysis of superior alleles and TF, all wheat cultivars were evaluated and dissected into different subgroups (Table 4). Five cultivars (CG15-009, Yannong 836, Yannong 0428, Jimai21 and Zhengmai 004), with an average TF value of 0.03, harbored seven or eight superior alleles, but carry only one or two inferior alleles. In contrast, five other cultivars (Jimai418, Hen 5835, Hen 1589, Cangmai 6055 and Liangxing 99) had an average TF value of 0.25 and carried six or seven inferior alleles (Table 4). A previous study demonstrated that Yannong 0428 was a low accumulator in grain Cd based on the results of two-year field experiments (Xia, 2018). A hydroponic experiment also showed that Yannong 0428 had a relatively lower TF (Fan, 2018). Thus, the cultivars with low Cd TFs identified in the current study, particularly Yannong 0428, can be considered as candidate materials for low-Cd breeding.

TABLE 4
Dissection of the diversity panel based on Cd translocation factor

Cultivar	Origin	Subgroup	TF	No. of superior alleles	No. of inferior alleles
CG15-009	Hebei	Low TF	0.03	7	2
Yannong 836	Shandong	Low TF	0.04	8	1
Yannong 0428	Shandong	Low TF	0.03	8	2
Jimai21	Shandong	Low TF	0.04	8	1
Zhengmai 004	Henan	Low TF	0.03	8	1
Jimai418	Hebei	High TF	0.26	3	6
Hen 5835	Hebei	High TF	0.22	1	6
Hen 1589	Hebei	High TF	0.32	1	6
Cangmai 6055	Hebei	High TF	0.27	0	7
Liangxing 99	Shandong	High TF	0.19	2	6

The high TF cultivar Liangxing 99 has been reported as a high grain Cd accumulator (Xiong *et al.*, 2018). This result is in accordance with our observation, in which Liangxing 99 showed a high TF of 0.19. High TF cultivars like Liangxing 99 should be deliberately rejected in low Cd breeding. Further comparison using functional genomics, transcriptomics or proteomics approaches may be necessary to elucidate key genes controlling Cd uptake and translocation of wheat in the future.

CONCLUSIONS

Cadmium responding phenotypic variations were observed among the 132 wheat cultivars collected from the North China Plain, and 17 significantly associated novel SNPs were identified in wheat genome via GWAS. These associated SNPs closely linked to the traits might be key markers in marker-assisted breeding of low-Cd wheat cultivars. The cultivars carrying the maximum number of superior alleles could be used as elite germplasm resources for low-Cd breeding. Future work may focus on assembling all superior alleles in wheat genomes by pyramiding or substitution breeding.

ACKNOWLEDGEMENT(S)

We thank Prof. Junming LI and Dr. Jun JI who provided germplasms and valuable comments on this study. This work was supported by the Hebei Science Fund for Distinguished Young Scholars [D2018503005]; the National Key Research and Development Program of China [2018YFD0800306]; and the CAS President's International Fellowship Initiative (PIFI) fund [2018PB0084].

CONTRIBUTION OF AUTHORS

Imdad Ullah ZAID, Sajid Hanif MUHAMMAD, and Na ZHANG contributed equally to this work.

SUPPLEMENTARY MATERIAL

Supplementary material for this article can be found in the online version.

