Multi-locus genome-wide association studies for five yield-related traits in rice

Hua Zhong[1], Shuai Liu[2], Tong Sun[1], Weilong Kong[1], Xiaoxiao Deng[1], Zhaohua Peng[2] and Yangsheng Li[1*]

Abstract

Background: Improving the overall production of rice with high quality is a major target of breeders. Mining potential yield-related loci have been geared towards developing efficient rice breeding strategies. In this study, one single-locus genome-wide association studies (SL-GWAS) method (MLM) in conjunction with five multi-locus genome-wide association studies (ML-GWAS) approaches (mrMLM, FASTmrMLM, pLARmEB, pKWmEB, and ISIS EM-BLASSO) were conducted in a panel consisting of 529 rice core varieties with 607,201 SNPs.

Results: A total of 152, 106, 12, 111, and 64 SNPs were detected by the MLM model associated with the five yield-related traits, namely grain length (GL), grain width (GW), grain thickness (GT), thousand-grain weight (TGW), and yield per plant (YPP), respectively. Furthermore, 74 significant quantitative trait nucleotides (QTNs) were presented across at least two ML-GWAS methods to be associated with the above five traits successively. Finally, 20 common QTNs were simultaneously discovered by both SL-GWAS and ML-GWAS methods. Based on genome annotation, gene expression analysis, and previous studies, two candidate key genes (*LOC_Os09g02830* and *LOC_Os07g31450*) were characterized to affect GW and TGW, separately.

Conclusions: These outcomes will provide an indication for breeding high-yielding rice varieties in the immediate future.

Keywords: *Oryza sativa*, Yield, ML-GWAS, MLM, QTNs

Background

Rice (*Oryza sativa* L.) is one of the three major food crops supporting more than 50% of the whole population worldwide [1]. In 2018, the total rice output accounted for 32.24% of the total grain production in China followed by maize (http://www.stats.gov.cn/). While in the world, the average production of rice from 1994 to 2019 is 654.78 million tonnes per year, accounting for 27.28% of total cereals output (http://www.fao.org/faostat/en/#data/QC/visualize). The growing global population and the deteriorating environment issue new challenges to the breeding of high-yielding crops [2]. Rice yield is a complex quantitative agronomic trait multiplicatively governed by three major components as the number of grains per panicle, thousand-grain weight, and the number of panicles per plant [3]. Besides, grain size (including grain length, width, and thickness) are also closely related to rice productivity [4]. A previous study reported that rice yield was a representative quantitative trait regulated by several minor genes, and a more efficient tool was needed to develop for exploiting these minor QTLs [5]. Many genes have been reported controlling grain size, grain number, and yield. For example, the *GS3* [6] is a major gene controlling rice grain length (GL). A mutation in the second exon changes a cysteine codon (TGC) to a termination codon (TGA) at the protein level, resulting in a diversity of rice GL. The *GW5* [7] is an IQ calmodulin-binding motif family protein, which regulates rice grain width and weight. Loss-of-function *gw5* showed wider grain compared to the wild type. The *WTG1/ OsOTUB1* [8] encodes an otubain-like protease with deubiquitination activity, which expresses in

* Correspondence: lysh2001@whu.edu.cn
[1]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, People's Republic of China 430072
Full list of author information is available at the end of the article

developing grains and panicles. The overexpression of *WTG1* showed narrow, thin, and long rice grains as a result of slim cells. The *OsSPL13* is a SQUAMOSA promoter-binding-like protein [9, 10], which was reported controlling rice grain length, grain number, grain size, and yield.

The genome-wide association study (GWAS) has become a powerful tool for mining QTL associated with complex traits [11, 12]. Single-locus GWAS (SL-GWAS) methods such as mixed linear model (MLM) [13], efficient mixed-model association eXpedited (EMMAX) [14], and factored spectrally transformed linear mixed models (FaST-LMM) [15] have been widely used to investigate tremendous genetic variants for agronomic traits. However, these SL-GWAS methods are limited in detecting marginal effects QTNs influenced by the polygenic background and stringent Bonferroni correction [16].

To address the shortcomings of SL-GWAS, multi-locus GWAS (ML-GWAS) has been developed as a multi-dimensional genome scan method in which the effects of all markers are estimated at the same time [17]. In particular, to solve the problem associated with co-factor selection in the ML-GWAS model when there are many markers, the mrMLM package was proposed, which containing the following six ML-GWAS methodologies: mrMLM (multi-locus random-SNP-effect MLM) [16], FASTmrMLM (fast mrMLM) [18], ISIS EM-BLASSO (iterative modified-sure independence screening expectation-maximization-Bayesian least absolute shrinkage and selection operator) [19], pKWmEB (integration of Kruskal-Wallis test with empirical Bayes) [20], FASTmrEMMA (fast multi-locus random-SNP-effect efficient mixed model analysis) [21], and pLARmEB (polygenic-background-control-based least angle regression plus empirical Bayes) [22]. ML-GWAS also has a lower false-positive rate and has been applied successfully to identify significant QTNs with subtle contributions for several agronomic [23–26]. But no studies have focused on ML-GWAS for yield-related traits in rice as yet. In general, the QTL (Quantitative Trait Loci) refers to the signal identified by single-locus methods, such as GLM, MLM, *etc*. In such QTL, it mostly contained numerous associated SNPs (Single nucleotide polymorphisms). While in multi-locus GWAS methods, when all the potentially associated markers were identified in the first step. These markers were submitted into a model in further analysis and true QTNs (Quantitative Trait Nucleotides) were further confirmed by the likelihood ratio test [16].

In the current study, a large-scale natural population of 529 rice accessions with five yield-related traits and 607,201 SNPs was conducted by a hybrid method of one SL-GWAS (MLM) algorithm and five ML-GWAS (mrMLM, FASTmrEMMA, pLARmEB, pKWmEB, and ISIS EM-BLASSO) models. We aim to investigate common QTNs via multiple methodologies and then deduce potential candidate genes to accelerate molecular marker-assisted breeding and boost rice production.

