

Genome wide association mapping for grain shape traits in *indica* rice

Yue Feng¹ · Qing Lu¹ · Rongrong Zhai² · Mengchen Zhang¹ · Qun Xu¹ ·
Yaolong Yang¹ · Shan Wang¹ · Xiaoping Yuan¹ · Hanyong Yu¹ · Yiping Wang¹ ·
Xinghua Wei¹

Received: 25 February 2016 / Accepted: 6 May 2016 / Published online: 19 May 2016
© The Author(s) 2016. This article is published with open access at Springerlink.com

Abstract

Main conclusion Using genome-wide association mapping, 47 SNPs within 27 significant loci were identified for four grain shape traits, and 424 candidate genes were predicted from public database.

Grain shape is a key determinant of grain yield and quality in rice (*Oryza sativa* L.). However, our knowledge of genes controlling rice grain shape remains limited. Genome-wide association mapping based on linkage disequilibrium (LD) has recently emerged as an effective approach for identifying genes or quantitative trait loci (QTL) underlying complex traits in plants. In this study, association mapping based on 5291 single nucleotide polymorphisms (SNPs) was conducted to identify significant loci associated with grain shape traits in a global collection of 469 diverse rice accessions. A total of 47 SNPs were located in 27 significant loci for four grain traits, and explained ~44.93–65.90 % of the phenotypic variation for each trait. In total, 424 candidate genes within a 200 kb extension region (± 100 kb of each locus) of these loci were predicted. Of them, the cloned genes *GS3* and *qSW5* showed very strong effects on grain length and grain width in our

study. Comparing with previously reported QTLs for grain shape traits, we found 11 novel loci, including 3, 3, 2 and 3 loci for grain length, grain width, grain length–width ratio and thousand grain weight, respectively. Validation of these new loci would be performed in the future studies. These results revealed that besides *GS3* and *qSW5*, multiple novel loci and mechanisms were involved in determining rice grain shape. These findings provided valuable information for understanding of the genetic control of grain shape and molecular marker assistant selection (MAS) breeding in rice.

Keywords Genome-wide association mapping · Grain shape · Rice · Single nucleotide polymorphism

Abbreviations

GL	Grain length
GW	Grain width
GT	Grain thickness
TGW	Thousand grain weight
LWR	Grain length–width ratio
QTL	Quantitative trait loci
LD	Linkage disequilibrium
SNP	Single nucleotide polymorphism
GWAS	Genome-wide association study
CTAB	Hexadecyltrimethyl ammonium bromide
MAF	Minor allele frequency
SE	Standard error
ANOVA	Analysis of variance
BLUP	Best linear unbiased prediction
cMLM	Compressed mixed linear model
PCA	Principle component analysis
G × E	Interaction of genotype and environment
GAPIT	Genome association and prediction integrated tool
MAS	Molecular assistant selection

Yue Feng and Qing Lu contributed equally to this work.

Electronic supplementary material The online version of this article (doi:10.1007/s00425-016-2548-9) contains supplementary material, which is available to authorized users.

✉ Xinghua Wei
weixinghua@caas.cn

¹ State Key Laboratory of Rice Biology, China National Rice Research Institute, Hangzhou 310006, Zhejiang, China

² Institute of Crop and Nuclear Technology Utilization, Zhejiang Academy of Agricultural Sciences, Hangzhou 310021, Zhejiang, China

Introduction

Rice (*Oryza sativa* L.) is one of the most important cereal crops and a staple food for more than one-half of the world's population. Rice yield is directly determined by four major components: number of effective tillers per plant, number of grains per panicle, the ratio of filled grains, and grain weight (Sakamoto and Matsuoka 2008; Xing and Zhang 2010). Grain size, a key factor determining grain weight, is specified by its three dimensional structures: grain length (GL), grain width (GW), and grain thickness (GT) (Fan et al. 2006; Zuo and Li 2014). In breeding applications, grain size and weight are important traits for rice yield and quality.

Grain size and weight are generally recognized as quantitative traits that are controlled by multiple genes and affected by environment factors. To date, more than 400 QTLs that control grain size and weight have been detected by using various mapping populations (<http://www.genome.org/qtl>). Several major QTLs controlling grain size and weight have been identified and functionally characterized, such as *GS3* (Fan et al. 2006; Mao et al. 2010), *GW2* (Song et al. 2007), *qSW5/GW5* (Shomura et al. 2008; Weng et al. 2008), *GS5* (Li et al. 2011), *qGL3/qGL3.1* (Qi et al. 2012; Zhang et al. 2012), *GW8* (Wang et al. 2012), *TGW6* (Ishimaru et al. 2013) and *GL7/GW7* (Wang et al. 2015a, b). Among these, *GS3* encoding a putative transmembrane protein and *qGL3/qGL3.1* encoding a protein phosphatase with Kelch-like repeat domain function as a negative regulator for grain length. For grain width, *GW2* encoding a RING-type E3 ubiquitin ligase and *qSW5/GW5* encoding a novel protein act as negative regulators, whereas *GS5* encoding a putative serine carboxypeptidase and *GW8* encoding a transcription factor with SBP domain are positive regulators. *TGW6* encodes a novel protein with indole-3-acetic acid (IAA)-glucose hydrolase activity and loss of *TGW6* function results in an increase in grain weight and yield. *GL7/GW7* encodes a protein homologous to *Arabidopsis thaliana* LONGIFOLIA proteins, which regulates longitudinal cell elongation. The *GL7/GW7* locus containing a 17.1-kb tandem duplication leads to upregulation of *GL7/GW7* and downregulation of its nearby negative regulator, resulting in an increase in grain length and improvement of grain appearance quality. Functional characterizations of these genes have greatly enriched our knowledge of the molecular mechanisms determining grain size and weight in rice. However, additional genes controlling grain size and weight remain to be identified.

Association mapping utilized linkage disequilibrium (LD) to examine the marker-trait associations, and enabled researchers to exploit natural variation and identify novel genes for complex traits (Zhu et al. 2008). With the

development of SNP assays, high throughput genotyping technologies and associated statistical methods, association mapping has been widely used in various plant species, including *A. thaliana* (Olsen et al. 2004; Ehrenreich et al. 2009), soybean (Jun et al. 2008), wheat (Brescghello and Sorrells 2006), maize (Palaisa et al. 2003; Wilson et al. 2004; Camus-Kulandaivelu et al. 2006), foxtail millet (Jia et al. 2013) and rice (Agrama et al. 2007; Wen et al. 2009; Huang et al. 2010; Zhao et al. 2011). Association mapping has potential advantages over conventional linkage analysis and QTL mapping, such as broader allele coverage, higher mapping resolution, shorter time-consuming and improving the cost effectiveness. Association mapping has become a useful and robust strategy complementary to classical bi-parental mapping and has the power to genetically map multiple traits simultaneously (Huang and Han 2014).

In the present study, association mapping of four traits (GL, GW, grain length–width ratio (LWR) and thousand-grain weight (TGW)) was performed with a panel of 469 accessions using a custom-designed array contained 5291 genomic SNPs. The objectives of this study were to (1) investigate the population structure and genetic diversity for grain shape in a global rice germplasm collection; (2) identify the novel SNPs and loci associated with grain shape in rice; (3) explore design of gene pyramiding breeding strategies for cultivar genetic improvement. These results will increase our understanding of the molecular mechanisms underlying grain size and weight, and may provide some new information for rice molecular design breeding.

