

Genetic dissection on rice grain shape by the two-dimensional image analysis in one *japonica* × *indica* population consisting of recombinant inbred lines

Changbin Yin¹ · Huihui Li¹ · Shanshan Li¹ · Lidong Xu² · Zhigang Zhao² · Jiankang Wang¹

Received: 23 March 2015 / Accepted: 13 June 2015 / Published online: 2 July 2015
© The Author(s) 2015. This article is published with open access at Springerlink.com

Abstract

Key message This article used seven characters from the 2D image analysis to dissect the genetic architecture underlying rice grain shape in one *japonica* × *indica* population consisting of 215 recombinant inbred lines.

Abstract Two-dimensional (2D) digital image analysis is efficient for investigating the rice grain shape characters in large genetic and breeding populations. In this study, we used 2D image analysis to investigate seven characters, i.e., grain length (GL), grain width (GW), length-to-width ratio (LW), grain area (GA), grain circumference (GC), grain diameter (GD), and grain roundness (GR), in one *japonica* × *indica* genetic population consisting of 215 recombinant inbred lines. GL and GW can be recorded manually as well, and have been extensively used together with LW (i.e., GL/GW) in genetic studies on grain shape. GC and GA can be hardly measured manually, and have not been used together with GD and GR. Results indicated

that the seven characters could be precisely measured by 2D image analysis, genotype by environment interaction was low, and heritability was high. Each character was controlled by a few major stable genes and multiple minor additive genes. A total of 51 QTL were detected for the seven characters across four diverse environments, 22 from GL, GW, and LW, the three traditional characters, and 29 from the other four characters. The 51 QTL were clustered in eighteen marker intervals. Comparing with previous studies and analyzing the stability of identified QTL, we found six non-reported marker intervals, one each on chromosomes 2 and 3, and two each on chromosomes 6 and 8. The newly identified loci and the large-scale phenotyping system would greatly improve our knowledge about the genetic architecture and the future rice breeding on grain shape.

Introduction

Rice (*Oryza sativa* L.) is a seed-eating cereal crop, and therefore grain shape is a vital appearance quality trait. In addition, rice grain shape is also a key determinant on grain yield (Huang et al. 2013). In genetics, grain shape has been widely accepted as a complex trait controlled by multiple genes with small effects. By phenotyping, it is complex because it could be evaluated in different ways. In conventional rice genetic study and breeding, grain shape is characterized by grain length (GL), grain width (GW), and the length-to-width ratio (LW), which greatly affect grain yield, grain appearance quality, and marketability (Wan et al. 2005; Fan et al. 2006). As a result, lots of QTL have been detected in the past two decades for the three characters relevant to grain shape (Xing et al. 2000; Tan et al. 2000; Li et al. 2003, 2004; Aluko et al. 2004;

Communicated by T. Sasaki.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-015-2560-7) contains supplementary material, which is available to authorized users.

✉ Jiankang Wang
jkwang@cgiar.org; wangjiankang@caas.cn

¹ The National Key Facility for Crop Gene Resources and Genetic Improvement, Institute of Crop Science and CIMMYT China Office, Chinese Academy of Agricultural Sciences, Beijing 100081, China

² National Key Laboratory for Crop Genetics and Germplasm Enhancement, Research Center of Jiangsu Plant Gene Engineering, Nanjing Agricultural University, Nanjing 210095, China

Wan et al. 2005; Bai et al. 2010; Shao et al. 2010; Huang et al. 2013).

Based on QTL identified with relatively large phenotypic effects, some genes related to grain shape have been isolated and cloned in rice, such as *GW2* (Song et al. 2007), *gw5/qSW5* (Shomura et al. 2008; Weng et al. 2008; Wan et al. 2008), *GIF1/OsCIN2* (Wang et al. 2008, 2010), *srs-3* (Tanabe et al. 2007; Kitagawa et al. 2010), *GS5* (Yu et al. 1997; Li et al. 2000, 2011), *GS3* (Fan et al. 2006; Takano-Kai et al. 2009, 2011; Mao et al. 2010; Wang et al. 2011), *qGW8/OsSPL16* (Wang et al. 2012a, b), *SG1* (Nakagawa et al. 2012), and *DEP1/qPE9-1* (Yan et al. 2007; Zhou et al. 2009; Huang et al. 2009; Yi et al. 2011; Taguchi-Shiobara et al. 2011; Sun et al. 2014). Positive regulator *PGL1* and *PGL2* were found in the network of grain shape genes (Heang and Sassa 2012a, b), and the relationship of four grain shape genes *GS3*, *GW2*, *gw5/qSW5*, and *GIF1* has been studied (Yan et al. 2011). However, the gene-to-trait pathway and relationship between genes identified by different characters are largely unknown (Huang et al. 2013).

Conventionally, the evaluation of grain shape is performed manually. In case of heavy workloads, long operating time, and short of experience, the manual measurement is less efficient and error-prone. Any mistakes in data collection may lead to incorrect and misleading results in genetic study and rice breeding. For this reason, more experienced workers and supervisors are needed to frequently check and verify the data to avoid the mistake. But subjective errors are still inevitable especially when the workers are at fatigue. Automated assessment of plant phenotypes is ideal and essential in the situation of large genetic and breeding populations. Recently, the two-dimensional (2D) digital image analysis has become available for the high-throughput phenotyping on traits like shoot biomass, yield and yield components, and grain shape (Yang et al. 2013). Taking grain shape for an example, the system can investigate more characters than GL, GW, and LW in much shorter time. However, the relationship between the conventional and novel 2D characters on grain shape is not clear. Genetic study on 2D grain shape characters is still lacking.

Our objectives in this study were (1) to investigate three conventional and four novel characters on grain shape in one *japonica* × *indica* genetic population by the 2D image analysis; (2) to study the relationship between the seven characters on grain shape; (3) to identify the genetic architecture and common stable QTL on grain shape measured by the seven characters; and (4) to discover novel QTL on grain shape which may contribute to future rice quality breeding.

Materials and methods

Population development and genotyping

Two parents of the genetic population used in this study are *Oryza sativa* ssp. *japonica* cv. Asominori and *Oryza sativa* ssp. *indica* cv. IR24. They were planted and the cross was firstly made in the 2007 summer season in the experimental field of Nanjing Agricultural University, Jiangsu Province, China. Their F₁ hybrids were planted in the 2007 winter season in Sanya, Hainan Province, China, and more than 500 F₂ seeds were harvested at maturity. The following generations were alternatively planted in the previous two locations till F₁₂, when no visual segregation was observed within each line in the field. Single seed descent was applied for generation advance during the repeated selfing process. At the end, a total of 215 recombinant inbred lines (RILs) were retained, and each RIL can be traced back to an individual plant in the F₂ generation.

DNA of each F₁₂ RIL was isolated and extracted for genotyping. DNA extraction was carried out according to the procedure described by Dellaporta et al. (1983). The polymerase chain reaction (PCR) was performed using the procedure of Chen et al. (1997), with minor modifications. The protocol of PCR was briefly described as follows. The template DNA was subjected to denaturation at 94 °C for 5 min, followed by 32 cycles of PCR amplification (denaturation at 94 °C for 1 min, primer annealing at 48–55 °C for 30 s, and primer extension at 72 °C for 1 min) and a final extension at 72 °C for 5 min. The PCR products were separated through electrophoresis on an 8 % non-denaturing polyacrylamide gel and detected using the silver staining method of Sanguinetti et al. (1994). A total of 933 pairs published SSR markers (McCouch et al. 2002) were firstly screened for Asominori and IR24, and 313 markers (33.55 %) showed polymorphism between the two parents. Referring to the rice consensus map (McCouch et al. 2002), we selected 143 evenly distributed markers to screen the 215 RILs.

Field experiments and trait measurement

The 215 RILs and their parents were grown from May to November, 2013 in four geologically and ecologically diverse locations in China, i.e., Guilin (24.18°N, 109.45°E), Guiyang (26.35°N, 106.42°E), Nanchang (28.38°N, 116.24°E), and Nanjing (31.95°N, 119.16°E). The four locations have rice as the major cultivated and consumed crop. A randomized complete block design was applied with two replications at each location. Each entry plot consisted of four rows, and each row was cultivated

with ten individual plants. Field managements during the growing season were similar to those adopted by local farmers. Three representative individual plants in the central part of each plot were used to measure various grain traits. Grains of the selected plants were harvested and air-dried, and then stored at the room temperature for at least 3 months before trait measurement or investigation.

Seven grain shape characters were evaluated by SC-G rice grain appearance quality image analysis system developed by Hangzhou WSeen Detection Technology Co., Ltd, China. Firstly, all filled grains from each selected rice plant were divided into a number of samples, each having more or less 1000 grains (about ~25 g). Each sample was spread as even as possible on one 21.00 cm × 30.00 cm flat-bed surface to be photographed. The grain image was acquired by an Eloam high-speed photographic apparatus S500A3B with a resolution of 4800 × 2400 × 24 bits. The image analysis system is able to distinguish individual grains when they are randomly spread on the flat-bed surface. It can distinguish individuals even when some grains are overlapped (Zhong et al. 2009).

Grain length (GL), grain width (GW), grain circumference (GC), and grain area (GA) were firstly deduced from pixel number on the projected image. GL is defined as the maximum Euclidean distance between two boundary points of the grain, and GW is defined as the maximum length of straight lines perpendicular to the line of GL. Grain length-to-width ratio (LW) is equal to the ratio of GL and GW. Grain diameter (GD) is equal to the diameter of a circle that has an area equal to GA, i.e., $GD = \sqrt{4 \times GA/\pi}$. Fitting a rice grain image as an ellipse that has the same area, and uniform distribution of points bounded by the perimeter of the profile of grain image, the grain roundness is calculated by $GR = \frac{4 \times GA}{\pi \times (\text{major axis})^2}$, where major axis is largest distance between antipodal points on the fitted ellipse.

