BMC Plant Biology

RESEARCH ARTICLE   **Open Access**

Genome-wide association studies of ionomic and agronomic traits in USDA mini core collection of rice and comparative analyses of different mapping methods

Shuai Liu[1†], Hua Zhong[2†], Xiaoxi Meng[1], Tong Sun[2], Yangsheng Li[2], Shannon R. M. Pinson[3], Sam K. C. Chang[4] and Zhaohua Peng[1*]

Abstract

Background: Rice is an important human staple food vulnerable to heavy metal contamination leading to serious concerns. High yield with low heavy metal contamination is a common but highly challenging goal for rice breeders worldwide due to lack of genetic knowledge and markers.

Results: To identify candidate QTLs and develop molecular markers for rice yield and heavy metal content, a total of 191 accessions from the USDA Rice mini-core collection with over 3.2 million SNPs were employed to investigate the QTLs. Sixteen ionomic and thirteen agronomic traits were analyzed utilizing two univariate (GLM and MLM) and two multivariate (MLMM and FarmCPU) GWAS methods. 106, 47, and 97 QTLs were identified for ionomics flooded, ionomics unflooded, and agronomic traits, respectively, with the criterium of p-value $< 1.53 \times 10^{-8}$, which was determined by the Bonferroni correction for p-value of 0.05. While 49 (~ 20%) of the 250 QTLs were coinciding with previously reported QTLs/genes, about 201 (~ 80%) were new. In addition, several new candidate genes involved in ionomic and agronomic traits control were identified by analyzing the DNA sequence, gene expression, and the homologs of the QTL regions. Our results further showed that each of the four GWAS methods can identify unique as well as common QTLs, suggesting that using multiple GWAS methods can complement each other in QTL identification, especially by combining univariate and multivariate methods.

Conclusions: While 49 previously reported QTLs/genes were rediscovered, over 200 new QTLs for ionomic and agronomic traits were found in the rice genome. Moreover, multiple new candidate genes for agronomic and ionomic traits were identified. This research provides novel insights into the genetic basis of both ionomic and agronomic variations in rice, establishing the foundation for marker development in breeding and further investigation on reducing heavy-metal contamination and improving crop yields. Finally, the comparative analysis of the GWAS methods showed that each method has unique features and different methods can complement each other.

Keywords: Rice, Ionomic traits, Agronomic traits, Multivariate GWAS

* Correspondence: zp7@bch.msstate.edu
†Shuai Liu and Hua Zhong contributed equally to this work.
[1]Department of Biochemistry, Molecular Biology, Entomology and Plant Pathology, Mississippi State University, Starkville, MS 39762, USA
Full list of author information is available at the end of the article

Liu *et al. BMC Plant Biology*   (2020) 20:441

Page 2 of 18

Background

Rice is an important cereal which feeds more than half the world's population [1]. With the rapid expansion of global population, food security has become a highly challenging task. Meanwhile, anthropogenic activities such as mining, smelting, chemical engineering, energy related industry, and broad application of pesticides & fertilizers in agriculture have led to frequent heavy metal contamination in soil, including Cadmium (Cd), Manganese (Mn), Nickel (Ni), and metalloid Arsenic (As) [2]. Soil with excessive heavy metal elements represses plant germination and growth, resulting in a decrease of crop yield [3, 4]. Meanwhile, plants uptake the toxic heavy metal elements from contaminated soil and accumulate them in edible plant tissues, leading to food contamination.

The anaerobic growing conditions of flooded rice paddies and the unique physiology of the rice plant allow rice to take up some heavy metals from water and soils in a highly efficient manner and sequester it in different organs within the plants, including the grains consumed by humans. The arsenic concentration in rice grains is roughly about 10 times higher than other crops grown in the same region even if the soil has no or limited anthropogenic contamination [5]. Rice has been reported to contribute substantially to inorganic and organic arsenic [6–8] intake by the human population in many regions of the world. As was ranked on the top of the US Agency for Toxic Substances and Disease Registry (ATSDR) Priority List of Hazardous Substances since 1997 (https://www.atsdr.cdc.gov/spl/index.html#2017spl). It has also been listed as a toxic component by many other countries and treated as a critical contaminant during food safety inspection. Cd is one of the most toxic heavy metals, and can easily reach the food chain due to strong assimilation by crops [9, 10]. Once absorbed, Cd is efficiently retained in the human body and may cause it to stay throughout the life span with an estimated half biology life between 6 to 38 years in kidney and between 4 to 19 years in the liver (ATSDR, 1999).

In contrast to heavy metals, many mineral elements are essential to humans but deficient in rice grains, for example, zinc, calcium, and iron [11]. Increasing the concentrations of these minerals can improve the nutritional value of rice thus promoting human health for those consuming rice as the staple food. However, it is highly challenging to either increase the essential minerals or reducing the heavy metal due to lack of understanding of the genetic bases and molecular mechanisms of the related traits. Further, it is still poorly explored whether the concentration of mineral or heavy metal is associated with agronomic traits. Although there are multiple rice association mapping studies with specific minerals, heavy metals, and agronomic traits, respectively, these studies used either different mapping populations or different statistical analyses [12–14]. Therefore, each of the studies revealed some but not all facets of the genetic bases of rice variations. Recent accessibility to comprehensive sequence data, and the development of software facilitating the use of more powerful statistical analytics, opens the opportunity for more comprehensive study and understanding of the genetic bases of these traits.

The USDA Rice Core Collection, containing about 10% of the whole NSGC (National Small Grains Collection) Rice Collection, was assembled by stratified random sampling method in 2002, which has been evaluated comprehensively for 25 characteristics and proven to be highly representative of the whole collection [15]. The Rice Mini-Core Collection contains approximately 10% of the Core Collection [16]. The grain mineral concentrations have been analyzed under flooded and unflooded growth conditions [17], and the agronomic traits have also been evaluated for the Core Collection [18]. But most of these researches were done before the genome sequencing data was available.

Biparental genetic mapping and Genome-wide-association-study (GWAS) are the two different tools for mapping Quantitative Trait Locus (QTL). GWAS involves studying natural populations with greater historical recombination events, which could detect more QTLs from broader genetic variation [19]. Therefore, it can explore genetic resources that cannot be revealed by studying the offspring of biparental crosses in linkage mapping [20]. GWAS has been applied successfully to a variety of plants, including *Arabidopsis* [21], maize [22, 23], barley [24], wheat [25], rice [26, 27], soybean [28], and cotton [29]. It is a critical tool for crop improvement.

Univariate GWAS is a mapping method that has been successfully used for gene mapping in plants and animals. However, a large number of genes may not be detected (false negative QTLs) due to the confounding problems between population structure, kinship, and markers. The population structure causes genome-wide linkage disequilibrium between unlinked loci, which leads to statistical confounding in genome-wide association studies. Mixed models have been shown to deal well with the confounding effects of a large number of small effect loci in the diffusion background, but they do not always account for large effect loci [30]. Multivariate GWAS method considers the confounding problem between covariates and test marker to detect more QTLs and previous reports showed that multivariate GWAS had lower FDR when using the same threshold compared with univariate GWAS method [30]. In recent years, a large number of multivariate GWAS methods have been developed, including MLMM (multi-locus mixed-model) [30], FarmCPU (Fixed and random model Circulating Probability Unification) [31], mrMLM (multi-locus random-SNP-

effect MLM) [32], FASTmrMLM (fast mrMLM) [33], FASTmrEMMA (fast multi-locus random-SNP-effect efficient mixed model analysis) [34], pLARmEB (polygenic-background-control-based least angle regression plus empirical Bayes) [35], pKWmEB (integration of Kruskal-Wallis test with empirical Bayes) [36], ISIS EM-BLASSO (iterative modified-sure independence screening expectation-maximization-Bayesian least absolute shrinkage and selection operator) [37], and GPWAS (Genome-Phenome Wide Association Study) [38]. The MLMM [30] uses forward-backward stepwise linear mixed-model regression, forward stepwise uses the most significant associated SNP as a new fixed-effect covariate (cofactor) and creates a new model until reaching a pre-specified maximum number of forward steps, backward stepwise means to remove least significant SNP and create a new smaller model until only one selected marker is left. Whereas, FarmCPU [31] performs marker tests with associated markers as covariates in a Fixed Effect Model (FEM), and then optimization on the associated covariate markers in a Random Effect Model separately. These multivariate GWAS methods were successfully applied to several different crop species, including cotton [39], rice [40, 41], foxtail millet [42], soybean [43, 44], maize [45, 46], and wheat [47, 48].

GWAS has been widely used for rice mapping research. However, most of the studies used univariate methods. The multivariate analysis methods and the deep DNA sequencing resources are still poorly explored. For example, Yan et al.(2014) [49] performed limited GWAS on agronomic traits using only 155 molecular markers (154 SSR makers and one indel marker). However, the related loci could not be precisely defined in the study due to limited number of markers. With the development of the sequencing technology, more markers, especially SNP markers, become available and have been shown to be successful in association mapping studies [50, 51]. The resequencing data of the mini core has been deposited to SRA database https://www.ncbi.nlm.nih.gov/sra (Accession: PRJNA301661) recently [52], which opened an opportunity for us to improve GWAS with higher density of genotypic data and evaluate the validity of different GWAS analysis methods. In order to identify all possible QTLs controlling the ionomic and agronomic traits for marker development and gene characterization, we employed two univariate GWAS methods (GLM and MLM) and two multivariate GWAS methods (MLMM and FarmCPU) to detect the related QTLs in USDA rice mini core collection. The analyzed ionomics traits included As, Ca, Co, Cd, Cu, Fe, K, Mg, Mn, Mo, Ni, P, Rb, S, Sr, and Zn) and the agronomic traits involved amylose [AMYLOSE], awn type [AWNTYPE], flowering time [DAYSFLOWER], hull color [HULLCOLOR], hull cover [HULLCOVER], kernel length [KERNELLEN], kernel width [KERNELWID], kernel rate [KERNELRAT], kernel weight [KERNEL WT], lodging [LODGING], panicle type [PANICLETYPE], plant height [PLANTHT], and plant type [PLANTTYPE]). Our results showed that we successfully remapped 49 loci/genes which have been shown to play essential roles in rice agronomic and ionomic trait control. Meanwhile, over 200 novel loci involved in heavy metal, minerals, and agronomic traits control were discovered. In addition, multiple candidate genes involved in ionomic and agronomic traits control were identified via DNA sequence, expression, and homologous analyses. These studies provided novel insights into the genetic basis of ionomic and agronomic trait variations in rice and possible correlations among these traits. The results will have critical reference value in further fine mapping the related genetic loci and in guiding rice breeding.
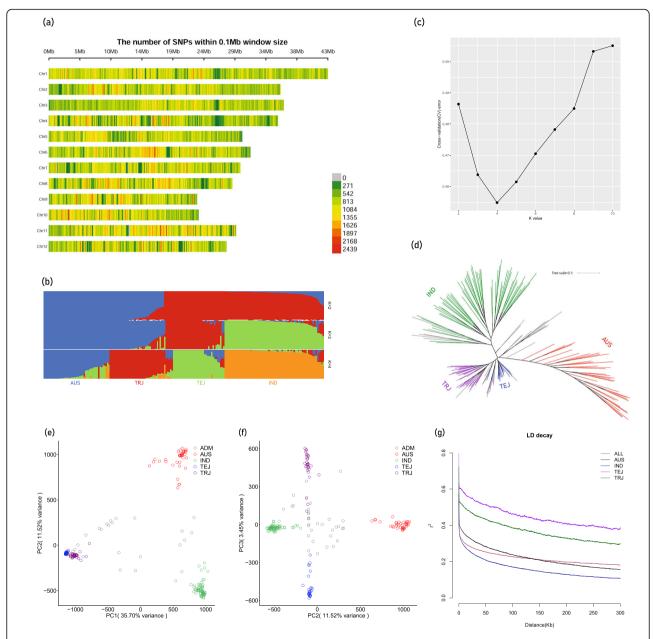
Results

