Natural variations in the P-type ATPase heavy metal transporter ZmCd1 controlling cadmium accumulation in maize grains

Bin Tang^{1,†}, Meijie Luo^{2,†}, Yunxia Zhang², Huanle Guo¹, Jingna Li², Wei Song², Ruyang Zhang², Zhen Feng², Mengsi Kong², Han Li¹, Zhongyang Cao¹, Xiaoduo Lu³, Delin Li^{4,5}, Jianhua Zhang⁶, Ronghuan Wang², Yuandong Wang², Zhihui Chen¹, Yanxin Zhao^{2,*}, Jiuran Zhao^{2,*}.

¹Crop Research Institute, Hunan Academy of Agricultural Sciences, Changsha 410125, China.

²Beijing Key Laboratory of Maize DNA Fingerprinting and Molecular Breeding, Maize Research Center, Beijing Academy of Agriculture and Forestry Sciences (BAAFS), Beijing 100097, China.

³Institute of Molecular Breeding for Maize, Qilu Normal University, Jinan 250200, Shandong, China.

⁴Department of Plant Genetics and Breeding, China Agricultural University, Beijing, 100193, China.

⁵The National Key Facility for Crop Gene Resources and Genetic Improvement (NFCRI)/Key Lab of Germplasm Utilization (MOA), Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing, 100081, China.

⁶Genetron Health (Beijing) Co. Ltd, Beijing, 102208, China.

⁷National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan 430070, China.

[†] Bin Tang and Meijie Luo contributed equally to this work.

*Correspondence: rentlang2003@163.com (Yanxin Zhao); maizezhao@126.com (Jiuran Zhao)

1 **Summary**

2 Cadmium (Cd) accumulation in maize grains is detrimental to human health. Developing
 3 maize varieties with low-Cd contents via marker-assisted selection is important for ensuring
 4 the production of maize grains safe for consumption. However, the key gene controlling
 5 maize grain Cd accumulation has not been cloned. In this study, we identified two major loci
 6 for maize grain Cd accumulation (*qCd1* and *qCd2*) on chromosome 2 during a genome-wide
 7 association study (GWAS). The *qCd1* locus was analyzed by bulked segregant RNA-seq and
 8 fine mapping with a biparental segregating population of Jing724 (low-Cd line) and Mo17
 9 (high-Cd line). The *ZmCd1* candidate gene in the *qCd1* locus encodes a vacuolar
 10 membrane-localized heavy metal P-type ATPase transporter, ZmHMA3, which is orthologous
 11 to the tonoplast Cd transporter OsHMA3. Genomic DNA sequence and transcript analyses
 12 suggested that a transposon in intron 1 of *ZmCd1* is responsible for the abnormal amino acid
 13 sequence in Mo17. An EMS mutant analysis and an allelism test confirmed *ZmCd1* influences
 14 maize grain Cd accumulation. The natural variations in *ZmCd1* were used to develop four
 15 PCR-based molecular markers, which revealed five *ZmCd1* haplotypes in the GWAS
 16 population. The molecular markers were also used to predict the grain Cd contents in
 17 commonly cultivated maize germplasms in China. The predicted Cd contents for 36 inbred
 18 lines and 13 hybrids were consistent with the measured Cd contents. Furthermore, several
 19 low-Cd elite inbred lines and hybrids were identified, including Jing2416, MC01,
 20 Jingnonke728, and Jingke968. Therefore, the molecular markers developed in this study are
 21 applicable for molecular breeding and developing maize varieties with low grain Cd contents.

22
 23 **Keywords:** cadmium accumulation, maize grain, *ZmCd1*, haplotype, molecular markers

24

25

26

27

28

29

30

1   **Introduction**

2   Cadmium (Cd) is one of the most deleterious heavy metals. It can accumulate in the food
3   chain, but even a very low Cd concentration can be fatal to humans because of its long
4   half-life (Yan et al., 2019). Specifically, Cd can damage bones, the liver, kidneys, and lungs,
5   and even cause cancer (Zhang et al., 2019). Unfortunately, Cd has been widely released into
6   soils because of increasing industrial and agricultural activities. Approximately 17% of the
7   total cultivated land in China is contaminated by Cd. Soil Cd pollution has become a serious
8   global issue (Li et al., 2018).

9   Maize is an important agricultural crop worldwide. It is the predominant cereal in terms
10   of planting area and global production. Maize makes a critical contribution to world food
11   security and it is the main staple food in Africa and South America (Li et al., 2019). Cadmium
12   in the soil can be easily absorbed by maize roots and then accumulate in plants and grains.
13   Therefore, elucidating the molecular mechanism underlying Cd accumulation in maize has
14   important implications for breeding maize varieties that accumulate low Cd levels, which will
15   be useful for minimizing the Cd intake by humans.

16   Plants have evolved a series of regulatory mechanisms that protect them from Cd stress,
17   including the following: (1) expelling or preventing the uptake of Cd by cells to minimize Cd
18   accumulation (Cao et al., 2019; Yan et al., 2019; Zhang et al., 2019; Zhao et al., 2018a); (2)
19   decreasing the Cd concentration in cells via chelation and complexation (Luo et al., 2018;
20   Zhang et al., 2018c); and (3) increasing antioxidant enzyme activities to eliminate reactive
21   oxygen species in cells (Gill and Tuteja, 2011; Rizwan et al., 2016). Among the identified Cd
22   tolerance-related genes, those encoding transporter proteins have critical effects on the Cd
23   accumulation in plants (Chen et al., 2019). The P_{1B}-type heavy metal ATPases (HMAs),
24   belonging to the P-type ATPase superfamily, are important for Cd transmembrane transport
25   (Chen et al., 2019; Pedersen et al., 2012; Qiao et al., 2018; Takahashi et al., 2012a). Previous
26   studies revealed that functionally diverse *HMA* genes encode proteins that vary regarding
27   subcellular localization and tissue distribution. In *Arabidopsis thaliana* (Arabidopsis),
28   AtHMA2 and AtHMA4 are localized in the plasma membrane and can promote Cd efflux
29   from cells (Verret et al., 2004). Both *AtHMA2* and *AtHMA4* are primarily expressed in tissues
30   surrounding the root vascular vessels and affect Cd root-to-shoot translocation (Verret et al.,

2004). Additionally, *AtHMA3* is localized in the vacuolar membrane, where it is responsible for detoxifying Cd through vacuolar sequestration (Morel et al., 2009). The *AtHMA3* gene is highly expressed in various tissues and cells, including the root apex, vascular tissues, guard cells, and hydathodes. The overexpression of *AtHMA3* can induce Cd accumulation in *Arabidopsis* shoots and roots (Morel et al., 2009). Similarly, *OshMA2* in the rice plasma membrane transports Cd out of cells (Takahashi et al., 2012b), whereas *OshMA3* functions in the vacuolar membrane to help sequester Cd in vacuoles (Miyadate et al., 2011; Ueno et al., 2010; Zhang et al., 2020). Although some genes regulating Cd accumulation have been functionally characterized in *Arabidopsis* and rice, the molecular mechanism regulating plant Cd tolerance remains unclear, and the key genes controlling Cd accumulation in maize have not been cloned.

The Cd accumulation level reportedly varies significantly among maize lines with differing genetic backgrounds (Zhao et al., 2018a). Applying molecular breeding technology to select and develop maize varieties that minimally accumulate Cd is an economically feasible and effective way to ensure the sustainable production of maize that is safe for human consumption (Jha and Bohra, 2016). In this study, we measured the maize grain Cd contents of an association panel comprising 513 lines. The results of a genome-wide association study (GWAS), bulked segregant RNA sequencing (BSR-seq), and fine mapping were combined to dissect the genetic loci controlling Cd accumulation in maize grains. By applying multiple approaches, including a gene functional annotation analysis, a sequence variation analysis, a transcript analysis, and an allelism test of the mutants, we revealed that a retrotransposon insertion in the *ZmCd1* gene affects the accumulation of Cd in maize grains. Phylogenetic and subcellular localization analyses indicated that *ZmCd1* might encode a P_{1B}-type HMA transporter. On the basis of the sequence differences in the *ZmCd1* promoter, intron 1, intron 4, and exon 5 between maize lines that accumulate Cd at high and low levels, four PCR-based molecular markers were developed and five *ZmCd1* haplotypes were detected in maize lines. The predicted maize grain Cd accumulation in widely used inbred lines and hybrids in China based on molecular markers was consistent with the measured Cd contents. Therefore, *ZmCd1* is an important new target gene related to maize grain Cd accumulation. Moreover, allele-specific PCR markers may be applicable for breeding maize lines with low Cd levels.

Results

Differences in the Cd contents of 513 maize accessions

We measured the Cd contents of dry mature maize grains from 513 inbred lines of an association population harvested in two planting seasons (spring and autumn) in two fields [Zhuzhou (ZZ) and Ningxiang (NX)]. Examinations of the soil in the NX and ZZ fields (0–20 cm depth) revealed that the soil Cd content in ZZ (2.01–2.24 mg/kg) was higher than that in NX (1.69–1.85 mg/kg), and the ZZ field (pH 5.68–6.08) was more acidic than the NX field (pH 7.51–8.14) (**Figure 1a, b**). The maize grain Cd contents were higher in ZZ than in NX (**Figure 1c**), possibly because of the higher soil Cd content and lower soil pH in ZZ than in NX (Ali et al., 2019). The average maize grain Cd contents were 0.1955 and 0.2436 mg/kg in ZZ in the spring (ZZC) and autumn (ZZQ), respectively, whereas they were 0.0690 and 0.0457 mg/kg in NX in the spring (NXC) and autumn (NXQ), respectively. A combined analysis of the data for the four environments using the best linear unbiased prediction (BLUP) method determined the average maize grain Cd content was 0.1355 mg/kg (**Figure 1c; Table 1**). The data for both locations and the BLUP data indicated the maize grain Cd content varied substantially in the association population (**Table 1; Figure S1**). For example, the maize grain Cd content based on the combined analysis of all four environments ranged from 0.0358 to 0.7325 mg/kg (**Table 1**). The Pearson correlation coefficient reflected a moderately strong relationship among the maize grain Cd contents of the four environments ($r = 0.54\text{--}0.68$) (**Figure 1d**). An analysis of variance (ANOVA) revealed significant variations in the maize grain Cd contents based on the genotype, environment, and the genotype $\times$ environment interaction. The broad-sense heritability for the maize grain Cd content was as high as 71.11% (**Table 2**).

Loci associated with maize grain Cd contents identified by a GWAS

A GWAS was performed using 1.25 M high-quality single nucleotide polymorphisms (SNPs) with a minor allele frequency greater than 0.05 and an association panel containing 513 maize inbred lines (Yang et al., 2014). A mixed linear model incorporating the population structure and relative kinship as well as the Cd concentration data for each environment were used. The BLUP data for the Cd concentration in two seasons at each location (BLUPZZ and BLUPNX) and the BLUP data for the Cd concentration in four environments (BLUP4EN) were also used

1 for the GWAS (**Figure 2a, b; Figure S2**).