REFERENCES

- AbuHammad W A, Mamidi S, Kumar A, Pirseyedi S, Manthey F A, Kianian S F, Alamri M S, Mergoum M, Elias E M. 2016. Identification and validation of a major cadmium accumulation locus and closely associated SNP markers in North Dakota durum wheat cultivars. *Mol Breeding*. **36**.
- Ahmad I, Akhtar M J, Asghar H N, Zahir Z A. 2013. Comparative efficacy of growth media in causing cadmium toxicity to wheat at seed germination stage. *Int J Agric Biol*. **15**: 517-522.
- Ahmad I, Akhtar M J, Zahir Z A, Jamil A. 2012. Effect of cadmium on seed germination and seedling growth of four wheat (*Triticum aestivum* L.) cultivars. *Pak J Bot*. **44**: 1569-1574.
- Ai J H, Liao X Y, Wang L Q, Zhao F H, Xue T, Gong X G, Zhao Y S. 2019. The selection of low cadmium accumulation wheat varieties under cadmium stress. *Journal of Nanchang University (Natural Science)*. **43**: 175-181.
- Arora S, Singh N, Kaur S, Bains N S, Uauy C, Poland J, Chhuneja P. 2017. Genome-wide association study of grain architecture in wild wheat *Aegilops tauschii*. *Front Plant Sci*. **8**.
- Ceserani T, Trofka A, Gandotra N, Nelson T. 2009. VH1/BRL2 receptor-like kinase interacts with vascular-specific adaptor proteins VIT and VIK to influence leaf venation. *Plant J*. **57**: 1000-1014.
- Chen Y R, Zhang Q F, Fu B S, Cai S B, Wu J Z, Chen Y H. 2017. Differences of lead, cadmium and zinc accumulation among chinese wheat mini-core collections germplasms and screening for low Pb, Cd and Zn accumulative cultivars in grains. *Journal of Nanjing Agricultural University*. **40**: 393-399.
- Ci D, Jiang D, Dai T, Jing Q, Cao W. 2009. Effects of cadmium on plant growth and physiological traits in contrast wheat recombinant inbred lines differing in cadmium tolerance. *Chemosphere*. **77**: 1620-1625.
- Ci D W, Jiang D, Li S S, Wollenweber B, Dai T B, Cao W X. 2012. Identification of quantitative trait loci for cadmium tolerance and accumulation in wheat. *Acta Physiol Plant*. **34**: 191-202.
- Clarke J M, Leisle D, Kopytko G L. 1997. Inheritance of cadmium concentration in five durum wheat crosses. *Crop Sci*. **37**: 1722-1726.