Characteristics of SNPs

High-quality re-sequencing raw data of 191 accessions derived from the USDA Rice Mini Core (Supplementary Table S1), was retrieved from NCBI SRA database (Accession: PRJNA301661) [52]. Genotyping of the 191 accessions were performed by GATK software. A total of 3,259,478 SNPs was obtained after filtration by minor allele frequencies (≥0.05) and integrity (≥0.4). Imputed SNPs, which were generated by Beagle 5.0 software [53], were used for further analyses. Distribution of these SNPs in the genome is summarized in Table 1 and Fig. 1a, and the overall

Table 1 Summary of the SNPs across 12 chromosomes of *Oryza sativa*

Chromosome	Number of SNPs	Length of Chromosome (bp)	Density of SNP (bp/SNP)
1	375,296	43,270,923	115.30
2	301,111	35,937,250	119.35
3	294,312	36,413,819	123.73
4	279,049	35,502,694	127.23
5	253,001	29,958,434	118.41
6	287,238	31,248,787	108.79
7	253,651	29,697,621	117.08
8	261,070	28,443,022	108.95
9	212,238	23,012,720	108.43
10	222,521	23,207,287	104.29
11	289,053	29,021,106	100.40
12	230,938	27,531,856	119.22
Total	3,259,478	373,245,519	114.51

Fig. 1 Sequence and structure analysis of USDA mini core collection. **a** Distribution of SNPs on the rice chromosomes. Number of SNPs per 0.1 Mb window was shown as a color index (bottom right), **b** Ancestries analysis for each individual was inferred using admixture, **c** Cross-validation error (CV) score across different K value. The best K value (K = 4) was chosen according to the lowest CV score for the admixture analysis, **d** Phylogenetic tree of 191 rice accessions. Green indicated Indica (IND) rice, Red indicated Aus (AUS) rice, Purple represented Tropical Japonica (TRJ) rice; Blue represented Temperate Japonica (TEJ) rice, **e** PCA showing genetic variation in the rice accessions with first and second PCs, the color was defined by current Admixture analysis. **f** PCA showing genetic variation in the rice accessions with second and third PCs, the color was defined by current Admixture analysis. **g** Genome-wide average LD decay estimated from the whole population and each subpopulation

SNP density in the genome was 114.51 (bp/SNP). The number of SNPs ranged from 212,238 to 375,296 across the twelve rice chromosomes. Chromosome 4 held the minimum marker density with 127.23 (bp/ SNP), while chromosome 11 exhibited a maximum marker density with one SNP per 100.40 bp.

Population structure and linkage disequilibrium

Admixture analysis divided the 191 accessions into four ancestries, including Indica (63 accessions), Aus (37 accessions), Temperate Japonica (28 accessions), and Tropical Japonica (31 accessions) under the best K model (K = 4) (Fig. 1b), which was determined by the lowest CV (cross-

Liu *et al. BMC Plant Biology*   (2020) 20:441

Page 5 of 18

validation error) score (Fig. 1c). Thirty-two accessions are classified as admixture (ADM) since the ratio from each single subpopulation is below 70%.

In order to reduce the amount of calculations, high-quality SNPs (SNPs integrity above 0.8) were selected to construct a maximum likelihood (ML) tree to illustrate the phylogenetic relationship of the 191 rice accessions (Fig. 1d). The population was divided into four subpopulations and the color for each clade was determined according to the Admixture analysis results. The relationship obtained from phylogenetic tree is in line with the Admixture analysis.

Principal component analysis (PCA) was performed based on the 3,259,478 SNPs. Four conceivable subpopulations were separated by PC1, PC2, and PC3. The first three principal components (PCs) explained over 50% of the genetic variation. The first PC separates Indica and Japonica subpopulations (35.70%), the second PC distinguishes the Aus and Indica varieties, while the third PC separates Temperate Japonica and Tropical Japonica varietal groups (Fig. 1e and f). Based on the results from the Admixture analysis, phylogenetic tree and PCA, the population was divided into four subgroups. In addition, the decay of LD with the physical distance between SNPs in all population occurred at 191 kb ($r^2 = 0.2$) (Fig. 1g), which is similar to that of a previous study [54]. Indica subpopulation exhibited the most rapid LD decay and Temperate Japonica showed the most extended LD.

Correlation of different traits

Correlation analyses between grain ionomics in flooded environment and agronomic traits (Supplementary Fig. S1a), between grain ionomics in unflooded environment and agronomic traits (Supplementary Fig. S1b), and between grain ionomics in flooded versus unflooded growth conditions (Supplementary Fig. S1c) were conducted. The results showed that days to flowering has strong correlation with Rb in flooded (0.53) and unflooded (0.57) (Supplementary Fig. S1a and b) environments. The accumulation of Cd, Mo, and Rb in rice grain in flooded environment and unflooded environment are correlated at $r = 0.52$, 0.81, and 0.60, respectively (Supplementary Fig. S1c).
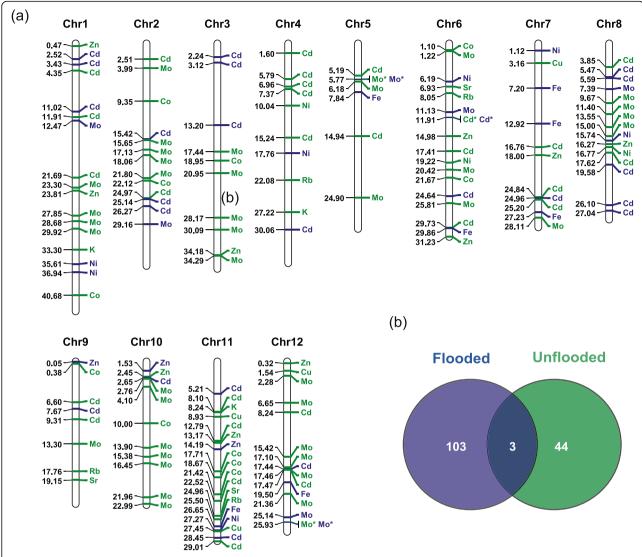
Genome-wide association study by univariate GWAS and multivariate GWAS

Sixteen grain ionomic traits (As, Ca, Co, Cd, Cu, Fe, K, Mg, Mn, Mo, Ni, P, Rb, S, Sr, and Zn) under flooded and unflooded conditions were the same as reported [17]. Thirteen agronomic traits, including AMYLOSE, AWNTYPE, DAYSFLOWER, HULLCOLOR, HULLCOVER, KERNELLEN, KERNELWID, KERNELRAT, KERNELWT, LODGING, PANICLETYPE, PLANTHT, and PLANTTYPE [18] were shared by Yan as reported

[55, 56] and recorded using the methods described by Li et al [57–59]. All these traits were analyzed using two univariate GWAS (GLM and MLM) and two multivariate GWAS (MLMM and FarmCPU) methods to identify QTLs. A total of 106 significant QTLs (*p*-value $< 1.53 \times 10^{-8}$) were detected to be associated with concentrations of 9 ionomic traits (Cd, Co, Cu, K, Mo, Ni, Rb, Sr, and Zn) in rice grain under flooded condition, in which 63, 68, 17, and 44 significant QTLs were identified by GLM, MLM, MLMM, and FarmCPU, respectively (Fig. 2 and Supplementary Fig. S4b). For Cd, twenty-eight significant QTLs were identified. Three of them located near published genes (*CAL1* [14], *OsHMA2* [60], *rgMT* [61]) which have shown to be related to Cd accumulation or resistance. Seven of them were identified in previous mapping studies using univariate methods (Supplementary Table S2). All of the seven QTLs were identified by univariate GWAS methods (GLM and MLM) but only two of the seven were also detected by multivariate methods (MLMM and FarmCPU) in our study. For Co, a total of eleven significant QTLs were detected. Two (one was identified by univariate methods and the other was detected by FarmCPU) of them co-located with previously reported QTLs. Nine of them were new QTLs discovered in the current study, MLMM method discovered 2 significant QTLs and FarmCPU method identified 7 QTLs, respectively. Three QTLs were detected to be significantly associated with K, one of which (only detected by FarmCPU) was also detected in previous studies [62]. For Zn, ten significant QTLs were identified, three of which co-located with previously reported loci [12, 13, 62, 63]. Among them, one significant QTL posited around 18,001,929 bp of Chromosome 7 was detected by both univariate and multivariate methods, which located near reported QTL *qZN-7* [13]. The other two QTLs were detected by FarmCPU method only.

In the unflooded environment, only 47 QTLs were detected to be significantly associated with Cd, Fe, Mo, Ni, and Zn concentration. Specifically, 29, 25, 10, and 20 significant QTLs were identified by GLM, MLM, MLMM, FarmCPU, respectively (Fig. 2 and Supplementary Fig. S4c). Twenty-three identified QTLs were related to Cd, one of which located near *CAL1* gene [14], eight QTLs co-located with previous studies (Supplementary Table S2). Among the eight co-localization QTLs, five were detected by univariate methods and three were identified by multivariate methods. For Fe, seven significant QTLs were identified, two of which were also reported by previous studies [62, 63], and both were detected by FarmCPU only in the current study. We noticed that for the traits that many QTLs were identified using GLM and MLM methods, the numbers of QTLs identified by MLMM and FarmCPU were less as shown in the case of Cd and Mo. When GLM and MLM method failed to
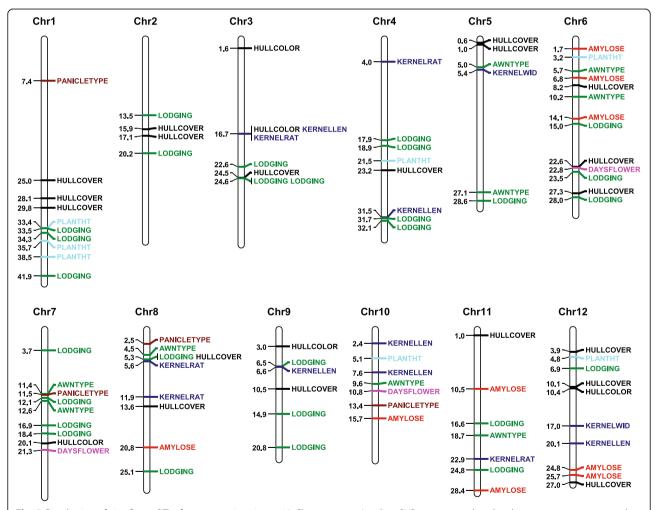
Fig. 2 QTLs related to ionomic traits. **a** Distribution of significant QTLs for ionomic traits across the 12 chromosomes of rice under flooded and unflooded environment. Leading SNP was mapped to the chromosome to represent the QTLs' physical location (Mb). The physical position of each lead SNP was shown on the left side and the corresponding ionomic traits displayed on the right side. QTLs from different growth conditions were distinguished by different colors: green, flooded condition; blue, unflooded condition. An asterisk indicates the locus which was detected from both conditions, **b** The Venn diagram shows the numbers of overlapped loci within or between different conditions

identify or only identified a few significant QTLs, QTLs were successfully identified by MLMM and FarmCPU methods as shown in the case of identifying QTLs for Co, Fe, K, and Zn concentration regulation (Supplementary Fig. S2). The QQ-plots in Fig. S2 (c) shows the power of MLMM and FarmCPU, which indicated no evidence for inflation but strong evidence for real effects. In contrast, the QQ-plots of GLM and MLM in Fig. S2 (c) shows the tendency of false positive peaks. This observation was further confirmed when mining the key candidate genes controlling ionomic and agronomic traits as shown in the section below. Interestingly, only 3 of the 106 (ionomic)

QTLs identified in flooded growth condition were shared with the QTLs identified in unflooded condition. The three QTLs (QTLs marked with an asterisk on Chromosome 5, 6, and 12; Fig. 2, Supplementary Fig. S2, and Supplementary Table S2) share by both growth condition were associated with Cd and Mo concentration regulation, indicating that the traits of these three QTLs were not impacted by water conditions. Furthermore, several loci were shown to be associated with more than one trait, indicating these QTLs may be pleiotropy. For example, the region around 15.5 Mb on chromosome 2 is associated with Cd and Mo (Fig. 2).