Materials and methods

Plant materials and phenotyping

To reduce spurious genetic associations caused by population structure, a set of 469 global diverse *indica* accessions with rice grain traits variation were chosen to construct this genome-wide association study (GWAS) population (Supplementary Table S1). The seeds of all accessions were collected, stored and supplied by China National Center for Rice Improvement, China National Rice Research Institute (CNRRI). All the rice accessions were planted in the following four environments: the Lingshui Experiment Station of CNRRI (LS; N 18°48', E 110°02') in Hainan, China, in the winter of 2013–2014; the Hangzhou Experiment Station of CNRRI (HZ; N 30°32', E 120°12') in Zhejiang, China, in the summer of 2013–2014. For field experiments, the accessions were grown in a 6 × 6 completely randomized block design with three

replications. The space was 20 cm between the plants and 25 cm between the rows. Four plants in the middle of each row were harvested individually to measure the grain traits. Twenty full-filled rice grains were randomly selected from each plant for trait measurement. GL was estimated by placing 20 grains one by one in a straight line along a ruler, and then arranged by breadth to measure grain width. The averaged GL and GW of 20 grains as the trait values of that line were used for data analysis. The LWR is equal to GL divided by its GW. TGW was calculated based on 200 grains and converted to 1000-grain weight.

DNA extraction and genotyping

Genomic DNA was extracted from leaf tissue of one plant per accession using the CTAB method (Murray and Thompson 1980). All the accessions were genotyped using the Illumina RiceSNP6k BeadChip containing 5291 SNPs, which were chosen from the Rice Haplotype Map Project Database (<http://www.ncgr.ac.cn/RiceHap2>) (Huang et al. 2010). Genotypes were called using the program GenomeStudio (Illumina, San Diego, Calif. USA). The quality of each SNP was confirmed manually, and low quality SNPs (call rate <80 % and minor allele frequency (MAF) <0.05) were removed from the dataset. Finally, 3951 SNPs were obtained for the association mapping (Supplementary Table S2).

Data analysis

The mean, standard error (SE) and broad-sense heritability (H_B^2) were calculated using the Excel 2007. The percentage of phenotypic variation explained by population structure (R_{PCA}^2) was computed using SAS system 9.0 (SAS, Inc., Cary, NC), as well as the ANOVA. Correlation coefficients were run in R “corrgram” (<https://cran.r-project.org/web/packages/corrgram/>), and the best linear unbiased prediction (BLUP) was carried out in R “lme4” (<https://cran.r-project.org/web/packages/lme4/>) for estimating phenotypic values of each line in four environments.

Genome-wide association mapping

For association analysis of our panel, genome association and prediction integrated tool (GAPIT) with compressed mixed linear model (cMLM) and population parameters previously defined (P3D) was performed under R environment (Lipka et al. 2012; Zhang et al. 2010). And the top 10 principle components were used as a covariant. The kinship matrix estimated by these SNPs data was combined with population structure to improve statistical power of our genome-wide association mapping. As Bonferroni

correction ($1/3951 = 2.5E-04$) was too conservative (Nakagawa 2004), a compromised threshold of $P = 0.001$ was used to detect the significant association signals. To obtain the loci with the lowest P value, redundant loci were filtered in a 200 kb genomic window. And the candidate gene prediction was performed from the Rice Haplotype Map Project Database (<http://202.127.18.221/RiceHap2/>). Finally, the allele with an increasing effect was defined as elite allele and vice versa. The elite alleles of association loci were used to evaluate the pyramiding effect.

Results

Grain shape variation among accessions

In this study, 469 *indica* accessions collected from 20 countries (Fig. 1a) were used as the genome wide association mapping panel. Four grain shape traits, namely GL, GW, LWR and TGW of 469 rice accessions were measured in four environments after harvested. Extensive and significant phenotypic variations were observed for the four grain traits in Hangzhou (HZ) and Lingshui (LS) during 2013–2014 (Table 1; Fig. 1b, c; Supplementary Figure S1; Supplementary Table S3). The mean of GL over the 469 accessions in each environment was 8.58, 8.51, 8.81 and 8.86, respectively. Over the four environments the minimum GL value was in 2013 HZ (6.67 cm), and the maximum value was in 2014 HZ (11.37 cm). GW had means of 2.69, 3.12, 2.81 and 2.89 in each environment, respectively. GW value ranged from 1.89 to 2.17 cm in the four environments. The mean of LWR over the 469 accessions in each environment was 3.24, 2.76, 3.19 and 3.10, respectively. Over the four environments the minimum LWR value was in 2013 LS (1.94), and the maximum value was in 2013 HZ (4.98). TGW had means of 23.94, 28.44, 25.04 and 26.88 g in each environment, respectively. TGW value ranged from 14.35 to 50.75 g in the four environments. The distribution of the four grain traits in all four environments showed continuous variation, and the phenotypic segregation approximately fit a normal distribution (Fig. 1d–g). This indicated that grain shape traits were governed by multiple loci in this association panel. The broad-sense heritability (H_B^2 %), averaged across four environments, of GL, GW, LWR and TGW, was 79.1, 84.4, 89.6 and 83.7 %, respectively. The ANOVA indicated that effects of genotype (G), environment (E) and their interaction were high significantly ($P < 0.001$) (Table 1).

The results of Pearson correlation analysis showed that the phenotypic correlations between traits were similar in the four testing environments (Fig. 2). GW showed significantly negative correlation with GL and

