 doi:10.20944/preprints201802.0075.v1

Natural variation OsCd1^{V449} contributes to reducing cadmium accumulation in rice grain

Huili Yan[#1], Wenxiu Xu[#1], Jianyin Xie[#2], Yiwei Gao[3], Lulu Wu[1], Liang Sun[4], Xu Chen[1], Tian Zhang[1], Changhua Dai[1], Xiuni Lin[1], Lu Feng[1], Xueqiang Wang[2], Fengmei Li[2], Xiaoyang Zhu[2], Jinjie Li[2], Zichao Li[2], Caiyan Chen[4], Mi Ma[1], Hongliang Zhang*[2] and Zhenyan He*[1]

[1]Key Laboratory of Plant Resources, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, P. R. China

[2]Key Lab of Crop Heterosis and Utilization, Beijing Key Lab of Crop Genetic Improvement, Ministry of Education, China Agricultural University

[3]College of Biological Sciences and Biotechnology, Beijing Forestry University, Beijing 100083, China

[4]Key Laboratory of Agro-Ecological Processes in Subtropical Region, Institute of Subtropical Agriculture, Chinese Academy of Sciences, Changsha, 410125, China.

[#]These authors contributed equally to this work

[*]Address correspondence to hezhenyan@ibcas.ac.cn, zhangl@cau.edu.cn

Abstract

Cadmium accumulation in rice grain poses a serious threat to people's health. Understanding the genetic basis on grain cadmium accumulation facilitates efforts to reduce it. Here, we show that *OsCd1* is involved in Cd uptake and contributes to grain accumulation in rice. Natural variation in *OsCd1* with a missense mutation Val449Asp is responsible for the divergence of rice grain cadmium accumulation between *japonica* and *indica*. Notably, near-isogenic line tests confirmed that the *indica* variety carrying the OsCd1^{V449} allele could reduce the grain cadmium accumulation. The favorable allele OsCd1^{V449} may be an important genetic resource to reduce grain cadmium accumulation for *indica*.

Keywords: OsCd1, Cadmium accumulation, rice grain, *indica*, *japonica*, GWAS

 doi:10.20944/preprints201802.0075.v1

Corresponding author contact details

Zhenyan He

Institute of Botany

Chinese Academy of Sciences

Xiangshan, Beijing 100093

China

Telephone: 0086-10-62836690

E-mail: hezhenyan@ibcas.ac.cn

Hongliang Zhang

Key Lab of Crop Heterosis and Utilization, Beijing Key Lab of Crop Genetic Improvement, Ministry of Education, China Agricultural University

No. 2 Yuanmingyuan West Road, Haidian District, Beijing, China

Telephone: 0086-10-62734018

E-mail: zhangl@cau.edu.cn

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 February 2018 doi:10.20944/preprints201802.0075.v1

Introduction

Cadmium (Cd) is a toxic heavy metal and can lead to Cd-related diseases such as renal tubular dysfunction and bone disease. Rice is a major component of diet for over half of the world's population. The accumulation of Cd is a serious threat to human being since it can be concentrated in body through the food chain[1] and the biological half-life is estimated to be nearly 30 year[2]. Molecular genetic tools have been urgently sought to develop 'low-Cd rice' to reduce potential health risks.

Generally in rice, Cd is first absorbed from soil by roots, and then translocated to shoots, and accumulated into grains finally[3]. Transport systems, which are responsible for the passage of Cd through the membrane, play crucial roles in the Cd accumulation of rice. So far, the characterization of membrane proteins, such as OsNramp1[4], OsNramp5[5], OsHMA2[6], OsHMA3[7] and OsLCT1[8], has enriched knowledge of Cd transport in several specific cell and issue types. However, the rice Cd transport mechanism is complex and grain Cd accumulation is a result of cooperative interactions among multiple cells and issues. It is therefore worthwhile to explore more genetic loci involving in Cd accumulation of rice grain.

Rice varieties exhibit substantial genetic variation with respect to Cd accumulation ability[9-12], which is a valuable resource for dissecting functional alleles and genetic improvement. It was reported that OsHMA3 showed different Cd transport ability[8] and OsNRAMP1 expressed in a different level among two cultivars[4,7]. Current understanding of the genetic basis of rice Cd accumulation diversity remains at the level of the identification of several quantitative trait loci (QTLs)[13-15]. The natural allelic variations responsible for rice varietal differences have not been fully explored and genetic basis of grain Cd accumulation differences remains unknown. The genome-wide association studies (GWAS) can be useful tools for identification of allelic variations underpinning the grain Cd accumulation diversity in rice.

Here, we identified a major facilitator superfamily (MFS) protein OsCd1 associated with Cd accumulation *via* GWAS. It diverges between *indica* and *japonica* and the natural variation OsCd1^{V449} in

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 February 2018 doi:10.20944/preprints201802.0075.v1

japonica was associated with a reduced Cd uptake ability and decreased grain Cd accumulation.

RESULTS

OsCd1 is associated with grain cadmium accumulation in the rice

In this study, we used a composite method, referred to as a 'bioinformatics tool-box' **(Fig S1)**, in which data from GWAS, gene annotation and the yeast assay were combined to find genes underlying rice grain Cd accumulation. A set of 127 rice cultivars including 41 *japonica* and 86 *indica* from mini-core collection of rice in Chian[16] and other wide-spread regions was used to carry out GWAS of grain Cd concentration (**Fig S2a-b**). After testing rice varieties with a Cd treatment assay, we found that *indica* varieties could be phenotypically distinguished from *japonica* varieties by their significantly higher grain Cd accumulation. (**Fig S2c and Table S1**). Using 3,291,150 single-nucleotide polymorphisms (SNPs) with a minor allele frequency (MAF) > 0.05 covering the whole rice genome, we performed GWAS to identify the genetic loci associating with grain Cd accumulation. Under the compressed general mixed linear model ($P<1\times10^{-5}$, MLM, threshold derived from 1000 permutation test) (**Fig 1a-b**), 12 QTLs were identified as significantly associated with grain Cd accumulation **(Table S2)**. According to the rice genome annotation project website (MSU-RGAP), 494 genes were annotated to be located in the candidate locus (**Table S3**). To identify the membrane transporter conferring Cd accumulation in rice, we then performed the gene ontology (GO) Slim analysis to selected candidate genes associated with the key words 'membrane' and 'transport' for further research **(Table S4)**. Altogether, 13 candidate genes located on chromosome 2, 3, 4, 5, 6 and 10 were found to be associated both with 'transport' and 'membrane' in GO Slim annotation **(Fig 1c)**. Four candidate genes were selected on QTL3 of chromosome 3 which explained approximately 20.7% of the phenotypic variation **(Fig 1d)** and transformed into *S. cerevisiae* to evaluate whether the candidate genes were involved in the transport of Cd.

Spot assays were tested with and without 30 μM $CdCl_2$ and the growth rate was normalized by non-transgenic strains. As a result, only the yeast transformed *C2* (*LOC_Os03g02380*) displayed sensitivity to Cd

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 February 2018 doi:10.20944/preprints201802.0075.v1

(**Fig 1e**). Under 24h Cd exposure, expression of *C2* caused enhanced Cd accumulation in a Cd dose-dependent manner while the other two showed no significant difference with the control (**Fig 1f**). These results indicated that *C2* was the most likely candidate transporter gene in QTL3 and it was christened '*OsCd1*'.

OsCd1 encoded a protein belonging to the MFS with 12 bilayer-spanning domains and a signature motif of MFS 'G(93)slaD(97)kqG(100)rkR(103)'[17] located in the second cytoplasmic loop (**Fig S3a-b**). There are 149 MFS superfamily members in rice and several of them were predicted to transport various substrates (**Table S5**). A phylogenetic tree was constructed for *OsCd1* along with all the MFS proteins in rice (**Fig S4**). It was clearly shown that the proteins facilitating the same substrate transport share a closer phylogenetic relationship, while *OsCd1* clusters separately and forms an unknown-function clade. Together with the yeast spot assay results, we deduced that *OsCd1* may form a new-function clade in MFS involving in Cd transport.

