

# Natural Variation in the Promoter of *GSE5* Contributes to Grain Size Diversity in Rice

Penggen Duan<sup>1,5</sup>, Jinsong Xu<sup>1,5</sup>, Dali Zeng<sup>2</sup>, Baolan Zhang<sup>1</sup>, Mufan Geng<sup>3</sup>, Guozheng Zhang<sup>1,4</sup>, Ke Huang<sup>1,4</sup>, Luojiang Huang<sup>1,4</sup>, Ran Xu<sup>1</sup>, Song Ge<sup>3</sup>, Qian Qian<sup>2</sup> and Yunhai Li<sup>1,\*</sup>

<sup>1</sup>State Key Laboratory of Plant Cell and Chromosome Engineering, CAS Centre for Excellence in Molecular Plant Biology, Institute of Genetics and Developmental Biology, Chinese Academy of Sciences, Beijing 100101, P. R. China

<sup>2</sup>State Key Laboratory of Rice Biology, China National Rice Research Institute, Hangzhou 310006, China

<sup>3</sup>State Key Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, P. R. China

<sup>4</sup>University of Chinese Academy of Sciences, Beijing 100039, P. R. China

<sup>5</sup>These authors contributed equally to this article.

\*Correspondence: Yunhai Li (yhli@genetics.ac.cn)

<http://dx.doi.org/10.1016/j.molp.2017.03.009>

## ABSTRACT

The utilization of natural genetic variation greatly contributes to improvement of important agronomic traits in crops. Understanding the genetic basis for natural variation of grain size can help breeders develop high-yield rice varieties. In this study, we identify a previously unrecognized gene, named *GSE5*, in the *qSW5/GW5* locus controlling rice grain size by combining the genome-wide association study with functional analyses. *GSE5* encodes a plasma membrane-associated protein with IQ domains, which interacts with the rice calmodulin protein, OsCaM1-1. We found that loss of *GSE5* function caused wide and heavy grains, while overexpression of *GSE5* resulted in narrow grains. We showed that *GSE5* regulates grain size predominantly by influencing cell proliferation in spikelet hulls. Three major haplotypes of *GSE5* (*GSE5*, *GSE5<sup>DEL1+IN1</sup>*, and *GSE5<sup>DEL2</sup>*) in cultivated rice were identified based on the deletion/insertion type in its promoter region. We demonstrated that a 950-bp deletion (DEL1) in *indica* varieties carrying the *GSE5<sup>DEL1+IN1</sup>* haplotype and a 1212-bp deletion (DEL2) in *japonica* varieties carrying the *GSE5<sup>DEL2</sup>* haplotype associated with decreased expression of *GSE5*, resulting in wide grains. Further analyses indicate that wild rice accessions contain all three haplotypes of *GSE5*, suggesting that the *GSE5* haplotypes present in cultivated rice are likely to have originated from different wild rice accessions during rice domestication. Taken together, our results indicate that the previously unrecognized *GSE5* gene in the *qSW5/GW5* locus, which is widely utilized by rice breeders, controls grain size, and reveal that natural variation in the promoter region of *GSE5* contributes to grain size diversity in rice.

**Key words:** rice, natural variation, *GSE5*, grain size, cell proliferation

Duan P., Xu J., Zeng D., Zhang B., Geng M., Zhang G., Huang K., Huang L., Xu R., Ge S., Qian Q., and Li Y. (2017). Natural Variation in the Promoter of *GSE5* Contributes to Grain Size Diversity in Rice. *Mol. Plant*. 10, 685–694.

## INTRODUCTION

Modern agriculture must meet the challenges of feeding an increasing population and decreasing area of arable land. Rice is a crop of major importance, providing food for more than half of the global population. Genetic variation in diverse rice varieties provides valuable resources for improvement of important agronomic traits in rice. Rice breeders have explored natural variation in genes involved in the regulation of yield-related traits to develop elite rice varieties (Zuo and Li, 2014). Rice grain yield is determined by grain weight, grain number per panicle, and panicle number per plant. Grain size is associated with grain weight, grain yield,

and quality of appearance. Several quantitative trait loci (QTL) genes for grain size have been identified in rice (Fan et al., 2006; Song et al., 2007; Shomura et al., 2008; Weng et al., 2008; Li et al., 2011; Qi et al., 2012; Wang et al., 2012; Zhang et al., 2012; Ishimaru et al., 2013; Che et al., 2015; Duan et al., 2015; Hu et al., 2015; Wang et al., 2015a, 2015b; Si et al., 2016), but only a few beneficial alleles are widely utilized by rice breeders (Zuo and Li, 2014; Li and Li, 2016).

Published by the Molecular Plant Shanghai Editorial Office in association with Cell Press, an imprint of Elsevier Inc., on behalf of CSPB and IPPE, SIBS, CAS.

## Molecular Plant

Asian cultivated rice includes *indica* and *japonica* subspecies, which show large variation in grain size and shape. Typical *indica* varieties produce long grains, whereas *japonica* varieties form round and short grains. Natural variation in several grain size genes has been reported to be selected by rice breeders. For example, natural variation in the major QTL for grain length (GS3) contributes to grain-length differences between *indica* varieties and *japonica* varieties (Fan et al., 2006; Mao et al., 2010). The *indica* varieties with long grains usually contain its loss-of-function allele, while *japonica* varieties with short grains often have the wild-type allele. By contrast, natural variation in the major QTL for grain width (*qSW5/GW5*) determines grain-width differences between *indica* and *japonica* varieties. Previous studies reported that the *qSW5/GW5* gene encodes an unknown protein that interacts with ubiquitin (Shomura et al., 2008; Weng et al., 2008). The 1212-bp deletion in most *japonica* varieties disrupts the *qSW5* gene, resulting in wide grains. By contrast, *indica* varieties do not contain this 1212-bp deletion in the *qSW5* gene, thereby producing narrow grains (Weng et al., 2008). In addition, genome-wide association studies (GWAS) have identified multiple association signals for grain size in cultivated rice (Huang et al., 2010). The QTL gene *GLW7/OsSPL13* has been recently identified using the GWAS approach (Si et al., 2016). High expression of *GLW7* is associated with large grains in tropical *japonica* rice. However, the grain size genes underlying natural variation have not been fully explored in rice.

In this study, we report the identification of a previously unrecognized gene (*GSE5*) in the *qSW5/GW5* locus controlling rice grain size. *GSE5* encodes a plasma membrane-associated protein with IQ domains (IQD), which regulates grain width by restricting cell proliferation in spikelet hulls. Two major types of deletions, DEL1 and DEL2, present in the promoter region of *GSE5* in some *indica* varieties and most *japonica* varieties, respectively, caused decreased expression of *GSE5* and, thereby, wide grains in cultivated rice. Wild rice accessions contain both DEL1 and DEL2, suggesting that these two deletions in cultivated rice are likely to have originated from different wild rice accessions during rice domestication.

## RESULTS AND DISCUSSION

### Identification of the *GSE5*-Containing Locus by GWAS Analysis

To identify natural variation in genes involved in grain size control, we performed a GWAS. We used 102 *indica* varieties (Supplemental Table 1), which showed large variation in grain size (Supplemental Figure 1). To detect nucleotide polymorphisms, we conducted whole-genome sequencing of these 102 *indica* varieties and obtained a total of 677.3 Gb of genomic sequence. The average sequencing depth is 15.4x, and 96.4% of the reference genome sequence is covered (International Rice Genome Sequencing Project, 2005). A total of 831050 single-nucleotide polymorphisms (SNPs) were detected among 102 *indica* varieties (Supplemental Figure 2). Based on these nucleotide polymorphisms, we conducted principal component analysis (PCA) to characterize the population structure of these 102 *indica* varieties. As shown in Supplemental Figure 3, these 102 *indica* varieties did not show a highly structured population. We then analyzed linkage disequilibrium (LD) for these 102 *indica* varieties using these SNPs. The average decay of LD was about 220 kb in

### Natural Variation in *GSE5* Controls Grain Size

this population ( $r^2 = 0.2$ ) (Supplemental Figure 4), similar to that reported in a previous study of rice (Huang et al., 2010).

