Dissecting the Genetic Basis of Grain Shape and Chalkiness Traits in Hybrid Rice Using Multiple Collaborative Populations

Dear Editor,

Through the efficient use of heterosis, hybrid rice varieties generally have higher grain yield potential than inbred varieties. With the significant advantage in grain yield, over the past 30 years approximately half of China's total rice-growing area is planted with rice hybrids. However, grain quality has now become one of the most important targets in hybrid rice breeding for meeting consumer demands. Grain shape and chalkiness are two important components of rice grain quality, in which slender grains (typically, grain length-to-width ratio >3) with low chalkiness are preferred by most consumers of hybrid rice. Especially, grain chalkiness not only brings down the grain appearance but also has a negative influence on milling and cooking properties. Previous studies have characterized several important genes controlling grain shape traits, e.g., *GS3* (Fan et al., 2006) affecting grain length, *qSW5/GW5/GSE5* (Shomura et al., 2008; Duan et al., 2017; Liu et al., 2017) affecting grain width, *GL7/GW7* (Wang et al., 2015a, 2015b) shaping both grain length and grain width. By contrast, only a few quantitative trait locus (QTLs) responsible for grain chalkiness have been finely mapped and functionally characterized. *Chalk5*, which encodes a vacuolar H$^+$-translocating pyrophosphatase, is the first cloned and functionally characterized gene that controls rice grain chalkiness (Li et al., 2014). However, the genetic architecture of the two traits, grain shape and grain chalkiness, in hybrid rice remains unclear.

For genetic dissection of these complex traits, two methods are commonly applied: QTL mapping in bi-parental recombinant populations and genome-wide association studies (GWAS) using diverse varieties. In general, the bi-parental linkage approach has limitations in genetic diversity and mapping resolution, while conventional GWAS is often perplexed by complicated population structure and low power to map the low-frequency alleles (Nordborg and Weigel, 2008; Myles et al., 2009). Recently, for genetic analysis of heterosis in hybrid rice, we selected 17 representative hybrid crosses and constructed large genetic populations consisting of totally 10 074 F_2 lines. The collaborative recombinant populations enabled powerful mapping of yield traits, because the problem of population structure was largely avoided and both historic and recent recombination events were utilized.

Here, we used the large populations of hybrid rice for genetic dissections of grain quality traits. We carefully phenotyped these F_2 lines for grain length, grain width, grain length-to-width ratio (an important characteristic of grain shape), chalky grain rate, chalk size, and degree of chalkiness, all of which showed a broad phenotypic distribution in the population (Supplemental Figure 1A). Hierarchical clustering of all six traits showed correlations among grain shape traits, and those among chalkiness traits were relatively strong (Supplemental Figure 1B). However, there were only very weak correlation between grain length-to-width ratio and degree of chalkiness, implying that the genetic architectures of the two traits should be different (Figure 1C). We generated a high-density genotype map for the rice lines through whole-genome resequencing, and a total of 1 482 139 single nucleotide polymorphisms (SNPs) could be used in GWAS after removing the SNPs with a high missing-data rate (>20%) and low minor allele frequency (<2%). Through conducting GWAS using the panel combining all the F_2 lines, we identified 12 GWAS loci underlying the grain length-to-width ratio and 11 GWAS loci underlying the chalky grain rate (Figure 1A and 1B, Supplemental Table 1). Manhattan plots for the other four traits are provided in Supplemental Figure 2. It was found that the 12 GWAS loci underling the grain length-to-width ratio together explained 79.6% of the phenotypic variation, while the heritability of this trait was estimated to be ~91.3% in hybrid rice, suggesting that the GWAS loci identified here may explain the majority of the heritable variation for grain shape in hybrid rice. In contrast, the GWAS loci for grain chalkiness could only explain 18.2% of the phenotypic variation for chalky grain rate, probably because grain chalkiness had a low heritability and many loci with minor effects for chalkiness were not identified.