2 The linkage disequilibrium (LD) decay for the association population was 100 kb ($r^2 =$
3 0.1) (Liu et al., 2017). Accordingly, the 100-kb region flanking both sides of each significant
4 SNP was defined as a quantitative trait locus (QTL). A total of 55, 44, 34, and 71 QTL related
5 to Cd accumulation were identified ($P \leq 7.97e-7$) for ZZC, ZZQ, NXC, and NXQ,
6 respectively (**Figure S2; Table S1**). Additionally, 68, 52, and 63 QTL were identified in the
7 GWAS of BLUP4EN, BLUPNX, and BLUPZZ, respectively (**Figure S2; Table S1**). The
8 phenotypic variance explained (PVE) for the QTL ranged from 5.24% to 23.78%. Moreover,
9 84.14% of the identified QTL were located on chromosome 2, with a PVE of 5.24%–23.78%.
10 The remaining QTL were on eight other chromosomes, with a PVE of 5.43%–12.04% (**Table**
11 **S1**). The QTL on chromosome 2 spanned a region from 5,468,732 bp to 189,546,923 bp (B73
12 RefGen_v2 reference genome). The PVE for these QTL gradually increased along the
13 chromosome, but decreased after peaking. The QTL with the highest PVE were between
14 157,504,643 bp and 158,846,650 bp (PVE of 20.03%) and between 160,628,736 bp and
15 163,940,592 bp (PVE of 23.78%) on chromosome 2. They were detected simultaneously in
16 all environments and in the BLUP data. Therefore, these two genomic regions harbor two
17 major loci related to maize grain Cd accumulation, *qCd1* (peak SNP: chr2.S_158408660) and
18 *qCd2* (peak SNP: chr2.S_161122814).

19 Three candidate genes (*GRMZM2G175576*, *GRMZM2G455491*, and *GRMZM2G018241*)
20 in the two major QTL were proposed based on gene functional annotations (**Table 3**). Both
21 *GRMZM2G175576* (*ZmHMA3*) and *GRMZM2G455491* (*ZmHMA4*) were detected in *qCd1*.
22 These genes are homologous to Arabidopsis *HMA3* and rice *HMA3* genes, which are involved
23 in sequestering Cd in root cell vacuoles. The other candidate gene, *GRMZM2G018241*
24 (*ZmCESA9*), was detected in *qCd2* and encodes a cellulose synthase.

25 **A recessive gene was responsible for high maize grain Cd contents**

26 Among the GWAS population inbred lines, Mo17 (0.4383 mg/kg) and Qi319 (0.4892 mg/kg)
27 had high grain Cd levels, in contrast to the low grain Cd contents of B73 (0.0248 mg/kg) and
28 Huangzaosi (0.0151 mg/kg) (**Figure 3a**). Maize inbred lines Jing724 (low Cd content; 0.0299
29 mg/kg) and D9H (high Cd content; 0.1283 mg/kg) were derived from maize hybrid x1132x,
30 and they have been widely used for hybrid seed production in China. Grains from lines

produced by crossing high-Cd maize lines (Mo17 and Qi319) (female parent) with low-Cd maize lines accumulated large amounts of Cd. Grains from lines generated by crossing low-Cd maize lines (B73, Huangzaosi, and Jing724) with high-Cd maize lines had low Cd levels (**Figure 3b**). These results suggested there was no pollen effect (xenia) on maize grain Cd accumulation.

To analyze the effects of genes regulating Cd accumulation, Mo17 and Qi319 (high-Cd lines) were crossed with Jing724 and B73 (low-Cd lines). The grain Cd contents of the self-pollinated hybrids (Jing724 × Mo17, Mo17 × Jing724, B73 × Qi319, and Qi319 × B73) were low (0.0158, 0.0122, 0.0124, and 0.0285 mg/kg, respectively) (**Figure 3c**). Moreover, the Jing724 × Mo17 hybrid was backcrossed with Mo17 (high-Cd line) to generate a BC₁ segregating population. The frequency distribution of the grain Cd content in the 218 BC₁ plants revealed two apparent peaks, with a Cd content of 0.1 mg/kg set as the threshold (**Figure 4d**). Of the 218 BC₁ plants, 102 had high grain Cd contents (0.1015–0.7678 mg/kg), whereas 116 had low grain Cd contents (0.0178–0.0938 mg/kg), which was consistent with the expected 1:1 ratio ($\chi^2_{1:3} = 0.899$, $P = 0.3430$) (**Figure 4d**). These results implied that the high maize grain Cd content was controlled by a single recessive major-effect gene.

Mapping of the genes regulating Cd accumulation via BSR-seq

To produce extremely high and low Cd accumulation sample pools, an F₂ segregating population was obtained by crossing Jing724 (low-Cd line) with Mo17 (high-Cd line). Total RNA samples extracted from the leaves of the high- and low-Cd plants in the F₂ population were divided into two separate pools for an RNA sequencing (RNA-seq) analysis using the HiSeq X instrument (Illumina, USA). A total of 158 million reads were produced, with an average length of 2 × 150 bp/read. The raw reads were trimmed and then aligned to the B73 RefGen_v3 reference genome using the Genomic Short-read Nucleotide Alignment Program (GSNAP) (Wu et al., 2016). Reads with a single unique alignment were used for SNP calling.

To map the gene responsible for the Cd accumulation in grains, the high-Cd grains were used as the recessive pool, whereas the low-Cd grains were used as the dominant pool. A total of 110,085 SNPs were identified. The empirical Bayesian approach was employed to estimate the probability of a linkage between a high confidence SNP and the causal gene (Liu et al., 2012). The BSR-seq-based mapping revealed 946 associated SNPs clustered on chromosome

2 (**Figure 4a; Table S2**). To delimit the interval, a sliding window scan was performed for the associated SNPs (window, 20 SNPs; step, 5 SNPs), after which the significant SNPs were localized to the 142–161 Mb region on chromosome 2 (B73 RefGen_v3) (**Figure 4b**). The *qCd1* major QTL was located within this interval, suggesting it may be related to the grain Cd content variations between Jing724 and Mo17.

Fine mapping of *qCd1* for maize grain Cd accumulation

To finely map *qCd1*, 11 SNP-based KASP (Kompetitive Allele-Specific PCR) markers (**Table S3**) were developed in the 142–161 Mb region on chromosome 2 for an analysis of 149 BC₁ [(Jing724 × Mo17) × Mo17] individuals with high grain Cd contents (> 0.2 mg/kg). We mapped *qCd1* to a 423-kb interval between markers KASP-7 and KASP-8 (**Figure 4c**). Using another five SNP markers and an additional 1,117 BC₁ individuals with high grain Cd contents (> 0.2 mg/kg), we further anchored *qCd1* to a 232-kb interval between markers KASP-12 and KASP-15 (**Figure 4d**). Six and seven genes were identified in this region in B73 and Mo17, respectively, based on the maize B73 RefGen_v3 reference genome (**Figure 4e, f**). Gene functional annotations indicated *GRMZM2G175576* (*ZmHMA3*) and *GRMZM2G455491* (*ZmHMA4*) encode P-type HMA transporters. Both genes were considered to be candidate genes in *qCd1*.

***GRMZM2G175576* was identified as the candidate gene for maize grain Cd accumulation**

To further characterize the candidate genes, the *GRMZM2G175576* and *GRMZM2G455491* genomic sequences and their expression were analyzed. The two gene sequences were identical in B73 and Jing724. Thus, B73 was used for further study because of the availability of the B73 reference genome. A comparison of the genomic sequences in Mo17 (high-Cd line) and B73 (low-Cd line) revealed 19 SNPs in the *GRMZM2G455491* coding sequence (CDS) that did not lead to translational termination. Regarding the comparison of the *GRMZM2G175576* sequences in Mo17 and B73, a 7,520-bp deletion and a 603-bp insertion were detected 300 bp upstream of the translation start codon as well as a 7,191-bp transposon insertion in intron 1 and a 628-bp insertion in intron 3 in Mo17 (**Figure 5a**). An analysis using the CENSOR software indicated the 603-bp insertion upstream of *GRMZM2G175576* was a gypsy-type LTR retrotransposon fragment, whereas the 7,191-bp insertion in exon 1

1 was a copia-type LTR retrotransposon and the 628-bp insertion in intron 3 was a LINE-1
2 retrotransposon.

3 The data in the Maize eFP browser (http://bar.utoronto.ca/efp_maize/cgi-bin/efpWeb.cgi)
4 suggested that *GRMZM2G175576* was highly expressed in the roots, leaves, internodes, and
5 whole seeds, whereas *GRMZM2G455491* was highly expressed exclusively in the anthers
6 (**Figure S3**). Reverse transcription polymerase chain reaction (RT-PCR) and quantitative
7 real-time polymerase chain reaction (qRT-PCR) experiments revealed that
8 *GRMZM2G175576* was expressed in maize seedling root and leaf tissues, in contrast to the
9 lack of *GRMZM2G455491* expression in Jing724, B73 and Mo17 (**Figure S4**). These results
10 implied *GRMZM2G175576* was the more likely target gene.

11 The GWAS analysis indicated *GRMZM2G175576* harbors 12 significant SNPs, with the
12 lead SNP (chr2.S_158408660, $P = 4.76e-20$) explaining 20.03% of the phenotypic variation
13 (**Figure 2c**). The phenotypic differences associated with the two alleles of chr2.S_158408660
14 were significant (**Figure 2d**). The 12 significant SNPs in *GRMZM2G175576* were clustered
15 ($D' > 0.94$), reflecting the strong LD of these SNPs (**Figure 2e**). This result further confirmed
16 that *GRMZM2G175576* was the candidate gene. Hereafter, we refer to *GRMZM2G175576* as
17 *ZmCd1*.

***ZmCd1* transcriptional variation between Mo17 and B73**

18 A qRT-PCR analysis proved that *ZmCd1* was expressed in the B73 and Mo17 maize seedling
19 leaf and root tissues. The *ZmCd1* expression level was significantly higher in Mo17 than in
20 Jing724 (**Figure 5d**), suggesting that the 7,520-bp deletion and transposon insertion in the
21 *ZmCd1* promoter resulted in increased expression in Mo17. To map the *ZmCd1* transcription
22 start site (TSS) in B73 and Mo17, a 5'-RACE was conducted using gene-specific primers
23 targeting exon 1 (**Table S4**). The *ZmCd1*^{B73} TSS was 191 or 89 bp upstream of the start codon,
24 whereas the *ZmCd1*^{Mo17} TSS was 274 bp upstream of the start codon. Thus, differences in the
25 *ZmCd1* promoter did not influence translation initiation in Mo17.
26

27 Using 7F/7R primers to amplify *ZmCd1* from exon 1 to exon 5, we determined that the
28 amplified Mo17 *ZmCd1* sequences were nonspecific and diffuse, in contrast to the distinct
29 2,737-bp amplified B73 *ZmCd1* sequence (**Figure 5b**). The amplified *ZmCd1*^{Mo17} fragments
30 were cloned and sequenced, resulting in the detection of 19 different transcripts in 109 clones,

of which 72 clones had transcripts shorter than 2,737 bp and the other 37 clones had transcripts longer than 2,737 bp (**Figure 5c; Figure S5**). The transcript analysis revealed that the retrotransposon inserted in intron 1 resulted in transcript and amino acid coding abnormalities in *ZmCd1*^{Mo17} (Dataset S1). Therefore, we proposed that the retrotransposon in intron 1 abolished the *ZmCd1* function in Mo17, leading to increased grain Cd accumulation.