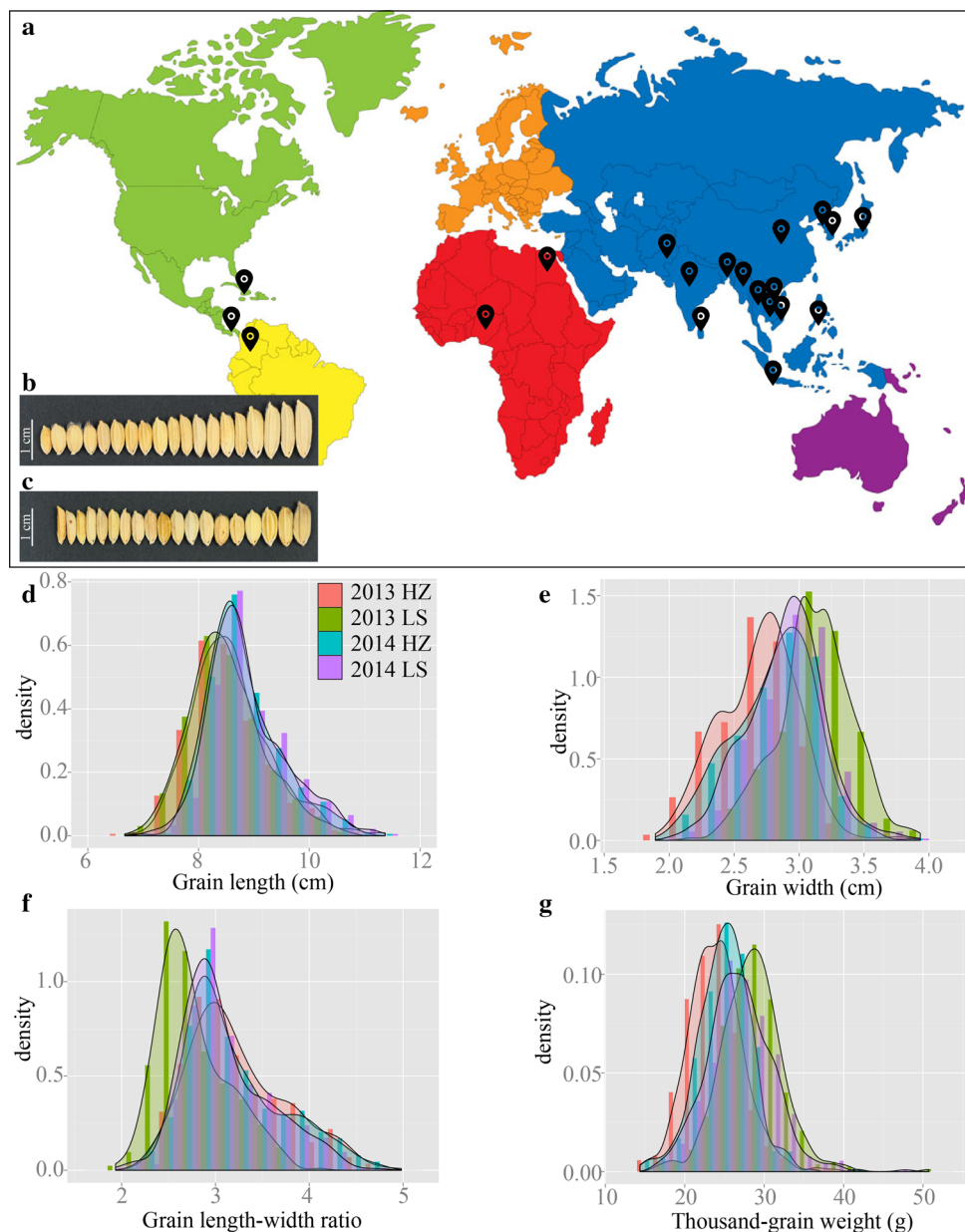


Fig. 1 Material distribution and grain shape traits diversity. **a** The worldwide distribution of 469 *indica* accessions. The *black dots* represent the country-specific distribution. **b** Different grain length in the germplasm. *Bar* 1 cm. **c** Different grain width in the germplasm. *Bar* 1 cm. **d** Histogram of grain length in multiple environments.

e Histogram of grain width in multiple environments. **f** Histogram of grain length–width ratio in multiple environments. **g** Histogram of thousand-grain weight in multiple environments. *LS* Lingshui, *HZ* Hangzhou

LWR, and highly positive correlation with TGW in all environments. TGW exhibited significantly or very significantly negative correlation with LWR, and highly positive correlation with GW. The correlation coefficients ranged from -0.88 between LWR and TGW, to 0.71 between LWR and GL (Fig. 2). These results demonstrated that rice grain traits were highly related to each other, and this provided valuable knowledge for rice grain shape modifications.

Genetic structure and linkage disequilibrium estimation

STRUCTURE software was used to assess the subpopulation genetic structure of the 469 *indica* rice accessions based on the distribution of the 3,951 SNPs across 12 rice chromosomes. Our previous study showed that the *indica* panel could be classified into four subpopulations (Lu et al. 2015). In this GWAS mapping, the population effect showed by a

Table 1 Phenotypic variation for grain traits across four environments in 469 *indica* accessions

Trait	Environment	Mean ± SE	Range	R ² _{PCA} (%) ^a	H ² _B (%) ^b	G × E ^c
GL (cm)	2013, HZ	8.58 ± 0.03	6.67–10.92	29.10	79.1	***
	2013, LS	8.51 ± 0.03	6.83–10.67	27.24		
	2014, HZ	8.81 ± 0.03	6.92–11.37	31.65		
	2014, LS	8.86 ± 0.03	7.11–11.20	32.22		
GW (cm)	2013, HZ	2.69 ± 0.01	1.89–3.79	37.50	84.4	***
	2013, LS	3.12 ± 0.01	2.17–3.93	31.10		
	2014, HZ	2.81 ± 0.01	1.92–3.90	39.37		
	2014, LS	2.89 ± 0.01	2.02–3.90	32.53		
LWR	2013, HZ	3.24 ± 0.02	2.17–4.98	41.23	89.6	***
	2013, LS	2.76 ± 0.02	1.94–4.11	38.52		
	2014, HZ	3.19 ± 0.02	2.15–4.74	43.73		
	2014, LS	3.10 ± 0.02	1.97–4.75	38.26		
TGW (g)	2013, HZ	23.94 ± 0.16	14.50–38.85	19.16	83.7	***
	2013, LS	28.44 ± 0.18	16.70–50.75	21.82		
	2014, HZ	25.04 ± 0.15	14.35–41.75	18.69		
	2014, LS	26.88 ± 0.19	16.45–47.80	23.96		

GL grain length, GW grain width, LWR grain length–width ratio, TGW thousand-grain weight, LS Lingshui, HZ Hangzhou, SE standard error

*** Significant at $P = 0.001$

^a Phenotypic variation explained by population structure

^b Heritability in the broad sense

^c G × E, interaction of genotype and environment

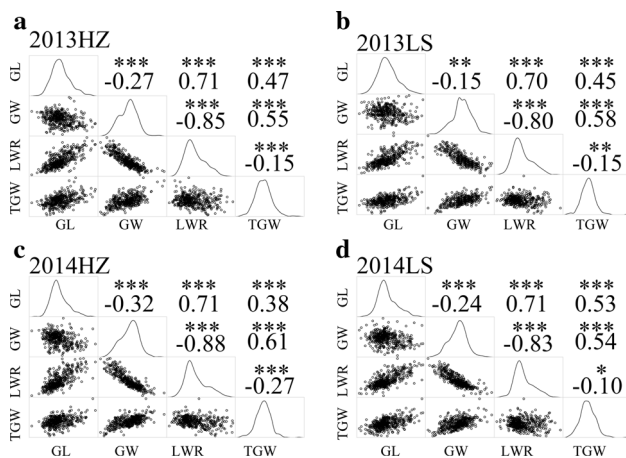


Fig. 2 Correlation of four grain shape traits in four environments. The upper panel contains the correlation coefficients, and the lower panel contains the distributions of the four traits. The diagonal represents the density line of the traits. *, ** and *** represent significant at 0.05, 0.01 and 0.001, respectively. GL, GW, LWR and TGW represent grain length, grain width, grain length–width ratio and thousand grain weight, respectively. HZ and LS represent Hangzhou and Lingshui, respectively

principal component analysis (PCA) component and relatedness estimated automatically from the genotypic data were evaluated using GAPIT as well as GWAS (Lipka et al. 2012;

Zhang et al. 2010). The scree plot generated through GAPIT indicated the first two PCs as informative, and then decreased gradually (Fig. 3a, b). Moreover, until the 10th PC component, the variance value almost unchanged, thus the top 10 PCs were used as a covariate to adjust the GWAS results (Fig. 3a). In addition, the heatmap of kinship value showed that most of the value concentrated at ~0.5 level indicating a weak relatedness in the complex GWAS panel (Fig. 3c). This evaluation is consistent with our previous study suggesting that the GAPIT is credible for population structure and relative kinship estimations (Lu et al. 2015). In the present study, the linkage disequilibrium (LD) decay distance was ~109.37 kb at which the LD parameter (r^2) dropped to half of its maximum value. In addition, the LD decay distance differed among chromosomes and ranged from 96.15 kb on chromosome 5 to 421.39 kb on chromosome 7.

