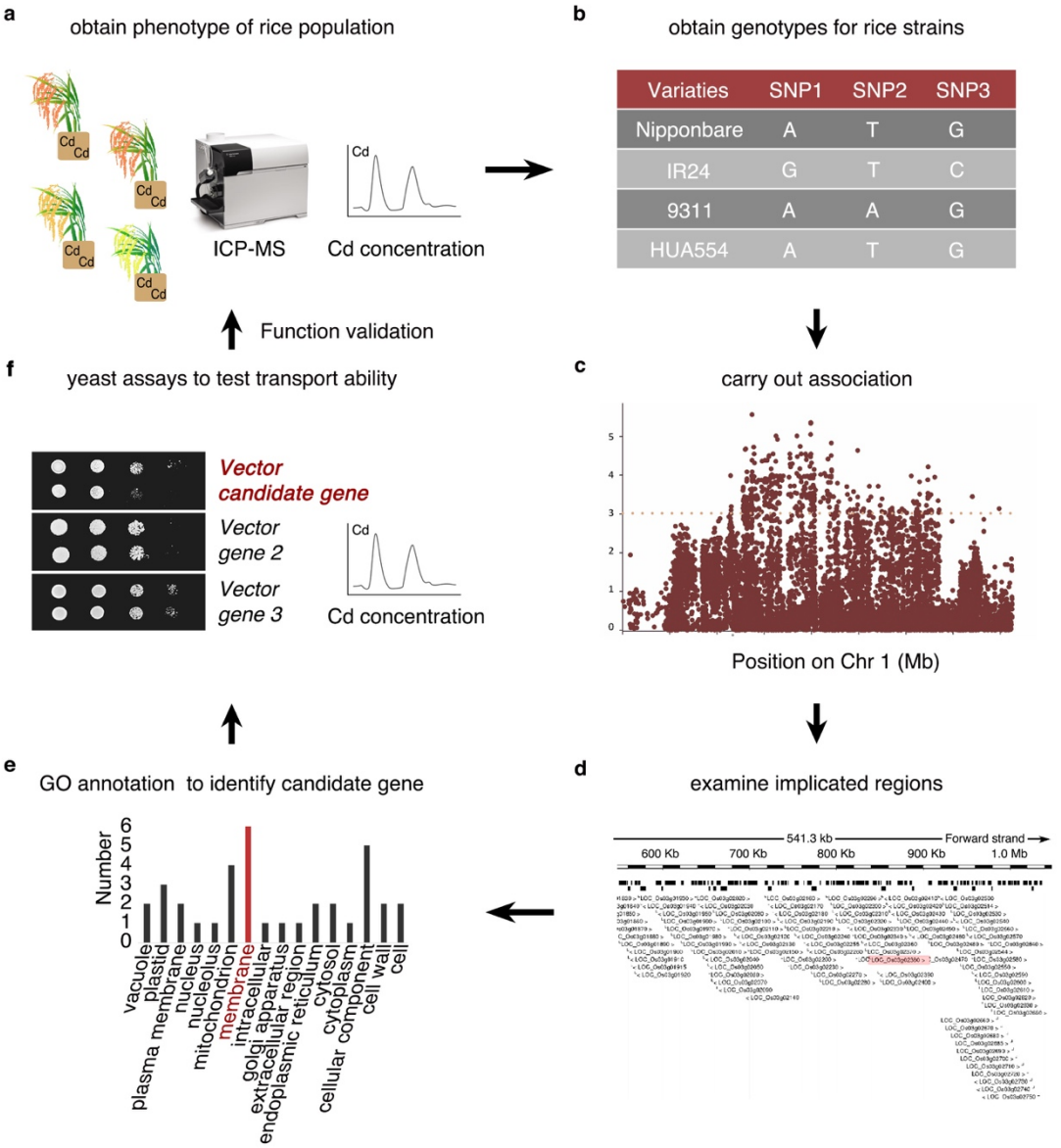
Table of contents

Supplementary Figures

Supplementary Tables

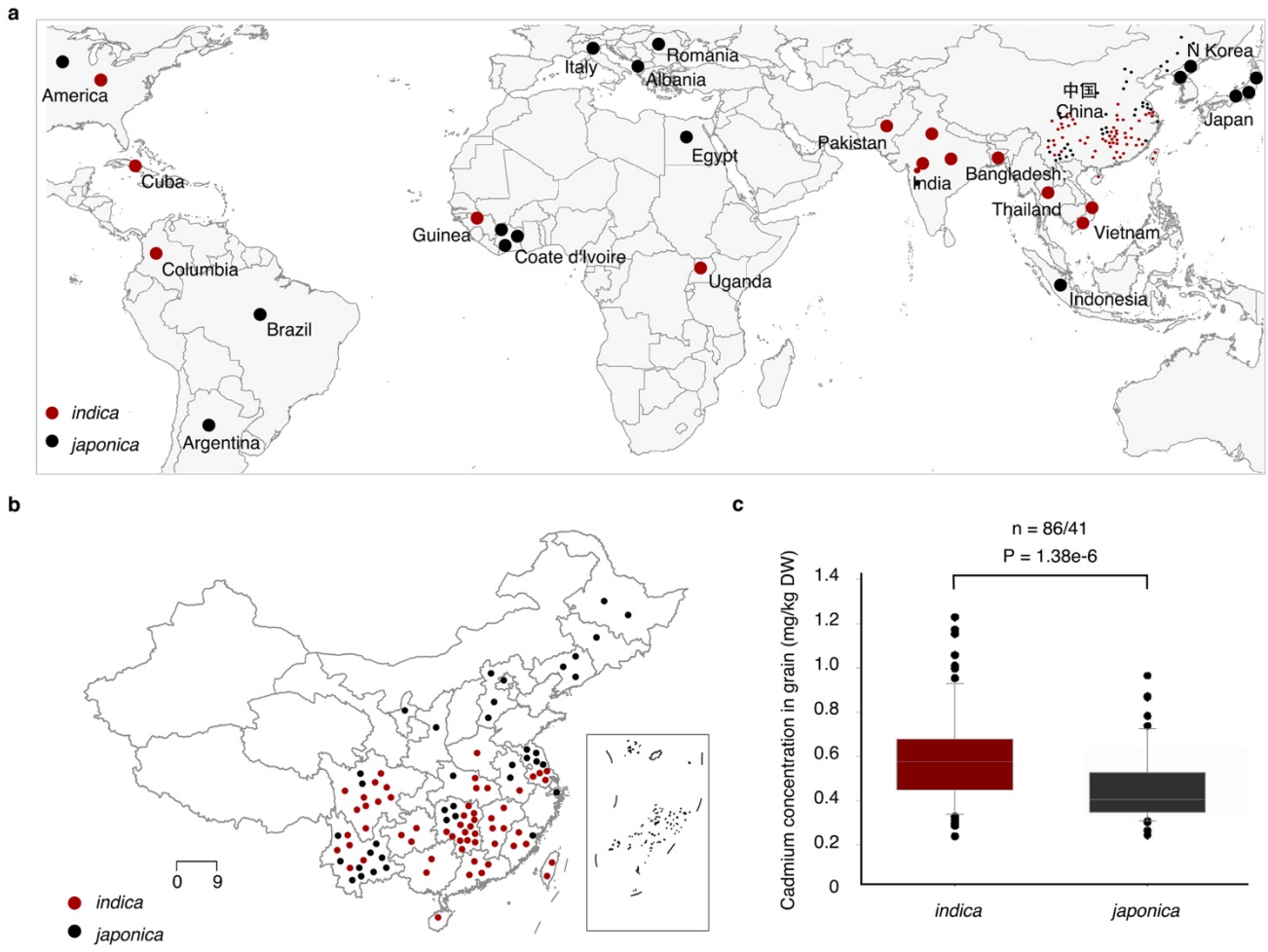
- Supplementary Table 1. The grain cadmium accumulation in *japonica* and *indica*
- Supplementary Table 2. The QTLs identified from Genome-wide association study of rice grain cadmium accumulation
- Supplementary Table 3. The annotated genes in the 12 QTLs
- Supplementary Table 4. The candidate genes and the Go Slim assignments for annotated genes in QTLs
- Supplementary Table 5. The members of Major Facilitator Superfamily (MFS) in rice
- Supplementary Table 6. *OsCd1* sequence variation in 127 rice cultivars
- Supplementary Table 7. The haplotype of *OsCd1* separated by SNP22
- Supplementary Table 8. The simple sequence repeat (SSR) primers in Chromosome 3 of NIL
- Supplementary Table 9. The SNP22 in 127 rice cultivars
- Supplementary Table 10. The SNP22 in 950 rice cultivars
- Supplementary Table 11. The SNP22 in 446 wild rice species
- Supplementary Table 12. The level of population differentiation (*Fst*)
- Supplementary Table 13. The sequences for primers used in this study
- Supplementary Table 14. SgRNA in *CRISPR-oscd1* vector construction



