



The Application of Multi-Locus GWAS for the Detection of Salt-Tolerance Loci in Rice

Yanru Cui, Fan Zhang* and Yongli Zhou*

Institute of Crop Sciences/National Key Facility for Crop Gene Resources and Genetic Improvement, Chinese Academy of Agricultural Sciences, Beijing, China

OPEN ACCESS

Edited by:

Yuan-Ming Zhang,
Huazhong Agricultural University,
China

Reviewed by:

Qishan Wang,
Shanghai Jiao Tong University, China
Chenwu Xu,
Yangzhou University, China
Gao Huijiang,
Chinese Academy of Agricultural
Sciences, China

*Correspondence:

Fan Zhang
zhangfan03@caas.cn
Yongli Zhou
zhouyongli@caas.cn

Specialty section:

This article was submitted to
Plant Breeding,
a section of the journal
Frontiers in Plant Science

Received: 31 July 2018

Accepted: 14 September 2018

Published: 04 October 2018

Citation:

Cui Y, Zhang F and Zhou Y (2018)
The Application of Multi-Locus GWAS
for the Detection of Salt-Tolerance
Loci in Rice. *Front. Plant Sci.* 9:1464.
doi: 10.3389/fpls.2018.01464

Improving the salt-tolerance of direct-seeding rice at the seed germination stage is a major goal of breeders. Efficiently identifying salt tolerance loci will help researchers develop effective rice breeding strategies. In this study, six multi-locus genome-wide association studies (GWAS) methods (mrMLM, FASTmrMLM, FASTmrEMMA, pLARmEB, pKWmEB, and ISIS EM-BLASSO) were applied to identify quantitative trait nucleotides (QTNs) for the salt tolerance traits of 478 rice accessions with 162,529 SNPs at the seed germination stage. Among the 371 QTNs detected by the six methods, 56 were identified by at least three methods. Among these 56 QTNs, 12, 6, 7, 4, 13, 12, and 12 were found to be associated with SSI-GI, SSI-VI, SSI-MGT, SSI-IR-24h, SSI-IR-48h, SSI-GR-5d, and SSI-GR-10d, respectively. Additionally, 66 candidate genes were identified in the vicinity of the 56 QTNs, and two of these genes (LOC_Os01g45760 and LOC_Os10g04860) are involved in auxin biosynthesis according to the enriched GO terms and KEGG pathways. This information will be useful for identifying the genes responsible for rice salt tolerance. A comparison of the six methods revealed that ISIS EM-BLASSO identified the most co-detected QTNs and performed best, with the smallest residual errors and highest computing speed, followed by FASTmrMLM, pLARmEB, mrMLM, pKWmEB, and FASTmrEMMA. Although multi-locus GWAS methods are superior to single-locus GWAS methods, their utility for identifying QTNs may be enhanced by adding a bin analysis to the models or by developing a hybrid method that merges the results from different methods.

Keywords: multi-locus, GWAS, QTNs, salt tolerance, rice

INTRODUCTION

A genome-wide association studies (GWAS) represents a powerful option for the genetic characterization of quantitative traits, and has been widely used for analyzing agronomic traits related to plants. Numerous genetic variants for complex traits have been identified based on single-locus GWAS methods, such as empirical Bayes, efficient mixed model association (EMMA), genome-wide efficient mixed linear model association (GEMMA), settlement of mixed linear

Abbreviations: GI, germination index; GR-10d, germination rate at the 10th day; GR-5d, germination rate at 5th day; IR-24h, imbibition rate at 24 h; IR-48h, imbibition rate at 48 h; MGT, mean germination time; N, normal condition; S, salt stress condition; SSI, stress-susceptibility index; VI, vigor index.

model under progressively exclusive relationship (SUPER), and mixed linear model (MLM) (Kang et al., 2008; Zhou and Stephens, 2012; Wang et al., 2014, 2016a). Although the statistical power of quantitative trait nucleotide (QTN) detection improves after controlling the polygenic background, most of the small effects associated with complex traits are still not captured by single-locus GWAS methods.

In a single-locus GWAS model, markers are tested individually in a one-dimensional genome scan. Moreover, the multiple test correction for the critical value of a significance test should be considered. Bonferroni correction is widely used to modify the threshold value to control the false positive rate (FPR). However, this type of correction method is so conservative that true QTNs may be eliminated. Therefore, the best way to solve this problem is to develop a multi-locus GWAS method that does not require a multiple test correction. Multi-locus GWAS methods involve a multi-dimensional genome scan, in which the effects of all markers are simultaneously estimated. Many penalized multi-locus GWAS methods have been developed, including the least absolute shrinkage and selection operator (LASSO), empirical Bayes LASSO, and adaptive mixed LASSO (Yi and Xu, 2008; Cho et al., 2009, 2010; Wu et al., 2009; Ayers and Cordell, 2010; Wang et al., 2010; Giglio and Brown, 2018). These methods can minimize some marker effects to zero when the number of single nucleotide polymorphisms (SNPs) is not much larger than the sample size. However, the rapid development of sequencing technologies has enabled the detection of many SNPs (i.e., the number of SNPs is hundreds of times larger than the sample size). Thus, the available methods for minimizing marker effects are ineffective. One option for addressing this issue involves decreasing the number of SNPs. Dr. Zhang's lab developed an R package called mrMLM, which includes the following six multi-locus GWAS methods: mrMLM, FASTmrMLM, FASTmrEMMA, pLARM EB, pKWmeB, and ISIS EM-BLASSO. All of these methods involve two-step algorithms. During the first step, a single-locus GWAS method is applied to scan the entire genome, and putative QTNs are detected according to a less stringent critical value, such as $P < 0.005$ or $P < 1/m$, where m is the number of markers. During the second step, all selected putative QTNs are examined by a multi-locus GWAS model to detect true QTNs (Wang et al., 2016a,b; Tamba et al., 2017; Zhang et al., 2017; Ren et al., 2018; Wen et al., 2018a,b; Zhang and Tamba, 2018). The mrMLM package solves the problem associated with co-factor selection in the multi-locus GWAS model when there are many markers.

Rice (*Oryza sativa* L.), which is one of the most important cereal crops worldwide, is sensitive to salt stress. With the increasing salinization of soils, salt stress is becoming a key abiotic factor limiting rice production that rice breeders must overcome (Hu et al., 2012). Developing salt-tolerant rice cultivars is an efficient way to minimize crop loss. Over the past several years, high density SNPs have been used to detect variants with GWAS methods to improve rice varieties (Han and Huang, 2013; Chen et al., 2014; Yang et al., 2014; Wei et al., 2017). However, most traits related to abiotic stress tolerance are controlled by several polygenes that are undetectable in single-locus GWAS models (Lee et al., 2003; Cui et al., 2015). Therefore, we should

apply multi-locus GWAS methods to identify loci related to salt tolerance. In this study, 478 rice accessions, each with seven salt stress susceptibility index (SSI)-related traits, and 162,529 SNPs were used to conduct a multi-locus GWAS. Our objectives were to identify the significant QTNs related to salt tolerance and provide recommendations regarding the selection of a multi-locus GWAS method by comparing the differences among the six multi-locus methods included in the mrMLM package.