We noticed that the identified GWAS loci included many well-characterized genes: 11 genes for grain shape traits and nine for grain chalk traits. Among them, five genes (*qSW5/GW5/GSE5*, *GL7/GW7*, *Ghd8/DTH8*, *Hd3a*, and *Tms5*) played a role in both traits. For *qSW5/GW5/GSE5* and *GL7/GW7*, the allele with higher length-to-width ratio showed a lower level of grain chalkiness (Figure 1E and Supplemental Figure 3A). However, the advantageous allele of *Ghd8/DTH8* (with higher grain yield and lower chalkiness) showed a lower length-to-width ratio, suggesting that the grain chalkiness trait is controlled by many complicated factors (Supplemental Figure 3B). There were still a few GWAS loci in which the causal genes remain to be characterized. For example, one strong association was identified on chromosome 12 (namely *CGR12q25/LWR12q25* here, $P = 3.3 \times 10^{-13}$ in the association by the mixed model), which to our knowledge has not been reported. However, ~20 genes are located around the association locus (Figure 1D). Further genetic and functional studies are needed to validate
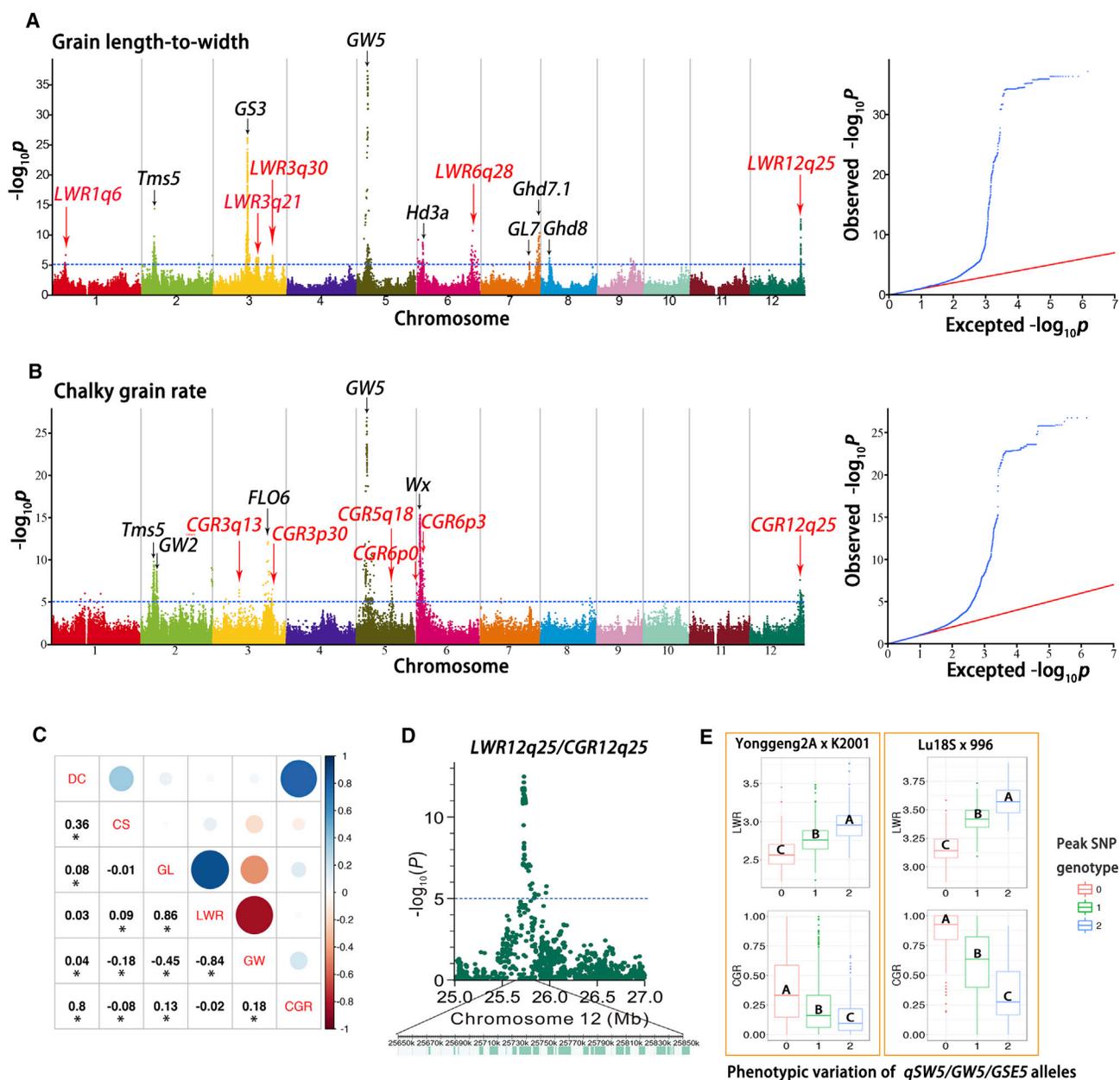
Figure 1. Genome-Wide Association Study of Grain Length-to-Width Ratio and Chalky Grain Rate Using the MLM Model.

