


ORIGINAL ARTICLE

Open Access

Grain Size Selection Using Novel Functional Markers Targeting 14 Genes in Rice



Lin Zhang^{1,2*†}, Bin Ma^{3†}, Zhong Bian¹, Xiaoyuan Li³, Changquan Zhang², Jiyun Liu⁴, Qun Li⁴, Qiaoquan Liu^{1,2*}  and Zuhua He^{2,4}

Abstract

Background: Grain size is an extremely important aspect of rice breeding, affecting both grain yield and quality traits. It is controlled by multiple genes and tracking these genes in breeding schemes should expedite selection of lines with superior grain yield and quality, thus it is essential to develop robust, efficient markers.

Result: In this study, 14 genes related to grain size (*GW2*, *GS2*, *qLGY3*, *GS3*, *GL3.1*, *TGW3*, *GS5*, *GW5*, *GS6*, *TGW6*, *GW6a*, *GLW7*, *GL7* and *GW8*) were selected for functional marker development. Twenty-one PCR-gel-based markers were developed to genotype the candidate functional nucleotide polymorphisms (FNPs) of these genes, and all markers can effectively recognize the corresponding allele types. To test the allele effects of different FNPs, a global collection of rice cultivars including 257 accessions from the Rice Diversity Panel 1 was used for allele mining, and four grain-size-related traits were investigated at two planting locations. Three FNPs for *GW2*, *GS2* and *GL3.1* were genotyped as rare alleles only found in cultivars with notably large grains, and the allele contributions of the remaining FNPs were clarified in both the *indica* and *japonica* subspecies. Significant trait contributions were found for most of the FNPs, especially *GS3*, *GW5* and *GL7*. Of note, *GW5* could function as a key regulator to coordinate the performance of other grain size genes. The allele effects of several FNPs were also tested by QTL analysis using an F_2 population, and *GW5* was further identified as the major locus with the largest contribution to grain width and length to width ratio.

Conclusions: The functional markers are robust for genotyping different cultivars and may facilitate the rational design of grain size to achieve a balance between grain yield and quality in future rice breeding efforts.

Keywords: Rice, Grain size, Functional nucleotide polymorphism, Functional marker, QTL analysis

Background

Grain size is the major determinant of grain weight, one of the three yield components (panicle number, grain number and grain weight) in rice, and it can be further categorized into grain length, grain width and length to width ratio (Xing and Zhang 2010). Grain size also correlates with grain quality traits, such as grain appearance, head rice yield, cooking and eating quality (Nelson

et al. 2011; Tan et al. 2000). The desirable grain appearance varies in different regions of the world; long and slender rice is preferred by consumers in America, Africa and Southeast Asia, while short and bold rice is preferred by northern China, Japan and Korea (Calingacion et al. 2014). Therefore, it is important to breed rice cultivars with various grain sizes to meet the needs of different market classes.

Grain size is controlled by multiple quantitative trait loci (QTLs), and the chromosomal regions of different loci have been repeatedly detected by genetic populations with different parent combinations (Segami et al. 2016; Xia et al. 2017). Many QTLs for grain size have each been fine-mapped to small chromosomal regions,

* Correspondence: zhangl@yzu.edu.cn; qqliu@yzu.edu.cn

†Lin Zhang and Bin Ma contributed equally to this work.

¹Jiangsu Key Laboratory of Crop Genomics and Molecular Breeding /Joint International Research Laboratory of Agriculture and Agri-Product Safety of the Ministry of Education, Yangzhou University, Yangzhou 225009, China
Full list of author information is available at the end of the article

laying the foundation for gene cloning (Dong et al. 2018; Qiu et al. 2012). In recent years, more than ten QTLs for grain size and weight have been cloned and functionally dissected, including *GW2* (Song et al. 2007), *GS2/GL2* (Che et al. 2016; Duan et al. 2016; Hu et al. 2015), *GS3* (Fan et al. 2006; Mao et al. 2010), *qLGY3/OsLG3b* (Liu et al. 2018; Yu et al. 2018), *TGW3/qTGW3/GL3.3* (Hu et al. 2018; Xia et al. 2018; Ying et al. 2018), *GL3.1/qGL3* (Qi et al. 2012; Zhang et al. 2012), *GS5* (Li et al. 2011), *GW5/qSW5* (Duan et al. 2017; Liu et al. 2017; Shomura et al. 2008; Weng et al. 2008), *TGW6* (Ishimaru et al. 2013), *GW6a* (Song et al. 2015), *GL7/GW7* (Wang et al. 2015a; Wang et al. 2015b) and *GW8* (Wang et al. 2012). The QTLs differ in their effects on determining the grain width and grain length, making an improvement either for grain yield or grain quality. *gw2*, *GS5* and *gw5* primarily contributed to increased grain width and grain weight (Li et al. 2011; Song et al. 2007; Weng et al. 2008), while *qLgy3*, *gs3*, *gl3.1* and *tgw3* primarily contributed to increased grain length and grain weight (Fan et al. 2006; Hu et al. 2018; Liu et al. 2018; Qi et al. 2012). *GS2* caused a simultaneous and substantial increase in grain length, grain width and grain weight, but this resulted in poor grain appearance with a significant increase in grain chalkiness (Hu et al. 2015). *GL7* contributed to increased grain length but decreased grain width without affecting grain weight and was designated a grain quality gene for its effect in reducing grain chalkiness (Wang et al. 2015b). *GW8* contributed to an increase in both grain width and weight for yield improvement, and the *gw8* allele reduced grain weight but increased grain length and grain quality (Wang et al. 2012). Genome-wide association studies (GWAS) have been shown to be another efficient method for QTL mapping, and through GWAS, the role played by *GLW7* in determining both rice grain length and weight was first identified (Si et al. 2016). *GS6* was cloned from an EMS mutant, and natural variations of the gene were found to increase grain width and weight (Sun et al. 2013).

Cloning of several different QTLs helped to elucidate the mechanism of grain size determination. Functional analysis demonstrated that grain size traits can be directly linked to the gene coding region or regulatory elements. *GW2*, *qLGY3*, *GS3*, *GL3.1*, *TGW3*, *GW5*, *GS6* and *TGW6* are negative regulators of grain size, while *GS2*, *GS5*, *GW6a*, *GLW7*, *GL7* and *GW8* are positive regulators. A 1-bp deletion in exon 4 of *GW2* resulted in increased grain width, and several SNPs in its promoter also contributed to trait variation (Lu et al. 2013; Song et al. 2007). Similar to *GW2*, one nucleotide deletion was identified in the coding region of *TGW6*, which led to loss of function for the gene and an increase in grain weight (Ishimaru et al. 2013). A T to A nucleotide substitution in exon 2 of *GS3* induced a premature codon, contributing to the increase in grain length (Fan et al.

2006). Additional variations in the coding region of *GS3* have been found in different cultivars, and their effects have not been clarified (Mao et al. 2010). A C to A nucleotide substitution led to an amino acid change in the kelch domain of *GL3.1* and was shown to be the causative polymorphism for long grains (Zhang et al. 2012). The 1212-bp deletion of *GW5* increased grain width directly, because the deletion reduced the expression of its causative gene (Liu et al. 2017; Weng et al. 2008). A truncating splice-site mutation of *qLGY3* was associated with slender grains, and the alternatively spliced protein became a dominant negative regulator of the normal protein (Liu et al. 2018). Functional loss of *TGW3* contributed to an increase in grain length, and several FNP s have been identified from different cultivars (Hu et al. 2018; Xia et al. 2018; Ying et al. 2018). Three haplotypes of *GS6* were identified in natural accessions, and the sequence polymorphism of its promoter most likely determined grain size due to the effect on gene expression (Sun et al. 2013). For the positive regulators, the contributing FNP s were primarily located in the regulatory regions, and polymorphisms in the promoter of *GS5* and *GW6a* and the 5'UTR (untranslated region) of *GLW7* and *GW8* possibly increased grain size by elevating gene expression (Li et al. 2011; Si et al. 2016; Song et al. 2015; Wang et al. 2012). *GS2* is under the control of miR396, and two SNPs have been identified on the miRNA targeting site to generate larger grains by promoting gene expression (Hu et al. 2015). The *GL7* locus harbored a new copy of the original gene, leading to gene upregulation and increased grain length (Wang et al. 2015b). Polymorphisms in the 5'UTR of *GL7* might also determine the trait variation, as they showed distinct ability in driving reporter expression (Wang et al. 2015a).