- Fan H J, Li G K, Luan W, Cai P X. 2005. [determination of cadmium, chromium and lead in *Lycoris radiata* with microwave digestion technique by graphite furnace atomic absorption spectrometry]. *Guang Pu Xue Yu Guang Pu Fen Xi*. **25**: 1503-1506.
- Fan J F. 2018. Genotypic differences of wheat in response to selenium, lead and cadmium stress *Huazhong agricultural university*.
- Gomez K A, Gomez A A, Gomez K A. 1984. Statistical procedures for agricultural research. 2nd. Wiley, New York.
- Gordon D, Huddleston J, Chaisson M J, Hill C M, Kronenberg Z N, Munson K M, Malig M, Raja A, Fiddes I, Hillier L W, Dunn C, Baker C, Armstrong J, Diekhans M, Paten B, Shendure J, Wilson R K, Haussler D, Chin C S, Eichler E E. 2016. Long-read sequence assembly of the gorilla genome. *Science*. **352**: aae0344.
- Greger M, Lofstedt M. 2003. Comparison of uptake and distribution of cadmium in different cultivars of bread and durum wheat. *Crop Sci*. **44**: 501-507.
- Guttieri M J, Baenziger P S, Frels K, Carver B, Arnall B, Wang S C, Akhunov E, Waters B M. 2015. Prospects for selecting wheat with increased zinc and decreased cadmium concentration in grain. *Crop Sci*. **55**: 1712-1728.
- Hanson C H, Robinson H F, Comstock R E. 1956. Biometrical studies of yield in segregating populations of korean lespedeza1. *Agronomy Journal*. **48**: 268-272.
- Khan N A, Singh S S, Nazar R. 2007. Activities of antioxidative enzymes, sulphur assimilation, photosynthetic activity and growth of wheat (*Triticum aestivum*) cultivars differing in yield potential under cadmium stress. *J Agron Crop Sci*. **193**: 435-444.
- Knox R E, Pozniak C J, Clarke F R, Clarke J M, Houshmand S, Singh A K. 2009. Chromosomal location of the cadmium uptake gene (*Cdu1*) in durum wheat. *Genome*. **52**: 741-747.
- Lei G J, Fujii-Kashino M, Wu D Z, Hisano H, Saisho S, Deng F L, Yamaji N, Sato K, Zhao F J, Ma J F. 2020. Breeding for low cadmium barley by introgression of a sukkula-like transposable element. *Nature Food*. **1**: 489-499.
- Li L L, Liu Y, Li B G, Hu C, Fan T, Zhao Z J, Wu D F, Wang D, Li Z Y. 2019. Screening of low-accumulation wheat varieties with cadmium. *Journal of Irrigation and Drainage*. **38**: 53-58.
- Li X F. 2019. Technical solutions for the safe utilization of heavy metal-contaminated farmland in China: A critical review. *Land Degrad Dev*. **30**: 1773-1784.
- Li X F, Zhou D M. 2019. A meta-analysis on phenotypic variation in cadmium accumulation of rice and wheat: Implications for food cadmium risk control. *Pedosphere*. **29**: 545-553.
- Liu J, Luo W, Qin N, Ding P, Zhang H, Yang C, Mu Y, Tang H, Liu Y, Li W, Jiang Q, Chen G, Wei Y, Zheng Y, Liu C, Lan X, Ma J. 2018. A 55 k SNP array-based genetic map and its utilization in QTL mapping for productive tiller number in common wheat. *Theor Appl Genet*. **131**: 2439-2450.
- Lu C N, Zhang L X, Tang Z, Huang X Y, Ma J F, Zhao F J. 2019. Producing cadmium-free indica rice by overexpressing *OsHMA3*. *Environ Int*. **126**: 619-626.
- Ma F F, Xu Y F, Ma Z Q, Li L H, An D G. 2018. Genome-wide association and validation of key loci for yield-related traits in wheat founder parent xiaoyan 6. *Mol Breeding*. **38**.
- Ma J F, Shen R F, Shao J F. 2020. Transport of cadmium from soil to grain in cereal crops: A review. *Pedosphere*. 1-13.
- Maccaferri M, Harris N S, Twardziok S O, Pasam R K, Gundlach H, Spannagl M, Ormanbekova D, Lux T, Prade V M, Milner S G, Himmelbach A, Mascher M, Bagnaresi P, Faccioli P, Cozzi P, Lauria M, Lazzari B, Stella A, Manconi A, Gnocchi M, Moscatelli M, Avni R, Deek J, Biyikliglu S, Frascaroli E, Corneti S, Salvi S, Sonnante G, Desiderio F, Mare C, Crosatti C,

