

RESEARCH ARTICLE

Open Access



Mining candidate gene for rice aluminum tolerance through genome wide association study and transcriptomic analysis

Peng Zhang^{*†} , Kaizhen Zhong[†], Zhengzheng Zhong and Hanhua Tong

Abstract

Background: The genetic mechanism of aluminum (Al) tolerance in rice is great complicated. Uncovering genetic mechanism of Al tolerance in rice is the premise for Al tolerance improvement. Mining elite genes within rice landrace is of importance for improvement of Al tolerance in rice.

Results: Genome-wide association study (GWAS) performed in EMMAX for rice Al tolerance was carried out using 150 varieties of Ting's core collection constructed from 2262 Ting's collections with more than 3.8 million SNPs. Within Ting's core collection of clear population structure and kinship relatedness as well as high rate of linkage disequilibrium (LD) decay, 17 genes relating to rice Al tolerance including cloned genes like *NRAT1*, *ART1* and *STAR1* were identified in this study. Moreover, 13 new candidate regions with high LD and 69 new candidate genes were detected. Furthermore, 20 of 69 new candidate genes were detected with significant difference between Al treatment and without Al toxicity by transcriptome sequencing. Interestingly, both qRT-PCR and sequence analysis in CDS region demonstrated that the candidate genes in present study might play important roles in rice Al tolerance.

Conclusions: The present study provided important information for further using these elite genes existing in Ting's core collection for improvement of rice Al tolerance.

Keywords: Rice aluminum tolerance, Candidate gene, Genome wide association study, Transcriptomic analysis

Background

Over 50% of the world's arable land is acidic and about 13% of global rice were produced on acidic soils [1–3]. Rice (*Oryza sativa* L.) is one of the most important crops in the world, nearly half of the world's population feed on rice. Aluminum (Al) which is the most abundant metal in the earth's crust could be solubilized into trivalent Al (Al³⁺) in acidic soils when pH value is below 5.0. Root growth might be inhibited and rice yield could be reduced significantly when Al³⁺ is at high concentration [4, 5] while beneficial at low level [6]. Therefore, improving Al tolerance is significantly useful for rice production.

Uncovering genetic mechanism of Al tolerance in rice is the premise for Al tolerance improvement. Plant physiologists and breeders have been focusing on revealing genetic mechanism of Al tolerance in rice [7–17]. Two detoxification mechanisms under Al³⁺ toxicity were illuminated in above studies, i.e. called exclusion of Al³⁺ producing in root cells and excreting chelating chemicals [18] and accumulation of Al³⁺ called internal detoxification [10, 14, 19, 20].

Many quantitative trait loci (QTL) for rice Al tolerance had been reported in previous researches by using different mapping populations [9, 17, 21–25]. Furthermore, several genes related to Al tolerance in rice were cloned, i.e. *ART1* [14], *STAR1* [10], *STAR2* [10], *Nrat1* [13], *OsFRDL4* [18], *OsALS1* [26], *OsMGT1* [27], *ASR5* [20] and *ART2* [28]. We summarized above QTL/genes and found that these QTL/genes were not identical for the greater part due to different mapping population or Al

* Correspondence: zhangpeng03@caas.cn.cn

[†]Peng Zhang and Kaizhen Zhong contributed equally to this work.

State Key Laboratory of Rice Biology, China National Rice Research Institute, Hangzhou 310006, China



toxicity concentration. Moreover, these also show that the genetic mechanism of Al tolerance in rice is great complicated and need more studies on it in the future.

There are several limitations of linkage mapping: 1. only two alleles at any given locus can be studied in biparental crosses; 2. high cost; 3. poor mapping resolution [29], whereas association mapping based on linkage disequilibrium (LD) could overcome these limitations [30] and enable researchers to use modern technologies to exploit natural variations [31]. A genome wide association study (GWAS) was reported in rice Al tolerance using about 36 thousand SNPs [9], and an association analysis was performed in rice Al tolerance in our previous study in which only 274 SSRs markers were used in association mapping [17]. However, no higher resolution GWAS on Al tolerance within natural population was performed in previous studies on Al tolerance in rice.

Ting's rice collection which is one of the earliest rice collection in China consists of 150 varieties constructed from 2262 of 7128 original landraces [32]. Ting's core collection had been used for association mapping on rice agronomic traits [33]. Moreover, Ting's core collection had been reported that it possessed a wide-range of phenotypic variation for Al tolerance and was used to perform association mapping on Al tolerance [17]. Therefore, Ting's core collection could be an appropriate population for GWAS on rice Al tolerance.

In the present study, a GWAS for rice Al tolerance based on relative root elongation (RRE) in seedling stage was carried out using Ting's core collection with more than 3.8 million high quality SNPs. Candidate regions identified by GWAS were compared with regions identified as QTL in

previous studies and with Al sensitive mutants and/or candidate genes. This study would provide important information of candidate genes for Al tolerance improvement in rice.

Results

Identification of phenotypic variations for Al tolerance

The rice landraces in Ting's core collection revealed a wide range of phenotypic variation and indicated a normal distribution for Al tolerance which was identified by RRE (Fig. 1a, Additional file 5: Table S1) as well as showed strong Al tolerance with an average of RRE > 0.50 (Additional file 1: Figure S1).

Genome re-sequencing and SNPs getting

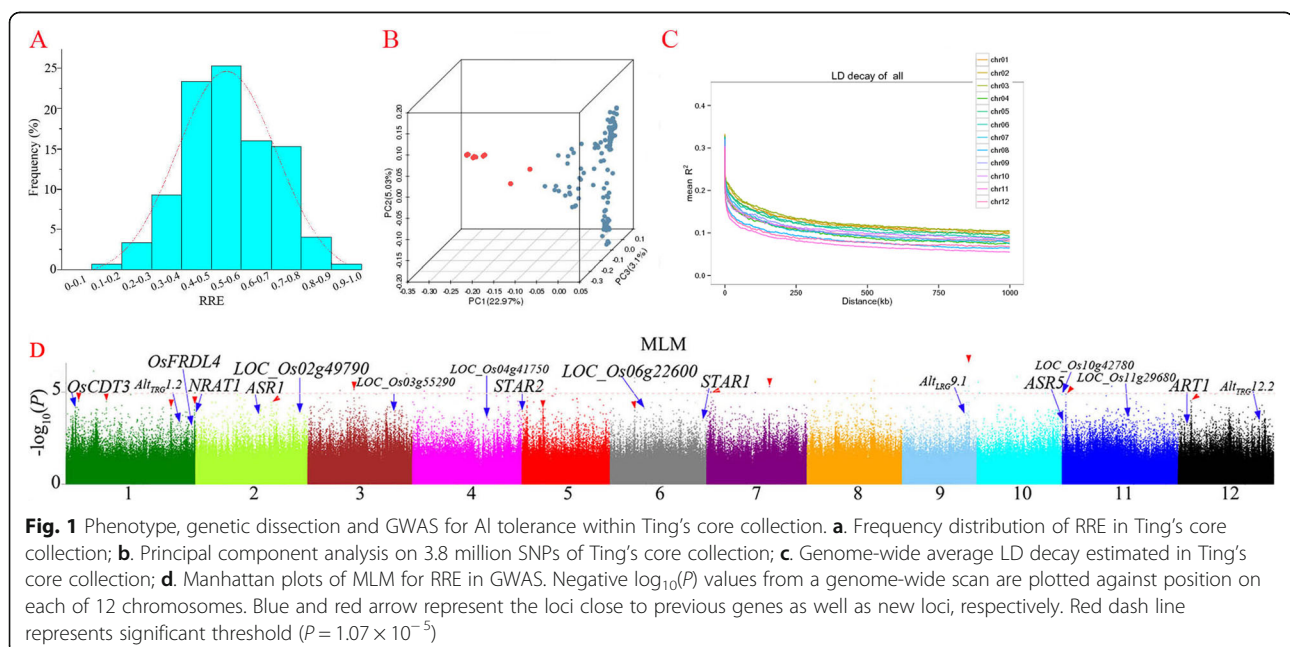
Ting's core collection was re-sequenced with whole genome re-sequencing method. High quality SNPs (Additional file 7: Table S3) distribution along position on each chromosome were shown in Additional file 2: Figure S2. 3,808,730 SNPs were generated totally, while there were 386,562 SNPs found in the CDS region (Additional file 7: Table S3).

