

## Research Paper

# Association between sequence variants in cadmium-related genes and the cadmium accumulation trait in thermo-sensitive genic male sterile rice

Xiaohua Hao<sup>1,2</sup>), Canming Wu<sup>1</sup>), Rong Wang<sup>1</sup>), Lianfu Tian<sup>1</sup>), Taoyu Song<sup>1</sup>), Hang Tan<sup>1</sup>), Yangcheng Peng<sup>2</sup>), Meng Zeng<sup>1</sup>), Liangbi Chen<sup>1</sup>), Manzhong Liang<sup>1</sup>) and Dongping Li<sup>\*1</sup>)

<sup>1</sup>) Hunan Province Key Laboratory of Crop Sterile Germplasm Resource Innovation and Application, College of Life Science, Hunan Normal University, Changsha 410081, P.R. China

<sup>2</sup>) College of Life and Environmental Science, Hunan University of Arts and Science, Changde 415000, China

Although cultivation of hybrid rice varieties has been increasing, there are risks that high levels of cadmium (Cd) will accumulate in grain when such rice is grown in Cd-polluted environments. To produce Cd-safe hybrid rice, one practical approach is the generation of low Cd-accumulating parental lines. In two-line hybrid breeding, thermosensitive genic male sterile (TGMS) lines function as female parents to yield hybrid seeds. Recently, Cd accumulation-related genes have been identified; however, the effect of these genes on Cd accumulation in the grains of TGMS lines has yet to be reported. Here, 174 TGMS lines were selected for Cd accumulation phenotyping, and 30 TGMS lines, including 15 stable low-Cd and 15 high-Cd lines, were selected for single-nucleotide polymorphism (SNP) genotyping and association analysis. Association studies were conducted to identify the relationship between Cd accumulation and variable sites within seven candidate Cd-associated genes using logistic models. Nine sequence variant sites in four of the candidate genes were found to be significantly associated with Cd accumulation, two of which in *OsNRAMP1* and *OsNRAMP5* are low-Cd favorable variants, explaining 46.4% and 22.6% of the phenotypic variation, respectively. These loci could be developed as new molecular markers for identification of Cd accumulation characteristics and low-Cd marker-assisted breeding.

**Key Words:** association analysis, hybrid rice, Cd accumulation, molecular marker, variant loci.

## Introduction

In recent years, cadmium (Cd) contaminated rice has become an issue of major concern in a number of rice-growing countries. For people living in China and East Asia, where rice is as staple food, this crop plant is the main source of ingested Cd. Cd over-accumulation in the human body can damage health, or even lead to illness, such as the well-known Itai-itai disease (Horiguchi *et al.* 1994, Takahashi *et al.* 2012a), and increase the risk of carcinogenesis in multiple human organs, including the pancreas (Buha *et al.* 2017), breast (Van Maele-Fabry *et al.* 2016), kidney (Huff *et al.* 2007, Il'yasova and Schwartz 2005, Song *et al.* 2015), bladder (Feki-Tounsi and Hamza-Chaffai 2014), prostate gland (Rapisarda *et al.* 2018, Vinceti *et al.* 2007), and testicle (Goyer *et al.* 2004). In attempts to lower Cd accumulation in rice, a number of remedial measures have been

taken, including physical, chemical, and microbiological methods, whereby Cd ions are fixed in soils or removed from fields (Brown *et al.* 1995, Chaney *et al.* 2004, Murakami *et al.* 2008, 2009). Nevertheless, it is generally very difficult to remediate Cd-polluted paddy soils to safe levels within a short period of time. Consequently, a more practical approach is to breed inbred rice varieties or hybrid combinations with a low-Cd accumulation trait.

Over the past few decades, high-yielding hybrid rice varieties have been planted extensively to meet the needs of increasing populations. To ensure that hybrid rice produces Cd-safe grains, it is crucial to breed low-Cd trait parent lines. In two-line hybrid rice breeding, thermosensitive genic male sterile (TGMS) line rice is used as the female parent to bear hybrid seeds with heterosis. TGMS lines are a new type of male sterile line, in which the sterility varies with temperature. At the booting stage, when temperatures are lower than 23°C, TGMS lines bear normal pollen grains, which give rise to TGMS seeds. However, when temperatures exceed 23°C, these lines bear abortive pollen grains, and can be used as the maternal parents to produce hybrid combinations. The discovery of TGMS has opened up new

Communicated by Qian Qian

Received December 6, 2018. Accepted May 9, 2019.

First Published Online in J-STAGE on July 10, 2019.

\*Corresponding author (e-mail: dli@hunnu.edu.cn)

avenues for the utilization of heterosis in rice and provided a germplasm basis for the transformation of hybrid rice from a “three-line method” to a “two-line method”. However, although the relative frequency of ideal hybrid rice varieties obtained is considerably higher when using the two-line method, information on natural variation in the major Cd-accumulation genes, including the reported link with the Cd-accumulation trait in the TGMS lines of brown rice, is relatively limited.

In recent years, several genes that affect grain Cd accumulation in rice have been reported, including *OsLCD* (Shimo *et al.* 2011), *OsHMA2* (Takahashi *et al.* 2012b), *OsHMA3* (Miyadate *et al.* 2011), *OsNRAMP1* (Takahashi *et al.* 2011), *OsNRAMP5* (Ishikawa *et al.* 2012, Tang *et al.* 2017), *OsMTP1* (Yuan *et al.* 2012), *OsLCT1* (Uraguchi *et al.* 2011), *OsABCG43* (Oda *et al.* 2011), and *OsCCX2* (Hao *et al.* 2018). *OsNRAMP5* is a root-expressed Mn transporter, and is a key factor involved in Cd uptake by roots (Sasaki *et al.* 2012). Knockout of this gene inhibits Cd uptake, and is accordingly a pivotal target for low-Cd breeding (Sasaki *et al.* 2012, Takahashi *et al.* 2014, Yang *et al.* 2014). Although *OsNRAMP1* functions primarily as an iron transporter, it also plays a role in Cd uptake (Takahashi *et al.* 2011). *OsHMA3* and *OsHMA2* are members of the ATPase transporter family in rice. *OsHMA3* can sequester Cd in the vacuoles of root cells (Miyadate *et al.* 2011, Ueno *et al.* 2010, Wu *et al.* 2015). Enhancing *OsHMA3* expression decreases grain Cd content (Ueno *et al.* 2010), whereas dysfunction of *OsHMA3* has been found to lead to an increase in the Cd content of aboveground parts (Miyadate *et al.* 2011, Ueno *et al.* 2010, 2011, Uraguchi *et al.* 2017, Yan *et al.* 2016). Almost all of the *japonica* rice cultivars contain the *OsHMA3* gene and have low grain Cd contents, whereas most of the *indica* rice cultivars contain a mutated *OsHMA3* gene and have higher grain Cd contents. *OsHMA2* is expressed in the vascular system of both roots and nodes, and Cd contents in the aerial parts of rice plants are reduced when *OsHMA2* is mutated (Satoh-Nagasawa *et al.* 2012, Takahashi *et al.* 2012a, Yamaji *et al.* 2013). However, although these Cd accumulation-related genes have been demonstrated to have certain effects on the Cd content in some materials, it remains to be clarified whether they have similar effects on Cd accumulation in the grains of TGMS lines.