Genome-wide association analysis

In this study, the association analysis was performed using BLUP method predicted for each accession to reduce the environment effects and simplify the calculations. The stringent criterion of $-\log_{10}(P) \geq 3.0$ under four environments was used for determining the association significance of the four grain traits. Using the GWAS approach,

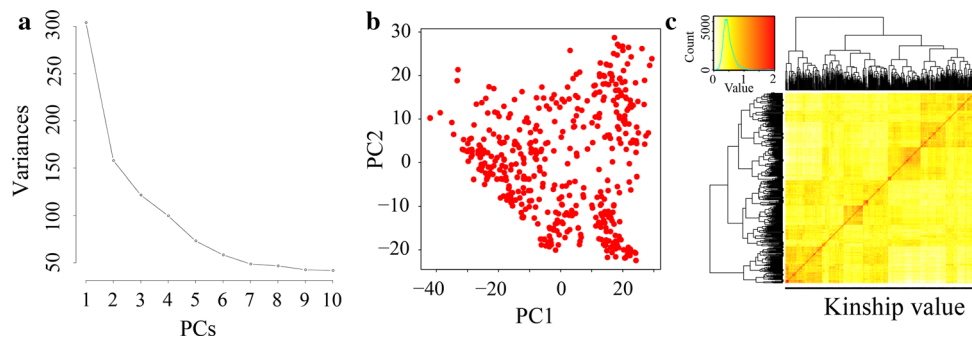


Fig. 3 Population structure of current association *panel* which consisted mostly of the *indica* accessions. **a** Scree plot from GAPIT showing the selection of PCs for association study. **b** Variation of the top two principal components. **c** Kinship values

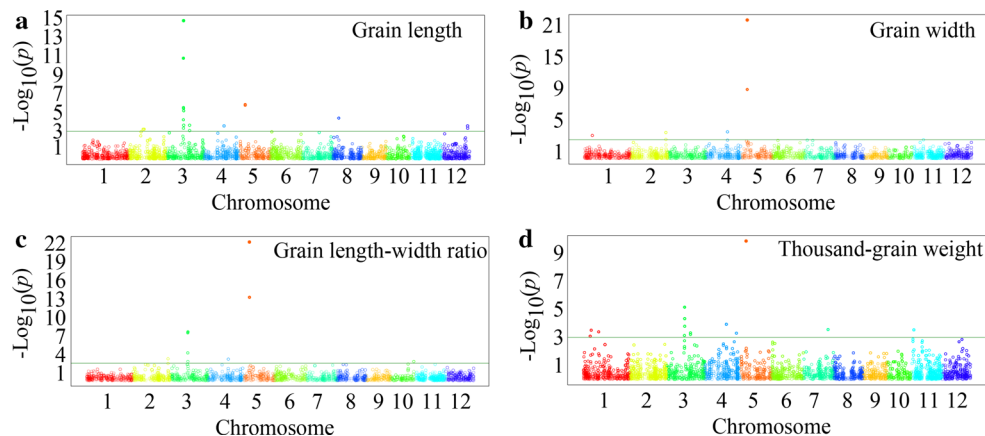


Fig. 4 Manhattan plots of genome wide association mapping for four rice grain traits. **a** Grain length, **b** grain width, **c** grain length–width ratio, **d** thousand grain weight. Green lines represent the threshold at $-\log_{10}(P) \geq 3$

we successfully identified both known genes and previously reported QTLs from rice as well as novel candidate loci in the rice genome. The results of our genome-wide association scans were summarized in Fig. 4 and Table 2 where we showed the SNP trait associations discovered in the association panel. A total of 27 significant loci were identified for the four grain traits. For the 27 significant loci, 424 candidate genes were obtained within a 200-kb genomic region (± 100 kb from the peak SNP) from the Rice Haplotype Map Project Database (<http://202.127.18.221/RiceHap2/>) (Supplementary Table S4). In addition, the distributions of the 3951 SNPs were analyzed and 47 SNPs were located in the 27 significant loci for the four grain traits (Supplementary Table S2, Table S4). Of the 27 loci, *GS3* and *qSW5* showed very strong effects on grain length and width. Simultaneously, in order to investigate environment-specific and common loci under multiple environments, we performed association mapping using TASSEL version 4.0 (Bradbury et al. 2007) under each environment (Supplementary Table S5).

For the GL trait, 8 major loci, explaining $\sim 50.23\%$ of the phenotypic variation, were identified on chromosomes 2, 3, 4, 5, 8 and 12 (Table 2). Among them, the *GS3* locus (seq-rs1614) was confirmed and explained $\sim 8.50\text{--}14.73\%$ of the total phenotypic variation under four environments (Fig. 5a; Supplementary Table S6), and this result was consistent with previous reports (Huang et al. 2010; Zhao et al. 2011). Furthermore, the loci, seq-rs1610, seq-rs2061 and seq-rs4660, on chromosomes 3, 4 and 10 were identified across three environments (Fig. 5a; Supplementary Table S6). Additionally, the loci, seq-rs1604 and seq-rs2427, on chromosomes 3 and 5 were detected in both 2013 and 2014 HZ environments (Fig. 5a; Supplementary Table S6). The locus (seq-rs2718) on chromosome 6 was identified in both 2014 HZ and LS environments (Fig. 5a; Supplementary Table S6).

For the GW trait, 4 major loci, explaining $\sim 58.57\%$ of the phenotypic variation, were identified on chromosomes 1, 2, 4 and 5 (Table 2). Among them, the previously cloned *qSW5* locus showed very strong effect on GW and had high

Table 2 Summary of SNPs significantly associated with grain traits using the BLUP method

Trait	Marker	Chr.	Position	Allele ^a	MAF	P value	R ² (%) ^b	Known QTL	Marker	References
GL	seq-rs918	2	13760905	G/A	0.17	6.68E–04	50.23			
	seq-rs919	2	14722011	T/G	0.17	6.68E–04				
	seq-rs1614	3	16939138	T/C	0.25	1.87E–15		<i>GS3</i>		Fan et al. (2006)
	seq-rs1697 ^c	3	22579680	G/T	0.21	2.63E–04		<i>qGL3a</i>	RM15456	Li et al. (2010)
	seq-rs2123	4	19733128	C/T	0.05	2.89E–04		<i>qGL4b</i>	RM5586	Kato et al. (2011)
	gwseq-rs14 ^d	5	5377176	T/C	0.21	1.91E–06		<i>qSW5</i>		Shomura et al. (2008)
	seq-rs3750	8	5964991	C/T	0.06	4.35E–05				
	seq-rs5920	12	25594275	G/A	0.36	2.91E–04		<i>qGL12</i>	RM2854	Li et al. (2010)
GW	seq-rs196	1	7562923	A/G	0.29	2.24E–04	58.57			
	seq-rs1303	2	33850349	T/C	0.25	7.62E–05				
	seq-rs2139	4	21062959	A/G	0.42	6.02E–05				
	seq-rs2427 ^d	5	5359498	G/A	0.32	1.12E–22		<i>qSW5</i>		Shomura et al. (2008)
LWR	seq-rs1303	2	33850349	T/C	0.25	2.75E–04	65.90			
	seq-rs1614	3	16939138	T/C	0.25	9.46E–09		<i>GS3</i>		Fan et al. (2006)
	seq-rs2123	4	19733128	C/T	0.05	3.03E–04		<i>qLWR4</i>	RM6997	Ying et al. (2012)
	seq-rs2427	5	5359498	G/A	0.32	1.48E–23		<i>qSW5</i>		Shomura et al. (2008)
	seq-rs4717	10	22631657	T/C	0.08	7.37E–04				
TGW	seq-rs184	1	6667379	A/G	0.45	9.17E–04	44.93	<i>Gw1-1</i>	RM10398	Yu et al. (2008)
	seq-rs196	1	7562923	A/G	0.29	3.70E–04		<i>QGwt1c, qBRW1a</i>	RM259	Xu et al. (2002), Wang et al. (2003)
	seq-rs288	1	14827946	C/T	0.30	4.76E–04				
	seq-rs1614	3	16939138	T/C	0.25	9.00E–06		<i>GS3</i>		Fan et al. (2006)
	seq-rs1698 ^c	3	22582110	C/A	0.21	5.42E–04		<i>qTGW3-4, qGW3.8</i>	RM16	Marathi et al. (2012), Zhang et al. (2013)
	seq-rs2139	4	21062959	A/G	0.42	1.41E–04				
	seq-rs2255	4	30599877	A/G	0.18	5.67E–04		<i>qTGW4-1</i>	RM3276	Marathi et al. (2012)
	seq-rs2427	5	5359498	G/A	0.32	2.03E–10		<i>qSW5</i>		Shomura et al. (2008)
	seq-rs3526	7	23662096	C/A	0.07	3.19E–04				
	seq-rs4745	11	3031650	T/C	0.34	3.40E–04		<i>GWt11</i>	RM332	Yoshida et al. (2002)

GL grain length, GW grain width, LWR grain length–width ratio, TGW thousand-grain weight

^a Major allele/minor allele

^b Phenotypic variation explained by all of the significant loci

^c The two SNPs can be considered as the same locus (~ 2 kb)

^d The two SNPs can be considered as the same locus (~ 17 kb)

R² values (~ 13.6–27.79 %) in the four environments (Supplementary Table S5), and this result was consistent with previous reports (Huang et al. 2010; Zhao et al. 2011). The locus (seq-rs2434) as well as the *qSW5* (seq-rs2427), on chromosome 5 was detected across four environments (Fig. 5b; Supplementary Table S6). In addition, eleven and eighteen significant loci were identified across two and three environments, respectively (Fig. 5b; Supplementary Table S6). Moreover, the loci, seq-rs196, seq-rs1303 and seq-rs2139, on chromosomes 1, 2 and 4 were also detected for TGW, LWR and TGW, respectively (Table 2).