Phenotypic data analysis

ANOVA was used to test the statistical significance of various sources of variation. In the combined ANOVA across the four locations, phenotype was partitioned into overall mean, replication (i.e., block) effect per environment, genotypic effect, environment effect, genotype by environment (GE) effect, and random error effect. Let y_{ijk} be the observed value of a trait in interest for the i th RIL in the k th replication in the j th environment (equivalent to location in this study). The linear model used in ANOVA is therefore,

$$y_{ijk} = \mu + R_{k/j} + G_i + E_j + GE_{ij} + \varepsilon_{ijk}, \text{ and } \varepsilon_{ijk} \sim N(0, \sigma_\varepsilon^2) \quad (1)$$

where $i = 1, 2, \dots, n$ ($n = 215$ in this study), $j = 1, 2, \dots, e$ ($e = 4$ in this study), $k = 1, 2, \dots, r$ ($r = 2$ in this study), μ is overall mean of the RIL population, $R_{k/j}$ is the k th

replication effect in the j th environment, G_i is genotypic effect of the i th RIL, E_j is environmental effect of the j th environment, GE_{ij} is interaction effect between the i th RIL and j th environment, and ε_{ijk} is random error effect which was assumed to be normally distributed with a mean of zero. Once the linear model of ANOVA is defined, total degree of freedom and total sum square can be partitioned into the components defined in the linear model, from which mean square (MS) of each source of variation can be calculated, and the significance test can be conducted.

Heritability is a useful genetic parameter. In the broad sense, heritability is the proportion of genetic variance compared with phenotypic variance. From the theoretical expectation of MS, genetic variance (σ_G^2), interaction variance (σ_{GE}^2), and error variance (σ_ε^2) can be estimated by the following equations, where $e = 4$ and $r = 2$ in this study.

$$\sigma_G^2 = \frac{1}{e \times r}(\text{MS}_G - \text{MS}_\varepsilon),$$

$$\sigma_{GE}^2 = \frac{1}{r}(\text{MS}_{GE} - \text{MS}_\varepsilon), \text{ and } \sigma_\varepsilon^2 = \text{MS}_\varepsilon \quad (2)$$

It is generally agreed that environmental variance should not be included in the calculation of heritability (Holland et al. 2003). Phenotypic variance per plot in multi-environmental trials can be written as $\sigma_P^2 = \sigma_G^2 + \sigma_{GE}^2 + \sigma_\varepsilon^2$. Therefore, the phenotypic variance on the mean performance across replications and environments can be written as $\sigma_P^2 = \sigma_G^2 + \frac{1}{e}\sigma_{GE}^2 + \frac{1}{er}\sigma_\varepsilon^2$. Heritability on the plot level and heritability on the mean performance of each RIL in the genetic population can be estimated from the following two equations, respectively.

$$H_P^2 = \frac{\sigma_G^2}{\sigma_P^2} = \frac{\sigma_G^2}{\sigma_G^2 + \sigma_{GE}^2 + \sigma_\varepsilon^2}, \text{ and}$$

$$H_P^2 = \frac{\sigma_G^2}{\sigma_P^2} = \frac{\sigma_G^2}{\sigma_G^2 + \frac{1}{e}\sigma_{GE}^2 + \frac{1}{er}\sigma_\varepsilon^2} \quad (3)$$

Genotypic variance is the same in calculating the two levels of heritability, but phenotypic variance is reduced in the mean performance across environments and replications. Obviously, heritability has a higher value on phenotypic mean. The method previously described has been implemented in tool “ANOVA” in the QTL IciMapping software (Meng et al. 2015).

Genotypic data analysis

Genetic linkage map construction and QTL mapping were conducted in the QTL IciMapping software (Meng et al. 2015), which is public and freely available (<http://www.isbreeding.net/software/>). For map construction, physical positions of SSR markers were used in grouping as anchor

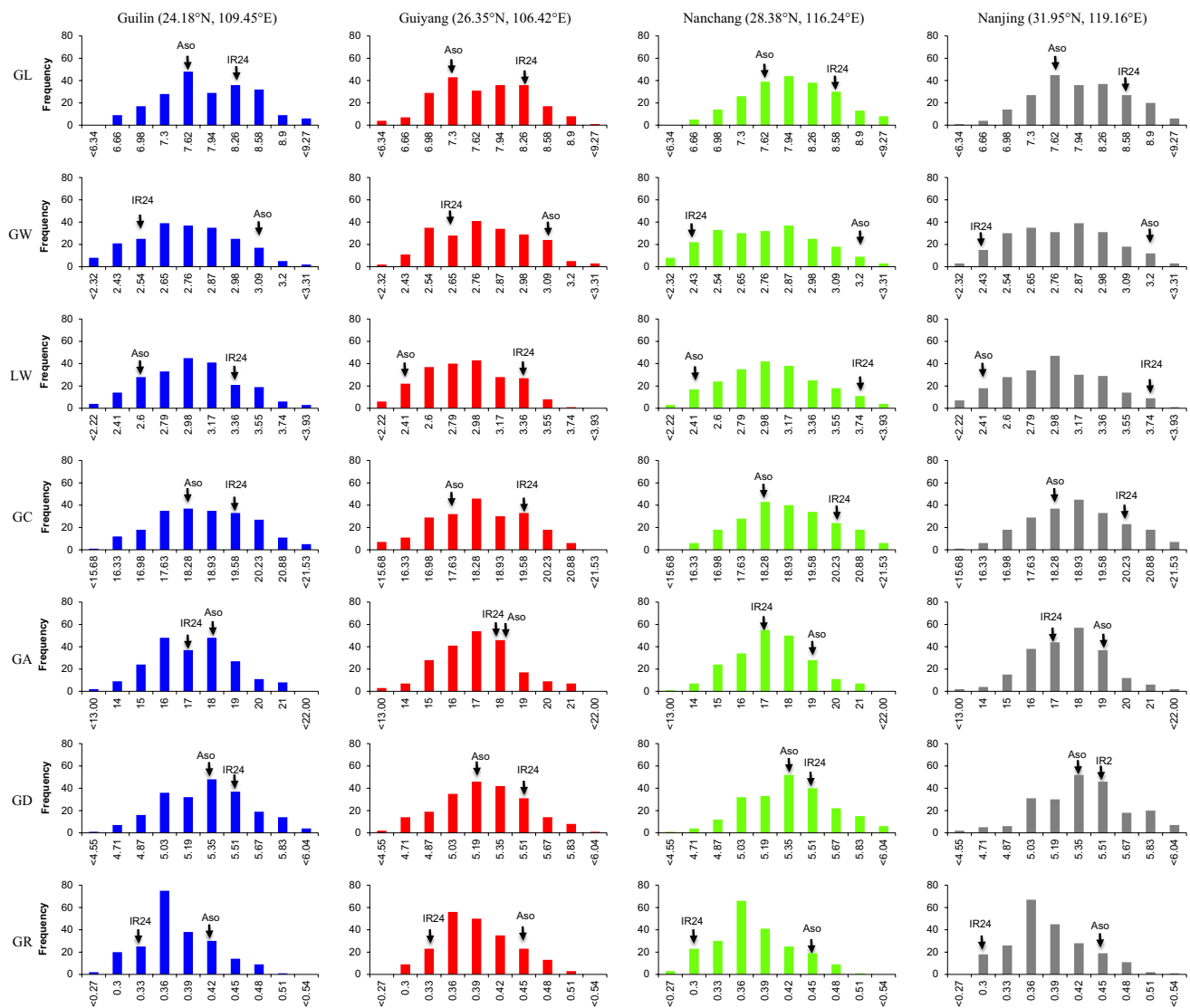


Fig. 1 Frequency distribution of the seven characters on rice grain shape in the RIL population grown in four environments. Aso and IR24 at the top of each histogram represented the two parents *Oryza sativa* ssp. *japonica* cv. Asominori and *Oryza sativa* ssp. *indica* cv.

information. Algorithm nnTwoOpt was used to acquire the preliminary order and positions of linked markers, where the nearest neighbor was used to construct an initial order and the two-opt algorithm in solving traveling salesman problems was used to improve the marker order. Rippling algorithm at a window size of 8 markers was used to fine-tune the linkage map with the objective to minimize the sum of adjacent recombination frequencies on each linkage map.

Inclusive Composite Interval Mapping, known as ICIM (Li et al. 2007; Wang 2009), was used for QTL identification. LOD threshold was obtained on a total of 7000 permutation tests for the seven characters and

IR24, respectively. *GL* grain length, *GW* grain width, *LW* grain length-to-width ratio, *GC* grain circumference, *GA* grain area, *GD* grain diameter, and *GR* grain roundness

a genome-wide type I error rate at 0.05. Probabilities for entering and leaving variables were set at 0.001 and 0.002, respectively, in the stepwise regression aiming to determine the linear relationship between phenotype and marker type. This linear model was then used for background genetic variation control in ICIM QTL mapping. The scanning step was set at 1 cM across the 12 rice chromosomes. A peak in a marker interval along the LOD profile was treated as a QTL, if there is at least one environment with the peak value higher than the LOD threshold. The identified QTL was named according to McCouch et al. (1997).

Table 1 Correlation coefficient between the seven characters on rice grain shape across the four locations

Character	GL	GW	LW	GC	GA	GD	GR
GL	1.000						
GW	−0.238**	1.000					
LW	0.777**	−0.790**	1.000				
GC	0.985**	−0.078*	0.666**	1.000			
GA	0.687**	0.529**	0.083*	0.793**	1.000		
GD	0.931**	0.124**	0.501**	0.974**	0.894**	1.000	
GR	−0.771**	0.793**	−0.986**	−0.657**	−0.087*	−0.494**	1.000

GL grain length, GW grain width, LW grain length-to-width ratio, GC grain circumference, GA grain area, GD grain diameter, and GR grain roundness

*, ** Significance at the level of 0.05 and 0.01, respectively

Results

Phenotypic distribution, correlation, and ANOVA

Phenotypic frequency distributions of the seven characters in the RIL population and four environments are shown in Fig. 1. Difference between the two parents varies by environment. But, IR24 has consistently greater values on GL, LW, GC, and GD, and Asominori has consistently greater values on GW, GA, and GR in all environments, indicating that the investigated characters may have high stability across environments. In the RIL population, transgressive segregation at both directions can be observed, but the transgressive level is different from the seven characters (Fig. 1). Higher level of transgressive segregation can be found from GL, GA, and GD.

For the seven characters, similar correlation coefficients were observed in the four environments (Table S1). Coefficients calculated from the phenotypic mean of each RIL across the four environments and two replications are shown in Table 1. Correlation coefficients are close to or >0.9 between GL and GC, GL and GD, GC and GD, and GA and GD (Table 1). This is understandable when thinking longer rice grain is always larger in grain size. GL, GC, GD, and GA reflect size of the rice grain in different ways, and one cannot be replaced by the other one. LW and GR have a correlation coefficient close to −1, indicating they measured grain shape in two opposite ways. When the 2D image of rice grain can be approximated by an ellipse, LW is actually ratio of long axis over short axis, and GR is ratio of short axis over long axis. This explains the highly negative correlation between LW and GR.