Clarification of variations controlling grain size traits provides valuable targets for molecular marker-assisted selection (MAS) breeding, and introgression of one or multiple genes has been successful in improving the grain yield and grain quality of different cultivars (Nan et al. 2018; Zhao et al. 2018). Moreover, a rational design breeding strategy was developed and achieved a breakthrough, highlighting the power of pyramiding suitable genes to quickly breed desirable rice cultivars (Zeng et al. 2017). Facilitated by markers linked to genes controlling grain size, eating quality and yield, the Chinese breeders introgressed multiple superior alleles of Nipponbare (NIP) and 9311 into the cultivar Teqing, and obtained improved cultivars with similar yield to Teqing while improving eating quality and appearance (Zeng et al. 2017). During MAS breeding, PCR agarose gel-based markers are popular in most breeding programs because this technology is easy and cost-effective to implement (Hu et al. 2020). Insertion/deletion (InDel) are major sequence variations targeted for marker development, and primers can be easily

designed for polymerase chain reaction (PCR). Nevertheless, single nucleotide polymorphisms (SNPs) can also be discriminated by agarose gel using the tetra primer PCR method or cleaved amplified polymorphic sequence (CAPS) markers (Fan et al. 2006; Kim et al. 2016). To date, some gene-linked random markers and functional markers have been developed, involving more than 60 known genes related to yield and biotic and abiotic stress tolerance (Hu et al. 2020; Kim et al. 2016). Compared with gene-linked random markers, functional markers are more efficient in MAS, as they can fix gene alleles in multiple genetic backgrounds without additional calibration (Bagge et al. 2007). Although FNP of different grain size genes have been identified, very few have been transferred into functional markers (Fan et al. 2006; Kim et al. 2016; Wang et al. 2011; Weng et al. 2008). Therefore, functional markers should be rapidly developed and continuously updated to facilitate breeding applications.

In this study, we surveyed 20 candidate FNPs from 14 cloned grain size genes, and transferred them into PCR-gel based markers. We attempted to clarify the markers' ability to discriminate the allelic distribution of different FNPs with 257 accessions from the Rice Diversity Panel 1 (RDP1), and evaluated their contribution to modifying grain dimensions (length, width, length to width ratio). Genotyping and phenotyping this large set of accessions made it possible to compare the effect of different FNPs synchronously, thus providing guidance for the rational design of grain size. Moreover, to determine the potential function of the marker set in locating QTLs, we tested it by an F_2 population. We believe these newly developed functional markers will facilitate both genetic study and breeding applications in the future.

Result

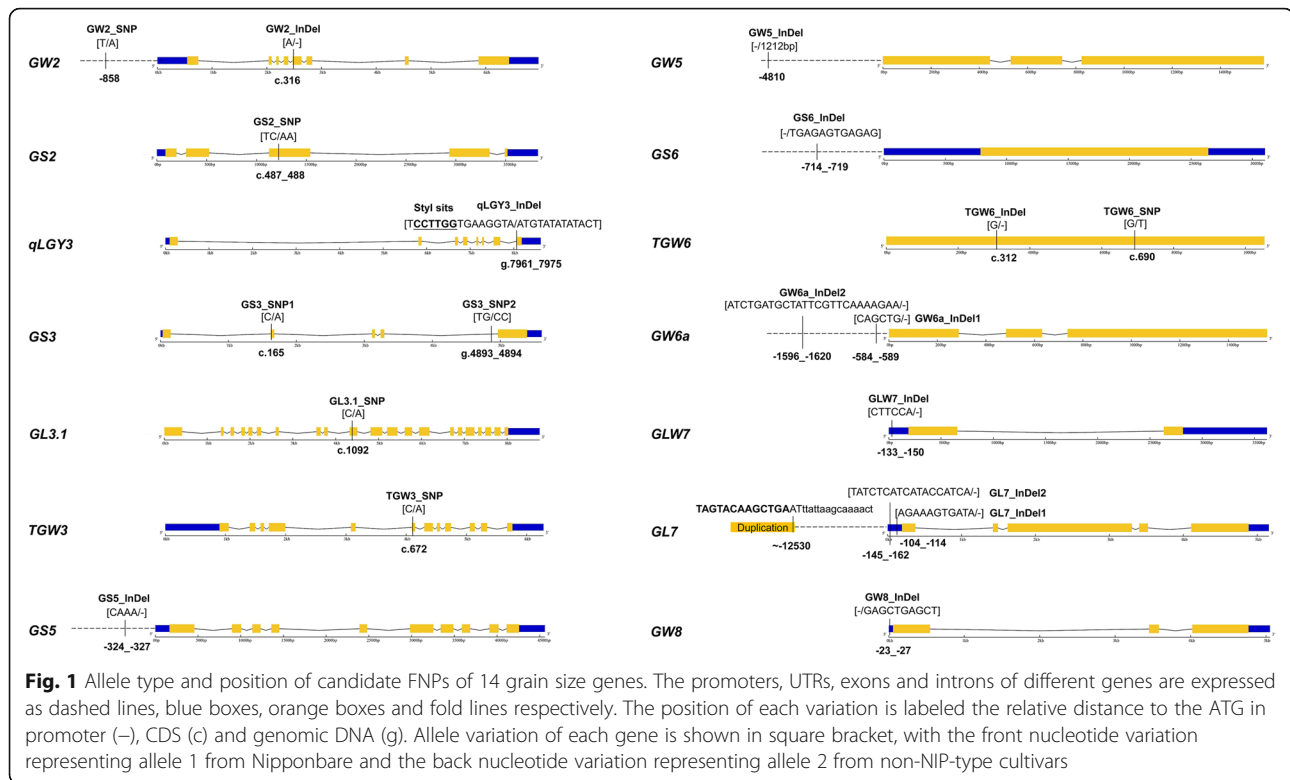
Survey of Candidate Functional Variations of 14 Grain Size Genes

According to previous reports, 14 grain size QTLs/genes located on six chromosomes were selected for sequence verification (Fig. 1 and Table S1). Using the Nipponbare (NIP) genome as the reference, we confirmed the polymorphic sequences and their positions on different genes, and 20 candidate FNPs including seven SNPs and 13 InDels were selected (Fig. 1). The contributions of several FNPs are known for respective genes, including the 1-bp deletion in exon 5 of *GW2*, the TC to AA substitution in exon 3 of *GS2*, the replacement of 15-bp sequence by 14-bp sequence in the joint region of intron 7 and exon 8 of *qLGY3*, the C to A substitution in exon 2 of *GS3*, the C to A substitution in exon 10 of *GL3.1*, the C to A substitution in exon 6 of *TGW3*, the 1212-bp deletion in promoter of *GW5*, the 1-bp deletion in the single exon of *TGW6*, the 6-bp/10-bp InDels in the 5'UTR of *GLW7/GW8* and the *GL7* duplication

approximately 12.5 kb upstream of its original copy (Fig. 1). The 4-bp InDel of the *GS5* promoter had been designated a functional variation without further validation (Kim et al. 2016) and was subjected to marker development to validate its effect. A T to A SNP at the -858-bp position of the *GW2* promoter was selected as a candidate functional variation, which was identified in a previously reported haplotype analysis (Lu et al. 2013). A 3-bp deletion with an unknown effect existed in exon 5 of *GS3* in the cultivar Zhenshan 97 (Mao et al. 2010), and we sequenced 43 RDP1 accessions to check its allele frequency. Ten accessions harbored the deletion tightly linked with the 2-bp SNP substitution in intron 4, which was used for marker design (Fig. 1 and Fig. S1). In addition to the 1-bp deletion of *TGW6*, six SNPs were also identified from the coding region between Kasalath and NIP (Ishimaru et al. 2013), and the G to T SNP at the 690-bp position of the exon was selected for marker design to test its effect (Fig. 1). Sequence polymorphisms from the -395 to -1415 position of the *GW6a* promoter have been suggested to contribute to trait variation (Song et al. 2015), and a 6-bp deletion in that interval and another 25-bp deletion located at approximately the -1.6 kb position were selected for marker design to compare their effects (Fig. 1). One 12-bp insertion in the *GS6* promoter and two deletions in the 5'UTR of *GL7* likely affect the grain size by changing their gene expression (Sun et al. 2013; Wang et al. 2015a), and were selected for marker development to validate their effect (Fig. 1). To facilitate citation, all the sequence variations were named the gene name with the mutant type (Fig. 1), allele 1 represented the reference allele based on the NIP sequence, and allele 2 represented another allele type (i.e. non-NIP-type illustrated in Fig. 1 legend).

Marker Development and Quality Evaluation for Gel Electrophoresis

To facilitate large-scale genotyping, we attempted to transfer all the candidate FNPs into PCR-gel-based molecular markers. InDels larger than 10 bp can be identified by PCR directly using flanking primers on the condition that the size of PCR products is no more than tenfold greater than the InDel size (Yamaki et al. 2013). With this criteria, six InDel markers for *GW5_InDel*, *GS6_InDel*, *GW6a_InDel2*, *GL7_InDel1*, *GL7_InDel2* and *GW8_InDel* were developed (Table 1). In addition to the 6-bp deletion of the *GLW7* 5'UTR, we found an additional 3-bp deletion in some accessions (Fig. S2), and based on this, an InDel marker covering the two small InDels was developed. We also developed new primers flanking the border of the *GL7* duplication with PCR products smaller than those previously reported (Wang et al. 2015b), which can save PCR amplification



time. CAPS or dCAPS markers were developed for the remaining variations including the 6-bp deletion of *GLW7*. In the target region of *qLGY3*, the reference allele harbors a *StyI* enzyme site, which is aborted in the mutated sequence (Fig. 1), and the CAPS marker was generated by a pair of flanking primers (Table 1). For dCAPS markers, one or two nucleotides in the forward primers were replaced by other types to generate a restriction enzyme site, and the strategy was successful for both SNPs and small InDels (Fig. S3). Finally, twelve forward dCAPS primers involving 10 restriction enzyme sites were selected (Fig. S3), and the reverse primers were designed accordingly. Taken together, twenty-one pairs of primers were developed for the 20 FNPs (Table 1).