For agronomic traits, a total of 97 significant QTLs (p-value $< 1.53 \times 10^{-8}$) were detected for the thirteen agronomic traits described above except for KERNELWT and PLANTTYPE (Fig. 3, Supplementary Table S2). In detail, 39, 16, 29, and 50 significant QTLs were identified by GLM, MLM, MLMM, and FarmCPU, respectively (Supplementary Fig. S4a). *Wax* [64] and *ALK* [65] genes were shown to be significantly associated with amylose content, which is consistent with previous reports. Gain size is a key agronomic trait that strongly linked to yield and quality. Many QTLs have been reported associating with rice grain size, which is decomposed into grain length, width, and thickness (*GS3* [66], *GS5* [67], *GW5* [68], *GW8* [69], *GL7* [70], *TGW6* [71], etc.). In this study, four types of rice grain size-related traits included kernel length (KERNELLEN), kernel width (KERNELWI

D), kernel rate (KERNELRAT), and kernel weight (KERNELWT) were analyzed. A total of 13 QTLs were detected by univariate and multivariate GWAS methods. Among them, three were detected by univariate (GLM or MLM) GWAS methods and twelve of them were detected by multivariate (MLMM and FarmCPU) GWAS methods. Six, two, and five of the 13 QTLs were found to be associated with KERNELLEN, KERNELWID, and KERNELRAT, respectively. No significant QTL was shown to be associated with KERNELWT. Remarkably, one QTL (Chromosome 3 position 16,733,441) was detected by all the four methods (Supplementary Fig. S3f). The QTL locates on gene *GS3* [66], which is a major gene regulating grain size and organ size. It is worth noting that, five more significant SNPs were identified by FarmCPU and one of them located on chromosome 4

Fig. 3 Distribution of significant QTLs for agronomic traits on 12 Chromosomes. Leading SNP was mapped to the chromosome to represent the QTLs' physical location (Mb). The physical position of each lead SNP was shown on the left side and the corresponding agronomic traits displayed on the right side. QTLs of different type of agronomic traits were distinguished by different colors: red, amylose; blue, grain size (kernel length, kernel width, and kernel rate); black, hull cover and hull color; purple, days to flower; brown, panicle type; green, lodging and awn type

Liu *et al. BMC Plant Biology*   (2020) 20:441

Page 8 of 18

situated nearby the *NAL1* gene, which has been shown to be related to rice yield [72]. For the trait of KERNEL WID, two significant QTLs were only detected by MLMM method solely and other methods failed to identify candidate QTLs. One of the identified QTLs positioned around 5,364,561 bp of chromosome 5, which was found approximately 0.56 kb apart from a well-known *GW5* gene controlling rice grain width (Supplementary Fig. S3g) [73]. These results demonstrated the power of GWAS, especially the power of the multivariate (MLMM and FarmCPU) GWAS methods.

Mining candidate genes of agronomic-related traits

Lodging and Plant height are both related to cell wall properties, which could impact rice yield. Appropriate plant height and the strong stem are required for stable production [74]. A cluster of SNPs around 33.4 Mb on chromosome 1 (Lodging: 33,010,693 to 33,975,764 with leading SNP at 33,469,251; Plant height: 33,181,529 to 33,730,067 with leading SNP at 33,363,796) is shown to be significantly associated with lodging and plant height (Supplementary Fig. S3j and k). Through LD block analysis, we defined a 72.37 kb blocks (33,458,683- 33,531, 049) containing 12 genes to be the candidate locus. Among these genes, *OsPME6* (*Os01g0788400*) is annotated as pectin methylesterase 6, which is related to the cell wall modification process. We further conducted blastP analysis with *Arabidopsis thaliana* and found that it shares high homology (E value = 3E-178) with Arabidopsis gene *PME18* (*AT1G11580*) (Supplementary Table S3). The expression of *PME18* increased dramatically under hyper gravity stimulus. It is speculated that pectin esterases induced pectin demethylation of carboxyl groups which increased the rigidity of pectin gel in the cell wall through calcium bridges [75]. Therefore, it is worth to test if *OsPME6* regulates rice lodging and height.
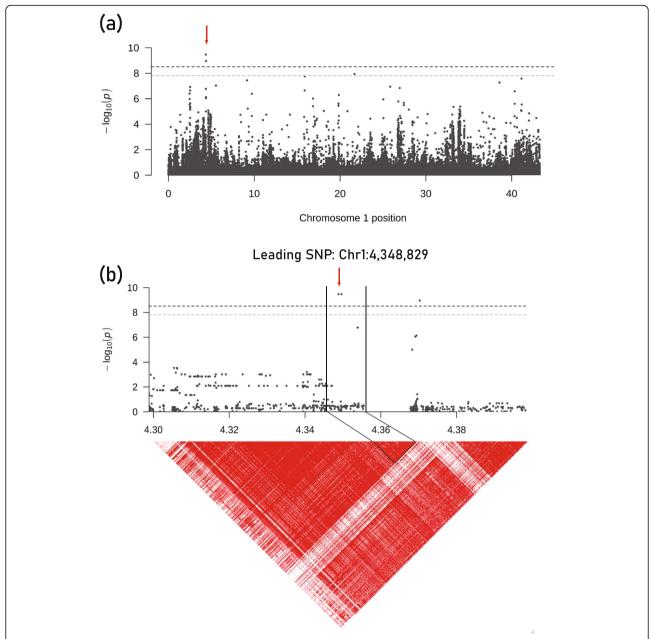
Flowering time is another important trait critical to rice production. Rice is a typical short-day (SD) flowering plant whose flowering is greatly affected by day length. A number of genes [76–79] have been reported to regulate rice flowering-time. In the current study, a total of three QTLs were significantly associated with the flowering time. Two of them were detected by FarmCPU exclusively on Chromosome 7 and 10. The other QTL on Chromosome 6 was detected by all the four different GWAS methods (Supplementary Fig. S3c). The haplotype analysis showed that this region only harbored 2 genes (*OsPLL9* and *OsPLL10*). Among them, *OsPLL9* (*Os06g0583900*) located 7.15 kb away from the leading SNP. This gene is a homolog of pectate lyase gene, which may play crucial roles during rice panicle development [80]. *OsPLL9* is highly expressed in Stamen (one day before flowering), Palea (one day before flowering),

and Panicle5 (heading stage) (Supplementary Fig. S5). Thus, *OsPLL9* has the potential to be a candidate gene with a critical role in rice flowering.

Mining candidate genes of ionomic traits

As a result, 28 and 23 significant QTLs were detected to be associated with Cd concentration in the flooded and unflooded environment, respectively. QTLs near *CAL1* (Chr2:25,190,487-25,191,188) were associated with rice grain Cd accumulation in both flooded (Leading SNP; Chr2: 24,968,588) and unflooded condition (Leading SNP; Chr2: 25,143,071). *CAL1* was annotated as a defensin-like protein, which could regulate Cd accumulation of rice leaves through translocating Cd from cytosol into extracellular spaces, but not rice grains [14]. We then further analyzed the genes around the QTLs and found that there is an ABC transporter (*Os01g0121700*), its phosphorylation level was up-regulated under high Cd treatment (100 μM CdCl$_2$·2.5H$_2$O) and it has been shown that the transporter reduces the concentration of Cd $^{2+}$ through transporting PCs-Cd into vacuole [81]. Another QTL (Chr6: 29,733,715) is also showed strongly related to Cd concentration in rice grain. This QTL locates near a known gene *OsHMA2* (about 253 kb away from the leading SNP, but not in LD region), which may decrease rice grain Cd concentration through suppressing the expression level of *OsHMA2* [60]. In addition, significantly associated SNP (Chr11: 29,014,045) posited near *rgMT* gene, which was a metallothionein protein responded to the Cd stress in *E. coli* [61]. Comparing the QTLs detected in this study with previously reported studies, we found that over fifteen QTLs were co-localized with reported loci. The details were shown in Supplementary Table S2. Meanwhile, thirty-two new QTLs were identified.

By applying the procedure mentioned in the methods section, we obtained a list of genes that represent plausible candidates of the causal gene for each of the loci controlling elemental concentrations in rice (Supplementary Table S3 and S5). We select three loci associated with Cd for further investigation with the aim of identifying the causal novel genes. A significant QTL associated with Cd was identified on Chromosome 1 around nucleotide at 4,348,829 bp with *p*-value 3.37E-10 (MLM method). This QTL posited within a 9.97 kb block (Chr1: 4,345,517 - 4,355,489) containing only one candidate gene *OsWRKY102* (*Os01g0182700*) (Fig. 4a and b). BlastP analysis showed that the *OsWRKY102* (*Os01g0182700*) was highly homologous (1.00E-58) to *Arabidopsis WRKY13* (*AT4G39410*) gene (Supplementary Table S4). *WRKY13* activates the expression of gene *PDR8* that encodes a Cd^{2+} extrusion pump, resulting in reduced Cd accumulation [82]. The expression profile from public data showed that *OsWRKY102* was
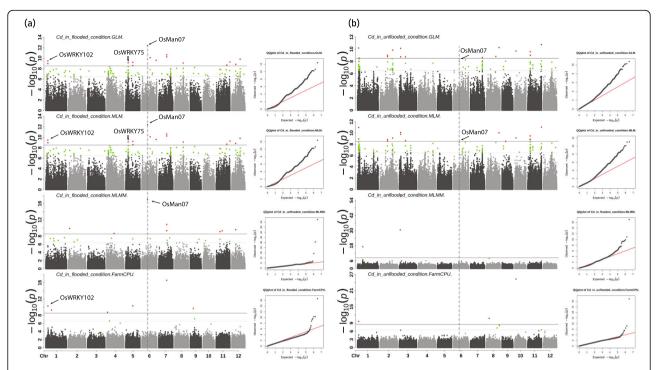
Fig. 4 Identification of *OsWRKY102* as a Cadmium concentration QTL in rice grain in flooded condition using MLM method. (**a**) Genome-wide association signals on chromosome 1, (**b**) Genome-wide association signals in the region at 4.299–4.399 Mb on chromosome 1 and LD heatmap (bottom)

intensively higher expressed in stem comparing to other tissues (Supplementary Fig. S6a). When treated with a high concentration of cadmium, the expression level of *OsWRKY102* increased rapidly in both shoot and root (Supplementary Fig. S6b). Overall, the results suggested that *OsWRKY102* responds at high-level cadmium treatment and regulates cadmium uptake and accumulation in rice. Another QTL (Chromosome 5 posited around 14,941,717) was identified in a flooded environment. Through LD analysis, we defined an

18.65 kb block (Chr5: 14,930,444 - 14,949,090) containing two genes, *Os05g0321600* and *Os05g0321900*. Among them, *Os05g0321900* (*OsWRKY75*) was annotated as DNA-binding WRKY domain-containing protein (Fig. 5a). BlastP analysis found that this gene shared high homology (4E-52) with *WRKY55* (*AT2G40740*) in *Arabidopsis thaliana*, which regulated gold uptake and tolerance. Remarkably, one QTL (Chromosome 6 around position 11,906,590) was identified under both growth environments (Fig.