For each character, ANOVA combining the four environments showed that there were significant variations from the four environments, the two replications (or two blocks) per environment, the 215 genotypes, and the genotype by environment (GE) interactions (Table S2).

Significance from the two replications in the four environments indicated that the block effect should be considered in the ANOVA linear model in Eq. (1) in order to reduce the random error, which actually represents one of the three basic principles in field experimental design. Four components of variance calculated by Eq. (2) and heritability in the broad sense calculated by Eq. (3) are shown in Table 2. Obviously, environment and GE interaction had much lower variances, compared with genotype, indicating that genotypic variation was the major part in the observed phenotypic variation for the seven characters. By plot, the characters had the heritability around 0.9 (Table 2). Much less GE interactions and random errors were included in the phenotypic mean across environments and replications. Therefore, the heritability was increased when estimated by the phenotypic mean (Table 2). High heritability was also found in other studies, for example see Huang et al. 2013.

Parental contribution, marker distortion, and linkage maps

Parental contribution is the proportion of the genome contributed by a parent to its progeny (Wang and Bernardo 2000). In RIL populations, each line is homozygous, and genotypic frequency is equivalent to gene (or allele) frequency at each locus. Therefore, for each of the 215 RILs, parental contribution from Asominori can be calculated by the proportion of Asominori marker type to the total marker number. Among the 215 RILs, it can be seen from Fig. 2 that Asominori had a contribution between 35 and 65 % to 156 (or 72.56 %) lines. There were 9 (or 4.19 %) RILs where Asominori had a contribution below 25 %, and 9 (or 4.19 %) RILs where Asominori had a contribution above 85 %. As single seed descent was strictly applied during the repeated selfing in developing the RIL population, the great variation on parental contribution observed in Fig. 2 may come from the random genetic drift,

Table 2 Variance components and heritability for the seven characters on rice grain shape estimated in the RIL population

Character	Variance components				Heritability	
	Environment	Genotype	G by E interaction	Random error	Plot level	Genotypic mean level
GL	0.0138	0.3501	0.0139	0.0111	0.9334	0.9944
GW	0.0006	0.0452	0.0021	0.0026	0.9073	0.9918
LW	0.0038	0.1255	0.0039	0.0040	0.9403	0.9989
GC	0.0623	1.4858	0.0648	0.0542	0.9258	0.9924
GA	0.0423	2.4288	0.1637	0.1579	0.8830	0.9838
GD	0.0035	0.0824	0.0049	0.0046	0.8971	0.9885
GR	0.0001	0.0022	0.0001	0.0001	0.9448	0.9954

GL grain length, GW grain width, LW grain length-to-width ratio, GC grain circumference, GA grain area, GD grain diameter, and GR grain roundness

which can be hardly controlled in breeding and genetic populations.

Regarding the 143 SSR markers, two alleles at each locus should be fitted by the 1:1 Mendelian ratio if there is no segregation distortion in the RIL population, i.e., each allele has the expected frequency of 0.5. Observed frequency of the allele from Asominori and the segregation distortion test are shown in Fig. 3. Allele frequencies at the 143 marker loci ranged from 0.2372 to 0.6558 (upper Fig. 3), and the average was 0.4722. When the significance level of 0.001 was applied, a total of 31 markers were identified to have segregation distortion, 4 markers on chromosome 1, 6 on chromosome 3, 2 on chromosome 4, 8 on chromosome 6, 2 on chromosome 9, 9 on chromosome 11, and 2 on chromosome 12 (lower Fig. 3).

Linkage maps of the 143 SSR markers constructed from the RIL population had a total length of 1474.31 cM (Fig. 4). Map length of two adjacent markers ranged from 0.24 to 32.09 cM, with an average of 10.31 cM. There were 83 marker intervals (or 58.04 %) shorter than 10 cM, and

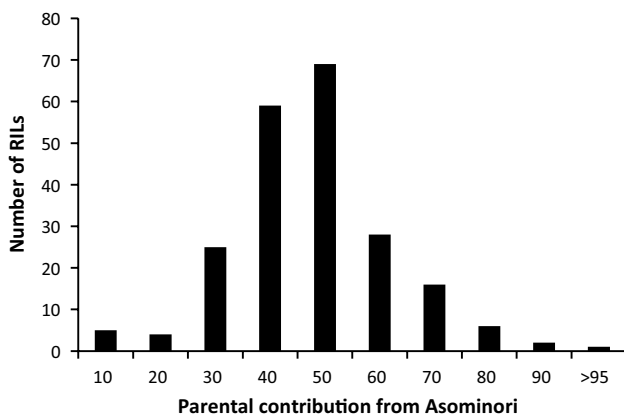


Fig. 2 Frequency distribution of the genetic contribution from parent *Oryza sativa* spp. *japonica* cv. Asominori to each line in the RIL population

6 intervals (or 4.20 %) longer than 25 cM. Larger gaps were observed on chromosomes 1 to 5 and 11, and markers were relatively evenly distributed on other chromosomes (Fig. 4).

QTL identified for the seven characters

A LOD threshold of 2.65 estimated from 7000 times of permutation test was used in QTL mapping for the seven characters. Seven QTL were detected for GL on four chromosomes (Table 3). qGL2-1 and qGL3-1 were detected with positive additive effects in four environments (Table 3), indicating the allele from IR24 increased GL. qGL2-1 had the LOD score from 2.78 to 6.31, and explained 4.46–9.65 % of the variation on phenotypic mean in each environment. qGL3-1 had the highest LOD scores and explained about 50 % of the variation. Three QTL were detected in three environments, among which qGL2-2 and qGL3-2 had negative effects and qGL8-2 had positive effects (Table 3). QTL not significant in all environments also showed peaks in non-significant environments, and the additive effects were at the same direction as those in the significant environments.

Seven QTL were detected for GW (Table 4). qGW3-1 and qGW5 were detected with negative additive effects in four environments (Table 4). qGL3-1 had the LOD score from 3.03 to 4.38, explaining 5.19–7.07 % of the variation on phenotypic mean in each environment. qGL5 had the highest LOD scores and explained about 20 % of the variation.

Eight QTL were detected for LW (Table 5). qLW3 and qLW5 were detected with positive additive effects in four environments (Table 5). qLW3 had the highest LOD score and explained about 30 % of the variation in each environment. qLW5 had the LOD score from 15.40 to 17.33, and explained 13.88–31.71 % of the variation. qLW2-2 was detected in three environments with positive additive effects (Table 5).

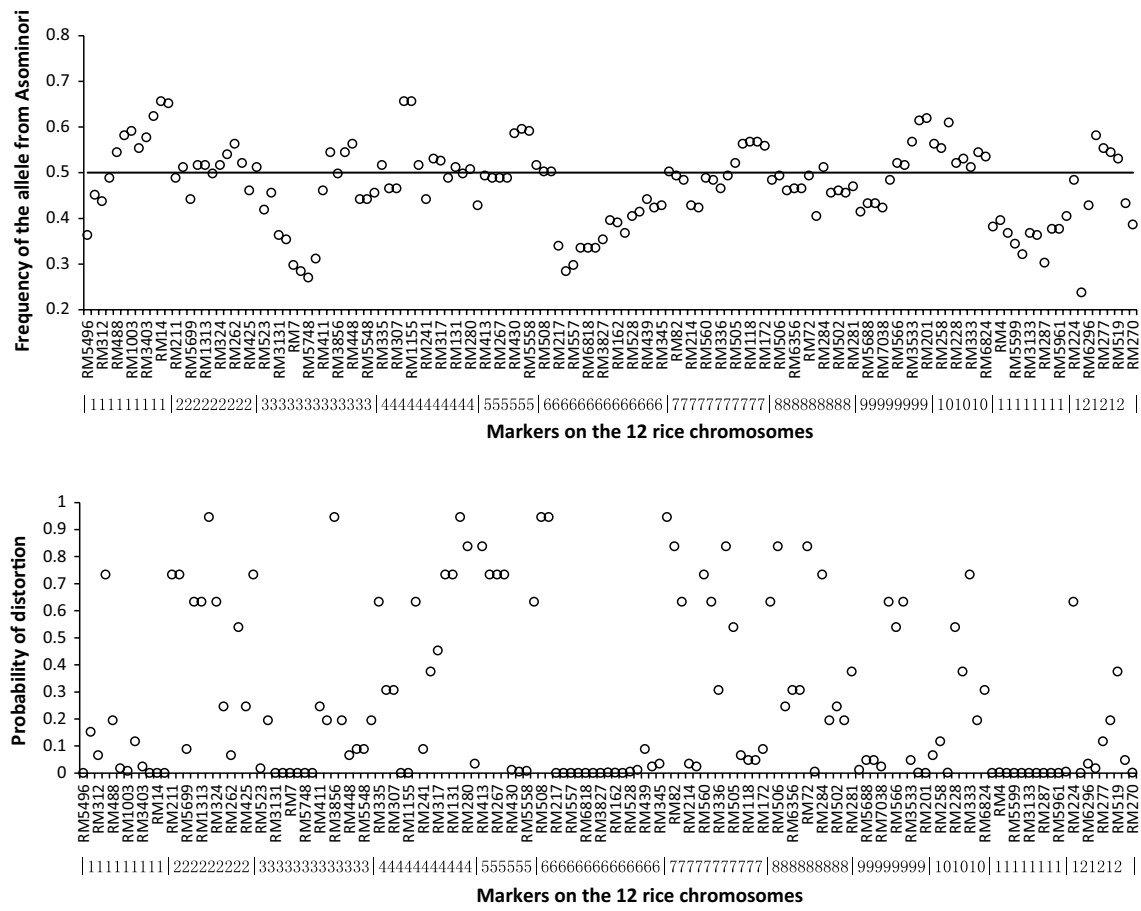


Fig. 3 Frequency of the allele from *Oryza sativa* spp. *japonica* cv. Asominori (*upper*) and the segregation distortion test (*lower*) in the RIL population at each marker locus

Six QTL were detected for GC (Table 6). qGC3-1 and qGC8 were detected with positive additive effects in four environments (Table 6). qGC3-1 had the highest LOD score and explained about 50 % of the variation in each environment. qGC8 had the LOD score from 3.87 to 6.80, and explained 4.06–7.10 % of the variation.

