Full Length Article

Genome-Wide Association Mapping for Grain Shape in Rice Accessions

Xianjin Qiu[1,2†], Shuangbing Zhu[1,3†], Hui Hu[1,2†], Chunchao Wang[4†], Wenkai Lv[1], Liping He[1], Junying Xu[1,5], Zhixin Li[1], Danying Xing[1], Longwei Yang[1], Wenliang Wei[1,2*] and Jianlong Xu[2,4*]

[1]*Hubei Key Laboratory of Waterlogging Disaster and Agricultural Use of Wetland (Yangtze University), Jingzhou 434025, China*
[2]*Hubei Collaborative Innovation Center for Grain Industry, Jingzhou 434025, China*
[3]*Shenzhen Institute of Breeding & Innovation, Chinese Academy of Agricultural Sciences, Shenzhen 518120, China*
[4]*Institute of Crop Science, Chinese Academy of Agricultural Sciences, Beijing 100081, China*
[5]*Hubei Key Laboratory of Food Crop Germplasm and Genetic Improvement, Wuhan 430070, China*
[*]*For correspondence: xujlcaas@126.com; whwenliang@163.com*
[†]Contributed equally to this work and are co-first authors

Abstract

Grain shape is a very important goal in the rice breeding program nowadays. Underlying the genetic basis of grain shape would largely speed the breeding process. In the present study, a diverse panel of 1016 rice accessions from 70 countries and areas worldwide were evaluated for their grain shapes. There were large variations for all three grain shape traits, including grain length, grain width and length to width ratio, and they had significantly positive or negative correlations with each other. Besides, high positive correlations were found between two experimental years. Genome-wide association mapping strategy was conducted to identify QTL for grain shape using 509219 high quality SNP genotypes and grain shape traits. A total of 70 QTL were identified in all chromosomes. Among them, 16 QTL were found on chromosome 1, 2, 3, 5, 7, 9 and 11 expressed in both years. Eight genomic regions on chromosome 1, 3, 4, 5, 7 and 9 had pleiotropic effects on two or three traits. Four genomic regions on chromosome 3, 5, 6 and 7 were identified in both year and controlled more than two traits. Besides, 25 QTL on all chromosomes except chromosome 8, 10 and 12 had large effect on the traits. In conclusion, these results helped us better understand the genetic basis of rice grain shape and provide valuable gene resources for improving rice grain for desired shape. © 2020 Friends Science Publishers

Keywords: Rice; Grain shape; Genome-wide association mapping; QTL

Introduction

In the past half century, rice (*Oryza sativa* L.) yield has been sharply increasing because of two green revolutions (Zhang 2007). However, quality of most rice varieties was very low. With rising living standard and development of economy at rice-consuming area, rice quality has been paid more and more attention by both consumers and producers. Thus, improving rice quality has become an important breeding goal in rice breeding (Tan *et al.* 2000; Zou *et al.* 2018).

Rice quality could be divided into four types. Among them, grain shape (belonging to appearance quality) is an especially important one. Peoples at difference countries and areas preferred different shapes with different prices (Luo *et al.* 2004). Besides, it could affect grain chalkiness and head milled rice rate. Further, grain shape is also an important factor for rice yield by controlling grain weight (Qiu *et al.* 2017). Grain shape consists of grain length (GL), grain width (GW) and length to width ratio (LWR). All of them are quantitative traits and controlled by a lot of gene located on all chromosomes (http://www.gramene.org/).

Among them, some genes have been successfully cloned, such as *GW2* (Song *et al.* 2007), *qSW5/GW5/GSE5* (Shomura *et al.* 2008; Weng *et al.* 2008; Liu *et al.* 2017), *GS3* (Mao *et al.* 2010), *GS5* (Li *et al.* 2011), *GL3.1* (Qi *et al.* 2012; Zhang *et al.* 2012), *GS2* (Che *et al.* 2015; Hu *et al.* 2015), *GL7/GW7/SLG7* (Wang *et al.* 2015; Zhou et al. 2015), *GLW7* (Si et al. 2016), *smg11* (Fang *et al.* 2016), *SGDP7* (Bai *et al.* 2017) and *GS9* (Zhao *et al.* 2018).

In recent years, genome-wide association mapping (GWAS) based on linkage disequilibrium (LD) has been used to identify a large number of quantitative trait loci (QTL) for rice grain shape. Seven QTL were detected for grain length and width in 517 landraces by resequencing (Huang *et al.* 2010; Wattoo et al. 2019). Zhao *et al.* (2011) identified 56 loci for grain shape of rice seed and brown rice respectively by 413 diverse inbred accessions using 44K SNP genotypes. Zhang *et al.* (2014) detected 35 QTL for three grain shape traits by 150 landraces with 274 simple sequence repeat markers. Nineteen QTL were identified for grain shape using 272 *xian* (*indica*) accessions with genotyping of 22266 SNPs (Qiu *et al.* 2015). Likewise, 45

QTL and 17 candidate genes were identified for rice appearance quality using 258 accessions from 3K RGP with high-through genotyping (Wang *et al.* 2016).