Fig. 5 Genome-wide association analysis of Cd concentration with GLM, MLM, MLMM, and FarmCPU methods. **a** in flooded condition and, **b** unflooded condition. Quantile-quantile plot of each model. Black arrows indicated candidate genes. The horizontal dot grey line and green dots indicated the Bonferroni-corrected significance thresholds and SNPs at -$\log_{10}(p)$ = 7.81. The horizontal solid grey line and red dots indicated the Bonferroni-corrected significance thresholds and SNPs at -log10(p) = 8.51. The vertical dash grey lines indicate the common QTL detected in flooded and unflooded condition

5a and b). This SNP is a singleton, and we failed to define haplotype blocks for it. For further mining causal gene, we expanded the region as we mentioned in the method section. Finally, we found an important gene *OsMan07* (*Os06g0311600*) was only 24.75 kb away from the leading SNP. This region was reported in previous study [83]. BlastP analysis found this gene had a high similarity (6E-108) to *Man3* (*AT3G10890*) gene in *Arabidopsis thaliana* (Supplementary Table S4). Overexpression of *MAN3* enhanced Cd accumulation and tolerance, whereas loss-of-function of *MAN3* led to decreased Cd accumulation and tolerance [84]. All the genes' expression patterns located in the haplotype region associated with Cd were showed in Supplementary Figure S7. Overall, thirty-two new QTLs were identified in addition to precise identification of the loci reported in previous Cd studies.

PIP2;6 has been suggested to be involved in arsenic concentration control [85]. Although no SNP with Bonferroni-corrected significant thresholds -log 10(*p*) above 7.81 was discovered, there was an SNP peak with -log(*p*) around 6 on the chromosome 4 near the published gene *PIP2;6* (Supplementary Fig. S2a), suggesting that *PIP2;6* is located near a candidate QTL revealed by GWAS analysis.

Comparison of univariate and multivariate GWAS methods

Our results demonstrated that there was not a single method that was able to detect all the QTLs while many loci were detected by all methods. The *GS3* gene was shown to be associated with grain length by all the four tested methods. However, the *GW5* gene was detected to be related to grain width only by multivariate GWAS. Similarly, the Cu related QTLs in flooded conditions and Fe related QTLs in unflooded condition were also detected by multivariate method only. Interestingly, when a large number of QTLs were identified by the univariate method, the QTLs identified by multivariate method were substantially reduced. For example, over 29 QTLs for Cd were identified by univariate methods in flooded and unflooded environments but only six QTLs were identified by multivariate. Further, it appeared that the multivariate methods were able to pinpoint the location of the QTLs more precisely on the chromosome compared with the univariate methods in many cases. As shown in Supplementary Figure 3f, the peaks identified by univariate method were much broad than the peaks identified by multivariate methods. On the other hand, the univariate methods also identified many loci exclusively. In the case of LODGING, the

QTL (candidate gene: *OsPME6*) located on chromosome 1 was only detected by GLM (Supplementary Figure 3j) and this gene was related with cell wall formation biological process, indicating that some of candidate QTLs might be ignored when we pursued lower FDR in multivariate methods.

Discussion

The reliability of the GWAS analyses

In this study, many previously reported loci important for agronomic and ionomic traits were rediscovered. The amylose controlling locus *Wax* [86] gene was rediscovered by all four analysis methods and the *ALK* [87] gene was mapped by the two univariate analysis methods. The hull cover genes *OsTCL2* [88] and *OsWOX3* [89] were mapped by the GLM and MLM methods. The kernel length genes *GS3* [66] were mapped by all four methods and the *NAL1* [72] was mapped by the FarmCPU method. The kernel rate gene *GS3* was mapped by GLM, MLMM, and FarmCPU methods and the kernel wide gene *GW5* [73] was mapped by the MLMM method. The two lodging genes (*OsSPL14* [90] and *OsCESA9* [91]) were also successfully mapped by univariate methods. Further, the three known Cd related genes (*CAL1* [14], *OsHMA2* [60], and *rgMT* [61]) were mapped by GLM and MLM methods. Since 49 genes/loci with known functions had been successfully remapped, the results confirmed the accuracy of the imputed SNP dataset and the power of mapping QTLs with GWAS. More importantly, these observations also suggested that the 201 new QTLs discovered in this study are worth to be further validated.

Different GWAS analysis methods possess unique features and complement each other

While many loci were mapped by all of the four tested GWAS methods, each of the four methods identified some of the function known loci exclusively. A kernel length QTL was only identified by FarmCPU method near the yield related gene *NAL1* [72]. The kernel width gene *GW5* [73] on chromosome 5 was only mapped by MLMM method. The lodging gene *OsSPL14* [90] on chromosome 8 was identified by GLM method alone, and the lodging gene *OsCESA9* [91] was only mapped by FarmCPU. There were many genes were mapped by only two or three of the four tested methods. For example, the amylose gene *ALK* [87] on chromosome 6 was mapped by univariate methods GLM and MLM; The kernel rate gene *GS3* [66] on chromosome 3 was mapped by GLM, MLMM, and FarmCPU. These results described above clearly demonstrated that all four methods can be effectively used to perform GWAS analysis and able to identify some of the known loci. However, none of the methods identified all the previously

reported loci. Meanwhile, every method successfully mapped some loci while the other three methods failed to identify. Therefore, it is worth to test every method when our goal is to identify as many of related loci as possible. In addition, it is worth noting that the differences within univariate or within multivariate methods are smaller compared to the differences between univariate and multivariate methods. Therefore, our results suggest that it is better to include at least one univariate and one multivariate method in GWAS analyses for the best coverage of the QTLs.

The relationship between rice grain ionome content and environment

In this study, we conducted the ionomic test of rice grain in flooded and unflooded environment. The results showed the accumulation of most minerals and heavy metals was significantly different in flooded and unflooded environment, indicating the element accumulation in rice grain was affected by water condition, which was coherent with previous studies [92]. However, the content of some minerals and heavy metals was less affected by water conditions. Among them, Mo showed the top consistent accumulation in both flooded and unflooded conditions with correlation efficiency of 0.81, followed by Rb of 0.60, and Cd of 0.52 (Supplementary Fig. S1c). For the identified QTLs, the similar results were obtained. Two Mo QTLs (Chr5:5.77 Mb and Chr12:25.93 Mb) were repeatedly detected in both flooded and unflooded growth conditions. This trend was in agreement with previous study [93], suggesting Mo accumulation in rice was steady and not significantly affected by environment. For the Rb, we only identified significant QTLs in flooded environment, which was not consistent with the correlation study results. For the third-ranked element - Cd, we repeatedly identified one QTL (Chr6: 11.91 Mb) in both flooded and unflooded environments, respectively. Other two Cd concentration related QTLs were detected in a proximate region around 25 Mb on chromosome 2 (24.97 Mb in flooded condition and 25.14 Mb in unflooded condition). The results above suggested that Mo and Cd accumulation in rice grain was barely affected by water conditions while the other elements analyzed in this study were dependent upon the irrigation conditions.

Potential application of the candidate new genes and significant QTLs in future research and breeding

Forty-nine QTLs or genes with known functions in mineral, heavy metal, or important agronomic trait control were rediscovered in the study. Additionally, 201 new QTL loci with critical roles in regulating these traits were identified. More importantly, multiple candidate genes were identified following programmed and manual

sequence analysis of the QTL loci in the mini core collection. Cd and As contamination are a major global concerns for rice production due to widespread contamination resulted from anthropogenic activities and their high toxicity to humans [2]. Cd has an extremely long half-life inside the human body especially in the kidney and liver [94] and inorganic As is a strong carcinogen. We identified 18 QTLs/genes which have been reported to play a role in Cd content control and 32 new QTLs. Moreover, our analyzing results suggested that genes *OsWRKY102* (*Os01g0182700*), *OsWRKY75* (*Os05g0321900*), and *OsMan07* (*Os06g0311600*) are potential candidate genes involved in Cd control in rice. Our results also suggested that the locus with *PIP2;6* gene is probably a QTL involved in As control, thus provided genetic evidence to support prior reports of *PIP2;6*'s potential function in rice, which were based on heterologous expression studies in *Xenopus laevis* oocytes and *Arabidopsis* [85]. The candidate genes involved in heavy metal content regulation are important resources for understanding the underlying mechanism of heavy metal control and agriculture breeding. We found that *OsPME6* has the potential to regulate rice lodging and height and *OsPLL9* has the potential to be a candidate gene with a critical role in rice flowering. Plant height, lodging, and flowering time regulation are critical traits directly related to crop production. Identifying these candidate genes provided new resources for marker development in rice breeding and molecular investigation on the control mechanism of lodging and flower timing. However, it is critical to note that all these candidate genes were identified base on sequence, expression, and homologous analysis of the QTLs, further tests are required to confirm the results before final conclusions.

Conclusion

In this study, comprehensive GWAS analyses for ionomic and agronomic traits based on 3,259,478 SNPs were performed using two univariate methods and two multivariate methods. Under the criterium *p*-value $< 1.53 \times 10^{-8}$, 106, 47, and 97 QTLs were identified for ionomics in flooded environment, unflooded environment, and agronomic traits, respectively. Under flooded environment, 28, 11, 4, 3, 40, 3, 4, 3, and 10 significant QTLs were shown to be associated with Cd, Co, Cu, K, Mo, Ni, Rb, Sr, and Zn, respectively. In unflooded condition, 23, 7, 7, 7, and 3 significant QTLs were detected to be associated with Cd, Fe, Mo, Ni, and Zn, respectively. In addition, 18, 3, 5, 19, 6, 5, 2, 28, 4, and 7 significant QTLs were tightly associated with amylose concentration, flowering time, hull color, hull cover, kernel length, kernel rate, kernel width, lodging, panicle type, and plant height, respectively. Detailed analysis of the QTLs revealed that 49 of the identified QTLs are co-localized or posited near the genes/QTLs with known functions in the related traits, respectively, and 201 QTLs are newly discovered. Moreover,

sequence, expression and homologous analyses of the QTLs suggested that three candidate genes (*OsWRKY102*, *OsWRKY75*, and *OsMan07*) are tightly associated with Cd concentration and *PIP2;6* gene may play role in As regulation in rice. Further, *OsPME6* or nearby gene may regulate plant height and *OsPLL9* or its nearby gene may play a role in flowering time control. Our results showed that each of the four GWAS methods can identify its unique as well as common QTLs and using multiple GWAS methods can complement each other in QTL identification. Using at least one univariate and one multivariate method in GWAS studies is highly recommended for better results. Our comprehensive GWAS analysis of the ionomic and agronomic traits with large scale DNA sequencing data of the USDA mini core collection sets a foundation for further genetic and molecular biology studies on mineral, heavy metal, and agronomic trait regulation.