Ten QTL were detected for GA (Table 7). qGA3-1 and qGA5 were both detected in four environments, but one had positive additive effect and the other had negative effect (Table 7). qGA3-1 had the LOD score from 8.59 to 17.31, and explained 15.21–30.27 % of the variation in each environment. qGA5 had the LOD score from 5.20 to 11.29, and explained 8.58–12.44 % of the variation. Three QTL were detected in three environments, among which qGA2-2 and qGA3-2 had negative additive effects, and qGL7 had positive additive effect (Table 7).

Six QTL were detected for GD (Table 8). Three QTL on chromosomes 2, 3, and 8 were detected in four environments, two of which had positive additive effects and one of which had negative effects in four environments (Table 8).

qGD2 had the LOD score from 4.48 to 9.34, and explained 6.04–12.74 % of the variation on phenotypic mean. qGD3-1, had the highest LOD score and explained about 50 % of the variation in each environment. qGD8 had the LOD score from 2.92 to 6.98, and explained 5.50–8.33 % of the variation.

Seven QTL were detected for GR (Table 9). qGR2-1, qGR3, and qGR5 were detected with negative effects in four environments (Table 9). qGR3 had the highest LOD scores from 18.93 to 25.25, and explained 28.58–34.00 % of the variation on phenotypic mean in the four environments. qGR5 had the LOD score from 12.23 to 18.87, and explained 14.88–17.67 % of the variation. qGR2-2 was detected in three environments with negative effect, and the other three, i.e., qGR1, qGR4, and qGR8, were detected in one environment (Table 9). Due to the highly negative correlation, QTL from GR and LW had effects at the opposite directions, but they were almost identical by position, LOD score, and PVE (Tables 5 and 9).

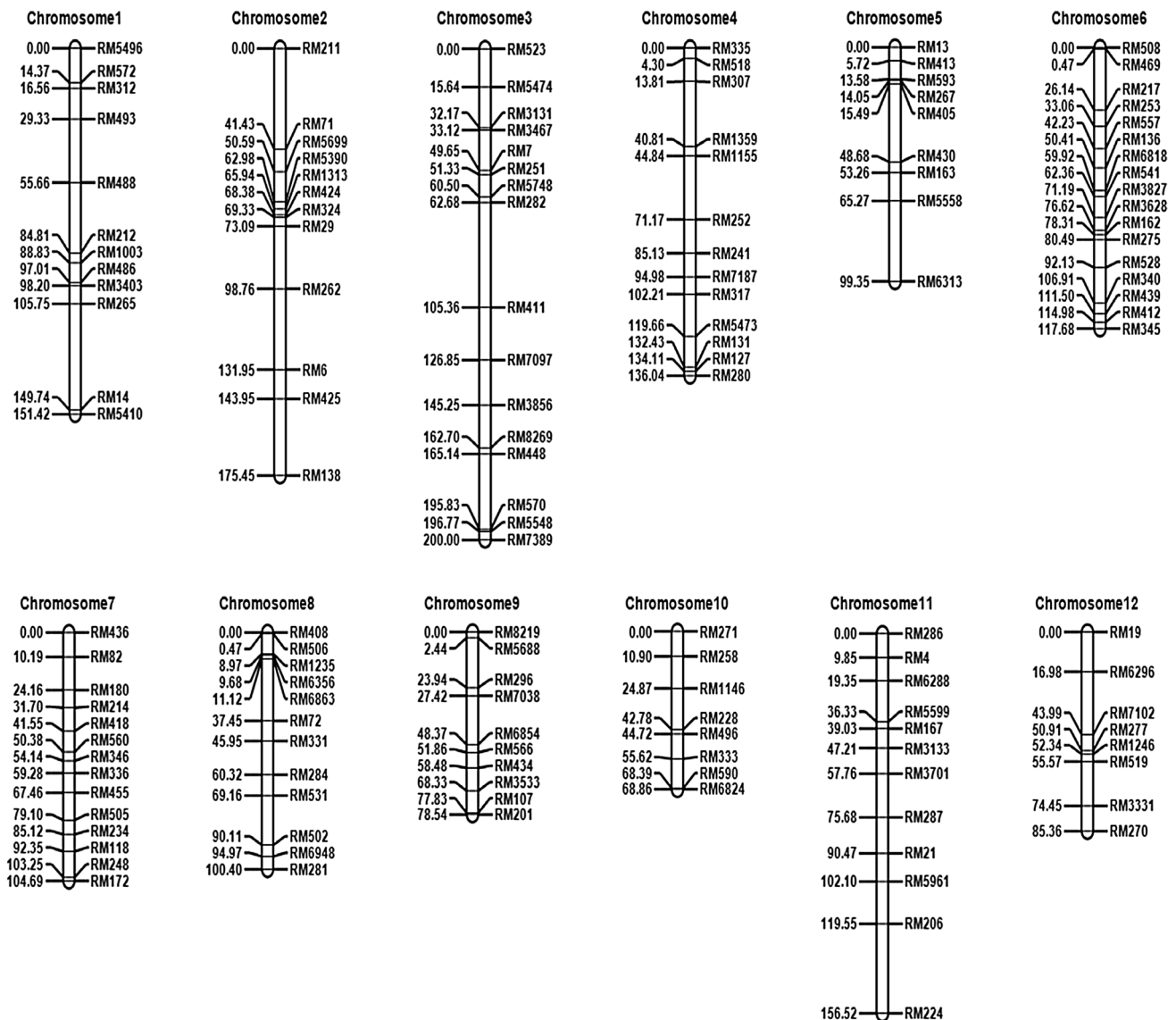


Fig. 4 The genetic linkage map of 143 SSR markers constructed in the RIL population derived from *Oryza sativa* ssp. *japonica* cv. Asominori and *Oryza sativa* ssp. *indica* cv. IR24

Potentially novel loci harboring QTL on grain shape

Combining mapping results of the seven characters (Tables 3, 4, 5, 6, 7, 8, and 9), we identified a total of 51 QTL which may affect the grain shape in various ways. Those loci were indicated by arrows on the LOD profiles shown in Fig. 5. However, due to high correlations between the seven characters (Table 1), it is highly likely that some loci have pleiotropic effects on several characters, and therefore the identified loci may not represent 51 totally different chromosomal locations. For examples, at similar position of qGL2-2, QTL was also detected for GC, GA, and GD, i.e., qGC2-2, qGA2-2, and qGD2 (Fig. 5). At similar position of qGL3-1, QTL was also detected

for the other six characters, i.e., qGW3-1, qLW3, qGC3-1, qGA3-1, qGD3-1 and qGR3 (Fig. 5). At similar position of qGW5, QTL was also detected for LW, GA, GD, and GR, i.e., qLW5, qGA5, qGD5, and qGR5 (Fig. 5). At similar position of qGL8-2, QTL was also detected for GC, GA, and GD, i.e., qGC8, qGA8, and qGD8 (Fig. 5). Each of the four chromosomal locations likely has pleiotropic effects on multiple characters, rather than closely linked loci affecting individual characters.

From Tables 3, 4, 5, 6, 7, 8, and 9 and the joint QTL mapping of four environments (results not shown), the 51 QTL were roughly clustered into 18 marker intervals on the first ten rice chromosomes (Table 10). Genes/QTL have been previously reported on 12 intervals but have not been

Table 3 QTL of grain shape identified by grain length (GL) in the four environments

QTL ^a	Chr.	Left marker	Right marker	Guijin (24.18°N, 109.45°E)				Guiyang (26.35°N, 106.42°E)				Nanchang (28.38°N, 116.24°E)				Nanjing (31.95°N, 119.16°E)			
				Pos. ^b	LOD ^c	PVE (%) ^d	Add ^e	Pos. ^b	LOD ^c	PVE (%) ^d	Add ^e	Pos. ^b	LOD ^c	PVE (%) ^d	Add ^e	Pos. ^b	LOD ^c	PVE (%) ^d	Add ^e
qGL2-1	2	RM29	RM262	87	2.78	4.67	0.13	83	6.31	9.25	0.18	90	5.20	7.58	0.17	89	4.40	7.89	0.17
qGL2-2	2	RM6	RM425	140	5.84	7.50	-0.17	139	5.65	5.99	-0.15	140	5.06	5.95	-0.15	140	2.11	2.62	-0.10
qGL3-1	3	RM411	RM7097	112	27.99	49.97	0.43	112	30.05	48.69	0.41	110	32.97	54.43	0.45	111	28.15	50.10	0.44
qGL3-2	3	RM8269	RM448	163	4.63	4.86	-0.14	163	0.81	0.70	-0.05	163	4.88	4.47	-0.13	163	4.47	4.68	-0.13
qGL7	7	RM505	RM234	80	3.16	3.37	0.11	83	4.54	4.29	0.12	83	1.37	1.28	0.07	81	1.67	1.81	0.08
qGL8-1	8	RM6863	RM72	21	1.77	3.17	0.11	21	3.51	5.81	0.14	12	0.81	0.74	0.05	17	0.73	1.10	0.07
qGL8-2	8	RM502	RM6948	92	5.47	6.12	0.15	91	1.17	1.03	0.06	92	5.97	5.99	0.15	93	3.29	3.70	0.12

^a A peak in a marker interval along the LOD profile was treated as a QTL, if there is at least one environment with the peak value higher than the LOD threshold

^b Chromosomal position (cM) of the peak

^c The number in bold was the value of a peak higher than the threshold value in the LOD profile

^d Percentage of the phenotypic variation explained by the locus at the peak position on the LOD profile

^e Additive effect of the identified QTL. Positive additive effect indicated the allele from IR24 increased grain length, and the allele from Aominori decreased grain length. Negative additive effect indicated the allele from IR24 decreased grain length, and the allele from Aominori increased grain length

reported on the other six (see the last column in Table 10). Detailed information on the reported genes and QTL collocated with QTL identified in this study are given in Table S3. QTL in interval RM411–RM7097 on chromosome 3 was detected by all characters in each environment, and has been fined-mapped by GL, GW, and LW, and cloned as *GS3* (Table 10). QTL in interval RM267–RM405 on chromosome 5 was detected by four characters in three environments and five characters in one environment. It has been fined-mapped by GW and LW, and cloned as *gw5/qSW5*, and *srs-3* (Table 10). QTL in interval RM502–RM6948 on chromosome 8 was detected by three characters in three environments and four characters in one environment. It has been fined-mapped by GW, and cloned as *qGW8/OsSPL16* (Table 10).