Recombinant inbred line-based (RIL) linkage mapping is an efficient method for quantitative trait locus analysis, which has the merit of simple genetic background (Huang and Han 2014). However, this method can only be used to explore sites with the strongest influence on the trait in question, and cannot be used to analyze phenotypes with low frequency in natural populations. In addition, when using this method, there are stringent requirements for the mapping population. Firstly, the population number must be sufficiently large, and, secondly, the genetic background of individual samples must be stable and homozygous, which can be achieved via selfing of the F<sub>2</sub> population for several

generations. Association analysis is an analytical method based on linkage disequilibrium (LD) that can be used to identify the relationship between a group of target traits and genetic markers (Gupta *et al.* 2005, Zondervan and Cardon 2004). It encompasses two analytical approaches, one of which is based on candidate gene sequences, and the other of which is based on whole-genome sequences (also referred to as genome-wide association study, GWAS) (Flint-Garcia *et al.* 2003, 2005, Zhu *et al.* 2008), which are efficient methods for exploring effective variant loci (Liu *et al.* 2015).

Over the past few decades, scientists have mapped many pathogenic loci by applying the association analysis technique. Inspired by genetic research on human diseases (Cardon and Palmer 2003), association analysis has also been adopted by plant breeders for research on crop genetics and breeding (Flint-Garcia *et al.* 2005). Using natural populations, researchers can analyze the association between candidate gene polymorphisms and target traits, making it possible to identify associations between polymorphisms of functional loci and phenotypes. Thornsberry and colleagues (2001), for example, performed association analysis for the first time in plants based on candidate gene strategies, whereby they examined the association between the polymorphic loci of the *Dwarf8* gene and the flowering time phenotype in *Zea mays*. Subsequently, Andersen *et al.* (2005) verified the association of maize *Dwarf8* polymorphism with flowering time and plant height in inbred lines under four different environmental conditions. Association analysis has also been shown to be an effect approach for gene mining and gene function verification, and has opened up new opportunities for quantitative trait research.

In this study, as experimental material, we used 174 TGMS lines, which were collected from different rice ecological regions and show wide genetic diversity. On the basis of a population structure analysis, we used a logistic model to analyze the relationship between the single-nucleotide polymorphism (SNP) variations of seven candidate genes (*OsLCD*, *OsNRAMP1*, *OsNRAMP5*, *OsHMA2*, *OsHMA3*, *OsMPT1*, *OsABCG43*) and the Cd accumulation trait associated with each gene. The findings of this analysis will contribute to gaining a better understanding of the association between sequence variants in Cd-related genes and Cd accumulation in TGMS lines of brown rice. Moreover, this approach could enable us to identify novel molecular marker(s) associated with the low-Cd trait in TGMS lines, which could in turn be applied in novel low-Cd rice resource breeding, including low-Cd TGMS lines.

## Materials and Methods

### Plant materials and growth conditions

The 174 TGMS lines analyzed in this study were obtained from the Hunan Province Key Laboratory of Crop Sterile Germplasm Resource Innovation and Application. Analyses were conducted over two successive years (2016

and 2017). Rice seeds were soaked in water for 2 days, and then germinated for 2 days in an incubator at 30°C. Fifteen-day-old seedlings were transplanted to paddy soil (containing 3.9 mg Cd per kg soil) at the end of June in each of the study years. Two seedlings were planted in individual buckets (30 cm × 30 cm) containing the paddy soil, and three replicates were assessed for each line. At the booting stage, the plants were exposed to low temperature (19–23°C) for 15 days in an intelligent cryogenic pool. As mentioned in the introduction, low temperature treatment below 23°C can promote the recovery of pollen grain fertility in TGMS varieties, thereby facilitating viable seed production via self-crossing. In the present study, we examined pollen fertility to ensure that it was greater than 80%. After the low temperature treatment, all materials were placed in the same pool for uniform moisture management. During the latter stage of grain filling, water content was controlled to maintain an appropriate level of soil moisture, and to ensure that no water accumulated in the barrel until the seeds were ripe. When the seeds were mature, the grain Cd content of plants with a seed setting rate of 70% or more was determined. Because of the different growth cycles, the materials were subjected to low temperature treatment in batches, their pollen fertility was checked separately, and water control management was carried out from 20 days after anthesis to seed maturity.

### Analysis of grain Cd concentrations

For grain Cd analysis, brown rice samples were placed in an oven at 105°C for 12 h and subsequently at 70°C until completely dehydrated. Each sample was then weighed and digested with 10 mL of HNO<sub>3</sub>–HClO<sub>4</sub> (4:1 v/v) solution at 160–180°C. Cd concentrations were determined by atomic absorption spectrophotometry (AAS, GBC3000).

### Simple sequence repeat (SSR) marker genotyping

Genomic DNA was extracted from leaf tissue according to a previously described method (Murray and Thompson 1980). The 48 pairs of SSR markers recommended by the Ministry of Agriculture of China for the identification of rice cultivars (NYT1433-2014) were used to genotype 30 selected TGMS lines. PCR was performed using 50-μL reaction mixtures containing 2 μL of 50 ng μL<sup>-1</sup> template DNA, 1 μL of 10 pmol μL<sup>-1</sup> forward primer, 1 μL of 10 pmol μL<sup>-1</sup> reverse primer, 25 μL of 2× PCR Mix (TaKaRa), and 11 μL of ddH<sub>2</sub>O. PCR amplifications were conducted using the following cycle program: an initial denaturation at 95°C for 5 min, followed by 31 cycles of denaturation at 95°C for 30 s, annealing at approximately 55°C for 30 s (depending on the T<sub>m</sub> values of primer pairs), and extension at 72°C for 30 s, and a final extension step at 72°C for 10 min. PAGE electrophoresis and silver staining procedures were performed as described by Liu *et al.* (2015).

### Population structure analysis

A total of 48 SSR primer pairs were used to analyze the

population structure. When a pair of SSR primers was used to detect single loci, each polymorphic band was regarded as a unique allele. Band analysis was based on the following criteria: if two bands had different intensities, the stronger band was read; if two bands had a similar intensity, they were both retained as common alleles. If neither of these two criteria could be applied, the data was considered as being missing. On the basis of the primer amplification results, samples with clear and polymorphic bands were selected for statistical analysis. According to the size of the amplified fragments, 1 and 0 databases were established. At the same migration location, if a band was present, it was labeled 1, whereas if no band was visible, it was labeled 0. Structure 2.3 software was used to analyze the population structure (Pritchard *et al.* 2000). Ten replicates were run according to the following procedure: the population number was set from 1 to 9, burn-in was set as 100,000, run length was set as 100,000, and a model allowing for admixture and correlated allele frequencies was adopted. The most probable value for K (for the number of clusters or subpopulations among individual samples) was calculated according to the methods described by Evanno *et al.* (2005). The analysis results were subsequently used to correct the deviation of association analysis.