Supplementary Figure 1

Overview of a composite method 'bioinformatics tool-box'

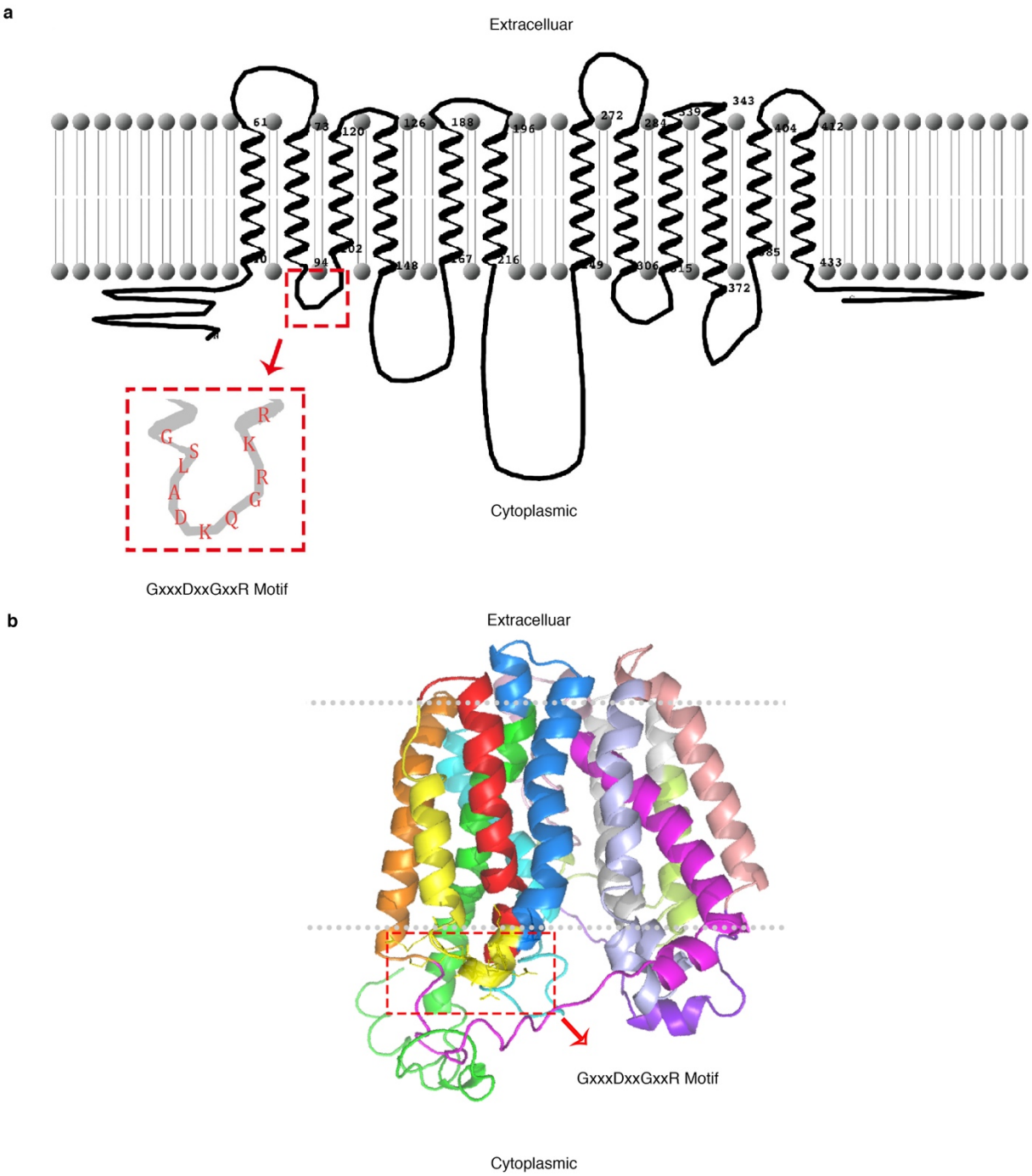
The genes underlying rice grain Cd accumulation was identified following a common general approach. **(a)** rice cultivars in the study are phenotyped for the grain Cd accumulation traits. **(b)** Genotypes for each rice cultivars are then obtained by direct genotyping. **(c)** GWAS is then carried out, typically using a statistical method for correcting for population structure. **(d)** Implicated regions are then examined via rice genome annotation project website (MSU-RGAP). **(e)** The examined genes were analyzed by GO annotation to select candidate genes **(f)** The candidate genes were then conducted with yeast assays to test the transport ability, the function gene which are then validated furtherly in rice.