In previous studies (Table 10), QTL on chromosome 1 has been fined-mapped by GL. The QTL in interval RM211–RM71 on chromosome 2 has been fined-mapped by GW and GL, and cloned as *GW2*. QTL in interval RM5473–RM131 on chromosome 4 has been fined-mapped by GL, QTL in interval RM136–RM6818 on chromosome 6 has been fined-mapped by GW, and QTL on chromosome 10 has been fined-mapped by GL. However, the five loci mentioned above were only detected in one environment by few characters in this study (Table 10). Considering also large size of the mapping population, high phenotyping precision of the seven characters (seen from heritability in Table 2), and high detection powers of the mapping method (Li et al. 2007, 2010; Wang 2009), we assume the six non-reported intervals may harbor novel loci on the grain shape-associated traits, and therefore are worth of further investigations. Two novel loci were located on chromosomes 2 and 3, and two each on chromosomes 6 and 8 (Table 10).

Discussions

Geometrical relationship of the seven characters on rice grain shape

If the 2D image of a rice grain could be fitted by an ellipse, GL is length of the long axis, GW is length of the short axis, and $LW = GL/GW$. In geometry, GC represents length of the circumference which can be approximated by $\pi [1.5(GL + GW) - \sqrt{GL \times GW}]$, GA represents area of the ellipse which is equal to $\pi \times GL \times GW$, $GD = \sqrt{GL \times GW}$ (i.e., the geometrical mean of GL and GW), and $GR = GW/GL$ (i.e., the reciprocal of LW). Therefore, it is not strange to see the correlation values shown in Table 1. For example, greater GL will result in greater LW, GC, GA, and GD, but smaller GW and GR. Therefore, GL was positively correlated with LW, GC, GA,

Table 4 QTL of grain shape identified by grain width (GW) in the four environments

QTL ^a	Chr.	Left marker	Right marker	Guilin (24.18°N, 109.45°E)			Guiyang (26.35°N, 106.42°E)			Nanchang (28.38°N, 116.24°E)			Nanjing (31.95°N, 119.16°E)						
				Pos. ^b	LOD ^c	PVE (%) ^d	Add ^e	Pos. ^b	LOD ^c	PVE (%) ^d	Add ^e	Pos. ^b	LOD ^c	PVE (%) ^d	Add ^e	Pos. ^b	LOD ^c	PVE (%) ^d	Add ^e
qGW2-1	2	RM211	RM71	10	2.39	6.50	-0.06	23	1.47	5.38	-0.05	5	1.66	2.57	-0.04	11	4.97	13.56	-0.08
qGW2-2	2	RM1313	RM424	66	3.61	4.66	-0.05	66	0.91	1.11	-0.02	68	0.01	0.01	0.00	66	0.01	0.01	0.00
qGW2-3	2	RM29	RM262	98	1.21	1.58	-0.03	88	3.95	9.62	-0.07	70	4.92	5.58	-0.05	70	5.45	6.10	-0.05
qGW3-1	3	RM411	RM7097	111	3.66	7.07	-0.06	109	3.03	5.19	-0.05	109	4.38	6.79	-0.06	112	4.08	7.07	-0.06
qGW3-2	3	RM8269	RM448	165	1.74	2.13	-0.03	165	0.49	0.61	-0.02	163	1.16	1.23	-0.03	163	3.13	3.38	-0.04
qGW5	5	RM267	RM405	15	19.62	30.50	-0.12	13	21.34	34.97	-0.12	15	22.59	31.24	-0.13	15	20.99	27.45	-0.12
qGW6	6	RM136	RM6818	59	0.66	0.88	-0.02	59	0.49	0.64	-0.02	58	4.44	5.86	-0.06	59	1.55	1.73	-0.03

^a A peak in a marker interval along the LOD profile was treated as a QTL, if there is at least one environment with the peak value higher than the LOD threshold

^b Chromosomal position (cM) of the peak

^c The number in bold was the value of a peak higher than the threshold value in the LOD profile

^d Percentage of the phenotypic variation explained by the locus at the peak position on the LOD profile

^e Additive effect of the identified QTL. Positive additive effect indicated the allele from IR24 increased grain width, and the allele from Asoinori decreased grain width. Negative additive effect indicated the allele from IR24 decreased grain width, and the allele from Asoinori increased grain width

Table 5 QTL of grain shape identified by grain length–width ratio (LW) in the four environments

QTL ^a	Chr.	Left marker	Right marker	Guilin (24.18°N, 109.45°E)			Guiyang (26.35°N, 106.42°E)			Nanchang (28.38°N, 116.24°E)			Nanjing (31.95°N, 119.16°E)						
				Pos. ^b	LOD ^c	PVE (%) ^d	Add ^e	Pos. ^b	LOD ^c	PVE (%) ^d	Add ^e	Pos. ^b	LOD ^c	PVE (%) ^d	Add ^e	Pos. ^b	LOD ^c	PVE (%) ^d	Add ^e
qLW1	1	RM488	RM212	75	1.56	2.59	0.06	74	1.13	1.68	0.04	84	1.99	1.59	0.05	84	2.84	2.17	0.05
qLW2-1	2	RM1313	RM424	66	6.54	6.21	0.09	66	0.06	0.04	0.01	66	0.08	0.06	0.01	66	3.32	2.41	0.06
qLW2-2	2	RM29	RM262	98	1.69	1.55	0.04	86	8.41	13.11	0.12	85	9.17	13.22	0.14	95	4.49	4.29	0.08
qLW3	3	RM411	RM7097	111	19.79	28.24	0.19	110	26.52	36.42	0.20	109	24.64	30.48	0.21	110	25.43	29.68	0.20
qLW4	4	RM5473	RM131	123	2.14	2.26	-0.05	121	2.91	2.61	-0.05	120	2.16	1.66	-0.05	122	1.77	1.38	-0.04
qLW5	5	RM267	RM405	26	17.33	31.71	0.20	15	15.40	14.99	0.13	15	17.22	16.13	0.15	15	16.55	13.88	0.14
qLW8	8	RM1235	RM6356	9	3.49	3.20	0.06	9	1.13	0.88	0.03	9	2.72	2.16	0.06	9	2.62	1.88	0.05
qLW9	9	RM434	RM3533	61	2.68	2.83	-0.06	66	2.38	2.14	-0.05	64	1.02	0.93	-0.04	63	1.50	1.26	-0.04

^a A peak in a marker interval along the LOD profile was treated as a QTL, if there is at least one environment with the peak value higher than the LOD threshold

^b Chromosomal position (cM) of the peak

^c The number in bold was the value of a peak higher than the threshold value in the LOD profile

^d Percentage of the phenotypic variation explained by the locus at the peak position on the LOD profile

^e Additive effect of the identified QTL. Positive additive effect indicated the allele from IR24 increased the grain length–width ratio, and the allele from Asoinori decreased the ratio. Negative additive effect indicated the allele from IR24 decreased the ratio, and the allele from Asoinori increased the ratio

Table 6 QTL of grain shape identified by grain circumference (GC) in the four environments

QTL ^a	Chr.	Left marker	Right marker	Guilin (24.18°N, 109.45°E)			Guiyang (26.35°N, 106.42°E)			Nanchang (28.38°N, 116.24°E)			Nanjing (31.95°N, 119.16°E)						
				Pos. ^b	LOD ^c	PVE (%) ^d	Add ^e	Pos. ^b	LOD ^c	PVE (%) ^d	Add ^e	Pos. ^b	LOD ^c	PVE (%) ^d	Add ^e	Pos. ^b	LOD ^c	PVE (%) ^d	Add ^e
qGC2-1	2	RM29	RM262	89	1.48	2.46	0.20	86	3.10	5.06	0.28	90	3.36	5.36	0.29	89	2.33	4.92	0.28
qGC2-2	2	RM6	RM425	140	7.76	9.92	-0.40	141	7.07	8.32	-0.36	140	5.65	6.97	-0.33	140	1.55	2.16	-0.19
qGC3-1	3	RM411	RM7097	113	27.50	48.83	0.90	113	28.37	51.29	0.88	111	31.53	54.09	0.91	111	26.48	52.31	0.92
qGC3-2	3	RM8269	RM448	163	5.86	6.09	-0.32	163	2.05	1.94	-0.17	163	5.87	5.70	-0.30	163	5.17	6.12	-0.32
qGC7	7	RM505	RM234	81	3.98	4.33	0.27	82	5.24	5.87	0.30	84	1.88	1.77	0.17	81	2.41	2.87	0.22
qGC8	8	RM502	RM6948	92	6.15	6.81	0.34	91	3.87	4.06	0.25	92	6.80	7.10	0.33	93	4.20	5.37	0.30

^a A peak in a marker interval along the LOD profile was treated as a QTL, if there is at least one environment with the peak value higher than the LOD threshold

^b Chromosomal position (cM) of the peak

^c The number in bold was the value of a peak higher than the threshold value in the LOD profile

^d Percentage of the phenotypic variation explained by the locus at the peak position on the LOD profile

^e Additive effect of the identified QTL. Positive additive effect indicated the allele from IR24 increased grain circumference, and the allele from Asominori decreased grain circumference. Negative additive effect indicated the allele from IR24 decreased grain circumference, and the allele from Asominori increased grain circumference

and GD, but negatively with GW and GR (Table 1). Greater GW will result in greater GA, GD, and GR, but smaller LW and GC. Therefore, GW was positively correlated with GA, GD, and GR, but negatively with LW and GC (Table 1). LW and GR are reciprocal from each other, and a completely negative correlation was observed (Table 1).

We understand that the 2D image of a rice grain may not completely be an ellipse, and therefore GL and GW may not completely determine GC and GA, the two characters which may most suitably represent size of the rice grain, but are hardly measured manually. In this study, GL, GW, GC, and GA were directly measured in the 2-D image system. The system also output LW, GR, and GD, but LW was actually calculated from GL and GW, and GD and GR were calculated from GA.

Advantages of the 2D image analysis in measuring grain shape

For manual measurement on grain shape, 10 or 20 filled grains were randomly selected, and then lined up lengthwise (or widthwise) along a vernier caliper in order to measure GL (or GW) (Table S3). Values of the filled grains were then averaged and used in genetic studies. The 2D image technology used in this study had advantages in measuring more characters directly, allowing a more complete description of the rice grain shape. The SC-G equipment was objective and high throughput. It takes at least 5 min to measure GL and GW of 10–20 grains manually. In comparison, SC-G can screen 800–1200 grains for the seven characters in 5–10 s. Measuring a larger number of grains from each RIL reduces the sampling errors associated with the phenotypic mean and therefore increases the estimated heritability (Table 2). In addition, the measurements from SC-G can be directly loaded into computer, so that some artificial errors in recording and transferring manual data can be greatly avoided. The image system is highly efficient in investigating grain shape characters and the WSeen product has been used in more than 300 institutes in China in recent 2 years.