In the present study, we evaluate the grain shapes of 1016 accessions from 3K Rice Genome Project (Wang *et al.* 2018). GWAS was performed to identify QTL for rice grain shape using high quality SNPs generated from 3K RGP through high-throughput sequencing. The association results are likely to help us better understand the genetic basis of rice grain shape and provide valuable gene resources for improving rice grain shape.

Materials and Methods

Association Mapping Materials

A total of 1016 rice accessions from 3K RGP (Wang *et al.* 2018) which could normally flower in Jinghzou city were used in this study. They came from 70 countries worldwide, including 262 (25.79%) from China, 132 (12.99%) from India, 78 (7.68%) from Philippine, 63 (6.20%) from Bangladesh, 38 (3.74%) from Japan, 201 (19.78%) from other Asian countries, 234 (23.03%) from America, Europe, Africa and Oceania, and 8 (0.79%) with unknown origin. They were divided into five subpopulations, including 543 (53.44) *Xian* (*indica*, 543, 53.44%), *Geng* (*japonica*, 318, 31.30%), *aus* (115, 11.32%), *basmati* (12, 1.18) and *admixture* (28, 2.76%).

Field Trails and Trait Measurement

All accessions were planted in the summer seasons of 2015 and 2016 on the experimental farm of Yangtze University at Jingzhou, China (30.2°N, 112.7°E). Seeds were sown at May 5 and transplanted at about 35 days after sowing. Each plot was grown in three or four rows with six individuals in each row. The spacing among different plants was 20 cm × 20 cm. The experiment design was randomized complete block design with two replications.

At maturity stage, all individuals of each plot were harvest and bulked. GL, GW and LWR were evaluated after storing for three months. One hundred full filled seeds were selected for measurement. GL and GW were determined by lined 10 grain end-to-end and breadth, respectively. They were both measured for three times, and the average values were used for analysis. LWR was calculated as GL divided by GW.

Data Analysis and Association Mapping

Statistica v5.5 was used for analyzing statistical description and correlations between two different traits and between two years for one trait (Morales 2001). A 2.3 M SNP data of 1016 accessions were generated from 3K RGP in RFGB v2.0 (Wang *et al.* 2020). For SNPs that had more than three alleles, only two highest frequency alleles were retained and the rest were considered missing. Heterozygous alleles were also considered missing. The SNP loci with missing rate over 20% or minor allele frequency (MAF) below 0.05 were removed. Finally, 509219 SNPs were used for GWAS.

The GWAS was conducted to detect QTL for grain shape by 509219 high quality SNPs and the trait values of the 1016 accessions using the SVS software package (SNP & Variation Suite, Version 8.4.0). The single-locus mixed linear model was applied to the marker dataset (Kang *et al.* 2010; Vilhjalmsson and Nordborg 2013). The threshold was set as $P < 2.0 \times 10^{-6}$. Since the LD decay in 3K rice germplasms were extend from 100–200 kb (Wang *et al.* 2018), peak SNPs within 200 kb were considered as a single QTL.

Results

Grain Shape Performances of 1016 Accessions

The grain shape of 1016 accessions is listed in Table 1. In both years, all three traits showed large variations (CV ranged from 11.49–22.11%), and LWR had larger variation than GL and GW (CV was above 20% for LWR and below 20% for both GL and GW). All traits appeared to be normally distributed in both years, suggesting that they were controlled by multiple genes. Interestingly, GL and GW in 2015 were significantly longer and wider than that of 2016.

Correlation among Different Traits

Correlation coefficients among and between different traits for two years are presented in Table 2. The correlations among different grain shape traits were similar in both years. GL had moderate negative correlation with GW (correlation coefficients were -0.50 and -0.46 in 2015 and 2016 respectively), and was highly positive correlated with LWR (correlation coefficients 0.83 and 0.81 in 2015 and 2016 respectively). GW was highly negatively correlated with LWR (correlation coefficients -0.88 in both 2015 and 2016). Moreover, highly positive correlations were shown between two years for all three grain shape traits, indicating that environments had low affection on them.

Genome-wide Association Mapping for Three Grain Shape Traits

A total of 70 QTL were identified for three grain shape traits, located on all 12 chromosomes (Table 3 and Fig. 1). For GL, 24 QTL were identified in the two years, and located on 12 chromosomes except chromosome 8 and 12. Among them, 8 QTL were detected only in 2015, including *qGL1*, *qGL2*, *qGL4.1*, *qGL5.2*, *qGL6.1*, *qGL7.3*, *qGL9.1* and *qGL9.3*, and they explained 1.2–30.1% of phenotypic variations. 9 QTL including *qGL3.2*, *qGL4.2*, *qGL4.3*, *qGL5.3*, *qGL6.4*, *qGL7.2*, *qGL9.2*, *qGL10* and *qGL11.1* were identified only in 2016, with the phenotypic variation rate varied from 0.9% to 13.1%. 7 QTL (*qGL3.1*, *qGL3.3*, *qGL5.1*, *qGL6.2*, *qGL6.3*, *qGL7.1* and *qGL11.2*) were found