MATERIALS AND METHODS

Rice Phenotypic Data Related to Salt Tolerance

We analyzed 478 rice accessions from 46 countries and regions regarding seven salt tolerance-related traits at the seed germination stage in a multi-locus GWAS. Phenotypic data were collected for control and stress-treated plants incubated in a growth chamber, with two independent experiments conducted for the control and stress treatments. Each independent experiment involved a randomized block design with two replicates. The dataset was published by Shi et al. (2017), and the seven salt tolerance-related traits were VI, GI, germination rate (GR) at days 5 and 10, MGT, and imbibition rate (IR) at 24 and 48 h. All salt tolerance-related traits were measured for plants treated with 60 mM NaCl or water (control) as follows: IR (mg/g) was calculated as $IR = (W_2 - W_1)/W_1 \times 1000$ at 24 and 48 h after starting the incubation, where W_1 represents the dry seed weight and W_2 represents the imbibed seed weight; GR was calculated as $GR = N_t/N_0 \times 100\%$ at days 5 and 10, where N_t is the number of germinated seeds at day t and N_0 is the total number of seeds; GI was calculated as $GI = \sum(G_t/T_t)$, where G_t is the accumulated number of germinated seeds at day t and T_t is the time (in days); MGT was calculated as $MGT = \sum T_i N_i / \sum N_i$, where N_i is the number of newly germinated seeds at day t and T_i is the time (in days); VI was calculated as $VI = GI \times SL$, where SL is the average shoot length of 10 germinated seeds at day 10. The salt tolerance level of rice during the germination stage was estimated with the following equation: $SSI = (1 - Y_s/Y_p)/D$, where Y_s is the performance of an individual under the stress condition, Y_p is the performance of an individual under the normal condition, and D is the stress intensity, which was calculated as $D = 1 - (\sum Y_s / \sum Y_p)$. Finally, 21 traits were included in this study. The abbreviated names of these 21 traits are provided in the abbreviations list.

Genotyping and Multi-Locus GWAS

The 478 rice accessions analyzed in this study were from the 3K rice genome project. The 3K rice genome project 404K coreSNP dataset from the Rice-Seek Database was downloaded from http://snp-seek.irri.org/_download.zul (Alexandrov et al., 2015). We used the PLINK program (version 1.9) (Purcell et al., 2007) to obtain a subset of 162,529 SNPs with a minor allele frequency > 5% and a missing data ratio < 0.1 for association analyses. The kinship matrix (K matrix) was calculated based on the genotype marker information described by Xu (2013). The mrMLM package, including six multi-locus GWAS methods,

was downloaded from <http://cran.r-project.org/web/packages/mrMLM/index.html>. Default values were used for all parameters.

Annotation of Candidate Genes and Pathway Enrichment Analysis

Synonymous and non-synonymous SNPs and SNPs associated with large-effect changes were annotated using the snpEff program (version 4.0) (Cingolani et al., 2012) based on the gene models of the annotated Nipponbare reference genome (IRGSP 1.0) (Kawahara et al., 2013). All putative SNPs located within genes and annotation details have been published (Kawahara et al., 2013). Enriched gene ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways were identified using the agriGO (version 2.0) (Tian et al., 2017) and EXPath 2.0 (Chien et al., 2015) programs, respectively.

RESULTS

Heritability and Variance

The heritability and residual errors estimated by the six multi-locus GWAS methods are presented in **Table 1**. The narrow sense heritability ranged from 0.17 for S_MGT and 0.57 for S_IR_48h. A comparison of the residual errors among the six multi-locus GWAS models revealed that the residual error estimated by FASTmrEMMA was the largest under the normal condition when the phenotypic variation was larger than 10. Under the salt stress condition, the largest residual errors for traits S_IR_24h and S_IR_48h were observed from FASTmrEMMA. Regarding the SSI-related traits, the largest residual error was estimated by FASTmrEMMA. The salt tolerance level was evaluated according to the SSI-related traits. Lower SSI values indicated a higher tolerance to salt stress. The results of the correlation analyses of the seven SSI-related traits are presented in **Figure 1A**. There were significant positive correlations among SSI_VI, SSI_GR_5d, SSI_GR_10d, and SSI_GI. The correlation coefficients between SSI-VI and the other three SSI-related traits, namely SSI_GR_5d, SSI_GR_10d, and SSI_GI, were 0.91, 0.91, and 0.96, respectively. Meanwhile, the correlation coefficients for SSI_GR_5d, SSI_GR_10d, and SSI_GI were 0.89, 0.95, and 0.96, respectively. The high correlation among the four SSI-related traits implied that some novel loci might be simultaneously detected for different traits.

QTNs Associated With Salt Tolerance at the Germination Stage Identified by a Multi-Locus GWAS

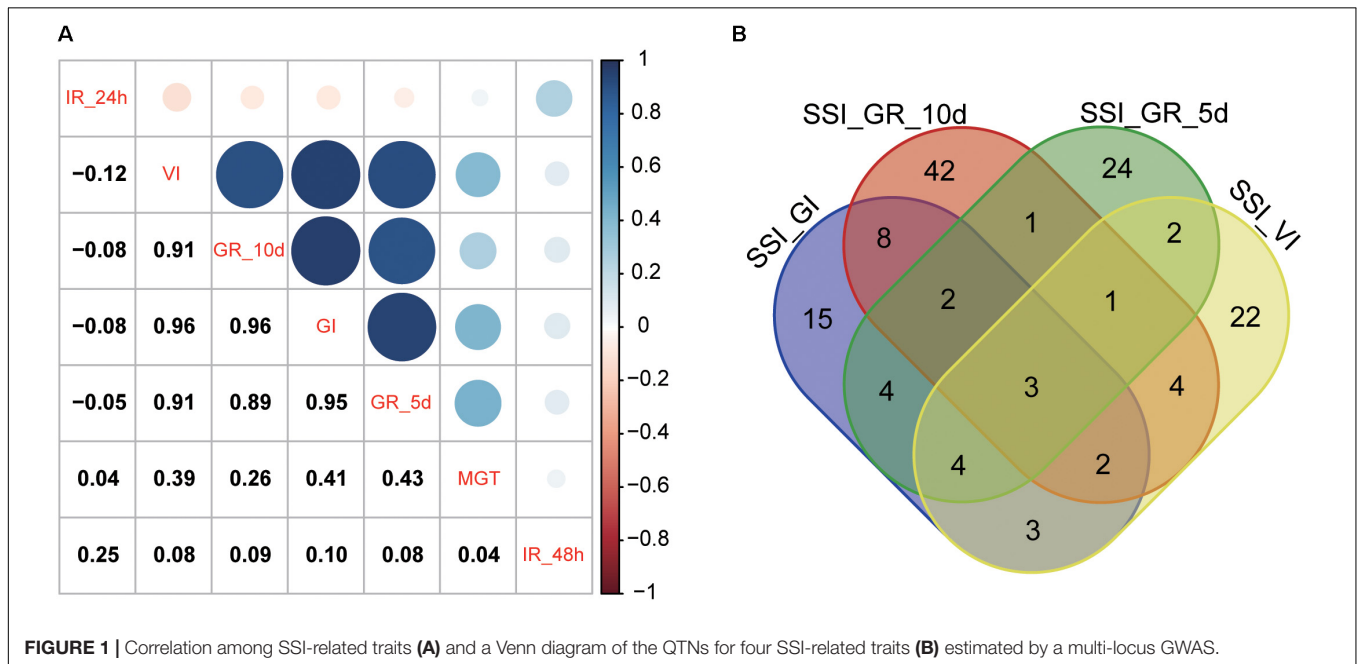
Using the six multi-locus GWAS methods in the mrMLM package (**Supplementary Table S1**), we identified 371 significant QTNs for the salt tolerance-related traits (SSI-VI, SSI-GR, SSI-IR, SSI-MGT, and SSI-GI) based on a logarithm of odds (LOD) threshold of ≥ 3 . Of these QTNs, 41, 41, 27, 63, 56, 41, and 151 were found to be associated with SSI-GI, SSI-VI, SSI-MGT, SSI-IR-24h, SSI-IR-48h, SSI-GR-5d, and SSI-GR-10d, respectively, with the QTNs explaining 0.57 ~ 9.80, 0.54 ~ 8.97, 0.64 ~ 8.21, 0.01 ~ 4.94, 0.37 ~ 8.93, 0.9 ~ 6.72, and 0.7 ~ 6.08