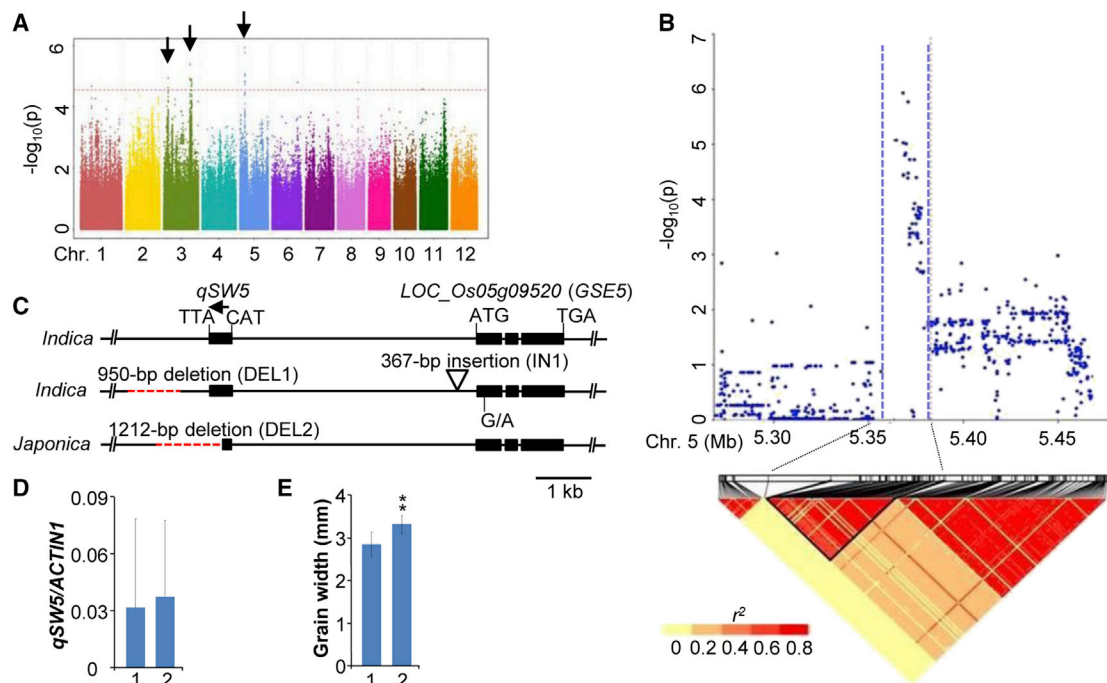
We then performed GWAS for grain width in this *indica* population using a mixed linear model with correction of kinship, a widely used method for GWAS analysis (Huang et al., 2010; Yano et al., 2016). As shown in Figure 1A, Supplemental Figure 5, and Supplemental Table 2, three loci were significantly associated with grain width. As this population structure is moderate and the population size is not large enough, it is possible that some other loci associated with grain width were insignificant in this population and thus were overlooked from our GWAS analysis. Surprisingly, one locus for grain width was found to be located in the region of *qSW5/GW5* on chromosome 5, which has been known to determine grain-width differences between *indica* and *japonica* varieties (Shomura et al., 2008; Weng et al., 2008). We analyzed the sequences of the *qSW5* locus in the *indica* varieties with wide grain and narrow grain, respectively. Most *indica* varieties with wide grain contained a 950-bp deletion (DEL1) in the 3' flanking region of *qSW5*, but most varieties with narrow grain had no such large deletion in this region (Figure 1C). As DEL1 was present in the 3' non-coding region of *qSW5*, *indica* varieties with wide grain still possessed the entire coding region of the *qSW5* gene (Figure 1C). If this DEL1 affects the function of *qSW5* in *indica* varieties, we presumed that it might decrease expression of *qSW5*. However, DEL1 was not associated with expression levels of *qSW5* in *indica* varieties (Figure 1D), suggesting that DEL1 might not affect the function of *qSW5*. Thus, it is unlikely that *qSW5* could be responsible for grain-width differences among these *indica* varieties. Importantly, DEL1 was strongly associated with grain width in *indica* varieties (Figure 1E), suggesting that the other gene in this locus could be responsible for grain-width variation in *indica* varieties. We therefore designated this gene GRAIN SIZE ON CHROMOSOME 5 (*GSE5*).

### Expression Level of *LOC\_Os05g09520* Is Associated with Grain Width

To identify the *GSE5* gene, we used pairwise LD correlations ( $r^2 > 0.6$ ) (Yano et al., 2016) to estimate a candidate region from 5.357 Mb to 5.379 Mb (22.42 kb) (Figure 1B). There are two genes within this 22.42-kb interval, *qSW5* and *LOC\_Os05g09520* (Figure 1B and 1C). This result suggests that *LOC\_Os05g09520* is a candidate gene for *GSE5*. We therefore sequenced the *LOC\_Os05g09520* gene in wide-grain and narrow-grain *indica* varieties, respectively. Although we found one SNP (G/A) in its coding region in wide-grain varieties, it did not cause an amino acid change (Figure 1C). We then selected 20 *indica* varieties with wide grain or narrow grain and examined expression levels of *LOC\_Os05g09520*. As shown in Figure 2A, expression levels of *LOC\_Os05g09520* were significantly associated with grain width. *LOC\_Os05g09520* showed lower expression in wide-grain *indica* varieties than in narrow-grain *indica* varieties, suggesting that decreased expression of *LOC\_Os05g09520* might cause wide grains.

### Both DEL1 in *Indica* Varieties and DEL2 in *Japonica* Varieties Decrease the Expression of *LOC\_Os05g09520*

To understand why the expression of *LOC\_Os05g09520* is decreased in wide-grain varieties, we examined the 5' flanking sequences of *LOC\_Os05g09520* in *indica* varieties and found that



**Figure 1. Identification of the *GSE5*-Containing Locus for Grain Size by GWAS.**

(A) Genome-wide association study of grain width. Manhattan plots for grain width. Dashed line represents the significance threshold ( $P = 2.78 \times 10^{-5}$ ). The arrows indicate the loci for grain width.

(B) Q local Manhattan plot (top) and LD heatmap (bottom) surrounding the peak on chromosome 5. Dashed lines indicate the candidate region for the peak.

(C) Schematic of the 22.42-kb genomic region. This region contains *qSW5* and *LOC\_Os05g09520* (*GSE5*). Most *japonica* varieties have a 1212-bp deletion (DEL2) in the *qSW5* gene. Most narrow-grain *indica* varieties have no deletion in *qSW5*, while most wide-grain *indica* varieties contain a 950-bp deletion (DEL1) in the 3' flanking region of *qSW5*, a 367-bp insertion (IN1) in the 5' flanking region of *LOC\_Os05g09520*, and a nucleotide change (G/A) in the first exon of *LOC\_Os05g09520*. The arrow shows the direction of the *qSW5* transcription. The red dashed lines represent the deletions in the genomic regions.

(D) Comparison of *qSW5* expression in young panicles of *indica* varieties without (1) or with (2) the 950-bp deletion (DEL1) in the 3' flanking region of *qSW5* ( $n = 34/36$ ).

(E) Correlation of the 950-bp deletion (DEL1) and 367-bp insertion (IN1) with grain width. Mature grains from the *indica* varieties without (1) or with DEL1 + IN1 (2) were measured ( $n = 68/65$ ).

Values in (D) and (E) are means  $\pm$  SD. Significance is determined using analysis of variance (ANOVA) (\*\* $P < 0.01$ ).