In the 2D image system used in this study, GC and GA were directly measured, in addition to GL and GW. LW, GD, and GR were not directly measured, which may be called mathematically derived traits (Wang et al. 2012a, b). The use of such traits increased gene number, caused higher-order gene interactions than observed in component traits, and possibly complicated the linkage relationship between QTL as well (Wang et al. 2012a, b). The increased complexity of genetic architecture in derived traits may reduce QTL detection power and increase false discovery rate. Therefore, additional characters which can be directly measured by the 2D image system, such as GC and GA, may also contribute to more efficient and precise dissection

Table 7 QTL of grain shape identified by grain area (GA) in the four environments QTL

QTL ^a	Chr.	Left marker	Right marker	Guilin (24.18°N, 109.45°E)			Guiyang (26.35°N, 106.42°E)			Nanchang (28.38°N, 116.24°E)			Nanjing (31.95°N, 119.16°E)						
				Pos. ^b	LOD ^c	PVE (%) ^d	Add ^e	Pos. ^b	LOD ^c	PVE (%) ^d	Add ^e	Pos. ^b	LOD ^c	PVE (%) ^d	Add ^e	Pos. ^b	LOD ^c	PVE (%) ^d	Add ^e
qGA2-1	2	RM211	RM71	19	1.72	6.93	-0.44	11	0.77	1.69	-0.21	13	0.83	2.95	-0.28	14	3.97	18.97	-0.70
qGA2-2	2	RM6	RM425	140	7.61	12.09	-0.58	139	6.23	7.61	-0.45	141	6.95	10.97	-0.53	140	1.45	2.76	-0.27
qGA3-1	3	RM411	RM7097	111	8.59	15.21	0.66	111	17.31	26.36	0.84	114	12.33	27.15	0.84	112	12.14	30.37	0.89
qGA3-2	3	RM8269	RM448	177	8.31	19.22	-0.74	183	2.52	5.43	-0.38	184	6.41	17.76	-0.68	163	4.56	7.39	-0.44
qGA5	5	RM267	RM405	15	8.28	10.89	-0.55	15	11.69	12.44	-0.58	15	8.36	11.90	-0.56	15	5.20	8.58	-0.47
qGA6-1	6	RM528	RM340	93	0.66	0.83	-0.16	95	6.16	7.50	-0.46	93	0.66	0.90	-0.16	93	0.35	0.56	-0.12
qGA6-2	6	RM439	RM412	112	1.15	1.37	0.20	112	7.54	7.76	0.46	112	1.12	1.45	0.20	112	0.91	1.44	0.19
qGA7	7	RM505	RM234	83	3.77	4.98	0.38	82	5.32	5.73	0.39	82	3.54	5.16	0.37	82	2.48	4.10	0.33
qGA8	8	RM502	RM6948	92	6.81	9.48	0.52	93	6.09	6.59	0.42	94	2.07	2.87	0.27	93	2.23	3.67	0.31
qGA10	10	RM333	RM590	68	1.65	1.96	-0.24	68	2.90	3.02	-0.29	68	0.91	1.17	-0.17	68	1.12	1.73	-0.21

^a A peak in a marker interval along the LOD profile was treated as a QTL, if there is at least one environment with the peak value higher than the LOD threshold

^b Chromosomal position (cM) of the peak

^c The number in bold was the value of a peak higher than the threshold value in the LOD profile

^d Percentage of the phenotypic variation explained by the locus at the peak position on the LOD profile

^e Additive effect of the identified QTL. Positive additive effect indicated the allele from IR24 increased grain area, and the allele from Asominori decreased grain area. Negative additive effect indicated the allele from IR24 decreased grain area, and the allele from Asominori increased grain area

Table 8 QTL of grain shape identified by grain diameter (GD) in the four environments

QTL ^a	Chr.	Left marker	Right markerGuilin (24.18°N, 109.45°E)			Guiyang (26.35°N, 106.42°E)			Nanchang (28.38°N, 116.24°E)			Nanjing (31.95°N, 119.16°E)						
			Pos. ^b	LOD ^c	PVE (%) ^d	Add ^e	Pos. ^b	LOD ^c	PVE (%) ^d	Add ^e	Pos. ^b	LOD ^c	PVE (%) ^d	Add ^e	Pos. ^b	LOD ^c	PVE (%) ^d	Add ^e
qGD2	2	RM6	141	9.34	12.74	-0.11	141	7.55	9.63	-0.09	143	6.19	7.52	-0.08	142	4.48	6.04	-0.07
qGD3-1	3	RM411	113	24.11	44.52	0.21	113	25.72	48.27	0.21	113	26.02	50.85	0.21	113	24.10	48.11	0.21
qGD3-2	3	RM18269	163	6.00	6.73	-0.08	163	1.89	1.93	-0.04	164	6.54	7.61	-0.08	163	5.76	6.86	-0.08
qGD5	5	RM267	15	1.77	1.82	-0.04	15	3.28	3.50	-0.06	15	1.38	1.47	-0.04	15	0.35	0.39	-0.02
qGD7	7	RM505	81	3.55	4.19	0.06	89	4.72	5.94	0.07	82	2.02	2.36	0.05	81	3.83	4.86	0.07
qGD8	8	RM502	92	6.98	8.33	0.09	92	4.55	5.48	0.07	93	6.39	8.09	0.08	91	2.92	3.50	0.06

^a A peak in a marker interval along the LOD profile was treated as a QTL, if there is at least one environment with the peak value higher than the LOD threshold

^b Chromosomal position (cM) of the peak

^c The number in bold was the value of a peak higher than the threshold value in the LOD profile

^d Percentage of the phenotypic variation explained by the locus at the peak position on the LOD profile

^e Additive effect of the identified QTL. Positive additive effect indicated the allele from IR24 increased grain diameter, and the allele from Aominori decreased grain diameter. Negative additive effect indicated the allele from IR24 decreased grain diameter, and the allele from Aominori increased grain diameter

Table 9 QTL of grain shape identified by grain roundness (GR) in the four environments

QTL ^a	Chr.	Left marker	Right markerGuilin (24.18°N, 109.45°E)			Guiyang (26.35°N, 106.42°E)			Nanchang (28.38°N, 116.24°E)			Nanjing (31.95°N, 119.16°E)						
			Pos. ^b	LOD ^c	PVE (%) ^d	Add ^e	Pos. ^b	LOD ^c	PVE (%) ^d	Add ^e	Pos. ^b	LOD ^c	PVE (%) ^d	Add ^e	Pos. ^b	LOD ^c	PVE (%) ^d	Add ^e
qGR1	1	RM488	78	1.44	2.28	-0.01	74	0.81	1.20	-0.01	84	1.92	1.50	-0.01	84	2.78	2.29	-0.01
qGR2-1	2	RM1313	66	5.96	6.55	-0.01	66	3.06	2.53	-0.01	66	4.33	3.43	-0.01	66	3.04	2.38	-0.01
qGR2-2	2	RM29	98	1.53	1.65	-0.01	95	3.58	3.81	-0.01	93	3.23	3.83	-0.01	98	4.60	3.87	-0.01
qGR3	3	RM411	109	18.93	28.58	-0.03	110	25.25	34.00	-0.03	110	23.45	28.86	-0.03	110	24.64	30.71	-0.03
qGR4	4	RM5473	121	1.18	1.32	0.01	121	3.11	2.81	0.01	120	2.18	1.64	0.01	121	1.34	1.10	0.01
qGR5	5	RM267	13	12.23	14.88	-0.02	15	16.06	15.83	-0.02	15	18.87	17.67	-0.02	15	16.44	15.10	-0.02
qGR8	8	RM1235	9	2.10	2.12	-0.01	9	1.09	0.85	0.00	9	2.84	2.22	-0.01	9	1.92	1.42	-0.01

^a A peak in a marker interval along the LOD profile was treated as a QTL, if there is at least one environment with the peak value higher than the LOD threshold

^b Chromosomal position (cM) of the peak

^c The number in bold was the value of a peak higher than the threshold value in the LOD profile

^d Percentage of the phenotypic variation explained by the locus at the peak position on the LOD profile

^e Additive effect of the identified QTL. Positive additive effect indicated the allele from IR24 increased grain roundness, and the allele from Aominori decreased grain roundness. Negative additive effect indicated the allele from IR24 decreased grain roundness, and the allele from Aominori increased grain roundness