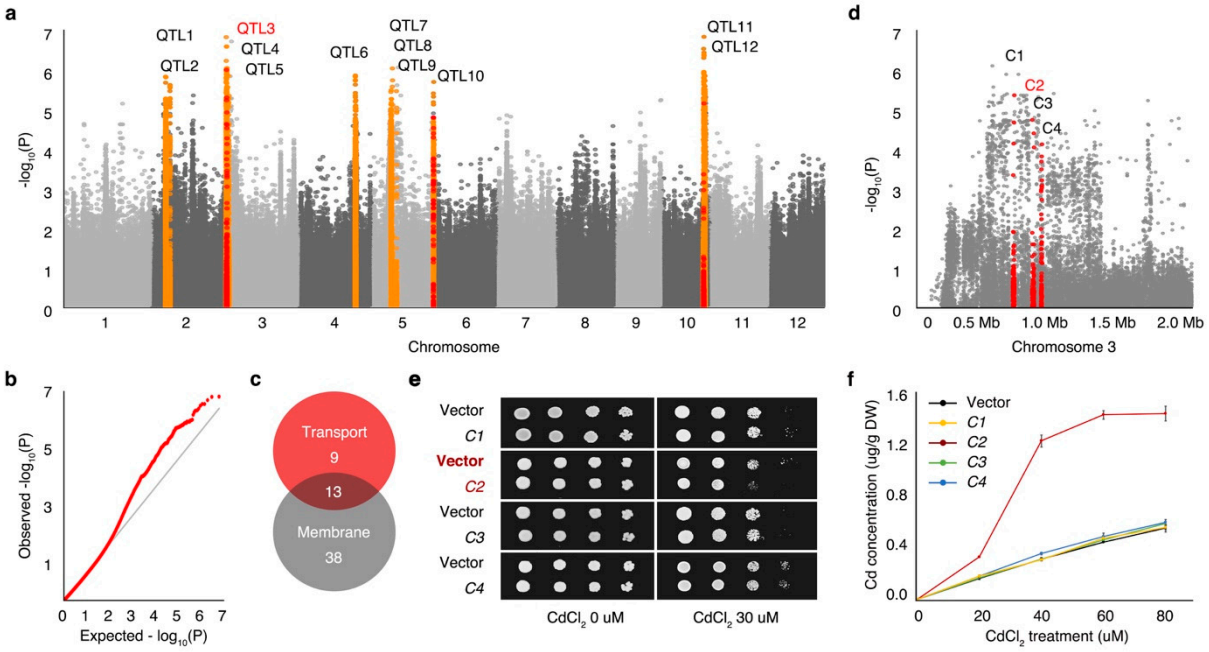

Figure 1 GWAS of the grain Cd accumulation and the function characterization of candidate genes

(a) Manhattan plots for grain Cd accumulation in diverse population. The orange pots indicate the 12 QTLs were identified as significantly associated with grain Cd accumulation. The red dots indicate the 14 candidate genes associated both with 'transport' and 'membrane' in GO Slim annotation. **(b)** The QQ plots for GWAS MLM model. **(c)** The numbers of genes annotated with transport and membrane. **(d)** The genome-

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 February 2018 doi:10.20944/preprints201802.0075.v1

wide association signals for grain Cd accumulation are shown in the region of 0-2 Mb on chromosome 3 (x axis). The locations of four candidate genes *C1*(*LOC_Os03g02150*), *C2* (*LOC_Os03g02380*), *C3* (*LOC_Os03g02390*) and *C4* (*LOC_Os03g02480*) are indicated with red pot respectively. **(e)** Dilution-series spot assays of yeast treated with and without 30 μM $CdCl_2$. **(f)** Cd accumulation in wild type W303 (black) and W303 expressing *C1* (yellow), *C2* (red), *C3* (green) and *C4* (blue) treated with $CdCl_2$ for 24h. Error bars, mean ±SE. OD, optical density.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 February 2018 doi:10.20944/preprints201802.0075.v1

OsCd1 altered the cadmium uptake and grain cadmium accumulation in rice

Considering the possible function of *OsCd1* in Cd transport, we transformed the *CRISPR/Cas9* constructs in rice callus and regenerated the transgenic plants to investigate the role of *OsCd1* in rice. Three Cas9-positive lines (*oscd1*$_{CR-1}$, *oscd1*$_{CR-2}$ and *oscd1*$_{CR-3}$), having either an insertion or deletion of one or few bases at the target sequences, were selected for further research (**Fig 2a**). In vegetative growth stage, the growth without $CdCl_2$ treatment was not affected by the mutation of *OsCd1*. After being treated with 20μM $CdCl_2$ for 10 days, the *CRISPR-oscd1* lines displayed a better growth (**Fig 2b**) and the Cd concentration both in root and shoot were much lower than that in the wild-type rice (**Fig 2c-d**). We then grew them in the field until ripening to investigate the physiological role of *OsCd1* in grain Cd accumulation. At harvest, both the growth and the yield were affected by the disruption of *OsCd1* **(Fig 2e)**. The biomass of the straw (**Fig 2f**) and the filled spikelets were reduced in the mutant lines due to decreased fertility (**Fig S5**). As to the Cd concentration, the result showed that there was a decrease in the straw of three *CRISPR-oscd1* lines compared with the wild type. The synchrotron radiation microscopic X-ray fluorescence (SR-μXRF) scanning was further used to reveal the spatial disparity in the distribution of Cd throughout the grain *in situ*. As it is shown in **Fig 2h**, Cd enriched both in the endosperm and aleurone layer, and its concentration was in a lower level in *CRISPR-oscd1* lines compared with the wild type. Moreover, the Cd concentration was significantly decreased in brown rice and husk in three *CRISPR-oscd1* lines, which was in well agreement with the SR-μXRF results above (**Fig 2i-g**).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 February 2018 doi:10.20944/preprints201802.0075.v1