### Full-length sequencing of candidate genes

The primers used for PCR were designed based on the reference sequence of the ‘Nipponbare’ variety of *japonica* rice in the NCBI database (**Supplemental Table 1**). The sequences of the Cd-related genes were amplified from the genomic DNA of the TGMS rice by PCR. High fidelity Taq DNA polymerase (AccuPrime™, Thermal Fisher) was used for the PCR amplifications. The PCR procedures were set as follows: an initial hold at 95°C for 4 min; 32 cycles of 95°C denaturation for 30 s, 55°C–60°C annealing 30 s depending on T<sub>m</sub> values of the primers, and 68°C elongation for 1 min, and a final extension of 10 min at 68°C. The amplified PCR products were sent for direct sequencing. In order to avoid base mismatch and amplification failure caused by the long gene sequence, we divided the candidate genes into several segments, and designed the primers according to the principle of terminal overlap (**Supplemental Fig. 1**). The downstream section of primer A1 of the first segment was located in the upstream section of primer S2 of the second segment, with the overlapping region between the two fragments being approximately 100 bp. Subsequently, the overlapping regions were removed and the sequence fragments were joined to obtain the complete full-length gene sequence. The full length of the Cd-related genes are defined in **Table 2**.

### Sequence polymorphism analysis

In order to ensure the accuracy and reliability of the sequences obtained, DNAMAN 7.0 software was used to compare the measured sequences with the reference sequence and to remove redundant sequences at the ends

of primers. Clustal X software was used to compare all the target sequences obtained (Thompson *et al.* 1997). The compared sequence was saved as a PHYLIP format file, which was used for subsequent association analysis.

TASSEL 5.0 software was used to analyze sequence diversity. When the LD structure was analyzed, the measured values  $D'$  (Standardized Disequilibrium Coefficients) and  $R^2$  (Squared Allele-frequency Correlations) were selected, and the Permutations value was set to 1000 (Balding 2006).

### Association analysis

TASSEL 5.0 software was used to analyze the association between low-Cd candidate genes and the low-Cd trait (Bradbury *et al.* 2007). The results of gene sequence alignment, population structure, and low-Cd data in TGMS lines were analyzed for association using a logistic regression model. For threshold selection, a  $P$ -value of  $<0.05$  was considered significant and  $P < 0.01$  considered extremely significant.

### Phenotypic effect of alleles

On the basis of the association analysis sites, we used the null allele method proposed by Brescghello and Sorrells (2006) to determine the phenotypic effects of each allele. The method used to calculate the phenotypic effect of allele variation is as follow:

$$a_i = \sum X_{ij}/N_i - N_k/n_k,$$

where  $a_i$  represents the phenotypic effect value of the  $i$  allele variation.  $X_{ij}$  is the phenotypic value of the material character of the  $j$  material carrying the  $i$  allele variation, and  $N_i$  is the number of materials with the  $i$  allelic variation.  $N_k$  is the phenotypic value of the  $k$  material with invalid alleles, and  $n_k$  is the number of materials with invalid alleles. If the  $a_i$  value is positive, it is considered that the allelic variation is a synergistic allelic variation and vice versa.

## Results

### Phenotypic variation in the Cd content of grains of different TGMS lines

Plants of the TGMS lines were grown in Cd-polluted soil in two successive years (2016 and 2017). At the booting stage, they were subjected to low temperature treatment (19–23°C) to reverse pollen sterility. The content of Cd in grains was determined for plants with a seed-setting rate higher than 70%. The grains were harvested and the Cd contents were determined after they had ripened. The results obtained in 2016 and 2017 years were found to be consistent. The frequency distribution of Cd contents in the 174 lines is presented in Fig. 1. The Cd content was found to vary from 0.063 mg/kg DW (dry weight) to 1.466 mg/kg DW, including 15 extremely low ( $<0.1$  mg/kg DW), and 15 extremely high ( $>0.4$  mg/kg DW) TGMS lines (Table 1). The average value was 0.23 mg/kg DW. Approximately 47% of the 174 lines accumulated less Cd than the accepted

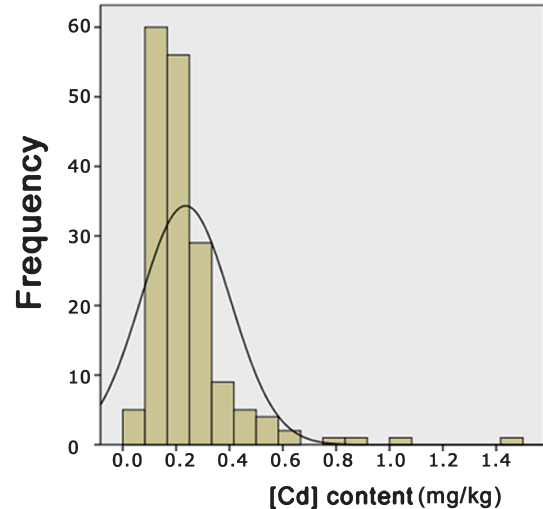


Fig. 1. Frequency distribution of brown rice Cd content.

Table 1. Grain Cd contents of 30 TGMS lines for cloning and sequencing

Varieties	[Cd] in grain (mg/kg DW)		Average
	2016	2017	
B8S	0.074 ± 0.03	0.052 ± 0.03	0.063
SD34-1S	0.068 ± 0.04	0.060 ± 0.04	0.064
Guangzhan63-4S	0.087 ± 0.04	0.051 ± 0.03	0.069
B14S	0.085 ± 0.03	0.073 ± 0.04	0.079
ZiBL02-1S	0.061 ± 0.04	0.097 ± 0.04	0.079
T91-22S	0.067 ± 0.02	0.113 ± 0.09	0.090
T95-5S	0.110 ± 0.04	0.074 ± 0.05	0.092
SD-40S	0.120 ± 0.05	0.066 ± 0.04	0.093
B15S	0.084 ± 0.04	0.104 ± 0.09	0.094
Xiangling628S	0.098 ± 0.03	0.094 ± 0.03	0.096
201507S	0.110 ± 0.04	0.086 ± 0.05	0.098
B76S	0.084 ± 0.04	0.112 ± 0.06	0.098
SD39S	0.120 ± 0.05	0.078 ± 0.05	0.099
B216S	0.140 ± 0.05	0.060 ± 0.04	0.100
SD39-1S	0.094 ± 0.04	0.106 ± 0.11	0.100
B206S	0.388 ± 0.12	0.406 ± 0.22	0.397
B152S	0.250 ± 0.09	0.598 ± 0.19	0.424
C1S10S	0.480 ± 0.13	0.412 ± 0.16	0.446
21504S	0.210 ± 0.08	0.742 ± 0.21	0.476
HSQ9S	0.610 ± 0.21	0.342 ± 0.11	0.476
B204S	0.350 ± 0.11	0.636 ± 0.22	0.493
Zhu77S	0.470 ± 0.23	0.536 ± 0.18	0.503
W6154S	0.580 ± 0.12	0.430 ± 0.26	0.505
B161S	0.410 ± 0.14	0.684 ± 0.23	0.547
640S	0.680 ± 0.16	0.452 ± 0.22	0.566
zhu25S	0.610 ± 0.21	0.584 ± 0.32	0.597
B63S	0.740 ± 0.22	0.460 ± 0.12	0.600
SD44S	0.650 ± 0.09	0.856 ± 0.23	0.753
B73S	0.920 ± 0.23	1.170 ± 0.33	1.045
C1S7S	1.200 ± 0.33	1.732 ± 0.65	1.466