Functional verification of *ZmCd1* by an allelism test of the mutants

Two B73 EMS mutants, *zmcd1-1* (Mut_Sample: EMS4-038f45) and *zmcd1-2* (Mut_Sample: EMS4-038f39), were obtained from the maize EMS mutant library (<http://www.elabcaas.cn/memd/>) for an allelism test. Both *zmcd1-1* and *zmcd1-2* have a termination mutation in the *ZmCd1* coding region (**Figure 6a, b**). The Cd contents in the *zmcd1-1* and *zmcd1-2* maize grains harvested from Cd-contaminated farmland were as high as 0.3507 and 0.3093 mg/kg, respectively (**Figure 6c**). Both maize mutants were crossed with the high-Cd maize line Mo17. The Cd contents of the grains from the self-pollinated hybrid maize plants grown on Cd-contaminated farmland were determined. Grains from Mo17 × *zmcd1-1* and *zmcd1-2* × Mo17 self-pollinated lines had high Cd levels of 0.4008 and 0.3421 mg/kg, respectively (**Figure 6c**). The allelism test results verified that *ZmCd1* was associated with the high maize grain Cd content.

Evolutionary analysis and subcellular localization of *ZmCd1*

A phylogenetic analysis divided 43 HMA-encoding genes from Arabidopsis, rice, sorghum, maize, *Brassica rapa*, barley, and durum wheat into four clades (**Figure 7a; Table S5**). Specifically, *ZmCd1/ZmHMA3* in the Zn/Co/Cd/Pb clade formed a subclade with *OsHMA3*, *SbHMA3*, *TdHMA3-B1a*, and *HvHMA3-B* (**Figure 7a**), indicating that *ZmCd1* is probably functionally similar to *OsHMA3* and *AtHMA3*. A protein sequence alignment revealed a 66.5% sequence match between *ZmCd1* and *OsHMA3* and a 51.9% sequence match between *ZmCd1* and *AtHMA3* (**Figure S6**).

Because *ZmCd1* was predicted to encode six putative transmembrane domains (**Figure 7b**), we performed a maize protoplast transient expression assay using the *ZmCd1*-eGFP fusion protein and the protoplast-localized fusion protein comprising Arabidopsis TIP2;3 and mCherry. The *ZmCd1*-eGFP fusion protein was co-localized with AtTIP2;3-mCherry in the maize protoplast vacuolar membrane (**Figure 7c**). We speculated that *ZmCd1* helps sequester

1 Cd in vacuoles, similar to AtHMA3 and OsHMA3.

2 **Natural variations in *ZmCd1* in diverse maize lines**

3 Using the structural variations of the *ZmCd1* gene in 38 maize inbred lines with sequenced
4 genomes, six simple natural *ZmCd1* variation types (haplotypes) were identified based on the
5 insertions/deletions (InDels) in the *ZmCd1* promoter, intron 1, intron 4, and exon 5 (compared
6 with the B73 sequence) (**Table S6; Figure 5a**). The first variation type (i.e., InDels in the
7 promoter) was detected only in the European EP1 and DK105 maize lines. The five other
8 types of *ZmCd1* variations were detected in the maize lines examined in this study, and the
9 representative maize lines were B73, Mo17, Ky21, D9H, and CML333, respectively (**Figure**
10 **5a; Figure 8**). Notably, D9H was used to represent SK, CML228, and F7 because the InDel
11 variation in the *ZmCd1*^{D9H} promoter was similar to that in the Mo17 *ZmCd1* promoter.

12 Compared with the corresponding sequence in B73 (low-Cd line, 0.0248 mg/kg), *ZmCd1*
13 in Ky21 (high-Cd line, 0.25 mg/kg) had a 7,520-bp deletion and a 603-bp insertion 300 bp
14 upstream of the translation start codon as well as a 12,968-bp gypsy-type LTR
15 retrotransposon insertion in exon 4 and a 1,002-bp MuDR DNA transposon insertion in exon
16 5 (**Figure 5a**). The full-length *ZmCd1* cDNA sequence was amplified by PCR using the
17 7F/7R primers, but no amplified fragment was detected for Ky21. However, a 1,355-bp
18 fragment was detected when the 8F/8R primers were used to amplify the sequence between
19 *ZmCd1* exons 1 and 4 (**Figure 5b**). These results indicated that the insertion mutation in
20 intron 4 resulted in a premature transcriptional termination in Ky21. The 3' ends of *ZmCd1*^{Ky21}
21 transcripts were obtained by 3'-RACE using gene-specific primers targeting exon 4. A
22 sequence analysis revealed two transcripts that were shorter in Ky21 than in B73, with no
23 exon 5. One transcript retained the 1–139 bp sequence of intron 4, whereas the other
24 transcript retained the 9,601–9,816 bp sequence of intron 4 (**Figure 5c**).

25 Compared with the B73 *ZmCd1* sequence, the *ZmCd1* gene in maize inbred line D9H
26 (high-Cd line, 0.13 mg/kg) had a 7,520-bp deletion and a 603-bp insertion 300 bp upstream of
27 the translation initiation codon as well as a 66-bp deletion in exon 5. In maize inbred line
28 CML333 (low-Cd line, 0.06 mg/kg), the *ZmCd1* sequence only had a 66-bp deletion in exon 5
29 (**Figure 5a**).

30

1 ***ZmCd1* molecular marker development**

2 Four PCR-based molecular markers were developed based on the *ZmCd1* differences in the
3 upstream (M1) sequence and in intron 1 (M2), intron 4 (M3), and exon 5 (M4) among B73,
4 CML333, Mo17, Ky21, and D9H (**Figure 5a; Figure 8a**). Using these markers, five
5 haplotypes were detected in the maize lines included in this study. In Hap_1, the PCR
6 amplified *ZmCd1* products were 776, 305, 529, and 188 bp in B73 (low-Cd line). In Hap_2,
7 the PCR amplified products were 776, 305, 529, and 122 bp in CML333 (low-Cd line). In
8 Hap_3, the PCR amplified products were 574, 630, 529, and 188 bp in Mo17 (high-Cd line).
9 In Hap_4, the PCR amplified products were 574, 305, 965, and 188 bp in Ky21 (high-Cd line).
10 In Hap_5, the PCR amplified products were 574, 305, 529, and 122 bp in D9H (high-Cd line)
11 (**Figure 8a, b**).

12 The genotyping of 454 maize lines from the GWAS population using the four molecular
13 markers developed for *ZmCd1* indicated that 293 lines carried Hap_1, 51 lines carried Hap_2,
14 82 lines carried Hap_3, 20 lines carried Hap_4, and 8 lines carried Hap_5. An analysis of the
15 BLUP data for the maize lines with haplotypes 1–5 determined that the average Cd contents
16 were 0.0993, 0.0939, 0.3163, 0.1214 and 0.1199 mg/kg, respectively. Moreover, there was a
17 significant difference in the Cd content between Hap_1 and Hap_3 ($P < 0.0001$), Hap_1 and
18 Hap_4 ($P < 0.05$), and Hap_1 and Hap_5 ($P < 0.05$) (**Figure 8c**).

19 **Prediction of maize grain Cd accumulation using the molecular markers developed in** 20 **this study**

21 The genotypes of 36 elite inbred lines and 13 elite hybrids widely used for maize production
22 were analyzed using the four developed markers. The 13 hybrids included the top three
23 varieties in China (Xianyu335, Zhengdan958, and Jingke968). Four *ZmCd1* haplotypes were
24 detected in the 36 inbred lines, including 20 lines carrying Hap_1, 7 lines carrying Hap_3, 3
25 lines carrying Hap_4, and 6 lines carrying Hap_5 (**Figure 9a**). We detected three haplotypes
26 in 13 hybrids, with Zhengdan958 carrying Hap_1/Hap_3, Xianyu335, MC812, and
27 Jingnongke736 carrying Hap_1/Hap_5, and Yuqingzhu23 and Kangnong999 carrying Hap_3
28 (**Figure 9b**). These maize inbred lines and hybrids were grown in Cd-contaminated soil, after
29 which their grain Cd contents were measured. The maize grain Cd contents of these materials
30 were consistent with the results predicted using the molecular markers (**Figure 9**). The grain

Cd contents of 20 inbred lines (e.g., Jing2416, Jing724, MC01, and Jing92) and 7 hybrids (e.g., Jingnongke728 and Jingke968) carrying Hap_1 were low (0.0096–0.0577 mg/kg), in contrast to the high grain Cd contents (0.1094–1.6760 mg/kg) of 16 inbred lines carrying Hap_3, Hap_4, or Hap_5 and two hybrids carrying Hap_3. Hybrids Zhengdan958, MC812, Xianyu335, and Jingnongke736 carried Hap_1/Hap_3 or Hap_1/Hap_5. Because Hap_1 is dominant over Hap_3 and Hap_5, these four hybrids had low grain Cd contents (**Table S7**). These results proved that the four molecular markers can be used to predict maize grain Cd accumulation characteristics, making them useful for screening low-Cd maize germplasms and for improving high-Cd maize lines by molecular marker-assisted selection.

Discussion

Minimizing maize grain Cd accumulation is critical for decreasing the risks of Cd toxicity in humans. Dissecting the genetic components underlying Cd accumulation may enable researchers and breeders to decrease maize grain Cd contents (Jha and Bohra, 2016; Liu et al., 2019). In this study, combining multiple genetic mapping techniques led to the identification of a candidate gene for maize grain Cd accumulation on chromosome 2. The allelism test of the mutants confirmed *ZmCd1* contributes to grain Cd accumulation. A DNA sequence analysis detected natural variations in the *ZmCd1* promoter, intron 1, intron 4, and exon 5. A transcript analysis revealed mutations in transcripts and the amino acid-encoding sequence of *ZmCd1* resulting from a retrotransposon insertion. Subsequent phylogenetic and subcellular localization analyses indicated that *ZmCd1* encodes a vacuolar membrane-localized P_{1B}-type HMA transporter. Moreover, four molecular markers for *ZmCd1* were developed and five *ZmCd1* haplotypes were detected. The utility of the molecular markers for predicting the maize grain Cd accumulation phenotype confirmed that these molecular markers are useful for maize breeding.

Cadmium accumulation is related to environmental factors (Gill and Tuteja, 2011; Zhang et al., 2018a). Several studies have demonstrated that plant Cd contents are influenced by soil Cd concentrations, pH, and other factors (Ali et al., 2019; Florijn and Van Beusichem; Gill and Tuteja, 2011). In rice, Cd accumulation increases as the pH of the culture medium increases from 3.5 to 6.0, but it decreases as the pH increases to 7.0 and 8.0 (Ali et al., 2019).