- Mica E, Ozkan H, Kilian B, De Vita P, Marone D, Joukhadar R, Mazzucotelli E, Nigro D, Gadaleta A, Chao S M, Faris J D, Melo A T O, Pumphrey M, Pecchioni N, Milanese L, Wiebe K, Ens J, MacLachlan R P, Clarke J M, Sharpe A G, Koh C S, Liang K Y H, Taylor G J, Knox R, Budak H, Mastrangelo A M, Xu S S, Stein N, Hale I, Distelfeld A, Hayden M J, Tuberosa R, Walkowiak S, Mayer K F X, Ceriotti A, Pozniak C J, Cattivelli L. 2019. Durum wheat genome highlights past domestication signatures and future improvement targets. *Nat Genet.* **51**: 885-+.
- Ministry of Environmental Protection of China, Ministry of Land and Resources of China. 2014. Report on the national general survey of soil contamination.
- Mirabella N E, Abbate P E, Ramirez I A, Pontaroli A C. 2016. Genetic variation for wheat spike fertility in cultivars and early breeding materials. *J Agr Sci-Cambridge.* **154**: 13-22.
- Myles S, Peiffer J, Brown P J, Ersoz E S, Zhang Z W, Costich D E, Buckler E S. 2009. Association mapping: Critical considerations shift from genotyping to experimental design. *Plant Cell.* **21**: 2194-2202.
- Naeem A, Saifullah, Rehman M Z U, Akhtar T, Ok Y S, Rengel Z. 2016. Genetic variation in cadmium accumulation and tolerance among wheat cultivars at the seedling stage. *Commun Soil Sci Plan.* **47**: 554-562.
- Oladzaad-Abbasabadi A, Kumar A, Pirseyedi S, Salsman E, Dobrydina M, Poudel R S, AbuHammad W A, Chao S A M, Faris J D, Elias E M. 2018. Identification and validation of a new source of low grain cadmium accumulation in durum wheat. *G3-Genes Genom Genet.* **8**: 923-932.
- Oliver D P, Gartrell J W, Tiller K G, Correll R, Cozens G D, Youngberg B L. 1995. Differential responses of australian wheat cultivars to cadmium concentration in wheat-grain. *Aust J Agr Res.* **46**: 873-886.
- Oono Y, Yazawa T, Kanamori H, Sasaki H, Mori S, Handa H, Matsumoto T. 2016. Genome-wide transcriptome analysis of cadmium stress in rice. *Biomed Res Int.* **2016**.
- Ouzounidou G, Moustakas M, Eleftheriou E P. 1997. Physiological and ultrastructural effects of cadmium on wheat (*Triticum aestivum* L) leaves. *Arch Environ Con Tox.* **32**: 154-160.
- Pariyar S R, Dababat A A, Sannemann W, Erginbas-Orakci G, Elashry A, Siddique S, Morgounov A, Leon J, Grundler F M W. 2016. Genome-wide association study in wheat identifies resistance to the cereal cyst nematode *Heterodera filipjevi*. *Phytopathology.* **106**: 1128-1138.
- Penner G A, Clarke J, Bezte L J, Leisle D. 1995. Identification of rapid markers linked to a gene governing cadmium uptake in durum-wheat. *Genome.* **38**: 543-547.
- Pritchard J K, Stephens M, Rosenberg N A, Donnelly P. 2000. Association mapping in structured populations. *Am J Hum Genet.* **67**: 170-181.
- Quraishi L B S F A S S M E Z A Z M L Y M M T M I S L U M. 2020. Genome-wide association study identifies five new cadmium uptake loci in wheat. *The Plant Genome.* **13**: e20030.
- Rai P K, Lee S S, Zhang M, Tsang Y F, Kim K H. 2019. Heavy metals in food crops: Health risks, fate, mechanisms, and management. *Environ Int.* **125**: 365-385.
- Rife T W, Wu S, Bowden R L, Poland J A. 2015. Spiked GBS: A unified, open platform for single marker genotyping and whole-genome profiling. *Bmc Genomics.* **16**: 248.
- Rizwan M, Meunier J D, Davidian J C, Pokrovsky O S, Bovet N, Keller C. 2016. Silicon alleviates Cd stress of wheat seedlings (*Triticum turgidum* L. cv. Claudio) grown in hydroponics. *Environ Sci Pollut R.* **23**: 1414-1427.
- Robert W A. 1999. Principles of plant breeding, 2nd edition. *John Wiley & Sons, Inc.*
- Salsman E, Kumar A, AbuHammad W, Abbasabadi A O, Dobrydina M, Chao S M, Li X H, Manthey F A, Elias E M. 2018. Development and validation of molecular markers for grain cadmium in durum wheat. *Mol Breeding.* **38**.