Supplementary Figure 2

***Indica* varieties show higher grain Cd accumulation level than *japonica* varieties**

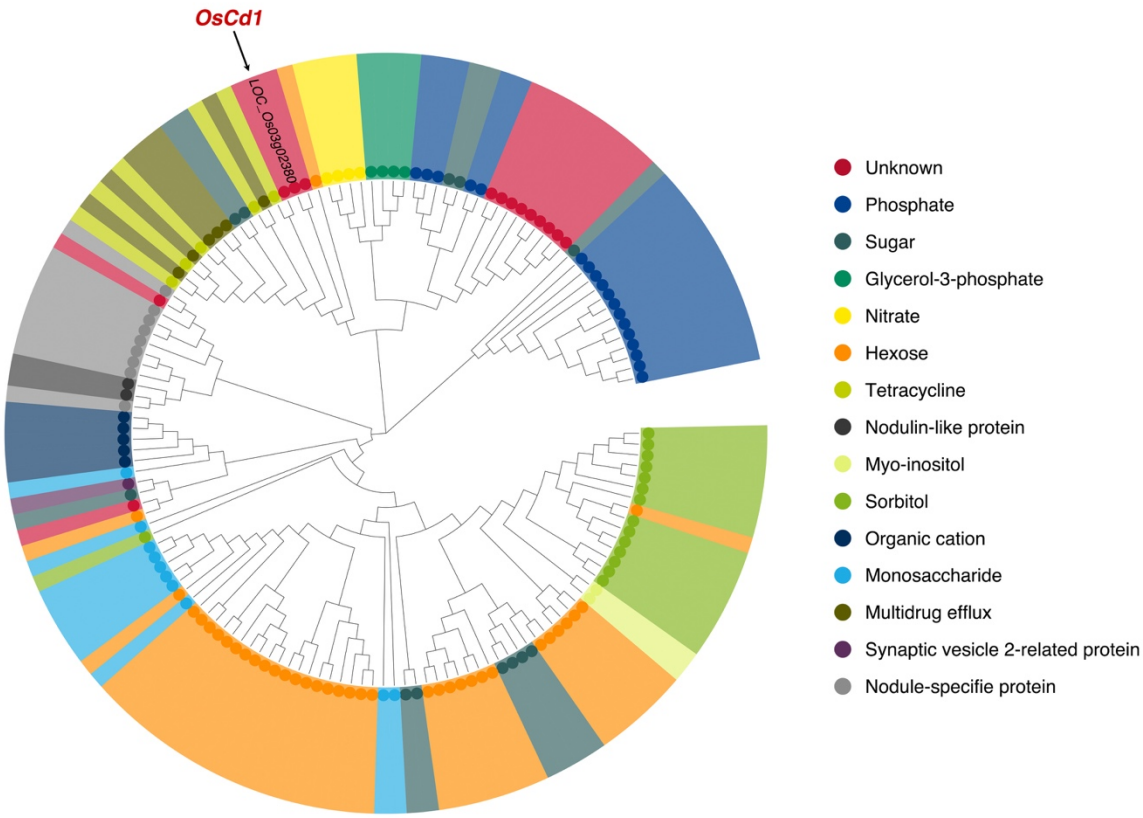
(a) and (b) The geographical distribution of the rice cultivars in the world **(a)** and China **(b)**. The red dots indicate the *indica* cultivars and the black dots indicate the *japonica* cultivars. **(c)** Comparison of grain Cd concentration between 86 *indica* and 41 *japonica* varieties. P values were generated from Student's t-test between *indica* and *japonica* varieties.