standard level in grains ( $<0.2$  mg/kg DW), whereas 8% of the lines showed an extremely high grain Cd concentration ( $>0.4$  mg/kg DW), and 45% of the lines had accumulated Cd contents of between 0.2 and 0.4 mg/kg DW. Among the TGMS lines examined, the following with grain Cd contents lower than 0.1 mg/kg DW were identified as low-Cd varieties: B8S, SD34-1S, guanzhan63-4S, B14S, T91-22S,

ZiseBLO2-1S, T95-5S, SD40S, B15S, Xiangling628S, 201507S, B76S, B216S, SD39S, and SD39-1S.

### Population structure analysis of the TGMS lines

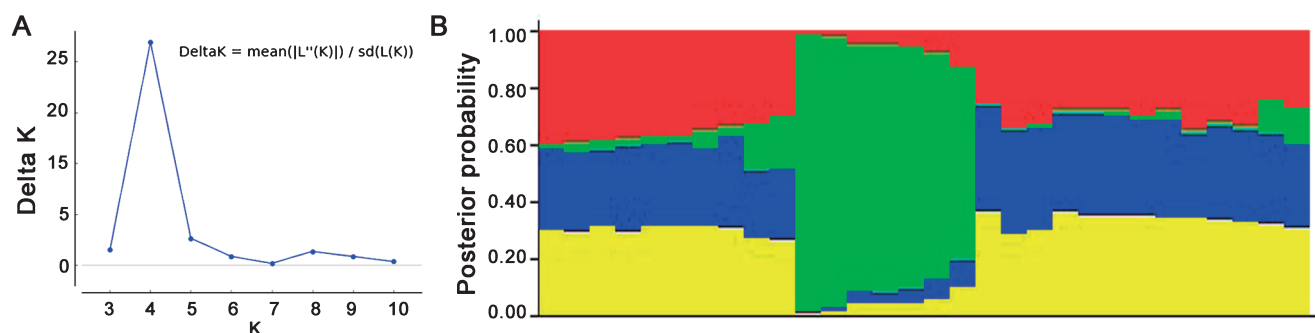
On the basis of the results obtained for Cd accumulation in grains, 30 TGMS lines (15 stable low-Cd and 15 high-Cd lines) were selected for further analysis. In order to make the association analysis more accurate, we initially performed a population structure analysis of the 30 TGMS lines prior to their use for candidate gene cloning. Genotype analysis was based on the use of 48 SSR markers that are evenly distributed throughout the rice genome, and which were revised and re-published by the Ministry of Agriculture of China in 2014. These markers were screened from 35 representative rice varieties with different genetic background characteristics. They cover 12 chromosomes of rice and have high polymorphism. The genetic structure of the 30 selected lines was analyzed using Structure 2.2 software (Pritchard *et al.* 2000). The results showed that the 48 pairs of SSR primers could be used to effectively divide the examined materials into different sub-populations, and K values with an obvious inflection point were obtained. The results showed that as the model parameter K increases, the log-likelihood also showed the same tendency. Furthermore, the  $\Delta K$  value was considerably higher for the model parameter K = 4 than for any other value of K (Fig. 2A), suggesting that the population studied comprised a mixture that can be divided into four sub-populations (Fig. 2B).

Analysis of the genetic diversity of the 30 TGMS lines

indicated that 46 SSR loci were polymorphic, which represents a polymorphic frequency of 95.8% among the 48 SSR loci examined. A total of 140 alleles were detected using the 46 SSR markers, ranging from two to five alleles per locus, with an average of 3.04 alleles at each SSR site. The value of Shannon's index for the population ranged from 0.0853 to 0.6926, the gene diversity index ranged from 0.0331 to 0.4995, and the allele frequency ranged from 0.0168 to 1.0. These data revealed that the TGMS lines have rich genetic diversity and were suitable for further sequence polymorphism analysis.

### Sequence polymorphism analysis

The full sequence information of seven low Cd-related genes were obtained from the web site <https://www.ncbi.nlm.nih.gov/>, and the full-length sequences were cloned by PCR. Polymorphism analysis revealed a total of 333 variant sites among all the sequences, comprising 268 SNPs and 65 INDELs (Insertion-Deletion sites), most of which were located in the non-coding regions of the gene sequences (Table 2). The *OsHMA2* gene was found to contain the largest number variable sites (on average, one site per 50 bp), whereas *OsMTP1* has the least number (on average, one site per 278 bp). We identified 62 variable sites in the coding regions of the candidate genes, with *OsHMA2* again being found to harbor the largest number of variable sites (up to 39), whereas *OsNRAMP5* contains just a single variable site in its coding regions. Compared with the 'Nipponbare' reference sequences, we identified 30 non-synonymous SNPs



**Fig. 2.** Population structure analysis of 30 TGMS rice accessions. (A) Changes in  $\Delta K$  (B) Bar plot of the genetic composition of 30 TGMS lines based on admixture model.

**Table 2.** Polymorphism of 7 low cadmium-related genes

Gene	Full length	No. of variations	No. of SNP		No. of INDEL		$\Pi \times 10^3$
			Coding region	Noncoding region	Coding seq. region	Noncoding seq. region	
<i>OsABCG43</i>	4890	25	7	12	–	6	2.28
<i>OsMTP1</i>	3621	13	1	11	–	1	2.49
<i>OsHMA3</i>	4057	16	9	6	–	1	2.97
<i>OsNRAMP1</i>	4846	46	4	31	–	11	3.05
<i>OsLCD1</i>	3922	25	1	21	1	2	3.62
<i>OsNRAMP5</i>	7281	54	1	42	–	11	4.03
<i>OsHMA2</i>	7690	154	37	85	2	30	6.67
Total	36307	333	60	208	3	62	

II: Nucleotide diversity.

spread across the *OsNRAMP1*, *OsABCG43*, *OsLCD1*, *OsHMA3*, and *OsHMA2* genes. The measured nucleotide diversity ( $\Pi$ ) values varied from  $2.28 \times 10^{-3}$  (*OsABCG43*) to  $6.67 \times 10^{-3}$  (*OsHMA2*).