- Sun H X, Xue P Y, Zhao Q L, Yang Z Z, Y. Y, Feng Y J, Liu F, Tang T C, Liu W J. 2015. Differences of cadmium and lead accumulation and transportation among winter wheat varieties. *Journal of Triticeae Crops*. **35**: 1161-1167.
- Talukdar P R, Rathi S, Pathak K, Chetia S K, Sarma R N. 2017. Population structure and marker-trait association in indigenous aromatic rice. *Rice Sci*. **24**: 145-154.
- Tan J, Wang J, Chai T, Zhang Y, Feng S, Li Y, Zhao H, Liu H, Chai X. 2013. Functional analyses of TaHMA2, a P(1B)-type ATPase in wheat. *Plant Biotechnol J*. **11**: 420-431.
- Tanhuanpaa P, Kalendar R, Schulman A H, Kiviharju E. 2007. A major gene for grain cadmium accumulation in oat (*Avena sativa* L.). *Genome*. **50**: 588-594.
- Ueda Y, Frimpong F, Qi Y T, Matthus E, Wu L B, Holler S, Kraska T, Frei M. 2015. Genetic dissection of ozone tolerance in rice (*Oryza sativa* L.) by a genome-wide association study. *J Exp Bot*. **66**: 293-306.
- Vassilev A, Lidon F C, Ramalho J C, Matos M D, Bareiro M G. 2004. Shoot cadmium accumulation and photosynthetic performance of barley plants exposed to high cadmium treatments. *J Plant Nutr*. **27**: 775-795.
- Wang C C, Tang Z, Zhuang J Y, Tang Z, Huang X Y, Zhao F J. 2020. Genetic mapping of ionic quantitative trait loci in rice grain and straw reveals *Osmot1;1* as the putative causal gene for a molybdenum QTL *qMo8*. *Mol Genet Genomics*. **295**: 391-407.
- Wiebe K, Harris N S, Faris J D, Clarke J M, Knox R E, Taylor G J, Pozniak C J. 2010. Targeted mapping of *Cdu1*, a major locus regulating grain cadmium concentration in durum wheat (*Triticum turgidum* L. var durum). *Theor Appl Genet*. **121**: 1047-1058.
- Winfield M O, Wilkinson P A, Allen A M, Barker G L, Coghill J A, Burrige A, Hall A, Brenchley R C, D'Amore R, Hall N, Bevan M W, Richmond T, Gerhardt D J, Jeddloh J A, Edwards K J. 2012. Targeted re-sequencing of the allohexaploid wheat exome. *Plant Biotechnol J*. **10**: 733-742.
- Wu D Z, Sato K, Ma J F. 2015. Genome-wide association mapping of cadmium accumulation in different organs of barley. *New Phytol*. **208**: 817-829.
- Xia Y T. 2018. Screening of high and low cadmium accumulation *Triticum aestivum* varieties and research on the difference and mechanism of cadmium absorption and transport *Huazhong agricultural university*.
- Xiong Z, Li J M, Zhao H W, Ma Y B. 2018. Accumulation and translocation of cadmium in different wheat cultivars in farmland. *Journal of Agro-Environment Science*. **37**: 36-44.
- Xue D W, Chen M C, Zhang G P. 2009. Mapping of qtls associated with cadmium tolerance and accumulation during seedling stage in rice (*Oryza sativa* L.). *Euphytica*. **165**: 587-596.
- Yu J M, Pressoir G, Briggs W H, Bi I V, Yamasaki M, Doebley J F, McMullen M D, Gaut B S, Nielsen D M, Holland J B, Kresovich S, Buckler E S. 2006. A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat Genet*. **38**: 203-208.
- Zaid I U, Zheng X, Li X F. 2018. Breeding low-cadmium wheat: Progress and perspectives. *Agronomy-Basel*. **8**.
- Zhang L, Gao C, Chen C, Zhang W, Huang X Y, Zhao F J. 2020. Overexpression of rice *OsHMA3* in wheat greatly decreases cadmium accumulation in wheat grains. *Environ Sci Technol*. **54**: 10100-10108.
- Zhang N, Zhang X L, Song L Q, Su Q N, Zhang S, Liu J J, Zhang W, Fu X Y, Zhao M H, Sun L J, Ji J, Qiao W C, Li H, He M Q, Li J M. 2020. Identification and validation of the superior alleles for wheat kernel traits detected by genome-wide association study under different nitrogen environments. *Euphytica*. **216**.
- Zhang X, Chen D, Zhong T, Zhang X, Cheng M, Li X. 2015. Assessment of cadmium (Cd) concentration in arable soil in China. *Environ Sci Pollut Res Int*. **22**: 4932-4941.