1 Similar to these findings, we determined that the average maize grain Cd content in two
2 seasons in ZZ was higher than that in NX, which is consistent with the observation that the
3 soil Cd concentration in ZZ was higher than that in NX. Additionally, the soil pH in ZZ was
4 close to 6.0, whereas it was higher than 7.5 in NX (**Figure 1b**).

5 In this study, a large genotypic variation in the maize grain Cd content was detected for
6 the GWAS population regardless of the season and location, suggesting that maize grain Cd
7 accumulation is genetically controlled. Previous studies identified several factors regulating
8 Cd tolerance, among which ATP-binding cassettes (Patrizia et al., 2015; Zhang et al., 2018b)
9 and HMAs (Jha and Bohra, 2016; Maccaferri et al., 2019; Sasaki et al., 2014; Yan et al., 2016)
10 have been widely investigated in Arabidopsis, rice, and wheat. However, the key gene
11 regulating Cd accumulation in maize remained uncharacterized. In a recent study, a major
12 QTL (153.75–167.58 Mb on chromosome 2) for maize leaf Cd accumulation, *qLCd2*, was
13 identified based on a GWAS and a DH population (Zhao et al., 2018a). The genomic position
14 of *qLCd2* was comparable to that of the major QTL identified by BSR-seq (142–161 Mb) and
15 GWAS mapping in this study, suggesting that this QTL controls Cd accumulation in maize
16 leaves and grains. The fine mapping results anchored this major QTL to the
17 159,015,366–159,246,974 bp interval on chromosome 2. Furthermore, gene functional
18 annotations and sequence analyses resulted in the identification of *GRMZM2G175576*
19 (*ZmCd1*) as a candidate gene. The *ZmCd1* function was further verified by an allelism test of
20 the maize *zmcd1* mutants.

21 The *ZmCd1* sequence of the high-Cd maize line Mo17 included long DNA deletion and
22 retrotransposon insertion variations in the promoter and introns. A previous investigation
23 proved that the differences in *OsHMA3* expression and Cd accumulation between *japonica*
24 and *indica* rice were due to 11 nucleotide differences in the promoter (Liu et al., 2020). A
25 Sukkula-like transposable element insertion upstream of *HvHMA3* reportedly promotes gene
26 expression (Lei et al., 2020). Similarly, we observed that *ZmCd1* expression was significantly
27 higher in Mo17 than in B73 (**Figure 5d**), possibly because of the insertion of a 603-bp
28 retrotransposon in the promoter. Although the *ZmCd1* expression level was higher in Mo17, a
29 7,191-bp retrotransposon insertion in intron 1 resulted in mutated transcripts (**Figure 5a–c**)
30 with abnormal amino acid-encoding sequences. The resulting loss of functional *ZmCd1*

1 increased the grain Cd content.

2 A phylogenetic analysis of maize, barley, durum wheat, sorghum, *B. rapa*, Arabidopsis,
3 and rice revealed that *ZmCd1* belongs to the Zn/Co/CD/Pb group and forms a subclade with
4 *OsHMA3*, *SbHMA3*, *TdHMA3-B1a*, and *HvHMA3-B*. The *OsHMA3*, *HvHMA3-B*, and
5 *TdHMA3-B1a* proteins are localized to the vacuolar membrane (Lei et al., 2020; Maccaferri
6 et al., 2019; Miyadate et al., 2011), where they help sequester Cd in vacuoles and influence
7 root-to-shoot Cd translocation (Lei et al., 2020; Maccaferri et al., 2019; Miyadate et al., 2011;
8 Ueno et al., 2010). These findings suggest the HMA3 transporters mediate a conserved Cd
9 tolerance mechanism in diverse species. In this study, a subcellular localization analysis
10 confirmed that *ZmCd1* is a vacuolar membrane protein (**Figure 7c**). The hybrid grain Cd
11 contents were correlated with the corresponding Cd content of the female parent (**Figure 3b**),
12 indicative of a lack of xenia effect on the maize grain Cd content. Consistent with the
13 distribution of Cd in various rice tissues (Rizwan et al., 2016), the rank order of the Cd
14 contents in maize tissues was as follows: leaf > stem > grain (**Figure S7**). These results
15 indicate that *ZmCd1* mediates the sequestration of Cd in vacuoles and restricts the Cd
16 accumulation in maize grains by modulating Cd root-to-shoot translocation.

17 We developed four molecular markers based on the natural variations in *ZmCd1* among
18 38 maize inbred lines with sequenced genomes. Five variation types (Hap_1–5) were
19 identified in the study materials, and three (Hap_3–5) were associated with high maize grain
20 Cd contents. The *ZmCd1* genes of the Hap_3 and Hap_4 lines had a transposon inserted in
21 introns, which adversely affected transcription. The Hap_5 lines had a 66-bp deletion in
22 *ZmCd1* exon 5. These three haplotypes had abnormal amino acid-encoding *ZmCd1* sequences,
23 resulting in nonfunctional *ZmCd1* (**Dataset S1**). However, the Hap_2 and Hap_5 both having
24 the 66-bp deletion in exon 5 showed different grain Cd contents probably due to the presence
25 of other loci in the GWAS panel controlling Cd accumulation in maize grains, which would
26 be figured out in the further study. The analysis of the GWAS population indicated most of
27 the maize inbred lines had low Cd levels. Additionally, Hap_3 was the main haplotype among
28 the high-Cd maize lines. The molecular markers were used to investigate the commonly
29 cultivated maize inbred lines and hybrids in China. Hybrids Jingke968, Zhengdan958, and
30 Xianyu335 are the predominant maize varieties grown in China. Jingnongke728 (MC01 ×

Jing2416) is suitable for mechanical harvesting. The Cd accumulation of 13 hybrids (including the above-mentioned hybrids) and 36 inbred lines was consistent with that predicted by the molecular markers (**Figure 9; Table S7**), implying the molecular markers are applicable for the early screening of low-Cd maize lines and for genetically improving high-Cd maize lines. Of the 13 hybrids, Yuqingzhu23 and Kangnong999 carried Hap_3 and had high grain Cd contents, indicating they may be inappropriate for the production of maize grains intended for human consumption. Molecular marker-assisted breeding may be useful for improving their Cd accumulation characteristics.

Experimental procedures

Plant materials and treatments

The association panel comprised 513 maize inbred lines, including 228 lines originating from temperate regions, 258 lines from tropical–subtropical regions, and 27 lines with an admixed origin (Yang et al., 2014). For Cd treatments, maize materials were planted in ZZ and NX. The soil at the study sites was contaminated with Cd because of Cd-rich surface water irrigation. Field trials for the association population were conducted in ZZ (N27°49'50.88", E113°07'41.23") and NX (N28°16'51.16", E112°32'47.13") in the spring and autumn of 2018. Plants were grown in a single-row plot with two replicates using a randomized complete block design. The F₂ population, a BC₁ segregating population and the parents, *zmcd1-1* and *zmcd1-2* (EMS mutants), as well as the F₁ hybrids (obtained by crossing the EMS mutants with test inbred lines) were planted in ZZ in the autumn of 2018. Each row plot had 10 plants in 3 m rows that were separated by 60 cm.

Soil sampling and composition analysis

Soil samples (0–20 cm depth) were collected from the ZZ and NX study sites using the five-point sampling method (Luo et al., 2017). The soil Cd content was measured by graphite furnace atomic absorption spectrophotometry according to the National Standards of the People's Republic of China (GB/T 17141-1997). Soil pH was measured using a pH meter.

Measurement and analysis of maize grain Cd contents

For each maize line replicate, grains were harvested and dried, after which they were ground to a powder. The ground material was passed through a 100-mesh sieve. The collected maize

1 grain powder was weighed and then 0.1–0.5 g was digested with HNO₃/HClO₄ (9:1, v/v).
 2 Finally, the Cd concentration of the digestion solution was determined using the ZEE nit700P
 3 atomic absorption spectrometer (Analytikjena, Germany) (Zhao et al., 2018a).

4 The BLUP data for the GWAS population across environments were estimated via
 5 mixed models using the following formula: $Y = (1|LINE) + (1|ENV) +$
 6 $(1|REP\%in\%LINE:ENV) + (1|LINE:ENV)$ (Y, trait data; “|”, groups; “:”, interactions; LINE,
 7 all testcrosses used; ENV, environments; and REP, replications in one ENV) (Zhao et al.,
 8 2018b). The combined ANOVA of four environments was completed and the heritability of
 9 the grain Cd content trait was calculated using the ANOVA tool of the IciMapping software
 10 (version 4.1) (Luo et al., 2017). The BLUP analysis was performed using the lme4 package of
 11 the R software (Zhao et al., 2018b), whereas the correlation analysis was conducted using the
 12 Corrplot package of the R software (Luo et al., 2019).

13 **GWAS analysis and candidate gene prediction**

14 The association panel was genotyped by combining RNA-seq and Illumina MaizeSNP50
 15 BeadChip data (Yang et al., 2014). A total of 1.25 million high-quality SNPs with a minor
 16 allele frequency exceeding 5% were used for the GWAS analysis (Yang et al., 2014). The
 17 GWAS was completed using TASSEL (version 5) (Bradbury et al., 2007). More specifically,
 18 we used a compressed mixed linear (cMLM) model that included the population structure (Q)
 19 and the relative kinship (K). A significance threshold of 7.98e-7 (1/independent marker
 20 number) was set to control the genome-wide type 1 error rate. The genes within 200-kb
 21 flanking regions (± 100 kb of every significant SNP) (Liu et al., 2017) in the B73 RefGen_v2
 22 reference genome in the MaizeGDB database (<https://www.maizegdb.org/>) were used for
 23 predicting candidate genes based on functional annotations.

24 **BSR-seq and fine mapping**

25 The Jing724 $\times$ Mo17 F₂ population was used for a BSR-seq analysis. Using the TRIzol
 26 reagent method (Luo et al., 2017), total RNA was extracted from the fresh leaves of each F₂
 27 seedling grown in a Cd-contaminated field in ZZ. After determining the grain Cd content of
 28 the F₂ population, the RNA samples of 30 plants with an extremely high grain Cd content and
 29 of 30 plants with an extremely low grain Cd content were divided into two equal pools (i.e.,
 30 high-Cd pool and low-Cd pool). The two RNA pools were sequenced using a HiSeq X

1 instrument (Illumina, USA). The RNA-seq data were deposited in the NCBI Sequence Read
2 Archive database (accession number PRJNA656212)
3 (<https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA656212>). The RNA-seq reads were
4 trimmed to remove low confidence nucleotides with quality scores < 15 out of 40. The
5 trimmed reads were aligned to the B73 RefGen_v3 reference genome using GSNAP (Wu et
6 al., 2016). The reads uniquely mapped to a single location in the reference genome were used
7 for SNP calling. A Bayesian approach was applied to evaluate the probability of a linkage
8 between each SNP and the causal gene. Associated SNPs were identified based on a
9 probability cutoff of 0.05 (Liu et al., 2012).