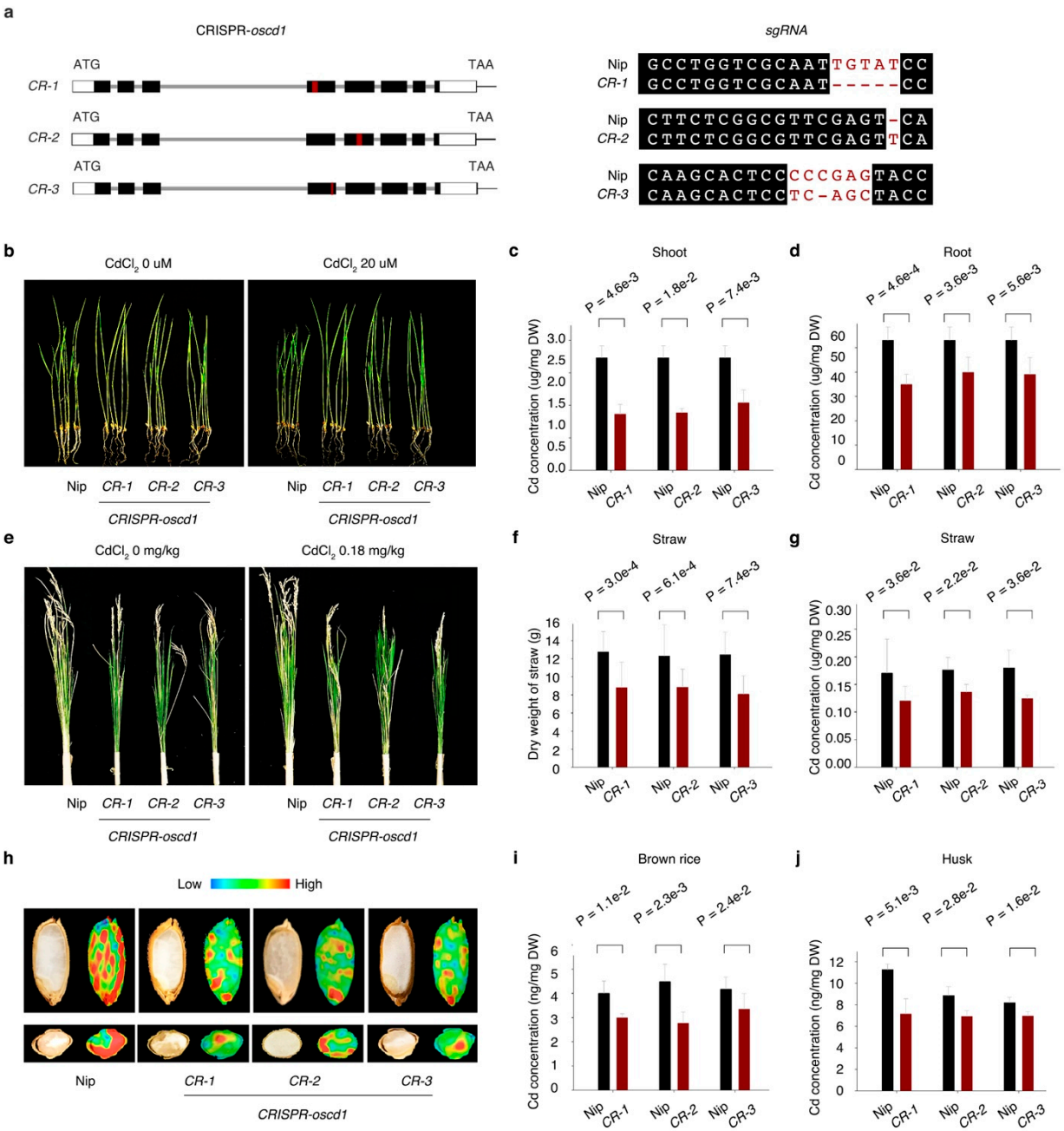

Figure 2 *OsCd1* contributes to the Cd uptake and grain Cd accumulation in rice

(a) Left: Schematic diagram of three *CRISPR-oscd1* lines. Black rectangles represent exons of *OsCd1*, red rectangle represents the exon of target sequence. Right: Sequences of *CRISPR-oscd1* alleles *CR-1*, *CR-2* and *CR-3*. sgRNA targets sequences are showed and deletions are indicated by red dashes. **(b)** Growth of wild-type rice and three *CRISPR-oscd1* lines after treated with and without 20 μM $CdCl_2$ for 10 days. **(c, d, g, i and j)** Concentration of Cd in the shoot (**c**), root (**d**), straw **(g)**, brown rice (**i**) and husk (**j**). The *CRISPR-oscd1* lines were shown in red and the wild-type line was shown in black. Error bars, mean ±SE. Statistical

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 February 2018 doi:10.20944/preprints201802.0075.v1

comparison was performed by Wilcoxon matched-pairs signed-rank. All data were compared with *Nipponbare*. **(e)** Growth of wild-type rice and three *CRISPR-oscd1* lines after treated with and without 0.18 mg/kg Cd at harvest. **(f)** Dry weight of straw. **(h)** SR-μXRF images of Cd distribution in the longitudinal (upper) and latitudinal (lower) sections of rice grain. The emission intensity of each pixel was normalized using the beam intensity as reference.

To determine the expression profile of *OsCd1* in rice, we analyzed the expression from root and shoot at vegetative growth stage in rice *via* qPCR. The results showed that *OsCd1* was mainly detected in roots, which was consistent with the GUS histochemical assay (**Fig 3a**). The cell-specific expression of *OsCd1* in rice root tissues was further analyzed using the GUS reporter fused to the *OsCd1* promoter. GUS activities were mainly detected in the root exodermis and parenchyma cells (**Fig 3b**). We then constructed the GFP fusion protein to examine the subcellular location of *OsCd1* in rice. The green fluorescence was localized mainly in the cytosol and nucleus of cells expressing GFP alone while the *OsCd1* fused with GFP was only observed at the periphery of the cells. The merged images of GFP and plasma membrane marker FM4-64 further confirmed the plasma membrane subcellular localization of *OsCd1* (**Fig 3c**). Together with the *CRISPR-oscd1* results, it is suggested that the *OsCd1*, which was a plasma membrane protein in root, involving in Cd uptake and contributed to grain Cd accumulation in rice in final.

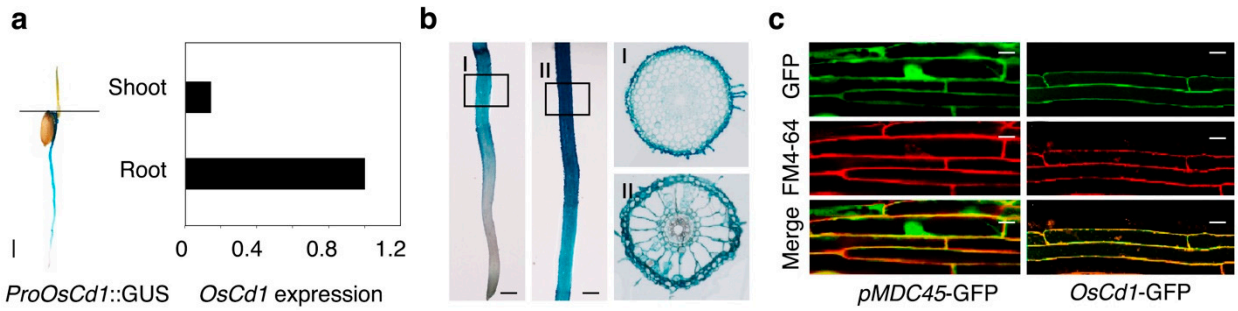

Figure 3 The expression pattern of *OsCd1* in the rice

(a) Tissue-dependent expression of *OsCd1* at vegetative stage. Plants were grown in a nutrient solution

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 February 2018 doi:10.20944/preprints201802.0075.v1

for 7 d. Left: GUS histochemical assay of transgenic plant with the GUS reporter driven by the *OsCd1* promoter. Right: the expression level of *OsCd1* in root and shoot *via* qPCR. Bars = 2 mm **(b)** Tissue-specific localization of *OsCd1* in the root of rice using the GUS reporter driven by the *OsCd1* promoter. **(c)** Subcellular location of *OsCd1* in the root of rice observed by confocal laser scanning microscopy. The GFP fluorescence, fluorescence of FM4-64 and overlay of FM4-64 and GFP of pMDC45-GFP (left) and OsCd1-GFP (right) were shown respectively. Bars = 20 μm.

A nonsynonymous mutation SNP22 altered the cadmium transport ability of *OsCd1*

The full-length sequences of 127 rice cultivars were used to investigate functional allelic variations in *OsCd1* locus. A total of 24 SNPs were identified in the *OsCd1* genomic region: 3 SNPs in the 5'UTR, 3 SNPs in the exon, 17 SNPs in the intron and 1 SNP in the 3'UTR, respectively (**Table S6**). In specific, only the SNP22 (g. 846037 T>A), a missense mutation in eighth exon, resulted in a negative amino acid valine corresponding to a neutral one aspartic (a p. Val449Asp substitution) **(Fig 4a)**. *OsCd1* was a typical MFS protein with 12 transmembrane a-helices. Additionally, there were three extra intracellular helical domains (ICHs) on the intracellular side, one at the C terminus (ICH1) and the other two (ICH2 and ICH3) locating between the amino- and C-terminal TM bundles. The amino acid substitution V449D located on the C-terminal and around the ICH1 from the cytoplasmic face of the plasma membrane (**Fig 4b**).

To find the haplotype analysis between SNP22 with rice grain Cd accumulation, we then conducted the haplotype analysis for SNP22 and grain Cd accumulation in 127 rice cultivars. The results showed that the *OsCd1* can be classified into two genotypes based on the SNP22: $OsCd1^{D449}$ with the A on SNP22 (n=82) and $OsCd1^{V449}$ with the T on SNP22 (n=34). Statistically, lines with $OsCd1^{D449}$ have an approximately 2-folds grain Cd concentration compared to those with $OsCd1^{V449}$ ($P=6.68\times10^{-8}$) (**Fig 4c and Table S7**) but with no obvious difference in expression levels. We further analyzed the Cd concentration in $OsCd1^{V449}$ and

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 February 2018 doi:10.20944/preprints201802.0075.v1

OsCd1^{D449} transgenic yeast strains to investigate if the variation caused by SNP22 might alter the Cd transport ability of *OsCd1*. As shown in **Fig 4d**, OsCd1^{D449} resulted in an approximately 2- to 3-folds Cd concentration compared to OsCd1^{V449} under Cd treatment. Meanwhile, the expression level showed that there was no significant difference between the OsCd1^{V449} and OsCd1^{D449} transgenic lines with or without Cd treatment (**Fig S6**). In this regard, we concluded that the nonsynonymous mutation SNP22 alters the Cd transport ability of *OsCd1*.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 February 2018 doi:10.20944/preprints201802.0075.v1