Table 1: Grain shape performance of 1016 accessions

Trait[a]	Year	Mean ± SD	Range	CV (%)	p[b]
GL	2015	8.66±1.01	5.57-11.85	11.64	$5.81×10^{-214}$
	2016	7.26±0.83	4.73-9.77	11.49	
GW	2015	3.04±0.42	1.92-4.19	13.75	$1.06×10^{-174}$
	2016	2.52±0.35	1.67-3.61	13.83	
LWR	2015	2.93±0.65	1.72-4.70	22.11	0.31
	2016	2.96±0.65	1.69-4.73	22.00	

a GL, grain length; GW, grain width; LWR, length to width ratio
b difference between two years

Table 2: Correlation coefficients among different grain shape traits in the two years

Grain characteristics	GL	GW	LWR
GL	0.91**	-0.46**	0.81**
GW	-0.50**	0.90**	-0.88**
LWR	0.83**	-0.88**	0.94**

Data under and above the diagonal were correlations among different grain shape traits in 2015 and 2016, respectively; data in the diagonal were correlations between two years of the same trait. GL, grain length; GW, grain width; LWR, length to width ratio. ** represented significant level at p<0.01
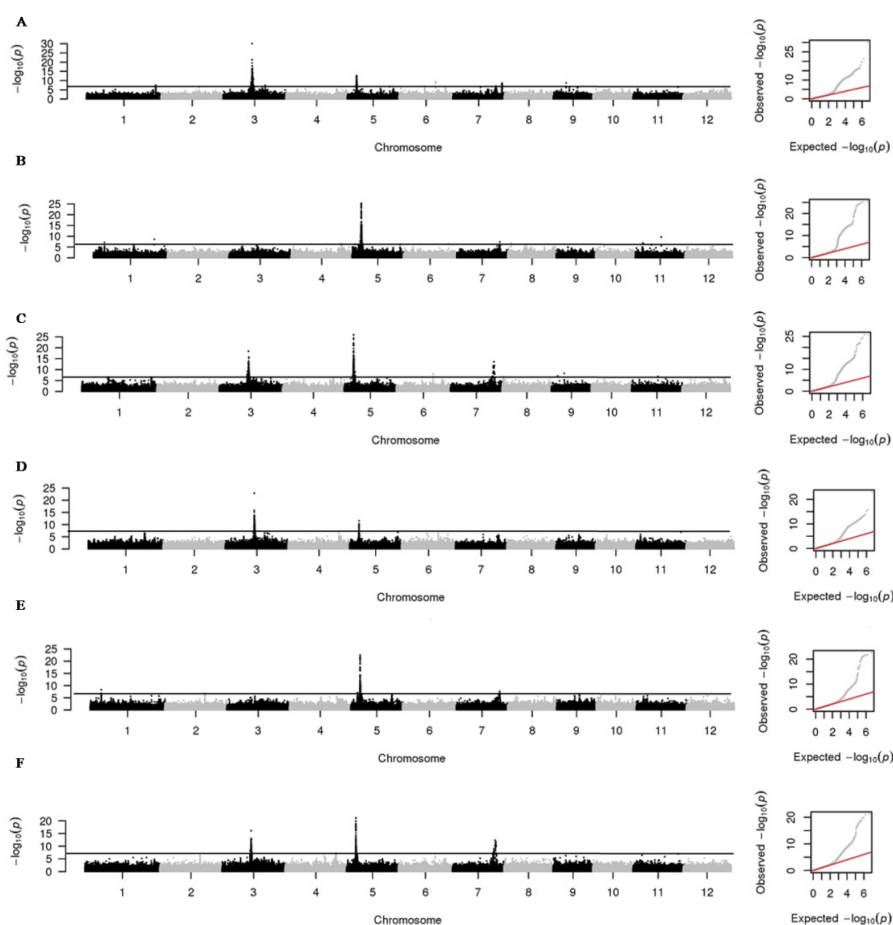

Fig. 1: Manhattan plot and QQ plot of grain shape traits in 2015 and 2016
A, GL in 2015; B, GW in 2015; C, LWR in 2015; D, GL in 2016; E, GW in 2016; F, LWR in 2016

in both years, with the average phenotypic rate of 14.9% and 14.2% in 2015 and 2016 respectively. Among all detected QTL, 9 QTL (*qGL2*, *qGL3.1*, *qGL4.1*, *qGL5.1*, *qGL5.2*, *qGL6.1*, *qGL7.2*, *qGL9.1* and *qGL11.1*) were main, which explained more than 10% phenotypic variations.