### Association analysis between gene sequence polymorphisms and the Cd accumulation trait

Analysis of the association between Cd accumulation and polymorphisms of the seven selected gene sequences from the TGMS lines was carried out using the Logistic module in TASSEL 5.0 software (Bradbury *et al.* 2007). General linear model (GLM) data revealed that a total of 42 variants, in four of the seven genes, were significantly or extremely significantly associated with Cd accumulation (Table 3). Furthermore, we found that most of the variant sites are located in non-coding regions, with only four sites being located in coding regions (one SNP at 3751 bp in the 10th exon of *OsNRAMP5*, and the other three SNPs located at 450, 513, and 568 bp in the 1st exon of *OsHMA2*).

We found that one SNP detected in the 5'UTR region of *OsHMA3*, at base pair position 840 (C-T), was significantly associated with Cd accumulation ( $P < 0.05$ ), and could explain 15.89% of the phenotypic variation.

Three polymorphic sites in *OsNRAMP1* were significantly related to Cd accumulation ( $P < 0.01$ ), whereas three SNPs in complete LD (124, 164, and 1076 bp), two SNPs in complete LD (1736 and 5781 bp), and one INDEL located at 2573 bp in *OsNRAMP1* explained 27.9%, 23.9%, and 46.4% of the phenotypic variation, respectively.

In *OsNRAMP5*, eight SNPs and three INDELS were significantly associated with Cd accumulation ( $P < 0.01$ ). One SNP at position 1306 bp explained 22.6% of phenotypic variation, where the remaining seven SNPs (3530, 3615, 3623–3624, 3627, 3751, 4016, and 4263 bp) and the three INDELS (2267–2279, 3543–3346, 4710–4727 bp) in complete LD (Table 3) explained 23.5% of the phenotypic variation.

Twenty SNPs and four INDELS in the *OsHMA2* genes

were shown to be significantly associated with Cd accumulation in the grains of rice plants subjected to Cd pollution ( $P < 0.05$ ). Two SNPs in complete LD (81 and 109 bp), and 17 SNPs and four INDELS in complete LD explained 23.8% and 15.5% of phenotypic variation, respectively. The other SNP located at 645 bp was found to explain 13.4% of the phenotypic variation in Cd accumulation.

### Combination of favorable alleles in TGMS lines

On the basis of the allelic effect on Cd accumulation in brown rice, each of the identified variants was classified as either favorable or unfavorable. The phenotypic allele effect was estimated through comparison between the average phenotypic value of accessions with the specific allele and that of accessions with a “null allele”.

Of the nine significantly associated loci, seven showed positive association with Cd accumulation in grain and the remaining two loci were negatively correlated (Table 4). These latter two loci in *OsNRAMP1* and *OsNRAMP5* showed a close association with a low-Cd phenotype and can thus be considered favorable alleles with polymorphisms for low-Cd grain. The other seven loci are associated with the accumulation of Cd in grain and can be classed as high-Cd alleles in grains. The aggregation of favorable alleles can contribute to enhancing certain favorable traits. In this regard, among the 30 TGMS lines examined, we found that B8S and B14S have two favorable alleles that synchronously contribute to low Cd accumulation. Accordingly, these two lines could be considered as low-Cd TGMS lines and used as putative resources for low-Cd breeding in the future. In contrast, the lines C1S7S and B152S harbor seven alleles that can be considered unfavorable with respect to low Cd accumulation, indicating that these lines would not be suitable as TGMS lines for low-Cd hybrid breeding (Table 4). However, it should be emphasized that these polymorphic sites might not necessarily have a direct influence on phenotypic characters, and thus our findings may simply reflect an indirect effect of strong LD with trait-linked sequence variants.

**Table 3.** Marker-trait associations with  $P$ -values of less than 0.05 by GLM analysis, and the  $r^2$  (phenotypic variance explained, PVE)

Gene	Site	$P$ -value	$r^2$
<i>OsHMA3</i>	840	0.0186	0.159
<i>OsNRAMP1</i>	3 sites in LD <sup>a</sup>	0.0014	0.279
	1736, 5781 <sup>b</sup>	0.0037	0.239
	2573	0.0003	0.464
<i>OsNRAMP5</i>	1306	0.0205	0.226
	10 sites in LD <sup>c</sup>	0.0039	0.235
<i>OsHMA2</i>	81, 109 <sup>d</sup>	0.0162	0.238
	21 sites in LD <sup>e</sup>	0.0234	0.155
	645	0.0367	0.134

a, 3 sites in complete LD: 124, 164, 1076; b, 2 sites in complete LD: 1736, 5781; c, 10 sites in complete LD: 2267–2279, 3530, 3543–3346, 3615, 3623–3624, 3627, 3751 (10 exon), 4016, 4263, 4710–4727; d, 2 sites in complete LD: 81, 109; e, 21 sites in complete LD: 141, 143–157, 177, 180, 182, 188–189, 203, 227, 242, 245–247, 250, 256, 258–280, 301, 316, 389, 417, 425–427, 450 (1 exon), 513 (1 exon), 568 (1 exon). <sup>a</sup>, <sup>b</sup>, <sup>c</sup>, <sup>d</sup>, and <sup>e</sup>, were in complete LD.

## Discussion

In this study, we examined the genetic diversity of 30 TGMS lines using 48 pairs of SSR primers. A total of 140 alleles were amplified using 46 SSR markers, with two to five alleles (average of 3.04 alleles) being detected at each SSR site. In addition, the values of Shannon's index for the population ranged from 0.0853 to 0.6926, the gene diversity index ranged from 0.0331 to 0.4995, the average Nei's gene diversity index was 0.28, and the allelic frequency ranged from 0.0168 to 1.0. On the basis of these data, it can be seen that the genetic diversity of the 30 TGMS lines is high, genetic relationships are relatively low, and the genetic background is diverse. Hence, it is suitable for further genotyping and association analysis though the population tested is not so big.