- Zhang X Y, Zhong T Y, Liu L, Ouyang X Y. 2015. Impact of soil heavy metal pollution on food safety in China. *Plos One*. **10**.
- Zhang Z W, Ersoz E, Lai C Q, Todhunter R J, Tiwari H K, Gore M A, Bradbury P J, Yu J M, Arnett D K, Ordovas J M, Buckler E S. 2010. Mixed linear model approach adapted for genome-wide association studies. *Nat Genet*. **42**: 355-U118.
- Zhao K, Aranzana M J, Kim S, Lister C, Shindo C, Tang C, Toomajian C, Zheng H, Dean C, Marjoram P, Nordborg M. 2007. An *Arabidopsis* example of association mapping in structured samples. *PLoS Genet*. **3**: e4.
- Zhao X W, Luo L X, Cao Y H, Liu Y J, Li Y H, Wu W M, Lan Y Z, Jiang Y W, Gao S B, Zhang Z M, Shen Y O, Pan G T, Lin H J. 2018. Genome-wide association analysis and QTL mapping reveal the genetic control of cadmium accumulation in maize leaf. *Bmc Genomics*. **19**.
- Zheng Z, Gao S, Zhou M X, Yan G J, Liu C J. 2017. Enhancing fusarium crown rot resistance by pyramiding large-effect QTL in common wheat (*Triticum aestivum* L.). *Mol Breeding*. **37**.
- Zhou M. 2011. Accurate phenotyping reveals better QTL for waterlogging tolerance in barley. *Plant Breeding*. **130**: 203-208.