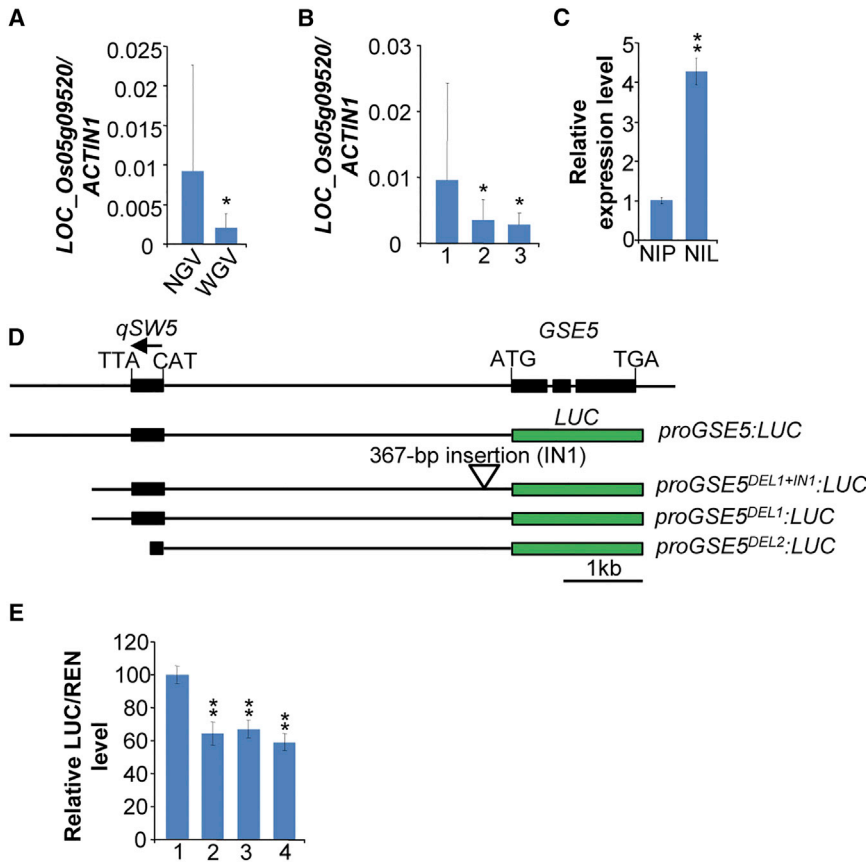
most wide-grain *indica* varieties contain a 950-bp deletion (DEL1) as well as a 367-bp insertion (IN1) (Figure 1C). Thus, it is possible that DEL1 and IN1 might cause the decreased expression of *LOC\_Os05g09520* in wide-grain *indica* varieties. As expected, DEL1 and IN1 negatively correlated with expression levels of *LOC\_Os05g09520* in *indica* varieties (Figure 2B).

The *japonica* varieties had a 1212-bp deletion (DEL2) that partially overlaps with DEL1 (Figure 1C) (Shomura et al., 2008; Weng et al., 2008). We therefore investigated whether DEL2 is also associated with expression levels of *LOC\_Os05g09520* in rice. As shown in Figure 2B, DEL2 was significantly associated with lower expression levels of *LOC\_Os05g09520*. To further investigate the effect of DEL2 on the expression of *LOC\_Os05g09520*, we obtained a near isogenic line (NIL), which contains the *LOC\_Os05g09520* allele from the narrow-grain *indica* variety 93-11 in the *japonica* variety Nipponbare background. Mature grains of NIL were narrower than those of Nipponbare with the deletion DEL2 (Figure 3I and 3J), indicating that DEL2 is linked to grain width. Expression of *LOC\_Os05g09520* in Nipponbare was significantly decreased compared with that in NIL (Figure 2C), further suggesting that DEL2 in *japonica* varieties might cause lower expression of *LOC\_Os05g09520*.

To determine whether DEL1 and IN1 in *indica* varieties and DEL2 in *japonica* varieties directly contribute to the reduced expression of *LOC\_Os05g09520*, we investigated the activity of the *GSE5* promoter (*proGSE5*) without or with DEL1 and IN1 (*proGSE5<sup>DEL1+IN1</sup>*), only DEL1 (*proGSE5<sup>DEL1</sup>*), and DEL2 (*proGSE5<sup>DEL2</sup>*), respectively (Figure 2D). We used these promoter variants to drive expression of the luciferase reporter gene (*LUC*). As shown in Figure 2E, the *proGSE5:LUC* had stronger luciferase activity than *proGSE5<sup>DEL1+IN1</sup>:LUC* and *proGSE5<sup>DEL2</sup>:LUC*, showing that DEL1 + IN1 and DEL2 decrease the promoter activity of *LOC\_Os05g09520*. The luciferase activity of *proGSE5<sup>DEL1+IN1</sup>:LUC* was similar to that of *proGSE5<sup>DEL1</sup>:LUC*, indicating that DEL1 decreases the promoter activity and that IN1 might not influence promoter activity. Thus, these results indicate that DEL1 in *indica* varieties and DEL2 in *japonica* varieties contribute to the decreased expression of *LOC\_Os05g09520*, respectively.

### Identification of *GSE5*

To confirm that *LOC\_Os05g09520* is the *GSE5* gene, we generated the loss-of-function mutant for *LOC\_Os05g09520* and performed a genetic complementation test. Zhonghua 11 (ZH11), a *japonica* variety, contains the deletion DEL2 in the promoter of



**Figure 2. Both *DEL1* in *Indica* Varieties and *DEL2* in *Japonica* Varieties Cause Decreased Expression of *GSE5*.**

**(A)** Comparison of *LOC\_Os05g09520* expression in young panicles of narrow grain (NGV) and wide grain (WGV) *indica* varieties. Values are means ± SD (n = 20/20). Significance is determined using ANOVA (\**P* < 0.05).

**(B)** Comparison of *LOC\_Os05g09520* expression in young panicles of rice varieties without (1) or with *DEL1* + *IN1* (2) and *DEL2* (3). Values are means ± SD (n = 34/36/31). Significance is determined using ANOVA (\**P* < 0.05).

**(C)** Expression levels of *LOC\_Os05g09520* expression in young panicles of the *japonica* variety Nipponbare (NIP) with *DEL2* and its near isogenic line (NIL). NIL contains the *LOC\_Os05g09520* allele from the narrow-grain *indica* variety 93-11 in the *japonica* variety Nipponbare background. Values are means ± SE (n = 3). Significance is determined using *t*-test (\*\**P* < 0.01).

**(D)** The constructs for each of the promoter-luciferase (LUC) fusions are shown. The arrow shows the direction of the *qSW5* transcription.

**(E)** Effects of *DEL1*, *IN1*, and *DEL2* on the activity of the *GSE5* promoter. *N. benthamiana* leaves were transformed by injection of *Agrobacterium* GV3101 cells harboring *proGSE5:LUC* (1), *proGSE5<sup>DEL1+IN1</sup>:LUC* (2), *proGSE5<sup>DEL1</sup>:LUC* (3), and *proGSE5<sup>DEL2</sup>:LUC* (4) plasmids, respectively. Relative reporter activity (LUC/REN) was calculated, and the value for *proGSE5:LUC* was set at 100. Values are means ± SE (n = 3). Significance is determined using *t*-test (\*\**P* < 0.01).

*LOC\_Os05g09520*. Although the ZH11 promoter (*proGSE5<sup>DEL2</sup>*) showed reduced activity compared with *proGSE5*, it still possessed transcriptional activity (Figure 2E). We therefore presumed that further disruption of the *LOC\_Os05g09520* gene using CRISPR/Cas9 technology could increase the width of ZH11 grains. A mutant of *LOC\_Os05g09520* generated by CRISPR/Cas9, named *gse5-cr*, had a 1-bp deletion in the first exon of *LOC\_Os05g09520*, resulting in a reading frame shift (Figure 3A). As expected, *gse5-cr* produced wider grains than ZH11 (Figure 3B and 3C). The length of *gse5-cr* grains was similar to that of ZH11 grains (Figure 3D). The 1000-grain weight of *gse5-cr* was significantly increased compared with that of ZH11 (Figure 3E). We then expressed the *LOC\_Os05g09520* gene driven by an *Actin* promoter (*proActin:GSE5*) in ZH11 background. Transgenic plants produced narrower grains than ZH11 (Figure 3F–3H), indicating that expression of *LOC\_Os05g09520* complemented the wide-grain phenotype of ZH11. We also observed that transgenic plants had long grains compared with ZH11. We further examined the grain size of a near isogenic line (NIL), which contains the *GSE5* locus from the narrow-grain *indica* variety 93-11 in the *japonica* variety Nipponbare background. This NIL also showed narrower and longer grains than Nipponbare (Figure 3I–3K), similar to those observed in *proActin:GSE5* transgenic lines. Taken together, these results indicate that *GSE5* is the *LOC\_Os05g09520* gene.