10 For the fine mapping of the major causal locus (*ZmCdl*) identified by BSR-seq, a
11 5,000-plant BC₁ population was generated by crossing the Jing724 × Mo17 F₁ plants with the
12 Mo17 plants. The causal locus genotype of the high-Cd plants should theoretically be
13 identical to that of Mo17. Thus, we developed new SNP-based KASP markers for the target
14 region with the SNPs identified by BSR-seq. These markers were used to delimit the interval
15 containing *ZmCdl* based on the recombination events in the extremely high-Cd BC₁ plants
16 (grain Cd content > 0.2 mg/kg). The KASP marker primers were designed and synthesized by
17 LGC Biosearch Technologies (UK) (**Table S3**). The markers were used to determine the
18 genotypes of the BC₁ plants with the LGC SNPlane PCR Genotyping System (UK). Candidate
19 genes were predicted based on functional annotations, gene expression, and target region
20 sequence variations between B73 and Mo17.

21 ***ZmCdl* sequence analysis**

22 Total DNA was extracted from the leaves of maize inbred line seedlings (e.g., B73, Jing724,
23 D9H, Ky21, CML333, and Mo17) using a published CTAB method (Luo et al., 2019).
24 Primers for amplifying *ZmCdl* (**Table S4**) were designed based on the B73 genome sequence
25 in the MaizeGDB database (<https://www.maizegdb.org/>). Phanta Max Super-Fidelity DNA
26 Polymerase (Vazyme Biotech Co., Ltd., Nanjing, China) was used for the PCR. The amplified
27 products were sequenced on the ABI3730 sequencer by TsingKe Biological Technology Co.
28 Ltd. (Beijing, China). Sequences were aligned and assembled using the MegAlign software
29 (<http://www.freedownload64.com/s/megalign>). Transposable elements were predicted for the
30 queried sequences using the CENSOR software (<https://www.girinst.org/censor/index.php>).

EMS mutant analysis and allelism test

By screening the Maize EMS induced Mutant Database (<http://www.elabcaas.cn/memd/>) for the B73 inbred line, we identified *zmcd1-1* (EMS4-038f45) and *zmcd1-2* (EMS4-038f39) as two *ZmCd1* EMS mutants harboring nonsense mutation sites. The mutation sites were validated by sequencing PCR products with specific primers (**Table S4**). The effects of the *zmcd1-1* and *zmcd1-2* mutations on grain Cd accumulation were evaluated by comparing the grain Cd contents of B73 (wild-type), *zmcd1-1*, and *zmcd1-2* plants grown in a Cd-contaminated field in ZZ. The F₁ hybrids generated by crossing the high-Cd *zmcd1-1* and *zmcd1-2* mutants with the high-Cd inbred line Mo17 were grown in a Cd-contaminated field in the autumn of 2018. The Cd contents of the kernels produced by the F₁ hybrids, Mo17 × *zmcd1-1* and *zmcd1-2* × Mo17, were measured to confirm whether the mutated gene in the *zmcd1-1* and *zmcd1-2* mutants was allelic to the candidate gene for high grain Cd contents in Mo17.

RT-PCR and qRT-PCR analyses of *ZmCd1*

Total RNA was extracted from whole maize seedlings using the TRIzol reagent (Thermo Fisher Scientific, USA) according to the manufacturer instructions. The extracted RNA was purified with RNase-free DNase 1 (Invitrogen, USA), after which it was used as the template for synthesizing cDNA using the PrimeScript™ II 1st strand cDNA Synthesis Kit (Takara Bio, Beijing, China). The full-length *ZmCd1* cDNA was amplified by PCR using Phanta Max Super-Fidelity DNA Polymerase (Vazyme Biotech). The PCR products were analyzed by 2% agarose gel electrophoresis. The RT-PCR primers are listed in **Table S4**. The PCR products were purified and then sequenced by TsingKe BioTech (Beijing, China). Regarding *ZmCd1*^{Mo17}, the RT-PCR products were inserted into the pEASY Blunt Simple vector (TransGen Biotech, Beijing, China) for sequencing.

To analyze *ZmCd1* expression in Jing724 and Mo17, total RNA extracted from seedling roots and leaves at the V3 stage was used to synthesize cDNA as described above. A qRT-PCR assay was performed using the QuantStudio™ 6 Flex system (Thermo Fisher Scientific) and TB Green® Premix Ex Taq™ II (Tli RNaseH Plus) (Takara Bio, Beijing, China) with gene-specific primers (**Table S4**). Three biological replicates were used to calculate relative expression levels according to the 2^{-ΔΔCt} method (Luo et al., 2017). The

1 *Actin1* gene (*GRMZM2G126010*) was used as the internal control.

2 **5'- and 3'-RACE**

3 The *ZmCd1*^{Jing724} and *ZmCd1*^{Mo17} transcription start sites were mapped using the
4 FirstChoice™ RLM-RACE Kit (Thermo Fisher Scientific) according to the manufacturer
5 instructions. Briefly, total RNA extracted from Jing724 and Mo17 seedlings was digested
6 with Calf Intestine Alkaline Phosphatase (CIP) to remove uncapped RNA sequences and then
7 treated with Tobacco Acid Pyrophosphatase (TAP) to remove the 5'-cap structure. An RNA
8 adapter was ligated to the CIP/TAP-treated RNA sequences using T4 RNA ligase, after which
9 cDNA was synthesized using M-MLV Reverse Transcriptase. A nested PCR was performed
10 using 5'-RNA adapter primers and gene-specific primers (**Table S4**). A 3'-RACE was
11 performed using total RNA from Ky21 seedlings to map the polyadenylation sites of
12 *ZmCd1*^{Ky21} mRNA. First-strand cDNA was synthesized using the 3'-RACE adapter
13 (5'-GCGAGCACAGAATTAATACGACTCACTATAGGT₁₂VN-3') as a reverse transcription
14 primer. A nested PCR was performed using the 3'-RNA adapter primers and gene-specific
15 primers (**Table S4**). The 5'- and 3'-RACE PCR products were cloned into the pEASY Blunt
16 Simple vector (TransGen Biotech) and the clones were sequenced by TsingKe Biological
17 Technology Co. Ltd. (Beijing, China).

18 **Phylogenetic analysis of plant HMAs**

19 The amino acid sequences of 43 previously reported plant HMA proteins were retrieved from
20 the Phytozome database (version 12) (<https://phytozome.jgi.doe.gov/pz/portal.html>) and the
21 NCBI GenBank database (<https://www.ncbi.nlm.nih.gov/genbank/>). The sequences included 8
22 AtHMAs in Arabidopsis (Baxter et al., 2003), 9 OsHMAs in rice (Baxter et al., 2003; Zhiguo
23 et al., 2018), 9 SbHMAs in sorghum (Zhiguo et al., 2018), 12 ZmHMAs in maize (Cao et al.,
24 2019; Zhiguo et al., 2018), BrHMA3 in *B. rapa* (Zhang et al., 2019), HvHMA3-B in barley
25 (Lei et al., 2020), and TdHMA3-B1a in durum wheat (Maccaferri et al., 2019) (**Table S5**).
26 The amino acid sequences of the 43 HMA proteins were aligned using MUSCLE
27 (<https://www.ebi.ac.uk/Tools/msa/muscle/>). A neighbor-joining phylogenetic tree was
28 constructed using MEGA 7 (Sudhir et al., 2016). Specifically, the Poisson amino acid
29 substitution model and 1,000 bootstrap replicates were used.

30 **Subcellular localization of ZmCd1**

The full-length *ZmCdl*^{B73} CDS was fused in frame with the enhanced green fluorescent protein (eGFP)-encoding CDS in the pM999-eGFP vector. The pM999-eGFP empty vector with a 35S-eGFP cassette served as the blank control. A vector containing a sequence encoding the AtTIP2;3-mCherry fusion protein (tonoplast-localized marker) was included in co-transformations. Maize protoplasts were isolated from leaf tissues and co-transformed with the vectors using the plant protoplast preparation and transformation kit (ZhongkeRuitai Biotechnology Co., Ltd., Beijing, China). Fluorescence was detected using the SP8 confocal laser scanning microscope (Leica, USA) with excitation/emission wavelengths of 488 nm/505 nm and 543 nm/610 nm for GFP and mCherry, respectively.

Development and application of PCR-based functional molecular markers for maize grain Cd accumulation

The genome sequences of 38 maize inbred lines were downloaded from the MaizeGDB database (<https://download.maizegdb.org/>) to obtain the *ZmCdl* genomic sequences. On the basis of the *ZmCdl* sequence differences among the maize inbred lines, four PCR-based molecular markers (M1–M4) were developed for four regions with varying gene structures (**Figure 5a; Table S6**). Three primers were used in one PCR amplification for the M1–M3 markers to distinguish the InDel variations (**Figure 5a**). For markers M2 and M3, the PCR extension step was less than 1 min to avoid amplifying long transposable elements. Details regarding the primers for these markers are listed in **Table S4**. The PCR products were analyzed by 2% agarose gel electrophoresis. Six haplotypes were detected in the genome-sequenced maize inbred lines based on different M1–M4 marker combinations (**Table S6**). Additionally, 36 elite inbred lines and 13 hybrid varieties (**Table S7**) commonly cultivated in China were grown in a Cd-contaminated field in ZZ in the autumn of 2018. The CTAB method (Luo et al., 2019) was used to extract the DNA from these materials. The M1–M4 markers were used for genotyping the inbred lines and hybrid varieties to predict the grain Cd contents.

1 Acknowledgments

2 We thank Prof. Jianbing Yan for kindly providing the GWAS panel. This research was financially
3 supported by the Beijing Municipal Natural Science Foundation (6204041), the Youth
4 Research Fund of Beijing Academy of Agriculture and Forestry Sciences (QNJJ202028), the
5 National Key Research and Development Program of China (2016YFD0300109), the
6 XiangCaiNongZhi (2019-33), the Construction of Collaborative Innovation Center of Beijing
7 Academy of Agricultural and Forestry Sciences (Collaborative Innovation Center of Crop
8 Phenomics, KJCX201917), and the Beijing Scholars Program (BSP041).

9

10 Conflicts of interest

11 The authors declare no competing interests.

12

13 Authors' contributions

14 Z.Y. (Yanxin Zhao), T.B., C.Z. (Zhihui Chen), and Z.J. (Jiuran Zhao) conceived the study. Z.Y.
15 (Yanxin Zhao), T.B., L.M., and S.W. designed the experiments. T.B., L.M., Z.Y. (Yunxia
16 Zhang), G.H., L.J., Z.R., F.Z., K.M., and L.H. performed the experiments. Z.Y. (Yanxin Zhao),
17 L.D., and Z.J. (Jianhua Zhang) analyzed the BSR-seq data. L.M., Z.Y. (Yanxin Zhao), C.Z.
18 (Zhongyang Cao), W.R., and W.Y. analyzed the experimental data. L.M., and Z.Y. (Yanxin
19 Zhao) wrote the manuscript. All authors revised and approved the final manuscript, and
20 agreed to be accountable for this study.