Methods
Plant materials and Phenotyping
Sixteen grain ionomic traits (As, Ca, Co, Cd, Cu, Fe, K, Mg, Mn, Mo, Ni, P, Rb, S, Sr, and Zn) [17] and thirteen agronomic traits (AMYLOSE, AWNTYPE, DAYS FLOWER, HULLCOLOR, HULLCOVER, KERNELLEN, KERNELWID, KERNELRAT, KERNELWT, LODGING, PANICLETYPE, PLANTHT, and PLANTTYPE) [18] of the mini core collection were the same as reported. Diverse rice accessions were grown over 2 years in Beaumont, Texas under both flooded (anaerobic) and unflooded (aerobic, flush irrigated) irrigation schemes for testing ionomics. The planting, field management, and harvest methods were as reported [55–59]. The correlations of the traits were calculated by Pearson's correlation and visualized with R corrgram package [95]. The details of the samples are listed in Supplementary Table S1.

Genotyping
In order to obtain high-quality sequencing data, the raw reads were downloaded from NCBI (Accession: PRJNA301661). Through comparing the materials between phenotype and available genotyping data, the overlap of 191 accessions were selected for further analysis. The raw data were firstly filtered by NGS QC Toolkit (v2.3.3) with default settings [96]. Then, the high-quality sequences were mapped to Nipponbare MSU7.0 genomic reference (http://rice.plantbiology.msu.edu/index.shtml, Release 7) with bwa program (version 0.7.17) using default parameters [97]. PCR duplicates were marked by Picard (version 2.18). Then, Haploype-Caller of GATK was used to identify SNPs. The raw SNPs were filtered by PLINK software with parameter '--maf 0.05 --geno 0.6 --snps-only'. Genotype imputation

was performed for the remaining 3,259,478 SNPs with Beagle 5.0 [97] for further analysis.

Population structure, genetic analysis, and linkage disequilibrium analysis

The raw SNPs with integrity higher than 0.8 (181,448 SNPs) were extracted for estimating individual ancestries and constructing a phylogenetic tree. A PLINK software tool [98] was used to calculate the potential unlinked SNPs with parameter --indep-pairwise 50 10 0.2. The potentially unlinked SNPs were submitted to ADMIXTURE [99] to assess the population structure with varying K from 2 to 10. Cross-validation error was calculated for each K, and the clustering model with the lowest cross-validation error (K = 4) was selected. Population structure was displayed using online software Pophelper (http://pophelper.com/). Each individual was assigned to one of the four subpopulations based on having ≥70% genetic ancestry derivation, the accessions that had < 70% ancestry from one specific subpopulation were assigned to a fifth group called 'Admix'. The matrix of pairwise genetic distance was used to construct phylogenetic trees using the software SNPhylo [100] with parameters set to 'default'. Principal component analysis (PCA) and kinship matrix (K matrix) were performed with 3,259,478 SNPs using default parameter by GAPIT [101]. The decay distance of LD (linkage disequilibrium) in each subpopulation and in the whole mini-core population were determined by software PopLDdecay [102].

Genome-wide association study (GWAS)

GWAS was performed among 191 rice accessions derived from USDA mini-core collection with 3,259,478 high-quality SNPs. Univariate GWAS methods (GLM and MLM) and multivariate GWAS methods (MLMM and FarmCPU) were employed to evaluate the trait-SNP associations for grain ionomic and agronomic traits using the Genomic Association and Prediction Integrated Tool (GAPIT) [101]. Principal component analysis (PCA) result was used as covariates to correct population structure due to subpopulations in the Mini Core. The genome-wide significant thresholds of the GWAS (p-value = 1.53×10^{-8}) was determined by 0.05/n (n is the number of markers) [93] and a higher significant threshold was set at 3.06×10^{-9} (0.01/n) [103]. The Manhattan and QQ plots for GWAS were visualized using the R package CMplot (https://github.com/YinLi-Lin/R-CMplot). Leading SNPs of each significant SNPs cluster (in 200 kb) were selected to display the location of the QTLs.

Haplotype block estimation

Haplotype blocks containing at least two SNPs were calculated with all imputed SNP using the PLINK software

[98] with the following parameters: ' --blocks no-pheno-req --blocks-max-kb 2000 --blocks-inform-frac 0.95 --blocks-strong-highci 0.98 --blocks-recomb-highci 0.9'. The haplotypic blocks of each significant SNP were determined by Confidence Intervals described by Gabriel [104]. The LD heatmap was visualized by software Haploview [105]. The annotated genes located in each haplotype block were extracted from RAP-DB (https://rapdb.dna.affrc.go.jp/) (Supplementary Table S5).

Gene expression data

The gene expression profile across 15 tissues (Endosperm, Callus, Seed, Radicle, Root, Plumule, Stem, Seedling, Shoot, Sheath, Leaf, Panicle, Spikelet, Lemma, and Stamen) was obtained from CREP (Collection of Rice Expression Profiles): http://crep.ncpgr.cn/crep-cgi/home.pl [106]. Gene expression data of rice plants treated with different cadmium concentration [107, 108] was adopted from TENOR (Transcriptome ENcyclopedia Of Rice): https://tenor.dna.affrc.go.jp/.

Mining causal candidate genes

The QTLs identified by four different GWAS methods provide important clues for understanding the genetic architecture of agronomic and ionomic in rice. To explore candidate genes responsible for each QTLs, we extracted all genes located in the haplotype block of leading SNPs and considering their annotations (Supplementary Table S5). For the leading SNP not posited in the haplotype block, we defined the boundary to within 190 kb of the locus. Besides, functions of homologous gene in *Arabidopsis* and gene expression changes under corresponding stress from published database such as TENOR were also used to narrow down the candidate genes.

Supplementary information

Additional file 1: Supplementary Figure 1. Pearson correlation among ionomics and agronomic traits. (**a**) Ionomics in flooded environment and agronomic traits. (**b**) Ionomics in unflooded environment and agronomic traits. (**c**) Ionomics in flooded environment (with 1 as suffix) and unflooded environment (with 2 as suffix).

Additional file 2: Supplementary Figure 2. (**a**). Genome-wide association analysis for As with GLM, MLM, MLMM, and FarmCPU methods (left) in flooded condition. Quantile-quantile plot of each model (right). Red arrow indicates published gene, black arrow indicates candidate gene. The horizontal dot grey line and green dots indicate the Bonferroni-corrected significance thresholds and SNPs at −log10(P) = 7.81. The horizontal solid grey line and red dots indicate the Bonferroni-corrected significance thresholds and SNPs at −log10(P) = 8.51. (**b**) Genome-wide association analysis for Cd with GLM, MLM, MLMM, and FarmCPU methods (left) in flooded condition. Quantile-quantile plot of each model (right). Black arrows indicate candidate genes. The horizontal dot grey line and green dots indicate the Bonferroni-corrected significance thresholds and SNPs at −log10(P) = 7.81. The horizontal solid

Liu *et al. BMC Plant Biology*   (2020) 20:441

Page 14 of 18

grey line and red dots indicate the Bonferroni-corrected significance thresholds and SNPs at −log10(*P*) = 8.51. (**c**) Genome-wide association analysis for Cd with GLM, MLM, MLMM, and FarmCPU methods (left) in unflooded condition. Quantile-quantile plot of each model (right). Black arrows indicate candidate genes. The horizontal dot grey line and green dots indicate the Bonferroni-corrected significance thresholds and SNPs at −log10(*P*) = 7.81. The horizontal solid grey line and red dots indicate the Bonferroni-corrected significance thresholds and SNPs at −log10(*P*) = 8.51. (**d**) Genome-wide association analysis for Cu with GLM, MLM, MLMM, and FarmCPU methods (left) in flooded condition. Quantile-quantile plot of each model (right). The horizontal dot grey line and green dots indicate the Bonferroni-corrected significance thresholds and SNPs at −log10(*P*) = 7.81. The horizontal solid grey line and red dots indicate the Bonferroni-corrected significance thresholds and SNPs at −log10(*P*) = 8.51. (**e**) Genome-wide association analysis for Co with GLM, MLM, MLMM, and FarmCPU methods (left) in flooded condition. Quantile-quantile plot of each model (right). The horizontal dot grey line and green dots indicate the Bonferroni-corrected significance thresholds and SNPs at −log10(*P*) = 7.81. The horizontal solid grey line and red dots indicate the Bonferroni-corrected significance thresholds and SNPs at −log10(*P*) = 8.51. (**f**) Genome-wide association analysis for Fe with GLM, MLM, MLMM, and FarmCPU methods (left) in unflooded condition. Quantile-quantile plot of each model (right). The horizontal dot grey line and green dots indicate the Bonferroni-corrected significance thresholds and SNPs at −log10(*P*) = 7.81. The horizontal solid grey line and red dots indicate the Bonferroni -corrected significance thresholds and SNPs at −log10(*P*) = 8.51. (**g**) Genome-wide association analysis for K with GLM, MLM, MLMM, and FarmCPU methods (left) in flooded condition. Quantile-quantile plot of each model (right). The horizontal dot grey line and green dots indicate the Bonferroni-corrected significance thresholds and SNPs at −log10(*P*) = 7.81. The horizontal solid grey line and red dots indicate the Bonferroni-corrected significance thresholds and SNPs at −log10(*P*) = 8.51. (**h**) Genome-wide association analysis for Mo with GLM, MLM, MLMM, and FarmCPU methods (left) in flooded condition. Quantile-quantile plot of each model (right). The horizontal dot grey line and green dots indicate the Bonferroni-corrected significance thresholds and SNPs at −log10(*P*) = 7.81. The horizontal solid grey line and red dots indicate the Bonferroni-corrected significance thresholds and SNPs at −log10(*P*) = 7.81. The horizontal solid grey line and red dots indicate the Bonferroni-corrected significance thresholds and SNPs at −log10(*P*) = 8.51. (**i**) Genome-wide association analysis for Mo with GLM, MLM, MLMM, and FarmCPU methods (left) in unflooded condition. Quantile-quantile plot of each model (right). The horizontal dot grey line and green dots indicate the Bonferroni-corrected significance thresholds and SNPs at −log10(*P*) = 7.81. The horizontal solid grey line and red dots indicate the Bonferroni-corrected significance thresholds and SNPs at −log10(*P*) = 8.51. (**j**) Genome-wide association analysis for Ni with GLM, MLM, MLMM, and FarmCPU methods (left) in flooded condition. Quantile-quantile plot of each model (right). The horizontal dot grey line and green dots indicate the Bonferroni-corrected significance thresholds and SNPs at −log10(*P*) = 7.81. The horizontal solid grey line and red dots indicate the Bonferroni-corrected significance thresholds and SNPs at −log10(*P*) = 8.51. (**k**) Genome-wide association analysis for Ni with GLM, MLM, MLMM, and FarmCPU methods (left) in unflooded condition. Quantile-quantile plot of each model (right). The horizontal dot grey line and green dots indicate the Bonferroni-corrected significance thresholds and SNPs at −log10(*P*) = 7.81. The horizontal solid grey line and red dots indicate the Bonferroni -corrected significance thresholds and SNPs at −log10(*P*) = 8.51. (**l**) Genome-wide association analysis for Rb with GLM, MLM, MLMM, and FarmCPU methods (left) in flooded condition. Quantile-quantile plot of each model (right). The horizontal dot grey line and green dots indicate the Bonferroni-corrected significance thresholds and SNPs at −log10(*P*) = 7.81. The horizontal solid grey line and red dots indicate the Bonferroni-corrected significance thresholds and SNPs at −log10(*P*) = 8.51. (**m**) Genome-wide association analysis for Sr with GLM, MLM, MLMM, and FarmCPU methods (left) in flooded condition. Quantile-quantile plot of each model (right). The horizontal dot grey line and green dots indicate the Bonferroni-corrected significance thresholds and SNPs at −log10(*P*) = 7.81. The horizontal solid grey line and red dots indicate the Bonferroni-corrected significance thresholds and SNPs at −log10(*P*) = 8.51. (**n**) Genome-wide association analysis for Zn with GLM, MLM, MLMM, and FarmCPU methods (left) in flooded condition.