Twenty one QTL were identified on all chromosomes except chromosome 10 and 12. Among them, 9 QTL including *qGW1.2*, *qGW3.2*, *qGW4*, *qGW5.3*, *qGW6*, *qGW8.1*, *qGW11.2*, *qGW11.4* and *qGW11.5* were detected only in 2015 and accounted for 1.1–17.1% of phenotypic variations. 8 QTL (*qGW2*, *qGW3.1*, *qGW5.2*, *qGW8.2*, *qGW9.1*, *qGW9.2*, *qGW11.1* and *qGW11.3*) were identified

Table 3: QTL detected for three grain shape traits in the two years

Trait	QTL	Year	Chr.	Peak Position (bp)	P value	allele[a]	MAF[b]	R^2(%)[c]
GL	qGL1	2015	1	40078994	2.10×10^{-8}	C/A	0.15	29.1
	qGL2	2015	2	21944316	3.98×10^{-7}	A/G	0.14	26.0
	qGL3.1	2015	3	16733441	8.48×10^{-32}	G/T	0.38	52.2
		2016	3	16733441	2.21×10^{-24}	G/T	0.37	44.5
	qGL3.2	2016	3	22660967	2.33×10^{-7}	C/T	0.28	4.4
	qGL3.3	2015	3	24310091	3.73×10^{-8}	G/A	0.11	3.3
		2016	3	24310073	3.57×10^{-7}	G/A	0.11	5.2
	qGL4.1	2015	4	11305816	1.71×10^{-6}	G/T	0.27	17.5
	qGL4.2	2016	4	22488406	1.23×10^{-6}	G/A	0.05	6.5
	qGL4.3	2016	4	29308935	1.20×10^{-6}	C/A	0.09	0.9
	qGL5.1	2015	5	5372955	6.37×10^{-14}	A/G	0.46	34.6
		2016	5	5372955	8.75×10^{-13}	A/G	0.49	30.6
	qGL5.2	2015	5	24521950	1.33×10^{-7}	C/T	0.12	23.5
	qGL5.3	2016	5	27836744	1.51×10^{-7}	G/A	0.13	1.0
	qGL6.1	2015	6	1103147	6.01×10^{-7}	T/C	0.09	13.0
	qGL6.2	2015	6	14837297	1.25×10^{-7}	G/A	0.15	3.8
		2016	6	14837319	4.62×10^{-7}	T/C	0.17	4.7
	qGL6.3	2015	6	21149455	4.81×10^{-10}	G/A	0.15	4.3
		2016	6	21149455	3.14×10^{-7}	G/A	0.17	4.8
	qGL6.4	2016	6	26822016	1.70×10^{-6}	G/A	0.06	1.6
	qGL7.1	2015	7	15943665	4.06×10^{-7}	A/G	0.10	3.1
		2016	7	15943665	9.63×10^{-7}	A/G	0.11	4.7
	qGL7.2	2016	7	24629753	8.82×10^{-7}	T/C	0.07	10.7
	qGL7.3	2015	7	28305040	8.42×10^{-9}	G/A	0.06	1.4
	qGL9.1	2015	9	7366713	1.11×10^{-9}	G/A	0.13	30.1
	qGL9.2	2016	9	13118071	1.02×10^{-6}	G/A	0.08	3.1
	qGL9.3	2015	9	14009731	9.70×10^{-7}	A/C	0.08	1.2
	qGL10	2016	10	6158569	7.02×10^{-7}	C/T	0.06	3.0
	qGL11.1	2016	11	2618388	1.19×10^{-6}	C/T	0.20	13.1
	qGL11.2	2015	11	26045504	3.21×10^{-7}	C/T	0.14	3.2
		2016	11	26045504	1.08×10^{-7}	C/T	0.15	5.3
GW	qGW1.1	2015	1	6410156	7.35×10^{-8}	G/A	0.06	2.1
		2016	1	6366660	2.49×10^{-8}	T/A	0.14	0.1
	qGW1.2	2015	1	23616332	1.53×10^{-6}	G/A	0.10	1.1
	qGW1.3	2015	1	35531397	1.65×10^{-9}	G/A	0.10	3.7
		2016	1	35531397	1.61×10^{-6}	G/A	0.10	1.8
	qGW2	2016	2	23182516	4.45×10^{-7}	A/G	0.05	0.6
	qGW3.1	2016	3	13492943	5.20×10^{-7}	C/T	0.12	37.3
	qGW3.2	2015	3	15979561	1.30×10^{-6}	T/C	0.49	17.1
	qGW4	2015	4	27872617	3.83×10^{-7}	A/T	0.07	1.1
	qGW5.1	2015	5	5371716	3.70×10^{-26}	G/A	0.49	42.2
		2016	5	5371952	6.95×10^{-24}	A/G	0.49	35.8
	qGW5.2	2016	5	23699793	4.98×10^{-7}	G/A	0.09	0.1
	qGW5.3	2015	5	28016351	1.54×10^{-7}	A/T	0.11	16.4
	qGW6	2015	6	26455250	2.90×10^{-7}	T/G	0.09	3.5
	qGW7	2015	7	25037095	1.04×10^{-8}	T/A	0.07	11.1
		2016	7	25037095	4.22×10^{-9}	T/A	0.07	9.6
	qGW8.1	2015	8	1963860	1.46×10^{-7}	C/T	0.07	2.4
	qGW8.2	2016	8	21506944	1.53×10^{-6}	T/A	0.06	0.9
	qGW9.1	2016	9	10526914	7.59×10^{-7}	G/A	0.06	2.7
	qGW9.2	2016	9	13076645	1.61×10^{-6}	A/G	0.08	3.1
	qGW11.1	2016	11	3947107	1.80×10^{-6}	C/T	0.37	36.5
	qGW11.2	2015	11	4371420	8.80×10^{-8}	G/A	0.09	12.0
	qGW11.3	2016	11	6313711	9.62×10^{-7}	G/A	0.06	8.0
	qGW11.4	2015	11	9348182	6.65×10^{-7}	C/T	0.06	1.2
	qGW11.5	2015	11	15063878	2.29×10^{-10}	C/T	0.06	2.4
LWR	qLWR1.1	2015	1	15292999	3.08×10^{-7}	A/G	0.24	0.7
	qLWR1.2	2016	1	27795538	1.62×10^{-6}	C/T	0.36	17.3
	qLWR1.3	2015	1	40079680	8.70×10^{-7}	A/G	0.23	28.7
	qLWR2	2016	2	7831173	1.74×10^{-6}	C/T	0.26	17.8
	qLWR3.1	2016	3	10757171	7.78×10^{-7}	T/A	0.07	4.8
	qLWR3.2	2016	3	14399796	2.78×10^{-7}	C/A	0.08	3.3
	qLWR3.3	2015	3	16733441	4.84×10^{-20}	G/T	0.37	48.5
		2016	3	16733441	4.98×10^{-18}	G/T	0.36	45.4
	qLWR3.4	2015	3	19643735	8.30×10^{-7}	G/A	0.05	0.8
	qLWR3.5	2015	3	29535523	7.76×10^{-7}	C/T	0.07	0.8