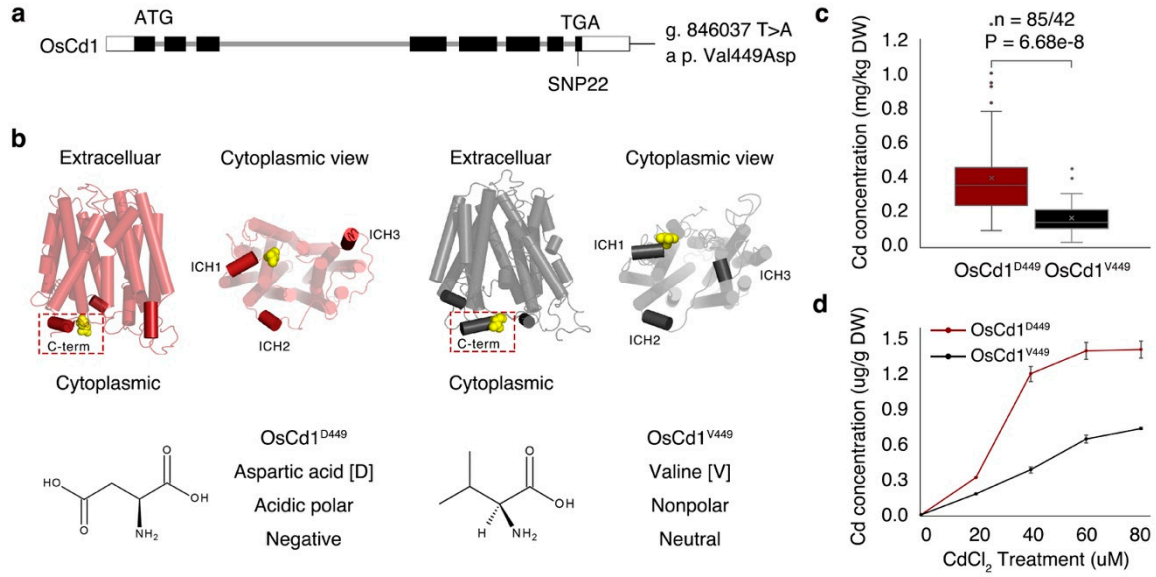

Figure 4 Functionally important amino acids in cadmium transporter *OsCd1*

(a) Schematic diagram of gene structure and SNPs in *OsCd1*. **(b)** Predicted three-dimensional structural model of *OsCd1*. The structures of OsCd1^{D449} (red) are viewed parallel (left one) and perpendicular (right one) to the membrane, respectively, and so for OsCd1^{V449} (grey). Residue Asp449 in OsCd1^{D449} (red) and Val449 in OsCd1^{V449} (grey) are labeled at the cytoplasmic end respectively. ICH indicates intracellular helix. All structure figures were prepared with PyMol. **(c)** Genotypes of *OsCd1* among rice natural variations. The grain Cd distribution of each genotype group is displayed by the box plot; n denotes the number of genotypes belonging to each genotype group. Statistical comparison was performed by student's t-test. **(d)** Cd accumulation in W303 expressing OsCd1^{D449} (red) and OsCd1^{V449} (black) yeast treated with $CdCl_2$ for 24 h. Error bars, mean ±SE. OD, optical density.

OsCd1^{V449} introgression reduced the grain cadmium accumulation

Considering that OsCd1^{V449} exhibit lower Cd transport ability than that of OsCd1^{D449}, we generated near-isogenic line (NIL) by the introgression OsCd1 ^{V449} of *Nipponbare* (*japonica*) into the background of 9311 (*indica* with OsCd1^{D449} genotype) for further research (**Fig S7a and Table S8**). The effect of the allele makeup at *OsCd1* on grain Cd accumulation was firstly investigated in pot-grown rice at greenhouse. As it was shown

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 February 2018 doi:10.20944/preprints201802.0075.v1

in **Fig S7**, 9311 displayed higher Cd concentration in rice grains than the female parent *Nipponbare* lines **(Fig S7b-c)**. With the introduction of OsCd1^{V449}, the Cd accumulation of NIL was remarkably reduced both in brown rice and husk compared with the 9311 backgrounds and no apparent difference in growth **(Fig 5a-c)**.

To evaluate the effect of OsCd1^{V449} allele in *vitro*, field trial in Cd-polluted fields at Changsha (Hunan Province, China, E112°, N28°)[18] was further performed. At harvest, there was no apparent difference in plant growth between the NIL and 9311. Moreover, the filled spikelets and actual yield per plot were also not affected by the introgression of OsCd1^{V449} (**Fig 5d-f**). Notably, the grain Cd concentration was detected to be reduced in NIL **(Fig 5h-i)**. According to the SR-μXRF results, the Cd distribution in whole grain of NIL, especially the edible part endosperm, were significantly less than that in 9311 **(Fig 5g)**. For other micronutrients, except for a slightly decrease of Mn, no obvious difference of the concentration of Zn and Fe was observed between 9311 and NIL **(Fig 5g-o)**. These results indicated that the allelic effect of OsCd1^{V449} contributes to the reduction of Cd in rice and, furthermore, may have a potential application value in rice genetic improvement.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 February 2018 doi:10.20944/preprints201802.0075.v1