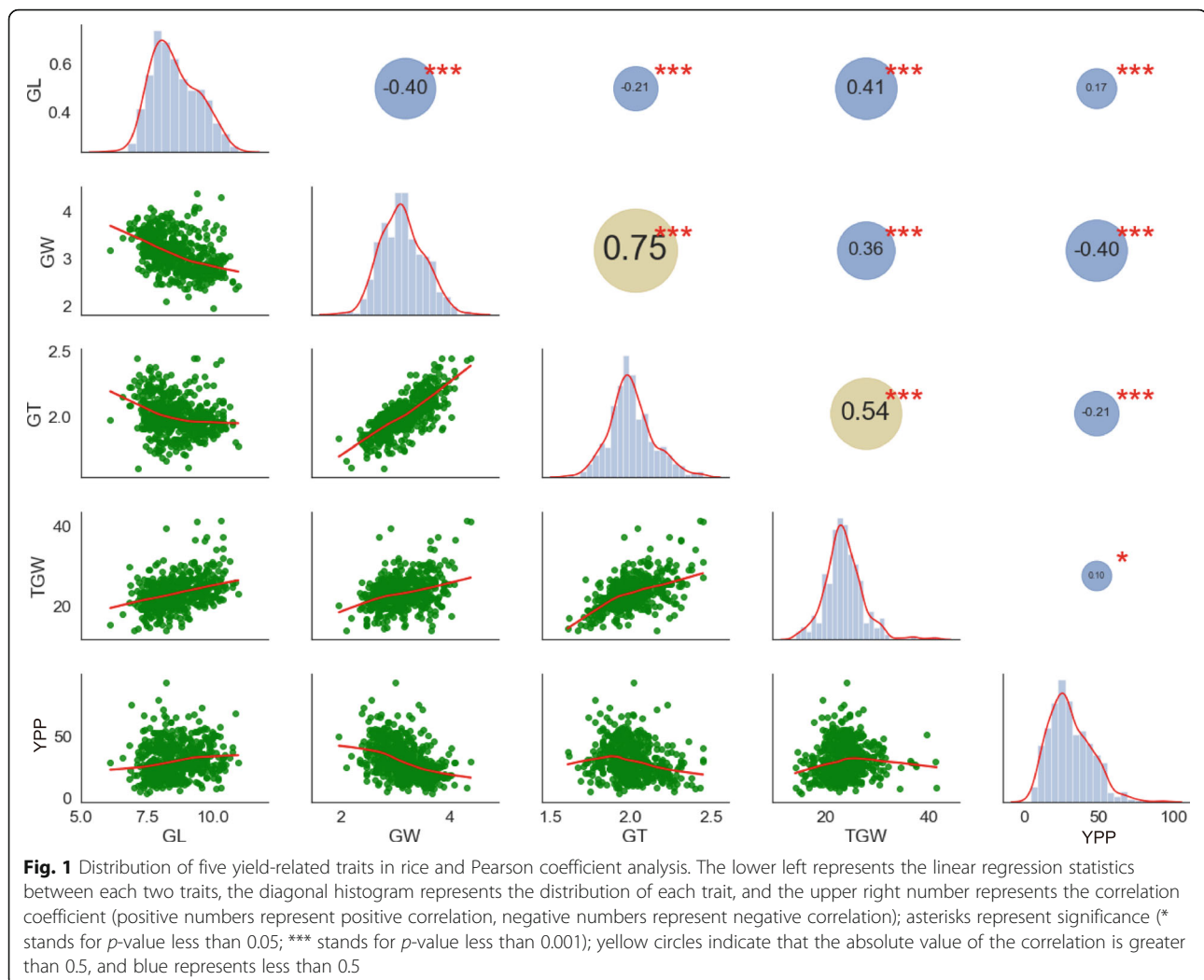## Results

Phenotypic variation

Five yield-related traits (including GL, GW, GT, TGW, and YPP) were selected to examine whether significant phenotypic variances exist in the yield among the 529 rice varieties. The results manifested that the parameters were varied for accessions to their corresponding traits (Table S1). For instance, the GL ranged from 6.13 to 10.97 mm, with a mean of 8.57 mm. The YPP had a great variation ranging from 4.41 to 92.66 g, whereas the GT possessed the smallest range from 1.61 to 2.45 mm with a CV of 6.86%. The CV of GL, GW, TGW, and YPP were 10.25, 12.36, 16.09, and 44.90%, respectively. Also, the frequency distributions of all five traits obeyed approximately the normal distributions.

Furthermore, the Pearson correlation coefficients (PCC) among the five traits were also estimated (Fig. 1). All paired traits showed statistically significant differences at the p-value< 0.001 except the relationship between TGW and YPP (p-value< 0.05). GL and GW had a negative relation with PPC = − 0.40, which was corresponding with a previous study [27]. GL was also associated negatively with GT (PCC = -0.21) while positively with TGW (PCC = 0.41) and YPP (PCC = 0.17), respectively. In addition, TGW was observed to positively correlate to GT (PCC = 0.54), GL (PCC = 0.41), and GW (PCC = 0.36), indicating grain size might make a major contribution to grain weight. These results exhibited that there was a close relationship among the five rice traits, which played an important role in regulating the rice grain shape and productivity.

Population structure and linkage disequilibrium analysis

To understand the population structure of the panel, PCA analysis was performed using 607,201 SNPs, which was mentioned in the Methods section. Five conceivable subpopulations were respectively distinguished via PC1, PC2, and PC3 (Fig. 2a and b). Next, a maximum likelihood phylogenetic tree was analysed by their genetic distances, which was derived from the SNP differences in these genotypes. The population could be divided into six distinct subgroups, 95 indica I (IndI), 74 indica II (IndII), 43 tropical japonica (TrJ), 93 temperate japonica (TeJ), 46 aus, and 178 admixture of the others (Adm), respectively (Fig. 2c). Based on the results from both the phylogenetic tree and PCA, the panel was separated into six groups.

The LD decay distance was further estimated using the identified SNPs. As delineated in Fig. 2d, the genome-wide LD decay rate of all populations was approximately 43 kb, where the r^2 dropped to half of the

Fig. 1 Distribution of five yield-related traits in rice and Pearson coefficient analysis. The lower left represents the linear regression statistics between each two traits, the diagonal histogram represents the distribution of each trait, and the upper right number represents the correlation coefficient (positive numbers represent positive correlation, negative numbers represent negative correlation); asterisks represent significance (* stands for *p*-value less than 0.05; *** stands for *p*-value less than 0.001); yellow circles indicate that the absolute value of the correlation is greater than 0.5, and blue represents less than 0.5

maximum value. Due to that, the theoretical average marker density was one SNP per 9 kb. In fact, the actual value of SNP density in the genome already reached 0.6 (kb/SNP) as the distribution of these SNPs within the whole genome was summarized in Fig. S1. Therefore, we concluded that these markers were sufficiently dense for detecting the associated QTNs.

SL-GWAS and ML-GWAS analyses
All five yield-related traits were analyzed using one SL-GWAS (MLM) to identify QTL and five ML-GWAS (mrMLM, FASTmrMLM, pLARmEB, pKWmEB, and ISIS EM-BLASSO) methods to identify QTNs (Fig. S2). As for MLM, 152, 106, 12, 111, and 64 SNPs corresponding to 3, 19, 8, 53, and 56 QTLs were found to be tightly associated with GL, GW, GT, TGW, and YPP under the cut-off criterion of *p*-value $=1.65 \times 10^{-6}$, respectively (Table S2). As figured in QQ plots (Fig. S2f and 2i), the curves of GL and TGW were consistent with

optimal trends, implying that the false-positive errors were controlled well and the results of the MLM model were reliable.