### *GSE5* Regulates Cell Proliferation in Spikelet Hulls

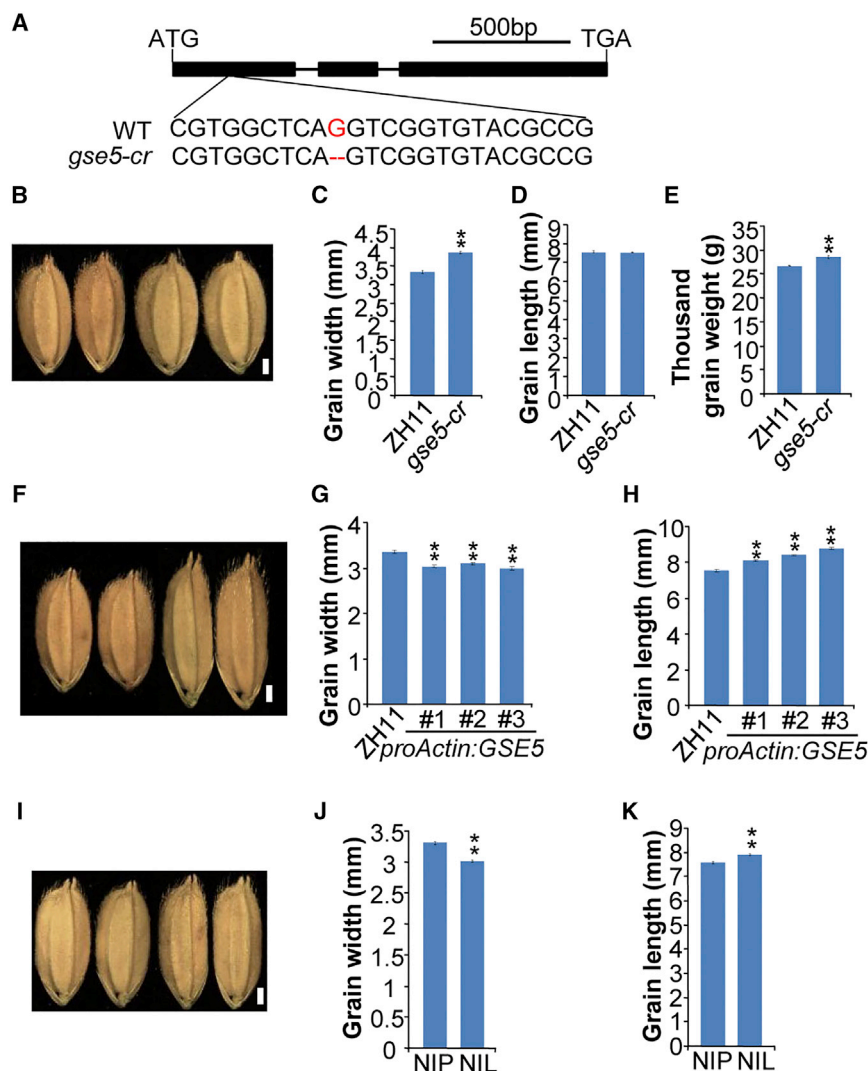
The spikelet hull restricts the growth of a grain, which has been proposed to influence grain size in rice (Li and Li, 2016). Cell

proliferation and cell expansion coordinately determine the growth of spikelet hulls. We therefore examined cell number and cell size in ZH11 and *gse5-cr* spikelet hulls. The *gse5-cr* spikelet hulls contained more epidermal cells than those of ZH11 in the grain-width direction (Figure 4A, 4B, and 4D), indicating that *GSE5* controls grain width by limiting cell proliferation. By contrast, epidermal cells in *gse5-cr* spikelet hulls were narrower than those in ZH11 spikelet hulls (Figure 4C), suggesting a possible compensation mechanism between cell proliferation and cell expansion. This compensation phenomenon was also found in several *Arabidopsis* seed size mutants (Xia et al., 2013).

We then investigated cell number and cell size in spikelet hulls of ZH11 and *proActin:GSE5* plants. As shown in Figure 4E–4H, the spikelet hulls of *proActin:GSE5* plants had fewer cells in the grain-width direction and more cells in the grain-length direction than those of ZH11, whereas epidermal cell length and width in *proActin:GSE5* spikelet hulls were similar to those in ZH11, consistent with the narrow- and long-grain phenotypes of *proActin:GSE5* transgenic plants. These results indicate that *GSE5* controls rice grain size predominantly by influencing cell proliferation in spikelet hulls.

### *GSE5* Encodes a Plasma Membrane-Associated Protein with IQ Domains

Sequence analysis showed that *GSE5* encodes a predicted protein with IQ domains (IQD) (Figure 5A). IQD proteins are an ancient family of calmodulin-binding proteins that regulate



**Figure 3. Identification of *GSE5*.**

**(A)** The *gse5-cr* mutant was generated by CRISPR/Cas9. In *gse5-cr* mutant, the 1-bp deletion happens in the first exon of *GSE5*, resulting in a reading frame shift.

**(B)** Grains of Zhonghua 11 (ZH11) (left) and *gse5-cr* (right).

**(C–E)** Grain width **(C)**, grain length **(D)**, and 1000-grain weight **(E)** of Zhonghua 11 (ZH11) and *gse5-cr*.

**(F)** Grains of Zhonghua 11 (ZH11) (left) and *proActin:GSE5* (right). *GSE5* was overexpressed in ZH11 background.

**(G and H)** Grain width **(G)** and grain length **(H)** of Zhonghua 11 (ZH11) and *proActin:GSE5*. *GSE5* was overexpressed in ZH11 background.

**(I)** Grains of Nipponbare (NIP) (left) and a near isogenic line (NIL) (right), which contains the *GSE5* locus from the narrow grain *indica* variety 93-11 in the *japonica* variety Nipponbare background.

**(J and K)** Grain width **(J)** and grain length **(K)** of Nipponbare (NIP) and a near isogenic line (NIL), which contains the *GSE5* locus from the narrow grain *indica* variety 93-11 in the *japonica* variety Nipponbare background.

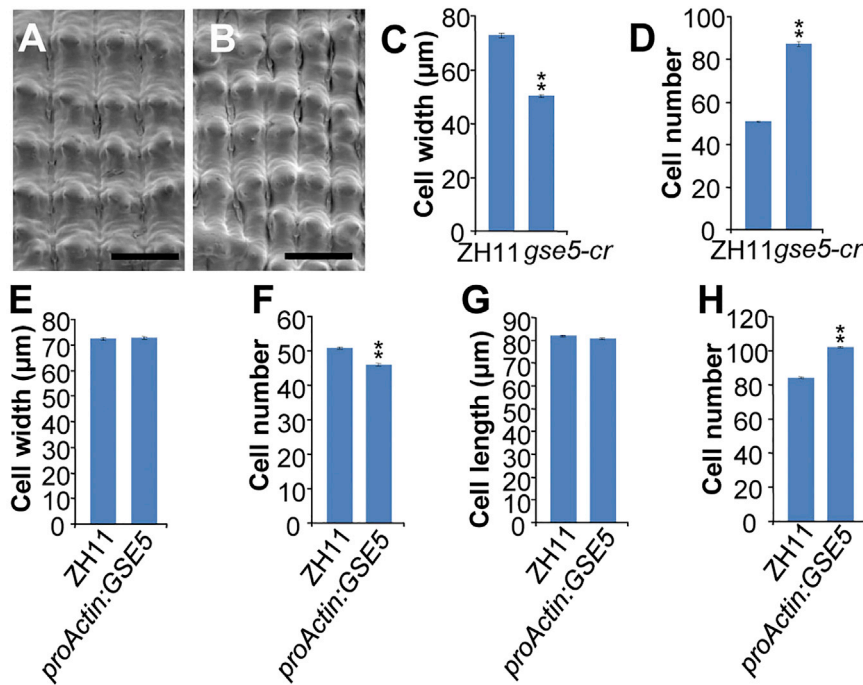
Values in **(C)** to **(E)**, **(G)**, **(H)**, **(J)**, and **(K)** are means ± SE. Significance is determined using *t*-test (\*\**P* < 0.01). Scale bars in **(B)**, **(F)**, and **(I)** represent 1 mm.