(A) Manhattan plot and quantile-quantile plot for grain length-to-width ratio. The blue dashed line represents the significant threshold ($P = 1 \times 10^{-5}$). For the significant loci identified, candidate genes are shown with black gene symbols and newly discovered loci are indicated by red symbols with red arrows. Detailed information on the significant loci for all traits is listed in Supplemental Table 1.

(B) Manhattan plot and quantile-quantile plot for chalky grain rate and the genome-wide significant threshold ($P = 1 \times 10^{-5}$, shown as blue dashed line).

(C) Pearson's correlation coefficients between phenotype values of six grain quality traits. The two-tailed t test was applied to test the significance of correlation coefficients (*$P < 0.01$). DC, degree of chalkiness; CS, chalk size; CGR, chalky grain rate; LWR, grain length-to-width ratio; GL, grain length; GW, grain width.

(D) Local Manhattan plot for grain length to width (*LWR12q25/CGR12q25* were detected in this narrow region).

(E) Measurement of the grain length-to-width ratio and chalky grain rate of *qSW5/GW5/GSE5*. Peak SNP genotype 0 is the homozygous reference type. Peak SNP genotype 1 represents the heterozygous type, while two stands represent the homozygous alternative type.

the involvement of these candidate genes in regulating rice grain quality.

The effect size of each GWAS locus was analyzed separately (Supplemental Table 1). As expected, for grain length-to-width

ratio, three major genes, *GS3*, *qSW5/GW5/GSE5*, and *LWR3p21*, showed very large phenotypic effects, while the other nine GWAS loci showed relatively modest effects. We investigated the co-location of GWAS loci for grain shape traits. Among them, two GWAS loci for grain length-to-width ratio

were also responsible majorly for grain length. Four GWAS loci for grain length-to-width had co-location with the loci for grain width (Supplemental Figure 4). The heterozygous state of the GWAS loci mostly showed partial dominance effects, ranging between the effects of two homozygous genotypes (Supplemental Figure 5). Hence, the 12 GWAS loci could create 3^{12} genotypic combinations in principle (n = 531 441), resulting in diverse grain shapes in hybrid rice. We screened the collection of all the F_2 lines, and found 8455 combinations with various phenotypes for grain length-to-width ratio. For chalky grain rate, four highly significant associated ($-\log_{10}P > 10$) GWAS loci ($qSW5/GW5/GSE5$, Wx, $FLO6$, $CGRR6p3$) and seven GWAS loci with relatively weak associations ($-\log_{10}P > 5$) were identified. The candidates for the GWAS loci underlying grain chalkiness include genes involving grain shape (e.g., $qSW5/GW5/GSE5$, $GL7/GW7$), starch synthesis (e.g., Wx, ALK), fertility (e.g., $Tms5$) and heading date (e.g., $Hd3a$, $Ghd8$), indicating that the molecular mechanism of grain chalkiness was rather complex. We investigated the frequency of low-chalkiness alleles in the rice hybrids and found that most of them have a low or modest frequency, showing that there was great potential for improving grain quality. The strong associations identified here could be important targets for marker-assisted selection of low-chalkiness grains in rice breeding.

Predicting the phenotypic performance accurately only using the information from a few genetic markers rather than genomic selection in which a large number of polymorphisms covering the whole genome are used is an attractive approach. According to the genotypic information of the GWAS loci, we were able to predict the grain length-to-width ratio, grain length, and grain width with 78.67%, 73.53%, and 58.10% accuracy, respectively (Supplemental Figure 6). The prediction accuracy of the three traits could be increased when the genetic interactions between loci at the statistical level were considered (Supplemental Figure 6 and Supplemental Tables 2–4), although the molecular mechanisms of the interactions (i.e., epistasis) were still largely unknown. We also tried to introduce a dummy variable (or binary variable) in our regression analysis to indicate the absence or presence of a reference type and alternative type of SNPs, which has the best performance for the fitness of the regression model compared with other models we used according to the cross-validation results (Supplemental Figure 6 and Supplemental Tables 5–7). However, the prediction ability was still weak for chalky grain rate when only the genetic data of the significant associations were used (Supplemental Figure 7).