A total of 74 significant QTNs (LOD $\geq$ 3) were simultaneously defined to be associated with the above five objective traits by at least two ML-GWAS methods (Table 1). Among these QTNs, 19, 9, 7, 22, and 17 were found to be associated with GL, GW, GT, TGW, and YPP, respectively. A total of 22 correlated QTNs were distinguished for TGW, which were widely located on all 12 chromosomes. For GW, 9 candidate QTNs were distributed on chromosomes 3, 4, 5, 9, 10, and 12. A total of 17 QTN hotspots were detected significantly related to YPP, spread over 2, 3, 4, 5, 6, 7, 8, 10, 11, and 12 chromosomes. Of these, six QTNs were found simultaneously using at least three ML-GWAS methods (*qYPP-3-1*, *qYPP-4-2*, *qYPP-5-2*, *qYPP-7-2*, *qYPP-10-1*, and *qYPP-10-2*). Notably, *qYPP-7-2* was determined across all five ML-GWAS approaches, explaining the
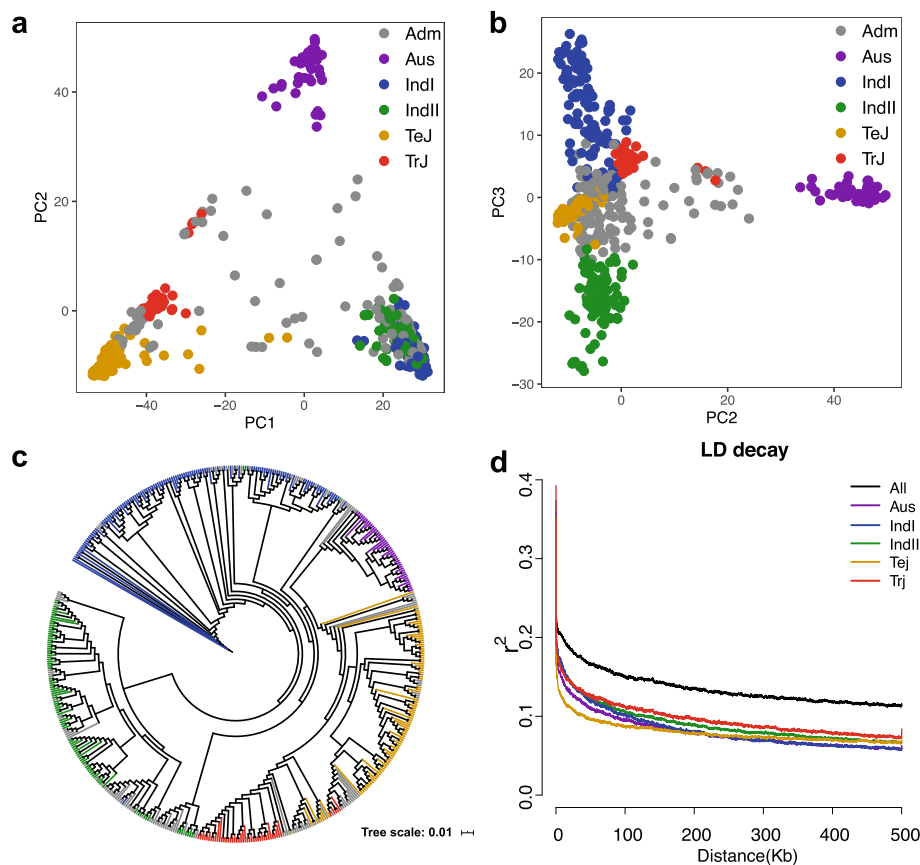
Fig. 2 Genetic structure of the 529 rice panel. (**a** - **b**) PCA plots of the 529 rice core varieties. PCA plots present the genetic variation in the rice accessions with PC1 and PC2, PC2 and PC3, separately. (**c**) Phylogenetic tree clustering of 529 rice core germplasm accessions. (**d**) Genome-wide LD decay is estimated from all population and subpopulations. The x-axis represents the physical distance and the y-axis represents the average pairwise correlation coefficient (r^2) of SNPs. The black, grey, purple, blue, green, orange, and red colors represent All, Adm, Aus, IndI, IndII, Tej, and Trj, successively

0.28 ~ 1.79% of the phenotypic alteration. Five QTNs (*qTGW-5-1*, *qTGW-5-2*, *qGW-5-2*, *qYPP-7-2*, and *qYPP-10-2*) were mapped by four or more ML-GWAS models.

Moreover, we compared 161 published rice yield-related genes' locations with 74 significant QTNs and their genomic ranges (300 kb up- and down-stream around the associated QTNs) (Fig. 3 and Table S3). Nearly one-third of QTNs overlapped with the known genes in total. For example, *qGL-3-3*, *qGL-3-4*, and *qGL-3-5* overlapped with *GS3* (*Grain Size 3*). Stunningly, three QTLs, *qGL-5-1*, *qGW-5-1*, and *qGW-5-2*, controlling multiple traits (GL and GW) simultaneously existed in the same region on chromosome 5, which were adjacent to *GW5* (*Grain Width 5*). These regions are generally regarded as pleiotropic. Additionally, we found 51 novel QTLs such as *qYPP-4-1*, *qGL-10-1*, *qTGW-11–2*, and *qTGW-12–1* without coinciding with or adjacent to the known genes.

In addition, we compared the results of ML-GWAS and SL-GWAS and a total of 20 common QTNs were identified (Table S4). Four, four, seven, and five QTNs

were discovered associated with GL, GW, TGW, and YPP, respectively. While no common QTN for GT was identified in this study.

Prediction of potential candidate genes

Only the QTNs simultaneously detected by both ML-GWAS and SL-GWAS were further analysed. The *qTGW-7-1*, which was located at 18,639,992 bp on chromosome 7, was identified associated with TGW using both SL-GWAS and ML-GWAS methods. This QTN was detected tightly related to TGW with LAR-mEB, pKWmEB, and ISIS EM-BLASSO methods with the LOD ranged from 3.57 to 11.64 (Table 1). In the MLM method, this SNP was also significantly (*p*-value = 3.62×10^{-9}) associated with TGW with an R^2 of 7.79% (Table S2). Then, a local LD block (18,623,910–18,644,333 bp) was defined (Fig. 4a) with the step we mentioned in the Method section. Based on genome-wide annotation information, *LOC_Os07g31440* and *LOC_Os07g31450* were extracted from this region. Among them, *LOC_Os07g31440* encodes an expressed

Table 1 The significant QTNs for five rice yield-related traits detected simultaneously by using two or more multi-locus GWAS methods

Trait	QTN	Chr	Position	LOD	R^2 (%)	Method	Reported genes
Grain Length	qGL-1-1	1	16,255,794	3.57–6.37	0.29–1.91	1,2,3	
	qGL-3-1	3	8,800,352	3.43–4.15	1.71–3.56	4,5	
	qGL-3-2	3	16,182,203	3.42–5.81	1.92–2.23	1,3	
	qGL-3-3	3	16,699,322	3.68–4.70	2.78–7.54	1,4	GS3
	qGL-3-4	3	16,717,839	3.47–3.51	2.68–6.97	2,3	GS3
	qGL-3-5	3	16,911,337	4.01–4.76	3.49–9.75	1,5	GS3
	qGL-3-6	3	35,509,618	3.22–7.14	2.79–3.68	1,3	qTGW3
	qGL-5-1	5	5,371,587	4.23–4.92	1.25–2.28	2,5	GSE5/GW5
	qGL-5-2	5	6,215,765	3.12–5.01	3.91–5.72	1,4	
	qGL-6-1	6	4,252,841	4.07–9.18	0.74–1.13	1,3	
	qGL-6-2	6	22,641,242	8.54–8.57	7.17–7.35	1,3	
	qGL-7-1	7	4,465,180	3.79–11.29	1.01–9.81	1,2,3	
	qGL-8-1	8	1,629,499	5.01–5.78	0.67–1.33	1,3	
	qGL-8-2	8	14,139,411	3.87–4.67	2.57–3.29	1,3,5	
	qGL-10-1	10	4,663,209	4.49–6.46	7.66–13.25	4,5	
	qGL-10-2	10	5,288,007	4.05–5.38	1.93–2.63	1,2	
	qGL-12–1	12	1,135,876	3.37–3.42	0.00–1.19	2,3	
	qGL-12-2	12	5,542,726	4.02–8.98	0.77–1.25	2,3	
	qGL-12-3	12	14,546,343	5.74–7.38	0.08–1.73	1,3	
Grain Width	qGW-3-1	3	10,817,310	3.50–4.81	0.34–3.64	2,3	
	qGW-3-2	3	21,604,259	3.02–3.65	1.62–5.28	1,4	
	qGW-4-1	4	26,662,080	3.95–5.95	1.77–2.83	3,4	STRK1
	qGW-5-1	5	5,359,498	4.12–6.93	1.68–6.34	2,4,5	GSE5/GW5
	qGW-5-2	5	5,371,587	4.92–11.55	2.83–7.20	1,3,4,5	GSE5/GW5
	qGW-9-1	9	191,910	3.59–5.35	2.30–2.68	3,4	
	qGW-9-2	9	1,318,664	6.74–7.74	5.97–13.97	3,4	BC12
	qGW-10-1	10	22,500,927	3.31–4.41	1.04–1.50	2,4	OsCAO1
	qGW-12–1	12	22,677,925	3.04–12.34	0.29–1.47	2,3	
Grain Thickness	qGT-5-1	5	4,830,996	12.94–13.56	4.93–7.23	1,4	
	qGT-5-2	5	7,022,361	4.94–5.14	1.85–3.20	2,4	
	qGT-5-3	5	7,036,290	4.34–13.62	3.32–3.38	1,5	
	qGT-5-4	5	23,605,308	3.03–4.41	0.85–5.69	1,3	OsSNAT1
	qGT-6-1	6	15,688,533	3.43–4.25	3.59–4.17	1,4	
	qGT-6-2	6	19,652,114	3.52–10.41	2.23–11.95	1,3	OsSPDS2
	qGT-12–1	12	17,685,177	3.02–5.90	2.09–3.77	1,4	
1000-Grain Weight	qTGW-1-1	1	4,853,002	6.27–6.60	1.97–2.11	1,4	GW5L
	qTGW-1-2	1	33,427,348	3.19–4.48	0.60–1.57	3,4	
	qTGW-2-1	2	2,529,182	4.94–10.21	0.56–5.62	2,3	
	qTGW-3-1	3	16,776,481	6.15–7.49	0.89–2.76	1,3,4	GS3
	qTGW-4-1	4	31,950,052	3.14–4.28	0.42–1.42	3,4	
	qTGW-4-2	4	32,409,784	4.54–5.22	1.47–2.41	2,4	FLO2
	qTGW-5-1	5	4,859,527	7.33–8.91	0.86–3.90	1,3,4,5	
	qTGW-5-2	5	7,115,594	4.35–12.71	0.34–4.51	1,3,4,5	