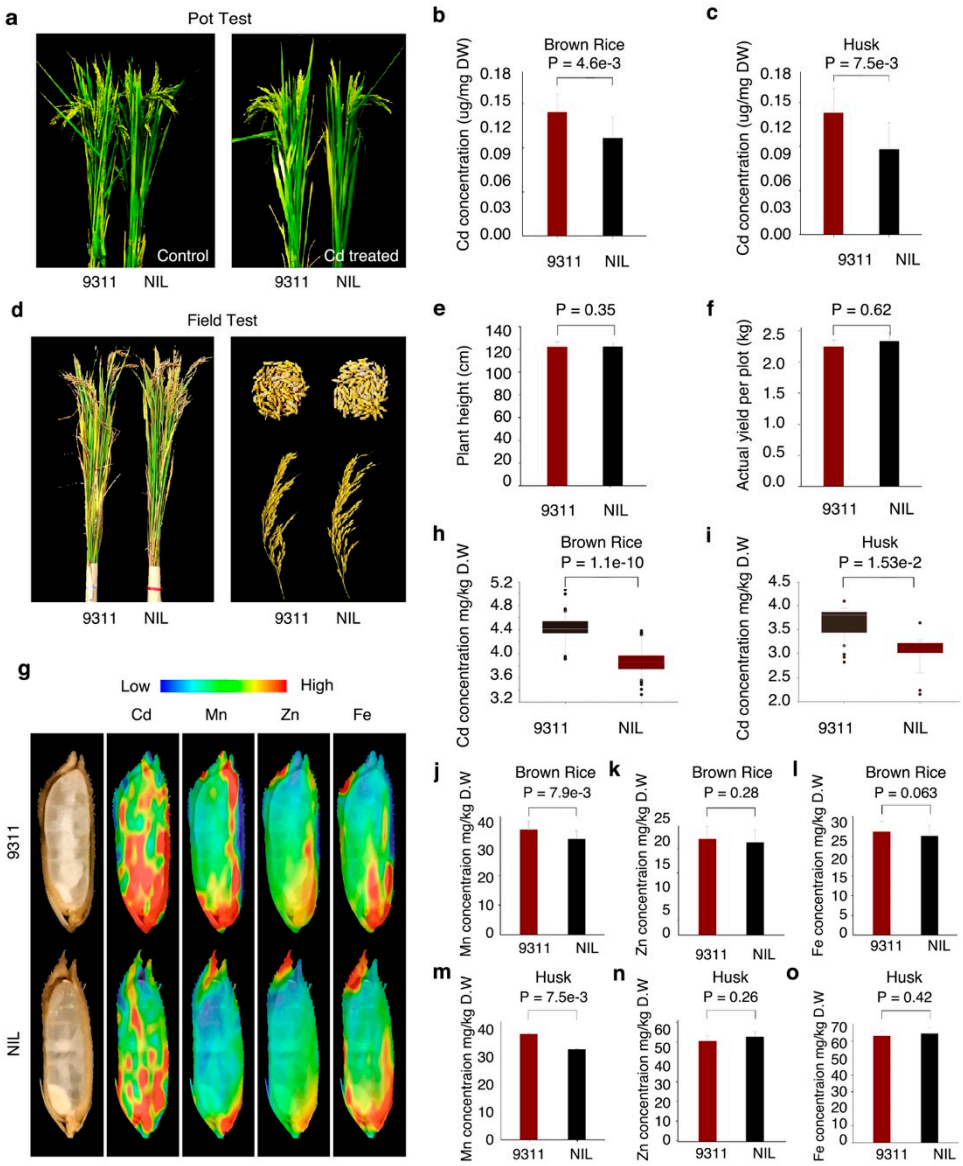

Figure 5 OsCd1^{V449} introgression reduces grain Cd accumulation

(a) Growth of 9311 and NIL after treated with or without 0.18 mg/kg Cd in greenhouse. **(b and c)** Concentration of Cd in the brown rice and husk. **(d)** Growth and grains morphologies of 9311 and NIL after treated with 1.8 mg/kg Cd in the field at harvest. **(e)** Plant height of 9311 and NIL. **(f)** Actual yield per plot of 9311 and NIL. **(g)** SR-μXRF images of Cd, Mn, Zn, Fe distribution in the longitudinal sections of rice grain. The emission intensity of each pixel was normalized using the beam intensity as reference. **(h, j, k and l)** Concentration of Cd, Mn, Zn, Fe in the brown rice. **(i, m, n and o)** Concentration of Cd, Mn, Zn, Fe in the husk. The 9311 were shown in red and the NIL was shown in black. Error bars, mean ±SE. Statistical analysis of pot test was performed by Wilcoxon matched-pairs signed-rank and filed test was inspected using student's

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 February 2018 doi:10.20944/preprints201802.0075.v1

t-test. All data were compared with 9311.

Natural variation of *OsCd1* contributed to grain cadmium accumulation divergence between *indica* and *japonica*

Interestingly, it was also observed that the OsCd1 in most of the *indica* accessions was with $OsCd1^{D449}$ (A at SNP22 site) and that in most of the *japonica* accessions was with the $OsCd1^{V449}$ (T at SNP22 site) **(Table S9)**. To explore the phylogenetic relationship of this allele, we then used whole-genome sequencing data for a large panel of accessions[19] containing 446 *O. rufipogon* accessions and 950 *O. sativa* varieties to determine the ancestral states of the SNP22 in *OsCd1*. Phylogenetic analysis using the 950 cultivar rice accessions showed that all the *japonica* accessions have nucleotide T (genotype $OsCd1^{V449}$) at SNP22 while 99% of the *indica* accessions have nucleotide A (genotype $OsCd1^{D449}$) **(Fig 6a, Table S10).** Moreover, both the nucleotide A and T in SNP22 retained in its ancestor *O. rufipogon* (**Fig 6a, Table S11)**.

Based on phylogenetic tree and PCA analysis, it was demonstrated that *OsCd1* diverged between *indica* and *japonica* subspecies in cultivated rice and between subgroup Or-I and Or-III in wild rice **(Fig 6b)**. Further analysis revealed that *OsCd1* in *indica* was descended from *O. rufipogon* Or-I, which mainly distribute in South and Southeast Asia, while that in *japonica* was from Or-III, which mainly distribute in China[19] (**Fig 6c and S8**). The level of population differentiation (*Fst*) was estimated to be 0.06 in *O. rufipogon* and increased to 0.11 in *O. sativa* during domestication **(Fig 6d and Table S12)**. No significant Tajima's *D* values were observed for *japonica* (Tajima's *D* 1.16) and *indica* (Tajima's *D* -1.32) cultivars, indicating that the divergence of *OsCd1* between *indica* and *japonica* did not subject to selection and derived from the divergence of their progenitor gene pools hundreds of thousands of years ago.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 February 2018 doi:10.20944/preprints201802.0075.v1

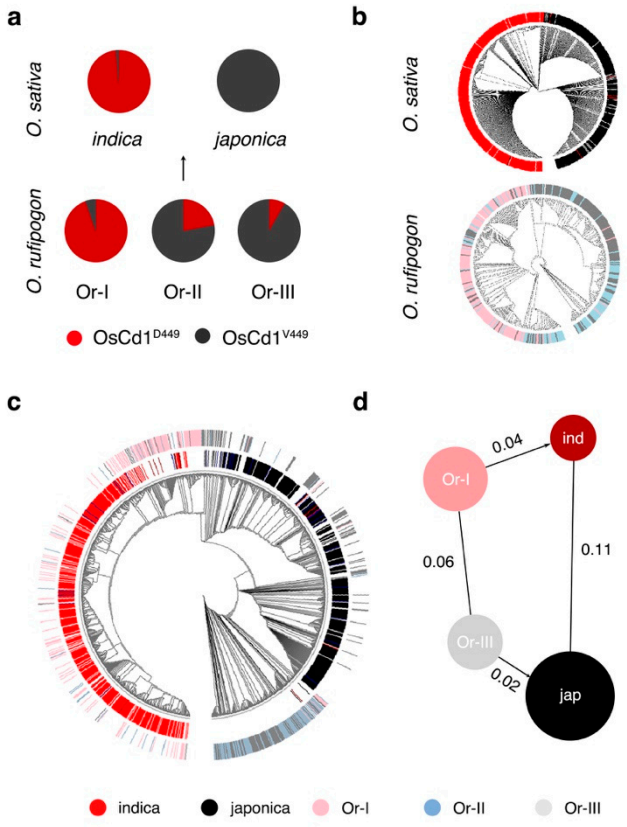

Figure 6 Genotypes and phylogenetic analysis of *OsCd1*

(a) The spectrum of allele frequencies at the causal polymorphisms of *OsCd1* in *O. rufipogon* (Or-I, Or-II and Or-III) and *O. sativa* (*japonica and indica*). **(b)** Phylogenetic tree of the full population calculated from the SNPs of *OsCd1* in *O. sativa* (*indica* and *japonica* subspecies, up) and *O. rufipogon* (Or-I, Or-II and Or-III, down). **(c)** Phylogenetic tree of the full population calculated from the SNPs of *OsCd1* in *O. rufipogon* (outer ring) and *O. sativa* (inner ring). **(d)** Illustration of genetic diversity and population differentiation in *O. rufipogon* and *O. sativa*. The *Fst* values between the groups are indicated. ind, *indica*; jap, *japonica*. Or-I, Or-II and Or-III are colored in pink, light blue and grey respectively; *indica* and *japonica* subspecies are in red and black respectively.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 February 2018 doi:10.20944/preprints201802.0075.v1

DISCUSSION

Reducing rice grain Cd accumulation is vital for human health. Identifying the genetic components underlying Cd transport is of great importance to reduce Cd accumulation in rice grain. In this study, a candidate locus on chromosome 3 was identified to be significantly associated with rice grain Cd accumulation *via* GWAS. We detected that *OsCd1* was associated with Cd transport *via* a composite 'bioinformatics tool-box' method (**Fig1 and S1**). *OsCd1* was a plasma membrane protein in rice root **(Fig 3)**. It is involve in the Cd uptake in root and, ultimately, contributing to grain Cd accumulation together with the reported gene *OsNRAMP5*[5] in rice **(Fig 2)**. OsCd1 belongs to the MFS, which is the largest group of secondary active membrane transporters. Its members transport a diverse range of substrates[17]. In previous research, MFS proteins were reported to be functioned as transporters of some metal elements such as zinc (Zn) transporter ZIF1[20] and cesium (Cs) and potassium (K) transporter ZIFL2[21] (**Fig S4**). Here, we add a new transport substrate member, Cd, to MFS.

We conducted research to excavate natural variation allele in *OsCd1*. Based on the sequences, we identified a nonsynonymous mutation (SNP22), locating at the ICH1 of C-terminal **(Fig 4b)**, can classify *OsCd1* into two genotypes: $OsCd1^{D449}$ and $OsCd1^{V449}$. MFS protein translocate the substrate *via* the "rocker-switch" mechanism[22-24]. N- and C-terminal, especially the charged or polar residues in the ICH domain, are likely to have a critical role to close the transporter on the intracellular side during the alternating access cycle[25]. When the V449 was substituted to D449 in SNP22, the polarity change caused a relatively higher Cd transport ability in $OsCd1^{D449}$ compared with $OsCd1^{V449}$ (**Fig 4d**). The difference in Cd transport ability, finally, endowed the rice cultivars with $OsCd1^{D449}$ with a higher grain Cd accumulation than the cultivars with $OsCd1^{V449}$ (**Fig 4c**). Thus, we believed that SNP22 in *OsCd1* can act as a functional allele for marker-assisted breeding.