21

22 Supporting information

23 **Figure S1. Frequency distribution of maize grain Cd contents in GWAS plants in four**
24 **environments.** ZZC, spring in ZZ. ZZQ, autumn in ZZ. NXC, spring in NX. NXQ, autumn in
25 NX.

26 **Figure S2. GWAS of maize grain Cd accumulation in four environments.** (a–f) Manhattan
27 plots for grain Cd accumulation in four environments and the best linear unbiased prediction
28 (BLUP) data. The red horizontal line represents the significance cutoff ($P = 7.97 \times 10^{-7}$). NX,
29 Ningxiang. (g–l) Quantile-Quantile plots for the GWAS MLM + Q + K model. BLUP4EN,
30 BLUP values for the grain Cd contents in four environments. BLUPZZ, BLUP values for the

1 grain Cd contents in two seasons in Zhuzhou (ZZ). NXC, spring in Ningxiang (NX). NXQ,
2 autumn in NX. ZZC, spring in ZZ. ZZQ, autumn in ZZ.

3 **Figure S3. *ZmHMA3* and *ZmHMA4* expression levels in diverse tissues.** Relative
4 *ZmHMA3* (a) and *ZmHMA4* (b) expression levels in diverse tissues based on the data in the
5 Maize eFP browser (http://bar.utoronto.ca/efp_maize/cgi-bin/efpWeb.cgi). Color intensity
6 from yellow to red represents the expression levels (low to high, respectively). (c) Differential
7 expression analysis of *ZmHMA3* and *ZmHMA4*. The linearized RMA expression signals in 60
8 diverse tissues were downloaded from the Maize eFP browser
9 (http://bar.utoronto.ca/efp_maize/cgi-bin/efpWeb.cgi). Error bar = standard deviation.

10 **Figure S4. Expression of *ZmHMA4* gene in B73, Jing724 and Mo17.** RT-PCR product
11 electrophoresis of *ZmHMA4* were run on 2% agarose gel. Root and leaf RNAs of B73,
12 Jing724 and Mo17 were isolated for cDNA synthesis. RT-PCR was performed with
13 gene-specific primers HMA61F and HMA2564R shown in **Table S4**. Maize *Actin1* gene was
14 amplified as an internal control with the primers actinF2 and actinR2. M, DL2000 DNA
15 Marker (Beijing TsingKe Biotech Co., Ltd, China). -RT, cDNA synthesized without reverse
16 transcriptase. The gene-specific PCR products were indicated by the orange arrows.

17 **Figure S5. Variations in the *ZmCd1^{Mo17}* transcript.** (a) Number of RNA-seq reads mapped
18 to the *ZmCd1* gene in Mo17. The RNA-seq reads of the high-Cd and low-Cd pools were
19 mapped to the Mo17 reference genome. The read counts for *ZmCd1^{Mo17}* were visualized using
20 the IGV browser (<https://software.broadinstitute.org/software/igv/>). The *ZmCd1^{Mo17}* gene
21 structure was generated based on the *ZmCd1^{B73}* coding sequence. (b) Structures of all 19
22 *ZmCd1^{Mo17}* transcripts. Transcript sequences were determined by sequencing 109 cDNA
23 clones, which were constructed from RT-PCR products using the 7F and 7R primers and
24 Mo17 leaf cDNA as the template. The number of clones and the length of the PCR products
25 for each transcript are indicated in parentheses. The transposable element-inserted regions in
26 introns 1 and 3 are highlighted in red.

27 **Figure S6. Alignment of HMA3 proteins from maize, sorghum, Arabidopsis, rice, wheat,**
28 **and barley.** The amino acid sequences of the following proteins were aligned using Clustal
29 Omega (<https://www.ebi.ac.uk/Tools/msa/clustalo/>) and displayed using GeneDoc
30 (<http://nrbsc.org/gfx/genedoc>): *ZmHMA3* (GRMZM2G175576), *SbHMA3a*

(Sobic.002G083000), AtHMA3 (AT4G30120), OsHMA3 (LOC_Os07g12900), TdHMA3-B1a (NCBI GenBank ID: KF683293), and HvHMA3-B (NCBI GenBank ID: LC523825). The conserved domains were predicted using the Pfam database (<http://pfam.xfam.org/>) and are framed by green, red, and blue lines. Green box, heavy metal-associated domain (PF00403). Red box, E1–E2 ATPase domain (PF00122). Blue box, haloacid dehalogenase-like hydrolase (PF00702).

Figure S7. Grain, stem, and leaf Cd contents of five maize inbred lines and six maize hybrid varieties. The inbred lines and hybrids were grown in a Cd-contaminated field in Zhuzhou in the autumn of 2018. After harvesting, the grains, stems, and leaves were collected and dried before measuring their Cd contents. n = 3, error bar = standard error.

Table S1. Summary of the QTL related to the maize grain Cd content identified in a GWAS.

Table S2. List of the SNPs on chromosome 2 significantly associated with the grain Cd content.

Table S3. SNP-based KASP markers used for fine mapping *ZmCd1*.

Table S4. List of the PCR primers used in this study.

Table S5. List of plant HMA-encoding genes used for constructing a phylogenetic tree.

Table S6. Information regarding the PCR-based markers in maize inbred lines with sequenced genomes.

Table S7. Maize inbred lines and hybrid varieties used in this study.

Supplemental Dataset S1. Alignment of ZmHMA3 from B73, D9H, CML333, Mo17, and Ky21.aln

References

- Ali, U., Zhong, M., Shar, T., Fiaz, S., Xie, L., Jiao, G., Ahmad, S. *et al.* (2019) The influence of pH on cadmium accumulation in seedlings of rice (*Oryza sativa* L.). *J. Plant Growth Regul.* **39**, 930–940.
- Baxter, I., Tchieu, J., Sussman, M.R., Boutry, M., Palmgren, M.G., Gribskov, M., Harper, J.F. *et al.* (2003) Genomic comparison of P-Type ATPase ion pumps in *Arabidopsis* and rice. *Plant Physiol.* **132**, 618–628.
- Bradbury, P.J., Zhang, Z., Kroon, D.E., Casstevens, T.M., Ramdoss, Y., Buckler, E.S. (2007) TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* **23**, 2633–2635.
- Cao, Y.H., Zhao, X.W., Liu, Y.J., Wang, Y.L., Wu, W.M., Jiang, Y.W., Liao, C.J. *et al.* (2019) Genome-wide identification of ZmHMAs and association of natural variation in ZmHMA2 and ZmHMA3 with leaf cadmium accumulation in maize. *PeerJ* **7**, e7877.
- Chen, J.G., Zou, W.L., Meng, L.J., Fan, X.R., Xu, G.H. and Ye, G.Y. (2019) Advances in the uptake and transport mechanisms and QTLs mapping of cadmium in rice. *Int. J. Mol. Sci.* **20**, 3417.
- Floriijn, P.J. and Van Beusichem, M.L. Uptake and distribution of cadmium in maize inbred lines. *Plant Soil* **150**, 25–32.
- Gill, S.S. and Tuteja, N. (2011) Cadmium stress tolerance in crop plants. *Plant Signal. Behav.* **6**, 215–222.
- Jha, U.C. and Bohra, A. (2016) Genomics enabled breeding approaches for improving cadmium stress tolerance in plants. *Euphytica* **208**, 1–31.
- Lei, G.J., Fujii-Kashino, M., Wu, D.Z., Hisano, H., Saisho, D., Deng, F., Yamaji, N. *et al.* (2020) Breeding for low cadmium barley by introgression of a Sukkula-like transposable element. *Nat. Food* **1**, 489–499.
- Li, C.H., Song, W., Luo, Y. F., Gao, S. H., Zhang, R. Y., Shi, Z., Wang, X. Q. *et al.* (2019) The HuangZaoSi maize genome provides insights into genomic variation and improvement history of maize. *Mol. Plant* **3**, 402–409.
- Li, S., Wang, M., Zhao, Z., Ma, C. and Chen, S. (2018) Adsorption and desorption of Cd by soil amendment: mechanisms and environmental implications in field-soil remediation. *Sustainability* **10**, 1–14.
- Liu, C., Guttieri, M.J., Waters, B.M., Eskridge, K.M. and Baenziger, P.S. (2019) Selection of bread wheat for low grain cadmium concentration at the seedling stage using hydroponics versus molecular markers. *Crop Sci.* **59**, 945–956.
- Liu, C.L., Gao, Z.Y., Shang, L.G., Yang, C. H., Ruan, B.P., Zeng, D.L., Biao, G.L. *et al.* (2020) Natural variation in the promoter of *OsHMA3* contributes to differential grain cadmium accumulation between *Indica* and *Japonica* rice. *J. Integ. Plant Biol.* **62**, 314–329.
- Liu, H.J., Luo, X., Niu, L.Y., Xiao, Y.J., Chen, L., Liu, J., Wang, X.Q. *et al.* (2017) Distant eQTLs and non-coding sequences play critical roles in regulating gene expression and quantitative trait variation in maize. *Mol. Plant* **10**, 414–426.
- Liu, S.Z., Yeh, C.T., Tang, H.M., Nettleton, D., Schnable, P.S. (2012) Gene mapping via bulked segregant RNA-Seq (BSR-Seq). *PLoS One* **7**(5):e36406.
- Luo, J.S., Huang, J., Zeng, D.L., Peng, J.S., Zhang, G.B., Ma, H.L., Guan, Y., *et al.* (2018) A defensin-like protein drives cadmium efflux and allocation in rice. *Nat. commun.* **9**, 645.
- Luo, M., Zhang, Y., Chen, K., Kong, M., Song, W., Lu, B., Shi, Y. *et al.* (2019) Mapping of

quantitative trait loci for seedling salt tolerance in maize. *Mol. Breeding* **39**, 64.

Luo, M., Zhao, Y., Zhang, R., Xing, J., Duan, M., Li, J., Wang, N. *et al.* (2017) Mapping of a major QTL for salt tolerance of mature field-grown maize plants based on SNP markers. *BMC Plant Biol.* **17**, 140.

Maccaferri, M., Harris, N.S., Twardziok, S.O., Pasam, R.K., Gundlach, H., Spannagl, M., Ormanbekova, D. *et al.* (2019) Durum wheat genome highlights past domestication signatures and future improvement targets. *Nat. Genet.* **51**, 885–895.

Miyadate, H., Adachi, S., Hiraizumi, A., Tezuka, K., Nakazawa, N., Kawamoto, T., Katou, K., *et al.* (2011) OsHMA3, a PIB-type of ATPase affects root-to-shoot cadmium translocation in rice by mediating efflux into vacuoles. *New Phytol.* **189**, 190–199.

Morel, M., Crouzet, J., Gravot, A., Auroy, P., Leonhardt, N., Vavasseur, A. and Richaud, P. (2009) AtHMA3, a PIB-ATPase allowing Cd/Zn/Co/Pb vacuolar storage in *Arabidopsis*. *Plant Physiol.* **149**, 894–904.