Zhong *et al. BMC Plant Biology* (2021) 21:364

Page 6 of 12

Table 1 The significant QTNs for five rice yield-related traits detected simultaneously by using two or more multi-locus GWAS methods *(Continued)*
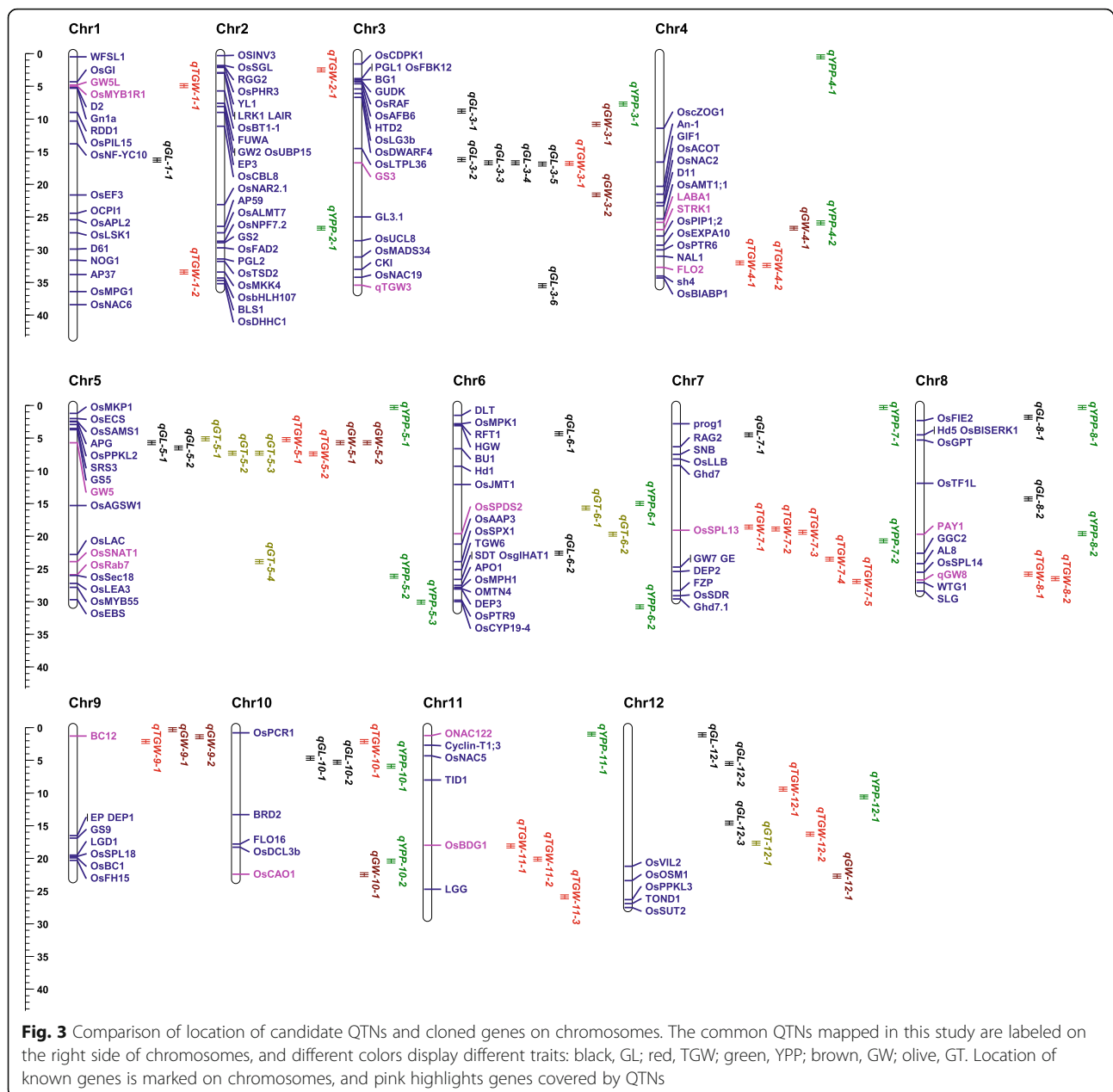
Trait	QTN	Chr	Position	LOD	R^2 (%)	Method	Reported genes
	qTGW-7-1	7	18,639,992	3.57–11.64	1.39–6.37	3,4,5	
	qTGW-7-2	7	18,895,502	4.22–7.48	1.70–4.99	3,5	*OsSPL13*
	qTGW-7-3	7	19,391,625	4.26–11.43	0.44–4.99	1,3,5	*OsSPL13*
	qTGW-7-4	7	23,476,357	3.77–5.02	0.48–3.88	1,3	
	qTGW-7-5	7	26,928,988	3.89–7.12	0.60–1.40	3,4	
	qTGW-8-1	8	25,590,176	3.10–4.53	0.34–1.35	3,4,5	
	qTGW-8-2	8	26,309,952	5.51–6.94	1.06–1.95	1,3	*GW8/OsSPL16/qGW8*
	qTGW-9-1	9	2,009,389	3.01–3.70	0.45–2.50	3,4	
	qTGW-10-1	10	2,107,100	5.81–11.33	1.29–5.77	3,4,5	
	qTGW-11–1	11	18,100,034	4.54–9.46	0.35–2.83	1,3,4	*OsBDG1*
	qTGW-11–2	11	20,067,792	3.52–6.31	0.87–2.95	1,2,4	
	qTGW-11-3	11	25,914,037	3.04–7.25	0.88–1.53	3,5	
	qTGW-12–1	12	9,394,697	8.15–10.58	1.72–4.77	1,3	
	qTGW-12-2	12	16,332,104	3.01–4.30	0.13–0.89	3,4	
Yield Per Plant	qYPP-2-1	2	26,731,138	3.01–3.80	1.13–3.13	4,5	
	qYPP-3-1	3	7,685,085	4.19–11.64	3.11–5.26	1,4,5	
	qYPP-4-1	4	513,812	4.81–5.88	2.67–3.98	1,5	
	qYPP-4-2	4	25,868,074	4.96–6.52	2.15–3.25	1,4,5	*LABA1*
	qYPP-5-1	5	28,202	3.72–6.29	3.59–4.01	1,5	
	qYPP-5-2	5	25,806,082	4.43–8.85	1.17–6.77	3,4,5	*OsRab7*
	qYPP-5-3	5	29,791,637	3.71–5.36	2.38–3.89	1,5	
	qYPP-6-1	6	15,037,982	5.50–6.92	3.75–3.87	4,5	
	qYPP-6-2	6	30,797,769	3.42–4.60	0.65–1.80	2,5	
	qYPP-7-1	7	337,826	8.93–9.51	1.86–5.81	1,3	
	qYPP-7-2	7	20,669,197	3.08–6.62	0.28–1.79	1,2,3,4,5	
	qYPP-8-1	8	79,975	4.58–5.62	1.65–1.86	2,5	
	qYPP-8-2	8	19,396,188	3.12–5.64	0.27–3.08	1,3	*PAY1*
	qYPP-10-1	10	5,854,315	3.21–4.64	0.33–2.03	2,3,5	
	qYPP-10-2	10	20,388,038	3.81–12.61	0.57–6.99	2,3,4,5	
	qYPP-11–1	11	994,705	3.40–3.62	0.59–2.97	3,4	*ONAC122*
	qYPP-12–1	12	10,578,786	3.15–4.20	1.65–3.60	4,5	