Supplementary Figure 3

Sequence Analysis of OsCd1

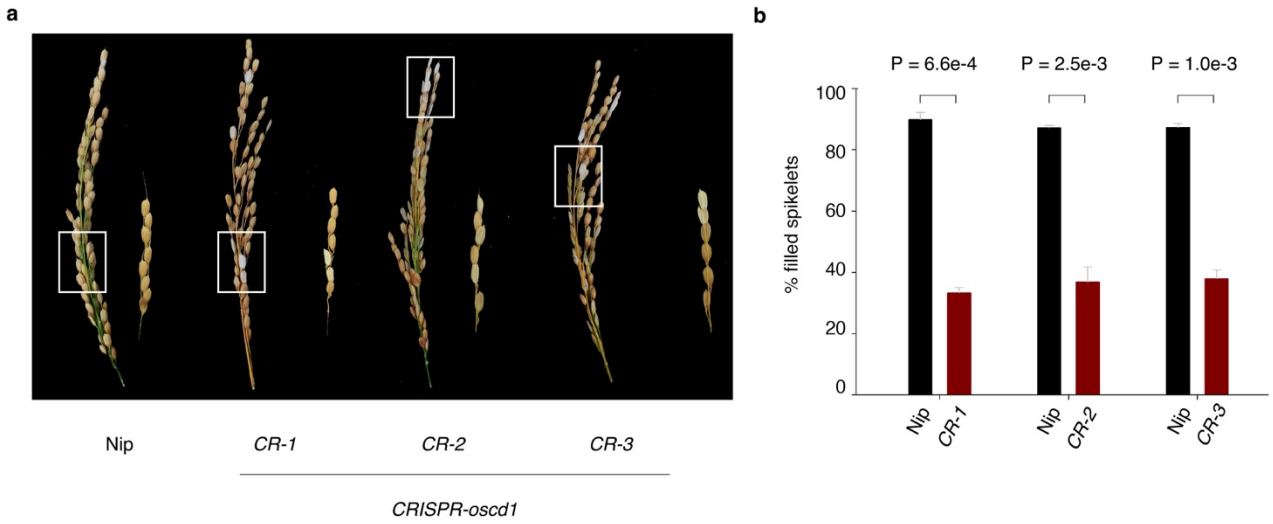
(a) The topology of OsCd1 protein was predicted using the Phyre2 software and showed by the TMRPres2D software. The conserved motif is indicated in the red box. **(b)** Predicted 3D model of OsCd1 using the Phyre2 software and showed by the PyMOL software. The conserved motif is indicated by the red box.



Supplementary Figure 4

Phylogenetic analysis of OsCd1 with rice MFS members

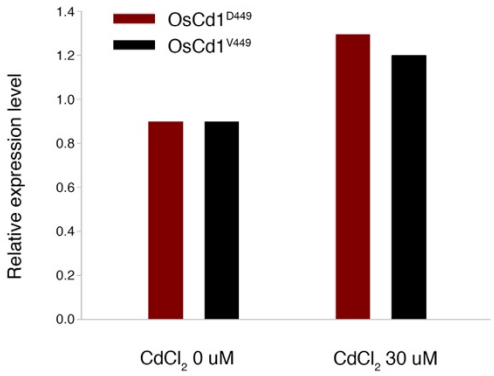
Neighbor-joining phylogenetic relationship of OsCd1 with MFS proteins in rice. OsCd1 is indicated and the MFS proteins with different predicted substrates are colored in differentially.