Population structure and LD estimation

PCA were performed with all SNPs for identifying the population structure of Ting's core collection, and two sub-populations were observed (Fig. 1b). Moreover, we measured LD through r^2 for Ting's core collection using the SNP data. We found that LD dropped to half of its maximum value at 100~350 kb on 12 chromosomes (Fig. 1c).

GWAS

A total of 3,808,730 SNPs were applied into GWAS for Al tolerance using GLM and MLM. We used $P =$



1.07×10^{-5} and $P = 2.63 \times 10^{-7}$ as the significant threshold for RRE in our study for MLM and GLM, respectively (Fig. 1d and Additional file 3: Figure S3). 5 and 38 significant loci distributing on 12 chromosomes were identified totally for Al tolerance in GLM and MLM, respectively (Table 1). The highest significant signal on each chromosome were shown in Table 1 in GLM and MLM. Our results indicated that the false positives were well controlled in this study (Additional file 4: Figure S4).

In our study, we identified 17 genes relating to rice Al tolerance which were reported in previous studies (Fig. 1d) with a modest significant threshold ($P < 0.001$). Besides, 13 new candidate regions with high LD and 69 new candidate genes for Al tolerance were identified firstly using MLM in our study (Additional file 8: Table S4). New significant association regions with smaller P value and higher consecutive peak detected by MLM for Al tolerance were shown in Fig. 1d. And detailed distribution of these new gene-based association signals were included in Additional file 9: Table S5.

Effects of allele variations

Allele analysis was performed for 13 high significant SNPs and frequency of alleles existing in Ting's core collection were higher than those in reference (Table 2). Furthermore, top five significant SNPs at the gene-based location were selected for detecting the effects of allelic variations on Al tolerance (Table 2). RRE of varieties with alternative alleles for above five SNPs were not larger than those with reference (Nipponbare) alleles in Ting's core collection (Fig. 2a). Moreover, staining by hematoxylin of Nipponbare (Al-tolerant) with allele C was lighter than Youzhan (an Al-sensitive variety) with

allele T at 10,953,291 on chromosome 1 after Al toxicity treatment (Fig. 2b and c).

Transcriptomic analysis

6948 genes were up-regulated for Al tolerant variety between treatment under (T) and without (CK) Al toxicity, while 1063 genes were up-regulated for Al sensitive variety and totally 796 genes were up-regulated both in Al tolerant and sensitive varieties (Fig. 3a). There were 4910, 544 and 322 genes which were down-regulated for Al tolerant, sensitive and both, respectively (Fig. 3b). Among 69 candidate genes identified through GWAS, 20 genes were also detected with significant difference between treatment under and without Al toxicity by transcriptome sequencing (Additional file 10: Table S6).

qRT-PCR and sequence analysis in CDS region

The expression of eight in above candidate genes detected by transcriptome sequencing were significant different in root tip between Nipponbare (Al tolerant) and Kasalath (Al sensitive) (Fig. 4). At the meanwhile, 20 candidate genes were chosen for sequencing analysis in CDS region and sequence variations including substitutions, deletions and insertions could be found in seven genes CDS region between Al-tolerant and Al-sensitive varieties, which lead to amino acid variations (Table 3).

Discussion

Al toxicity is one of the most limits for rice production. QTL mapping based on segregating population is the main method for most of previous studies for Al tolerance. Linkage mapping which were constructed using typical Al tolerant and sensitive varieties. However, only two alleles at any given locus can be studied in linkage

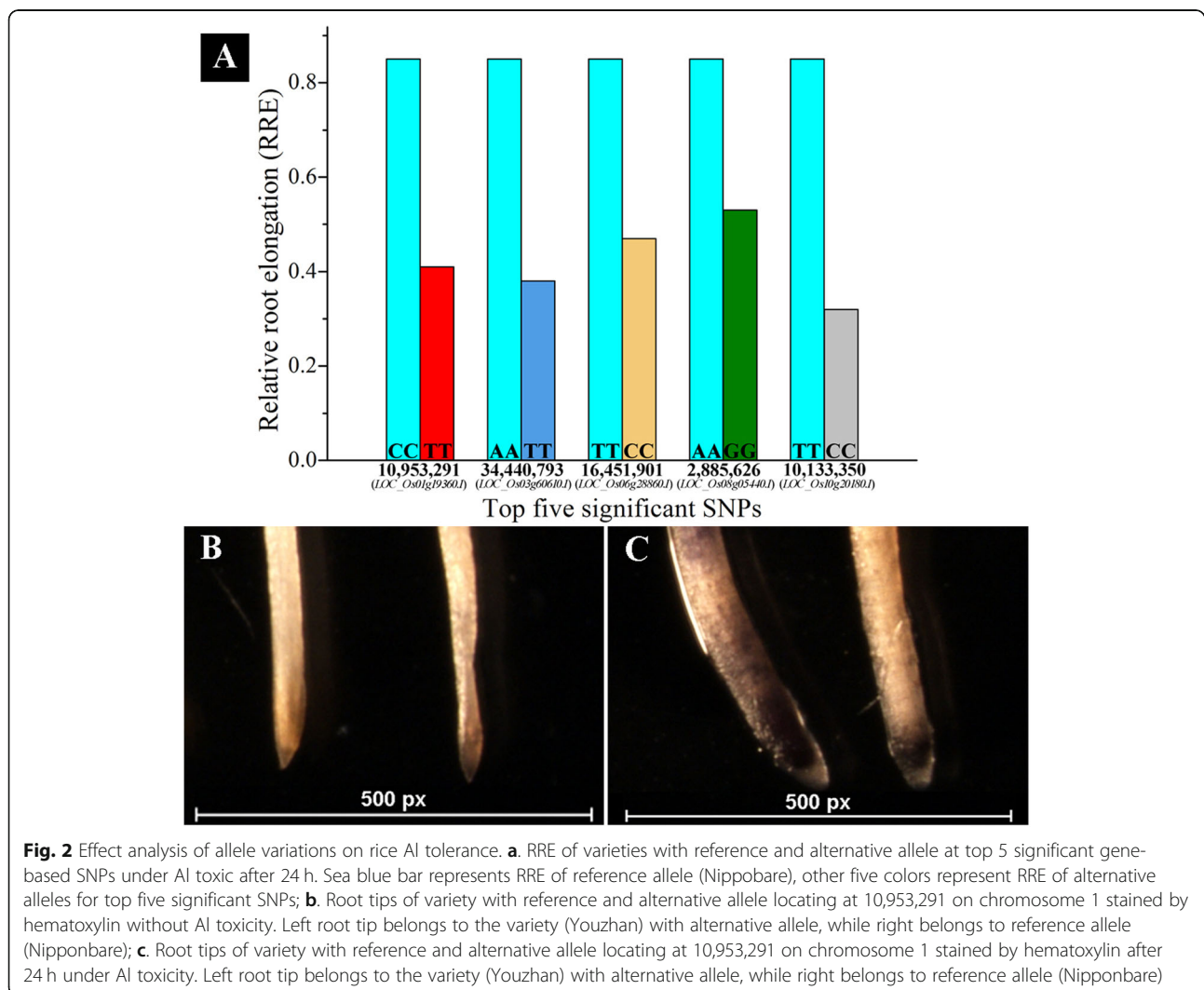
Table 1 Summary of association mapping results for relative root elongation using GLM and MLM