Quantile-quantile plot of each model (right). The horizontal dot grey line and green dots indicate the Bonferroni-corrected significance thresholds and SNPs at −log10(*P*) = 7.81. The horizontal solid grey line and red dots indicate the Bonferroni-corrected significance thresholds and SNPs at −log10(*P*) = 8.51. (**o**) Genome-wide association analysis for Zn with GLM, MLM, MLMM, and FarmCPU methods (left) in unflooded condition. Quantile-quantile plot of each model (right). The horizontal dot grey line and green dots indicate the Bonferroni-corrected significance thresholds and SNPs at −log10(*P*) = 7.81. The horizontal solid grey line and red dots indicate the Bonferroni-corrected significance thresholds and SNPs at −log10(*P*) = 8.51.

Additional file 3: Supplementary Figure 3. (a) Genome-wide association analysis for AMYLOSE with GLM, MLM, MLMM, and FarmCPU methods (left). Quantile-quantile plot of each model (right). Red arrows indicate published genes. The horizontal dot grey line and green dots indicate the Bonferroni-corrected significance thresholds and SNPs at −log10(*P*) = 7.81. The horizontal solid grey line and red dots indicate the Bonferroni-corrected significance thresholds and SNPs at −log10(*P*) = 8.51. **(b)** Genome-wide association analysis for AWNTYPE with GLM, MLM, MLMM, and FarmCPU methods (left). Quantile-quantile plot of each model (right). The horizontal dot grey line and green dots indicate the Bonferroni-corrected significance thresholds and SNPs at −log10(*P*) = 7.81. The horizontal solid grey line and red dots indicate the Bonferroni-corrected significance thresholds and SNPs at −log10(*P*) = 8.51. **(c)** Genome-wide association analysis for DAYSFLOWER with GLM, MLM, MLMM, and FarmCPU methods (left). Quantile-quantile plot of each model (right). Black arrow indicates candidate gene. The horizontal dot grey line and green dots indicate the Bonferroni-corrected significance thresholds and SNPs at −log10(*P*) = 7.81. The horizontal solid grey line and red dots indicate the Bonferroni-corrected significance thresholds and SNPs at −log10(*P*) = 8.51. **(d)** Genome-wide association analysis for HULLCOLOR with GLM, MLM, MLMM, and FarmCPU methods (left). Quantile-quantile plot of each model (right). The horizontal dot grey line and green dots indicate the Bonferroni-corrected significance thresholds and SNPs at −log10(*P*) = 7.81. The horizontal solid grey line and red dots indicate the Bonferroni-corrected significance thresholds and SNPs at −log10(*P*) = 8.51. **(e)** Genome-wide association analysis for HULLCOVER with GLM, MLM, MLMM, and FarmCPU methods (left). Quantile-quantile plot of each model (right). The horizontal dot grey line and green dots indicate the Bonferroni-corrected significance thresholds and SNPs at −log10(*P*) = 7.81. The horizontal solid grey line and red dots indicate the Bonferroni-corrected significance thresholds and SNPs at −log10(*P*) = 8.51. **(f)** Genome-wide association analysis for KERNELLEN with GLM, MLM, MLMM, and FarmCPU methods (left). Quantile-quantile plot of each model (right). Red arrow indicates published gene. The horizontal dot grey line and green dots indicate the Bonferroni-corrected significance thresholds and SNPs at −log10(*P*) = 7.81. The horizontal solid grey line and red dots indicate the Bonferroni-corrected significance thresholds and SNPs at −log10(*P*) = 8.51. **(g)** Genome-wide association analysis for KERNELWID with GLM, MLM, MLMM, and FarmCPU methods (left). Quantile-quantile plot of each model (right). Red arrow indicates published gene. The horizontal dot grey line and green dots indicate the Bonferroni-corrected significance thresholds and SNPs at −log10(*P*) = 7.81. The horizontal solid grey line and red dots indicate the Bonferroni-corrected significance thresholds and SNPs at −log10(*P*) = 8.51. **(h)** Genome-wide association analysis for KERNELRAT with GLM, MLM, MLMM, and FarmCPU methods (left). Quantile-quantile plot of each model (right). Red arrow indicates published gene. The horizontal dot grey line and green dots indicate the Bonferroni-corrected significance thresholds and SNPs at −log10(*P*) = 7.81. The horizontal solid grey line and red dots indicate the Bonferroni-corrected significance thresholds and SNPs at −log10(*P*) = 8.51. **(i)** Genome-wide association analysis for KERNELWTB with GLM, MLM, MLMM, and FarmCPU methods (left). Quantile-quantile plot of each model (right). The horizontal dot grey line and green dots indicate the Bonferroni-corrected significance thresholds and SNPs at −log10(*P*) = 7.81. The horizontal solid grey line and red dots indicate the Bonferroni-corrected significance thresholds and SNPs at −log10(*P*) = 8.51. **(j)** Genome-wide association analysis for LODGING with GLM, MLM, MLMM, and FarmCPU methods (left). Quantile-quantile plot of each model (right). Black arrow indicates candidate gene. The

horizontal dot grey line and green dots indicate the Bonferroni-corrected significance thresholds and SNPs at $-\log10(P) = 7.81$. The horizontal solid grey line and red dots indicate the Bonferroni-corrected significance thresholds and SNPs at $-\log10(P) = 8.51$. **(k)** Genome-wide association analysis for PLANTHT with GLM, MLM, MLMM, and FarmCPU methods (left). Quantile-quantile plot of each model (right). Black arrow indicates candidate gene. The horizontal dot grey line and green dots indicate the Bonferroni-corrected significance thresholds and SNPs at $-\log10(P) = 7.81$. The horizontal solid grey line and red dots indicate the Bonferroni-corrected significance thresholds and SNPs at $-\log10(P) = 8.51$.

Additional file 4: Supplementary Figure 4. The number of QTLs identified by four different methods. (a) Agronomic traits; (b) Ionomic traits in flooded environment; (c) Ionomic traits in unflooded environment. Different color represents different methods; blue, GLM; yellow, MLM; green, MLMM; red, FarmCPU. Venn diagrams were visualized by online software Venny 2.1: http://bioinfogp.cnb.csic.es/tools/venny/.

Additional file 5: Supplementary Figure 5. Heat map of the putative candidate gene expression patterns (log2-transformed) in 15 tissues in Minghui 63, Shanyou 63, and Zhenshan 97 varieties. Darkblue indicates high expression, white indicates low expression, and grey indicates NA.

Additional file 6: Supplementary Figure 6. (**a**) The expression pattern of a candidate gene (OsWRK102) in different tissues from public data. (**b**) The candidate gene expression profile in different treatment (High Cd [(50 μM $CdSO_4$)]), Low Cd [(1 μM $CdSO_4$)], and Very Low Cd [(0.2 μM $CdSO_4$)] in Shoot and Root.

Additional file 7: Supplementary Figure 7. Heat map of the genes' (located in the blocks) expression patterns under cadmium stress. Red indicates high expression, and blue indicates low expression. Genes with red color are candidate genes in the current study.

Additional file 8: Supplementary Table 1. The sample list and structure information of *O. sativa* accessions.

Additional file 9: Supplementary Table 2. The QTLs associated with agronomic traits, ionomic in flooded and unflooded conditions with four different methods.

Additional file 10: Supplementary Table 3. List of gene orthologous with *Arabidopsis thaliana*.

Additional file 11: Supplementary Table 4. The distribution of LD blocks in the 12 chromosomes of rice.

Additional file 12: Supplementary Table 5. List of candidate genes located in the LD blocks for agronomic, ionomic in flooded, and ionomic in unflooded traits.

Abbreviations
ADM: Admixture; AMYLOSE: Amylose; As: Arsenic; ATSDR: Agency for Toxic Substances and Disease Registry; AWNTYPE: Awn type; Cd: Cadmium; CREP: Collection of Rice Expression Profiles; DAYSFLOWER: Flowering time; FarmCPU: Fixed and random model Circulating Probability Unification; FASTmrEMMA: Fast multi-locus random-SNP-effect efficient mixed model analysis; FASTmrMLM: Fast mrMLM; FEM: Fixed Effect Model; GAPIT: Genomic Association and Prediction Integrated Tool; GPWAS: Genome-Phenome Wide Association Study; GWAS: Genome-wide-association-study; HULLCOLOR: Hull color; HULLCOVER: Hull cover; ISIS EM-BLASSO: Iterative modified-sure independence screening expectation-maximization-Bayesian least absolute shrinkage and selection operator; KERNELLEN: Kernel length; KERNELRA T: Kernel rate; KERNELWID: Kernel width; KERNELWT: Kernel weight; LD: Linkage disequilibrium; LODGING: Lodging; ML: Maximum likelihood; MLMM: Multi-locus mixed-model; Mn: Manganese; mrMLM: Multi-locus random-SNP-effect MLM; Ni: Nickel; NSGC: National Small Grains Collection; PANICLETYPE: Panicle type; PCA: Principal component analysis; PCs: Principal components; pKWmEB: Integration of Kruskal-Wallis test with empirical Bayes; PLANTHT: Plant height; PLANTTYPE: Plant type; pLARmEB: Polygenic-background-control-based least angle regression plus empirical Bayes; QTL: Quantitative Trait Locus; SD: Short-day; TENOR: Transcriptome ENcyclopedia Of Rice

Author details
[1]Department of Biochemistry, Molecular Biology, Entomology and Plant Pathology, Mississippi State University, Starkville, MS 39762, USA. [2]State Key Laboratory of Hybrid Rice, Key Laboratory for Research and Utilization of Heterosis in Indica Rice, Ministry of Agriculture, College of Life Sciences, Wuhan University, Wuhan, China. [3]Dale Bumpers National Rice Research Center, USDA ARS, Stuttgart, AR 72160, USA. [4]Experimental Seafood Processing Laboratory, Coastal and Research Extension Center, Mississippi State University, Pascagoula, MS 39567, USA.

References
1. Izawa T, Shimamoto K. Becoming a model plant: the importance of rice to plant science. Trends Plant Sci. 1996;1:95–9.
2. Fasani E, Manara A, Martini F, Furini A, DalCorso G. The potential of genetic engineering of plants for the remediation of soils contaminated with heavy metals. Plant Cell Environ. 2018;41:1201–32.
3. Panaullah GM, Alam T, Hossain MB, Loeppert RH, Lauren JG, Meisner CA, et al. Arsenic toxicity to rice (Oryza sativa L.) in Bangladesh. Plant Soil. 2009; 317:31–9.
4. Rizwan M, Ali S, Adrees M, Rizvi H, Zia-ur-Rehman M, Hannan F, et al. Cadmium stress in rice: toxic effects, tolerance mechanisms, and management: a critical review. Environ Sci Pollut Res. 2016;23:17859–79. https://doi.org/10.1007/s11356-016-6436-4.
5. Williams PN, Villada A, Deacon C, Raab A, Figuerola J, Green AJ, et al. Greatly enhanced arsenic shoot assimilation in rice leads to elevated grain levels compared to wheat and barley. Environ Sci Technol. 2007;41:6854–9.
6. Gilbert-Diamond D, Cottingham KL, Gruber JF, Punshon T, Sayarath V, Gandolfi AJ, et al. Rice consumption contributes to arsenic exposure in US women. Proc Natl Acad Sci U S A. 2011;108:20656–60.
7. Azizur Rahman M, Hasegawa H, Mahfuzur Rahman M, Mazid Miah MA, Tasmin A. Arsenic accumulation in rice (Oryza sativa L.): human exposure through food chain. Ecotoxicol Environ Saf. 2008;69:317–24.
8. Tuli R, Chakrabarty D, Trivedi PK, Tripathi RD. Recent advances in arsenic accumulation and metabolism in rice. Mol Breed. 2010;26:307–23.
9. Meharg AA, Norton G, Deacon C, Williams P, Adomako EE, Price A, et al. Variation in rice cadmium related to human exposure. Environ Sci Technol. 2013;47:5613–8.