**Table 4.** Estimates of the allelic effect of Cd-related genes for phenotypic variation of the TGMS lines

Gene Site	Allele	Allelic effect	Typical carrier variety
<i>OsHMA3</i> 840	C	0.0	
	T	0.493	C1S7S, W6154S, B152S
<i>OsNRAMP1</i> 124 <sup>a</sup>	C	0.0	
	G	0.53	C1S7S, B63S, B152S
	C	0	
	G	0.634	C1S7S, B152S
2573	–	0.0	
	INS T	–0.303	<u>B8S, B14S</u>
<i>OsNRAMP5</i> 1306	T	0	
	DEL	–0.207	<u>B8S, SD34-1S, Guangzhan63-4S, B14S, ZiBL02-1S, T95-5S, B15S, Xianglin628S, B76S, 201507S, SD39-1S, B206S, 21504S, Zhu77S, Zhu25S</u>
	C	0.39	C1S7S, W6154S, B152S
	–	0.0	
2267–2279 <sup>c</sup>	Indel 12	0.945	C1S7S, B152S
<i>OsHMA2</i> 81 <sup>d</sup>	A	0.0	
	C	0.156	<u>T91-22S, SD39-1S, B216S, SD39S, HSQ9S, C1S7S, SD44S, Zhu25S, 640S, Zhu77S, B152S, B206S</u>
	A	0.0	
	T	0.197	<u>SD39-1S, B216S, SD39S, HSQ9S, C1S7S, SD44S, Zhu25S, B161S, 640S, Zhu77S, B152S, B206S</u>
	T	0.0	
645	DEL	0.157	<u>T91-22S, SD39-1S, B216S, SD39S, HSQ9S, C1S7S, SD44S, Zhu25S, B161S, 640S, Zhu77S, B152S, B206S</u>

a, 3 sites in complete LD: 124, 164, 1076; b, 2 sites in complete LD: 1736, 5781; c, 10 sites in complete LD: 2267–2279, 3530, 3543–3346, 3615, 3623–3624, 3627, 3751 (10 exon), 4016, 4263, 4710–4727; d, 2 sites in complete LD: 81, 109; e, 21 sites in complete LD: 141, 143–157, 177, 180, 182, 188–189, 203, 227, 242, 245–247, 250, 256, 258–280, 301, 316, 389, 417, 425–427, 450 (1 exon), 513 (1 exon), 568 (1 exon). <sup>a, b, c, d, e</sup>, were in complete LD. The varieties with low Cd were underlined.

Association analysis revealed 42 variable sites within the genes *OsHMA2*, *OsHMA3*, *OsNRAMP1*, and *OsNRAMP5* that are highly associated with Cd accumulation, of which four are located in exonic regions, and the others are located in intronic and 5'UTR regions. However, no highly Cd-associated variant sites were detected in 3'UTR regions. The four polymorphic sites located in the coding regions of *OsNRAMP5* and *OsHMA2* were shown to be significantly associated with Cd accumulation, of which the two polymorphic sites in *OsHMA2* are non-synonymous SNPs at 450 bp and 568 bp that result in the conversion of the amino acids Arg and Ser to Gly and Ala, respectively. In both cases, a polar hydrophilic amino acid is replaced with a non-polar hydrophobic amino acid. Furthermore, the Ser residue might be a putative phosphorylation site. Accordingly, the presence of these non-synonymous SNPs may contribute to a conformation change in the *OsHMA2* protein, and thereby influence its function. The other two exonic polymorphic sites are synonymous SNPs that are not associated with amino acid conversion.

Noteworthy are the SNPs and INDELS detected in the 5'UTR sequences. The three INDEL sites located in the 5'UTR region of *OsHMA2* (two large INDEL regions of 15 bp and 22 bp, and a small INDEL of 4 bp) were significantly associated with Cd accumulation traits. Differences in the SNPs and INDELS in gene promoter regions might be involved in gene expression regulation. For example, the intron in the *Gmubi* gene promoter region regulates the ex-

pression of this gene in *Glycine max* (De La Torre and Finer 2015). Similarly, two SNPs in the promoter region of the *GS5* gene show close association with the expression levels of this gene in developing young panicles of rice to regulate the size of grain width (Xu *et al.* 2015).

We also identified certain intronic variable sites with extremely significant associations. There are 13 variable sites located in the *OsNRAMP1*, *OsNRAMP5*, and *OsHMA2* genes, including two large INDEL regions of 13 bp and 17 bp in *OsNRAMP5*. In the past, the importance of intron sequences in the genome went unrecognized, and for a long time, introns were considered simply as “garbage” DNA. However, more recently, introns have begun to attract increasing attention and the important roles played by these regions are gradually being recognized. Introns can have a significant influence on gene expression in plants and many other eukaryotes; for example, by providing auxiliary *cis*-acting elements for the regulation of gene expression, and playing roles in modifying gene transcription efficiency (Rose 2004, 2008, Rose *et al.* 2016). Analysis has indicated that introns in the 5'UTR regions of two E3 ligase genes are essential for the expression of these genes in *Arabidopsis* (Aguilar-Hernández and Guzmán 2013). However, whether intronic variable loci play regulatory roles in Cd-associated gene function needs further study. In this regard, it is conceivable that new markers could be developed for intronic variants to facilitate the identification of Cd-accumulating rice varieties.

By determining the positive and negative effects of nine loci that we identified to be significantly associated with Cd accumulation in grain, we found that only two (in *OsNRAMP1* and *OsNRAMP5*) of the nine loci have a negative effect on Cd accumulation, and are thereby non-conducive to Cd accumulation and can thus be considered as low-Cd markers. In contrast, the other seven polymorphic loci appear to make a positive contribution to Cd accumulation in grain and can accordingly be considered as high-Cd markers. Among the 30 TGMS lines we examined, B8S and B14S contain two sites with a negative influence on Cd accumulation and can be considered as potential low Cd-accumulating varieties. These markers can provide a basis for Cd accumulation trait identification and molecular marker-assisted breeding of rice. Using a combination of molecular marker-assisted selection and progeny identification with B8S and B14S as female parents, we believe that it will be possible to develop new varieties characterized by low Cd accumulation and other excellent agronomic traits.

### Acknowledgments

We are grateful to Ms. Lian Chen and Jiayong Xiao, Inspection and Quarantine Technology Center, Hunan Entry - Exit Inspection and Quarantine Bureau, Changsha, China, for assistance in sample Cd detection. This work was supported by grants of NSFC (31570316; 31371244), Projects of Hunan Provincial Education Department (17C0974), the Cooperative Innovation Center of Engineering and New Products for Developmental Biology of Hunan Province (20134486), and Hunan Provincial Innovation Foundation for Postgraduate (CX2017B230, CX2018B300).

### Literature Cited

Aguilar-Hernández, V. and P. Guzmán (2013) Spliceosomal introns in the 5' untranslated region of plant BTL RING-H2 ubiquitin ligases are evolutionary conserved and required for gene expression. *BMC Plant Biol.* 13: 179.

Andersen, J.R., T. Schrag, A.E. Melchinger, I. Zein and T. Lübberstedt (2005) Validation of *Dwarf8* polymorphisms associated with flowering time in elite European inbred lines of maize (*Zea mays* L.). *Theor. Appl. Genet.* 111: 206–217.

Balding, D.J. (2006) A tutorial on statistical methods for population association studies. *Nat. Rev. Genet.* 7: 781–791.