Supplementary Figure 5

Effect of OsCd1 mutation on rice productivity

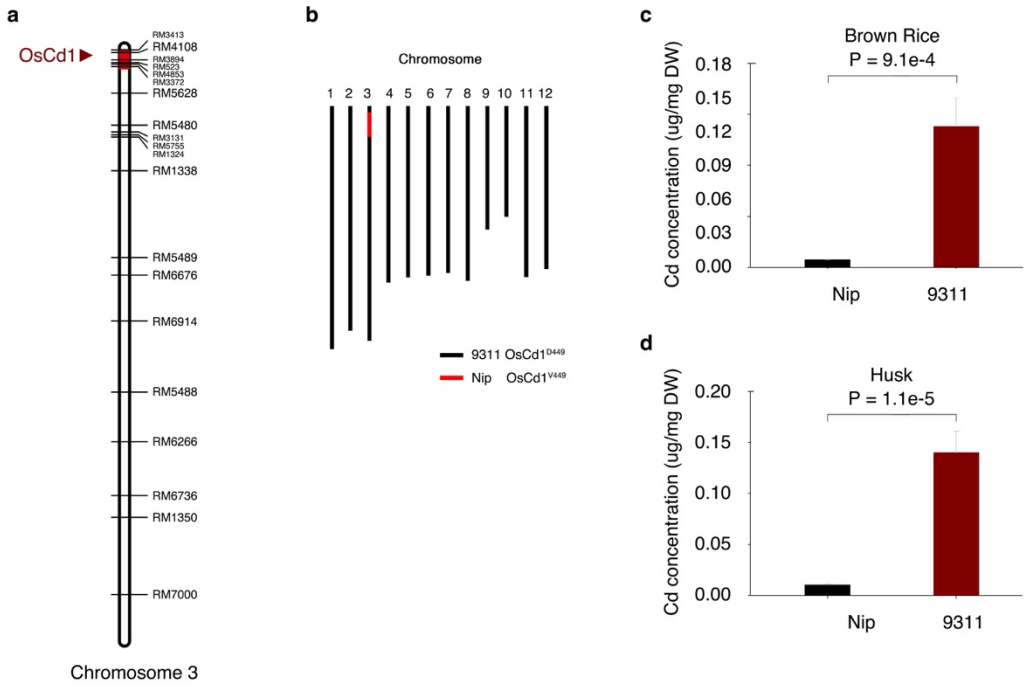
(a) Growth of the spikelets in wild-type rice and three *CRIPSR-oscd1* lines at harvest. **(b)** Fertility of the seeds. The *CRIPSR-oscd1* lines were shown in red and the *Nipponbare* was shown in black. Error bars, mean ±SE. Statistical comparison was performed by Wilcoxon matched-pairs signed-rank. All data were compared with *Nipponbare*.