10. Liu J, Li K, Xu J, Liang J, Lu X, Yang J, et al. Interaction of cd and five mineral nutrients for uptake and accumulation in different rice cultivars and genotypes. F Crop Res. 2003;83:271–81.
11. Bashir K, Takahashi R, Akhtar S, Ishimaru Y, Nakanishi H, Nishizawa NK. The knockdown of OsVIT2 and MIT affects iron localization in rice seed. Rice. 2013;6:1–6.
12. Kumar J, Jain S, Jain RK. Linkage mapping for grain iron and zinc content in F2population derived from the cross between PAU201 and Palman 579 in rice (*Oryza sativa* L.). Cereal Res Commun. 2014;42:389–400.
13. Lu K, Li L, Zheng X, Zhang Z, Mou T, Hu Z. Quantitative trait loci controlling cu, Ca, Zn, Mn and Fe content in rice grains. J Genet. 2008;87:305–10.
14. Luo JS, Huang J, Zeng DL, Peng JS, Bin ZG, Ma HL, et al. A defensin-like protein drives cadmium efflux and allocation in rice. Nat Commun. 2018;9. https://doi.org/10.1038/s41467-018-03088-0.
15. Yan W, Rutger JN, Bryant RJ, Bockelman HE, Fjellstrom RG, Chen MH, et al. Development and evaluation of a core subset of the USDA rice germplasm collection. Crop Sci. 2007;47:869–78.
16. Agrama HA, Yan WG, Lee F, Robert F, Chen MH, Jia M, et al. Genetic assessment of a mini-core subset developed from the USDA rice genebank. Crop Sci. 2009;49:1336–46.
17. Pinson SRM, Tarpley L, Yan W, Yeater K, Lahner B, Yakubova E, et al. Worldwide genetic diversity for mineral element concentrations in Rice grain. Crop Sci. 2015;55:294.
18. Yan W. Genetic characterization of global rice germplasm for sustainable agriculture. Food Prod Approach Challenges Tasks. 2012:243–70.
19. Abdurakhmonov IY, Abdukarimov A. Application of association mapping to understanding the genetic diversity of plant germplasm resources. Int J Plant Genomics. 2008;2008:574927. https://doi.org/10.1155/2008/574927.
20. Nordborg M, Weigel D. Next-generation genetics in plants. Nature. 2008; 456:720–3.
21. Cao J, Schneeberger K, Ossowski S, Günther T, Bender S, Fitz J, et al. Whole-genome sequencing of multiple Arabidopsis thaliana populations. Nat Genet. 2011;43:956–65.
22. Tian F, Bradbury PJ, Brown PJ, Hung H, Sun Q, Flint-Garcia S, et al. Genome-wide association study of leaf architecture in the maize nested association mapping population. Nat Genet. 2011;43:159–62.
23. Li H, Peng Z, Yang X, Wang W, Fu J, Wang J, et al. Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels. Nat Genet. 2013;45:43–50.
24. Cai S, Yu G, Chen X, Huang Y, Jiang X, Zhang G, et al. Grain protein content variation and its association analysis in barley. BMC Plant Biol. 2013;13:35. https://doi.org/10.1186/1471-2229-13-35.
25. Jiang Y, Schmidt RH, Zhao Y, Reif JC. Quantitative genetic framework highlights the role of epistatic effects for grain-yield heterosis in bread wheat. Nat Genet. 2017;49:1741–6.
26. Yano K, Yamamoto E, Aya K, Takeuchi H, Lo PC, Hu L, et al. Genome-wide association study using whole-genome sequencing rapidly identifies new genes influencing agronomic traits in rice. Nat Genet. 2016;48:927–34.
27. Huang X, Wei X, Sang T, Zhao Q, Feng Q, Zhao Y, et al. Genome-wide asociation studies of 14 agronomic traits in rice landraces. Nat Genet. 2010;42:961–7.
28. Zhou Z, Jiang Y, Wang Z, Gou Z, Lyu J, Li W, et al. Resequencing 302 wild and cultivated accessions identifies genes related to domestication and improvement in soybean. Nat Biotechnol. 2015;33:408–14.
29. Du X, Huang G, He S, Yang Z, Sun G, Ma X, et al. Resequencing of 243 diploid cotton accessions based on an updated a genome identifies the genetic basis of key agronomic traits. Nat Genet. 2018;50:796–802. https://doi.org/10.1038/s41588-018-0116-x.
30. Segura V, Vilhjálmsson BJ, Platt A, Korte A, Seren Ü, Long Q, et al. An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. Nat Genet. 2012;44:825–30.
31. Liu X, Huang M, Fan B, Buckler ES, Zhang Z. Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. PLoS Genet. 2016;12:e1005767.
32. Wang S-B, Feng J-Y, Ren W-L, Huang B, Zhou L, Wen Y-J, et al. Improving power and accuracy of genome-wide association studies via a multi-locus mixed linear model methodology. Sci Rep. 2016;6:19444. https://doi.org/10.1038/srep19444.
33. Tamba CL, Zhang Y. A fast mrMLM algorithm for multi-locus genome-wide association studies; 2018.
34. Wen YJ, Zhang H, Ni YL, Huang B, Zhang J, Feng JY, et al. Methodological implementation of mixed linear models in multi-locus genome-wide association studies. Brief Bioinform. 2018;19:700–12.
35. Zhang J, Feng JY, Ni YL, Wen YJ, Niu Y, Tamba CL, et al. PLARmEB: integration of least angle regression with empirical Bayes for multilocus genome-wide association studies. Heredity (Edinb). 2017;118:517–24.
36. Ren WL, Wen YJ, Dunwell JM, Zhang YM. PKWmEB: integration of Kruskal-Wallis test with empirical Bayes under polygenic background control for multi-locus genome-wide association study. Heredity (Edinb). 2018;120:208–18.
37. Tamba CL, Ni YL, Zhang YM. Iterative sure independence screening EM-Bayesian LASSO algorithm for multi-locus genome-wide association studies. PLoS Comput Biol. 2017;13:1–20.
38. Liang Z, Qiu Y, Schnable JC. Genome-phenome wide association in maize and Arabidopsis identifies a common molecular and evolutionary signature. Mol Plant. 2020;13:907–22.
39. Liu R, Gong J, Xiao X, Zhang Z, Li J, Liu A, et al. Gwas analysis and qtl identification of fiber quality traits and yield components in upland cotton using enriched high-density snp markers. Front Plant Sci. 2018; 9(September):1–15.
40. Naveed SA, Zhang F, Zhang J, Zheng TQ, Meng LJ, Pang YL, et al. Identification of QTN and candidate genes for salinity tolerance at the germination and seedling stages in Rice by genome-wide association analyses. Sci Rep. 2018;8:1–11. https://doi.org/10.1038/s41598-018-24946-3.
41. Wu Y, Sun Y, Sun S, Li G, Wang J, Wang B, et al. Aneuploidization under segmental allotetraploidy in rice and its phenotypic manifestation. Theor Appl Genet. 2018;131:1273–85. https://doi.org/10.1007/s00122-018-3077-7.
42. Jaiswal V, Bandyopadhyay T, Gahlaut V, Gupta S, Dhaka A, Ramchiary N, et al. Genome-wide association study (GWAS) delineates genomic loci for ten nutritional elements in foxtail millet (Setaria italica L.). J Cereal Sci. 2019; 85:48–55. https://doi.org/10.1016/j.jcs.2018.11.006.
43. Zhao X, Teng W, Li Y, Liu D, Cao G, Li D, et al. Loci and candidate genes conferring resistance to soybean cyst nematode HG type 2.5.7. BMC Genomics. 2017;18:1–10.
44. Ziegler G, Nelson R, Granada S, Krishnan HB, Gillman JD, Baxter I. Genomewide association study of ionomic traits on diverse soybean populations from germplasm collections. Plant Direct. 2018;2:e00033.
45. Sanchez DL, Liu S, Ibrahim R, Blanco M, Lübberstedt T. Genome-wide association studies of doubled haploid exotic introgression lines for root system architecture traits in maize (*Zea mays* L.). Plant Sci. 2018;268:30–8 https://doi.org/10.1016/j.plantsci.2017.12.004.
46. Hu S, Sanchez DL, Wang C, Lipka AE, Yin Y, Gardner CAC, et al. Brassinosteroid and gibberellin control of seedling traits in maize (*Zea mays* L.). Plant Sci. 2017;263(July):132–41. https://doi.org/10.1016/j.plantsci.2017.07.011.
47. Kumar J, Saripalli G, Gahlaut V, Goel N, Meher PK, Mishra KK, et al. Genetics of Fe, Zn, β-carotene, GPC and yield traits in bread wheat (Triticum aestivum L.) using multi-locus and multi-traits GWAS. Euphytica. 2018;214:1–17. https://doi.org/10.1007/s10681-018-2284-2.
48. Jaiswal V, Gahlaut V, Meher PK, Mir RR, Jaiswal JP, Rao AR, et al. Genome wide single locus single trait, multi-locus and multi-trait association mapping for some important agronomic traits in common wheat (T. aestivum L.). PLoS One. 2016;11:1–25.
49. Yan W, Jackson A, Jia M, Zhou W, Xiong H, Bryant R. Association mapping of four important traits using the USDA rice mini-core collection. Rice Germplasm Genet Improv. 2014:105–42.
50. Cui Y, Zhang F, Zhou Y. The application of multi-locus GWAS for the detection of salt-tolerance loci in rice. Front Plant Sci. 2018;9:1464.
51. Sun D, Cen H, Weng H, Wan L, Abdalla A, El-Manawy AI, et al. Using hyperspectral analysis as a potential high throughput phenotyping tool in GWAS for protein content of rice quality. Plant Methods. 2019;15:54. https://doi.org/10.1186/s13007-019-0432-x.
52. Wang H, Vieira FG, Crawford JE, Chu C, Nielsen R. Asian wild rice is a hybrid swarm with extensive gene flow and feralization from domesticated rice. Genome Res. 2017;27:1029–38.
53. Browning BL, Zhou Y, Browning SR. A one-penny imputed genome from next-generation reference panels. Am J Hum Genet. 2018;103:338–48. https://doi.org/10.1016/j.ajhg.2018.07.015.
54. Liu H, Zhan J, Li J, Lu X, Liu J, Wang Y, et al. Genome-wide association study (GWAS) for mesocotyl elongation in rice (*Oryza sativa* L.) under multiple

Liu *et al. BMC Plant Biology*   (2020) 20:441

Page 17 of 18

culture conditions. Genes (Basel). 2020;11:49. https://doi.org/10.3390/genes11010049.

55. Yan WG, Rutger JN, Bockelman HE, Tai TH. Agronomic evaluation and seed stock establishment of the USDA rice core collection. BR Wells Rice Res Stud Stuttgart Univ Arkansas Agri Exp Sta Res Ser. 2005. p. 63–8.