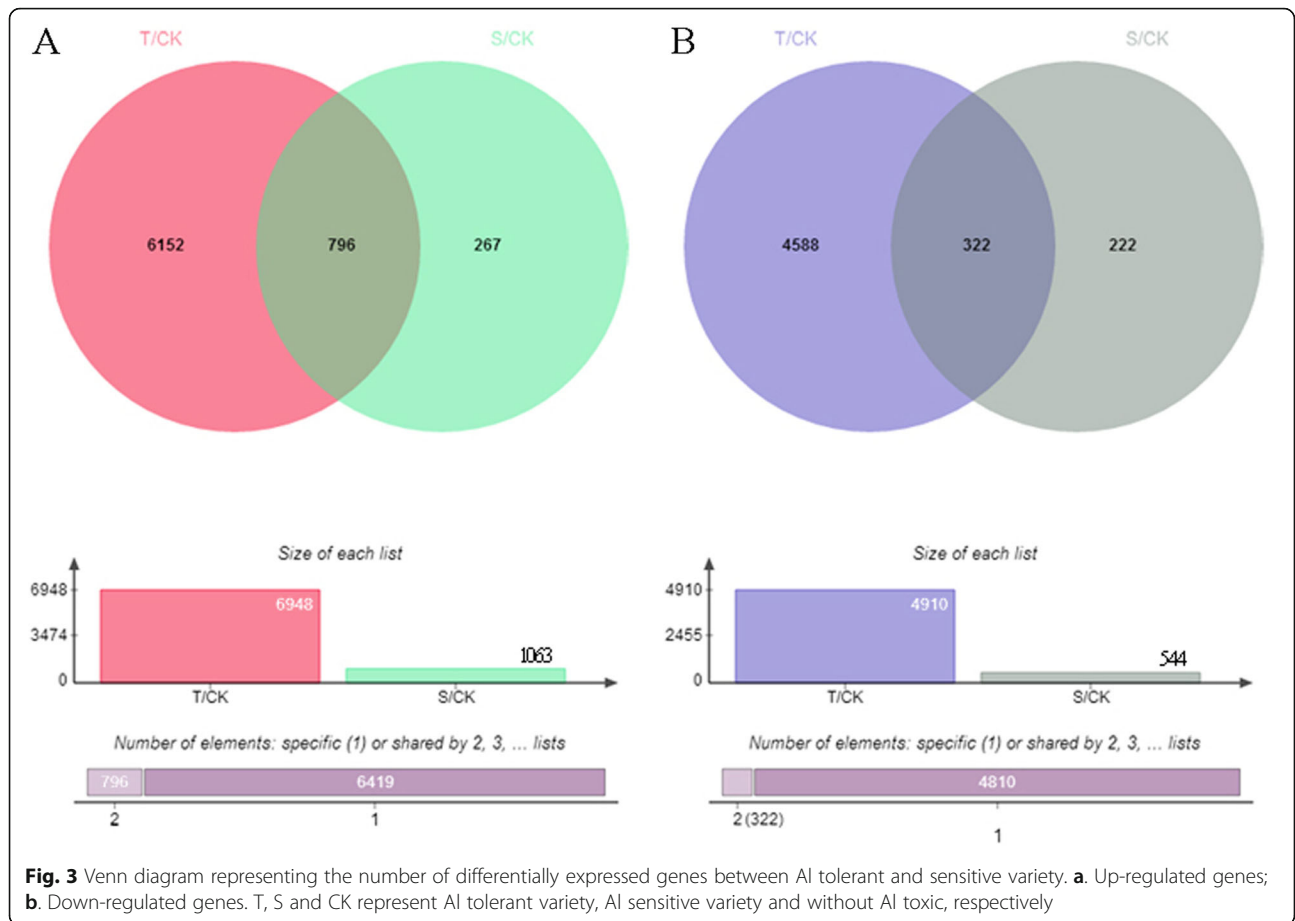
Chr.	Number of significant loci		SNP with the highest $-\log_{10}P$ (IRGSP-1.0)	
	GLM ($-\log_{10}P > 6.58$)	MLM ($-\log_{10}P > 4.97$)	Position (GLM)	Position (MLM)
1	1	2	10,953,291 ($-\log_{10}P = 6.7$)	10,953,291 ($-\log_{10}P = 5.6$)
2	1	1	14,694,840 ($-\log_{10}P = 7.2$)	14,694,840 ($-\log_{10}P = 5.0$)
3	0	4	34,440,793 ($-\log_{10}P = 6.5$)	34,440,793 ($-\log_{10}P = 5.5$)
4	0	3	6,783,812 ($-\log_{10}P = 5.9$)	6,783,812 ($-\log_{10}P = 5.1$)
5	0	3	6,223,962 ($-\log_{10}P = 6.1$)	6,223,962 ($-\log_{10}P = 5.1$)
6	1	5	7,814,182 ($-\log_{10}P = 7.2$)	7,814,182 ($-\log_{10}P = 5.3$)
7	0	5	17,113,274 ($-\log_{10}P = 6.1$)	17,113,274 ($-\log_{10}P = 5.1$)
8	1	6	2,885,626 ($-\log_{10}P = 6.8$)	2,885,626 ($-\log_{10}P = 5.1$)
9	0	4	19,838,960 ($-\log_{10}P = 6.5$)	19,838,960 ($-\log_{10}P = 6.0$)
10	1	4	10,133,350 ($-\log_{10}P = 6.8$)	10,133,350 ($-\log_{10}P = 5.5$)
11	0	1	13,457,541 ($-\log_{10}P = 6.1$)	13,457,541 ($-\log_{10}P = 5.1$)
12	0	0	11,951,228 ($-\log_{10}P = 6.0$)	10,528,920 ($-\log_{10}P = 4.5$)
Total	5	38	-	-