(%) of the phenotypic variations, respectively [i.e., phenotypic variation explained (PVE) values] (**Supplementary Table S1** and **Supplementary Figure S1**). Additionally, 3, 9, and 22 QTNs were associated with four, three, and two salt tolerance-related traits, respectively, which explained the high correlation among SSI_VI, SSI_GR_5d, SSI_GR_10d, and SSI_GI (**Figure 1B**).

In this study, 110 and 56 QTNs were co-detected by at least two and three methods, respectively (**Supplementary Table S2** and **Table 2**). Among the 56 QTNs, 12 that were located on chromosomes 1, 2, 3, 6, 8, 9, 11, and 12 were identified to be associated with SSI-GI, of which 11 were identified by ISIS EM-BLASSO, while 10, 9, 8, 7, and 3 were detected by FASTmrMLM, mrMLM, pKWmEB, pLARmEB, and FASTmrEMMA, respectively. Four of the 12 QTNs were simultaneously detected by five methods. Of these four QTNs, rs3_29294598, rs6_30827714, and rs8_24915626, were simultaneously detected by mrMLM, FASTmrMLM, pLARmEB, pKWmEB, and ISIS EM-BLASSO, with PVE values of 2.45 ~ 5.01, 1.19 ~ 2.82, and 1.44 ~ 4.48 (%), respectively. Meanwhile, rs8_27233581 was simultaneously detected by mrMLM, FASTmrMLM, FASTmrEMMA, pKWmEB, and ISIS EM-BLASSO, with a PVE value of 2.28 ~ 6.28 (%). Six QTNs related to SSI-VI were detected on chromosomes 5, 6, 8, 10, and 11, five of which were identified by mrMLM and pKWmEB, with LOD values of 3.22 ~ 7.16 and 3.11 ~ 7.11, respectively. Only one QTN was detected by ISIS EM-BLASSO, with an LOD value of 8.59. Seven QTNs located on chromosomes 1, 2, 4, 6, 9, and 11 were correlated with SSI-MGT. All seven of these QTNs were detected by ISIS EM-BLASSO and pKWmEB, with LOD values of 3.18 ~ 7.97 and 3.54 ~ 6.62, respectively. The mrMLM, FASTmrMLM, FASTmrEMMA, and pLARmEB methods detected 3, 5, 1, and 2 QTNs related to SSI-MGT, respectively. Among the seven QTNs, rs1_15357371 was identified by all methods, except for mrMLM, with a PVE value of 2.95 ~ 5.64 (%). For SSI-IR-24h, four significant QTNs were detected on chromosomes 4, 6, and 9 by mrMLM, pKWmEB, and ISIS EM-BLASSO, with LOD values of 6.97 ~ 18.97, 3.42 ~ 7.16, and 3.90 ~ 10.18, respectively. Two of these QTNs were identified by FASTmrMLM, while none of the QTNs were detected by FASTmrEMMA and pLARmEB. Thirteen QTNs located on chromosomes 1, 2, 3, 4, 6, 7, 10, 11, and 12 were associated with SSI-IR-48h, including 10 that were detected by ISIS EM-BLASSO, with LOD values of 3.54 ~ 10.0, and nine QTNs that were identified by FASTmrMLM, pLARmEB, and pKWmEB, with LOD values of 3.29 ~ 6.51, 3.58 ~ 6.1, and 5.04 ~ 9.04, respectively. The mrMLM and FASTmrEMMA methods separately detected eight and six QTNs, with LOD values of 3.14 ~ 6.68 and 3.39 ~ 6.97, respectively. Of the 13 QTNs, rs1_5453364, rs11_28865880, and rs12_19111880 were identified by all six methods, with PVE values of 0.86 ~ 2.16, 1.38 ~ 4.83, and 0.62 ~ 2.97 (%), respectively. Moreover, 12 QTNs associated with SSI-GR-5d were detected on chromosomes 1, 3, 5, 7, 8, 9, 10, and 11. Of these QTNs, nine, eight, seven, six, six, and four QTNs were separately detected by pLARmEB, FASTmrMLM, mrMLM, pKWmEB, FASTmrEMMA, and ISIS EM-BLASSO, respectively, with LOD values of 3.26 ~ 7.57, 3.61 ~ 5.96, 3.03 ~ 6.43, 3.34 ~ 6.13,

TABLE 1 | Phenotypic variance, estimated residual error, and heritability of 21 rice traits.