9311, which showed a relatively higher grain Cd content compared with *Nipponbare*, was selected as the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 February 2018 doi:10.20944/preprints201802.0075.v1

background of NIL. When the allele containing OsCd1^{V449} was introduced into the 9311, a lower grain Cd accumulation NIL was obtained, indicating a great potential in reducing rice grain Cd accumulation by introducing the OsCd1^{V449} allele via molecular breeding techniques (**Fig 5 b-c and h-i**). Meanwhile, it is noteworthy that the NIL with lower grain Cd accumulation has no apparent differences in plant growth and spikelet fertility compared with 9311 (**Fig 5d-f**). The deletion of *OsCd1* had a negative impact on plant growth and reduced yield at harvest (**Fig 2e**), likely due to the involvement of *OsCd1* in the shoot development of later stages as reported previously[26], indicating that the disruption of *OsCd1* is not an optimal strategy to establish low-Cd rice, while the favorable allele OsCd1^{V449} may be an important genetic resource to reduce grain Cd accumulation for rice.

Remarkably, phylogenetic and population genetic analyses demonstrated the natural variation in *OsCd1* diverged between *indica* and *japonica* subspecies: the *japonica* accessions with genotype OsCd1^{V449} displayed a relatively lower Cd accumulation in rice grain than the *indica* accessions were with genotype OsCd1^{D449}. Asian cultivated rice is classified into *indica* and *japonica* subspecies. *Indica* cultivars, which are the major food and widely grown in the South and Southeast Asia, generally exhibit higher grain Cd accumulation than *japonica* cultivars[10,27] (**Fig S2c**). Since Cd is one of the most serious soil heavy metal pollutant in these regions[28,29], the ingestion of *indica* rice grain bring an unneglectable risk of Cd to people living there. Here, by integrating the 'low-Cd allele' OsCd1^{V449} into *indica* cultivar 9311, the grain Cd accumulation were successfully reduced, suggesting the application value of OsCd1^{V449} in a wider range of *indica* backgrounds.

URLs. Rice HapMap3, http://202.127.18.221/RiceHap3/; Rice Genome Annotation Project (MSU-RGAP), http://rice.plantbiology.msu.edu/cgi-bin/gbrowse/rice/; TransportDB 2.0 http://www.membranetransport.org/transportDB2/index.html

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 February 2018 doi:10.20944/preprints201802.0075.v1

Accession codes. This study made use of GenBank accessions for *OsCd1* (AP014959.1).

ACKOWLEGEMENT

This work was supported by the grants from Chinese Academy of Sciences (XDA08010406), the Ministry of Science and Technology of China (2015FY111300, 2017YFD0800901) and Innovation Training Programs for Undergraduates, CAS (201717001751). We thank Fengqin Dong for help in the resin semi-thin section and Jingquan Li for microscopy experiments. The μ-XRF beam time was granted by 4W1B beamline of Beijing Synchrotron Radiation Facility, Institute of High Energy Physics, Chinese Academy of Sciences. The staff members of 4W1B are acknowledged for their support in measurements and data reduction.

AUTHOR CONTRIBUTION

Z.H.＆H.Z. designed the research and M.M. ＆C.C. conducted the experiments. H.Y., W.X., Y.G., L.W., X.C., C.D., T.Z., X.L. and L.F. contributed to research. J.X., W.X., Q.W., M.L., Y.Z., J.L. and C.L. performed the GWAS analysis. H.Y., W.X. and L.S. contributed to the field management. H.Y. and Y.G. analyzed data. Z.H. and H.Y. wrote the manuscript.

Methods

GWAS analysis and gene annotation

A diverse worldwide collection of 127 *O. sativa* accessions including both landraces and elite varieties was obtained from a core/mini-core collection of *Oryza sativa* L. for GWAS analysis[30]. Each accession was germinated and planted in cultivation pot treated with 1.8 mg/kg Cd (Beijing, China; 39.9°N, 116.3°E). The grain was collected after harvest to test the Cd concentration of each genotype. In total, we used 3,291,150 SNPs with a minor allele frequency of >0.05 to carry out GWAS. Association analysis using a mix model was performed with the GAPIT software package. The top three principal components were used as fixed effects, and the matrix of genetic similarity based on simple SNP matching coefficients was used to model the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 February 2018 doi:10.20944/preprints201802.0075.v1

variance-covariance matrix of the random effect. The analyses were performed in GAPIT and the parameters for each trait was optimized automatically.

Permutation tests were used to help define the genome-wide significant P value threshold[31]. For all the examined trait, we reshuffled the original phenotype data, and then performed association analysis using GAPIT with the same parameters. There ought to be no real associations between the SNPs and the 'simulated' phenotypes, so all the SNPs passing the permutation threshold should be positives. In total, after 1000 permutation analyses were conducted, we clearly found that the permutation threshold was varied across genomic region and some of the 'association signals' even passing the significant cutoff 10^{-5}. Thus, we adopted a threshold P =10^{-5} at genome wide level.

For QTL and gene annotation, we firstly selected the SNP signal that passing the threshold at P =$10^{-5,}$ if the distance between adjacent SNP was less than 170 kb[32]. We then merged it into the same QTL, all the genes that located in the QTL region were predicted by the Rice Genome Annotation Project (MSU-RGAP, *Nipponbare version 6.1*) and annotated by gene ontology (GO) Slim annotation with the key words 'membrane' and 'transport'.

Primers

The sequences for all primers used in this study are listed in **Supplementary Table 13**.

Yeast cadmium transport activities assay

RNA was extracted from 9311 and *Nipponbare* accessions. For candidate gene selection assay, the coding sequence of *C1*, *C2*, *C3* and *C4* was amplified from 9311 cDNA. The haplotypes OsCd1^{V449} and OsCd1^{D449} were amplified from 9311 and *Nipponbare* cDNA, respectively. The sequences were then cloned into *pAG413GAL* vector to construct yeast expression vector. The constructed vectors along with the empty vector *pAG413GAL* were transformed into *Saccharomyces cerevisiae* W303. The methods related to yeast cultures,

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 February 2018 doi:10.20944/preprints201802.0075.v1

transformations and growth assays mainly referred to Adams et al., 1997[33]. Yeast cells were grown at 30°C in synthetic defined (SD) medium (0.67% yeast nitrogen base, Sigma) without amino acids, containing 2% (w/v) glucose or 2% (w/v) galactose (induction medium), supplemented with yeast synthetic dropout without histidine (Clontech, CA, USA), pH 5.8.

For the cadmium toxicity assay, yeast was diluted by sterile water to an OD_{600} of 1.0, 0.1, 0.01 and 0.001. The drop assays were performed on SD-His plates (with 2% (w/v) galactose/glucose) containing 0 and 30 μM $CdCl_2$ for the yeast transformants and grown at 30°C for 2 d for phenotype observation. For the cadmium accumulation test, the yeast transformants were grown in liquid SD medium (with 2% (w/v) galactose) to an OD_{600} of 1.0 containing 20, 40, 60, 80 μM $CdCl_2$ for 24 h. The yeast samples were collected after centrifuging and washing for three times with sterile water. All the assays were performed at least three times.

Three-dimensional structure modeling and protein structure analysis

The OsCd1 sequence was analyzed by Phyre2[34] and secondary structure displayed using TMRPres2D[35] to determine the position of the transmembrane helices. Three-dimensional model of $OsCd1^{V449}$ and $OsCd1^{D449}$ were constructed using the Phyre2 and MODELLER software[36]. Three high-resolution aquaporin structures GlpT (PDB entry 1pw4), YajR (PDB entry 3wdo), and POT (PDB entry 4iky) were used as templates simultaneously in the comparative modelling procedure. Three-dimensional models were visualized within PyMOL[37].