plant stress responses and plant development (Abel et al., 2005; Xiao et al., 2008). We therefore asked whether *GSE5* could interact with rice calmodulin. As shown in Figure 5B, bimolecular fluorescence complementation assays showed that *GSE5* physically associated with a rice calmodulin, OsCaM1-1, *in vivo*. Thus, it is possible that *GSE5* is involved in calcium signaling to regulate grain size in rice. In plants, how calcium signaling is involved in seed size control is totally unknown. Identification of *GSE5* provides a good starting point for future studies on the role of calcium signaling in seed size control. In addition, genes whose products share significant sequence homology with *GSE5* are found in diverse plant species such as rice, wheat, maize, soybean, and sorghum, but not animals (Supplemental Figure 6), suggesting that *GSE5* homologs may play a conserved role in controlling seed size in plants.

The transcripts of *GSE5* could be detected in developing panicles using quantitative real-time RT-PCR analysis (Figure 5C). To examine its tissue-specific expression patterns, we also generated the *GSE5* promoter:*GSE5-GUS* fusion (*proGSE5*:

but disappeared at the late stages of panicle and grain development (Figure 5D–5H). The expression patterns of *GSE5* are consistent with its role in cell proliferation.

To determine the subcellular localization of *GSE5*, we expressed a *GSE5-GFP* fusion protein under its own promoter (*proGSE5*:*GSE5-GFP*) in the *japonica* variety ZH11. As the *proGSE5*:*GSE5-GFP* transgenic plants produced narrow grains compared with ZH11 (Supplemental Figure 7), we presumed that *GSE5-GFP* is a functional fusion protein. We observed GFP fluorescence in the *proGSE5*:*GSE5-GFP* transgenic plants predominantly in the cell periphery (Figure 5I). We further used the plasmolysis induced by a high sucrose level to determine whether *GSE5-GFP* is associated with the plasma membrane or cell walls. The result showed that *GSE5-GFP* was detected in the shrunken plasma membrane (Figure 5J). Extensive sequence analyses and predications failed to identify any predicted transmembrane domain in *GSE5*, suggesting that *GSE5* is not a canonical transmembrane protein and may be associated with the plasma membrane by a non-conventional mechanism or by associating with other membrane-localized proteins.



**Figure 4. *GSE5* Regulates Cell Proliferation in Spikelet Hulls.**

(A and B) The outer epidermal surface of ZH11 (A) and *gse5-cr* (B).

(C and D) The outer epidermal cell width (C) and the calculated outer epidermal cell number (D) of ZH11 and *gse5-cr* lemma in the grain-width direction.

(E and F) The outer epidermal cell width (E) and the calculated outer epidermal cell number (F) of ZH11 and *proActin:GSE5* lemma in the grain-width direction.

(G and H) The outer epidermal cell length (G) and the calculated outer epidermal cell number (H) of ZH11 and *proActin:GSE5* lemma in the grain-length direction.

Values in (C) to (H) are means  $\pm$  SE. Significance is determined using *t*-test (\*\**P* < 0.01). Scale bars in (A) and (B) represent 100  $\mu$ m.

likely to have originated from different *O. rufipogon* accessions during rice domestication.

### Evolutionary Analysis of the *GSE5* Locus

Based on the deletion/insertion type in the promoter of *GSE5*, we identified three major haplotypes (*GSE5*, *GSE5<sup>DEL1+IN1</sup>*, and *GSE5<sup>DEL2</sup>*) in cultivated rice (Figure 1C). As both DEL1 in *indica* varieties carrying the *GSE5<sup>DEL1+IN1</sup>* haplotype and DEL2 in *japonica* varieties carrying the *GSE5<sup>DEL2</sup>* haplotype contribute to wide grains, we genotyped 141 *indica* and 91 *japonica* varieties of cultivated rice. Among 141 *indica* varieties, 48.2%, 46.1%, and 5.7% of them were *GSE5*, *GSE5<sup>DEL1+IN1</sup>*, and *GSE5<sup>DEL2</sup>* haplotypes, respectively (Figure 6A). By contrast, among 91 *japonica* varieties, 11%, 7.7%, and 81.3% of them contained *GSE5*, *GSE5<sup>DEL1+IN1</sup>*, and *GSE5<sup>DEL2</sup>* haplotypes, respectively (Figure 6B). These results indicate that both DEL1 in *indica* varieties and DEL2 in *japonica* varieties were widely utilized by rice breeders.

Cultivated rice has been proposed to be domesticated from wild rice (*Oryza rufipogon*). We therefore asked whether wild rice accessions could contain these two deletions (DEL1 and DEL2) in the promoter region of *GSE5*. We genotyped 41 wild rice accessions (*O. rufipogon*) (Supplemental Table 3) and observed that five accessions contained the *GSE5<sup>DEL1+IN1</sup>* haplotype; one from Hunan province in southern of China had the *GSE5<sup>DEL2</sup>* haplotype, whereas the majority of accessions had the *GSE5* haplotype (Figure 6C). This result suggests that both DEL1 and DEL2 deletions might have occurred before the domestication of cultivated rice. We further sequenced ~8.4-kb sequences including 6320-bp 5' flanking sequence, the *GSE5* gene, and 531-bp 3' flanking sequence from 63 cultivated rice and 26 *O. rufipogon* accessions. Phylogenetic analyses showed that several wild rice accessions were clustered together with cultivated rice varieties carrying *GSE5*, *GSE5<sup>DEL1+IN1</sup>*, or *GSE5<sup>DEL2</sup>* haplotypes, respectively (Figure 6D). These results further suggest that the *GSE5*, *GSE5<sup>DEL1+IN1</sup>* and *GSE5<sup>DEL2</sup>* haplotypes in cultivated rice are

Natural variation in genes involved in the regulation of important agronomic traits has been utilized by breeders to improve crop yield and quality. Grain size is associated with grain weight, thus influencing grain yield in rice. Several QTL genes for grain size have been characterized in rice, but only a few of the beneficial alleles are widely used in rice production. In this study, we identify a previously unrecognized *GSE5* gene in the *qSW5/GW5* locus controlling grain size by a GWAS combined with functional analysis. *GSE5*, which encodes a plasma membrane-associated IQD protein that interacts with the rice calmodulin (OsCaM1-1), controls grain width by restricting cell proliferation in spikelet hulls. We demonstrate that natural variation in the promoter of *GSE5* contributes to grain size diversity in cultivated rice. Both the DEL1 and DEL2 deletions are widely utilized by rice breeders and could have originated from different wild rice accessions during rice domestication. Thus, these results provide significant insights into the genetic basis for natural variation in rice grain size control, and suggest that *GSE5* and its homologs in other crops such as maize, wheat, and soybean may be used to improve the yield.