Trait	PV	Heritability (%)	Residual error					
			FASTmrEMMA	FASTmrMLM	ISIS EM-BLASSO	mrMLM	pKWmEB	pLARmEB
S_GI	0.17	42	0.13	0.13	0.11	0.10	0.12	0.13
S_VI	2.64	41	2.11	1.63	1.39	1.55	1.48	1.70
S_MGT	1.03	17	0.85	0.74	0.72	0.86	0.66	0.85
S_IR_24h	53351.6322	51	36552.52	23546.99	22673.23	26406.72	21410.21	24658.42
S_IR_48h	52655.13	56	32734.46	24011.68	21900.10	23275.70	24914.75	19703.57
S_GR_5d	1158.11	34	900.85	935.09	690.32	765.26	746.93	815.23
S_GR_10d	1233.41	35	844.53	961.70	893.90	796.73	719.22	907.11
N_GI	0.07	43	0.05	0.04	0.04	0.04	0.04	0.04
N_VI	5.84	34	3.72	3.10	3.00	3.86	2.72	3.18
N_MGT	0.82	50	0.57	0.48	0.36	0.41	0.43	0.49
N_IR_24h	63786.10	55	41714.56	30057.04	26812.22	28640.95	30489.47	31072.55
N_IR_48h	66260.99	48	44506.99	30736.87	23802.86	34238.02	32400.73	26220.37
N_GR_5d	452.67	27	326.57	307.38	233.09	330.44	267.86	285.44
N_GR_10d	86.61	31	66.58	58.11	48.46	59.51	51.38	52.48
SSI_GI	0.41	33	0.31	0.25	0.22	0.27	0.25	0.28
SSI_VI	0.10	22	0.08	0.06	0.05	0.08	0.06	0.07
SSI_MGT	4.07	32	3.61	3.02	2.57	2.75	2.66	3.22
SSI_IR_24h	199.2613	15	176.41	164.03	143.96	23.17	121.67	173.06
SSI_IR_48h	13.76	42	10.59	8.46	6.62	7.95	7.18	8.67
SSI_GR_5d	0.54	27	0.44	0.40	0.32	0.39	0.30	0.38
SSI_GR_10d	0.4509	34	0.34	0.31	0.23	0.30	0.24	0.28



3.26 ~ 6.57, and 3.09 ~ 5.76, respectively. Three of the 12 QTNs, rs3_4264086, rs5_29609065, and rs11_27392033, were detected by five methods, with PVE values of 1.42 ~ 4.47, 1.07 ~ 4.65, and 0.96 ~ 3.86 (%), respectively. For SSI-GR-10d, 12 QTNs were detected on chromosomes 1, 2, 4, 6, 7, 8, 9, 10, and 11. Of these 12 QTNs, rs10_22754603 and rs11_27380577 were identified by five methods, with

PVE values of 0.93 ~ 3.08 and 1.11 ~ 4.4 (%), respectively (Table 2).

Validation of the Common QTNs

Among the 56 QTNs, 14 were identified by at least five methods, of which four, three, two, four, and one were associated with SSI_GI, SSI_GR_5d, SSI_GR_10d, SSI_IR_48h,

TABLE 2 | Significant QTNs for SSI-related traits in rice co-detected by at least three multi-locus GWAS methods.

Trait	SNPs ¹	Chromosome	Position	QTN effect	LOD score	PVE (%) ²	Method ³
SSI_GI	rs1_11882948	1	11882948	0.1 ~ 0.11	3.67 ~ 4.68	0.98 ~ 1.55	2,3,5
	rs2_22250136	2	22250136	-0.16 ~ -0.09	3.27 ~ 4.43	0.61 ~ 2.19	1,2,3,4
	rs2_24480757	2	24480757	0.08 ~ 0.08	3.66 ~ 4.01	0.98 ~ 1.5	2,3,6
	rs3_29294598	3	29294598	0.08 ~ 0.17	3.97 ~ 6.21	2.45 ~ 5.01	1,2,3,5,6
	rs6_30827714	6	30827714	-0.13 ~ -0.09	3.15 ~ 6.15	1.19 ~ 2.82	1,2,3,5,6
	rs8_7832802	8	7832802	0.1 ~ 0.21	3.48 ~ 4.94	2.17 ~ 4.69	1,3,4,6
	rs8_24915626	8	24915626	0.09 ~ 0.16	3.12 ~ 7.04	1.44 ~ 4.48	1,2,3,5,6
	rs8_25014297	8	25014297	-0.35 ~ -0.21	5.31 ~ 10.4	4.56 ~ 8.91	1,2,3,6
	rs8_27233581	8	27233581	0.1 ~ 0.29	3.71 ~ 7.67	2.28 ~ 6.28	1,2,3,4,6
	rs9_5893568	9	5893568	0.05 ~ 0.08	3.17 ~ 3.43	0.57 ~ 1.09	2,3,5
	rs11_17680260	11	17680260	0.19 ~ 0.26	6.74 ~ 10.53	4.9 ~ 9.8	1,3,5,6
	rs12_21121298	12	21121298	-0.13 ~ -0.09	3.04 ~ 4.35	1.03 ~ 2.06	1,2,5
SSI_VI	rs5_29590002	5	29590002	-0.1 ~ 0.09	3.5 ~ 4.36	1.26 ~ 5.77	1,2,4,5
	rs6_26952785	6	26952785	0.1 ~ 0.11	3.89 ~ 8.59	3.22 ~ 3.57	1,3,6
	rs8_27233581	8	27233581	0.06 ~ 0.13	5.12 ~ 5.45	2.38 ~ 5.17	1,4,5,6
	rs10_2806159	10	2806159	0.08 ~ 0.12	3.16 ~ 5.42	1.85 ~ 3.55	1,2,6
	rs10_11718859	10	11718859	-0.13 ~ 3.56	4.28 ~ 6.36	1.74 ~ 3.2	2,4,5,6
	rs11_17680260	11	17680260	0.07 ~ 0.1	3.12 ~ 5.68	3.03 ~ 7.16	1,5,6
SSI_MGT	rs1_15357371	1	15357371	0.43 ~ 1.37	4.07 ~ 6.54	2.95 ~ 5.64	2,3,4,5,6
	rs2_23991498	2	23991498	0.26 ~ 0.33	3.06 ~ 6.62	1.61 ~ 3.48	2,3,5,6
	rs4_13696726	4	13696726	-0.98 ~ -0.44	3.54 ~ 6.03	1.79 ~ 4.88	1,3,6
	rs6_27962052	6	27962052	-0.43 ~ -0.26	3.75 ~ 4.55	1.06 ~ 4.81	1,2,3,6
	rs9_4258702	9	4258702	-0.53 ~ -0.41	5.71 ~ 7.96	2.76 ~ 3.35	2,3,6
	rs9_11450011	9	11450011	-0.36 ~ -0.31	3.18 ~ 5.53	1.94 ~ 3.55	2,3,6
SSI_IR_24h	rs11_24660808	11	24660808	-0.79 ~ -0.44	3.3 ~ 4.19	2.47 ~ 2.88	1,3,6
	rs4_31794832	4	31794832	2.59 ~ 3.07	3.9 ~ 6.97	0.11 ~ 2.65	1,3,6
	rs6_5699431	6	5699431	2.33 ~ 4.08	3.42 ~ 10.18	0.09 ~ 6.9	1,2,3,6
SSI_IR_48h	rs9_12353804	9	12353804	-4.57 ~ -3.5	4.95 ~ 18.97	0.3 ~ 5.3	1,3,6
	rs9_6746183	9	6746183	-10.84 ~ -4.22	3.29 ~ 16.12	0.97 ~ 5.91	1,2,3,6
	rs1_2103242	1	2103242	0.64 ~ 0.96	3.39 ~ 6.24	1.46 ~ 3.92	1,2,4
SSI_IR_84h	rs1_5453364	1	5453364	-1.26 ~ 0.48	3.23 ~ 6.79	0.86 ~ 3.51	1,2,3,4,5,6
	rs1_31748567	1	31748567	-0.83 ~ -0.56	4.04 ~ 6.43	0.98 ~ 3.55	1,3,5,6
	rs2_24073194	2	24073194	-0.62 ~ 0.72	3.58 ~ 5.04	0.62 ~ 2.2	3,5,6
	rs3_20204466	3	20204466	0.89 ~ 1.02	3.81 ~ 6.73	0.82 ~ 2.76	3,5,6
	rs4_4695323	4	4695323	-1.5 ~ -0.78	3.29 ~ 5.61	1.48 ~ 1.9	2,3,6
	rs4_31202952	4	31202952	-0.67 ~ -0.52	3.43 ~ 5.63	1.61 ~ 2.37	1,2,3
	rs6_1459330	6	1459330	0.62 ~ 1.1	3.34 ~ 5.38	1.51 ~ 1.71	1,2,4
	rs7_21649301	7	21649301	1.08 ~ 1.6	4.4 ~ 6.08	1 ~ 3.78	1,2,5
	rs10_10209541	10	10209541	0.67 ~ 1.03	3.54 ~ 6.61	1.51 ~ 3.98	3,5,6
	rs11_28865880	11	28865880	0.7 ~ 1.65	5.83 ~ 10	1.38 ~ 4.83	1,2,3,4,5,6
	rs12_7176832	12	7176832	-1.43 ~ -0.84	4.43 ~ 8.64	1.2 ~ 4.91	2,3,4,5,6
	rs12_19111880	12	19111880	-1.66 ~ -0.66	3.14 ~ 7.23	0.62 ~ 2.97	1,2,3,4,5,6
SSI_GR_5d	rs1_11882948	1	11882948	0.12 ~ 0.15	3.03 ~ 3.94	1.69 ~ 1.99	1,2,6
	rs1_22648607	1	22648607	0.17 ~ 0.23	4.49 ~ 6.57	1.36 ~ 1.85	3,5
	rs3_4264086	3	4264086	0.11 ~ 0.21	3.97 ~ 6.43	1.42 ~ 4.47	1,2,4,5,6
	rs3_29294598	3	29294598	0.1 ~ 0.11	3.59 ~ 3.61	1.34 ~ 1.72	2,5
	rs5_29609065	5	29609065	0.08 ~ 0.22	3.34 ~ 5.74	0.96 ~ 3.86	1,2,4,5,6
	rs7_1171356	7	1171356	0.15 ~ 0.28	4.48 ~ 4.77	3.26 ~ 3.58	1,4
	rs8_24915626	8	24915626	0.11 ~ 0.15	3.31 ~ 4.1	1.5 ~ 3.01	2,3,5
	rs8_27233581	8	27233581	0.09 ~ 0.28	3.09 ~ 4.66	1.36 ~ 3.33	3,4,6
	rs9_8174432	9	8174432	0.1 ~ 0.15	3.26 ~ 4.23	1.29 ~ 2.84	1,2,5
	rs9_21139613	9	21139613	0.11 ~ 0.19	3.58 ~ 5.02	1.61 ~ 4.79	1,5,6
	rs10_11718859	10	11718859	-0.23 ~ -0.11	4.32 ~ 5.76	1.17 ~ 2.5	2,3,4,5