Phylogenetic reconstruction of OsCd1

149 protein sequence of MFS members from rice were obtained from online databases as cited. Multiple-sequence alignment was optimized by MUSCLE. The phylogeny of OsCd1 in the MFS was constructed by MEGA7[38] using the neighbor-joining method with No. of differences model, pairwise deletion for missing

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 February 2018 doi:10.20944/preprints201802.0075.v1

data and 1,000 bootstrap pseudo replicates. The software EvolView[39] were used for visualizing the phylogenetic trees.

CRISPR-oscd1 mutant lines construction

The sgRNA-Cas9 plant expression vectors were constructed as previously described[40]. The oligos used in constructing the sgRNA vectors for *OsCd1* are listed in **Supplementary Table 14**. The rice variety *Nipponbare* (*Oryza sativa* L. ssp. *japonica*) were used as hosts in agrobacterium-mediated transformation as previously described. The transgenic rice lines were grown in the paddy field in Hainan (18.48° N, 110.02° E, China), during normal rice-growing seasons. Mature seeds were collected from T_0 plants and geminated for genotyping. Total DNA was isolated from transgenic plants and PCR was performed to amplify the genomic region surrounding the CRISPR target sites using the specific primers. The PCR fragments were directly sequenced by Sanger method to identify mutations.

Hydroponic experiments

For hydroponic experiments, seeds of wild-type rice and *CRISPR-oscd1* lines were soaked in water for 3 days at 30°C in the dark and then transferred to Kimura B solution for 5 days. Then seedlings were treated with 20 μM $CdCl_2$ and grown at 30°C for 10 days. The shoot and root were collected for metal concentration determination. The Kimura B solution contained the macronutrients $(NH_4)_2SO_4$ (0.18 mM), $MgSO_4 \cdot 7H_2O$ (0.27 mM), KNO_3 (0.09 mM), $Ca(NO_3)_2 \cdot 4H_2O$ (0.18 mM), and KH_2PO_4 (0.09 mM) and the micronutrients $MnCl_2 \cdot 4H_2O$ (0.5 μM), H_3BO_3 (3 μM), $(NH_4)_6Mo_7O_{24} \cdot 4H_2O$ (1 μM), $ZnSO_4 \cdot 7H_2O$ (0.4 μM), $CuSO_4 \cdot 5H_2O$ (0.2 μM), and Fe-EDTA (20 μM). The pH of this solution was adjusted to 5.6, and the nutrient solution was renewed every 2 d.

Pot and Field experiments

Pot experiments were carried out at the greenhouse of Institute of Botany, Beijing. Three-week seedlings of

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 February 2018 doi:10.20944/preprints201802.0075.v1

Nipponbare, *CRISPR-oscd1* lines, 9311 and NILs pre-cultured hydroponically were transplanted to the pot with 0.18 mg/kg Cd in May and harvested at the end of September. All experiments were repeated with 30 biological replicates.

Field tests of 9311 and the NIL were performed during the regular rice cultivation season in 2017. All the cultivars were planted in the Cd-polluted paddy fields of the Institute of Subtropical Agriculture, CAS. The soil Cd concentration was 1.80 mg/kg, with pH of 5.4. The experiments were arranged in a randomized complete block design with three replicates with a spacing of 17 cm between plants and a distance of 20 cm between rows. Each block contained 100 plants in field. Field management followed normal agricultural practices except that intermittent irrigation was adopted to maximize the phenotypic differences.

The grain metal concentration was determined as described following. Important agronomic traits including plant height, seed number per panicle and grain yield per plant were measured on a single-plant basis. Plant height was determined as the height of the main tiller. Filled spikelet rate (filled grains / (filled grains + unfilled grains) × 100%) were calculated. All filled grains from plant were collected and dried at 50 °C in an oven for measurements of grain yield per plot.

Determination of metal accumulation using ICP-MS

Yeast and plant samples for Cd, Zn, Mn and Fe determination were dried at 80°C for 6 h and digested with HNO_3 at 200°C for 8 h. The concentrated acid was diluted with distilled water to the final concentration of 5%. Blank digests with no samples and the certified standard material samples (CRM rice; GBW100348) were also treated the same way. The solution Cd, Zn, Mn and Fe concentrations were determined using an Inductive Coupled Plasma Emission Spectrometer (ICP-OES) (iCAP6300, Thermo Electron Corp., MA, USA). Each sample was run in duplicate to ensure the repeatability of the results.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 February 2018 doi:10.20944/preprints201802.0075.v1

In situ analyses using synchrotron X-ray fluorescence (SR-*μ*XRF)

Approximately 500 μm thick longitudinal and latitudinal sections of brown rice were prepared and placed on Kapton tape for SR-*μ*XRF analysis. The micro-X-ray fluorescence (μ-XRF) microspectroscopy experiment is performed at 4W1B beamline, Beijing synchrotron Radiation Facility, which runs 2.5 GeV electron with current from 150 to 250 mA. The incident X-ray energy is monochromatized by W/B4C Double-Multilayer-Monochromator (DMM) at 15 keV and is focused down to 50 μm in diameter by the polycapillary lens. The two-dimensional mapping is acquired by step-mode: the sample is held on a precision motor-driven stage, scanning 100 μm stepwise. The Si(Li) solid state detector is used to detect X-ray fluorescence emission lines with livetime of 30 s. The data reduction and processing were performed using PyMca package[41].

qRT-PCR

Total RNA of root and shoot from *Nipponbare* varieties was extracted using TRIzol reagent (Invitrogen). Approximately 2 μg of the total RNA treated with DNase I was used to synthesize the first-strand cDNA using oligo$(dT)_{18}$ as a primer. The product of first-strand cDNA was used as the template for the qPCR. RT-PCR was performed on an ABI 7500 instrument using SYBR Green PCR Master mix according to the manufacturer's instructions (Takara, RR420). The cycling conditions included incubation for 30 s at 95°C followed by 40 cycles of amplification (95°C for 5 s and 60°C for 31 s). The *Ubiquitin* gene was used as an endogenous control to normalize detected gene expression. The primers used in qRT-PCR are listed in **Supplementary Table 13**.

ProOsCd1::GUS assays

A 2.5-kb upstream DNA fragment from the ATG start codon of OsCd1 was amplified from *Nipponbare* and cloned into pCAMBIA1391 to generate $OsCd1_{promoter}$::GUS construct. The resulting vector was transformed into *Nipponbare*. Tissues from the root and shoot of transgenic plants were sampled for histochemical

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 February 2018 doi:10.20944/preprints201802.0075.v1

detection of GUS expression. To measure GUS activity, rice samples were incubated in a solution of 1 mM 5-bromo-4-chloro-3-indolyl-d-glucuronic acid (X-gluc), 10% methanol, 0.5% Triton X-100 and 50 mM Na_3PO_4 at 37 °C in the dark for 24 h. After GUS staining, chlorophyll was removed using 75% ethanol for resin semi-thin section to take picture.

Subcellular-localization assay

The CDs of *OsCd1-Nipponbare* were fused in frame with GFP via cloning into the binary vector pMDC45. The resulting vectors were transformed into *Nipponbare* (*Oryza sativa* L. ssp. *japonica*) and used as hosts in agrobacterium-mediated transformation as previously described. For FM4-64 staining, rice root was transferred to 2.5 μM FM4-64 diluted in 1/2 MS medium for 3 min. The subcellular localization and co-localization was evaluated using a confocal laser scanning microscope (FluoView FV1000, Olympus).

Population genetics analysis

12 kb of the sequence centered on *OsCd1* in 446 *O. rufipogon* accessions and 950 *O. sativa* varieties were obtained from rice HapMap3 data set. SNPs in the 12-kb region were used for the *OsCd1* variety phylogenetic reconstruction analysis by MEGA7 using the neighbor-joining method with No. of differences model, pairwise deletion for missing data and 1,000 bootstrap pseudo replicates. Principal component analysis of the SNPs was performed using the software TASSEL. The level of population differentiation (F_{st}) [42] and Tajima's D[43]. Statistic was calculated by a custom PERL script using a 100 bp window. The F_{st} between Or-I and Or-III in *O. rufipogon* and between *japonica* and *indica* of *O. sativa* was calculated respectively.