Supplementary Figure 6

Transcript expression analysis of *OsCd1* in transgenic yeast

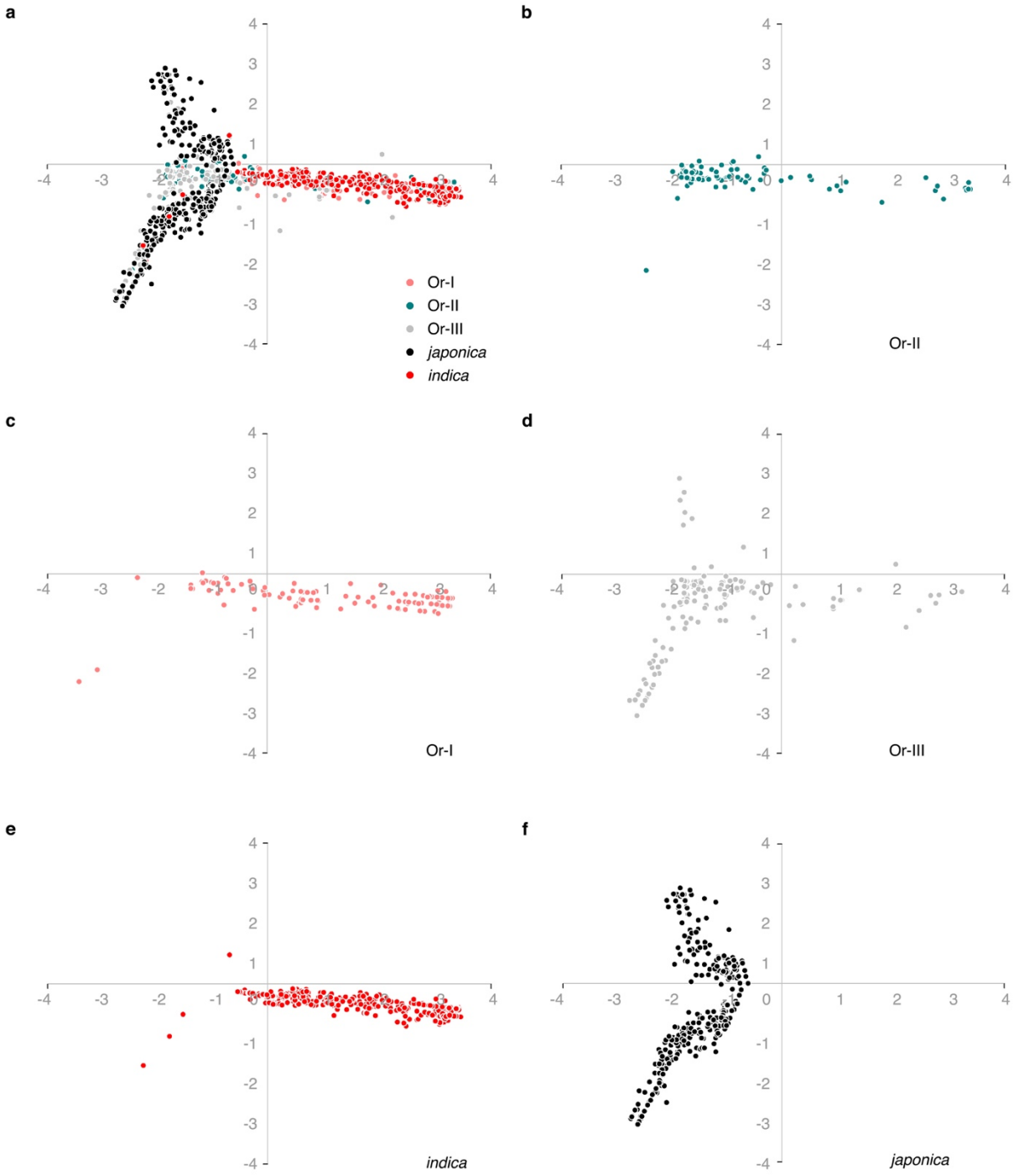
Transcript expression assay of $OsCd1^{D449}$ and $OsCd1^{V449}$ in transgenic yeast. The transcript level was determined by qPCR.



Supplementary Figure 7

The Schematic of NIL genotype and grain Cd accumulation in the parent 9311 and *Nipponbare*

(a) Graphical genotyping of NIL involving the OsCd1 on chromosome 3. Red bars represent the substituted segments in NIL. The numbers on the right of thick black lines were the simple sequence repeat (SSR) molecular makers. **(b)** Schematic of NIL genotype. Black bar, genomic region from 9311; red bar, genomic region from *Nipponbare*. **(c and d)** Concentration of Cd in the brown rice (**c**) and husk (**d**). The 9311 is shown in red and the *Nipponbare* was shown in black. Error bars, mean ±SE. Statistical comparison was performed by Wilcoxon matched-pairs signed-rank. All data were compared with 9311.



Supplementary Figure 8

PCA plots of the first two components of a full population (446 *O. rufipogon* accessions plus 950 *O. sativa* varieties)

The three types of *O. rufipogon* (Or-I **(c)**, Or-II **(b)** and Or-III **(d)**) were color-coded as pink, light blue and light grey. The two subspecies of *O. sativa* (*indica* **(e)** and *japonica* **(f)**) were colored in black and red, respectively.