Bradbury, P.J., Z. Zhang, D.E. Kroon, T.M. Casstevens, Y. Ramdoss and E.S. Buckler (2007) TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23: 2633–2635.

Breseghello, F. and M.E. Sorrells (2006) Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars. *Genetics* 172: 1165–1177.

Brown, S.L., J.S. Angle, R.L. Chaney and A.J.M. Baker (1995) Zinc and cadmium uptake by hyperaccumulator *Thlaspi caerulescens* grown in nutrient solution. *Environ. Sci. Technol.* 29: 1581–1585.

Buha, A., D. Wallace, V. Matovic, A. Schweitzer, B. Oluic, D. Micic and V. Djordjevic (2017) Cadmium exposure as a putative risk factor for the development of pancreatic cancer: three different lines of evidence. *Biomed Res. Int.* 2017: 1981837.

Cardon, L.R. and L.J. Palmer (2003) Population stratification and spurious allelic association. *Lancet* 361: 598–604.

Chaney, R.L., P.G. Reeves, J.A. Ryan, R.W. Simmons, R.M. Welch and J.S. Angle (2004) An improved understanding of soil Cd risk to humans and low cost methods to phytoextract Cd from contaminated soils to prevent soil Cd risks. *Biometals* 17: 549–553.

De La Torre, C.M. and J.J. Finer (2015) The intron and 5' distal region of the *soybean Gmubi* promoter contribute to very high levels of gene expression in transiently and stably transformed tissues. *Plant Cell Rep.* 34: 111–120.

Evanno, G., S. Regnaut and J. Goudet (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. *Mol. Ecol.* 14: 2611–2620.

Feki-Tounsi, M. and A. Hamza-Chaffai (2014) Cadmium as a possible cause of bladder cancer: a review of accumulated evidence. *Environ. Sci. Pollut. Res.* 21: 10561–10573.

Flint-Garcia, S.A., J.M. Thornsberry and E.S. IV Buckler (2003) Structure of linkage disequilibrium in plants. *Annu. Rev. Plant Biol.* 54: 357–374.

Flint-Garcia, S.A., A.C. Thuillet, J.M. Yu, G. Pressoir, S.M. Romero, S.E. Mitchell, J. Doebley, S. Kresovich, M.M. Goodman and E.S. Buckler (2005) Maize association population: a high-resolution platform for quantitative trait locus dissection. *Plant J.* 44: 1054–1064.

Goyer, R.A., J. Liu and M.P. Waalkes (2004) Cadmium and cancer of prostate and testis. *Biometals* 17: 555–558.

Gupta, P.K., S. Rustgi and P.L. Kulwal (2005) Linkage disequilibrium and association studies in higher plants: Present status and future prospects. *Plant Mol. Biol.* 57: 461–485.

Hao, X., M. Zeng, J. Wang, Z. Zeng, J. Dai, Z. Xie, Y. Yang, L. Tian, L. Chen and D. Li (2018) A node-expressed transporter OsCCX2 is involved in grain cadmium accumulation of rice. *Front. Plant Sci.* 9: 476.

Horiguchi, H., H. Teranishi, K. Niiya, K. Aoshima, T. Katoh, N. Sakuragawa and M. Kasuya (1994) Hypoproduction of erythropoietin contributes to anemia in chronic cadmium intoxication: clinical study on Itai-itai disease in Japan. *Arch. Toxicol.* 68: 632–636.

Huang, X. and B. Han (2014) Natural variations and genome-wide association studies in crop plants. *Annu. Rev. Plant Biol.* 65: 531–551.

Huff, T.B., L. Tong, Y. Zhao, M.N. Hansen, J.-X. Cheng and A. Wei (2007) Hyperthermic effects of gold nanorods on tumor cells. *Nanomedicine (Lond)* 2: 125–132.

Il'yasova, D. and G.G. Schwartz (2005) Cadmium and renal cancer. *Toxicol. Appl. Pharmacol.* 207: 179–186.

Ishikawa, S., Y. Ishimaru, M. Igura, M. Kuramata, T. Abe, T. Senoura, Y. Hase, T. Arao, N.K. Nishizawa and H. Nakanishi (2012) Ion-beam irradiation, gene identification, and marker-assisted breeding in the development of low-cadmium rice. *Proc. Natl. Acad. Sci. USA* 109: 19166–19171.

Liu, E., X. Liu, S. Zeng, K. Zhao, C. Zhu, Y. Liu, M.C. Breria, B. Zhang and D. Hong (2015) Time-course association mapping of the grain-filling rate in rice (*Oryza sativa* L.). *PLoS ONE* 10: e0119959.

Miyadate, H., S. Adachi, A. Hiraizumi, K. Tezuka, N. Nakazawa, T. Kawamoto, K. Katou, I. Kodama, K. Sakurai, H. Takahashi *et al.* (2011) OsHMA3, a P-1B-type of ATPase affects root-to-shoot cadmium translocation in rice by mediating efflux into vacuoles. *New Phytol.* 189: 190–199.

Murakami, M., N. Ae, S. Ishikawa, T. Ibaraki and M. Ito (2008) Phyto-extraction by a high-Cd-accumulating rice: reduction of Cd content