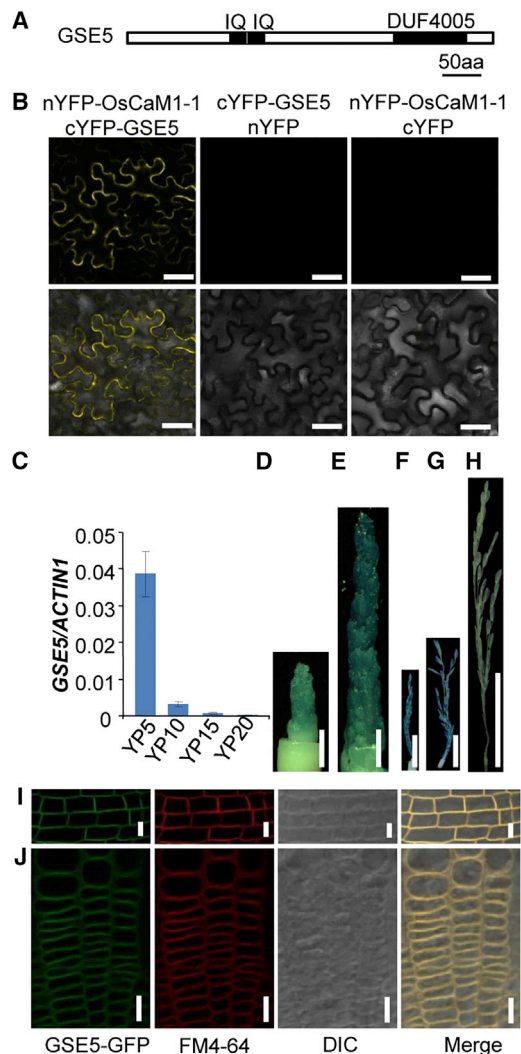
### METHODS

#### Plant Materials and Growth Conditions

The cultivated rice varieties were obtained from a collection of cultivated rice preserved at the China National Rice Research Institute. The common wild rice varieties (*O. rufipogon*) were obtained from the Institute of Botany, Chinese Academy of Sciences (Zhu et al., 2007; Zheng and Ge, 2010). The *indica* and *japonica* varieties used in this study were cultivated in the paddy fields at Hangzhou (China) and Hainan (China).

#### Morphological and Cellular Analyses

Grain size of the 102 *indica* varieties was measured using the SC Detection and Analysis System of Rice Seeds (Hangzhou WSeen Detection Technology). Dry grains of Zhonghua 11 (ZH11) and *gse5-cr* were weighed using an electronic analytical balance (Mettler Moledo AL104, China).



**Figure 5. *GSE5* Encodes a Plasma Membrane-Associated Protein with IQ Domains.**

**(A)** The *GSE5* protein contains two IQ motifs and an unknown DUF4005 domain.

**(B)** Bimolecular fluorescence complementation assays show that *GSE5* associated with OsCaM1-1 in *N. benthamiana*. nYFP-OsCaM1-1 and cYFP-*GSE5* were coexpressed in leaves of *N. benthamiana*.

**(C)** Quantitative real-time RT-PCR analysis of *GSE5* expression in young panicles of 5 cm (YP5), 10 cm (YP10), 15 cm (YP15), and 20 cm (YP20). Values are mean ± SE (n = 3).

**(D–H)** *GSE5* expression activity was monitored using *proGSE5*:*GSE5*-*GUS* transgenic plants. GUS activity was detected in developing panicles.

**(I)** Subcellular localization of *GSE5*-GFP in *proGSE5*:*GSE5*-GFP transgenic plants. GFP fluorescence in *proGSE5*:*GSE5*-GFP transgenic plants was detected in the cell periphery. FM4-64 was used to stain the membrane.

**(J)** Cells were plasmolyzed with 30% sucrose. *GSE5*-GFP was detected in the shrunken plasma membrane. FM4-64 was used to stain the membrane.

Scale bars represent 50 μm in **(B)**, 1 mm in **(D)** and **(E)**, 1 cm in **(F)** and **(G)**, 5 cm in **(H)**, and 10 μm in **(I)** and **(J)**.

For observation of cell size and cell number, grain hulls of Zhonghua 11 (ZH11), *gse5-cr*, and *proActin*:*GSE5* transgenic plants were sputter-coated with platinum and observed using a scanning electron microscope

(Hitachi S-3000N). ImageJ software was used to measure the size of epidermal cells.

### DNA Isolation, Genome Sequencing, and Sequence Analysis

NuClean PlantGen DNA kits (CWBI, China) were used for the genomic DNA extraction. For each cultivated rice variety, a single individual was used for genome sequencing on the Illumina Hiseq 2500. Library construction and sample indexing were performed as described previously (Huang et al., 2009). The libraries were loaded into the Illumina Hiseq 2500 for 100-bp paired-end sequencing. Image analysis and base calling were conducted using the Illumina Genome Analyzer processing pipeline (v1.4). PERL scripts in the SEG-Map pipeline were used to sort raw sequences on the basis of the 5' indexes.

A total of  $6.773 \times 10^9$  paired-end 100-bp reads were obtained for the cultivated accessions. Firstly, quality control was performed, and the average Q30 was 89.94%, which means that the reads were reliable. The reads were then aligned to Os-Nipponbare-Reference-MSU7.0 pseudomolecules using bwa-mem with the -M option of BWA software (Li and Durbin, 2010). The mapped reads were realigned using RealignerTargetCreator and indelRealigner of GATK software (DePristo et al., 2011). To label SNPs, we used UnifiedGenotyper of GATK with the -glm BOTH option. All nucleotide polymorphisms were analyzed according to their location in the reference genome.

### Population Genetic Analyses

The population structure of the 102 *indica* varieties (PCA) was estimated using the software PLINK version 1.9 (<http://pngu.mgh.harvard.edu/~purcell/plink/>). The LD between SNPs in the 102 varieties was evaluated using squared Pearson's correlation coefficient ( $r^2$ ) as calculated with the -r<sup>2</sup> command in the software PLINK version 1.9. The LD heatmaps surrounding peaks in the GWAS were constructed using the R package "LD heatmap" (Shin et al., 2006). We estimated the candidate regions using an  $r^2 > 0.6$  (Yano et al., 2016).

### Genome-Wide Association Study

The population structure (Q) was inferred using Admixture (Alexander et al., 2009), and the best one was selected when cross-validation error was minimum. The relative kinship matrix (K) of the natural population was calculated using TASSEL 5.2.1 (Bradbury et al., 2007). GWAS was performed using the Q + K model in TASSEL 5.2.1. The genome-wide significance threshold was determined using permutation-based false discovery rate-adjusted *P* values (Dudbridge and Gusnanto, 2008). The permutation tests were repeated 1000 times.

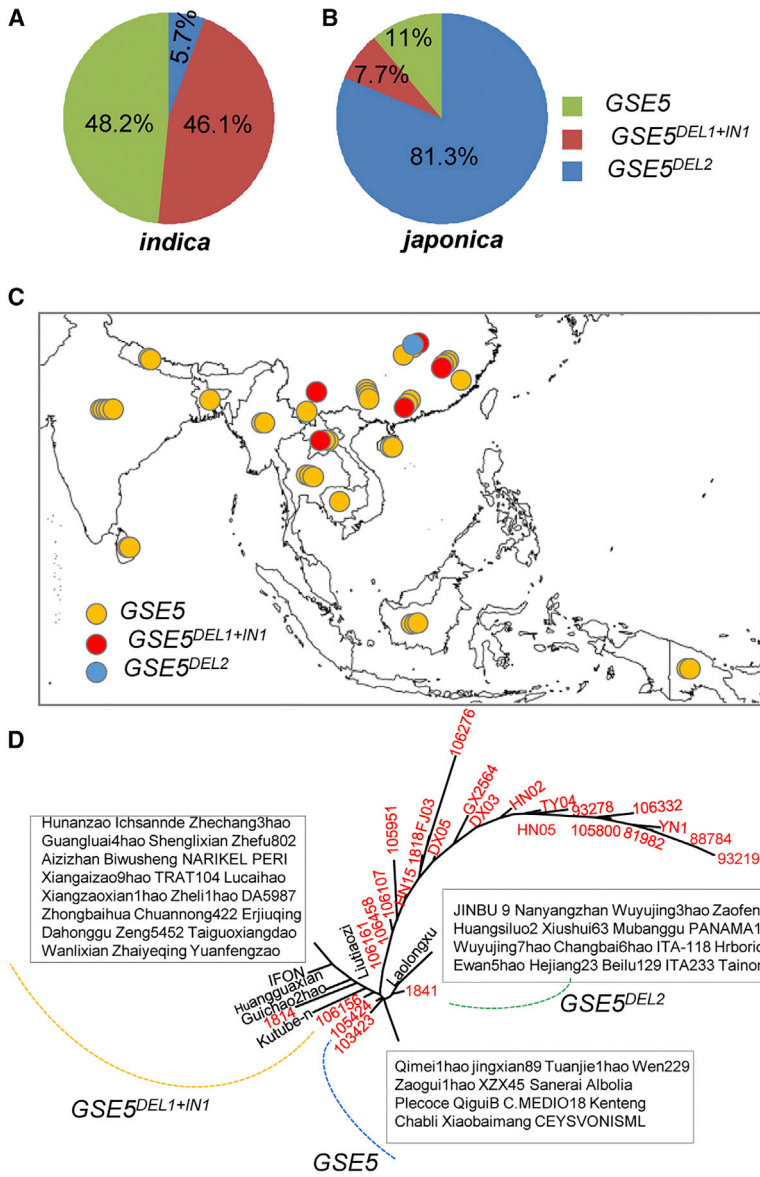
### Plasmid Construction and Plant Transformation