In summary, we performed genome-wide high-resolution mapping for the traits of grain shape and grain chalkiness in hybrid rice using multiple collaborative populations for joint analyses. The segregation genetic populations had three genotypes present with expected proportions (i.e., 1:2:1 in these F_2 lines), facilitating the estimates of the genetic effects of the GWAS loci and the predictions of the grain appearance quality traits. The contents and components of starch, protein, and lipid in rice grains also strongly affect rice grain quality, especially cooking and eating quality. However, these traits cannot be measured using the F_2 populations because the seeds of a single plant were not enough for the commonly used chemical

assay (typically 30–40 g). In order to map and utilize gene alleles for more complex grain quality traits, in future it may be quite useful to develop some micro-determination methods for measuring grain quality and build multiple sets of backcross inbred lines, aiming to generate hybrid rice varieties with better performance for appearance quality, nutritional quality, and eating quality.

Junyi Gong[1,5], Jiashun Miao[2,3,5], Yan Zhao[2],
Qiang Zhao[2], Qi Feng[2], Qilin Zhan[2],
Benyi Cheng[1], Junhui Xia[1], Xuehui Huang[4,*],
Shihua Yang[1,*] and Bin Han[2,*]

[1]State Key Laboratory of Rice Biology, China National Rice Research Institute, Chinese Academy of Agricultural Sciences, Hangzhou 310006, China
[2]National Center for Gene Research, CAS Center for Excellence of Molecular Plant Sciences, Institute of Plant Physiology and Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200233, China
[3]University of Chinese Academy of Sciences, Beijing 100039, China
[4]College of Life and Environmental Sciences, Shanghai Normal University, Shanghai 200234, China
[5]These authors contributed equally to this article.
*Correspondence: Xuehui Huang (xhhuang@shnu.edu.cn), Shihua Yang (yangshihua6308@163.com), Bin Han (bhan@ncgr.ac.cn)
http://dx.doi.org/10.1016/j.molp.2017.07.014

REFERENCES

Duan, P., Xu, J., Zeng, D., Zhang, B., Geng, M., Zhang, G., Huang, K., Huang, L., Xu, R., Ge, S., et al. (2017). Natural variation in the promoter of GSE5 contributes to grain size diversity in rice. Mol. Plant **10**:685–694.

Fan, C., Xing, Y., Mao, H., Lu, T., Han, B., Xu, C., Li, X., and 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.

Li, Y., Fan, C., Xing, Y., Yun, P., Luo, L., Yan, B., Peng, B., Xie, W., Wang, G., Li, X., et al. (2014). Chalk5 encodes a vacuolar H(+)-translocating pyrophosphatase influencing grain chalkiness in rice. Nat. Genet. **46**:398–404.

Liu, J., Chen, J., Zheng, X., Wu, F., Lin, Q., Heng, Y., Tian, P., Cheng, Z., Yu, X., Zhou, K., et al. (2017). GW5 acts in the brassinosteroid signalling pathway to regulate grain width and weight in rice. Nat. Plants **3**:17043.

Myles, S., Peiffer, J., Brown, P.J., Ersoz, E.S., Zhang, Z., Costich, D.E., and Buckler, E.S. (2009). Association mapping: critical

considerations shift from genotyping to experimental design. Plant Cell **21**:2194–2202.

Nordborg, M., and Weigel, D. (2008). Next-generation genetics in plants. Nature **456**:720–723.

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

Wang, S., Li, S., Liu, Q., Wu, K., Zhang, J., Wang, S., Wang, Y., Chen, X., Zhang, Y., Gao, C., et al. (2015a). The OsSPL16-GW7 regulatory module determines grain shape and simultaneously improves rice yield and grain quality. Nat. Genet. **47**:949–955.

Wang, Y., Xiong, G., Hu, J., Jiang, L., Yu, H., Xu, J., Fang, Y., Zeng, L., Xu, E., Xu, J., et al. (2015b). Copy number variation at the GL7 locus contributes to grain size diversity in rice. Nat. Genet. **47**:944–948.