Methods 1–5 represnet mrMLM, FASTmrEMMA, pLARmEB, pKWmEB, and ISIS EM-BLASSO, respectively
R^2 (%): the proportion of total phenotypic variance explained by each QTN

protein with unknown function. The *LOC_Os07g31450*, also known as *CHR729/CRL6*, is a CHD (Chromodomain, helicase/ATPase, and DNA-binding domain) protein. Then, we defined four haplotypes of *LOC_Os07g31450* (HapA, HapB, HapC, and HapD) based on the missense mutations in the gene. The accessions with the favorable HapD displayed significantly higher TGW than those with the HapA, HapB, and HapC types (Fig. 4b). These findings revealed that the grain weights of the accessions with favorable haplotype variations were predominantly improved compared to those with unfavourable variations. A previous

study reported that *CHR729* expresses ubiquitously, such as in stems, leaves, leaf sheaths, young panicles, and flower organs [28]. To further explore the expression pattern of *LOC_Os07g31450* in different tissues, we utilized the CREP database to analyse and found *LOC_Os07g31450* had the highest expression level in the young panicles (< 1 mm, 3–5 mm, and 10–15 mm) (Fig. 4c).

The *qGW-9-2* was another QTN simultaneously detected by both SL-GWAS and ML-GWAS. This QTN was identified significantly associated with GW using pLARmEB and pKWmEB methods with the LOD value

Fig. 3 Comparison of location of candidate QTNs and cloned genes on chromosomes. The common QTNs mapped in this study are labeled on the right side of chromosomes, and different colors display different traits: black, GL; red, TGW; green, YPP; brown, GW; olive, GT. Location of known genes is marked on chromosomes, and pink highlights genes covered by QTNs

of 6.74 and 7.74, respectively. Moreover, this site was also detected by the MLM method with the *p*-value of $\times 10^{-8}$. A 50.7 kb LD block (1,314,833–1,365,534 bp) was defined for this QTN. In this region, *LOC_Os09g02830* is annotated as *OsMADS78* which belongs to the MADS-box family. *OSMADS78* has been confirmed to be an important regulator of early seed developmental transition and impacts both rice seed size and quality [29]. Nevertheless, the biological function of *OsMADS78* is far from being understood. Here, our study speculated that *OsMADS78* might contribute to rice grain width regulation.

Discussion
Comparison of SL-GWAS and ML-GWAS results
The conventional single-locus methods like the general linear model (GLM) and MLM have been widely implemented to identify genetic variants in many cereals [30–32]. However, these models have certain shortcomings as they neglect the overall effects of multiple loci and suffer from the problem of multiple test corrections for critical values. For example, the stringent threshold leads to missing many robust QTLs, particularly small-effect QTLs in MLM [16]. ML-GWAS methodologies therewith have been developed, such as

Fig. 4 (**a**) Local linkage disequilibrium for *qTGW-7-1*. (**b**) Box plot of TGW traits about four haplotypes of *LOC_Os07g31450*. The x-axis represents four haplotypes of *LOC_Os07g31450* and the y-axis represents 1000-grain weight. The table below is the detailed information of four haplotypes. (**c**) Heatmap of the expression pattern of *LOC_Os07g31450* in various tissues among three local rice species. The y-coordinate indicates three species and relative expression, and x-coordinate indicates 39 different parts and development stages of rice tissue. Red represents higher gene expression and green indicates lower gene expression level, the gene expression levels are log_2 transformed

mrMLM, FASTmrMLM, LASSO [33], and FarmCPU (Fixed and random model Circulating Probability Unification) [34]. After comparing the statistical power of ML-GWAS with SL-GWAS methods, several studies demonstrated that multi-locus methods have lower false-positive error and higher statistical power than single-locus ones [17, 35, 36]. According to each ML-GWAS algorithm has its own characteristics and different QTL detection, investigators generally combine the merits of several ML-GWAS methods to mine target QTL for complex agronomic traits [26, 37, 38].

In the present study, we adopted the MLM model and five ML-GWAS methods to analyse five yield-related traits of 529 rice core germplasms. Consequently, 152, 106, 12, 111, and 64 significant SNPs, while 3, 19, 8, 53, and 56 QTLs were detected by MLM underlying GL, GW, GT, TGW, and YPP, respectively (Table S2).

Likewise, 161, 136, 160, 189, and 171 significant QTNs were identified using ML-GWAS methods linked with the above five traits successively. We noted that the number of QTLs mapped by MLM was less than the QTNs identification of five ML-GWAS algorithms, especially of those about GT and YPP. The previous study observed similar findings in GWAS analysis of soybean seed size, suggesting that the recognition results of the five ML-GWAS methods outperformed those of the two SL-GWAS programs. In addition, the QTNs distribution detected by ML-GWAS approaches were more dispersed compared with MLM. For instance, as described in Fig. S2a, the significant loci identified by MLM were concentrated near *GS3* on chromosome 3. Whereas, many loci on other chromosomes failed to meet the threshold, indicating that it was difficult to find new loci from other chromosomes when applied traditional MLM. Afterward, a lot of significant QTNs presented across ML-GWAS models were not only situated on chromosome 3, but also widely distributed on other chromosomes, among which QTNs examined by at least three methods are worthy of further research. These data explicated that the ML-GWAS algorithms are considered more effective, powerful, and robust when applying to investigate the small-effect QTNs for yield-related traits.

Comparison of QTLs or QTNs detected in our study and previous studies

Over the past decade, numerous rice yield-related genes such as *GS3*, *GW2*, and *GW5* have been identified and their functional roles were deeply elucidated [39, 40]. Among them, *GS3*, the first molecular characterized QTL for grain size, controls grain weight and length, with minor impacts on grain width and thickness [41]. *GW2* (Grain width 2), negatively regulates grain width [42] and *GW2* homologs in common wheat plays a critical role in the genetic control of grain weight and protein content traits [43]. *GW5* influences grain width and weight acting in the brassinosteroid

pathway and overexpression of *GSE5/GW5* resulted in narrow grains [44, 45].

In the current study, we characterized 74 QTNs for five yield-related traits that were simultaneously identified using two or more ML-GWAS methods. Compared with the mapped loci of the previous studies, 23 QTNs and its ±300 kb genomic ranges overlapped the known genes. On chromosome 3, the QTN hotspot (*qGL-3-3*, *qGL-3-4*, and *qGL-3-5*) was located near the region of *GS3* for grain length. The QTN cluster on chromosome 5 (*qGW-5-1*, *qGW-5-2*, and *qGL-5-1*) was mapped nearby *GW5*. Moreover, there were 51 novel QTNs excluded in the genomic regions of the past reports. Therefore, these identified makers may be the potential QTNs controlling rice grain and productivity.