Table: 3 Continued

qLWR4.1	2016	4	22488406	8.58×10^{-8}	G/A	0.05	5.8
qLWR4.2	2016	4	29343825	4.01×10^{-7}	G/A	0.14	1.5
qLWR5.1	2016	5	5373491	2.15×10^{-22}	A/G	0.49	41.0
	2015	5	5373492	2.63×10^{-27}	T/G	0.46	48.0
qLWR5.2	2015	5	24521950	6.01×10^{-7}	C/T	0.12	21.1
qLWR6	2015	6	21149455	2.25×10^{-8}	G/A	0.16	1.2
	2016	6	21149455	1.69×10^{-6}	G/A	0.18	1.6
qLWR7	2016	7	24943997	6.86×10^{-14}	G/A	0.06	13.7
	2015	7	25037095	3.42×10^{-15}	T/A	0.07	15.4
qLWR8.1	2015	8	870364	5.26×10^{-7}	G/T	0.09	1.6
qLWR8.2	2016	8	24258192	1.78×10^{-6}	G/A	0.05	5.5
qLWR9.1	2016	9	1030669	8.35×10^{-7}	A/G	0.05	5.0
qLWR9.2	2015	9	3539945	3.06×10^{-7}	C/T	0.40	15.1
qLWR9.3	2015	9	7366713	4.10×10^{-9}	G/A	0.13	31.2
	2016	9	7366713	3.57×10^{-7}	G/A	0.13	27.7
qLWR11.1	2016	11	5198328	8.70×10^{-7}	G/A	0.09	0.1
qLWR11.2	2015	11	15063878	9.10×10^{-8}	C/T	0.06	6.1
qLWR11.3	2016	11	18267418	5.70×10^{-7}	C/T	0.12	5.9
qLWR11.4	2016	11	26045504	1.82×10^{-7}	C/T	0.16	2.7
qLWR12	2015	12	15924561	1.18×10^{-6}	T/C	0.06	1.0

a Major/Minor allele
b minor allele frequency
c phenotypic variation explained

only in 2016, with average phenotypic variation rate of 11.2%. Four QTL, *qGW1.1*, *qGW1.3*, *qGW5.1* and *qGW7* were found in both years, accounting for 2.1–42.2% and 0.1–35.8% of phenotypic variations in the two years respectively. Seven QTL including *qGW3.1*, *qGW3.2*, *qGW5.1*, *qGW5.3*, *qGW7*, *qGW11.1* and *qGW11.2* with phenotypic variations explained above 10% were the main QTL. For LWR, 25 QTL were detected on all chromosomes except chromosome 10. 9 QTL were found only in 2015, including *qLWR1.1*, *qLWR1.3*, *qLWR3.4*, *qLWR3.5*, *qLWR5.2*, *qLWR8.1*, *qLWR9.2*, *qLWR11.2* and *qLWR12*. They explained 0.7–28.7% of phenotypic variations. 11 QTL (*qLWR1.2*, *qLWR2*, *qLWR3.1*, *qLWR3.2*, *qLWR4.1*, *qLWR4.2*, *qLWR8.2*, *qLWR9.1*, *qLWR11.1*, *qLWR11.3* and *qLWR11.4*) were identified only in 2016, with phenotypic variations rates varied from 0.1 to 17.8%. 5 QTL including *qLWR3.3*, *qLWR5.1*, *qLWR6*, *qLWR7* and *qLWR9.3* were found in both two years, with average phenotypic variations rate of 28.9 and 25.9% in 2015 and 2016 respectively. Among them, nine QTL with phenotypic variations above 10% were main QTL, including *qLWR1.2*, *qLWR1.3*, *qLWR2*, *qLWR3.3*, *qLWR5.1*, *qLWR5.2*, *qLWR7*, *qLWR9.2* and *qLWR9.3*.