Table 2 Top highest genome-wide significant association signals of RRE using MLM

Trait	Chr.	Position (IRGSP-1.0)	Reference allele	Alternative allele	Alternative allele frequency	$-\log_{10}(P)$	Candidate/known gene ^a
RRE	1	10,953,291	C	T	0.87	5.58	<i>LOC_Os01g19360.1</i>
	2	14,694,840	A	G	0.85	5.04	
	3	34,440,793	A	T	0.88	5.45	<i>LOC_Os03g60610.1</i>
	6	7,814,182	A	G	0.73	5.24	
	6	7,814,184	C	T	0.74	5.24	
	6	7,814,207	C	T	0.75	5.06	
	6	16,451,901	T	C	0.86	5.02	<i>LOC_Os06g28860.1</i>
	8	18,696,483	T	C	0.67	5.11	
	8	2,885,626	A	G	0.63	5.10	<i>LOC_Os08g05440.1</i>
	8	2,885,642	A	T	0.61	5.04	
	9	19,838,960	G	A	0.88	5.99	
	10	10,133,350	T	C	0.58	5.30	<i>LOC_Os10g20180.1</i>
	10	8,299,577	T	C	0.71	5.03	

Note: ^aGene ID of MSU rice genome annotation project (<http://rice.plantbiology.msu.edu/>)





mapping. In this study, Ting’s core collection consisted of rice landraces which represents an intermediate stage in domestication between wild and elite cultivars [34] was used for association mapping for AI tolerance. A wide-range of phenotypic variations and a normal distribution for AI tolerance were indicated in Ting’s core

collection. Therefore, Ting’s core collection could be used as for GWAS on AI tolerance.

There were two researches on rice AI tolerance using low resolution GWAS [9, 17]. More than 3.8 million SNPs were applied into GWAS for AI tolerance in the present study and much higher in resolution than above

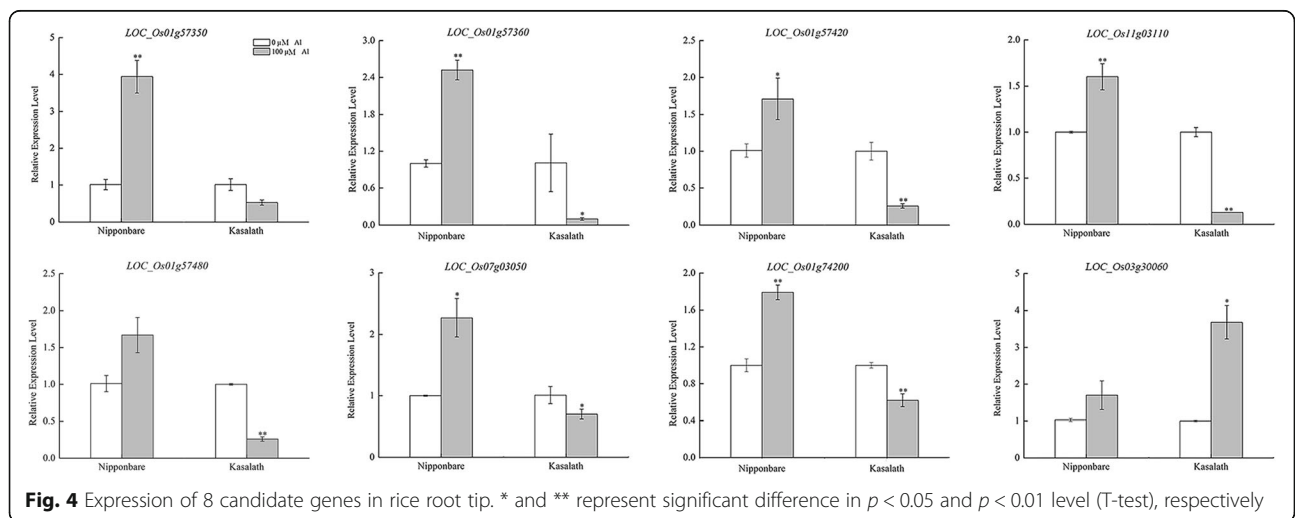


Table 3 Detection of frame shift mutation in varieties with different Al tolerance

Candidate gene	Varieties with different Al tolerance	Position from the initial codon in CDS region	Codon	Amino acid
<i>LOC_Os01g57350.1</i>	Nipponbare	182nd	CAG	Glutamine
	Kasalath		GAG	Glutamic acid
<i>LOC_Os03g30060.1</i>	Nipponbare	261st	TAC	Tyrosine
	Kasalath		CAC	Histidine
	Nipponbare	20th	AAG	Lysine
	Kasalath		CAG	Glutamine
	Nipponbare	66th	GCC	Alanine
	Kasalath		ACC	Threonine
	Nipponbare	88th	GTG	Valine
	Kasalath		ATG	Methionine
	Nipponbare	125th	GCA	Alanine
	Kasalath		GGA	Glycine
	Nipponbare	126th	GCG	Alanine
	Kasalath		TCG	Serine
	Nipponbare	155th	AAC	Asparagine
	Kasalath		AGC	Serine
	Nipponbare	184th	GCG	Alanine
	Kasalath		ACG	Threonine
	Nipponbare	220th	CAG	Glutamine
	Kasalath		GAG	Glutamic acid
<i>LOC_Os07g03050.1</i>	Nipponbare	246th	GCC	Alanine
	Kasalath		Deletion	Deletion
	Nipponbare	260th	ATG	Methionine
	Kasalath		GTG	Valine
	Nipponbare	276th	ATC	Isoleucine
	Kasalath		GTC	Valine
	Nipponbare	277th	GCG	Alanine
	Kasalath		ACG	Threonine
	Nipponbare	7th	TAT	Tyrosine
	Kasalath		TCT	Serine
	Nipponbare	344th	TGT	Cysteine
	Kasalath		AGT	Serine
<i>LOC_Os01g57360.1</i>	Nipponbare	32nd	AAG	Lysine
	IR64		AAT	Asparagine
	Nipponbare	348th	GGC	Glycine
	IR64		TGC	Cysteine
<i>LOC_Os11g03110.1</i>	Nipponbare	261st~269th	Deletion	Deletion
	Kasalath		Insertion	Insertion
	Nipponbare	608th	GTG	Valine
	Kasalath		GCG	Alanine
<i>LOC_</i>	Nipponbare	12th	TCG	Cysteine

Table 3 Detection of frame shift mutation in varieties with different Al tolerance (Continued)

Candidate gene	Varieties with different Al tolerance	Position from the initial codon in CDS region	Codon	Amino acid
<i>Os01g74200.1</i>	IR64		TCA	Serine
<i>LOC_Os09g33550.1</i>	Nipponbare	50th	CAC	Histidine
	IR64		CCC	Proline
	Nipponbare	58th	GCG	Alanine
	IR64		ACG	Threonine
<i>LOC_</i>	Nipponbare	65th	GCG	Alanine
	IR64		GTG	Valine

two studies. Moreover, LD dropped to half of its maximum value at 100~350 kb (Fig. 1c) which is agreement with the previous measurements [35–37].