Dissecting two candidate genes of yield-related traits

Using the efficient mixed-model association, two robust QTLs (*qTGW-7-1* and *qGW-9-2*) were validated with major effects associated with yield-related traits using both SL- and ML-GWAS approaches. In the candidate *qTGW-7-1* related to grain weight, *LOC_Os07g31450/ CHR729* is a kind of CHD protein, which encoded protein contains 2259 amino acids, belonging to the CHD3 subfamily [28]. *CHR729* has been reported to play an essential role in multiple aspects of rice root and seed development. As an example, *CHR729* can control seedling development through the gibberellin pathway [28] and affect crown root formation through the auxin signalling pathway [46]. In this study, GWAS results inferred that *CHR729* might be related to grain weight, and transcriptional expression analysis reflected that *CHR729* was highly expressed in young panicles. The role played by *CHR729* in regulating rice grain and even productivity is worthwhile for further study and confirmation.

In the candidate *qGW-9-2* involved in grain width, gene *LOC_Os09g02830* (*OsMADS78*) belongs to the MADS transcription factor family. MADS family is a large family with conserved MADS-box domains, whose members widely take part in the key regulatory pathways of plant growth and reproduction (including flower formation) [47]. The OsMADS family in rice participates in controlling flowering time, development of root and seed, especially of flower organs [48, 49]. For instance, *OsMADS16/SPW1*, which is homologous to *APETALA3* in *Arabidopsis*, belongs to the Class B in the ABC model of flower organ development, determining the properties of slates and stamens in rice flower organs [50]. Moreover, *OsMADS13* controls ovule identity [51]. *OsMADS26* negatively regulates resistance to rice blast and drought tolerance [52]. And *OsMADS23*, *OsMADS25*, *OsMADS27*, *OsMADS57*, and *OsMADS61* determine root development [48]. Although there are 75 members in the OsMADS

family [53], nearly half of the members' functions are still unknown. In this study, our findings give a clue that *OsMADS78* may be related to grain width in rice. Recently, Paul et al. reported that *OSMADS78* modulates early seed developmental transition and impacts rice grain length, grain width, 1000 grain weight, and grain quality [29]. This conclusion showed the high reliability of our results, corresponding with this published research which verified the biological function of OSMADS78 by overexpression experiment [29].

Conclusions

In this study, a total of 74 QTN hotspots were simultaneously detected for five yield-related traits by two or more ML-GWAS methods. Among them, *qTGW-7-1* and *qGW-9-2*, closely associated with TGW and GW separately, were presented across both SL-GWAS and ML-GWAS analyses. Besides, two key annotated genes (*LOC_Os07g31450* and *LOC_Os09g02830*) underlying the above two target genomic ranges were mined. In summary, many robust QTLs and two candidate genes were supposed to potentially modulate grain shape and productivity in rice. This research made a beneficial attempt by a combinatory approach of ML-GWAS methods and will facilitate the detection of yield-related QTNs.

Methods

Phenotyping data and statistical analyses

The complete phenotypic records of 529 rice core accessions were downloaded from RiceVarMap v2.0 [54], an integrated dataset of rice genomic variations denoted by Huazhong Agricultural University. This set of germplasm contains diverse rice cultivars. Thereinto, the materials were classified into 299 indica (95 indica I, 74 indica II, 13 indica III, and 117 indica intermediate types), 156 japonica (93 temperate japonica, 43 tropical japonica, and 20 japonica intermediate types), 46 aus, 14 aromatic, and 14 intermediate types.

The yield-related agronomic characters include grain length (GL), grain width (GW), grain thickness (GT), thousand-grain weight (TGW), and yield per plant (YPP) were downloaded from RiceVarMap2 website (http:// ricevarmap.ncpgr.cn/phenos/). The detailed information, including experimental design, years, replicate, could be found in a study [55]. Meanwhile, the minimum, maximum, mean, standard deviation, range, and coefficient of variation (CV) for each trait were calculated in Table S1. Pearson correlation analysis for phenotypic data was conducted using the SAS 9.4 software (http://www.sas.com/).

Genotyping data processing

High-quality re-sequencing raw data of 529 germplasms were derived from RiceVarMap (http://ricevarmap.

ncpgr.cn/v1/) [54]. The detailed information of data processing and BCF files could be found on the RiceVarMap webpage. Raw single nucleotide variants (SNV) were processed by PLINK 1.9 software with parameter --maf 0.05 --geno 0.05 --snps-only. Then the genotypic data was imputed by Beagle 5.0 and a total of 607,201 SNPs were left for further analysis [56].

Clustering analysis, population structure, and LD analysis
The matrix of pairwise genetic distance derived from 607,201 SNPs was implemented to construct phylogenetic trees by SNPhylo with the default parameter [57]. Principal component analysis (PCA) and kinship matrix were performed by the Tassel 5.2 program to estimate the population structure [58]. Linkage disequilibrium (LD) between SNPs was estimated as the squared correlation coefficient (r^2) of alleles, meanwhile the whole population and sub-populations were implemented by software PopLDdecay [59].

Genome-wide association study
In our case, GWAS was performed in 529 rice varieties with 607,201 high-quality SNPs. A mixed linear model (MLM) was carried out for single-locus method to evaluate the trait-SNP association analysis for agriculture traits using the Tassel software. The first three principal components (PCs) and kinship matrix were used as co-variates to correct population structure for decreasing false-positive rate in MLM. The genome-wide significance threshold (p-value = 1.65×10^{-6}) was calculated by negative log(1/n, n is the number of SNPs).

Five ML-GWAS methods within the mrMLM R package (https://cran.r-project.org/web/packages/mrMLM/index.html) were used to map candidate QTNs, including mrMLM, FASTmrMLM, FASTmrEMMA, pLARmEB, and ISIS EM-BLASSO. All parameters were set at default values, and the critical LOD score was set to 3 for robust QTNs at the last stage. All these five methods used the PCA and kinship matrices in our study. The Manhattan and QQ plots for GWAS were displayed using the R package CMplot (https://github.com/YinLiLin/R-CMplot). QTNs were named as Q + the initial letter of traits name abbreviations + chromosome number + occurrence sequence [60].

Identification of putative genes
The QTNs identified by at least two different ML-GWAS methods were regarded as the putative candidate loci. Local LD blocks containing at least two SNPs were calculated with all imputed SNP using the PLINK 1.9 software [61]. The local LD blocks of each significant QTN were determined via confidence intervals described by Gabriel [62]. The LD heatmap was visualized using Haploview software [63]. All the genes located in the LD block of QTNs were extracted for further analysis. By comprehensive analysis of gene annotation (MSU 6.1, http://rice.plantbiology.msu.edu/pub/data/Eukaryotic_Projects/o_sativa/annotation_dbs/pseudomolecules/version_6.1/), protein domain function in previous reports, and transcriptome information (data were deposited in the CREP database, Collections of Rice Expression Profiling, http://crep.ncpgr.cn/), the candidate genes for each trait were further mined.

Phenotypic difference of candidate genes
The haplotypes of the candidate gene were determined based on the missense SNP, then the Wilcox-test was used to test the phenotypic difference among each haplotype. The characters of each trait were visualized with box plots using the R 3.6.1 language [64].

Supplementary Information
The online version contains supplementary material available at https://doi.org/10.1186/s12870-021-03146-8.

Additional file 1: Figure S1. Distribution of SNP markers on Chromosomes. The x-axis represents the number of SNPs in the 1 Mb window, and the y-axis represents 12 chromosomes of rice. Different colors represent different numbers of SNPs.