Supplementary Table 2. The QTLs identified from Genome-wide association study of rice grain cadmium accumulation

Name	Chromosome	start	stop	Contribution
QTL1	chr2	6756894	7048723	17.0%
QTL2	chr2	8915797	8951020	16.3%
QTL3	chr3	519263	995454	20.7%
QTL4	chr3	1222782	1321781	16.1%
QTL5	chr3	1681658	1907570	15.1%
QTL6	chr4	27363626	27685989	17.1%
QTL7	chr5	9283326	9577230	15.8%
QTL8	chr5	9889823	10112084	17.8%
QTL9	chr5	12006893	12200901	14.1%
QTL10	chr6	175373	304839	16.6%
QTL11	chr10	20003546	20361883	20.7%
QTL12	chr10	20363169	20860856	17.1%

Supplementary Table 8 The simple sequence repeat (SSR) primers in Chromosome 3 of NIL

Name	motif	repeat	length	position
RM3413	TC	9	182	599763
RM4108	AT	35	230	770202
RM3894	AC	6	171	1336332
RM523	TC	6	235	1530589
RM4853	AT	12	103	1586664
RM3372	CT	7	230	1652913
RM5628	AAG	9	388	3593958
RM5480	CT	18	166	5805929
RM3131	AC	21	165	6350074
RM5755	GTA	15	164	6455920
RM1324	CT	15	214	6538028
RM1338	AG	10	163	8929903
RM3297	AG	20	315	14910690
RM5489	TC	13	275	14928299
RM6676	AAT	18	328	16136255
RM6914	AAT	29	188	19299263
RM5488	GA	9	115	24198502
RM6266	TCC	4	140	27617006
RM6736	ATT	10	267	31331873
RM1350	AG	36	234	32815272
RM7000	ATT	8	351	38140768

Supplementary Table 13 The sequences for primers used in this study

name	sequence
OsCd1 S	ATGGAGGTGTTCTACTACCTCGTG
OsCd1 R	TTAAGGATTCAGTGGCTCATCTTCATCATC
pAG413-OsCd1 S	gaaaaaaccccggattctagaATGGAGGTGTTCTACTACCTCGTG
pAG413-OsCd1 R	taactaattacatgactcgagTTAAGGATTCAGTGGCTCATCTTCATCATC
pMDC45-OsCd1 S	gtaaaacgacggccagtgccGTGCAGCGTGACCCGGTCGTGCCCC
pMDC45-OsCd1 R	actcatttttctaccggtaTTGAAGCGGAGGTGCCGACGGGTGG
pCAMBIA1391Z-OsCd1 S	cccaagcttTACAGGTCATTCAACCTCACAGCCT
pCAMBIA1391Z-OsCd1 R	ccggaattcTTGGCGGCGTTACTGACACAATTCT
qPCR-OsCd1 S	TCAGCTGCATCACCAAGCACT
qPCR-OsCd1 R	TCTCTTGTTGTGCTCCGCGA
qPCR-Histon S	AGTTTGGTCGCTCTCGATTTCG
qPCR-Histon R	TCAACAAGTTGACCACGTCACG
CRISPR-OsCd1-CR1 S	CCTGTGAGGCTGTGAGTCTG
CRISPR-OsCd1-CR1 R	TGCAACCTTGAACTGGGACA
CRISPR-OsCd1-CR2 S	CTGTGTGTTTCAGGGGGAGG
CRISPR-OsCd1-CR2 R	AAAGCATCAGTGTGAGGGGG
CRISPR-OsCd1-CR3 S	CTTGGCGCATTTTCGTCCTC
CRISPR-OsCd1-CR3 R	GAAATGTGTGAGCGTCCAGT

Supplementary Table 14 SgRNA in *CRISPR-oscd1* vector construction

name	sgRNA
CRISPR-oscd1 CR-1	GCCTGGTCGCAATTGTATCC
CRISPR-oscd1 CR-2	GCTTCTCGGCGTTCGAGTCA
CRISPR-oscd1 CR-3	GATGAGGATCTTGTACTCGG