The research of Famoso et al. (2011) is the first GWAS on Al tolerance in rice [9]. In this study, 48 regions associated with Al tolerance were identified by GWAS. In total, three regions/loci were detected in both study of Famoso et al. (2011) and present. Ting's core collection was used to perform GWAS in our previous study using 274 SSR markers [17], and 23 significant trait-marker associations were discovered. However, no identical region/locus was identified in both our previous and present study. There might be three reasons for explaining this: 1. Different markers with distinct resolutions used in two researches; 2. No identical significant threshold used in two researches. In our previous study, $P < 0.05$ was applied to judge a significant region/locus. The significant trait-marker associations might not be detected in present study; 3. RRE of 18 varieties were missing in present study while there were only three missing in our previous studies.

In the present study, we found that there was no signals with a strong significant threshold in previously reported genes or regions. It is probably because: 1. Different materials were used in previous studies; 2. No same statistical model was used in previous studies; 3. Most of QTL were identified with mutations in previous studies; 4. Al tolerance were affected by numerous alleles with minor effects, so previous QTL/genes cannot be detected in the present study. Thus, a modest significant threshold ($P < 0.001$) was used in the present study which is same to the study of Huang et al. (2015) for grain number [38]. We identified 17 genes ($P < 0.001$) relating to rice Al tolerance which were reported in previous studies (Fig. 1d). Significant association signals close to *OsCDT3* on chromosome 1 [39], *Alt_{TRG1.2}* on chromosome 1 [9], *OsFRDL4* on chromosome 1 [15], *NRAT1* on chromosome 2 [13], *ASRI* on chromosome 2 [7], *LOC_Os02g49790* on chromosome 2 (Rice Genome Annotation Project, RGAP), *LOC_Os03g55290* on chromosome 3 (RGAP), *LOC_Os04g41750* on chromosome 4

(RGAP), *STAR2* on chromosome 5 [10], *LOC_Os06g22600* on chromosome 6 (RGAP), *STAR1* on chromosome 6 [10], *Alt_{LRG}9.1* on chromosome 9 [9], *LOC_Os10g42780* on chromosome 10 (RGAP), *ASR5* on chromosome 11 [20], *LOC_Os11g29680* on chromosome 11 (RGAP), *ART1* on chromosome 12 [14] and *Alt_{TRG}12.2* on chromosome 12 [9] were detected in our study. It is necessary to notice that $-\log_{10}(P)$ value of these significant signals both in our study and the first as well as the only one GWAS research [9] for rice Al tolerance are not large as those in GWAS researches of other rice traits. In our opinion, this might be due to high complexity of rice Al tolerance mechanism.

In our opinion, those peaks with consecutive loci which are below the significant threshold ($-\log_{10}(P) = 4.97$) are also valuable to be analyzed deeply because the reported genes for Al tolerance are below 4.97, hence, we set $-\log_{10}(P) = 3.0$ as the threshold for above peaks. 13 new candidate regions with high LD and 69 new candidate genes (Additional file 8: Table S4) for Al tolerance were identified firstly using MLM in our study. Among 69 candidate genes for Al tolerance, 48 genes could be not predicted their performance on Al detoxification through annotation, while the remaining 21 genes could be speculated their detoxifying Al mechanism with annotation information and previous studies (Additional file 8: Table S4). For instance, *LOC_Os01g09370.1* encodes an ankyrin repeat domain-containing protein 28, and ankyrin repeat domain-containing protein was reported that it is probably involved in the regulation of antioxidation metabolism shared by stress responses [40]. Therefore, *LOC_Os01g09370.1* might be one gene which is related to Al stress. *LOC_Os01g57350.1*, *LOC_Os01g57360.1*, *LOC_Os01g57420.1* and *LOC_Os07g29220.1* encode a diacylglycerol kinase, an acyltransferase, a diacylglycerol kinase as well as a cyclopropane-fatty-acyl-phospholipid synthase respectively which are necessary to lipid metabolism in plant [41]. Moreover, lipids were reported that they might participate in rice Al detoxification [42, 43]. Interestingly, few scientists paid attention to study on lipid in rice Al tolerance. Pentatricopeptide repeat domain containing protein encoded by *LOC_Os01g57410.1* was identified that this protein was a regulator for abiotic stress [44]. *LOC_Os01g25090.1*, *LOC_Os01g74160.1* and *LOC_Os03g30080.1* encode a transposon protein, putative, CACTA, En/Spm sub-class [45], a carboxyl-terminal peptidase [46] as well as a transposon protein, putative, CACTA, En/Spm sub-class [44] which were associated with metal accumulation. *LOC_Os09g33559.1* encodes a proline-rich family protein, while proline was reported it relating to Al metabolism [47].

To further test positive capability of candidate genes identified in our study, we chose two varieties (Al tolerant and sensitive) to perform transcriptome sequencing. Among 69 candidate genes identified through GWAS, 20 genes were also detected with significant differences

between treatments under and without Al toxicity by transcriptome sequencing (Additional file 10: Table S6). There were three candidate genes (i.e. *LOC_Os01g57350.1*, *LOC_Os01g57420.1* and *LOC_Os07g29220.1*) which were up regulated (T versus CK) obviously in Al tolerant variety while no significant difference in Al sensitive variety among candidate genes in group lipid metabolism, and another candidate gene (i.e. *LOC_Os01g57360.1*) was up regulated in Al sensitive variety. Both *LOC_Os01g57450.1* and *LOC_Os01g57480.1* which are in group abiotic stress response were up regulated in Al tolerant variety while no significant difference in Al sensitive variety. In group membrane protein, both *LOC_Os01g74180.1* and *LOC_Os06g14030.1* were up regulated in Al tolerant variety while no significant difference in Al sensitive variety. Expression of candidate gene *LOC_Os09g33559.1* was down regulated in Al tolerant variety while no significant difference in Al sensitive variety. Moreover, remaining 11 candidate genes in group unknown were up or down regulated for two varieties. Furthermore, qRT-PCR and sequence analysis in CDS region also gave more supports to our candidate genes.

In addition, after GWAS experiment, we agree that it is a better way to carry out further studies by using varieties with extreme values in Ting's core collection. In present study, after GWAS transcriptome analysis, histological observation, sequence comparisons of CDS and qRT-PCR were used to support the results of GWAS. Histological observation was performed to identify the effects of allele variations of significant SNPs from GWAS. Transcriptome analysis was to try to find some identical genes to GWAS. Both sequence comparisons of CDS and qRT-PCR were designed to try to support the candidate genes identified by GWAS and transcriptome. We chose different materials for further study because the candidate genes in present study were supposed to be supported by different materials. And we tried to prove the candidate genes from GWAS which are natural variations existing in Ting's core collection would have effects to rice Al tolerance in other materials.

Conclusions

In this study, 69 new candidate genes for Al tolerance were identified using GWAS and some among 69 candidate genes were also detected through transcriptome sequencing, qRT-PCR and checking of sequence variation of CDS. It is worth to perform deeper research on them.

Methods

Plant material

Ting's core collection with 150 rice landraces [48] was used in present study. The information of 150 landraces are shown in Additional file 5: Table S1.

Phenotyping for Al tolerance

Seeds of Ting's core collection were harvested from the farm of China National Rice Research Institute, Hangzhou (30°3 N, 120°2E), during the late season (May–October) in 2016. Relative root elongation (RRE) was used to evaluate Al tolerance of all varieties, please see more details in our previous study [17]. RRE of 10 seedlings of each variety growing in parallel were measured with a ruler before and after the treatments (24 h) in one replicate, and six replications were performed. Fu et al. (2010) referenced that $RRE \geq 0.5$ should be used as a criterion to find out Al tolerant varieties [49].