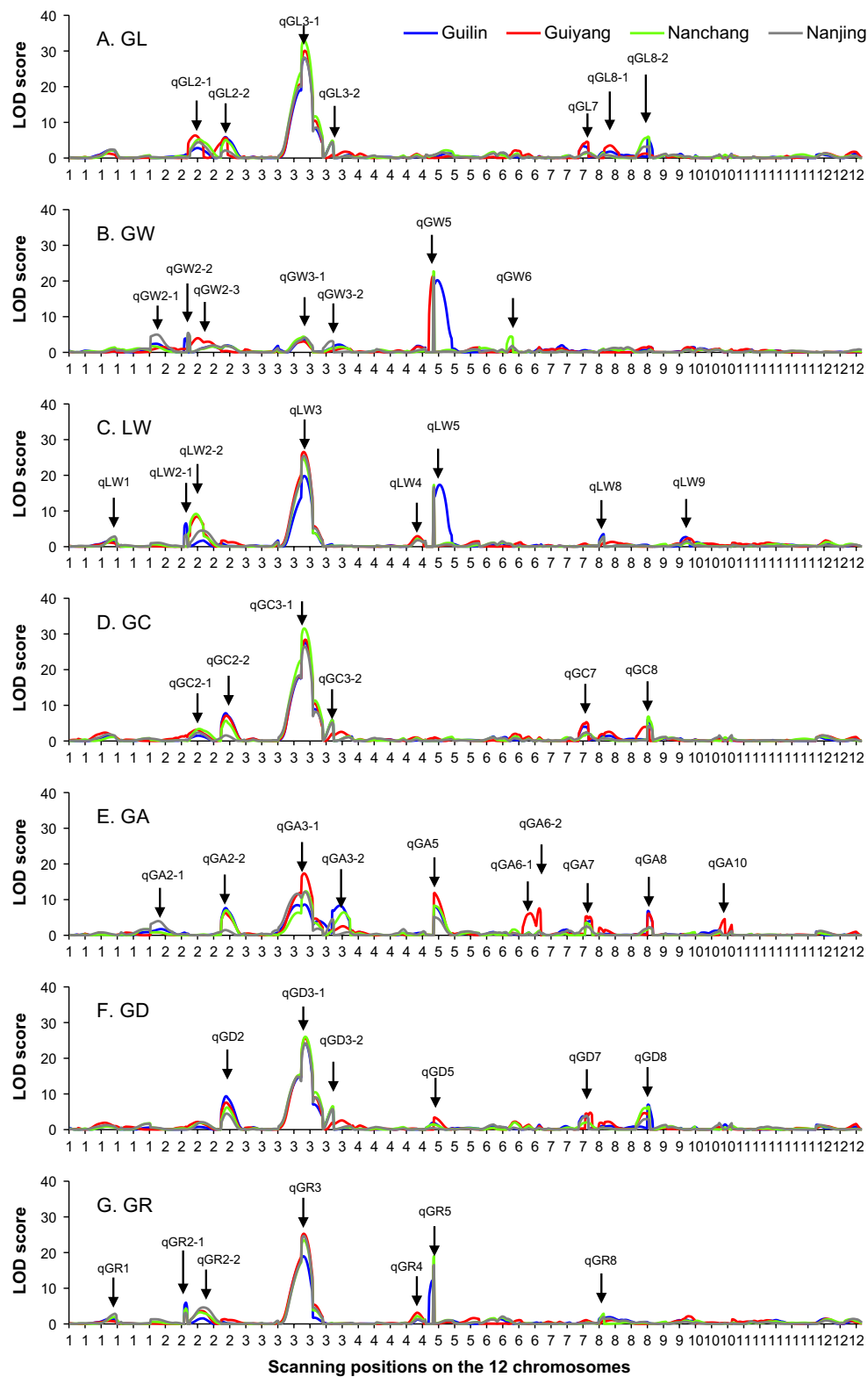


Fig. 5 LOD profile of QTL mapping for the seven characters in four environments. Each identified QTL was indicated by an arrow pointing to the peak on the LOD profile. *GL* grain length, *GW* grain width,

LW grain length-to-width ratio, *GC* grain circumference, *GA* grain area, *GD* grain diameter, and *GR* grain roundness

Table 10 Summary of the identified QTL in the four environments, and reported genes/QTL from previous studies

Marker interval	Chr.	Guilin (24.18°N, 109.45°E)	Guiyang (26.35°N, 106.42°E)	Nanchang (28.38°N, 116.24°E)	Nanjing (31.95°N, 119.16°E)	Reported genes/QTL
RM488–RM212	1				qLW1, qGR1	kl1.1, gl1
RM211–RM71	2				qGW2-1, qGA2-1	GW2, qGL-2b
RM1313–RM424	2	qLW2-1, qGW2-2, qGR2-1	qGR2-1	qGL2-1, qGW2-3, qLW2-2, qGC2-1	qLW2-1, qGR2-1	Not reported
RM29–RM262	2	qGL2-1	qGL2-1, qGW2-3, qLW2-2, qGC2-1	qGL2-1, qGW2-3, qLW2-2	qGL2-1, qGW2-3, qLW2-2	qGL-2, <i>PGL2</i> , qLWLR-2
RM6–RM425	2	qGL2-2, qGC2-2, qGA2-2, qGD2	qGL2-2, qGC2-2, qGA2-2, qGD2	qGL2-2, qGC2-2, qGA2-2, qGD2	qGD2	qGL-2a, <i>PGL2</i>
RM411–RM7097	3	qGL3-1, qGW3-1, qLW3, qGC3-1, qGA3-1, qGD3-1, qGR3	qGL3-1, qGW3-1, qLW3, qGC3-1, qGA3-1, qGD3-1, qGR3	qGL3-1, qGW3-1, qLW3, qGC3-1, qGA3-1, qGD3-1, qGR3	qGL3-1, qGW3-1, qLW3, qGC3-1, qGA3-1, qGD3-1, qGR3	<i>GS3</i> , qGL-3, qLWLR-3, qGW-3
RM8269–RM448	3	qGL3-2, qGC3-2, qGA3-2, qGD3-2	qGL3-2, qGC3-2, qGA3-2, qGD3-2	qGL3-2, qGC3-2, qGA3-2, qGD3-2	qGL3-2, qGW3-2, qGC3-2, qGA3-2, qGD3-2	Not reported
RM5473–RM131	4	qLW4, qGR4	qLW4, qGR4			qGL-4
RM267–RM405	5	qGW5, qLW5, qGA5, qGR5	qGW5, qLW5, qGA5, qGD5, qGR5	qGW5, qLW5, qGA5, qGR5	qGW5, qLW5, qGA5, qGR5	<i>gw5/qSW5</i> , <i>srs-3</i> , qLWLR-5
RM136–RM6818	6				qGW6	qGW-6
RM528–RM340	6	qGA6-1	qGA6-1			Not reported
RM439–RM412	6	qGA6-2	qGA6-2			Not reported
RM505–RM234	7	qGL7, qGC7, qGA7, qGD7	qGL7, qGC7, qGA7, qGD7		qGD7	qGL7-2
RM1235–RM6356	8	qLW8	qLW8, qGR8			Not reported
RM6863–RM72	8	qGL8-1	qGL8-1			Not reported
RM502–RM6948	8	qGL8-2, qGC8, qGA8, qGD8	qGL8-2, qGC8, qGA8, qGD8		qGL8-2, qGC8, qGD8	<i>qGW8/OsSPL16</i>
RM434–RM3533	9	qLW9-1				<i>DEP1/qPE9-1</i> , <i>SG1</i>
RM333–RM590	10	qGA10	qGA10			kl10.1

of the genetic architecture on grain shape. If GL and GW can be viewed as one-dimensional (1D) characters, GC and GA are 2D characters. Intuitively, 2D characters may better describe the grain shape regarding the size and volume. The 2D image analysis can not only facilitate the traditional genetic studies on 1D characters, it also allows the genetic dissection of grain shape from directly measured 2D characters.

Genetic architecture on grain shape

Grain shape is complex by phenotyping, as there is no single character which can completely quantify the shape of a grain. It is complex by genetics, as there is no single gene which can completely determine the shape of a grain. Fortunately, there are some characters which are closely associated with grain shape. These characters can be precisely measured in large scale, and have low genotype by environmental interactions, low random errors, high heritability, and high correlation relationships (Tables 1 and 2). These features explained why such a complex trait has received so much attention in rice genetics studied in past two decades.

From the seven characters investigated in the 215 RILs, a total of 51 QTL were identified to have additive effects (Tables 3, 4, 5, 6, 7, 8, and 9). Major and minor loci both exist. The major locus explained more than 20 %, while the minor locus explained a few percentage of the phenotypic variation. The identified QTL have varied stabilities across the four environments. Some were detected in four environments, but some were detected in three, two, or just one environment. Interestingly, QTL not significant in four environments also showed peaks in non-significant environments, and the additive effects were always at the same direction as those in the significant environments. This may represent another important feature of grain shape, i.e., there may not be cross-over GE interactions at the identified locus. Though each QTL has different effects in different environments, the difference will not change which allele is favorable and which allele is unfavorable. This feature of grain shape was confirmed by the low GE interactions from ANOVA (Table S2).

RIL populations are widely used in QTL mapping, where each line is homozygous in genotype and can be grown in multi-locations with replications for precision phenotyping. Additive QTL can be mapped by one-dimensional scanning, and additive by additive epistatic QTL can be mapped by 2D scanning. No dominance and dominance-associated epistasis can be studied in RIL populations. We conducted epistatic mapping (Li et al. 2008) for the 215 RIL, but did not detect significant epistatic effects. Similar results have been reported in previous studies (Huang et al. 2013). To summarize, grain shape can be precisely

measured by various 1D and 2D characters, GE interaction is low, and heritability is high. It is controlled by a few major stable genes and multiple minor additive genes.

Novelty of the six non-reported chromosomal intervals

The 51 QTL on the seven characters were clustered into 18 chromosomal intervals flanked by SSR markers (Table 10). We went through previous literatures for QTL and genes on grain shape, and compared with QTL identified in this study by physical locations or associated markers (Tables 10 and S3). Genes/QTL have been reported in 12 intervals but not yet in the other six intervals. Obviously, intervals showing QTL from multiple characters and multiple environments have been previously reported (Table 10). But this is not always the case, for example see intervals RM488–RM212 on chromosome 1, RM211–RM71 on chromosome 2, RM5473–RM131 on chromosome 4, RM136–RM6818 on chromosome 6, and RM333–RM590 on chromosome 10. We assume the six non-reported intervals may harbor novel loci on grain shape, which are worth of further investigations. Three most promising intervals are RM1313–RM424 on chromosome 2, RM8269–RM448 on chromosome 3, and RM1235–RM6356 on chromosome 8, where QTL showed up for multiple characters and in multiple environments.

Of course, we cannot exclude the possibility that some of the 51 QTL may be false positives. This problem cannot be solved in the current mapping population. In the meantime of developing the RIL population, we also developed two-way chromosome segment substitution lines (CSSLs). We are using CSSLs to confirm QTL identified in the RIL population.

Author contribution statement C. Yin developed the genetic population, conducted the SSR genotyping and field experiments, constructed the genetic linkage map, conducted the QTL analysis, and wrote the manuscript draft; H. Li and S. Li conducted phenotypic and genotypic data analysis; L. Xu and Z. Zhao investigated the 2D characters of rice grain in the genetic population; J. Wang designed the research, and finalized the manuscript.

Acknowledgments We thank Shijia Liu, Liangming Chen, Xi Liu, Yunlu Tian, Susong Zhu, Dashuang Zhang, Danting Li, Zhiqian Wang, and Kun Zhang for their assistance in field management and sample preparation. This work was supported by the National Basic Scientific Research (973) Programs (Project No. 2011CB100106) and the National Natural Science Foundation (Project No. 31271798) of China.

Conflict of interest The authors declare that they have no conflict of interest.