NIL construction

Oryza sativa near-isogenic line (NIL) were selected from a BC_5F_2, which was generated by crossing between *japonica* variety *Nipponbare* × *indica* variety 9311 then backcrossing to 9311. 120 SSR markers distributed evenly throughout 12 chromosomes were used for identification and selection of the candidate lines containing

 doi:10.20944/preprints201802.0075.v1

454 the target donor segment. The size of the introgression fragment in the NIL was about 1 Mb between RM3413
455 and RM3372. The SSR primers are listed in **Supplementary Table 8**.
456
457 **REFERENCE**

458 1. Tsukahara, T. *et al.* Rice as the most influential source of cadmium intake among general Japanese
459 population. *Sci. Total Environ.* **305**, 41-51 (2003).
460 2. Phillipp, R. Cadmium in the environment and cancer registration. *J. Epidemiol. Community Health* **34**,
461 151-151 (1980).
462 3. Uraguchi, S. & Fujiwara, T. Rice breaks ground for cadmium-free cereals. *Curr. Opin. Plant Biol.* **16**,
463 328-34 (2013).
464 4. Takahashi, R. *et al.* The OsNRAMP1 iron transporter is involved in Cd accumulation in rice. *J. Exp.*
465 *Bot.* **62**, 4843-50 (2011).
466 5. Sasaki, A., Yamaji, N., Yokosho, K. & Ma, J.F. Nramp5 is a major transporter responsible for
467 manganese and cadmium uptake in rice. *Plant Cell* **24**, 2155-67 (2012).
468 6. Yamaji, N., Xia, J., Mitani-Ueno, N., Yokosho, K. & Feng Ma, J. Preferential delivery of zinc to
469 developing tissues in rice is mediated by P-type heavy metal ATPase OsHMA2. *Plant Physiol.* **162**,
470 927-39 (2013).
471 7. Miyadate, H. *et al.* OsHMA3, a P1B-type of ATPase affects root-to-shoot cadmium translocation in
472 rice by mediating efflux into vacuoles. *New Phytol.* **189**, 190-9 (2011).
473 8. Uraguchi, S. *et al.* Low-affinity cation transporter (OsLCT1) regulates cadmium transport into rice
474 grains. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 20959-64 (2011).
475 9. Morishita, T., Fumoto, N., Yoshizawa, T. & Kagawa, K. Varietal differences in cadmium levels of rice
476 grains of japonica, indica, javanica, and hybrid varieties produced in the same plot of a field. *Soil Sci.*

Plant Nutr. **33**, 629-637 (1987).

10. Arao, T. & Ae, N. Genotypic variations in cadmium levels of rice grain. *Soil Sci. Plant Nutr.* **49**, 473-479 (2003).

11. Liu, J. *et al.* Interaction of Cd and five mineral nutrients for uptake and accumulation in different rice cultivars and genotypes. *Field Crop. Res.* **83**, 271-281 (2003).

12. Uraguchi, S. *et al.* Root-to-shoot Cd translocation via the xylem is the major process determining shoot and grain cadmium accumulation in rice. *J. Exp. Bot.* **60**, 2677-2688 (2009).

13. Ishikawa, S., Ae, N. & Yano, M. Chromosomal regions with quantitative trait loci controlling cadmium concentration in brown rice (Oryza sativa). *New Phytol.* **168**, 345-350 (2005).

14. Ueno, D. *et al.* A major quantitative trait locus controlling cadmium translocation in rice (Oryza sativa). *New Phytol.* **182**, 644-653 (2009).

15. Kato, H. *et al.* Structural diversity and evolution of the Rf-1 locus in the genus Oryza. *Heredity* **99**, 516-524 (2007).

16. Zhang, H. *et al.* A core collection and mini core collection of Oryza sativa L. in China. *Theor. Appl. Genet.* **122**, 49-61 (2011).

17. Pao, S.S., Paulsen, I.T. & Saier, M.H., Jr. Major facilitator superfamily. *Microbiol. Mol. Biol. Rev.* **62**, 1-34 (1998).

18. Sun, L. *et al.* Genetic Diversity, Rather than Cultivar Type, Determines Relative Grain Cd Accumulation in Hybrid Rice. *Front. Plant Sci.* **7**(2016).

19. Huang, X. *et al.* A map of rice genome variation reveals the origin of cultivated rice. *Nature* **490**, 497-501 (2012).

20. Haydon, M.J. & Cobbett, C.S. A novel major facilitator superfamily protein at the tonoplast influences zinc tolerance and accumulation in Arabidopsis. *Plant Physiol.* **143**, 1705-1719 (2007).

21. Remy, E. *et al.* The Major Facilitator Superfamily Transporter ZIFL2 Modulates Cesium and Potassium Homeostasis in Arabidopsis. *Plant Cell Physiol.* **56**, 148-162 (2015).

22. Structure and mechanism of the lactose permease of Escherichia coli. *Eur. Biophys. J. Biophys Lett.* **34**, 570-570 (2005).

23. Dang, S. *et al.* Structure of a fucose transporter in an outward-open conformation. *Nature* **467**, 734-U130 (2010).

24. Jiang, D. *et al.* Structure of the YajR transporter suggests a transport mechanism based on the conserved motif A. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 14664-14669 (2013).

25. Lu, J.M.Y. & Bush, D.R. His-65 in the proton-sucrose symporter is an essential amino acid whose modification with site-directed mutagenesis increases transport activity. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 9025-9030 (1998).

26. Hibara, K.-I. *et al.* Abnormal shoot in youth, a homolog of molybdate transporter gene, regulates early shoot development in rice. *Am. J. Plant Sci.*, 1-9 (2013).

27. He, J., Zhu, C., Ren, Y., Yan, Y. & Jiang, D. Genotypic variation in grain cadmium concentration of lowland rice. *J. Plant Nutr. Soil Sci.* **169**, 711-716 (2006).

28. Liu, X., Tian, G., Jiang, D., Zhang, C. & Kong, L. Cadmium (Cd) distribution and contamination in Chinese paddy soils on national scale. *Environ. Sci. Pollut. Res.* **23**, 17941-52 (2016).

29. Duan, Q., Lee, J., Liu, Y., Chen, H. & Hu, H. Distribution of Heavy Metal Pollution in Surface Soil Samples in China: A Graphical Review. *Bull. Environ. Contam. Toxicol.* **97**, 303-9 (2016).

30. Zhang, H. *et al.* A core collection and mini core collection of Oryza sativa L. in China. *Theor. Appl. Genet.* **122**, 49-61 (2011).

31. Churchill, G.A. & Doerge, R.W. Empirical threshold values for quantitative trait mapping. *Genetics* **138**, 963-971 (1994).

32. Huang, X. *et al.* Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat. Genet.* **42**, 961-U76 (2010).

33. Roberts, C. Methods in yeast genetics. *Methods Med. Res.* **3**, 37-50 (1950).

34. Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N. & Sternberg, M.J.E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845-858 (2015).

35. Spyropoulos, I.C., Liakopoulos, T.D., Bagos, P.G. & Hamodrakas, S.J. TMRPres2D: high quality visual representation of transmembrane protein models. *Bioinformatics* **20**, 3258-3260 (2004).

36. Fiser, A. & Sali, A. MODELLER: Generation and refinement of homology-based protein structure models. *Macromolecular Crystallography,* **374**, 461-491 (2003).

37. DeLano, W.L. & Lam, J.W. PyMOL: A communications tool for computational models. *Abstr. Pap. Am. Chem. Soc.* **230**, U1371-U1372 (2005).

38. Tamura, K. *et al.* MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. and Evol.* **28**, 2731-2739 (2011).

39. Zhang, H., Gao, S., Lercher, M.J., Hu, S. & Chen, W.-H. EvolView, an online tool for visualizing, annotating and managing phylogenetic trees. *Nucleic Acids Res.* **40**, W569-W572 (2012).

40. Shan, Q., Wang, Y., Li, J. & Gao, C. Genome editing in rice and wheat using the CRISPR/Cas system. *Nat. Protoc.* **9**, 2395-2410 (2014).

41. Sole, V.A., Papillon, E., Cotte, M., Walter, P. & Susini, J. A multiplatform code for the analysis of energy-dispersive X-ray fluorescence spectra. *Spectroc. Acta Pt. B-Atom. Spectr.* **62**, 63-68 (2007).

42. Wright, S. The interpretation of population-structure by f-statistics with special regard to systems of mating. *Evolution* **19**, 395-420 (1965).

43. Tajima, F. Statistical-method for testing the neutral mutation hypothesis by dna polymorphism. *Genetics* **123**, 585-595 (1989).