The root tips (1~2 cm) of rice seedlings cultivated for 24 h in the presence or absence of Al were gently shaken in a $0.5 \text{ mmol l}^{-1} \text{ CaCl}_2$ (pH = 4.0) solution for 10 min. The root tips were submerged into hematoxylin solution (0.2% hematoxylin and 0.02% potassium iodide). After 45 min, the root tips were soaked into deionized water till no clear color on root tips. Then the root tips were photographed using stereoscopic and light microscopes. We chose Youzhan as one of the varieties which had different alleles (alternative allele) to Nipponbare (reference allele) for some SNPs to be stained by hematoxylin solution.

Genome re-sequencing and SNP calling

Genomic DNA from a single plant was used for sequencing. Re-sequencing was performed by Illumina HiSeq™ 4000 with 6~7 folds of genome coverage. We performed the mapping to the rice reference genome (IRGSP-1.0) for the 150-bp reads by using *bwamem* with the *-M* option of BWA software [50], and the mapped reads were realigned by using *RealignerTargetCreator* in GATK [51]. SNPs were labeled with the *-glm BOTH* option of *UnifiedGenotyper* in GATK. After filtering the SNPs with low minor-allele frequency (5%), a total of 3,808,730 SNPs was remained for GWAS in our study. Please read more details in our previous study [52].

Genetic analyses

Genetic analyses were performed following our previous study [52]. Principal component analysis (PCA), and LD decay analysis among Ting's core collection were performed based on all SNPs. SPSS 17.0 was used for PCA analysis. The LD was evaluated using squared Pearson's correlation coefficient (r^2) calculated with the *-r²* command of the software PLINK [53]. The Q-matrix was obtained with the membership probability of each variety using ADMIXTURE Version 1.22. The Loisel algorithm was chosen for calculating kinship matrix by software SPAGeDi [54].

GWAS

A total of 3,808,730 SNPs were used for GWAS. General linear model (GLM, without Q-matrix) and mixed linear model (MLM, K + Q) were performed using TASSEL

software (www.maizegenetics.net/tassel). $P \leq 2.63 \times 10^{-7}$ ($P = 1/n$; n = total markers, which is a rough Bonferroni correction corresponding to $-\log_{10}(P) = 6.58$) should be defined as the suggestive threshold which is same to that in our previous study [52] in present study, but significant locus can not be identified according to this threshold. Thus, another significance threshold was calculated, i.e., a minimum Bayes factor (mBF). The mBF was calculated using the following formula: $mBF = -e^{-P} \cdot \ln(P)$ [55], thus, the significance threshold in present study was $-\log_{10}(P) = 4.97$.

Transcriptomic analysis

Bashizi, one Al-tolerant variety (RRE = 0.65) and Heikedanuo, one Al-sensitive (RRE = 0.34) from Ting's core collection (Additional file 5: Table S1) were chosen for transcriptome analysis. Total RNA of root tip after one day under $100 \mu\text{mol l}^{-1} \text{ AlCl}_3$ was extracted with the miRNeasy Kit (QIAGEN, USA). The RNA samples were evaluated on agarose gels, quantified in a spectrophotometer and stored at -80°C . RNA sequencing was performed by Illumina NextSeq 500 for 2×125 -bp paired-end sequencing. Samples under Al toxic treatment and without Al toxic in three times as well as one time respectively were collected for RNA sequencing [56].

Real-time qRT-PCR

Nipponbare (RRE = 0.85), one well known Al-tolerant variety and Kasalath (RRE = 0.30), one well known Al-sensitive were chosen for real-time qRT-PCR analysis. Total RNA of root tip after one day without and under $100 \mu\text{mol l}^{-1} \text{ AlCl}_3$ was extracted with the miRNeasy Kit (QIAGEN, USA). RNA was converted to cDNA using the protocol supplied by the manufacturer of ReverTra Ace qPCR RT Master Mix with gDNA remover (TOYOBO). The expression was determined with THUNDERBIRD™ SYBR® qPCR Mix without ROX (TOYOBO) by Roche Light Cycler® 480II. The primer sequences for qRT-PCR were listed in Additional file 6: Table S2. Ubiquitin was used as an internal control. Relative expression levels were calculated by $2^{-\Delta\Delta\text{Ct}}$ method. Three independent biological replicates were made for each treatment. The volume of the qRT-PCR reaction system was 10 μl : SYBR Premix Ex Taq II, 2x, 5 μl ; PCR Primer (Forward+Reverse, 10uM), 2 μl ; cDNA, 2 μl ; ddH₂O, 1 μl . The profile of PCR program is: 95 °C for 3 min; 45 cycles of 95 °C for 10 s, 58 °C for 15 s, 72 °C for 25 s; 95 °C 10 s; 65 °C 1 min; 40 °C 1 min.

CDS sequence variation analysis of candidate genes

Nipponbare, one well known Al-tolerant variety. Kasalath and IR64 (RRE = 0.29), two well known Al-sensitive varieties were chosen for sequence variation analysis. Genomic DNA was extracted using modified SDS

method. Several primers were designed for CDS sequencing of each gene. Then, CDS were amplified by PCR, the volume of the PCR reaction system was 10 μ l. The profile of PCR program is: 94 °C for 5mins; 31 cycles of 94 °C for 1 min, 52–58 °C for 1 min, 72 °C for 1–2 min; and a 5mins final extension at 72 °C. Amplified products were size separated by 1% agarose gel electrophoresis and purified using QIAquick Gel Extraction Kit (QIAGEN, Germantown, MD, USA). After sequencing of PCR products, sequences of same CDS were assembled and CDS sequences of Nipponbare, Kasalath and IR64 were aligned using software BioEdit.

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s12870-019-2036-z>.

Additional file 1: Figure S1. Box chart of Al tolerance (RRE) in Ting's core collection.

Additional file 2: Figure S2. SNP and Indel distribution along position in each chromosome. Longitudinal axis represent the number of SNP and Indel.

Additional file 3: Figure S3. Manhattan plots of GLM for RRE in GWAS. Negative $\log_{10}(P)$ values from a genome-wide scan are plotted against position on each of 12 chromosomes. Red dash line represents significant threshold ($P = 1 \times 10^{-5}$).

Additional file 4: Figure S4. Plots of observed versus expected P -values using MLM and GLM for RRE. A. MLM; B. GLM. Red symbol represents expected P -values, and Blue symbol represents observed P -values.

Additional file 5: Table S1. Accessions, variety names, origin and RRE of 150 rice varieties in Ting's core collection.

Additional file 6: Table S2. Primers used in qRT-PCR in this study.

Additional file 7: Table S3. Summary of categorized SNPs and Indels.

Additional file 8: Table S4. List of new loci in association analysis of root relative elongation under Al toxic.

Additional file 9: Table S5. List of all P value ranked genes in the gene-based association analysis of root relative elongation using MLM.

Additional file 10: Table S6. List of new gene's expression in association analysis of root relative elongation under Al toxic and without Al toxic.