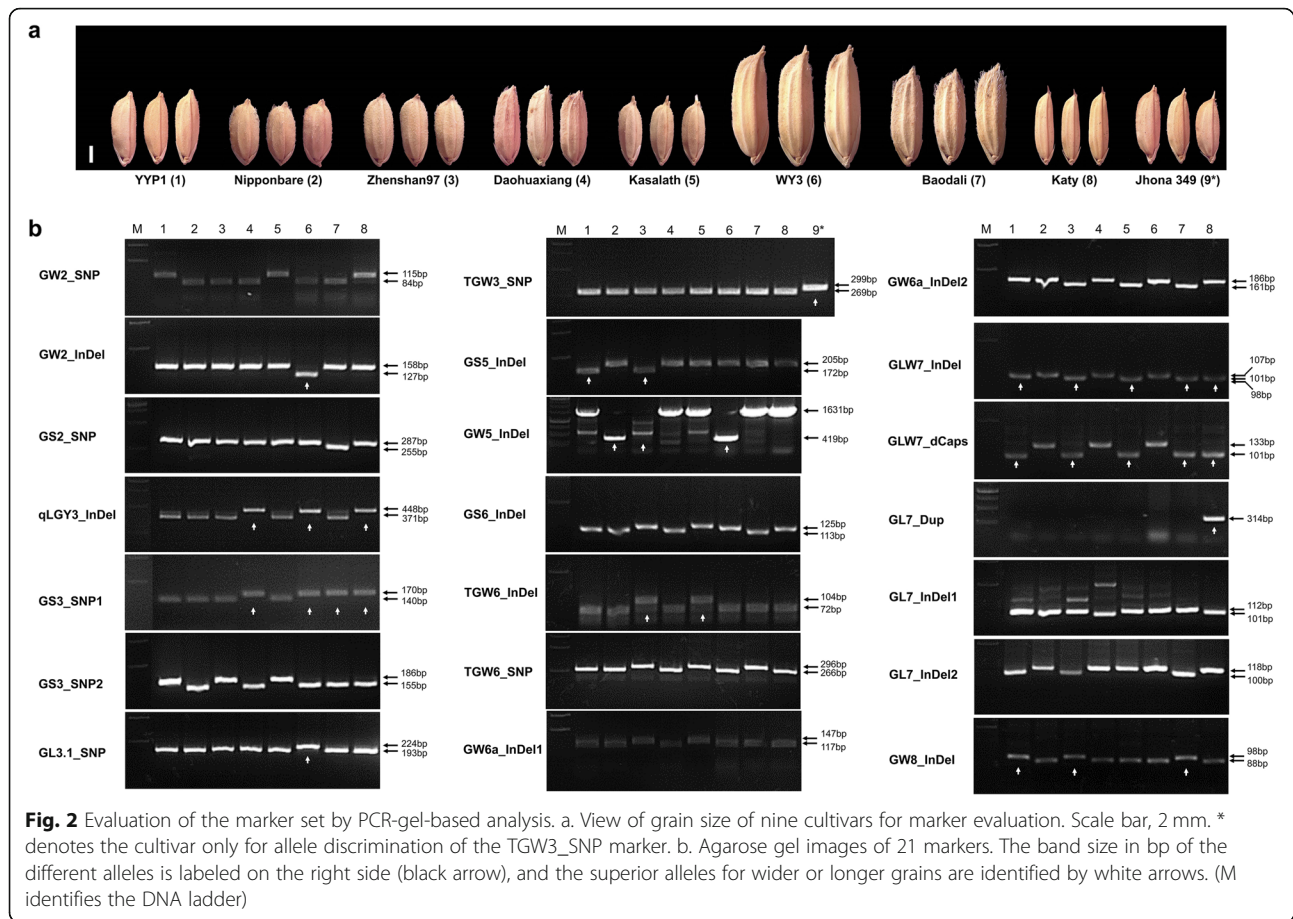
To check the markers' ability to discriminate different alleles, eight cultivars were genotyped with the markers. As no allele variation was found for *TGW3_SNP* among the eight cultivars, thus an additional cultivar with the target allele was added for the marker (Fig. 2). It is worth noting that *WY3* and *Baodali* are two cultivars with notably large grains (Fig. 2a), which were the original cultivars for cloning *GL3.1/GW2* and *GS2*, respectively. All the markers showed good amplicon quality and polymorphism among the tested cultivars (Fig. 2b). For the *GL7_Dup* marker, a 303-bp band can only be amplified in the U.S. long grain cultivar *Katy* which has a slender

grain, indicating its specificity in amplifying the border region of the new copy. The variation of large InDels can be reflected directly by the amplicon size of InDel markers (Fig. 2b). In the NIP cultivar, smaller bands were amplified for *GW5_InDel*, *GS6_InDel* and *GW8_InDel*, and larger bands were amplified for *GW6a_InDel2*, *GLW7_InDel*, *GL7_InDel1* and *GL7_InDel2*, closely matching the respective InDel size (Fig. 1). As *Zhenshan97* harbors another type of *GW5* mutation (Duan et al. 2017), it was identified as a null band by the *GW5_InDel* marker (Fig. 2b). For the CAPS marker of *qLGY3*, smaller bands were found in the cultivars with the NIP-type allele after enzyme digestion, thereby confirming the existence of the *StyI* site on the target region (Fig. 1 and Fig. 2b). Eight dCAPS markers showed digestible bands for NIP, including *GW2_SNP*, *GS3_SNP1*, *GS3_SNP2*, *GL3.1_SNP*, *TGW3_SNP*, *TGW6_InDel*, *TGW6_SNP*, and *GW6_InDel2* (Fig. 2b). In contrast, resistant bands for NIP were generated by *GW2_InDel*, *GS2_SNP*, *GS5_InDel* and *GLW7_InDel* (dCAPS) markers (Fig. 2b). Compared with the InDel marker of *GLW7*, its dCAPS marker was better at discriminating the allele difference (Fig. 2b). Therefore, all these newly developed markers correctly identify allele variations of different grain size genes.

To facilitate comparisons between the markers, we defined the allele for wider or longer grains as the

Table 1 Primer information on the 21 functional markers targeting 14 grain size genes

Primer Names	Primers sequence (5'-3')	Enzyme sites	Amplicon size (Digested) bp
GW2_SNP-F	AAAACCAAAACCTAACACGTGGATACAACA	Nspl	115 (84)
GW2_SNP-R	GGCGGTGAAGATAGATGTACT		
GW2_InDel-F	CTCACACTGCTCAGCCTACA	PstI	158 (127)
GW2_InDel-R	CACGATACTCCACAGCATAACTGGGAGTCT		
GS2_SNP-F	GAGCGCCACATGCACCGCGCCGCATACGT	SnaBI	287 (255)
GS2_SNP-R	TTGCCTGTTCCACCACCAACAGC		
qLGY3_InDel-F	GGGTATTCTTCTTTCTTCCA	StyI	448 (371)
qLGY3_InDel-R	TGCTGCTGCTTCATTGCT		
GS3_SNP1-F	CGAAGGGATCCACGCTGCCTCCAGATGATT	HinfI	170 (140)
GS3_SNP1-R	AGTTGCTTAAAAGATAACGGTCAAA		
GS3_SNP2-F	CAAACATGAAAACCTTGCT	HaeIII	186 (155)
GS3_SNP2-R	AAATAAAACGTGTGATTTAATCGTTAACGGC		
GL3.1_SNP-F	TGCACGATTCTATCTGGTTCAGTGGTCCA	SalI	224 (193)
GL3.1_SNP-R	CTAAACAACAGGTTTTCTTAC		
TGW3_SNP-F	TGCACTGCCAAAATCAGATATTTTGGCCG	HpaII	299 (269)
TGW3_SNP-R	TCACTACCACTAGGTCCAC		
GS5_InDel-F	ACTCCCATGGAATTACTAGAGAAGCCAAGC	AluI	262 (205/172)
GS5_InDel-R	AGGAGAAGAAAAGGTGAAAAGT		
GW5_InDel-F	TGCGTCGGTCGTTGGAGG		419/1631
GW5_InDel-R	GCGAGCGAGCGTGTGTAGG		
GS6_InDel-F	TCTTCTCCCTCCCTTCCAC		113/125
GS6_InDel-R	TGCCTCACCATTACCACTCTCT		
TGW6_InDel-F	AGCCCCAGCTACACGAAAACAAGTGCCCG	HpaII	154 (104/72)
TGW6_InDel-R	GACCAACTCGCATCAATCCC		
TGW6_SNP-F	ATGCGAGTTGGTCCAAAAGG	SacI	296 (266)
TGW6_SNP-R	GCAACGATCAGATGTGTTCCGTCAGAGCT		
GW6a_InDel1-F	TGCACCTCTCTCGTGTCT	BclI	147 (117)
GW6a_InDel1-R	TTATTTTCGTTCAATTTTCTATGGCTGAT		
GW6a_InDel2-F	ATTACTCTTGATTTTGTCTGTT		186/161
GW6a_InDel2-R	GAGTTTGCATGTTCTGTTCT		
GLW7_InDel-F	CTCGAGCTCGAGCTCATG		107/101/98
GLW7_InDel-R	TCCCTTCAACCTTTTCCA		
GLW7_InDel_dCaps-F	CCTCTCTCTCCGCTTCCACTTCGACT	HinfI	133 (101)
GLW7_InDel_dCaps-R	TGCTACTGTGCTGTGTGCT		
GL7_Dup-F	GATGAGCTGACAAGCAAAAG		0/314
GL7_Dup-R	AGAGGATGGGGATGAAAGAAT		
GL7_InDel1-F	CAGCTCACGCACATCCAAC		112/101
GL7_InDel1-R	ACCATACCACATCTCATCTCAC		
GL7_InDel2-F	GGACGCGTGAGATGAGATGTG		118/100
GL7_InDel2-R	CTCCCGCTTATTTCAACCCC		
GW8_InDel-F	GTGCGTGCGTCAACACACAG		88/98
GW8_InDel-R	TGAGATCCCACTCCATGGCC		