(Continued)

TABLE 2 | Continued

Trait	SNPs ¹	Chromosome	Position	QTN effect	LOD score	PVE (%) ²	Method ³
SSI_GR_10d	rs11_27392033	11	27392033	-0.31 ~ 0.12	3.93 ~ 6.37	1.07 ~ 4.65	1,2,4,5,6
	rs1_3401561	1	3401561	-0.24 ~ -0.14	3.21 ~ 4.65	2.29 ~ 6.08	1,3,5
	rs1_11882948	1	11882948	0.1 ~ 0.12	3.35 ~ 4.56	1.31 ~ 1.72	2,3,6
	rs2_8009453	2	8009453	0.08 ~ 0.15	3.34 ~ 4.96	1.18 ~ 3.31	1,2,3
	rs2_22247315	2	22247315	-0.17 ~ -0.12	3.54 ~ 6.51	0.95 ~ 3.19	3,5,6
	rs4_19568498	4	19568498	0.16 ~ 0.21	5.39 ~ 6.37	1.9 ~ 2.96	2,3,6
	rs6_26597879	6	26597879	0.11 ~ 0.13	3.83 ~ 5.41	2.44 ~ 2.82	2,3,5
	rs7_3788168	7	3788168	0.22 ~ 0.23	5.46 ~ 7.54	2.54 ~ 3.59	2,3,6
	rs7_22276671	7	22276671	0.09 ~ 0.1	3.65 ~ 4.66	1.57 ~ 2.33	2,3,6
	rs8_24915626	8	24915626	0.13 ~ 0.14	3.88 ~ 5.8	2.07 ~ 3.16	1,2,3
	rs9_21139613	9	21139613	0.1 ~ 0.18	3.55 ~ 8.69	1.92 ~ 4.59	1,2,5,6
rs10_22754603	10	22754603	0.08 ~ 0.17	3.46 ~ 5.94	0.93 ~ 3.08	1,2,3,5,6	
rs11_27380577	11	27380577	-0.21 ~ -0.13	3.24 ~ 7.31	1.11 ~ 4.4	1,2,3,5,6	

¹SNPs in bold font are pleiotropic QTNs which were detected associate with multiple traits.

²PVE: Phenotypic variation explained.

³1:mrMLM; 2:FASTmrMLM; 3:ISIS EM-BLASSO; 4:FASTmrEMMA; 5:pLARmEB; 6:pKWmEB.

and SSI_MGT, respectively. We divided the population into two groups according to allelic genotypes to test whether the mean phenotypes of the two groups were significantly different. The mean value of the group carrying the favorable allele was less than that of the other group (Figure 2).