Abbreviations

CDS: Coding sequence; DNA: Deoxyribonucleic acid; EMMAX: Efficient mixed model association eXpedited; IRGSP: International Rice Genome Sequencing Project; qRT-PCR: Quantitative reverse transcription-PCR; QTL: Quantitative trait locus; SDS: Sodium dodecyl sulfate; SNPs: Single Nucleotide Polymorphisms; SSR: Simple Sequence Repeats

Acknowledgements

We are grateful to Dr. Jinquan Li from Max Planck Institute for Plant Breeding Research and Dr. Xiangdong Liu from South China Agricultural University.

Authors' contributions

Conceived and designed the experiments: PZ. Performed the experiments: PZ and KZ. Analyzed the data: PZ, KZ, and ZZ. Contributed reagents/materials/analysis tools: PZ and HT. Wrote the paper: PZ and KZ. All authors read and approved the final manuscript.

Funding

This work was supported by a fund of the National Natural Science Foundation of China grant 31701401, a fund of Zhejiang Province Public Welfare Technology Application Research Project (LGN19C130005), and a

fund from the State Key Laboratory for Conservation and Utilization of Subtropical Agro-bioresources (SKLCUSA-b201713). The funding bodies did not play any role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Availability of data and materials

The datasets used during the current study are available from the corresponding author on reasonable request.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 21 February 2019 Accepted: 12 September 2019

Published online: 12 November 2019

References

- Kochian LV, Hoekenga OA, Pineros MA. How do crop plants tolerate acid soils? Mechanisms of aluminum tolerance and phosphorous efficiency. *Annu Rev Plant Biol.* 2004;55:459–93.
- Ryan PR, Tyerman SD, Sasaki T, Furuichi T, Yamamoto Y, Zhang WH, Delhaize E. The identification of aluminium-resistance genes provides opportunities for enhancing crop production on acid soils. *J Exp Bot.* 2011; 62:9–20.
- Vonuekkull HR, Mutert E. Global extent, development and economic-impact of acid soils. *Plant Soil.* 1995;171:1–15.
- Ismail AM, Heuer S, Thomson MJ, Wissuwa M. Genetic and genomic approaches to develop rice germplasm for problem soils. *Plant Mol Biol.* 2007;65:547–70.
- Liu J, Luo X, Shaff J, Liang C, Jia X, Li Z, Magalhaes J, Kochian LV. A promoter-swap strategy between the *AtALMT* and *AtMATE* genes increased Arabidopsis aluminum resistance and improved carbon-use efficiency for aluminum resistance. *Plant J.* 2012;71:327–37.
- Pilon-Smits EA, Quinn CF, Tapken W, Malagoli M, Schiavon M. Physiological functions of beneficial elements. *Curr Opin Plant Biol.* 2009;12:267–74.
- Arenhart RA, Schunemann M, Bucker NL, Margis R, Wang ZY, Margis-Pinheiro M. Rice ASR1 and ASR5 are complementary transcription factors regulating aluminium responsive genes. *Plant Cell Environ.* 2016;39:645–51.
- Che J, Yamaji N, Shen RF, Ma JF. An Al-inducible expansin gene, *OsEXPA10* is involved in root cell elongation of rice. *Plant J.* 2016;88:132–42.
- Famoso AN, Zhao K, Clark RT, Tung CW, Wright MH, Bustamante C, Kochian LV, McCouch SR. Genetic architecture of aluminum tolerance in rice (*Oryza sativa*) determined through genome-wide association analysis and QTL mapping. *PLoS Genet.* 2011;7:e1002221.
- Huang CF, Yamaji N, Mitani N, Yano M, Nagamura Y, Ma JF. A bacterial-type ABC transporter is involved in aluminum tolerance in rice. *Plant Cell.* 2009; 21:655–67.
- Li J, Liu J, Dong D, Jia X, McCouch SR, Kochian LV. Natural variation underlies alterations in Nramp aluminum transporter (NRAT1) expression and function that play a key role in rice aluminum tolerance. *Proc Natl Acad Sci U S A.* 2014;111:6503–8.
- Liu S, Gao H, Wu X, Fang Q, Chen L, Zhao F, Huang C. Isolation and characterization of an aluminum-resistant mutant in rice. *Rice.* 2016;9:11–2016.
- Xia J, Yamaji N, Kasai T, Ma JF. Plasma membrane-localized transporter for aluminum in rice. *Proc Natl Acad Sci U S A.* 2010;107:18381–5.
- Yamaji N, Huang CF, Nagao S, Yano M, Sato Y, Nagamura Y, Ma JF. A zinc finger transcription factor ART1 regulates multiple genes implicated in aluminum tolerance in rice. *Plant Cell.* 2009;21:3339–49.
- Yokosho K, Yamaji N, Fujii-Kashino M, Ma JF. Functional analysis of a MATE gene *OsFRDL2* revealed its involvement in Al-induced secretion of citrate, but a lower contribution to Al tolerance in rice. *Plant Cell Physiol.* 2016;57:976–85.
- Yokosho K, Yamaji N, Fujii-Kashino M, Ma JF. Retrotransposon-mediated aluminum tolerance through enhanced expression of the citrate transporter *OsFRDL4*. *Plant Physiol.* 2016;172:2327–36.
- Zhang P, Zhong K, Tong H, Shahid MQ, Li J. Association mapping for aluminum tolerance in a Core collection of Rice landraces. *Front Plant Sci.* 2016;7:1415.