Patrizia, B., Letizia, Z., Angelo, D.P., Di, L.D., Valentina, C., Giuseppina, F., Maurizio, B. *et al.* (2015) Cadmium-inducible expression of the ABC-type transporter *AtABCC3* increases phytochelatin-mediated cadmium tolerance in *Arabidopsis*. *J. Exp. Bot.* **66**, 3815–3829

Pedersen, C.N.S., Axelsen, K.B., Harper, J.F. and Palmgren, M.G. (2012) Evolution of plant p-Type ATPases. *Front. Plant Sci.* **3**, 31.

Qiao, K., Gong, L., Tian, Y.B., Wang, H. and Chai, T.Y. (2018) The metal-binding domain of wheat heavy metal ATPase 2 (TaHMA2) is involved in zinc/cadmium tolerance and translocation in *Arabidopsis*. *Plant Cell Rep.* **37**, 1343–1352.

Rizwan, M., Ali, S., Adrees, M., Rizvi, H., Zia-Ur-Rehman, M., Hannan, F., Qayyum, M.F. *et al.* (2016) Cadmium stress in rice: toxic effects, tolerance mechanisms, and management: a critical review. *Environ. Sci. Pollut. R.* **23**, 17859–17879.

Sasaki, A., Yamaji, N. and Ma, J.F. (2014) Overexpression of *OsHMA3* enhances Cd tolerance and expression of Zn transporter genes in rice. *J. Experimental Bot.* **65**, 6013–6021.

Sudhir, K., Glen, S. and Koichiro, T. (2016) MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **33**, 1870–1874.

Takahashi, R., Bashir, K., Ishimaru, Y., Nishizawa, N.K. and Nakanishi, H. (2012a) The role of heavy-metal ATPases, HMAs, in zinc and cadmium transport in rice. *Plant Signal. Behav.* **7**, 1605–1607.

Takahashi, R., Ishimaru, Y., Shimo, H., Ogo, Y., Senoura, T., Nishizawa, N.K., Nakanishi H. *et al.* (2012b) The OsHMA2 transporter is involved in root-to-shoot translocation of Zn and Cd in rice. *Plant Cell Environ.* **35**, 1948–1957.

Ueno, D., Yamaji, N., Kono, I., Huang, C.F., Ando, T., Yano, M. and Ma, J.F. (2010) Gene limiting cadmium accumulation in rice. *P. Natl. Acad. Sci. USA* **107**, 16500–16505.

Verret, F., Gravot, A., Auroy, P., Leonhardt, N., David, P., Nussaume, L., Vavasseur, A. *et al.* (2004) Overexpression of *AtHMA4* enhances root-to-shoot translocation of zinc and cadmium and plant metal tolerance. *FEBS Lett.* **576**, 306–312.

Wu, T.D., Jens Reeder, J., Lawrence, M., Becker, G., Brauer, M.J. (2016) GMAP and GSNAP for genomic sequence alignment: enhancements to speed, accuracy, and functionality. *Methods Mol. Biol.* 1418, 283–334.

Yan, H.L., Xu, W.X., Xie, J.Y., Gao, Y.W., Wu, L.L., Sun, L., Feng, L. *et al.* (2019) Variation of a

1 major facilitator superfamily gene contributes to differential cadmium accumulation between
2 rice subspecies. *Nat. commun.* **10**, 2562.

3 Yan, J.L., Wang, P.T., Wang, P., Yang, M., Lian, X.M., Zhong, T., Feng, H.C. *et al.* (2016) A
4 loss-of-function allele of *OsHMA3* associated with high cadmium accumulation in shoots and
5 grain of Japonica rice cultivars. *Plant Cell Environ.* **39**, 1941–1954.

6 Yang, N., Lu, Y.L., Yang, X.H., Huang, J., Zhou, Y., Ali, F., Wen, W. *et al.* (2014) Genome wide
7 association studies using a new nonparametric model reveal the genetic architecture of 17
8 agronomic traits in an enlarged maize association panel. *Plos Genet.* **10**, e1004573.

9 Zhang, B.L., Ouyang, Y.N., Xu, J.Y. and Liu, K. (2018a) Cadmium remobilization from shoot to grain
10 is related to pH of vascular bundle in rice. *Ecotox. Environ. Safe.* **147**, 913–918.

11 Zhang, L., Gao, C., Chen, C., Zhang, W., Huang, X.Y. and Zhao, F.J. (2020) Overexpression of rice
12 *OsHMA3* in wheat greatly decreases cadmium accumulation in wheat grain. *Environ. Sci.*
13 *Technol.* **Just accepted**.

14 Zhang, L.X., Wu, J., Tang, Z., Huang, X.Y., Wang, X.W., Salt, D.E. and Zhao, F.J. (2019) Variation in
15 the *BrHMA3* coding region controls natural variation in cadmium accumulation in *Brassica*
16 *rapa* vegetables. *J. Exp. Bot.* **70**, 5865–5878.

17 Zhang, X.D., Zhao, K.X. and Yang, Z.M. (2018b) Identification of genomic ATP binding cassette
18 (ABC) transporter genes and Cd-responsive ABCs in *Brassica napus*. *Gene* **664**, 139–151.

19 Zhang, X.X., Rui, H.Y., Zhang, F.Q., Hu, Z.B., Xia, Y. and Shen, Z.G. (2018c) Overexpression of a
20 functional *Vicia sativa* PCS1 homolog increases cadmium tolerance and phytochelatin
21 synthesis in *Arabidopsis*. *Front. Plant Sci.* **9**, 107.

22 Zhao, X.W., Luo, L.X., Cao, Y.H., Liu, Y.J., Li, Y.H., Wu, W.M., Lan, Y.Z. *et al.* (2018a)
23 Genome-wide association analysis and QTL mapping reveal the genetic control of cadmium
24 accumulation in maize leaf. *BMC Genomics* **19**, 91.

25 Zhao, Z., Zhang, H.S., Fu, Z.J., Chen, H., Lin, Y.N., Yan, P.S., Li, W.H. *et al.* (2018b) Genetic-based
26 dissection of arsenic accumulation in maize using a genome-wide association analysis method.
27 *Plant Biotechnol. J.* **16**, 1085–1093.

28 Zhiguo, E., Tingting, L.I., Chen, C. and Lei, W. (2018) Genome-wide survey and expression analysis of
29 P_{1B}-ATPases in rice, maize and sorghum. *Rice Sci.* **25**, 208–217.

1

2 **Figure Legends**

3 **Figure 1. Maize grain Cd contents and correlations in the GWAS population in four**

4 **environments.** (a, b) Soil Cd content and pH in the spring and autumn in Zhuzhou (ZZ) and
5 Ningxiang (NX) fields. Error bar = standard error, $n = 5$. (c, d) Maize grain Cd contents and
6 correlations in the GWAS population in four environments. Correlation coefficients are
7 provided in (d). Circle size represents the extent of the correlation. Blue and red indicate
8 positive and negative correlations between traits, respectively. The color intensity is
9 proportional to the extent of the correlation. ZZC, spring in ZZ. ZZQ, autumn in ZZ. NXC,
10 spring in NX. NXQ, autumn in NX. *** $P < 0.001$.

11 **Figure 2. GWAS of the maize grain Cd accumulation.** (a) Manhattan plots for the BLUP

12 data of grain Cd accumulation across two seasons in NX. The red horizontal line represents
13 the significance cutoff ($P = 7.97\text{e-}7$). NX, Ningxiang. (b) Quantile-Quantile plots for the
14 GWAS MLM + Q + K model. (c) The GWAS signals for grain Cd accumulation are presented
15 on chromosome 2. The peak SNP (i.e., SNP with the lowest P -value for the identified loci) for
16 candidate gene *GRMZM2G175576* is indicated by a red arrow, whereas the SNPs within the
17 identified QTL interval for grain Cd accumulation are highlighted in green. (d) Phenotypic
18 differences resulting from the two alleles of the lead SNP (Chr2.S_158408660) of
19 *GRMZM2G175576*. (e) Association between the *GRMZM2G175576* genetic variation and the
20 grain Cd content as well as the linkage disequilibrium (LD) between the SNPs in
21 *GRMZM2G175576*. The extent of the LD was determined based on the D' statistic. Red
22 squares without numbers represent complete LD ($D' = 1$, $P < 0.01$). D' values are presented in
23 the squares. The red horizontal line represents the significance cutoff ($P = 7.97\text{e-}7$).

24 **Figure 3. Grain Cd contents of maize inbred lines, hybrids, and the BC₁ population of**

25 **Jing724 and Mo17.** (a) Grain Cd contents of seven maize inbred lines. (b) The Cd contents of
26 heterozygous kernels were unaffected by foreign pollen grains. High-Cd lines (Mo17 and
27 Qi319) were pollinated with the pollen grains of low-Cd lines (B73, Jing724, and
28 Huangzaosi), resulting in heterozygous kernels with high Cd contents. The low-Cd lines were
29 pollinated with the pollen grains from high-Cd lines, resulting in heterozygous kernels with
30 low Cd contents. The inbred lines that provided the pollen grains are in parentheses. (c)

1 Dominance of low grain Cd contents following the selfing of F₁ hybrid plants. Hybrids were
2 obtained by crossing the high-Cd inbred lines (Mo17 and Qi319) with the low-Cd inbred lines
3 (Jing724 and B73). (d) Frequency distribution of the grain Cd contents of 218 plants in the
4 BC₁ population (Jing724 × Mo17) × Mo17. The curve generated by CurveExpert (version 1.4)
5 (<http://www.curveexpert.net/>) revealed two apparent Cd content peaks using 0.1 mg/kg as the
6 threshold: low Cd accumulation peak (n = 116) and high Cd accumulation peak (n = 102).
7 Data were analyzed by Student's *t*-test, n = 3 (a–c).

8 **Figure 4. BSR-seq and fine mapping of the major QTL for the high grain Cd content in**
9 **Mo17.** (a) Probability of the association between 110,085 SNPs and the grain Cd
10 accumulation. A BSR-seq analysis was performed using the RNA of two extreme Cd content
11 pools: 30 plants with a high grain Cd content and 30 plants with a low grain Cd content from
12 the F₂ segregating population obtained by crossing Jing724 with Mo17 followed by selfing.
13 The B73 RefGen_v3 genome was used as the reference sequence (<https://www.maizgedb.org/>)
14 for the BSR-seq analysis and the subsequent fine mapping. (b) Significantly associated SNPs
15 on chromosome 2. The median association probabilities for the SNPs on chromosome 2 were
16 calculated using a sliding window method, with window and step sizes of 20 and 5 SNPs,
17 respectively. (c, d) Fine mapping of *qCd1*. The *qCd1* QTL was mapped to the interval
18 between KASP-7 and KASP-8 using 149 BC₁ plants with a high grain Cd content (more than
19 0.2 mg/kg). New SNP-based KASP markers were developed to narrow the QTL interval to a
20 232-kb region between KASP-12 and KASP-15 using 1,117 high-Cd BC₁ plants. (e) Synteny
21 of the QTL intervals within the B73 and Mo17 genomes. Similar regions in the B73 and
22 Mo17 sequences determined by a local BLAST analysis (e-value of 1e-100) are linked. Blue
23 and red boxes represent the genes annotated in the corresponding regions of the B73 and
24 Mo17 genomes. (f) Genes in the QTL intervals annotated in the B73 and Mo17 reference
25 genomes. The candidate gene *GRMZM2G175576* is highlight in red (e, f).