Additional file 2: Figure S2. Manhattan plots of the SL-GWAS and ML-GWAS for yield. (A-E) Manhattan maps representing five traits of GL, GW, GT, TGW, and YPP, respectively; The x-axis displays the chromosome label, and the y-axis displays -log10 (*p*-value). The dotted and solid gray lines show significant associations between SNPs and phenotype value with threshold levels of *p*-value < 1.65×10^{-6} and *p*-value < 8.25×10^{-8}, separately. Red, green, blue, purple, yellow, and grey dots represent mrMLM, FASTmrEMMA, pLARmEB, pKWmEB, ISIS EM-BLASSO, and MLM models. (F-J) QQ plots represent MLM analysis of the above five traits.

Additional file 3: Table S1. Statistical analysis of five rice yield-related traits.

Additional file 4: Table S2. SNPs and QTLs mapping information in MLM analysis.

Zhong *et al. BMC Plant Biology*   (2021) 21:364

Page 11 of 12

Additional file 5: Table S3. The 161 published genes and their functional information of yield traits in rice.

Additional file 6: Table S4. QTN simultaneously detected by both ML-GWAS and MLM methods.

Availability of data and materials
The genotype and phenotype datasets analysed during the current study are available in the RiceVarMap (http://ricevarmap.ncpgr.cn/). The raw sequence data are available in NCBI with accession number PRJNA171289.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare that they have no competing interests

Author details
[1]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, People's Republic of China 430072. [2]Department of Biochemistry, Molecular Biology, Entomology and Plant Pathology, Mississippi State University, Starkville, MS 39762, USA.

References
1. Muthayya S, Sugimoto JD, Montgomery S, Maberly GF. An overview of global rice production, supply, trade, and consumption. Ann N Y Acad Sci. 2014;1324(1):7–14. https://doi.org/10.1111/nyas.12540.
2. Challinor AJ, Watson J, Lobell DB, Howden SM, Smith DR, Chhetri N. A meta-analysis of crop yield under climate change and adaptation. Nat Clim Chang. 2014;4(4):287–91. https://doi.org/10.1038/nclimate2153.
3. Xing Y, Zhang Q. Genetic and molecular bases of rice yield. Annu Rev Plant Biol. 2010;61(1):421–42. https://doi.org/10.1146/annurev-arplant-042 809-112209.
4. Zhong H, Liu C, Kong W, Zhang Y, Zhao G, Sun T, et al. Effect of multi-allele combination on rice grain size based on prediction of regression equation model. Mol Gen Genomics. 2020;295(2):465–74. https://doi.org/10.1007/s00438-019-01627-y.
5. Bai X, Wu B, Xing Y. Yield-related QTLs and their applications in Rice genetic improvement. J Integr Plant Biol. 2012;54(5):300–11. https://doi.org/10.1111/j.1744-7909.2012.01117.x.
6. Fan C, Yu S, Wang C, Xing Y. A causal C-A mutation in the second exon of GS3 highly associated with rice grain length and validated as a functional marker. Theor Appl Genet. 2009;118(3):465–72. https://doi.org/10.1007/s00122-008-0913-1.
7. 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(12):1199–209. https://doi.org/10.1038/cr.2008.307.
8. Wang S, Wu K, Qian Q, Liu Q, Li Q, Pan Y, et al. Non-canonical regulation of SPL transcription factors by a human OTUB1-like deubiquitinase defines a new plant type rice associated with higher grain yield. Cell Res. 2017;27(9):1142–56. https://doi.org/10.1038/cr.2017.98.
9. Si L, Chen J, Huang X, Gong H, Luo J, Hou Q, et al. OsSPL13 controls grain size in cultivated rice. Nat Genet. 2016;48(4):447–56. https://doi.org/10.1038/ng.3518.
10. Zhong H, Kong W, Gong Z, Fang X, Deng X, Liu C, et al. Evolutionary analyses reveal diverged patterns of Squamosa promoter binding protein-like (Spl) gene family in Oryza genus. Front Plant Sci. 2019 May:1–10.
11. Prom LK, Ahn E, Isakeit T, Magill C. GWAS analysis of sorghum association panel lines identifies SNPs associated with disease response to Texas isolates of Colletotrichum sublineola. Theor Appl Genet. 2019;132(5):1389–96. https://doi.org/10.1007/s00122-019-03285-5.
12. Li N, Zheng H, Cui J, Wang J, Liu H, Sun J, et al. Genome-wide association study and candidate gene analysis of alkalinity tolerance in japonica rice germplasm at the seedling stage. Rice. 2019;12(1):24. https://doi.org/10.1186/s12284-019-0285-y.
13. Zhang YM, Mao Y, Xie C, Smith H, Luo L, Xu S. Mapping quantitative trait loci using naturally occurring genetic variance among commercial inbred lines of maize (Zea mays L.). Genetics. 2005;169:2267–75.
14. Kang HM, Sul JH, Service SK, Zaitlen NA, Kong SY, Freimer NB, et al. Variance component model to account for sample structure in genome-wide association studies. Nat Genet. 2010;42(4):348–54. https://doi.org/10.1038/ng.548.
15. Lippert C, Listgarten J, Liu Y, Kadie CM, Davidson RI, Heckerman D. FaST linear mixed models for genome-wide association studies. Nat Methods. 2011;8(10):833–5. https://doi.org/10.1038/nmeth.1681.
16. Wang SB, Feng JY, Ren WL, Huang B, Zhou L, Wen YJ, et al. Improving power and accuracy of genome-wide association studies via a multi-locus mixed linear model methodology. Sci Rep. 2016;6.
17. 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 October:1–9. https://doi.org/10.3389/fpls.2018.01464.
18. Tamba CL, Zhang YM. A fast mrMLM algorithm for multi-locus genome-wide association studies. bioRxiv. 2018;October.
19. 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):e1005357. https://doi.org/10.1371/journal.pcbi.1005357.
20. 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(3):208–18. https://doi.org/10.1038/s41437-017-0007-4.
21. 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(4):700–12. https://doi.org/10.1093/bib/bbw145.
22. 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(6):517–24. https://doi.org/10.1038/hdy.2017.8.
23. Xu Y, Yang T, Zhou Y, Yin S, Li P, Liu J, et al. Genome-wide association mapping of starch pasting properties in maize using single-locus and multi-locus models. Front Plant Sci. 2018;9 September:1–10. https://doi.org/10.3389/fpls.2018.01311.
24. Misra G, Badoni S, Domingo CJ, Cuevas RPO, Llorente C, Mbanjo EGN, et al. Deciphering the genetic architecture of cooked rice texture. Front Plant Sci. 2018;9 October:1–16. https://doi.org/10.3389/fpls.2018.01405.
25. Lü H, Yang Y, Li H, Liu Q, Zhang J, Yin J, et al. Genome-wide association studies of photosynthetic traits related to phosphorus efficiency in soybean. Front Plant Sci. 2018;9 August. https://doi.org/10.3389/fpls.2018.01226.
26. Liu S, Zhong H, Meng X, Sun T, Li Y, Pinson SRM, et al. Genome-wide association studies of ionomic and agronomic traits in USDA mini core collection of rice and comparative analyses of different mapping methods. BMC Plant Biol. 2020;20(1):441. https://doi.org/10.1186/s12870-020-02603-0.
27. Ngangkham U, Samantaray S, Yadav MK, Kumar A, Chidambaranathan P, Katara JL. Effect of multiple allelic combinations of genes on regulating

grain size in rice. PLoS One. 2018;13(1):1–20. https://doi.org/10.1371/journal.pone.0190684.