18. Yokosho K, Yamaji N, Ma JF. An Al-inducible MATE gene is involved in external detoxification of Al in rice. *Plant J.* 2011;68:1061–9.
19. Arenhart RA, Lima JC, Pedron M, Carvalho FE, Silveira JA, Rosa SB, Caverzan A, Andrade CMB, Schunemann M, Margis R, Margis-Pinheiro M. Involvement of ASR genes in aluminium tolerance mechanisms in rice. *Plant Cell Environ.* 2013;36:52–67.
20. Arenhart RA, Bai Y, de Oliveira LF, Neto LB, Schunemann M, Maraschin FS, Maraschin FS, Mariath J, Silverio A, Sachetto-Martins G, et al. New insights into aluminium tolerance in rice: the ASR5 protein binds the STAR1 promoter and other aluminum responsive genes. *Mol Plant.* 2014;7:709–21.
21. Ma JF, Furukawa J. Recent progress in the research of external Al detoxification in higher plants: a mini review. *J Inorg Biochem.* 2003;97:46–51.
22. Mao C, Yi K, Yang L, Zheng B, Wu Y, Liu F, Wu P. Identification of aluminium-regulated genes by cDNA-AFLP in rice (*Oryza sativa* L.): aluminium-regulated genes for the metabolism of cell wall components. *J Exp Bot.* 2004;55:137–43.
23. Mao C, Yang L, Zheng B, Wu Y, Liu F, Yi K, Wu P. Comparative mapping of QTL for Al tolerance in rice and identification of positional Al-induced genes. *J Zhejiang Univ Sci.* 2004;5:634–43.
24. Nguyen VT, Nguyen BD, Sarkarung S, Martinez C, Paterson AH, Nguyen HT. Mapping of genes controlling aluminium tolerance in rice: comparison of different genetic backgrounds. *Mol Gen Genomics.* 2002;267:772–80.
25. Xue Y, Jiang L, Su N, Wang JK, Deng P, Ma JF, Wan J. The genetic basic and fine-mapping of a stable quantitative-trait loci for aluminium tolerance in rice. *Planta.* 2007;227:255–62.
26. Huang CF, Yamaji N, Chen ZC, Ma JF. A tonoplast-localized half-size ABC transporter is required for internal detoxification of aluminum in rice. *Plant J.* 2012;69:857–67.
27. Chen ZC, Yamaji N, Motoyama R, Nagamura Y, Ma JF. Up-regulation of a magnesium transporter gene *OsmGT1* is required for conferring aluminum tolerance in rice. *Plant Physiol.* 2012;159:1624–33.
28. Che J, Tsutsui T, Yokosho K, Yamaji N, Ma JF. Functional characterization of an aluminum (Al)-inducible transcription factor, ART2, revealed a different pathway for Al tolerance in rice. *New Phytol.* 2018. <https://doi.org/10.1111/nph.15252>.
29. Flint-Garcia SA, Thornsberry JM, Buckler ET. Structure of linkage disequilibrium in plants. *Annu Rev Plant Biol.* 2003;54:357–74.
30. Kraakman AT, Niks RE, Van den Berg PM, Stam P, Van Eeuwijk FA. Linkage disequilibrium mapping of yield and yield stability in modern spring barley cultivars. *Genetics.* 2004;168:435–46.
31. Zhu CS, Gore M, Buckler ES, Yu JM. Status and prospects of association mapping in plants. *Plant Genome.* 2008;1:5–20.
32. Li JQ, Zhang P. Assessment and utilization of the genetic diversity in Rice, in genetic diversity in plants (M Ç ed). Hard cover: InTech-Open Access Publisher; 2012.
33. Zhang P, Liu X, Tong H, Lu Y, Li J. Association mapping for important agronomic traits in core collection of rice (*Oryza sativa* L.) with SSR markers. *PLoS One.* 2014;9:e111508.
34. Londo JP, Chiang YC, Hung KH, Chiang TY, Schaal BA. Phylogeography of Asian wild rice, *Oryza rufipogon*, reveals multiple independent domestications of cultivated rice, *Oryza sativa*. *Proc Natl Acad Sci U S A.* 2006;103:9578–83.
35. Huang X, Wei X, Sang T, Zhao Q, Feng Q, Zhao Y, Li C, Zhu C, Lu T, Zhang Z, et al. Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat Genet.* 2010;42:961–7.
36. Yano K, Yamamoto E, Aya K, Takeuchi H, Lo P, Hu L, Matsuoka M. Genome-wide association study using whole-genome sequencing rapidly identifies new genes influencing agronomic traits in rice. *Nat Genet.* 2016;48:927.
37. Zhao KY, Tung CW, Eizenga GC, Wright MH, Ali ML, Price AH, McCouch SR. Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. *Nat Commun.* 2011;2:467.
38. Huang X, Yang S, Gong J, Zhao Y, Feng Q, Gong H, Li W, Zhan Q, Cheng B, Xia J, et al. Genomic analysis of hybrid rice varieties reveals numerous superior alleles that contribute to heterosis. *Nat Commun.* 2015;6:6258.
39. Xia J, Yamaji N, Ma J. A plasma membrane-localized small peptide is involved in rice aluminum tolerance. *Plant J.* 2013;76:345–55.
40. Yan J, Wang J, Zhang H. An ankyrin repeat-containing protein plays a role in both disease resistance and antioxidation metabolism. *Plant J.* 2002;29:193–202.
41. Busi MV, Maliandi MV, Valdez H, Clemente M, Zabaleta EJ, Araya A, Gomez-Casati DF. Deficiency of *Arabidopsis thaliana* frataxin alters activity of mitochondrial Fe-S proteins and induces oxidative stress. *Plant J.* 2006;48:873–82.
42. Mao X, Zhang H, Tian S, Chang X, Jing R. TaSnRK2.4, an SNF1-type serine/threonine protein kinase of wheat (*Triticum aestivum* L.), confers enhanced multistress tolerance in *Arabidopsis*. *J Exp Bot.* 2010;61:683–96.
43. Wagatsuma T, Khan MS, Watanabe T, Maejima E, Sekimoto H, Yokota T, Nakano T, Toyomasu T, Tawarayama K, Koyama H, et al. Higher sterol content regulated by *CYP51* with concomitant lower phospholipid content in membranes is a common strategy for aluminium tolerance in several plant species. *J Exp Bot.* 2015;66:907–18.
44. Tan J, Tan Z, Wu F, Sheng P, Heng Y, Wang X, Ren Y, Wang J, Guo X, Zhang X, et al. A novel chloroplast-localized pentatricopeptide repeat protein involved in splicing affects chloroplast development and abiotic stress response in Rice. *Mol Plant.* 2014;7:1329–49.
45. Hanikenne M, Talke IN, Haydon MJ, Lanz C, Nolte A, Motte P, Kroymann J, Weigel D, Kraemer U. Evolution of metal hyperaccumulation required cis-regulatory changes and triplication of *HMA4*. *Nature.* 2008;453:344–91.
46. Grill E, Winnacker EL, Zenk MH. Phytochelatins, a class of heavy-metal-binding peptides from plants, are functionally analogous to metallothioneins. *Proc Natl Acad Sci U S A.* 1987;84:439–43.
47. Giannakoula A, Moustakas M, Mylona P, Papadakis I, Yupsanis T. Aluminum tolerance in maize is correlated with increased levels of mineral nutrients, carbohydrates and proline, and decreased levels of lipid peroxidation and Al accumulation. *J Plant Physiol.* 2008;165:385–96.
48. Li X, Lu Y LJ, Xu H, Muhammad Q. Strategies on sample size determination and qualitative and quantitative traits integration to construct core collection of rice (*Oryza sativa*). *Rice Sci.* 2011;18:46–55.
49. Fu XL, Chen WD, Pan CH, Feng JH, Liu XD, Lu YG. A comparative study on aluminum tolerance between *Oryza sativa* and *O. rufipogon* in Gaozhou. *Sci Agric Sin.* 2010;4:661–9.
50. Li H, Durbin R. Fast and accurate long-read alignment with burrows-wheeler transform. *Bioinformatics.* 2010;26:589–95.
51. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, Del Angel G, Rivas MA, Hanna M, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet.* 2011;43:491–8.
52. Zhang P, Zhong K, Zhong Z, Tong H. Genome-wide association study of important agronomic traits within a core collection of rice (*Oryza sativa* L.). *BMC Plant Biol.* 2019;19:259.
53. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira M, Bender D, Maller J, Sklar P, Bakker PI, Daly MJ, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007;81:559–75.
54. Hardy OJ, Vekemans X. SPAGeDi: a versatile computer program to analyze spatial genetic structure at the individual or population levels. *Mol Ecol Notes.* 2002;2:618–20.
55. Goodman SN. Of p-values and Bayes: a modest proposal. *Epidemiology.* 2001;12:295–7.
56. Zhang P, Ding Z, Zhong Z, Tong H. Transcriptomic analysis for *Indica* and *Japonica* rice varieties under aluminum toxicity. *Int J Mol Sci.* 2019;20:997.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