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

- Aluko G, Martinez C, Tohme J, Castano C, Bergman C, Oard JH (2004) QTL mapping of grain quality traits from the interspecific cross *Oryza sativa* × *O. glaberrima*. *Theor Appl Genet* 109:630–639
- Bai X, Luo L, Yan W, Kovi MR, Zhan W, Xing Y (2010) Genetic dissection of rice grain shape using a recombinant inbred line population derived from two contrasting parents and fine mapping a pleiotropic quantitative trait locus qGL7. *BMC Genet* 11:16
- Chen X, Temnykh S, Xu Y, Cho YG, McCouch SR (1997) Development of a microsatellite framework map providing genome-wide coverage in rice (*Oryza sativa* L.). *Theor Appl Genet* 95:553–567
- Dellaporta SL, Wood J, Hicks JB (1983) A plant DNA miniprep: version II. *Plant Mole Biol Report* 1:19–21
- 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
- Heang D, Sassa H (2012a) Antagonistic actions of HLH/bHLH proteins are involved in grain length and weight in rice. *Plos One* 7(2):e31325
- Heang D, Sassa H (2012b) An atypical bHLH protein encoded by *POSITIVE REGULATOR OF GRAIN LENGTH 2* is involved in controlling grain length and weight of rice through interaction with a typical bHLH protein APG. *Breed Sci* 62:133–141
- Holland JB, Nyquist WE, Cervantes-Martinez CT (2003) Estimating and interpreting heritability for plant breeding: an update. *Plant Breed Rev* 22:9–112
- Huang X, Qian Q, Liu Z, Sun H, He S, Luo D, Xia G, Chu C, Li J, Fu X (2009) Natural variation at the *DEP1* locus enhances grain yield in rice. *Nat Genet* 41:494–497
- 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
- Kitagawa K, Kurinami S, Oki K, Abe Y, Ando T, Kono I, Yano M, Kitano H, Iwasaki Y (2010) A novel kinesin 13 protein regulating rice seed length. *Plant Cell Physiol* 51:1315–1329
- Li J, Yu S, Xu C, Tan Y, Gao Y, Li X, Zhang Q (2000) Analyzing quantitative trait loci for yield using a vegetatively replicated F2 population from a cross between the parents of an elite rice hybrid. *Theor Appl Genet* 101:248–254
- Li Z, Wan J, Xia J, Zhai H (2003) Mapping quantitative trait loci underlying appearance quality of rice grains (*Oryza sativa* L.). *Acta Genetica Sin* 30:251–259
- Li J, Xiao J, Grandillo S, Jiang L, Wan Y, Deng Q, Yuan L, McCouch SR (2004) QTL detection for rice grain quality traits using an interspecific backcross population derived from cultivated Asian (*O. sativa* L.) and African (*O. glaberrima* S.) rice. *Genome* 47:697–704
- Li H, Ye G, Wang J (2007) A modified algorithm for the improvement of composite interval mapping. *Genetics* 175:361–374
- Li H, Ribaut JM, Li Z, Wang J (2008) Inclusive composite interval mapping (ICIM) for digenic epistasis of quantitative traits in biparental populations. *Theor Appl Genet* 116:243–260
- Li H, Hearne S, Bänziger M, Li Z, Wang J (2010) Statistical properties of QTL linkage mapping in biparental genetic populations. *Heredity* 105:257–267
- Li Y, Fan C, Xing Y, Jiang Y, Luo L, Sun L, Shao D, Cu 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
- 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 Nat Acad Sci* 107:19579–19584
- McCouch SR, Cho YG, Yano M, Paul E, Blinstrub M, Morishima H, Kinoshita T (1997) Report on QTL nomenclature. *Rice Genet Newsl* 14:11–131
- McCouch SR, Teytelman L, Xu Y, Lobos KB, Clare K, Walton M, Fu B, Maghirang R, Li Z, Xing Y, Zhang Q, Kono I, Yano M, Fjellstrom R, DeClerck G, Schneider D, Cartinhour S, Ware D, Stein L (2002) Development and mapping of 2240 new SSR markers for rice (*Oryza sativa* L.). *DNA Res* 9:199–207
- Meng L, Li H, Zhang L, Wang J (2015) QTL IciMapping: integrated software for genetic linkage map construction and quantitative trait locus mapping in bi-parental populations. *Crop J* 3:265–279
- Nakagawa H, Tanaka A, Tanabata T, Ohtake M, Fujioka S, Nakamura H, Ichikawa H, Mori M (2012) *SHORT GRAINS* decreases organ elongation and brassinosteroid response in rice. *Plant Physiol* 158:1208–1219
- Sanguinetti CJ, Dias NE, Simpson AJ (1994) Rapid silver staining and recovery of PCR products separated on polyacrylamide gels. *Biotechniques* 17:914–921
- Shao G, Tang S, Luo J, Jiao G, Wei X, Tang A, Wu J, Zhuang J, Hu P (2010) Mapping of qGL7-2, a grain length QTL on chromosome 7 of rice. *J Genet Genomics* 37:523–531
- Shomura A, Izawa T, Ebana K, Ebitani T, Kanegae 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 X, Huang W, Shi M, Zhu M, Lin H (2007) A QTL for rice grain width and weight encodes a previously unknown RING-type E3 ubiquitin ligase. *Nat Genet* 39:623–630
- Sun H, Qian Q, Wu K, Luo J, Wang S, Zhang C, Ma Y, Liu Q, Huang X, Yuan Q, Han R, Zhao M, Dong G, Guo L, Zhu X, Gou Z, Wang W, Wu Y, Lin H, Fu X (2014) Heterotrimeric G proteins regulate nitrogen-use efficiency in rice. *Nat Genet* 46:652–656
- Taguchi-Shiobara F, Kawagoe Y, Kato H, Onodera H, Tagiri A, Hara N, Miyao A, Hirochika H, Yano M, Toki S (2011) A loss-of-function mutation of rice *DENSE PANICLE 1* causes semi-dwarfness and slightly increased number of spikelets. *Breed Sci* 61:17–25
- Takano-Kai N, Jiang H, Kubo T, Sweeney M, Matsumoto T, Kanamori H, Padhukasahasram B, Bustamante C, Yoshimura A, Dio K, McCouch S (2009) Evolutionary history of *GS3*, a gene conferring grain length in rice. *Genetics* 182:1323–1334
- Takano-Kai N, Doi K, Yoshimura A (2011) *GS3* participates in stigma exertion as well as seed length in rice. *Breed Sci* 61:244–250
- Tan Y, Xing Y, Li J, Yu S, Xu C, Zhang Q (2000) Genetic bases of appearance quality of rice grains in Shanyou 63, an elite rice hybrid. *Theor Appl Genet* 101:823–829
- Tanabe S, Kurinami S, Ashikari M, Kitano H, Iwasaki Y (2007) Mapping of small and round seed 3 gene in rice. *Rice Genet Newsl* 23:56–58
- Wan X, Wan J, Weng J, Jiang L, Bi J, Wang C, Zhai H (2005) Stability of QTL for rice grain dimension and endosperm chalkiness characteristics across eight environments. *Theor Appl Genet* 110:1334–1346
- Wan X, Weng J, Zhai H, Wang J, Lei C, Liu X, Guo T, Jiang L, Sun N, Wan J (2008) Quantitative trait loci (QTL) analysis for rice grain width and fine mapping of an identified QTL allele *gw-5*

- in a recombination hotspot region on chromosome 5. *Genetics* 179:2239–2252
- Wang J (2009) Inclusive composite interval mapping of quantitative trait genes. *Acta Agron Sin* 35:239–245
- Wang J, Bernardo R (2000) Variance of marker estimates of parental contribution to F_2 and BC_1 -derived inbreds. *Crop Sci* 40:659–665
- Wang E, Wang J, Zhu X, Hao W, Wang L, Li Q, Zhang L, Lu B, Lin H, Ma H, Zhang G, He Z (2008) Control of rice grain-filling and yield by a gene with a potential signature of domestication. *Nat Genet* 40:1370–1374
- Wang E, Xu X, Zhang L, Zhang H, Lin L, Wang Q, Li Q, Ge S, Wang W, He Z (2010) Duplication and independent selection of cell-wall invertase genes *GIF1* and *OsCINI* during rice evolution and domestication. *BMC Evol Biol* 10:108
- Wang C, Chen S, Yu S (2011) Functional markers developed from multiple loci in *GS3* for fine marker-assisted selection of grain length in rice. *Theor Appl Genet* 122:905–913
- Wang S, Wu K, Yuan Q, Liu X, Liu Z, Lin X, Zeng R, Zhu H, Dong G, Zhang G, Fu X (2012a) Control of grain size, shape and quality by *OsSPL16* in rice. *Nat Genet* 44:950–954
- Wang Y, Li H, Zhang L, Lü W, Wang J (2012b) On the use of mathematically-derived traits in QTL mapping. *Mol Breed* 29:661–673
- 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
- Xing Y, Tan Y, Xu C, Hua J, Sun X (2000) Mapping quantitative trait loci for grain appearance traits of rice using a recombinant inbred line population. *Acta Botanica Sin* 43:840–845
- Yan C, Zhou J, Yan S, Chen F, Yeboah M, Tang S, Liang G, Gu M (2007) Identification and characterization of a major QTL responsible for erect panicle trait in japonica rice (*Oryza sativa* L.). *Theor Appl Genet* 115:1093–1100
- Yan S, Zou G, Li S, Wang H, Liu H, Zhai G, Guo P, Song H, Yan C, Tao Y (2011) Seed size is determined by the combinations of the genes controlling different seed characteristics in rice. *Theor Appl Genet* 123:1173–1181
- Yang W, Duan L, Chen G, Xiong L, Liu Q (2013) Plant phenomics and high-throughput phenotyping: accelerating rice functional genomics using multidisciplinary technologies. *Curr Opin Plant Biol* 16:180–187
- Yi X, Zhang Z, Zeng S, Tian C, Peng J, Li M, Lu Y, Meng Q, Gu M, Yan C (2011) Introgression of *qPE9-1* allele, conferring the panicle erectness, leads to the decrease of grain yield per plant in japonica rice (*Oryza sativa* L.). *J Genet Genomics* 38:217–223
- Yu S, Li J, Xu C, Tan Y, Gao Y, Li X, Zhang Q, Maroof M (1997) Importance of epistasis as the genetic basis of heterosis in an elite rice hybrid. *Proc Natl Acad Sci* 94:9226–9231
- Zhong Q, Zhou P, Yao Q, Mao K (2009) A novel segmentation algorithm for clustered slender-particles. *Comput Electron Agric* 69:118–127
- Zhou Y, Zhu J, Li Z, Yi C, Liu J, Zhang H, Tang S, Gu M, Liang G (2009) Deletion in a quantitative trait gene *qPE9-1* associated with panicle erectness improves plant architecture during rice domestication. *Genetics* 183:315–324