GO and KEGG Pathway Enrichment Analyses

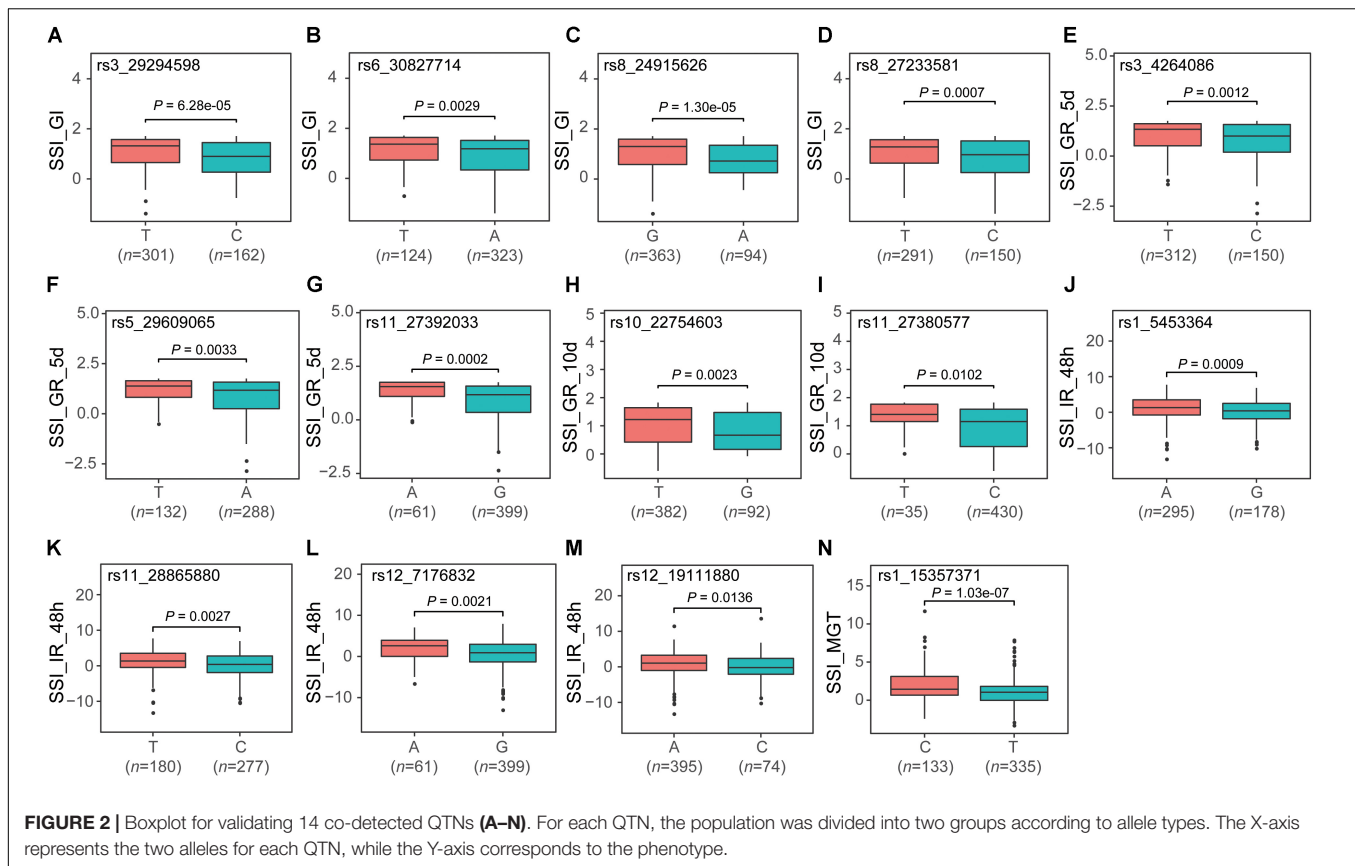
According to the Nipponbare reference genome, the 371 identified QTNs for traits related to salt tolerance were part of or were adjacent to 581 genes (Supplementary Table S1). These genes were significantly enriched for GO biological processes related to the plant lipid metabolic process and transmembrane transport process (Supplementary Table S3). They were also significantly enriched for the plant tryptophan metabolism pathway ($P < 0.03$). Moreover, two genes (LOC_Os01g45760 and LOC_Os10g04860) were associated with auxin biosynthesis. A total of 66 genes were identified around the 56 QTNs based on the enriched GO terms and KEGG pathways as well as the functional annotations (Supplementary Table S4). This information may be very useful for identifying the genes responsible for salt tolerance in rice.

DISCUSSION

Multi-locus GWAS models, which are relatively close to the true genetic models of plants and animals, are superior to single-locus GWAS models because of their higher statistical power and lower FPR (Segura et al., 2012; Wang et al., 2016a). These models were developed by geneticists, who added the polygenic effect and population structure to the single-locus GWAS model to decrease the bias in effect estimations by controlling the genetic background (Zhang et al., 2005; Yu et al., 2006; Zhang et al., 2010). Although advancements in the single-locus GWAS models have improved the detection accuracy to some extent, the multiple test correction for the threshold value of the significance

test in single-locus models (e.g., Bonferroni correction) is too stringent to capture all true QTNs. Another unavoidable problem is that single-locus GWAS methods are inappropriate when the target traits are controlled by a series of polygenes. In this study, 478 rice accessions with 162,529 SNPs were used to identify QTNs for traits related to salt tolerance based on six multi-locus GWAS methods. We compared the QTNs identified by the multi-locus GWAS methods in our study with the previously reported QTNs detected by the efficient mixed-model EMMA eXpedited (EMMAX) program comprising a single-locus GWAS method. The comparison revealed that four of the previously reported six QTNs related to SSI-VI were detected by a multi-locus GWAS, and two QTNs associated with SSI-MGT overlapped with the previously reported QTNs. Additionally, 12, 4, 13, 12, and 12 QTNs separately associated with SSI-GI, SSI-IR-24h, SSI-IR-48h, SSI-GR-5d, and SSI-GR-10d, respectively, were simultaneously detected by at least three multi-locus GWAS methods. In contrast, none of the QTNs associated with the five traits were identified by a single-locus GWAS method. These observations were as expected, and can be explained by the following two points: (i) salt tolerance is a quantitative genetic characteristic that is controlled by multiple genes with small effects, which are difficult to detect in a single-locus GWAS model (Wang et al., 2011; Kumar et al., 2015); (ii) some true QTNs for traits related to salt tolerance are missed by a single-locus GWAS model because of an overly conservative critical value. Furthermore, our results suggest that a multi-locus GWAS model may be useful for detecting loci with small effects.

In this study, we used six multi-locus GWAS methods included in the mrMLM package to detect QTNs. The six methods involve two-step algorithms, and marker effects are treated as random effects in each method. However, each method has its own characteristics. We observed that mrMLM detected the most QTNs (Supplementary Table S1), but this method has one shortcoming. When the number of putative QTNs is much larger than the sample size, the multi-locus model in this method will be over-fitted. The residual error estimated by mrMLM was



much smaller than that estimated by the five other methods (Table 1). During the first step, 7,588 QTNs with a threshold value $P < 0.01$ were selected, which is 16 times larger than the sample size. Over-fitting may occur when too many variables are added to a multi-locus model. This issue was solved by using FASTmrMLM, in which the least angle regression (LARS) algorithm is implemented between the first single-locus scanning step and the EM-Empirical Bayes estimation in the second step. The LARS algorithm (Efron et al., 2004) is a flexible method for selecting variables, and can be applied in the lars package¹. In this method, $n-1$ variables (n is the number of samples), which are most likely associated with the target traits, are added to the multi-locus model.