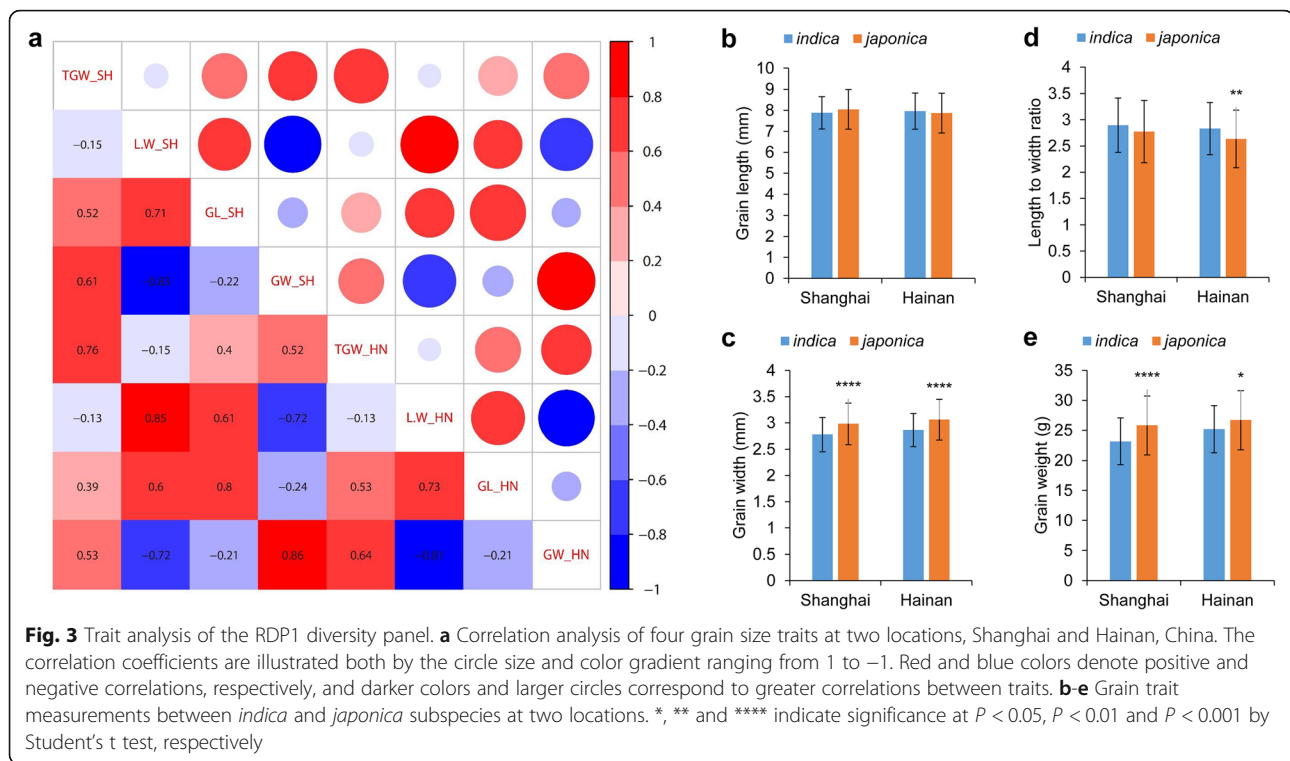
“superior allele”, although the contrasting allele would also be useful depending on the grain shape preference for a particular market. With this in mind, we compared the number of known superior alleles among the tested cultivars and found that each cultivar contained no more than five superior alleles (Fig. 2b; white arrows). Two cultivars with large grains harbored five and four superior alleles, specifically *gw2*, *qlgy3*, *gs3*, *gl3.1*, *gw5* for WY3 and *GS2*, *gs3*, *GLW7* and *GW8* for Baodali (Fig. 2b). Although Zhenshan97 also contained five superior alleles, including *GS5*, *gw5*, *tgw6*, *GLW7* and *GW8* (Fig. 2b), its grain size was significantly smaller than that of WY3 and Baodali (Fig. 2a). It can be inferred that other unknown genes contributed to the increase in grain size for these two large grain cultivars.

Grain Size Performance of a Large Rice Diversity Panel

To clarify the allele contribution of different FNPs in shaping grain size, we analyzed 257 accessions in the RDP1 from different regions of the world (Table S2). The diversity panel was planted at Shanghai and Hainan, two locations corresponding to the temperate and tropical climates respectively, which should facilitate the

comparison of environmental effects. Mature grains were harvested for trait analyses, including grain length, grain width, grain length to width ratio and grain weight. Pearson correlation analysis was performed for the different traits, and significant correlations were found for the same traits between the two locations, suggesting consistent performance of most accessions in different environments (Fig. 3a). Among the four traits, grain length showed the strongest positive correlation to length to width ratio and a weak negative correlation to grain width, while grain width showed the most significant negative correlation to length to width ratio. Grain weight showed a similar positive correlation with grain length and grain width, and a weak negative correlation with length to width ratio (Fig. 3a). This result suggests that the length to width ratio can be modified easily by changing either grain length or grain width, but it may be difficult to increase the grain weight by improving grain length and width simultaneously.

Based on the previously reported population structure analysis (Table S2; McCouch et al. 2016), we divided RDP1 accessions into the two well recognized rice subspecies, *indica* and *japonica* and compared the trait



variations between the subspecies at Shanghai and Hainan. No significant difference in grain length was found between the two subspecies, but a considerable increase in grain width was identified in the *japonica* subspecies at both locations (Fig. 3b, c). For the length to width ratio, a significant decrease in *japonica* subspecies was only found at the Hainan (Fig. 3d). Similar to grain width, the grain weight was increased in the *japonica* subspecies at both locations (Fig. 3e), suggesting that the higher grain weight of these accessions was largely attributable to the grain width.

Clarification of Allele Effects by Different FNP

Using the newly-developed functional markers, we obtained all the allele types of the 20 FNPs for the RDP1 accessions (Table S4). We did not find the superior alleles of GS2_SNP, GW2_InDel and GL3.1_SNP identified from the large grain cultivars, indicating that they belonged to rare alleles. The other polymorphic sites showed relatively balanced allele distribution, with the minor allele frequencies ranging from 8.2% to 48.6%, and GL7_Dup exhibiting the fewest superior alleles (Fig. 4a). In addition, heterozygotes were found in a few accessions (Fig. 4a). We further compared the homozygous allele distribution between the two subspecies and found an uneven allele ratio in either subspecies for most target variations (Fig. 4b). Most *indica* accessions carried allele 1 of qLGY3_SNP, GS3_SNP1, GL7_Dup and GL7_InDel1 variations, and allele 2 of GS6_InDel,

TGW6_SNP, GLW7_InDel and GW8_InDel variations. Most *japonica* accessions contain allele 1 of GS3_SNP2, TGW3_SNP, GS5_InDel, TGW6_InDel, TGW6_SNP, GW6a_InDel1, GW6a_InDel2, and GL7_InDel2. Therefore, some useful alleles were enriched only in a specific subspecies.