Pleiotropic QTL for Three Grain Shape Traits

In this study, a total of 70 QTL were identified for three grain shape traits (Table 3 and Fig. 1). Among them, eight genomic regions controlled more than two traits and defined as pleiotropic QTL. Two QTL (*qGL1* and *qLWR1.3*) in the 40.79 Mb on chromosome 1, two QTL (*qGL3.1* and *qLWR3.3*) with the same peak SNP at 16.73 Mb on chromosome 3, two QTL (*qGL4.2* and *qLWR4.1*) with the same peak SNP at 22.49 Mb and two QTL (*qGL4.3* and *qLWR4.2*) in the 29.3 Mb on chromosome 4 and two QTL

(*qGL5.2* and *qLWR5.2*) with the same peak SNP at 24.53 Mb on chromosome 5 controlled both GL and LWR. Two QTL (*qGL9.2* and *qGW9.2*) in the 13.1 Mb on chromosome 9 had pleiotropic effect on both GL and GW. Two QTL (*qGW7* and *qLWR7*) in the 25.0 Mb on chromosome 7, two QTL (*qGW11.5* and *qLWR11.2*) with the same peak SNP at 15.06 Mb on chromosome 11 affected both GW and LWR. Three QTL (*qGL5.1*, *qGW5.1* and *qLWR5.1*) with a mass of 5.37 Mb on chromosome 5 had pleiotropic effect on all three traits.

Discussion

With the development of marker technology, a large number of QTL for grain shape have been mapped or even cloned (http://www.gramene.org/). However, most of them were identified by segregating populations derived from bi-parents, such as F_2, backcross population, recombinant inbred lines, chromosome segment substitution lines, introgression lines. They are very important populations for mining QTL/genes for all important traits, but there were still some notable limitations (Qiu *et al.* 2015). Firstly, selecting suitable parents was the key factor for research success. Besides, researchers had to develop segregating populations (Yang *et al.* 2007). All these processes would require a lot of time, labor and money. Moreover, because of limited sample number and meiosis, the mapping resolutions of segregating populations were low. Finally, as each parent carried only one allele, we could only detect no more than two alleles. Thus, we could not find the most favorable allele in the germplasms.

In recent years, GWAS has been becoming a popular strategy for identifying QTL for complex traits, and the number of QTL detected by this method has been sharply increasing. The most important factors influencing effect on

QTL detection were sample number, marker resolution and genetic diversity (Yang *et al.* 2007). With the increase in sample number, maker resolution and diversity, the detection effect increased sharply. A total of 1016 rice accessions from 3K RGP were used in the present study. Compared with previous studies, the number was much more than most other studies. Besides, they came from 70 countries and areas worldwide and had large genetic variation (Wang *et al.* 2018). They had high-throughput SNP genotyping with an average of 1.27 SNP markers per 1 kb. Thus, the sample number, genetic diversity and marker resolution were all high, and a large number of QTL for rice grain shape were identified in a single study. Moreover, almost all grain shape genes with large effect were identified in this study, such as *GS3*, *GL3.1*, *qSW5/GW5/GSE5* and *GL7/GW7/SLG7*. All above results indicated that our study was greatly pragmatic for detecting QTL for grain shape.

However, there was another limitation in the present study. Before conducting GWAS, the SNP markers were screened and markers with rare alleles were removed to decrease false positive probability. For rice grain shape, rare alleles in some important genes were the responsible factors for phenotypic variations, such as *GW2* (Song *et al.* 2007), *GS2* (Che *et al.* 2015; Hu *et al.* 2015) and *SGDP7* (Bai *et al.* 2017). Because of removal of these SNPs, some QTL with rare alleles could not be identified.

In the present study, a total of 70 QTL for three grain shape traits were identified on all chromosomes. Among them, 16 QTL stably expressed in both years, and eight genomic regions had pleiotropic effect on more than two traits. Besides, 25 QTL with phenotypic variation rates above 10% had large effects on grain shapes. They were important QTL regions for molecular breeding to improve grain shape. Among these QTL, four regions (16.73 Mb on chromosome 3, 5.37 Mb on chromosome 5, 21.49 Mb on chromosome 6 and 25.0 Mb on chromosome 7) controlling two or three traits and stably expressing in both years were the most important gene resources for improving grain shape. Among them, three regions (16.73 Mb on chromosome 3, 5.37 Mb on chromosome 5 and 25.0 Mb on chromosome 7) with large effects on grain shape were cloned as *GS3* (Mao *et al.* 2010), *qSE5/GW5/GSE5* (Shomura *et al.* 2008; Weng *et al.* 2008; Liu *et al.* 2017) and *GL7/GW7/SLG7*(Wang *et al.* 2015; Zhou *et al.* 2015). We would further mine the most favorable alleles for molecular breeding. For the region of 21.49 Mb on chromosome 6, we should firstly clone the responsible gene and mine the favorable alleles, then use it in rice breeding program.