28. Ma X, Ma J, Zhai H, Xin P, Chu J, Qiao Y, et al. CHR729 is a CHD3 protein that controls seedling development in rice. PLoS One. 2015;10.

29. Paul P, Dhatt BK, Miller M, Folsom JJ, Wang Z, Krassovskaya I, et al. MADS78 and MAdS79 are essential regulators of early seed development in RICE. Plant Physiol. 2020;182(2):933–48. https://doi.org/10.1104/pp.19.00917.

30. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet. 2006;38(8):904–9. https://doi.org/10.1038/ng1847.

31. He L, Xiao J, Rashid KY, Yao Z, Li P, Jia G, et al. Genome-wide association studies for pasmo resistance in flax (*Linum usitatissimum* L.). Front Plant Sci. 2019;9 January:1982. https://doi.org/10.3389/fpls.2018.01982.

32. Sant'Ana GC, Pereira LFP, Pot D, Ivamoto ST, Domingues DS, Ferreira RV, et al. Genome-wide association study reveals candidate genes influencing lipids and diterpenes contents in Coffea arabica L. Sci Rep. 2018;8:1–12.

33. Xu Y, Xu C, Xu S. Prediction and association mapping of agronomic traits in maize using multiple omic data. Heredity (Edinb). 2017;119(3):174–84. https://doi.org/10.1038/hdy.2017.27.

34. 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(2):e1005767. https://doi.org/10.1371/journal.pgen.1005767.

35. 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(7):825–30. https://doi.org/10.1038/ng.2314.

36. Hu X, Zuo J, Wang J, Liu L, Sun G, Li C, et al. Multi-locus genome-wide association studies for 14 main agronomic traits in Barley. Front Plant Sci. 2018;871(November):1–14. https://doi.org/10.3389/fpls.2018.01683.

37. Wu X, Li Y, Shi Y, Song Y, Zhang D, Li C, et al. Joint-linkage mapping and GWAS reveal extensive genetic loci that regulate male inflorescence size in maize. Plant Biotechnol J. 2016;14(7):1551–62. https://doi.org/10.1111/pbi.12519.

38. Ma Z, He S, Wang X, Sun J, Zhang Y, Zhang G, et al. Resequencing a core collection of upland cotton identifies genomic variation and loci influencing fiber quality and yield. Nat Genet. 2018;50(6):803–13. https://doi.org/10.1038/s41588-018-0119-7.

39. Sakamoto T, Matsuoka M. Identifying and exploiting grain yield genes in rice. Curr Opin Plant Biol. 2008;11(2):209–14. https://doi.org/10.1016/j.pbi.2008.01.009.

40. Zuo J, Li J. Molecular genetic dissection of quantitative trait loci regulating rice grain size. Annu Rev Genet. 2014;48(1):99–118. https://doi.org/10.1146/annurev-genet-120213-092138.

41. 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(6):1164–71. https://doi.org/10.1007/s00122-006-0218-1.

42. Song XJ, Huang W, Shi M, Zhu MZ, Lin HX. A QTL for rice grain width and weight encodes a previously unknown RING-type E3 ubiquitin ligase. Nat Genet. 2007;39(5):623–30. https://doi.org/10.1038/ng2014.

43. Zhang Y, Li D, Zhang D, Zhao X, Cao X, Dong L, et al. Analysis of the functions of TaGW2 homoeologs in wheat grain weight and protein content traits. Plant J. 2018;94(5):857–66. https://doi.org/10.1111/tpj.13903.

44. Duan P, Xu J, Zeng D, Zhang B, Geng M, Zhang G, et al. Natural variation in the promoter of GSE5 contributes to grain size diversity in Rice. Mol Plant. 2017;10(5):685–94. https://doi.org/10.1016/j.molp.2017.03.009.

45. 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(5). https://doi.org/10.1038/nplants.2017.43.

46. Wang Y, Wang D, Gan T, Liu L, Long W, Wang Y, et al. CRL6, a member of the CHD protein family, is required for crown root development in rice. Plant Physiol Biochem. 2016;105:185–94. https://doi.org/10.1016/j.plaphy.2016.04.022.

47. Callens C, Tucker MR, Zhang D, Wilson ZA. Dissecting the role of MADS-box genes in monocot floral development and diversity. J Exp Bot. 2018;69(10):2435–59. https://doi.org/10.1093/jxb/ery086.

48. Yu C, Su S, Xu Y, Zhao Y, Yan A, Huang L, et al. The effects of fluctuations in the nutrient supply on the expression of five members of the AGL17 clade of MADS-box genes in rice. PLoS One. 2014;9.

49. Jeon JS, Lee S, Jung KH, Yang WS, Yi GH, Oh BG, et al. Production of transgenic rice plants showing reduced heading date and plant height by ectopic expression of rice MADS-box genes. Mol Breed. 2000;6(6):581–92. https://doi.org/10.1023/A:1011388620872.

50. Xiao H, Wang Y, Liu D, Wang W, Li X, Zhao X, et al. Functional analysis of the rice AP3 homologue OsMADS16 by RNA interference. Plant Mol Biol. 2003;52(5):957–66. https://doi.org/10.1023/A:1025401611354.

51. Dreni L, Jacchia S, Fornara F, Fornari M, Ouwerkerk PBF, An G, et al. The D-lineage MADS-box gene OsMADS13 controls ovule identity in rice. Plant J. 2007;52(4):690–9. https://doi.org/10.1111/j.1365-313X.2007.03272.x.

52. Khong GN, Pati PK, Richaud F, Parizot B, Bidzinski P, Mai CD, et al. OsMADS26 negatively regulates resistance to pathogens and drought tolerance in rice. Plant Physiol. 2015;169(4):2935–49. https://doi.org/10.1104/pp.15.01192.

53. Arora R, Agarwal P, Ray S, Singh AK, Singh VP, Tyagi AK, et al. MADS-box gene family in rice: genome-wide identification, organization and expression profiling during reproductive development and stress. BMC Genomics. 2007;8(1):242. https://doi.org/10.1186/1471-2164-8-242.

54. Zhao H, Yao W, Ouyang Y, Yang W, Wang G, Lian X, et al. RiceVarMap: a comprehensive database of rice genomic variations. Nucleic Acids Res. 2015;43(D1):D1018–22. https://doi.org/10.1093/nar/gku894.

55. Xie W, Wang G, Yuan M, Yao W, Lyu K, Zhao H, et al. Breeding signatures of rice improvement revealed by a genomic variation map from a large germplasm collection. Proc Natl Acad Sci U S A. 2015;112(39):E5411–9. https://doi.org/10.1073/pnas.1515919112.

56. Browning BL, Zhou Y, Browning SR. A one-penny imputed genome from next-generation reference panels. Am J Hum Genet. 2018;103(3):338–48. https://doi.org/10.1016/j.ajhg.2018.07.015.

57. 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):162. https://doi.org/10.1186/1471-2164-15-162.

58. Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES. TASSEL: software for association mapping of complex traits in diverse samples. Bioinformatics. 2007;23(19):2633–5. https://doi.org/10.1093/bioinformatics/btm308.

59. 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(10):1786–8. https://doi.org/10.1093/bioinformatics/bty875.

60. McCouch SR. Gene nomenclature system for rice. Rice. 2008;1(1):72–84. https://doi.org/10.1007/s12284-008-9004-9.

61. 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(3):559–75. https://doi.org/10.1086/519795.

62. Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, et al. The structure of haplotype blocks in the human genome. Science. 2002;296(5576):2225–9. https://doi.org/10.1126/science.1069424.

63. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics. 2005;21(2):263–5. https://doi.org/10.1093/bioinformatics/bth457.

64. R Core Team. R: A language and environment for statistical computing: R Foundation for Statistical Computing; 2019.

Publisher's Note