We then compared the contribution of different FNPs in determining grain size, and 16 FNPs were subjected to analysis in either subspecies with a relatively even allele distribution (Fig. 4b). To show the allele contribution directly, we determined the changing rate of mean values of different traits between two alleles, using allele 1 as the reference (Fig. 5). We found that nine FNPs contributed largely to grain length, grain width or length to width ratio, namely GW2_SNP, qLGY3_SNP, GS3_SNP1, GS3_SNP2, TGW3_SNP, GW5_InDel, GLW7_InDel, GL7_Dup and GL7_InDel1 (Fig. 5a-c). Among these FNPs, allele 1 of GS3_SNP2 and allele 2 of the remaining FNPs increased both the grain length and length to width ratio but decreased the grain width. GS5_InDel contributed only to the change in grain width, with allele 2 for the wider grains. GS6_InDel contributed to the change in grain length only at Hainan, with allele 1 for the longer grains. TGW6_InDel contributed to both grain width and length to width ratio, with allele 2 increasing grain width but reducing the length to width ratio. No significant contribution to the three abovementioned traits can be found for GW6a_InDel1, GW6a_InDel2, GL7_InDel2 and GW8_InDel, suggesting

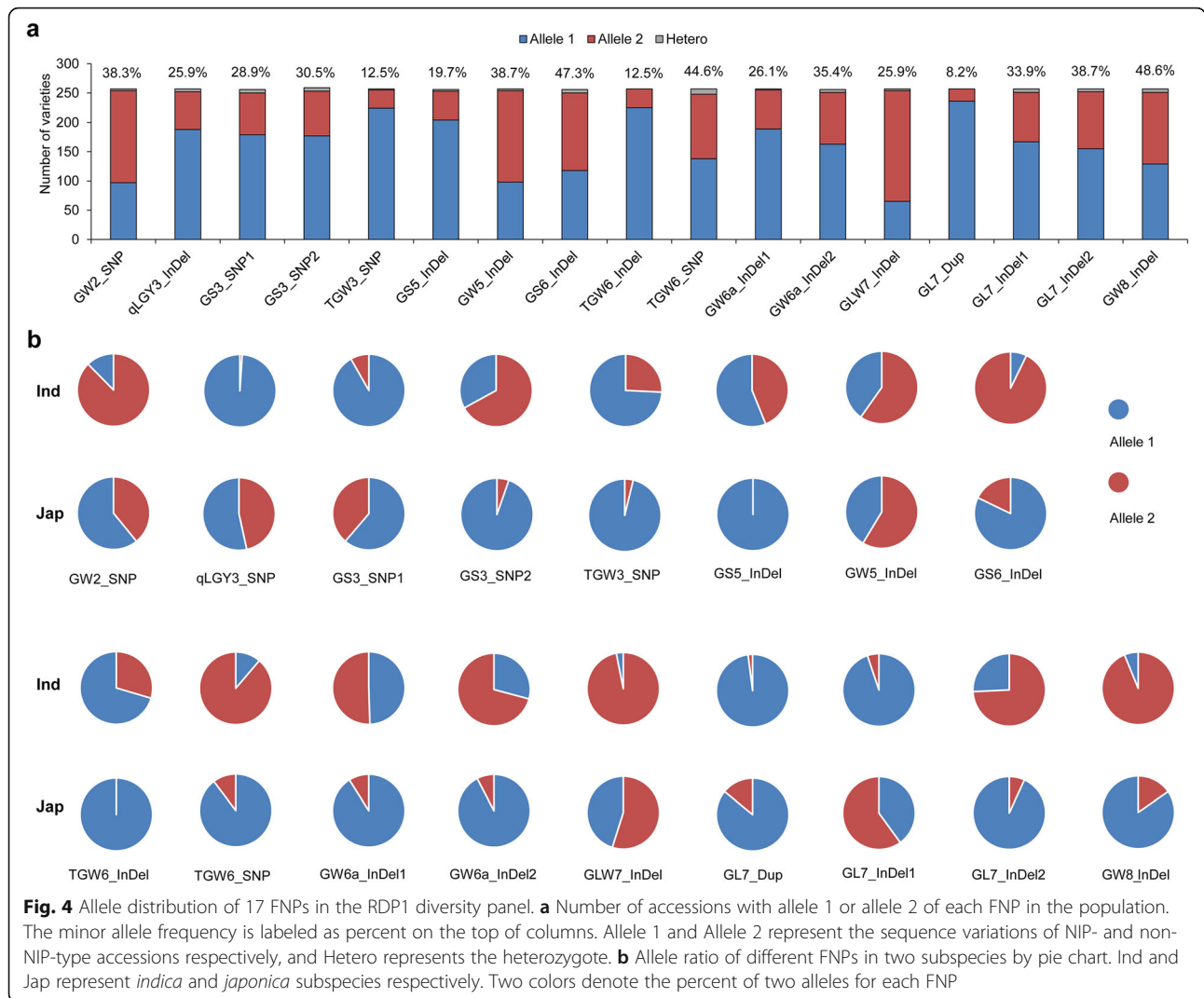


Fig. 4 Allele distribution of 17 FNPs in the RDP1 diversity panel. **a** Number of accessions with allele 1 or allele 2 of each FNP in the population. The minor allele frequency is labeled as percent on the top of columns. Allele 1 and Allele 2 represent the sequence variations of NIP- and non-NIP-type accessions respectively, and Hetero represents the heterozygote. **b** Allele ratio of different FNPs in two subspecies by pie chart. Ind and Jap represent *indica* and *japonica* subspecies respectively. Two colors denote the percent of two alleles for each FNP

relatively weak effects of these variations. Regarding grain weight, stable trait contributions can only be found for GS5_InDel and GW5_InDel in the *indica* subspecies and GL7_Dup in the *japonica* subspecies, with allele 2 of GS5_InDel and allele 1 of GW5_InDel and GL7_Dup increasing the trait (Fig. 5d). Five variations contributed to the trait change only at one location, namely, an increase in grain weight by allele 1 of qLGY3_SNP and GL7_InDel1 and allele 2 of GS3_SNP1 in Hainan, and allele 2 of GW6a_InDel2 and GW8_InDel in Shanghai (Fig. 5d). Therefore, grain weight determination could be complicated due to changes in multiple grain dimensions and easily affected by the environment. Moreover, the effect of GW5_InDel on grain weight can only be found in *indica* subspecies, suggesting that genetic background could affect the performance of different genes.

Among the variations detected, GW5_InDel showed balanced allele distribution and stable effect in

determining grain width for both subspecies (Fig. 4b and Fig. 5b), therefore we tested its allelic interactions with other variations, including GS3_SNP2, qLGY3_InDel, GS5_InDel, TGW6_InDel, GLW7_InDel and GL7_InDel1 (Fig. 6). Allele 1 of GW5_InDel had a stable effect by increasing grain width when combined with either alleles of other variations. However, the promoting effect of other tested variations can only be identified when combined with allele 2 of GW5_InDel, especially for GS3_SNP2, qLGY3_InDel and GL7_InDel1. Therefore, allele 2 of GW5, namely the functional allele with the 1212-bp sequence, could underlie the regulation of grain size for other QTLs.

Evaluation of the Functional Markers by QTL Analysis

We then tested the allele contribution using an available F₂ population with 119 individuals derived from a cross of YYP1 and NIP. We found that the two parents had

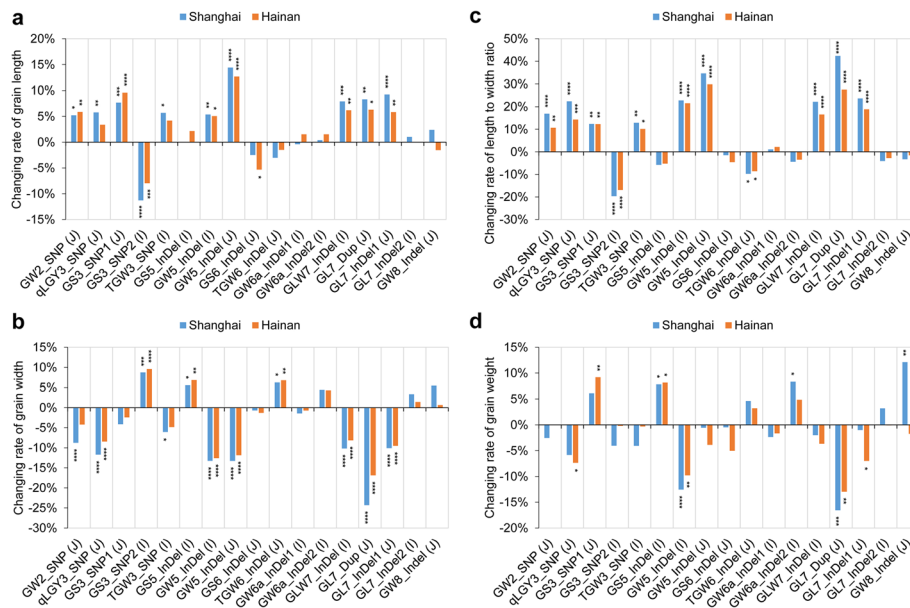


Fig. 5 Contribution of different alleles to four grain-size-related traits. The columns show the relative changing rate of allele 2 (non-NIP-type) exceeding allele 1 (NIP-type) for grain length (a), grain width (b), length to width ratio (c) and grain weight (d). The I and J in parentheses represent the *indica* and *japonica* subspecies respectively. The values are shown as percentages, and the significance level of the difference between the two alleles of each SNP was calculated by Student's t test. *, **, *** and **** indicate significance at $P < 0.05$, $P < 0.01$, $P < 0.001$ and $P < 0.001$, respectively

different allele types for GW2_SNP, GS3_SNP2, GS5_InDel, GW5_InDel, GLW7_InDel, GL7_InDel2 and GW8_InDel, and NIP carried allele 1 of GW5_InDel for wider grains (Fig. 2). QTL analysis for the four grain size traits was performed with 123 markers evenly distributed across the 12 chromosomes, which covered the

seven polymorphic functional markers (Table S5). In all, 11 QTLs were identified for the four traits, including three for grain length on chromosomes 5, 8 and 11, three for grain width on chromosomes 3 and 5, four for length to width ratio on chromosomes 2, 4, 5 and 10, and one for grain weight on chromosome 2 (Fig. 7 and