Conclusion

In the present study, 25 large effects QTL, 16 stably expressed QTL and 8 pleiotropic QTL were identified. They were all valuable gene resources in molecular breeding for improving rice grain for desired shape.

References

Bai X, Y Huang, Y Hu, H Liu, B Zhang, C Smaczniak, G Hu, Z Han, Y Xing (2017). Duplication of an upstream silencer of *FZP* increases grain yield in rice. *Nat Plants* 3:885–893

Che R, H Tong, B Shi, Y Liu, S Fang, D Liu, Y Xiao, B Hu, L Liu, H Wang, M Zhao, C Chu (2015). Control of grain size and rice yield by GL2-mediated brassinosteroid responses. *Nat Plants* 2:15195

Fang N, R Xu, L Huang, B Zhang, P Duan, N Li, Y Luo, Y Li (2016). *SMALL GRAIN 11* controls grain size, grain number and grain yield in rice. *Rice (N.Y.)* 9:64

Hu J, Y Wang, Y Fang, L Zeng, J Xu, H Yu, Z Shi, J Pan, D Zhang, S Kang, L Zhu, G Dong, L Guo, D Zeng, G Zhang, L Xie, G Xiong, J Li, Q Qian (2015). A rare allele of *GS2* enhances grain size and grain yield in rice. *Mol Plant* 8:1455–1465

Huang X, X Wei, T Sang, Q Zhao, Q Feng, Y Zhao, C Li, C Zhu, T Lu, Z Zhang, M Li, D Fan, Y Guo, A Wang, L Wang, L Deng, W Li, Y Lu, Q Weng, K Liu, T Huang, T Zhou, Y Jing, W Li, Z Lin, ES Buckler, Q Qian, QF Zhang, J Li, B Han (2010). Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat Genet* 42:961–967

Kang HM, JH Sul, SK Service, NA Zaitlen, SY Kong, NB Freimer, C Sabatti, E Eskin (2010). Variance component model to account for sample structure in genome-wide association studies. *Nat Genet* 42:348–354

Li Y, C Fan, Y Xing, Y Jiang, L Luo, L Sun, D Shao, C Xu, X Li, J Xiao, Y He, Q Zhang (2011). Natural variation in *GS5* plays an important role in regulating grain size and yield in rice. *Nat Genet* 43:1266–1269

Liu JF, J Chen, XM Zheng, FQ Wu, QB Lin, YQ Heng, P Tian, ZJ Cheng, XW Yu, KN Zhou, X Zhang, XP Guo, JL Wang, HY Wang, JM Wan (2017). *GW5* acts in the brassinosteroid signalling pathway to regulate grain width and weight in rice. *Nat Plants* 3:17043

Luo YK, ZW Zhu, N Chen, BW Duan, LP Zhang (2004). Grain types and related quality characteristics in China. *Chin J Rice Sci* 18:135–139

Mao H, S Sun, J Yao, C Wang, S Yu, C Xu, X Li, Q Zhang 2010. Linking differential domain functions of the *GS3* protein to natural variation of grain size in rice. *Proc Natl Acad Sci USA* 107:19579–19584

Morales M (2001). Statistica, Vresion 5.5. *Bull Ecol Soc Amer* 82:126–127

Qi P, YS Lin, XJ Song, JB Shen, W Huang, JX Shan, MZ Zhu, L Jiang, JP Gao, HX Lin (2012). The novel quantitative trait locus *GL3.1* controls rice grain size and yield by regulating Cyclin-T1;3. *Cell Res* 22:1666–1680

Qiu X, K Chen, W Lv, X Ou, Y Zhu, D Xing, L Yang, F Fan, J Yang, J Xu, T Zheng, Z Li 2017. Examining two sets of introgression lines reveals background-independent and stably expressed QTL that improve grain appearance quality in rice (*Oryza sativa* L.). *Theor Appl Genet* 130:951–967

Qiu X, Y Pang, Z Yuan, D Xing, J Xu, M Dingkuhn, Z Li, G Ye (2015). Genome-wide association study of grain appearance and milling quality in a worldwide collection of indica rice germplasm. *PLoS One* 10:e0145577

Shomura A, T Izawa, K Ebana, T Ebitani, H Kanegae, S Konishi, M Yano (2008). Deletion in a gene associated with grain size increased yields during rice domestication. *Nat Genet* 40:1023–1028

Si L, J Chen, X Huang, H Gong, J Luo, Q Hou, T Zhou, T Lu, J Zhu, Y Shangguan, E Chen, C Gong, Q Zhao, Y Jing, Y Zhao, Y Li, L Cui, D Fan, Y Lu, Q Weng, Y Wang, Q Zhan, K Liu, X Wei, K An, G An, B Han (2016). *OsSPL13* controls grain size in cultivated rice. *Nat Genet* 48:447–456

Song XJ, W Huang, M Shi, MZ Zhu, HX Lin (2007). A QTL for rice grain width and weight encodes a previously unknown RING-type E3 ubiquitin ligase. *Nat Genet* 39:623–630