The 7897-bp *GSE5* genomic sequence was amplified from the *indica* variety 93-11 using the primers gGUS-F/R and gGFP-F/R and cloned into the *pMDC164* and *pMDC107* vectors using in-fusion enzyme (Genebank Biosciences, China), respectively. The coding sequence of *GSE5* was amplified by the specific primers cGSE5-F/R and cloned into the *pIpkb003* vector using in-fusion enzyme (Genebank Biosciences) to generate *proActin*:*GSE5* plasmid. The 488-bp sequence was amplified from the PCR products of crGSE5-1 and crGSE5-2 using the primers crGSE5-1F and crGSE5-2R, and cloned into the vector *pMDC99-Cas9* using in-fusion enzyme (Genebank Biosciences) to generate the *CRISPR/Cas9-gse5* plasmid. The plasmids were introduced into *Agrobacterium tumefaciens* strain GV3101 by electroporation, and rice transformation was carried out according to a previously published method (Hiei et al., 1994).

### GUS Staining and GFP Fluorescence Observations

The developing panicles of *proGSE5*:*GSE5*-*GUS* transgenic plants were stained in a GUS buffer according to the method described previously (Wang et al., 2016). The roots of *proGSE5*:*GSE5*-GFP transgenic plants were used to investigate the subcellular localization of *GSE5*. Plasma

## Molecular Plant



## Natural Variation in GSE5 Controls Grain Size

### Figure 6. Evolutionary Analysis of the GSE5 Locus.

(A and B) The percentages of GSE5, GSE5<sup>DEL1+IN1</sup>, and GSE5<sup>DEL2</sup> haplotypes in *indica* (A) and *japonica* (B) varieties. A total of 141 *indica* varieties and 91 *japonica* varieties were genotyped.

(C) Geographical origin of wild rice accessions used in this study. Wild rice accessions (*O. rufipogon*) contained GSE5, GSE5<sup>DEL1+IN1</sup>, and GSE5<sup>DEL2</sup> haplotypes.

(D) Phylogenetic tree. The approximate 8.4-kb sequences including 6320-bp 5' flanking sequence, the GSE5 gene, and 531-bp 3' flanking sequence from 63 cultivated rice varieties with GSE5, GSE5<sup>DEL1+IN1</sup>, and GSE5<sup>DEL2</sup> haplotypes and 26 *O. rufipogon* with GSE5, GSE5<sup>DEL1+IN1</sup>, and GSE5<sup>DEL2</sup> haplotypes were used to construct the phylogenetic tree. Bootstrap values over 60% are given on the branches. The red letters represent *O. rufipogon* accession.

with SuperScript III Reverse Transcriptase (Invitrogen). A Lightcycler 480 machine (Roche) was used to conduct quantitative real-time PCR. Relative amounts of *qSW5* and GSE5 were calculated using the comparative threshold (Wang et al., 2016). The primers for quantitative real-time RT-PCR are shown in Supplemental Table 4.

### Real-Time Detection of Promoter Activation

The promoter sequences of 6320 bp, 5310 bp, and 4547 bp were amplified from *indica* variety 93-11 genomic DNA using the specific primers of pLUC1-F/R, pLUCM-F/R, and pLUCS-F/R, and constructed into the vector *pGreenII0800-LUC* (Hellens et al., 2005) to generate *proGSE5:LUC*, *proGSE5<sup>DEL1</sup>:LUC*, and *proGSE5<sup>DEL2</sup>:LUC* plasmids, respectively. For *proGSE5<sup>DEL1+IN1</sup>:LUC* construction, the 5677-bp PCR fragment was amplified from *indica* variety Zhefu802 using the specific primers pLUCM-F/R and cloned into the vector *pGreenII0800-LUC* using in-fusion

enzyme (Genebank Biosciences). The plasmids were transferred into the *A. tumefaciens* strain GV3101 by electroporation and coinfiltrated into *N. benthamiana* leaves. Firefly and Renilla luciferase activities were measured using a Dual-Luciferase Reporter Assay System (Promega).

### Phylogenetic Analysis

For analysis of the evolutionary history, the approximate 8.4-kb genomic fragments including 6320-bp 5' flanking sequence, the GSE5 gene, and 531-bp 3' flanking sequence from 63 cultivated rice and 26 wild rice (*O. rufipogon*) varieties were amplified and sequenced. The DNA sequences were aligned using the CLUSTAL X 2.1 program. The evolutionary history was inferred using the neighbor-joining method with the MEGA7.0 program.

### SUPPLEMENTAL INFORMATION

Supplemental Information is available at *Molecular Plant Online*.

### FUNDING

This work was supported by grants from the Ministry of Science and Technology of China (2013CBA01401; 2016YFD0100402; 2016YFD0100501), Chinese Academy of Sciences (XDA08020108), the National Natural

membranes were stained using FM4-64 (5 µg/mL), and samples were observed under a Zeiss LSM 710 NLO confocal microscope.

### Bimolecular Fluorescence Complementation Assay

The coding sequence of GSE5 were amplified by specific primers ycGSE5-F/R, fused with the C-terminal fragment of YFP (cYFP), and subcloned into the *pGWB414* vector (Invitrogen) using in-fusion enzyme (Genebank Biosciences). The N-terminal fragment of YFP (nYFP) was amplified from *pSY736* using the primers YN-736-F and YN-736-R, fused with the *OsCaM1-1* gene, and subcloned into the *pGWB414* vector (Invitrogen) using in-fusion enzyme (Genebank Biosciences). *nYFP-OsCaM1-1* and *cYFP-GSE5* constructs were transformed into *Agrobacterium* strain GV3101. Transient expression of *nYFP-OsCaM1-1* and *cYFP-GSE5* in *Nicotiana benthamiana* leaves and fluorescence observation were conducted as described previously (Wang et al., 2016).

### RT-PCR and Quantitative Real-Time PCR

Developing panicles were used to extract total RNA using an RNAprep pure Plant Kit (Tiangen, China). Total RNA was used for cDNA synthesis



Science Foundation of China (91535203; 31425004), and the Ministry of Agriculture of China (2014ZX08009-003).

### AUTHOR CONTRIBUTIONS

P.D., J.X., and Y.L. designed the research. P.D., J.X., D.Z., B.Z., M.G., G.Z., K.H., L.H., and R.X. performed experiments. P.D., J.X., S.G., Q.Q., and Y.L. analyzed data. P.D., J.X., and Y.L. wrote the paper.

### ACKNOWLEDGMENTS

We would like to thank Dr. Qiaoquan Liu for the NIL line and the National Mid-term Genebank for Rice of the China National Rice Research Institute for providing the rice germplasm collection. No conflict of interest declared.