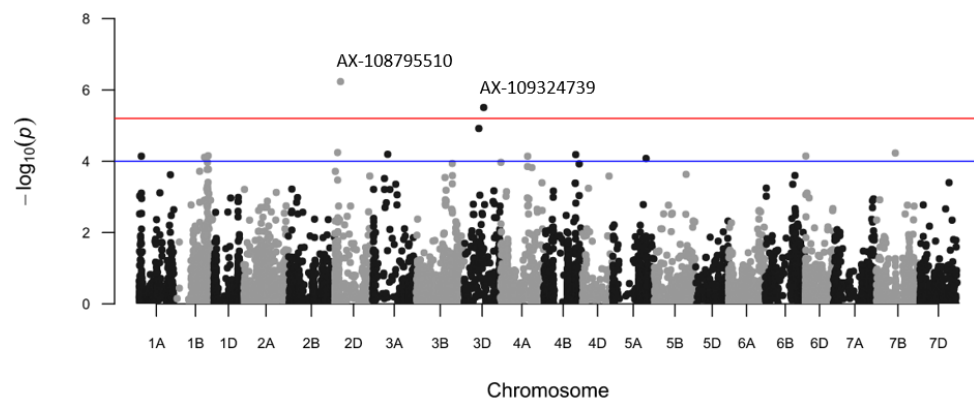


Fig. 1

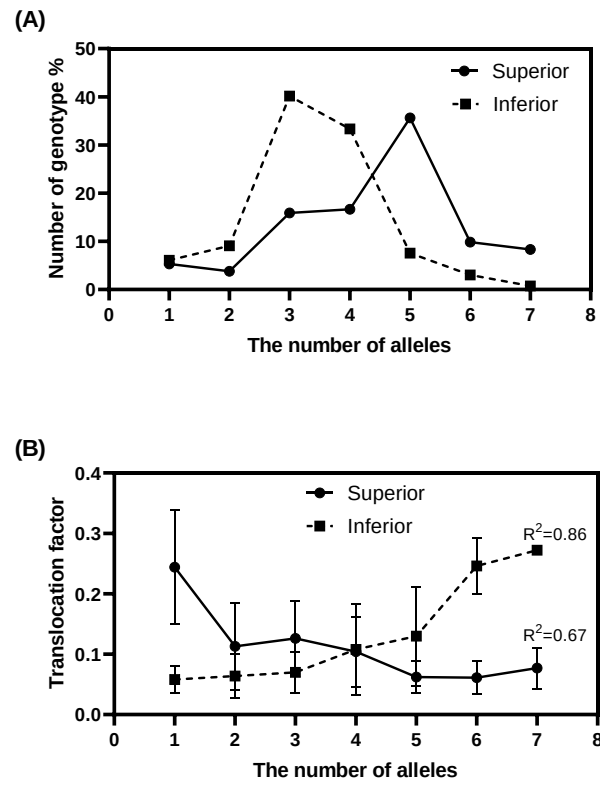


Fig. 2

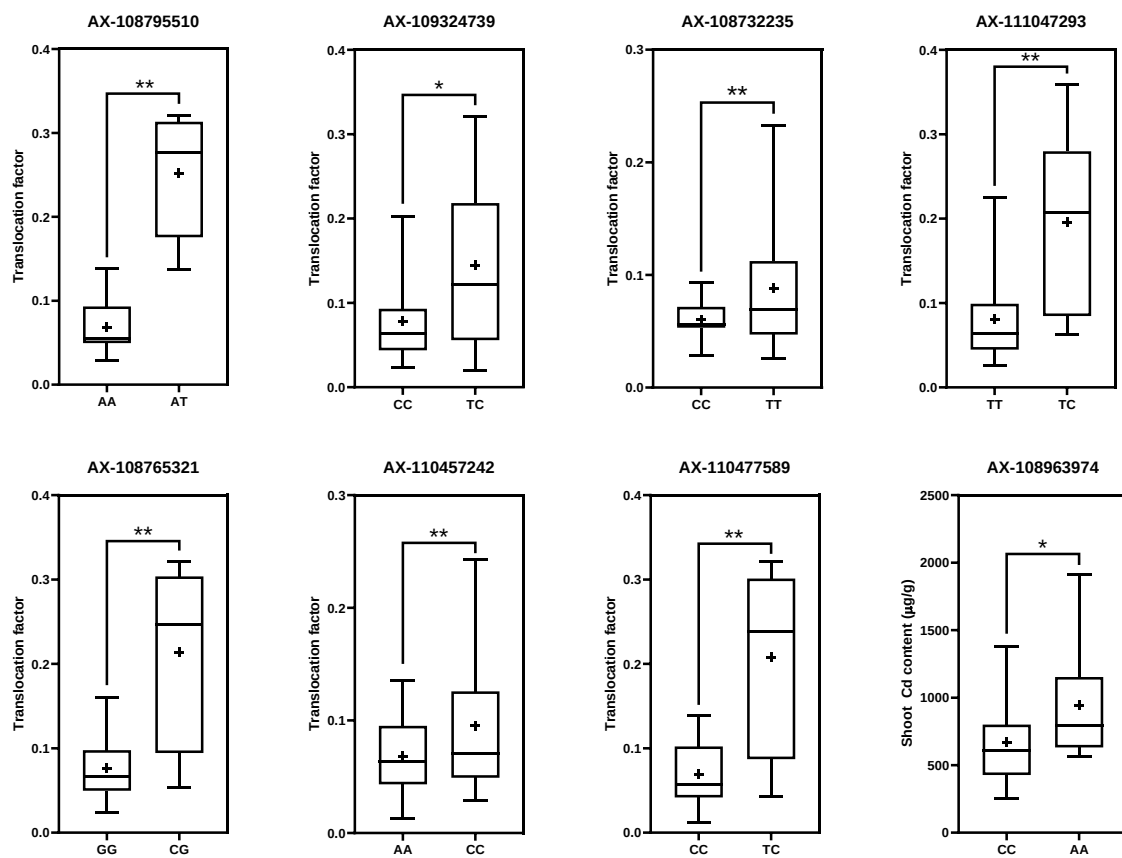


Fig. 3