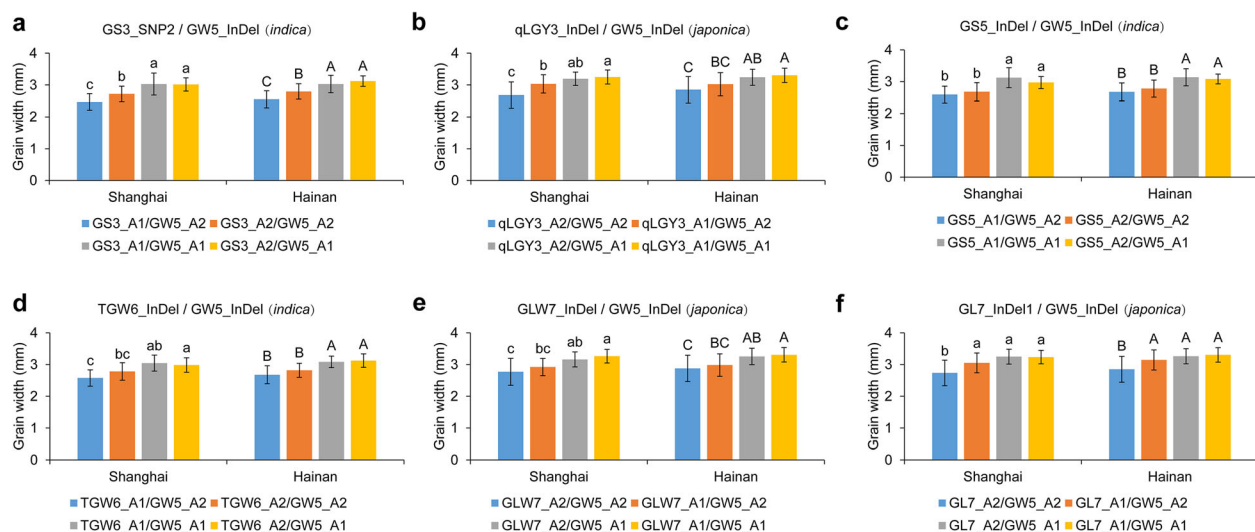


Fig. 6 Impact of GW5 alleles on grain width control of six other loci. Four allele combinations can be obtained between GW5_InDel and six other FNPs in either the *indica* or *japonica* subspecies, namely GS3_SNP2 (a), qLGY3_InDel (b), GS5_InDel (c), TGW6_InDel (d), GLW7_InDel (e) and GL7_InDel1 (f). A1 represents allele 1 (NIP-type) and A2 represents allele 2 (non-NIP-type). Multiple comparisons by Tukey's test were performed for data from Shanghai and Hainan, and different letters indicate significant differences among the four allele combinations

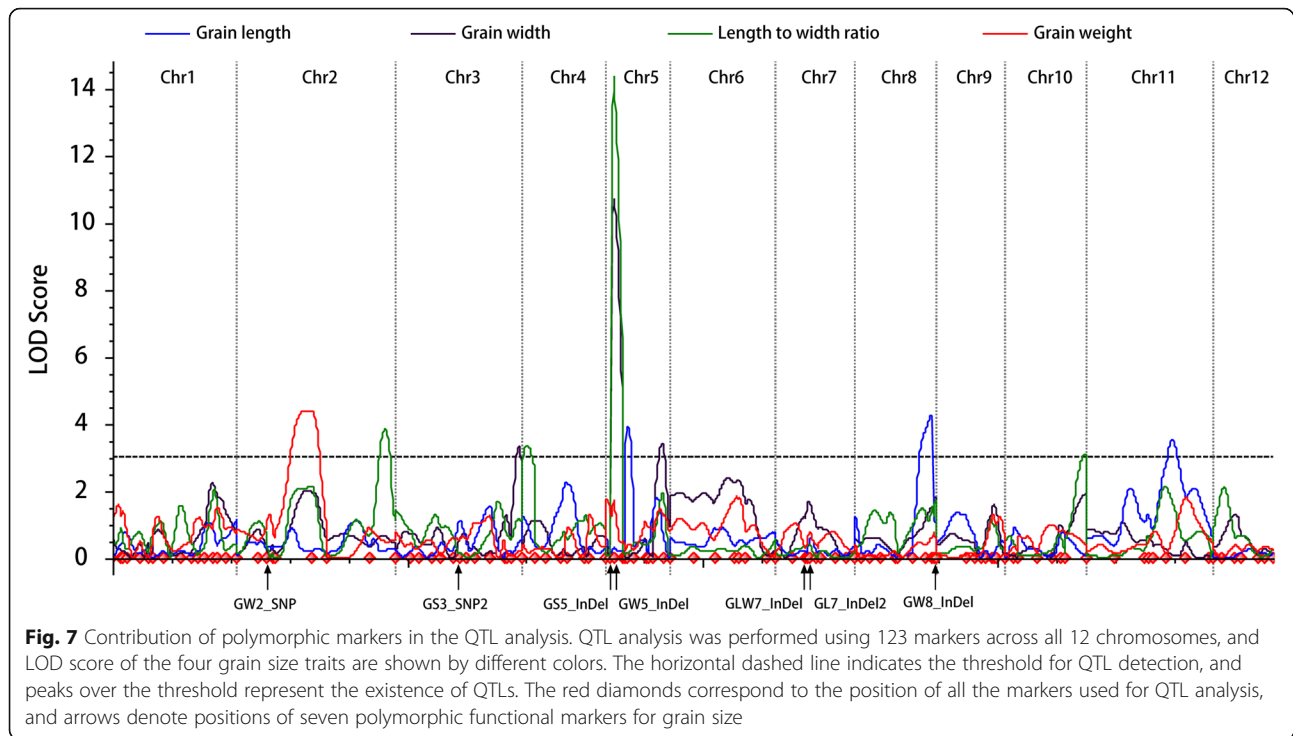


Table S6). The QTL on the short arm of chromosome 5 presented the largest LOD value for length to width ratio and grain width, and explained ~24% and 22% of the phenotypic variation respectively (Table S6). It can be inferred that *GW5* underlies the major QTL, as the *GW5_InDel* marker colocalizes with the QTL region and the promoting effect is contributed by the NIP allele. No exact QTL colocalization was found for the other functional markers, possibly due to the weak effect of these variations or their epistatic interaction with *GW5* as suggested above. Therefore, *GW5* could be the key locus for controlling grain size and had a stable contribution to grain width and length to width ratio in nearly all genetic backgrounds.

Discussion

Gene-associated markers are important when implementing the MAS strategy in the breeding process. This was illustrated in rice where gene-linked markers were developed for single or multiple genes for various traits and these were used to develop improved cultivars with purified genetic backgrounds in approximately six rounds of backcrossing without field trait selection (Liu et al. 2016; Wang et al. 2017). Many MAS breeding practices have been performed to target disease or pest resistance genes, and restorer or male sterility lines with improved resistance have been generated (Jiang et al. 2019; Jiang et al. 2015; Wang et al. 2016). However, limited research has been focused on the MAS of yield-

related traits in recent decades. Recently, beneficial alleles of the *GS3*, *Gn1a*, and *Hd1* genes were introgressed into Kongyu131, a high-quality cultivar widely planted in the northeastern region of China, which improved both the grain yield and regional adaptation of the newly developed lines (Feng et al. 2017; Nan et al. 2018; Wang et al. 2019). To date, many gene-linked markers have been developed to facilitate MAS breeding but these gene-linked markers are not associated with the target alleles for the genes and can only be used for allele mining after their association with the target allele is confirmed (Hu et al. 2020). In contrast, functional markers developed from the causative variations are diagnostic of the desired trait allele thus are the preferred over gene-linked markers (Liu et al. 2012b). Although several functional markers have been developed, the functional marker set needs to be broadened as additional significant genes are characterized. A functional marker set was developed for eight yield-enhancing genes, but only covered two genes for grain size (Kim et al. 2016). In order to implement more extensive use of MAS breeding, we developed a more extensive functional marker set covering the main cloned QTLs for grain size and proved the effects of these markers in discriminating grain size traits.