56. Yan WG, Rutger JN, Bockelman HE, Tai TH. Evaluation of kernel characteristics of the USDA rice core collection. BR Wells Rice Res Stud Stuttgart Univ Arkansas, Agric Exp Stn Res Ser Agri Exp Sta Res Ser. 2005. p. 69–74.

57. Li X, Yan W, Agrama H, Hu B, Jia L, Jia M, et al. Genotypic and phenotypic characterization of genetic differentiation and diversity in the USDA rice mini-core collection. Genetica. 2010;138:1221–30.

58. Li X, Yan W, Agrama H, Jia L, Shen X, Jackson A, et al. Mapping QTLs for improving grain yield using the USDA rice mini-core collection. Planta. 2011;234:347–61.

59. Li X, Yan W, Agrama H, Jia L, Jackson A, Moldenhauer K, et al. Unraveling the complex trait of harvest index with association mapping in rice (Oryza sativa L.). PLoS One. 2012;7:1–10.

60. Satoh-Nagasawa N, Mori M, Nakazawa N, Kawamoto T, Nagato Y, Sakurai K, et al. Mutations in rice (oryza sativa) heavy metal ATPase 2 (OsHMA2) restrict the translocation of zinc and cadmium. Plant Cell Physiol. 2012;53: 213–24.

61. Jin S, Cheng Y, Guan Q, Liu D, Takano T, Liu S. A metallothionein-like protein of rice (rgMT) functions in E. coli and its gene expression is induced by abiotic stresses. Biotechnol Lett. 2006;28:1749–53.

62. Zhang M, Pinson SRM, Tarpley L, Huang XY, Lahner B, Yakubova E, et al. Mapping and validation of quantitative trait loci associated with concentrations of 16 elements in unmilled rice grain. Theor Appl Genet. 2014;127:137–65.

63. Du J, Zeng D, Wang B, Qian Q, Zheng S, Ling HQ. Environmental effects on mineral accumulation in rice grains and identification of ecological specific QTLs. Environ Geochem Health. 2013;35:161–70.

64. Wang Y, Wan L, Zhang L, Zhang Z, Zhang H, Quan R, et al. An ethylene response factor OsWR1 responsive to drought stress transcriptionally activates wax synthesis related genes and increases wax production in rice. Plant Mol Biol. 2012;78:275–88.

65. Gao Z, Zeng D, Cheng F, Tian Z, Guo L, Su Y, et al. ALK, the key gene for gelatinization temperature, is a modifier gene for gel consistency in Rice. J Integr Plant Biol. 2011;53:756–65.

66. Fan C, Xing Y, Mao H, Lu T, Han B, Xu C, et al. GS3, a major QTL for grain length and weight and minor QTL for grain width and thickness in rice, encodes a putative transmembrane protein. Theor Appl Genet. 2006;112: 1164–71.

67. Li Y, Fan C, Xing Y, Jiang Y, Luo L, Sun L, et al. Natural variation in GS5 plays an important role in regulating grain size and yield in rice. Nat Genet. 2011; 43:1266–9.

68. Liu J, Chen J, Zheng X, Wu F, Lin Q, Heng Y, et al. GW5 acts in the brassinosteroid signalling pathway to regulate grain width and weight in rice. Nat Plants. 2017;3:17043.

69. Wang S, Wu K, Yuan Q, Liu X, Liu Z, Lin X, et al. Control of grain size, shape and quality by OsSPL16 in rice. Nat Genet. 2012;44:950–4. https://doi.org/10.1038/ng.2327.

70. Wang Y, Xiong G, Hu J, Jiang L, Yu H, Xu J, et al. Copy number variation at the GL7 locus contributes to grain size diversity in rice. Nat Genet. 2015;47:944–8.

71. Ishimaru K, Hirotsu N, Madoka Y, Murakami N, Hara N, Onodera H, et al. Loss of function of the IAA-glucose hydrolase gene TGW6 enhances rice grain weight and increases yield. Nat Genet. 2013;45:707–11.

72. Fujita D, Trijatmiko KR, Tagle AG, Sapasap MV, Koide Y, Sasaki K, et al. NAL1 allele from a rice landrace greatly increases yield in modern indica cultivars. Proc Natl Acad Sci U S A. 2013;110:20431–6.

73. Weng J, Gu S, Wan X, Gao H, Guo T, Su N, et al. Isolation and initial characterization of GW5, a major QTL associated with rice grain width and weight. Cell Res. 2008;18:1199–209.

74. Rebetzke GJ, Richards RA. Gibberellic acid-sensitive dwarfing genes reduce plant height to increase kernel number and grain yield of wheat. Aust J Agric Res. 2000;51:235–45.

75. Tamaoki D, Karahara I, Nishiuchi T, De Oliveira S, Schreiber L, Wakasugi T, et al. Transcriptome profiling in Arabidopsis inflorescence stems grown under hypergravity in terms of cell walls and plant hormones. Adv Sp Res. 2009;44:245–53. https://doi.org/10.1016/j.asr.2009.03.016.

76. Gao H, Zheng XM, Fei G, Chen J, Jin M, Ren Y, et al. Ehd4 encodes a novel and Oryza-genus-specific regulator of photoperiodic flowering in Rice. PLoS Genet. 2013;9:e1003281.

77. Xue W, Xing Y, Weng X, Zhao Y, Tang W, Wang L, et al. Natural variation in Ghd7 is an important regulator of heading date and yield potential in rice. Nat Genet. 2008;40:761–7.

78. Yano M, Katayose Y, Ashikari M, Yamanouchi U, Monna L, Fuse T, et al. Hd1, a major photoperiod sensitivity quantitative trait locus in rice, is closely related to the Arabidopsis flowering time gene CONSTANS. Plant Cell. 2000; 12:2473–83.

79. Takahashi Y, Shomura A, Sasaki T, Yano M. Hd6, a rice quantitative trait locus involved in photoperiod sensitivity, encodes the α subunit of protein kinase CK2. Proc Natl Acad Sci U S A. 2001;98:7922–7.

80. Zheng Y, Yan J, Wang S, Xu M, Huang K, Chen G, et al. Genome-wide identification of the pectate lyase-like (PLL) gene family and functional analysis of two PLL genes in rice. Mol Gen Genomics. 2018;293:1317–31. https://doi.org/10.1007/s00438-018-1466-x.

81. Zhong M, Li S, Huang F, Qiu J, Zhang J, Sheng Z, et al. The phosphoproteomic response of rice seedlings to cadmium stress. Int J Mol Sci. 2017;18:2055. https://doi.org/10.3390/ijms18102055.

82. Sheng Y, Yan X, Huang Y, Han Y, Zhang C, Ren Y, et al. The WRKY transcription factor, WRKY13, activates PDR8 expression to positively regulate cadmium tolerance in Arabidopsis. Plant Cell Environ. 2019;42:891–903.

83. Ishikawa S, Ae N, Yano M. Chromosomal regions with quantitative trait loci controlling cadmium concentration in brown rice (Oryza sativa). New Phytol. 2005;168:345–50.

84. Chen J, Yang L, Gu J, Bai X, Ren Y, Fan T, et al. MAN3 gene regulates cadmium tolerance through the glutathione-dependent pathway in Arabidopsis thaliana. New Phytol. 2015;205:570–82.

85. Mosa KA, Kumar K, Chhikara S, Mcdermott J, Liu Z, Musante C, et al. Members of rice plasma membrane intrinsic proteins subfamily are involved in arsenite permeability and tolerance in plants. Transgenic Res. 2012;21: 1265–77.

86. Wang Z-Y, Zheng F-Q, Shen G-Z, Gao J-P, Snustad DP, Li M-G, et al. The amylose content in rice endosperm is related to the post-transcriptional regulation of the waxy gene. Plant J. 1995;7:613–22.

87. Wang X, Pang Y, Zhang J, Wu Z, Chen K, Ali J, et al. Genome-wide and gene-based association mapping for rice eating and cooking characteristics and protein content. Sci Rep. 2017;7:1–10.

88. Zheng K, Tian H, Hu Q, Guo H, Yang L, Cai L, et al. Ectopic expression of R3 MYB transcription factor gene OsTCL1 in Arabidopsis, but not rice, affects trichome and root hair formation. Sci Rep. 2016;6:19254.

89. Li J, Yuan Y, Lu Z, Yang L, Gao R, Lu J, et al. Glabrous Rice 1, encoding a homeodomain protein, regulates trichome development in rice. Rice. 2012;5:32.

90. Jiao Y, Wang Y, Xue D, Wang J, Yan M, Liu G, et al. Regulation of OsSPL14 by OsmiR156 defines ideal plant architecture in rice. Nat Genet. 2010;42: 541–4.

91. Li F, Xie G, Huang J, Zhang R, Li Y, Zhang M, et al. OsCESA9 conserved-site mutation leads to largely enhanced plant lodging resistance and biomass enzymatic saccharification by reducing cellulose DP and crystallinity in rice. Plant Biotechnol J. 2017;15:1093–104.

92. Hu P, Huang J, Ouyang Y, Wu L, Song J, Wang S, et al. Water management affects arsenic and cadmium accumulation in different rice cultivars. Environ Geochem Health. 2013;35:767–78.

93. Yang M, Lu K, Zhao FJ, Xie W, Ramakrishna P, Wang G, et al. Genome-wide association studies reveal the genetic basis of ionomic variation in rice. Plant Cell. 2018;30:2720–40.

94. Services USD of H and H. Agency for Toxic Substances and Disease Registry-ATSDR. 1999.

95. Friendly M. Corrgrams: exploratory displays for correlatigon matrices. Am Stat. 2002;56:316–24.

96. Patel RK, Jain M. NGS QC toolkit: a toolkit for quality control of next generation sequencing data. PLoS One. 2012;7:e30619.

97. Li H, Durbin R. Fast and accurate short read alignment with burrows-wheeler transform. Bioinformatics. 2009;25:1754–60.

98. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81:559–75.

99. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 2009;19:1655–64.

100. Lee TH, Guo H, Wang X, Kim C, Paterson AH. SNPhylo: a pipeline to construct a phylogenetic tree from huge SNP data. BMC Genomics. 2014;15:1–6.

101. Lipka AE, Tian F, Wang Q, Peiffer J, Li M, Bradbury PJ, et al. GAPIT: genome association and prediction integrated tool. Bioinformatics. 2012;28:2397–9.

102. Zhang C, Dong SS, Xu JY, He WM, Yang TL. PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. Bioinformatics. 2019;35:1786–8.

103. Zhao Y, Zhao W, Jiang C, Wang X, Xiong H, Todorovska EG, et al. Genetic architecture and candidate genes for deep-sowing tolerance in rice revealed by non-syn GWAS. Front Plant Sci. 2018;9(March):1–14.

104. Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, et al. The structure of haplotype blocks in the human genome. Science (80-). 2002;296:2225–9.

105. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics. 2005;21:263–5.

106. Wang L, Xie W, Chen Y, Tang W, Yang J, Ye R, et al. A dynamic gene expression atlas covering the entire life cycle of rice. Plant J. 2010;61:752–66.

107. Oono Y, Yazawa T, Kawahara Y, Kanamori H, Kobayashi F, Sasaki H, et al. Genome-wide transcriptome analysis reveals that cadmium stress signaling controls the expression of genes in drought stress signal pathways in rice. PLoS One. 2014;9:e96946.

108. Kawahara Y, Oono Y, Wakimoto H, Ogata J, Kanamori H, Sasaki H, et al. TENOR: database for comprehensive mRNA-Seq experiments in rice. Plant Cell Physiol. 2016;57:e7.

Publisher's Note