- of soybean seeds. *Environ. Sci. Technol.* 42: 6167–6172.
- Murakami, M., F. Nakagawa, N. Ae, M. Ito and T. Arao (2009) Phyto-extraction by rice capable of accumulating Cd at high levels: reduction of Cd content of rice grain. *Environ. Sci. Technol.* 43: 5878–5883.
- Murray, M.G. and W.F. Thompson (1980) Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res.* 8: 4321–4325.
- Oda, K., M. Otani, S. Uruguchi, T. Akihiro and T. Fujiwara (2011) Rice ABCG43 is Cd inducible and confers Cd tolerance on yeast. *Biosci. Biotechnol. Biochem.* 75: 1211–1213.
- Pritchard, J.K., M. Stephens and P. Donnelly (2000) Inference of population structure using multilocus genotype data. *Genetics* 155: 945–959.
- Rapisarda, V., E. Miozzi, C. Loreto, S. Matera, C. Fenga, R. Avola and C. Ledda (2018) Cadmium exposure and prostate cancer: insights, mechanisms and perspectives. *Front. Biosci. (Landmark Ed)* 23: 1687–1700.
- Rose, A.B. (2004) The effect of intron location on intron-mediated enhancement of gene expression in *Arabidopsis*. *Plant J.* 40: 744–751.
- Rose, A.B. (2008) Intron-mediated regulation of gene expression. *Curr. Top. Microbiol. Immunol.* 326: 277–290.
- Rose, A.B., A. Carter, I. Korf and N. Kojima (2016) Intron sequences that stimulate gene expression in *Arabidopsis*. *Plant Mol. Biol.* 92: 337–346.
- Sasaki, A., N. Yamaji, K. Yokosho and J.F. Ma (2012) Nramp5 is a major transporter responsible for manganese and cadmium uptake in rice. *Plant Cell* 24: 2155–2167.
- Satoh-Nagasawa, N., M. Mori, N. Nakazawa, T. Kawamoto, Y. Nagato, K. Sakurai, H. Takahashi, A. Watanabe and H. Akagi (2012) Mutations in rice (*Oryza sativa*) heavy metal ATPase 2 (OsHMA2) restrict the translocation of zinc and cadmium. *Plant Cell Physiol.* 53: 213–224.
- Shimo, H., Y. Ishimaru, G. An, T. Yamakawa, H. Nakanishi and N.K. Nishizawa (2011) Low cadmium (LCD), a novel gene related to cadmium tolerance and accumulation in rice. *J. Exp. Bot.* 62: 5727–5734.
- Song, J.K., H. Luo, X.H. Yin, G.L. Huang, S.Y. Luo, D.R. Lin, D.B. Yuan, W. Zhang and J.G. Zhu (2015) Association between cadmium exposure and renal cancer risk: a meta-analysis of observational studies. *Sci. Rep.* 5: 17976.
- Takahashi, R., Y. Ishimaru, H. Nakanishi and N.K. Nishizawa (2011) Role of the iron transporter OsNRAMP1 in cadmium uptake and accumulation in rice. *Plant Signal. Behav.* 6: 1813–1816.
- Takahashi, R., K. Bashir, Y. Ishimaru, N.K. Nishizawa and H. Nakanishi (2012a) The role of heavy-metal ATPases, HMAs, in zinc and cadmium transport in rice. *Plant Signal. Behav.* 7: 1605–1607.
- Takahashi, R., Y. Ishimaru, H. Shimo, Y. Ogo, T. Senoura, N.K. Nishizawa and H. Nakanishi (2012b) The OsHMA2 transporter is involved in root-to-shoot translocation of Zn and Cd in rice. *Plant Cell Environ.* 35: 1948–1957.
- Takahashi, R., Y. Ishimaru, H. Shimo, K. Bashir, T. Senoura, K. Sugimoto, K. Ono, N. Suzui, N. Kawachi, S. Ishii *et al.* (2014) From laboratory to field: OsNRAMP5-knockdown rice is a promising candidate for Cd phytoremediation in paddy fields. *PLoS ONE* 9: e98816.
- Tang, L., B. Mao, Y. Li, Q. Lv, L. Zhang, C. Chen, H. He, W. Wang, X. Zeng, Y. Shao *et al.* (2017) Knockout of OsNramp5 using the CRISPR/Cas9 system produces low Cd-accumulating indica rice without compromising yield. *Sci. Rep.* 7: 14438.
- Thompson, J.D., T.J. Gibson, F. Plewniak, F. Jeanmougin and D.G. Higgins (1997) The CLUSTAL\_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* 25: 4876–4882.
- Thornberry, J.M., M.M. Goodman, J. Doebley, S. Kresovich, D. Nielsen and E.S. IV. Buckler (2001) Dwarf8 polymorphisms associate with variation in flowering time. *Nat. Genet.* 28: 286–289.
- Ueno, D., N. Yamaji, I. Kono, C. F. Huang, T. Ando, M. Yano and J.F. Ma (2010) Gene limiting cadmium accumulation in rice. *Proc. Natl. Acad. Sci. USA* 107: 16500–16505.
- Ueno, D., E. Koyama, N. Yamaji and J.F. Ma (2011) Physiological, genetic, and molecular characterization of a high-Cd-accumulating rice cultivar, Jarjan. *J. Exp. Bot.* 62: 2265–2272.
- Uraguchi, S., T. Kamiya, T. Sakamoto, K. Kasai, Y. Sato, Y. Nagamura, A. Yoshida, J. Kyozuka, S. Ishikawa and T. Fujiwara (2011) Low-affinity cation transporter (*OsLCT1*) regulates cadmium transport into rice grains. *Proc. Natl. Acad. Sci. USA* 108: 20959–20964.
- Uraguchi, S., Y. Sone, Y. Ohta, N. Ohkama-Ohtsu, C. Hofmann, N. Hess, R. Nakamura, Y. Takanezawa, S. Clemens and M. Kiyono (2017) Identification of C-terminal regions in *Arabidopsis thaliana* phytochelatin synthase 1 specifically involved in activation by arsenite. *Plant Cell Physiol.* 59: 500–509.
- Van Maele-Fabry, G., N. Lombaert and D. Lison (2016) Dietary exposure to cadmium and risk of breast cancer in postmenopausal women: A systematic review and meta-analysis. *Environ. Int.* 86: 1–13.
- Vinceti, M., M. Venturelli, C. Sighinolfi, P. Trerotoli, F. Bonvicini, A. Ferrari, G. Bianchi, G. Serio, M. Bergomi and G. Vivoli (2007) Case-control study of toenail cadmium and prostate cancer risk in Italy. *Sci. Total Environ.* 373: 77–81.
- Wu, D., K. Sato and J.F. Ma (2015) Genome-wide association mapping of cadmium accumulation in different organs of barley. *New Phytol.* 208: 817–829.
- Xu, C., Y. Liu, Y. Li, X. Xu, C. Xu, X. Li, J. Xiao and Q. Zhang (2015) Differential expression of GS5 regulates grain size in rice. *J. Exp. Bot.* 66: 2611–2623.
- Yamaji, N., J. Xia, N. Mitani-Ueno, K. Yokosho and J.F. Ma (2013) Preferential delivery of zinc to developing tissues in rice is mediated by P-Type heavy metal ATPase OsHMA2. *Plant Physiol.* 162: 927–939.
- Yan, J., P. Wang, P. Wang, M. Yang, X. Lian, Z. Tang, C.-F. Huang, D.E. Salt and F.J. Zhao (2016) A loss-of-function allele of OsHMA3 associated with high cadmium accumulation in shoots and grain of Japonica rice cultivars. *Plant Cell Environ.* 39: 1941–1954.
- Yang, M., Y. Zhang, L. Zhang, J. Hu, X. Zhang, K. Lu, H. Dong, D. Wang, F.-J. Zhao, C.-F. Huang *et al.* (2014) OsNRAMP5 contributes to manganese translocation and distribution in rice shoots. *J. Exp. Bot.* 65: 4849–4861.
- Yuan, L., S. Yang, B. Liu, M. Zhang and K. Wu (2012) Molecular characterization of a rice metal tolerance protein, OsMTP1. *Plant Cell Rep.* 31: 67–79.
- Zhu, C., M. Gore, E.S. Buckler and J. Yu (2008) Status and prospects of association mapping in plants. *Plant Genome* 1: 5–20.
- Zondervan, K.T. and L.R. Cardon (2004) The complex interplay among factors that influence allelic association. *Nat. Rev. Genet.* 5: 89–100.