All the markers we developed are PCR-gel-based and can be easily implemented by most breeding programs with reasonable equipment expenditures. Among the different DNA polymorphisms, InDels can be easily

transferred into molecular markers, but only work efficiently when the InDel size is large enough. As indicated by the GLW7_InDel marker in our analysis, the smallest InDel size could be 9 bp for agarose gel discrimination on the condition that the PCR product is also small. However, four FNP for the grain size genes have InDels smaller than 9 bp, making them difficult to discriminate as InDel markers. As shown by Kim et al. (2016), the InDel marker for the 4-bp deletion of *GS5* can only be discriminated by polyacrylamide gel electrophoresis or capillary electrophoresis, thus more efficient markers for these small InDels need to be developed by a new strategy. For SNP variations, PCR-gel-based markers can be developed by either the allele-specific PCR method or the CAPS/dCAPS method (Drenkard et al. 2000), but the PCR method has the limitations of nonspecific amplification, being efficient for specific SNP types and a low overall success rate in generating primers (Drenkard et al. 2000; Liu et al. 2012a). Moreover, this method requires two runs of PCR, thus increasing the cost for gel electrophoresis and data collection. Therefore, we used the CAPS/dCAPS method for marker development and successfully transferred all the SNPs into dCAPS markers. Noteworthy, small InDels were also transferred into dCAPS markers, providing a novel strategy for marker design of such FNPs. Utilizing this strategy, it was possible to transfer nearly all the FNPs for different genes into PCR-gel-based markers, thus facilitating breeding efforts to utilize a large number of genes in MAS strategies. The cost of each marker was also considered and priority was given to inexpensive restriction enzymes during primer design. Compared with the previous functional markers, the newly-developed marker set may reduce either the economic cost or the time commitment for PCR and gel-running. This marker set will facilitate the identification of the functional variations for grain size genes to quickly genotype a large number of cultivars. As shown in the study, a large set of RDP1 accessions was genotyped by these markers to simultaneously compare the effects of different genes, with considerably more capacity than Sanger sequencing (Wang et al. 2015b). In the future, this functional marker set will ultimately facilitate accurate allele mining for both genetic research and breeding applications.

Due to their global origin, the RDP1 accessions represent a wide range of genetic diversity (McCouch et al. 2016), as well as phenotypic diversity for the four grain size traits evaluated in this study, with the trait performance being relatively consistent between the two planting locations. Previously, these same traits were analyzed for the RDP1 grown in Arkansas, USA (Zhao et al. 2011), so we compared our dataset to these results. The correlation coefficients for grain length, grain width, length to width ratio and grain weight were 0.88, 0.89, 0.91 and

0.80, respectively, between the Arkansas and Shanghai data, and 0.72, 0.79, 0.76 and 0.66, respectively, between the Arkansas and Hainan data. Therefore, our data were well correlated with the previous data and expands the germplasm information for public use. The diversity panel can be categorized into the *indica* and *japonica* subspecies based on genotyping with 700,000 SNPs (McCouch et al. 2016). We found that the values of both grain width and grain weight in *japonica* were significantly larger than those of *indica*, suggesting that *japonica* cultivars might contain superior genes/alleles for these two traits. However, the main reason for this difference might not be attributable to the preferential allele distribution of the tested FNPs because the enriched alleles only consistently had a promoting effect for GW2_SNP, GS3_SNP1 and GLW7_InDel. In the future, it will be important to clarify the novel loci for these subspecies differences in segregating populations. Although there is a known close relationship between the *aus* and *indica* subpopulations composing the *indica* subspecies and similarly between *temperate japonica* (*Tej*) and *tropical japonica* (*Trj*) subpopulations composing the *japonica* subspecies, significant difference in grain size were noted between the *Tej* and *Trj* subpopulations, suggesting grain size is an important driver of subpopulation structure (Ali et al. 2011; McClung et al. 2020; Zhao et al. 2011). To investigate this relationship, we used our data to compare the grain size traits of four subpopulations (Table S2 and Table S3), and found significant differences between the *Tej* and *Trj* subpopulations for all four grain size traits but no difference between the *aus* and *indica* subpopulations, consistent with the previous results from different locations. We concluded four FNPs might contribute to the grain size difference between the two *japonica* subpopulations, namely, qLGY3_SNP, GW5_InDel, GLW7_InDel and GL7_InDel1, because the allele ratio is differentiated between the subpopulations (96%, 76%, 86% and 84% for allele 1 in *Tej* and 90%, 83%, 89% and 94% for allele 2 in *Trj* for four FNPs, respectively). This allele distribution might explain why *Tej* accessions have wider but shorter grains in contrast to *Trj* accessions as shown in previous studies (Ali et al. 2011; McClung et al. 2020; Zhao et al. 2011).

Previous GWAS repeatedly identified *GS3* and *GW5* as the major loci for grain size using different collections (Huang et al. 2010; McCouch et al. 2016; Zhao et al. 2011). We also consistently found a significant correlation between grain size traits to the two genes, with a major contribution of *GW5* to grain width and length to width ratio in both two subspecies at both locations. The large contribution of *GW5* was also confirmed in the segregating population, indicating that *GW5* can function as a key target to modulate grain size during

breeding. Regarding *GS3*, we found that the C to A mutation was not the only functional allele for grain length because a large effect also was identified for the TG to CC mutation linked to a 3-bp deletion in exon 5. Different deletions in exon 5 of *GS3* were identified to generate short grain rice (Takano-Kai et al. 2013), suggesting the role of the 3-bp deletion in reducing grain length. However, we did not identify a contribution of this allele in the segregating F_2 population, possibly due to the interaction between the two loci, as discussed below. In addition to the 3-bp deletion, another 1-bp deletion was previously identified in a specific cultivar Chuan 7 (Mao et al. 2010), but we did not find this allele by sequencing 43 accessions, suggesting that it belongs to a rare allele. In the GWAS analysis, *GLW7* could only be identified in the *japonica* cultivars (Si et al. 2016), and we did find a significant contribution of its FNP in the subspecies. In addition, we found a significant contribution of two *GL7* FNPs for nearly all four grain size traits, which was rarely identified by GWAS. This finding suggested that the FNPs might not be linked to SNPs used for GWAS, and our marker set can be used to enhance such analysis in the future.

The performance of some QTLs relied on interactive factors to shape regulatory pathways, and such phenomena have been found for *qLGY3*, *GW6a*, *GLW7*, *GL7/GW7* and *GW8* (Liu et al. 2018; Si et al. 2016; Song et al. 2015; Wang et al. 2012). The expression of *GL7/GW7* was negatively regulated by *GW8*, and the contribution of *GW8* could only be identified when it was pyramided with the responsive *GW7* allele (Wang et al. 2015a). In our analysis, we found similar situations when *GW5* was combined with several other loci, and the contribution of other loci could only be identified with the responsive *GW5* allele. Therefore, *GW5* might be an important downstream gene coordinating the performance of other loci, and it is important to consider these interactions between these genes when breeding for a particular market class. Such interactions might also affect the performance of other loci during QTL analysis, because we only found an exact colocalization of the QTL with *GW5* but not the other six FNPs tested. Evidence from QTL analysis by Wang et al. (2012) concludes some loci can only be detected in the populations without *GW5* segregation, thus it would be worthwhile to test the contribution of other loci combined with different *GW5* alleles using near isogenic lines in the future.

With the availability of these newly developed functional markers, it is now possible to rationally design the grain size by introgressing or pyramiding different alleles as part of the breeding process. Deciding which genes and alleles to use depends on the desired grain size and allele types of recipient cultivars, since either short, medium or long grain cultivars could be the target

depending on the market class. Based on the allele contribution obtained, allele 1 of *GW2_SNP*, *qLGY3_SNP*, *GS3_SNP1*, *TGW3_SNP*, *GW5_InDel*, *GLW7_InDel*, *GL7_Dup*, and *GL7_InDel1* and allele 2 of *GS3_SNP2*, *GS5_InDel*, and *TGW6_InDel* could be used to generate short and bold rice, and their contrasting alleles can be used to generate long and slender rice. To implement MAS, it is first necessary to confirm the allele type absent in the recipient using the marker set developed in this study, and determine which gene(s) to use. The superior alleles of *GW2_InDel*, *GS2_SNP* and *GL3.1_SNP* for large grains can be used for nearly all recipients due to their low frequency. Theoretically, the extent of trait change relies on the number of target alleles used, and a medium grain type could be obtained using some of these target alleles. We listed the allele information of 257 RDP1 accessions for all the tested FNPs (Table S4), thus it will be easier for breeders to select suitable donor accessions based on these FNPs. In China, most of the *japonica* elite cultivars were bred with short and round grains, which could lead to chalkiness, an undesirable grain appearance. Zhao et al. (2018) suggested *japonica* cultivars with a better grain appearance might be bred by introgressing the *gs9* mutation without reducing grain yield and grain yield further improved by pyramiding the *gs3* allele, suggesting higher grain yield and better grain quality could be obtained simultaneously by the proper allele combination. In the future, using the marker set of 14 grain size genes, it will be important to introgress and pyramid the different alleles into the same *japonica* background and explore the best allele combination to design cultivars with a balance between grain yield and grain quality.