Received: March 11, 2017

Revised: March 23, 2017

Accepted: March 24, 2017

Published: March 30, 2017

### REFERENCES

- Abel, S., Savchenko, T., and Levy, M. (2005). Genome-wide comparative analysis of the IQD gene families in *Arabidopsis thaliana* and *Oryza sativa*. *BMC Evol. Biol.* **5**:72.
- Alexander, D.H., Novembre, J., and Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**:1655–1664.
- Bradbury, P.J., Zhang, Z., Kroon, D.E., Casstevens, T.M., Ramdoss, Y., and Buckler, E.S. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* **23**:2633–2635.
- Che, R., Tong, H., Shi, B., Liu, Y., Fang, S., Liu, D., Xiao, Y., Hu, B., Liu, L., Wang, H., et al. (2015). Control of grain size and rice yield by GL2-mediated brassinosteroid responses. *Nat. Plants* **2**:15195.
- DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**:491–498.
- Duan, P., Ni, S., Wang, J., Zhang, B., Xu, R., Wang, Y., Chen, H., Zhu, X., and Li, Y. (2015). Regulation of OsGRF4 by OsmiR396 controls grain size and yield in rice. *Nat. Plants* **2**:1.
- Dudbridge, F., and Gusnanto, A. (2008). Estimation of significance thresholds for genomewide association scans. *Genet. Epidemiol.* **32**:227–234.
- 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.
- Hellens, R.P., Allan, A.C., Friel, E.N., Bolitho, K., Grafton, K., Templeton, M.D., Karunairetnam, S., Gleave, A.P., and Laing, W.A. (2005). Transient expression vectors for functional genomics, quantification of promoter activity and RNA silencing in plants. *Plant Methods* **1**:13.
- Hiei, Y., Ohta, S., Komari, T., and Kumashiro, T. (1994). Efficient transformation of rice (*Oryza sativa* L.) mediated by *Agrobacterium* and sequence analysis of the boundaries of the T-DNA. *Plant J.* **6**:271–282.
- Hu, J., Wang, Y., Fang, Y., Zeng, L., Xu, J., Yu, H., Shi, Z., Pan, J., Zhang, D., Kang, S., et al. (2015). A rare allele of GS2 enhances grain size and grain yield in rice. *Mol. Plant* **8**:1455–1465.
- Huang, X., Feng, Q., Qian, Q., Zhao, Q., Wang, L., Wang, A., Guan, J., Fan, D., Weng, Q., Huang, T., et al. (2009). High-throughput genotyping by whole-genome resequencing. *Genome Res.* **19**:1068–1076.
- Huang, X., Wei, X., Sang, T., Zhao, Q., Feng, Q., Zhao, Y., Li, C., Zhu, C., Lu, T., Zhang, Z., et al. (2010). Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat. Genet.* **42**:961–967.
- International Rice Genome Sequencing Project. (2005). The map-based sequence of the rice genome. *Nature* **436**:793–800.
- Ishimaru, K., Hirotsu, N., Madoka, Y., Murakami, N., Hara, N., Onodera, H., Kashiwagi, T., Ujiie, K., Shimizu, B., Onishi, A., et al. (2013). Loss of function of the IAA-glucose hydrolase gene TGW6 enhances rice grain weight and increases yield. *Nat. Genet.* **45**:707–711.
- Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**:589–595.
- Li, N., and Li, Y. (2016). Signaling pathways of seed size control in plants. *Curr. Opin. Plant Biol.* **33**:23–32.
- Li, Y., Fan, C., Xing, Y., Jiang, Y., Luo, L., Sun, L., Shao, D., Xu, C., Li, X., Xiao, J., et al. (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., and Zhang, Q. (2010). Linking differential domain functions of the GS3 protein to natural variation of grain size in rice. *Proc. Natl. Acad. Sci. USA* **107**:19579–19584.
- Qi, P., Lin, Y.S., Song, X.J., Shen, J.B., Huang, W., Shan, J.X., Zhu, M.Z., Jiang, L., Gao, J.P., and Lin, H.X. (2012). The novel quantitative trait locus GL3.1 controls rice grain size and yield by regulating Cyclin-T1;3. *Cell Res.* **22**:1666–1680.
- Shin, J.-H., Blay, S., McNeney, B., and Graham, J. (2006). LDheatmap: an R function for graphical display of pairwise linkage disequilibria between single nucleotide polymorphisms. *J. Stat. Softw.* **16**, Code Snippet 3.
- 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.
- Si, L., Chen, J., Huang, X., Gong, H., Luo, J., Hou, Q., Zhou, T., Lu, T., Zhu, J., Shangguan, Y., et al. (2016). OsSPL13 controls grain size in cultivated rice. *Nat. Genet.* **48**:447–456.
- Song, X.J., Huang, W., Shi, M., Zhu, M.Z., and Lin, H.X. (2007). A QTL for rice grain width and weight encodes a previously unknown RING-type E3 ubiquitin ligase. *Nat. Genet.* **39**:623–630.
- Wang, S., Wu, K., Yuan, Q., Liu, X., Liu, Z., Lin, X., Zeng, R., Zhu, H., Dong, G., Qian, Q., et al. (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, Y., Chen, X., Zhang, Y., Gao, C., Wang, F., et al. (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., Ye, W., et al. (2015b). Copy number variation at the GL7 locus contributes to grain size diversity in rice. *Nat. Genet.* **47**:944–948.
- Wang, Z., Li, N., Jiang, S., Gonzalez, N., Huang, X., Wang, Y., Inze, D., and Li, Y. (2016). SCF<sup>SAP</sup> controls organ size by targeting PPD proteins for degradation in *Arabidopsis thaliana*. *Nat. Commun.* **7**:11192.
- Weng, J., Gu, S., Wan, X., Gao, H., Guo, T., Su, N., Lei, C., Zhang, X., Cheng, Z., Guo, X., et al. (2008). Isolation and initial characterization of GW5, a major QTL associated with rice grain width and weight. *Cell Res.* **18**:1199–1209.
- Xia, T., Li, N., Dumenuil, J., Li, J., Kamenski, A., Bevan, M.W., Gao, F., and Li, Y. (2013). The ubiquitin receptor DA1 interacts with the E3 ubiquitin ligase DA2 to regulate seed and organ size in *Arabidopsis*. *Plant Cell* **25**:3347–3359.

## Molecular Plant

**Xiao, H., Jiang, N., Schaffner, E., Stockinger, E.J., and van der Knaap, E.** (2008). A retrotransposon-mediated gene duplication underlies morphological variation of tomato fruit. *Science* **319**:1527–1530.

**Yano, K., Yamamoto, E., Aya, K., Takeuchi, H., Lo, P.C., Hu, L., Yamasaki, M., Yoshida, S., Kitano, H., Hirano, K., et al.** (2016). Genome-wide association study using whole-genome sequencing rapidly identifies new genes influencing agronomic traits in rice. *Nat. Genet.* **48**:927–934.

**Zhang, X., Wang, J., Huang, J., Lan, H., Wang, C., Yin, C., Wu, Y., Tang, H., Qian, Q., Li, J., et al.** (2012). Rare allele of OsPPKL1 associated with grain length causes extra-large grain and a

## Natural Variation in GSE5 Controls Grain Size

significant yield increase in rice. *Proc. Natl. Acad. Sci. USA* **109**:21534–21539.

**Zheng, X.M., and Ge, S.** (2010). Ecological divergence in the presence of gene flow in two closely related *Oryza* species (*Oryza rufipogon* and *O. nivara*). *Mol. Ecol.* **19**:2439–2454.

**Zhu, Q., Zheng, X., Luo, J., Gaut, B.S., and Ge, S.** (2007). Multilocus analysis of nucleotide variation of *Oryza sativa* and its wild relatives: severe bottleneck during domestication of rice. *Mol. Biol. Evol.* **24**:875–888.

**Zuo, J., and Li, J.** (2014). Molecular genetic dissection of quantitative trait loci regulating rice grain size. *Annu. Rev. Genet.* **48**:99–118.