The FASTmrEMMA method detected the fewest QTNs. This method involves an approximation algorithm in which the covariance matrix of the polygenic matrix K and environmental noise are whitened by a matrix transformation to increase the computing speed. In the pLARmEB method, the same transformed model as that used in FASTmrEMMA is implemented to control the polygenic background, and the LARS algorithm is applied to select potential SNPs related to the target trait for the subsequent multi-locus GWAS detection. Among the six multi-locus GWAS methods, ISIS-EM-BLASSO had the shortest running time and the smallest estimated residual errors (Supplementary Figure S2 and Table 1). In the first step of

this method, an iterative-modified sure independence screening (ISIS) approach is used to decrease the number of SNPs to a moderate level, after which the Expectation-Maximization (EM)-Bayesian least absolute shrinkage and selection operator (BLASSO) is used to estimate all of the selected SNP effects to detect true QTNs. The last method, pKWmEB, is a non-parametric method, in which a Kruskal–Wallis test and the LARS algorithm are used to identify potential SNPs. All identified markers are added to the multi-locus model to detect true QTNs.

The two-step multi-locus GWAS methods included in this study significantly improved the statistical power and decreased the FPR. Moreover, ISIS EM-BLASSO identified the most co-detected QTNs, followed by pKWmEB, while FASTmrEMMA identified the fewest QTNs (Table 2). Additionally, ISIS EM-BLASSO performed best, with the smallest estimated residual errors and highest computing speed. However, selecting an appropriate critical value is still problematic for the two-step multi-locus GWAS model. A threshold value that is too stringent will lead to the omission of loci information, whereas a relaxed threshold value will result in numerous loci being selected, which may lead to the over-fitting of multi-locus models. A simple solution to this problem involves developing a hybrid method that combines the results from different methods. Directly decreasing the number of SNPs instead of applying a single-locus GWAS scanning step represents another potential solution. We are currently developing a new bin analysis method that can be applied to any type of population. In the bin analysis method,

¹<http://cran.r-project.org/web/packages/lars/>

the number of markers is decreased, but the information for all markers is fully retained. Adding a bin analysis to the multi-locus GWAS model represents a new option.

CONCLUSION

In this study, six multi-locus GWAS methods were used to detect loci related to rice salt tolerance at the seed germination stage. A total of 371 QTNs were identified, with 56 QTNs co-detected by at least three methods. Moreover, 66 genes were identified in the vicinity of the 56 QTNs based on functional annotations. Two of these genes (LOC_Os01g45760 and LOC_Os10g04860) are involved in auxin biosynthesis according to the enriched GO terms and KEGG pathways. These observations may be useful for identifying the genes responsible for rice salt tolerance.

AUTHOR CONTRIBUTIONS

YC drafted the manuscript. FZ and YC analyzed the data. YZ and FZ conceived the study and were in charge of the direction and planning. All authors read and approved the final version of this manuscript.

FUNDING

This research was supported by grants from the National Key Research and Development Program of China (Project No. 2016YFD0100101), the National Natural Science Foundation of China (Grant No. 31771762), the National High-tech

Program of China (No. 2014AA10A603), the Shenzhen Peacock Plan (No. 20130415095710361), the Bill & Melinda Gates Foundation (OPP1130530), and the CAAS Agricultural Science and Technology Innovative Program. YC was supported by the Beachell-Borlaug International Student Fellowship from Monsanto.

ACKNOWLEDGMENTS

We thank Liwen Bianji, Edanz Editing China (www.liwenbianji.cn/ac) for editing the English text of a draft of this manuscript.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpls.2018.01464/full#supplementary-material>

FIGURE S1 | Bar plot of the number of QTNs associated with seven salt tolerance traits detected by different methods.

FIGURE S2 | Computing time of the six multi-locus GWAS methods.

TABLE S1 | Significant salt tolerance QTNs detected by six multi-locus GWAS methods.

TABLE S2 | Significant QTNs detected by at least two multi-locus GWAS methods.

TABLE S3 | Results of a GO enrichment analysis.

TABLE S4 | Gene annotations for the 56 significant QTNs associated with salt tolerance traits.

REFERENCES

- Alexandrov, N., Tai, S., Wang, W., Mansueto, L., Palis, K., Fuentes, R. R., et al. (2015). SNP-seek database of SNPs derived from 3000 rice genomes. *Nucleic Acids Res.* 43(Database issue), D1023–D1027. doi: 10.1093/nar/gku1039
- Ayers, K. L., and Cordell, H. J. (2010). SNP selection in genome-wide and candidate gene studies via penalized logistic regression. *Genet. Epidemiol.* 34, 879–891. doi: 10.1002/gepi.20543
- Chen, W., Gao, Y. Q., Xie, W. B., Gong, L., Lu, K., Wang, W. S., et al. (2014). Genome-wide association analyses provide genetic and biochemical insights into natural variation in rice metabolism. *Nat. Genet.* 46, 714–721. doi: 10.1038/ng.3007
- Chien, C., Chow, C., Wu, N., Chiang-Hsieh, Y., Hou, P., and Chang, W. (2015). EXPath: a database of comparative expression analysis inferring metabolic pathways for plants. *BMC Genomics* 16(Suppl. 2):S6. doi: 10.1186/1471-2164-16-S2-S6
- Cho, S., Kim, H., Oh, S., Kim, K., and Park, T. (2009). Elastic-net regularization approaches for genome-wide association studies of rheumatoid arthritis. *BMC Proc.* 3(Suppl. 7):S25. doi: 10.1186/1753-6561-3-s7-s25
- Cho, S., Kim, K., Kim, Y. J., Lee, J. K., Cho, Y. S., Lee, J. Y., et al. (2010). Joint identification of multiple genetic variants via elastic-net variable selection in a genome-wide association analysis. *Annu. Hum. Genet.* 74, 416–428. doi: 10.1111/j.1469-1809.2010.00597
- Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., et al. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly* 6, 80–92. doi: 10.4161/fly.19695
- Cui, Y. R., Zhang, F., Xu, J. L., Li, Z. K., and Xu, S. Z. (2015). Mapping quantitative trait loci in selected breeding populations: a segregation distortion approach. *Heredity* 115, 538–546. doi: 10.1038/hdy.2015.56
- Efron, B., Hastie, T., Johnstone, I., and Tibshirani, R. (2004). Least angle regression. *Ann. Stat.* 32, 407–451. doi: 10.1214/009053604000000067
- Giglio, C., and Brown, S. D. (2018). Using elastic net regression to perform spectrally relevant variable selection. *J. Chemom.* 32:e3034. doi: 10.1002/cem.3034
- Han, B., and Huang, X. H. (2013). Sequencing-based genome-wide association study in rice. *Curr. Opin. Plant Biol.* 16, 133–138. doi: 10.1016/j.pbi.2013.03.006
- Hu, S. K., Tao, H. J., Qian, Q., and Guo, L. B. (2012). Genetics and molecular breeding for salt tolerance in rice. *Rice Genom. Genet.* 3, 39–49.
- Kang, H. M., Zaitlen, N. A., Wade, C. M., Kirby, A., Heckerman, D., Daly, M. J., et al. (2008). Efficient control of population structure in model organism association mapping. *Genetics* 178, 1709–1723. doi: 10.1534/genetics.107.080101
- Kawahara, Y., de la Bastide, M., Hamilton, J. P., Kanamori, H., McCombie, W. R., Ouyang, S., et al. (2013). Improvement of the *Oryza sativa* nipponbare reference genome using next generation sequence and optical map data. *Rice* 6:4. doi: 10.1186/1939-8433-6-4
- Kumar, V., Singh, A., Mithra, S., Krishnamurthy, S., Parida, S., and Jain, S. (2015). Genome-wide association mapping of salinity tolerance in rice (*Oryza sativa*). *DNA Res.* 22, 133–145. doi: 10.1093/dnares/dsu046
- Lee, K., Choi, W., Ko, J., Kim, T., and Gregorio, G. (2003). Salinity tolerance of japonica and indica rice (*Oryza sativa* L.) at the seedling stage. *Planta* 216, 1043–1046. doi: 10.1007/s00425-002-0958-3

- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., et al. (2007). Plink: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 81, 559–575. doi: 10.1086/519795
- Ren, W. L., Wen, Y. J., Dunwell, J. M., and Zhang, Y. M. (2018). pKWmEB: integration of Kruskal–Wallis test with empirical Bayes under polygenic background control for multi-locus genome-wide association study. *Heredity* 120, 208–218. doi: 10.1038/s41437-017-0007-4
- Segura, V., Vilhjalmsón, B. J., Platt, A., Korte, A., Seren, U., Long, Q., et al. (2012). An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. *Nat. Genet.* 44, 825–830. doi: 10.1038/ng.2314
- Shi, Y. R., Gao, L. L., Wu, Z. C., Zhang, X. J., Wang, M. M., Zhang, C. S., et al. (2017). Genome-wide association study of salt tolerance at the seed germination stage in rice. *BMC Plant Biol.* 17:92. doi: 10.1186/s12870-017-1044-0
- Tamba, C. L., Ni, Y. L., and Zhang, Y. M. (2017). Iterative sure independence screening EM-Bayesian LASSO algorithm for multi-locus genome-wide association studies. *PLoS Comput. Biol.* 13:e1005357. doi: 10.1371/journal.pcbi.1005357
- Tian, T., Liu, Y., Yan, H., You, Q., Yi, X., Du, Z., et al. (2017). agriGO v2.0: a GO analysis toolkit for the agricultural community. *Nucleic Acids Res.* 45, W122–W129. doi: 10.1093/nar/gkx382
- Wang, D., Eskridge, K. M., and Crossa, J. (2010). Identifying QTLs and epistasis in structured plant populations using adaptive mixed LASSO. *J. Agric. Biol. Environ. Stat.* 16, 170–184. doi: 10.1007/s13253-010-0046-2
- Wang, Q., Tian, F., Pan, Y., Buckler, E. S., and Zhang, Z. (2014). A super powerful method for genome wide association study. *PLoS One* 9:e107684. doi: 10.1371/journal.pone.0107684
- Wang, S. B., Feng, J. Y., Ren, W. L., Huang, B., Zhou, L., Wen, Y. J., et al. (2016a). Improving power and accuracy of genome-wide association studies via a multi-locus mixed linear model methodology. *Sci. Rep.* 6:19444. doi.org/10.1038/srep19444 doi: 10.1038/srep19444
- Wang, S. B., Wen, Y. J., Ren, W. L., Ni, Y. L., Zhang, J., Feng, J. Y., et al. (2016b). Mapping small-effect and linked quantitative trait loci for complex traits in backcross or DH populations via a multi-locus GWAS methodology. *Sci. Rep.* 6:29951. doi: 10.1038/srep29951
- Wang, Z., Wang, J., Bao, Y., Wu, Y., and Zhang, H. (2011). Quantitative trait loci controlling rice seed germination under salt stress. *Euphytica* 178, 297–307. doi: 10.1007/s10681-010-0287-8
- Wei, J. L., Wang, A. G., Li, R. D., Qu, H., and Jia, Z. Y. (2017). Metabolome-wide association studies for agronomic traits of rice. *Heredity* 120, 342–355. doi: 10.1038/s41437-017-0032-3
- Wen, Y. J., Zhang, H., Ni, Y. L., Huang, B., Zhang, J., Feng, J. Y., et al. (2018a). Methodological implementation of mixed linear models in multi-locus genome-wide association studies. *Brief. Bioinform.* 19, 700–712. doi: 10.1093/bib/bbw145
- Wen, Y. J., Zhang, Y. W., Zhang, J., Feng, J. Y., Dunwell, J. M., and Zhang, Y. M. (2018b). An efficient multi-locus mixed model framework for the detection of small and linked QTLs in F2. *Brief. Bioinform.* bby058. doi: 10.1093/bib/bby058
- Wu, T. T., Chen, Y. F., Hastie, T., Sobel, E., and Lange, K. (2009). Genome-wide association analysis by lasso penalized logistic regression. *Bioinformatics* 25, 714–721. doi: 10.1093/bioinformatics/btp041
- Xu, S. (2013). Mapping quantitative trait loci by controlling polygenic background effects. *Genetics* 195, 1209–1222. doi: 10.1534/genetics.113.157032
- Yang, W., Guo, Z., Huang, C., Duan, L., Chen, G., Jiang, N., et al. (2014). Combining highthroughput phenotyping and genome-wide association studies to reveal natural genetic variation in rice. *Nat. Commun.* 5:5087. doi: 10.1038/ncomms6087
- Yi, N., and Xu, S. (2008). Bayesian LASSO for quantitative trait loci mapping. *Genetics* 179, 1045–1055. doi: 10.1534/genetics.107.085589
- Yu, J., Pressoir, G., Briggs, W. H., Bi, I. V., Yamasaki, M., and Doebley, J. F. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat. Genet.* 38, 203–208. doi: 10.1038/ng1702
- Zhang, J., Feng, J. Y., Ni, Y. L., Wen, Y. J., Niu, Y., Tamba, C. L., et al. (2017). pLARM EB: integration of least angle regression with empirical Bayes for multilocus genome-wide association studies. *Heredity* 118, 517–524. doi: 10.1038/hdy.2017.8
- Zhang, Y. M., Mao, Y., Xie, C., Smith, H., Luo, L., and Xu, S. (2005). Mapping quantitative trait loci using naturally occurring genetic variance among commercial inbred lines of maize. *Genetics* 169, 2267–2275. doi: 10.1534/genetics.104.033217
- Zhang, Y. M., and Tamba, C. L. (2018). A fast mrMLM algorithm for multi-locus genome-wide association studies. *bioRxiv* [Preprint]. doi: 10.1101/341784
- Zhang, Z., Ersoz, E., Lai, C. -Q., Todhunter, R. J., Tiwari, H. K., and Gore, M. A. (2010). Mixed linear model approach adapted for genome-wide association studies. *Nat. Genet.* 42, 355–360. doi: 10.1038/ng.546
- Zhou, X., and Stephens, M. (2012). Genome-wide efficient mixed-model analysis for association studies. *Nat. Genet.* 44, 821–824. doi: 10.1038/ng.2310

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Cui, Zhang and Zhou. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.