In our study, marker-trait association analyses revealed that 5 loci associated with LWR, locating on chromosomes 2, 3, 4, 5 and 10 (Table 2; Fig. 4c). These loci could

explain up to 65.90 % of the total phenotypic variation (Table 2). Among them, the known genes *GS3* and *qSW5* were showed clear signal for LWR in all the four environments (Fig. 5c; Supplementary Table S6). The locus (seq-rs1610) on chromosome 3 was also detected across four environments (Supplementary Table S6). Additionally, the loci (seq-rs944 and seq-rs2123) on chromosomes 2 and 4 were detected in both 2013 and 2014 HZ, and the loci (seq-rs3533 and seq-rs3537) on chromosome 7 were identified in both 2013 and 2014 LS (Fig. 5c; Supplementary Table S6).

For the TGW trait, 10 major loci, explaining ~ 44.93 % of phenotypic variation, were detected on chromosomes 1, 3, 4, 5, 7 and 11, and six loci were located in the adjacent

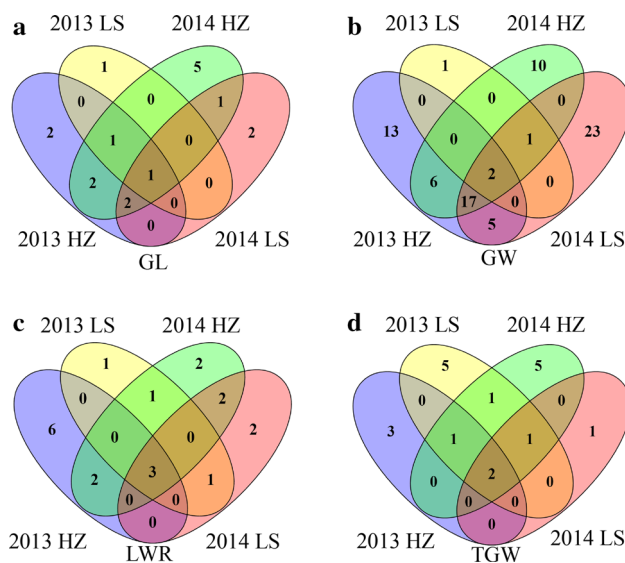


Fig. 5 Venn diagrams showing unique and shared SNPs for grain shape traits among four environments. **a** *GL* grain length. **b** *GW* grain width. **c** *LWR* grain length–width ratio. **d** *TGW* thousand grain weight. *LS* and *HZ* represent Lingshui and Hangzhou, respectively

or overlapping regions previously reported to be associated with *TGW* (Table 2). Moreover, the loci (seq-rs288 and seq-rs2139) on chromosomes 1 and 4 were identified across three environments (Fig. 5d; Supplementary Table S6). Furthermore, we also identified the known genes (*GS3* and *qSW5*) for *TGW* in all four environments (Fig. 5d; Supplementary Table S6). Additionally, the locus (seq-rs5793) on chromosome 12 was detected in both 2013 *LS* and 2014 *HZ* (Fig. 5d; Supplementary Table S6).

Gene linkage or pleiotropy

Gene linkage and pleiotropy are common phenomena in plant genetics. A matrix summarizing the linkage or pleiotropy of the loci associated with four grain traits were shown in Fig. 6. In our study, 12 known or new candidate SNPs showed clear signal linkage or pleiotropy that was associated with multiple grain traits in the four environments. Among them, the loci seq-rs1614 and seq-rs2427 underlying grain size (*GS3* and *qSW5*) were shown to be significantly associated with *GL*, *LWR*, *TGW* and all four grain traits, respectively (Fig. 6). Previous studies showed that *GS3* had pleiotropic effects on grain length and weight (Fan et al. 2006; Mao et al. 2010), and *qSW5* was associated with grain width and weight (Shomura et al. 2008; Weng et al. 2008). These associations were also supported by Pearson correlation analysis based on traits measured in all four environments (Fig. 2). In addition, we detected the locus (seq-rs2123) on chromosome 4 that were

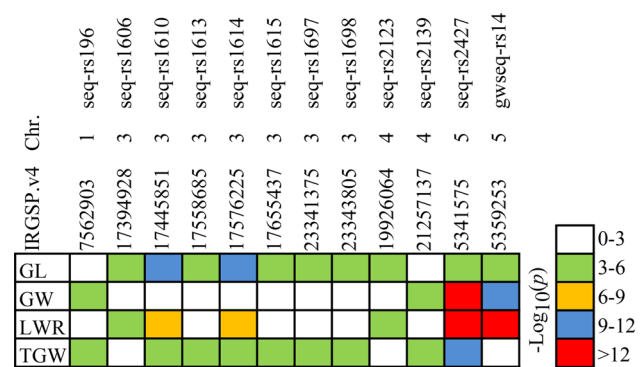


Fig. 6 Summary of significant trait-marker associations for four grain traits across genomic regions. *GL*, *GW*, *LWR* and *TGW* represent grain length, grain width, grain length–width ratio and thousand grain weight, respectively

significantly associated with both *GL* and *LWR*, and we found a locus (seq-rs196) on chromosome 2 that was significantly associated with *GW* and *TGW*. Moreover, the average correlation coefficients of *GL*–*LWR* ($r = 0.71$, $P < 0.001$) and *GW*–*TGW* ($r = 0.57$, $P < 0.001$) were highly significant in all four environments. Additionally, some new candidate loci possessing gene linkage or pleiotropy were observed as well (Fig. 6). Exploring and utilizing the linkage or pleiotropy of the loci underlying grain traits would be beneficial for rice grain yield improvement.

Pyramiding elite alleles for breeding

The objective of gene pyramiding in molecular breeding is to combine a series of target alleles in a specific line or variety (Servin et al. 2004). We examined the efficacy of pyramiding elite alleles into an individual plant for four grain traits (Supplementary Table S7). Besides *GS3* (seq-rs1614) and *qSW5* (seq-rs2427), such as the elite allele of the locus seq-rs1698 showed potential breeding value for *GL* and *TGW* (Supplementary Table S7). This result was also confirmed by classical QTL mapping in different biparental mapping populations (Li et al. 2010; Marathi et al. 2012; Zhang et al. 2013). Without considering the interaction effects among these loci, mean values for phenotypic traits except *GL* significantly increased linearly with pyramiding more elite alleles in rice cultivars (Fig. 7). For the *GL* trait, when pyramiding 0–3 elite alleles, the mean phenotypic values had no significant changes, and the mean phenotypic values increased significantly with pyramiding 4 elite alleles, and the mean phenotypic values were significantly higher when pyramiding 5–7 elite alleles in rice varieties (Fig. 7). These results indicated that enhancing the frequency of elite alleles has a significant phenotypic effect on rice grain traits improvements.