26 **Figure 5. *ZmCd1* gene structural variations and expression.** (a) *ZmCd1* genomic sequence
27 variations in representative maize inbred lines. The backbone of gene structure was drawn
28 based on the *ZmCd1*^{B73} CDS and genomic sequences. Black boxes indicate exons, whereas
29 black broken lines indicate introns. Insertions in the *ZmCd1* promoter, introns, and exons are
30 labeled with triangles, whereas deletions in the promoter and exon 5 are marked in blue. (b)

1 Results for the RT-PCR analysis of *ZmCdl* expression in B73, Mo17, and Ky21. The cDNA
2 samples synthesized from the leaf RNA of B73, Mo17, and Ky21 seedlings were used for the
3 PCR amplifications with primer pairs 7F/7R and 8F/8R. Primers 7F and 8F target exon 1,
4 whereas primers 7R and 8R target exons 4 and 5, respectively, as shown in (c). Maize *Actin1*
5 cDNA (*GRMZM2G126010*) was amplified as a control. The primers are listed in **Table S4**.
6 The PCR products were analyzed by 2% agarose gel electrophoresis. M, DL5000 DNA
7 Marker (Beijing TsingKe Biotech Co., Ltd, China). –RT, cDNA synthesis without a reverse
8 transcriptase. (c) *ZmCdl* gene structures in B73, Mo17, and Ky21 determined by aligning
9 DNA and transcript sequences. Boxes indicate exons, black boxes indicate coding regions,
10 and black hat lines indicate introns. Only two representative *ZmCdl*^{Mo17} transcripts are
11 presented (**Figure S5**). The transposable element insertion regions within *ZmCdl*^{Mo17} and
12 *ZmCdl*^{Ky21} genes are highlighted in red and green, respectively. (d) *ZmCdl* expression in
13 Jing724 and Mo17 seedlings. Root and leaf total RNA samples for Jing724 and Mo17 at the
14 V3 seedling stage were used for the qRT-PCR using primers targeting exon 1 (**Table S4**).
15 Three biological replicates were used to calculate relative expression levels according to the
16 $2^{-\Delta\Delta Ct}$ method. The *Actin1* gene served as the internal control. Data were analyzed by
17 Student's *t*-test. Error bar = standard error.

18 **Figure 6. High grain Cd contents of *ZmCdl* EMS mutants and results of the allelism test.**

19 (a) *ZmCdl* gene mutation sites in the *zmcd1-1* (EMS4-038f45) and *zmcd1-2* (EMS4-038f39)
20 EMS mutants. (b) Validation of mutation sites by DNA sequencing. Altered nucleotides in the
21 EMS mutants are framed by red dashed lines. (c) Allelism test results for the *ZmCdl* mutants
22 and Mo17. Maize plants grown in a Cd-contaminated field in Zhuzhou in the autumn of 2018
23 were harvested for grain Cd content measurements. The grain Cd contents were significantly
24 higher for the *zmcd1-1* and *zmcd1-2* mutants than for B73 (wild-type). The kernels produced
25 by the F₁ hybrids, Mo17 × *zmcd1-1* and *zmcd1-2* × Mo17, had high Cd contents. Data were
26 analyzed by Student's *t*-test (n = 4). Error bar = standard error.

27 **Figure 7. Phylogenetic relationships among plant HMAs and the subcellular localization**

28 **of *ZmCdl*.** (a) Phylogenetic tree of plant HMAs. The neighbor-joining tree was constructed
29 using MEGA 7 based on the alignment of 43 plant HMA proteins, including 8 AtHMAs in
30 Arabidopsis (Baxter et al., 2003), 9 OsHMAs in rice (Baxter et al., 2003; Zhiguo et al., 2018),

9 SbHMAs in sorghum (Zhiguo et al., 2018), 12 ZmHMAs in maize (Cao et al., 2019; Zhiguo et al., 2018), BrHMA3 in *Brassica rapa* (Zhang et al., 2019), HvHMA3-B in barley (Lei et al., 2020), and TdHMA3-B1a in durum wheat (Maccaferri et al., 2019). The Poisson amino acid substitution model and 1,000 bootstrap replicates were used. The *ZmCd1* (*ZmHMA3*) gene is highlighted in red. (b) *ZmCd1* transmembrane domains predicted by the TMHMM server (version 2.0) (<http://www.cbs.dtu.dk/services/TMHMM/>). (c) Transient expression of the *ZmCd1*-eGFP fusion protein in maize protoplasts. The pM999-*ZmCd1*-eGFP and pM999-AtTIP2;3-mCherry recombinant plasmids were used for the co-transformation of maize protoplasts, which were then examined for fluorescence. The Arabidopsis TIP2;3 protein, which served as a tonoplast-localized control, was anchored to the vacuolar membrane (Zhang et al., 2020). Maize protoplasts co-transformed with pM999-eGFP (blank) and pM999-AtTIP2;3-mCherry were used as the negative control. Scale bar = 10 μ m.

Figure 8. *ZmCd1* gene haplotypes in the GWAS panel. (a) Analysis of the PCR-based molecular markers by 2% agarose gel electrophoresis. The following primers for the M1–M4 molecular markers were designed based on the gene structural variations (**Figure 5a** and **Table S6**): 1F/2F//1R for M1, 3F//2R/3R for M2, 4F//4R/5R for M3, and 5F/6R for M4. Regarding M1–M3, three primers were used in each PCR amplification and the PCR extension time was less than 1 min. The left lane presents the DL2000 DNA Marker (Beijing TsingKe Biotech Co., Ltd, China). The genomic DNA from five representative maize inbred lines (B73, CML333, Mo17, Ky21, and D9H) harboring five different haplotypes of the M1–M4 molecular marker combination was used for the PCR amplification. (b) The M1–M4 molecular marker lengths in B73, CML333, Mo17, Ky21, and D9H. Black boxes indicate the molecular markers that differed in B73. (c) Grain Cd contents for five haplotypes in the GWAS panel. Student's *t*-test was used to compare haplotypes with hap_1. *P < 0.05, ***P < 0.001.

Figure 9. Application of PCR-based markers for evaluating the grain Cd contents of maize inbred lines and hybrid varieties. (a, b) Grain Cd contents (upper panel) and detection of the M1–M4 molecular markers (lower panel) for 24 maize inbred lines (a) and 13 maize hybrid varieties (b). Maize plants were grown in a Cd-contaminated field in Zhuzhou in the autumn of 2018 for an analysis of grain Cd contents. Details regarding the maize inbred

lines and hybrid varieties are listed in **Table S7**. M, DL2000 DNA Marker (Beijing TsingKe Biotech Co., Ltd, China). (c) Grain Cd contents in inbred lines with differing haplotypes. The grain Cd contents are listed in **Table S7**. Data were analyzed by Student's *t*-test. Error bar = standard error.

Tables

Table 1. Summary of the maize grain Cd contents of the GWAS population

Location	Mean±SE (mg/kg)	Range (mg/kg)	Kurt	Skew
NXC	0.06904±0.005653	0.0002833-0.8412	11.68	3.152
NXQ	0.04571±0.002713	0.00078-0.4013	10.72	2.882
ZZC	0.1955±0.01859	0.00504-2.242	11.79	3.319
ZZQ	0.2436±0.02102	0.00754-2.298	6.816	2.603
BLUPZZ	0.2155±0.01056	0.06721-1.337	6.144	2.550
BLUPNX	0.05705±0.002473	0.01900-0.4168	11.21	3.078
BLUP4EN	0.1355±0.005328	0.03575-0.7325	6.341	2.532

Table 2. Analysis of variance for the Cd contents of the GWAS population in four environments

Variance	MS	F	P value	H ² (%)
E	6.8679	1083.4769	0.0000E+000	71.11
G	0.2409	38.0122	0.0000E+000	
G*E	0.0953	15.0383	0.0000E+000	

1 **Table 3. Candidate genes for the maize grain Cd content revealed by a GWAS**

Traits	Peak SNP	Allele	Chr	QTL interval ^a	P value	R ² (%)	Candidate gene	Position	Annotation
BLUP4EN, BLUPNX, BLUPZZ, NXC, NXQ, ZZC, ZZQ	chr2.S_158408660	G/A	2	157504643- 158846650	4.76E -20	20. 03	GRMZM2 G175576	158406991- 158410957	Heavy Metal- Transporting P _{1B} -Type ATPase
BLUP4EN, BLUPNX, BLUPZZ, NXC, NXQ, ZZC, ZZQ	chr2.S_158408660	G/A	2	157504643- 158846650	4.76E -20	20. 03	GRMZM2 G455491	158386097- 158389415	Heavy Metal- Transporting P _{1B} -Type ATPase
BLUP4EN, BLUPNX, BLUPZZ, NXC, NXQ, ZZC, ZZQ	chr2.S_161122814	G/A	2	160628736- 163940592	3.28E -16	23. 78	GRMZM2 G018241	161123950- 161130108	Cellulose synthase-9

2 ^aQTL intervals and candidate gene positions are based on the B73 AGPv2 genome sequence.

3 The QTL intervals comprise the 200-kb sequence ($\pm$ 100 kb) flanking the peak SNP.

4

5

(f)

Gene	Gene Model (B73)	Gene Model (Mo17)	Gene Annotation
1	<i>GRMZM2G455491</i>	<i>Zm00014a038725</i>	P-type heavy metal ATPase
2	<i>GRMZM2G175576</i>	<i>Zm00014a038724</i>	P-type heavy metal ATPase
3	<i>GRMZM5G845339</i>	No homolog	Alpha/beta-Hydrolases superfamily protein, fragment of <i>GRMZM5G820727</i>
4	<i>GRMZM5G876616</i>	<i>Zm00014a038730</i>	Protein of unknown function (DUF567)
5	<i>GRMZM2G333095</i>	<i>Zm00014a038721</i>	Unknown protein
6	<i>GRMZM2G026301</i>	<i>Zm00014a038720</i>	Putative serine/threonine-protein kinase
7	No homolog	<i>Zm00014a038728</i>	Diacylglycerol kinase catalytic domain-containing protein
8	No homolog	<i>Zm00014a038723</i>	Unknown protein

(a)

(b)

(c)

(d)

(a)

(b)

(c)

(a)

(b)

(c)

(a)

(b)

	M1 (in promoter)	M2 (in intron 1)	M3 (in intron 4)	M4 (in exon 5)
B73	776 bp	305 bp	529 bp	188 bp
CML333	776 bp	305 bp	529 bp	122 bp
Mo17	574 bp	630 bp	529 bp	188 bp
Ky21	574bp	305 bp	965 bp	188 bp
D9H	574 bp	305 bp	529 bp	122 bp

(c)