Tan YF, YZ Xing, JX Li, SB Yu, CG Xu, QF Zhang (2000). Genetic bases of appearance quality of rice grains in Shanyou 63, an elite rice hybrid. *Theor Appl Genet* 101:823–829

Vilhjalmsson BJ, M Nordborg (2013). The nature of confounding in genome-wide association studies. *Nat Rev Genet* 14:1–2

Wang CC, H Yu, J Huang, WS Wang, M Faruquee, F Zhang, XQ Zhao, BY Fu, K Chen, HL Zhang, SS Tai, C Wei, KL McNally, N Alexandrov, XY Gao, J Li, ZK Li, JL Xu, TQ Zheng (2020). Towards a deeper haplotype mining of complex traits in rice with RFGB v2.0. *Plant Biotechnol J* 18:14–16

Wang S, S Li, Q Liu, K Wu, J Zhang, S Wang, Y Wang, X Chen, Y Zhang, C Gao, F Wang, H Huang, X Fu (2015). The *OsSPL16-GW7* regulatory module determines grain shape and simultaneously improves rice yield and grain quality. *Nat Genet* 47:949–954

Wang W, R Mauleon, Z Hu, D Chebotarov, S Tai, Z Wu, M Li, T Zheng, RR Fuentes, F Zhang, L Mansueto, D Copetti, M Sanciangco, KC Palis, J Xu, C Sun, B Fu, H Zhang, Y Gao, X Zhao, F Shen, X Cui, H Yu, Z Li, M Chen, J Detras, Y Zhou, X Zhang, Y Zhao, D Kudrna, C Wang, R Li, B Jia, J Lu, X He, Z Dong, J Xu, Y Li, M Wang, J Shi, J Li, D Zhang, S Lee, W Hu, A Poliakov, I Dubchak, VJ Ulat, FN Borja, JR Mendoza, J Ali, J Li, Q Gao, Y Niu, Z Yue, MEB Naredo, J Talag, X Wang, J Li, X Fang, Y Yin, JC Glaszmann, J Zhang, J Li, RS Hamilton, RA Wing, J Ruan, G Zhang, C Wei, N Alexandrov, KL McNally, Z Li, H Leung (2018). Genomic variation in 3,010 diverse accessions of Asian cultivated rice. *Nature* 557:43–49

Wang X, Y Pang, C Wang, K Chen, TQ Zhu, C Shen, J Ali, J Xu, Z Li (2016). New candidate genes affecting rice grain appearance and milling quality detected by genome-wide and gene-based association analyses. *Front Plant Sci* 7:1998

Wattoo JI, S Liaqat, H Mubeen, M Ashfaq, MN Shahid, A Farooq, M Sajjad, M Arif (2019). Genetic mapping of grain nutritional profile in rice using basmati derived segregating population revealed by SSRs. *Intl J Agric Biol* 21:929–935

Weng J, S Gu, X Wan, H Gao, T Guo, N Su, C Lei, X Zhang, Z Cheng, X Guo, J Wang, L Jiang, H Zhai, J Wan (2008). Isolation and initial characterization of *GW5*, a major QTL associated with rice grain width and weight. *Cell Res* 18:1199–1209

Yang XH, JB Yan, YP Zheng, JM Yu, JS Li (2007). Reviews of associatin analysis for quantitative traits in plants. *Acta Agron Sin* 33:523–530

Zhang P, X Liu, H Tong, Y Lu, J Li (2014). Association mapping for important agronomic traits in core collection of rice (*Oryza sativa* L.) with SSR markers. *PLoS One* 9:e111508

Zhang QF (2007). Strategies for developing green uuper rice. *Proc Natl Acad Sci USA* 104:16402–16409

Zhang XJ, JW Wang, J Huang, HX Lan, CL Wang, CF Yin, YY Wu, HJ Tang, Q Qian, JY Li, HY Zhang (2012). Rare allele of *OsPPKL1* associated with grain length causes extra-large grain and a significant yield increase in rice. *Proc Natl Acad Sci USA* 109:21534–21539

Zhao DS, QF Li, CQ. Zhang, C Zhang, QQ Yang, LX Pan, XY Ren, J Lu, MH Gu, QQ Liu (2018). *GS9* acts as a transcriptional activator to regulate rice grain shape and appearance quality. *Nat Commun* 9:1240

Zhao K, CW Tung, GC Eizenga, MH Wright, ML Ali, AH Price, GJ Norton, MR Islam, A Reynolds, J Mezey, AM McClung, CD Bustamante, SR McCouch (2011). Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa. Nat Commun* 2:467

Zhou Y, J Miao, HY. Gu, XR Peng, M Leburu, FH Yuan, HW Gu, Y Gao, YJ Tao, ZY Gong, CD Yi, MH Gu, ZF Yang, GH Liang (2015). Natural variation in *SLG7* regulate grain shape in rice. *Genetics* 201:1591–1599

Zou Z, G Deng, X Peng, H Xie, X Li, L Peng, Z Zhang (2018). Effects of exogenous amino acids on nitrogen accumulation, yield and grain quality in rice. *Intl J Agric Biol* 20:2053–2058