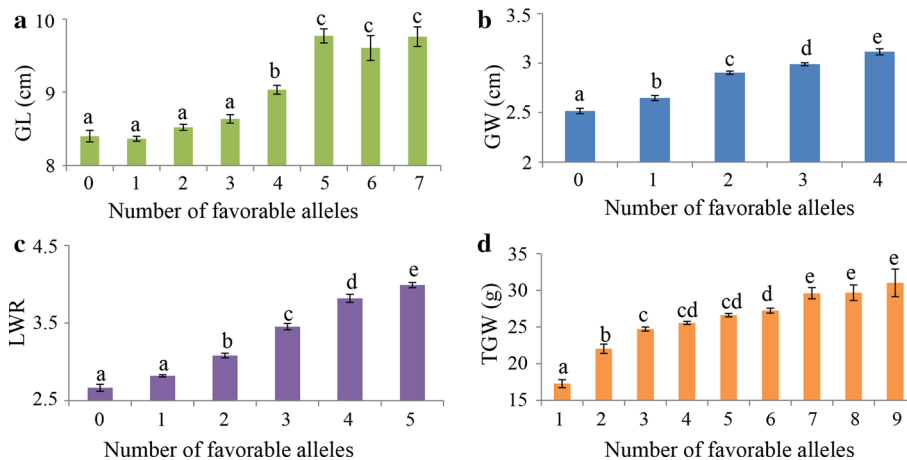


Fig. 7 Pyramid effect analysis for different numbers of elite alleles. *X*-axis represents the number of elite alleles carried by the accessions and *Y*-axis represents trait mean value *GL*, *GW*, *LWR* and *TGW* represent grain length, grain width, grain length–width ratio and thousand grain weight, respectively. All bars represent standard error

Discussion

Rice is one of the most staple foods in the world, feeding more than half of the global population especially in Asian countries. It has been estimated that a 40 % increase in rice production by 2030 will be needed to meet the predicted demand of the growing world population (Khush 2005). Rice grain shape is one of the most important factors determining rice yield (Huang et al. 2013). Thus, identification of major QTLs for grain shape and grain weight is an important objective of rice genetic research and breeding programs. In this study, genome-wide association mapping was performed to identify the loci for four grain shape traits in 469 rice accessions. The sample size of our study is larger or similar to the number of accessions used in genome wide association study in rice (Huang et al. 2010; Zhao et al. 2011). The 469 diverse accessions were from 20 countries, which represented all major rice growing regions of the world (Fig. 1a). Four grain shape traits including *GL*, *GW*, *LWR* and *TGW* were measured in four environments. We successfully identified 27 loci for the four grain traits, comprising 17 distinct regions distributed on all 12 chromosomes except chromosomes 6 and 9.

In rice, more than 400 QTLs that control grain shape traits have been detected by using various mapping population (Huang et al. 2013). To date, 20 genes associated with grain size and weight have been isolated by map-based cloning strategies (Huang et al. 2013; Zuo and Li 2014; Duan et al. 2013; Hu et al. 2015; Liu et al. 2015a, b; Wang et al. 2015a, b). We compared the positions of the significant loci identified in this study with the positions of the reported genes or QTLs for the four grain traits. Of the 27 identified loci, the positions of 16 loci were close to

those reported in previous studies, including 5, 1, 3 and 7 loci for *GL*, *GW*, *LWR* and *TGW*, respectively (Table 2).

As is known, *GS3* and *qSW5* are major QTLs for *GL* and *GW* (Fan et al. 2006; Shomura et al. 2008), which showed very strong effect for the same phenotypes in our study. This result was consistent with the GWAS results for *GL* and *GW* by Huang et al. (2010) and Zhao et al. (2011). Kato et al. (2011) mapped a *GL* QTL (*qGLAb*) to a 3.0 Mb region on chromosome 4 between marker RM5586 and RM3524. Interestingly, we detected a locus (seq-rs2123) for both *GL* and *LWR* in this chromosome region (Table 2), within which overlapped with a QTL for *LWR* reported by Ying et al. (2012). In addition, we detected 2 loci (seq-rs1697 and seq-rs5920) for *GL* on chromosome 3 and 12, respectively (Table 2), which have also been located in similar positions in different mapping population (Li et al. 2010). Moreover, Yu et al. (2008) mapped a *TGW* QTL (*Gw1-1*) to a 392.9 kb region on chromosome 1 between marker RM10376 and RM10398. We also detected a locus (seq-rs184) for *TGW* in the overlapped region. In the present study, 2 significant loci (seq-rs1698 and seq-rs2255) for *TGW* were located in the same genomic regions as those previously reported by Marathi et al. (2012). Additionally, three out of eight associated loci for *TGW* were mapped in adjacent intervals on the same chromosome as those reported in previous studies (Table 2). These results demonstrated that our 6K SNP array could capture the major common variants responsible for grain shape traits. However, no significant loci for grain traits were detected on chromosomes 6 and 9. One possible explanation was that the QTL effects were too small to be detected in our association panel, and another reason might be due to the lack of SNP coverage in the potential QTL regions. On the other hand, we identified 11 novel loci,

including 3, 3, 2 and 3 loci for GL, GW, LWR and TGW, respectively (Table 2). Validation of these new loci would be performed in the future studies.

QTL \times environment interaction is an important component for quantitative traits (Xing et al. 2002). In the present study, we observed some significant loci were only detected in one environment. For example, the associated loci, seq-rs3156, seq-rs4859, seq-rs768 and seq-rs2255 for GL, GW, LWR and TGW, respectively, were only detected in 2013 HZ. Result of ANOVA also showed that the $G \times E$ interactions were highly significant for all rice grain traits. Therefore, the QTL \times environment interaction should not be ignored if molecular marker assistant selection (MAS) is applied for the manipulation of rice grain shape traits. However, the stable expression of a QTL across a broad range of agrometeorological conditions is a critical factor when breeding for wide adaptation (Maccaferri et al. 2008). In this study, the major loci (*GS3* and *qSW5*) were detected across all four environments. These results were also proved by those obtained NILs carrying various alleles in the same background of ZhenShan97 (Lu et al. 2013) and supported by other GWAS experiments (Huang et al. 2010; Zhao et al. 2011). These results suggested that the major loci were first fixed in cultivated variety free of environmental influences due to strong human selection. Additionally, 3, 18 and 2 significant loci for GL, GW and TGW, respectively, were detected across three environments (Fig. 5; Supplementary Table S6). Moreover, 3, 11, 6 and 1 significant loci for GL, GW, LWR and TGW, respectively, were detected across two environments (Fig. 5; Supplementary Table S6). These stable significant loci affecting grain traits across different environments may be helpful in MAS of rice grain shape breeding.

The colocation of QTLs for multiple traits may be due to either pleiotropy or tightly linked genes controlling related traits (Brown 2002; Chen and Lubberstedt 2010). Our results showed that the associated loci *GS3* and *qSW5* exhibited strong pleiotropic effects on grain size and grain weight. Besides, 5 overlapped regions (seq-rs2123, seq-rs184, seq-rs196, seq-rs1698 and seq-rs2255) were found to be associated with grain traits in this study (Table 2). Among them, the locus (seq-rs2123) was detected for both GL and LWR, which was closely to the position identified by Kato et al. (2011) for GL and Ying et al. (2012) for LWR. In addition, the seq-rs1697 locus was detected for GL and TGW on chromosome 3. Its position was coincident with previously identified QTLs (Li et al. 2010; Marathi et al. 2012). Moreover, we identified a new locus (seq-rs2139) on chromosome 4 for GW and TGW across three environments. The Pearson correlation analysis also proved that the traits with co-localized loci were correlated with each other in all four environments (Fig. 2). Accumulation of favorable alleles in linkage blocks is very important and useful for

breeders to implement grain shape improvement programs. Characterization and validation studies involving joint linkage and association mapping, combining with fine mapping to identify the novel genes and alleles underlying our association hits, would help us more clearly understand the relationship between these candidate genes and the phenotypes observed in our study, and provide breeders with the effective genetic tools to break unfavorable linkages and exploit these valuable alleles.

In the present study, we identified 27 loci associated with four grain traits via GWAS and convincingly demonstrated that the rice grain shape is complex trait controlled by many genes of major or minor effect. We also refined candidate chromosomal regions for the known QTLs, including the cloned genes *GS3* and *qSW5*. Moreover, new candidate loci were obtained, and genetic variations combined with phenotypic variances were observed. Results of the present study demonstrated that genome-wide association mapping in rice could complement and enhance the information from linkage-based QTL studies toward the implementation of MAS in breeding programs. Considering the effect and distribution of novel loci associated with grain shape in our study, further validation is needed to unearth the relationship between these candidate loci and the phenotypes and facilitate exploring the molecular mechanisms governing grain shape suitable for rice breeding programs.

Author contribution statement YF and XHW designed the experiments. YF, QL, RRZ and MCZ performed the phenotyping. QX, YF, RRZ and YLY carried out the genetic studies. XPY, HYY and YPW managed the materials. XHW designed the overall project. YF analyzed the data and drafted the manuscript. SW and YLY helped to revise the manuscript. All of the authors read and approved the final manuscript.

Acknowledgments We sincerely thank all the participants. This work was supported by Zhejiang Provincial Natural Science Foundation of China (LY15C130005), the National Natural Science Foundation of China (31200916), the Agricultural Sciences and Technologies Innovation Program of the Chinese Academy of Agricultural Sciences (CAAS) and the Ministry of Science and Technology of China (2013CBA01404 and 2013BAD01B02-14).

Compliance with ethical standards

Conflict of interest The authors declare that they have no competing interests.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Agrama HA, Eizenga GC, Yan W (2007) Association mapping of yield and its components in rice cultivars. *Mol Breed* 19:341–356
- Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* 23:2633–2635
- Breseghello F, Sorrells ME (2006) Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars. *Genetics* 172:1165–1177
- Brown JKM (2002) Yield penalties of disease resistance in crops. *Curr Opin Plant Biol* 5:339–344
- Camus-Kulandaivelu L, Veyrieras JB, Madur D, Combes V, Fourmann M, Barraud S, Dubreuil P, Gouesnard B, Manicacci D, Charcosset A (2006) Maize adaptation to temperate climate: relationship between population structure and polymorphism in the *Dwarf8* gene. *Genetics* 172:2449–2463
- Chen Y, Lubberstedt T (2010) Molecular basis of trait correlations. *Trends Plant Sci* 15:454–461
- Duan P, Rao Y, Zeng D, Yang Y, Xu R, Zhang B, Dong G, Qian Q, Li Y (2013) *SMALL GRAIN 1*, which encodes a mitogen-activated protein kinase kinase 4, influences grain size in rice. *Plant J* 77:547–557
- Ehrenreich IM, Hanzawa Y, Chou L, Roe JL, Kover PX, Purugganan MD (2009) Candidate gene association mapping of Arabidopsis flowering time. *Genetics* 183:325–335
- Fan C, Xing Y, Mao H, Lu T, Han B, Xu C, Li X, Zhang Q (2006) *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* 112:1164–1171
- Hu J, Wang Y, Fang Y, Zeng L, Xu J, Yu H, Shi Z, Pan J, Zhang D, Kang S, Zhu L, Dong G, Guo L, Zeng D, Zhang G, Xie L, Xiong G, Li J, Qian Q (2015) A rare allele of *GS2* enhances grain size and grain yield in rice. *Mol Plant* 8:1455–1465
- Huang X, Han B (2014) Natural variations and genome-wide association studies in crop plants. *Ann Rev Plant Biol* 65:531–551
- Huang X, Wei X, Sang T, Zhao Q, Feng Q, Zhao Y, Li C, Zhu C, Lu T, Zhang Z, Li M, Fan D, Guo Y, Wang A, Wang L, Deng L, Li W, Lu Y, Weng Q, Liu K, Huang T, Zhou T, Jing Y, Li W, Lin Z, Buckler ES, Qian Q, Zhang Q, Li J, Han B (2010) Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat Genet* 42:961–967
- Huang R, Jiang L, Zheng J, Wang T, Wang H, Huang Y, Hong Z (2013) Genetic bases of rice grain shape: so many genes, so little known. *Trends Plant Sci* 18:218–226
- Ishimaru K, Hirotsu N, Madoka Y, Murakami N, Hara N, Onodera H, Kashiwagi T, Ujiie K, Shimizu B, Onishi A, Miyagawa H, Katoh E (2013) Loss of function of the IAA-glucose hydrolase gene *TGW6* enhances rice grain weight and increases yield. *Nat Genet* 45:707–711
- Jia G, Huang X, Zhi H, Zhao Y, Zhao Q, Li W, Chai Y, Yang L, Liu K, Lu H, Zhu C, Lu Y, Zhou C, Fan D, Weng Q, Guo Y, Huang T, Zhang L, Lu T, Feng Q, Hao H, Liu H, Lu P, Zhang N, Li Y, Guo E, Wang S, Wang S, Liu J, Zhang W, Chen G, Zhang B, Li W, Wang Y, Li H, Zhao B, Li J, Diao X, Han B (2013) A haplotype map of genomic variations and genome-wide association studies of agronomic traits in foxtail millet (*Setaria italica*). *Nat Genet* 45:957–961
- Jun TH, Van K, Kim MY, Lee SH, Walker DR (2008) Association analysis using SSR markers to find QTL for seed protein content in soybean. *Euphytica* 162:179–191
- Kato T, Segami S, Toriyama M, Kono I, Ando T, Yano M, Kitano H, Miura K, Iwasaki Y (2011) Detection of QTLs for grain length from large grain rice (*Oryza sativa* L.). *Breed Sci* 61:269–274
- Khush GS (2005) What it will take to feed 5.0 billion rice consumers in 2030. *Plant Mol Biol* 59:1–6
- Li M, Xu L, Ren J, Cao G, Yu L, He H, Han L, Koh H (2010) Identification of quantitative trait loci for grain traits in japonica rice. *Agric Sci China* 9:929–936
- Li Y, Fan C, Xing Y, Jiang Y, Luo L, Sun L, Shao D, Xu C, Li X, Xiao J, He Y, Zhang Q (2011) Natural variation in *GS5* plays an important role in regulating grain size and yield in rice. *Nat Genet* 43:1266–1269
- Lipka AE, Tian F, Wang Q, Peiffer J, Li M, Bradbury PJ, Gore MA, Buckler ES, Zhang Z (2012) GAPIT: genome association and prediction integrated tool. *Bioinformatics* 28:2397–2399
- Liu L, Tong H, Xiao Y, Che R, Xu F, Hu B, Liang C, Chu J, Li J, Chu C (2015a) Activation of *Big Grain1* significantly improves grain size by regulating auxin transport in rice. *Proc Natl Acad Sci* 112:11102–11107
- Liu S, Hua L, Dong S, Chen H, Zhu X, Jiang J, Zhang F, Li Y, Fang X, Chen F (2015b) OsMAPK6, a mitogen-activated protein kinase, influences rice grain size and biomass production. *Plant J* 84:672–681
- Lu L, Shao D, Qiu X, Sun L, Yan W, Zhou X, Yang L, He Y, Yu S, Xing Y (2013) Natural variation and artificial selection in four genes determine grain shape in rice. *New Phytol* 200:1269–1280
- Lu Q, Zhang M, Niu X, Wang S, Xu Q, Feng Y, Wang C, Deng H, Yuan X, Yu H, Wang Y, Wei X (2015) Genetic variation and association mapping for 12 agronomic traits in *indica* rice. *BMC Genom* 16:1067
- Maccafferri M, Sanguineti MC, Corneti S, Ortega JLA, Salem MB, Bort J, DeAmbrogio E, Moral LFG, Demontis A, El-Ahmed A, Maalouf F, Machlab H, Martos V, Moragues M, Motawaj J, Nachit M, Nserallah N, Quabbou H, Royo C, Slama A, Tuberosa R (2008) Quantitative trait loci for grain yield and adaptation of durum wheat (*Triticum durum* Desf.) across a wide range of water availability. *Genetics* 178:489–511
- Mao H, Sun S, Yao J, Wang C, Yu S, Xu C, Li X, Zhang Q (2010) Linking differential domain functions of the *GS3* protein to natural variation of grain size in rice. *Proc Natl Acad Sci* 107:19579–19584
- Marathi B, Guleria S, Mohapatra T, Parsad R, Mariappan N, Kurungara VK, Atwal SS, Prabhu KV, Singh NK, Singh AK (2012) QTL analysis of novel genomic regions associated with yield and yield related traits in new plant type based recombinant inbred lines of rice (*Oryza sativa* L.). *BMC Plant Biol* 12:137
- Murray MG, Thompson WF (1980) Rapid isolation of high molecular weight plant DNA. *Nucl Acid Res* 8:4321–4326
- Nakagawa S (2004) A farewell to Bonferroni: the problems of low statistical power and publication bias. *Behav Ecol* 15:1044–1045
- Olsen KM, Halldorsdottir SS, Stinchcombe JR, Weinig C, Schmitt J, Purugganan MD (2004) Linkage disequilibrium mapping of Arabidopsis *CRY2* flowering time alleles. *Genetics* 167:1361–1369
- Palaisa KA, Morgante M, Williams M, Rafalski A (2003) Contrasting effects of selection on sequence diversity and linkage disequilibrium at two phytoene synthase loci. *Plant Cell* 15:1795–1806
- Qi P, Lin YS, Song XJ, Shen JB, Huang W, Shan JX, Zhu MZ, Jiang L, Gao JP, Lin HX (2012) The novel quantitative trait locus *GL3.1* controls rice grain size and yield by regulating *Cyclin-T1;3*. *Cell Res* 22:1666–1680
- Sakamoto T, Matsuoka M (2008) Identifying and exploiting grain yield genes in rice. *Curr Opin Plant Biol* 11:209–214
- Servin B, Martin OC, Mezard M, Hospital F (2004) Toward a theory of marker-assisted gene pyramiding. *Genetics* 168:513–523

- Shomura A, Izawa T, Ebana K, Ebitani T, Kagegae H, Konishi S, Yano M (2008) Deletion in a gene associated with grain size increased yields during rice domestication. *Nat Genet* 40:1023–1028
- Song XJ, Huang W, Shi M, Zhu MZ, Lin HX (2007) A QTL for rice grain width and weight encodes a previously unknown RING-type E3 ubiquitin ligase. *Nat Genet* 39:623–630
- Wang B, Lan T, Wu W (2003) Identification of QTLs underlying grain traits in rice using SSLP linkage map. *Fujian J Agric Sci* 18:11–15
- Wang S, Wu K, Yuan Q, Liu X, Liu Z, Lin X, Zeng R, Zhu H, Dong G, Qian Q, Zhang G, Fu X (2012) Control of grain size, shape and quality by *OsSPL16* in rice. *Nat Genet* 44:950–954
- Wang S, Li S, Liu Q, Wu K, Zhang J, Wang S, Wang Y, Chen X, Zhang Y, Gao C, Wang F, Huang H, Fu X (2015a) The *OsSPL16-GW7* regulatory module determines grain shape and simultaneously improves rice yield and grain quality. *Nat Genet* 47:949–954
- Wang Y, Xiong G, Hu J, Jiang L, Yu H, Xu J, Fang Y, Zeng L, Xu E, Xu J, Ye W, Meng X, Liu R, Chen H, Jing Y, Wang Y, Zhu X, Li J, Qian Q (2015b) Copy number variation at the *GL7* locus contributes to grain size diversity in rice. *Nat Genet* 47:944–948
- Wen W, Mei H, Feng F, Yu S, Huang Z, Wu J, Chen L, Xu X, Luo L (2009) Population structure and association mapping on chromosome 7 using a diverse panel of Chinese germplasm of rice (*Oryza sativa* L.). *Theor Appl Genet* 119:459–470
- Weng J, Gu S, Wan X, Gao H, Guo T, Su N, Lei C, Zhang X, Cheng Z, Guo X, Wang J, Jiang L, Zhai H, Wan J (2008) Isolation and initial characterization of *GW5*, a major QTL associated with rice grain width and weight. *Cell Res* 18:1199–1209
- Wilson LM, Whitt SR, Ibanez AM, Rocheford TR, Goodman MM, Buckler ES (2004) Dissection of maize kernel composition and starch production by candidate gene association. *Plant Cell* 16:2719–2733
- Xing Y, Zhang Q (2010) Genetic and molecular bases of rice yield. *Annu Rev Plant Biol* 61:421–442
- Xing Y, Tan Y, Hua J, Sun X, Xu C, Zhang Q (2002) Characterization of the main effects, epistatic effects and their environmental interactions of QTLs on the genetic basis of yield traits in rice. *Theor Appl Genet* 105:248–257
- Xu J, Xue Q, Luo L, Li Z (2002) Genetic dissection of grain weight and its related traits in rice (*Oryza sativa* L.). *Chin J Rice Sci* 16:6–10
- Ying JZ, Gao JP, Shan JX, Zhu MZ, Shi M, Lin HX (2012) Dissecting the genetic basis of extremely large grain shape in rice cultivar ‘JZ1560’. *J Genet Genom* 39:325–333
- Yoshida S, Ikegami M, Kuze J, Sawada K, Hashimoto Z, Ishii T, Nakamura C, Kamijima O (2002) QTL analysis for plant and grain characters of *sake*-brewing rice using a doubled haploid population. *Breed Sci* 52:309–317
- Yu SW, Yang CD, Fan YY, Zhuang JY, Li XM (2008) Genetic dissection of a thousand-grain weight quantitative trait locus on rice chromosome 1. *Chin Sci Bull* 53:2326–2332
- Zhang Z, Ersoz E, Lai CQ, Todhunter RJ, Tiwari HK, Gore MA, Bradbury PJ, Yu J, Arnett DK, Ordovas JM, Buckler ES (2010) Mixed linear model approach adapted for genome-wide association studies. *Nat Genet* 42:355–360
- Zhang X, Wang J, Huang J, Lan H, Wang C, Yin C, Wu Y, Tang H, Qian Q, Li J, Zhang H (2012) Rare allele of *OsPPKL1* associated with grain length causes extra-large grain and a significant yield increase in rice. *Proc Natl Acad Sci* 109:21534–21539
- Zhang H, Wang H, Qian Y, Xia J, Li Z, Shi Y, Zhu L, Ali J, Gao Y, Li Z (2013) Simultaneous improvement and genetic dissection of grain yield and its related traits in a backbone parent of hybrid rice (*Oryza sativa* L.) using selective introgression. *Mol Breed* 31:181–194
- Zhao K, Tung CW, Eizenga GC, Wright MH, Ali ML, Price AH, Norton GJ, Islam MR, Reynolds A, Mezey J, McClung AM, Bustamante CD, McCouch SR (2011) Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. *Nat Commun* 2:467
- Zhu C, Gore M, Buckler ES, Yu J (2008) Status and prospects of association mapping in plants. *Plant Genome* 1:5–20
- Zuo J, Li J (2014) Molecular genetic dissection of quantitative trait loci regulating rice grain size. *Ann Rev Genet* 48:99–118