Conclusion

In this study, 21 functional markers were developed for 14 genes regulating grain size and their ability to discriminate different alleles was clarified. A large diversity panel was genotyped with this marker set to determine the allele distribution of the different FNPs. Furthermore, the panel was evaluated for four grain size traits at two planting locations and the contribution of the FNPs to these traits was compared in the *indica* or *japonica* subspecies. Eleven markers were found to be significantly associated with grain size traits, and allele 2 of *GW5_InDel* could sustain the effects of several other loci. QTL analysis was performed to test the effect of seven polymorphic markers between two cultivars and confirmed the major contribution of *GW5* in determining grain size. This marker set may facilitate both association and linkage analyses for future genetic study, and provides efficient tools for the rational design of grain size as part rice breeding strategies in the future.

Methods

Plant Materials

Of the 261 rice cultivars evaluated in this study, 257 were part of the Rice Diversity Panel 1 (RDP1) and obtained from the U.S. Department of Agriculture Genetic Stocks-*Oryza* (Eizenga et al. 2014). The seeds of the RDP1 accession seeds were planted at the Songjiang Experiment Station near Shanghai, China (N 31.03°, E 121.22°) and Lingshui Experiment Station near Hainan, China (N 18.25°, E 109.50°) for trait analysis. Four additional cultivars obtained from China National Rice Research Institute and Zhejiang Academy of Agricultural Sciences were only used to evaluate the marker quality, included *indica* cultivars YYP1 and Baodali and *japonica* cultivars Daohuaxiang and WY3. YYP1 is a new plant type landrace with strong culms and large panicles. Daohuaxiang is a modern elite cultivar, widely planted in China. WY3 and Baodali are cultivars with notably large grains that contain the superior alleles *GW2/GL3.1* and *GS2*, respectively. The four cultivars are available from Dr. Lin Zhang upon reasonable request for research purposes. YYP1 was crossed with Nipponbare to generate F₁ and then selfed to generate F₂ seeds, and F₂ seeds were planted at Shanghai for QTL analysis.

Validation of Allele Variations

Sanger sequencing was performed for some loci to confirm the sequence variation, including the *TGW6* coding region and *GW6a* promoter region in Kasalath, the coding region of *GW2* and *GL3.1* in WY3, and the duplication border of *GL7* in Katy. Using randomly selected RDP1 accessions, four accessions were sequenced in the 5'UTR region of *GLW7* and 43 in the intron 4-exon 5 region of *GS3*. For the remaining allele variations, the sequences were retrieved either from the published data or RiceVarMap2 database (<http://ricevarmap.ncpgr.cn/v2/>). The candidate functional variations and their flanking sequences were aligned to the reference sequence of Nipponbare (NIP) to locate their positions.

Development of Allele Specific Markers

For InDel markers, the primers were designed by software Premier primer 5 directly, with a pair of primers spanning the InDel region, and the amplicon size was set to no more than ten times that of the InDel. As the mutation of *qLGY3* led to the loss of the StyI restriction enzyme site, a CAPS marker was designed with primers flanking this mutation. The dCAPS Finder 2.0 tool (<http://helix.wustl.edu/dcaps/dcaps.html>) was used to design dCAPS markers with permission of one or two mismatches on the primers, and forward primers with inexpensive restriction enzyme sites were selected. To match with the forward primers, Premier Primer 5 was used to select reverse primers. As dCAPS Finder 2.0

cannot recognize small InDels, we inputted the modified SNP-type sequence to facilitate enzyme site finding.

DNA Extraction, PCR and Gel Electrophoresis

The DNA was extracted by the mini-prep protocol following a previous publication (Xu et al. 2004). A 20 µL system with 2 µL of template DNA, 2 µL of dNTPs (2 mM), 1 µL of each primer (10 µM), 0.3 µL of Taq polymerase, 2 µL of 10 × PCR buffer (Dingguo Biotech, Beijing) and 12.7 µL of distilled water was subjected to PCR for most of the markers. For the GW5_InDel and GS3_SNP1 markers, 10 µL of 2 × PCR premixture (Yisheng Biotech, Shanghai) was used because it amplified better, and supplemented with 2 µL of template DNA, 1 µL of each primer (10 µM) and 6 µL of distilled water. The PCR program was set as follows: an initial denaturation at 94 °C for 3 min, 35 cycles of denaturation at 94 °C for 10 s, annealing at the proper temperature for 20 s, extension at 72 °C for 1 min per kb, and a final extension at 72 °C for 5 min. For dCAPS or CAPS markers, 10 µL of PCR products were used for enzyme analysis with 2 µL of 10 × reaction buffer, 0.3 µL of restriction enzyme (Thermo Fisher Scientific) and 7.7 µL of distilled water to obtain the 20 µL system. After digestion at 37 °C for one hour, 10 µL of product was used for electrophoresis. During gel electrophoresis, 1% agarose gel was used for GW5_InDel and GL7_dup markers, and 3% agarose gel was used for the remaining markers.

Analysis of Grain Size Related Traits

The seeds of RDP1 accessions and YYP1/NIP F₂ progeny were harvested at full maturity, and approximately 100 filled grains of each were scattered evenly on a white board equipped with a footlight and imaged with a high-speed scanner (Eloam S500, Shenzhen Eloam Technology Co.,Ltd., Shenzhen P.R.C). The images were analyzed by the SC-G program package (Wanshen, Hangzhou P.R.C), which automatically generated the grain number, grain length, grain width and length to width ratio information. The same set of grains was then weighed, and 1000 grain weight was calculated as follows: (grain weight/grain number) * 1000. The Pearson's correlation coefficients among traits at two locations were calculated by the R package "corrplot", and the heatmap was drawn accordingly.

Trait-Marker Association Analysis

The genotypes of all markers were assigned to each accession of the RDP1 and the allele frequency was calculated accordingly. To facilitate trait comparison, we grouped the RDP1 into *indica* and *japonica* subspecies based on the published population structure (McCouch et al. 2016). The *indica* subspecies included the *admix-indica*, *aus* and *indica* subpopulations and the *japonica* subspecies the *admix-japonica*, *temperate japonica* and

tropical japonica subpopulations. Twenty-four accessions were dropped for analysis, including five admixed accessions, eight *aromatic* accessions and eleven accessions without subpopulation assignment. Either the *indica* or *japonica* subspecies was separated into two allele types for each marker, and only markers with more than 15 accessions for minor alleles in one subspecies were chosen to compare the allele effect. Student's *t* test was performed by Graphpad Prism 6 software to calculate the significance level of trait differences between alleles, and the trait change rate between alleles was calculated as follows: (allele 2 value – allele 1 value)/allele 1 value. To test the possible interaction between *GW5* and other loci, six markers showing a stable effect on grain width control were selected and each marker will form a four allele combination with the *GW5* marker. The trait value of four allele combinations was subjected to one-way ANOVA, followed by Tukey's test for multiple comparisons.

QTL Analysis

One hundred nineteen individuals of the F_2 population from YYP1 and NIP were genotyped by 123 markers including seven of the newly developed functional markers for grain size (*GW2_SNP*, *GS3_SNP2*, *GS5_InDel*, *GW5_InDel*, *GLW7_InDel*, *GL7_InDel2* and *GW8_InDel*), 40 InDels, 62 SSRs and 14 SNPs, which were evenly distributed on the 12 chromosomes. QTL analysis was conducted using QTL IciMapping 4.2 software (Meng et al. 2015), and the genetic distance between markers was calculated by the Kosambi mapping function of the software. The inclusive composite interval mapping of additive (ICIM-ADD) QTL method was used to detect the QTLs. A significant logarithm of the odds (LOD) threshold value was set as 3 for each trait to claim the existence of QTLs. Locations of QTLs were expressed by flanking markers.

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s12284-020-00427-y>.

Additional file 1: Figure S1. Alignment of the *GS3* sequence covering the 3-bp deletion and the linked SNPs from 43 randomly selected Rice Diversity Panel 1 accessions. Note that the SNPs are absolutely linked with the 3-bp deletion.

Additional file 2: Figure S2. Alignment of the *GLW7* sequence covering two types of deletions from four randomly selected Rice Diversity Panel 1 accessions.

Additional file 3: Figure S3. Sketch map illustrating the restriction enzyme sites induced by the forward dCAPS primers. The primer sequences are highlighted in blue, and nucleotide substitutions generating enzyme sites are highlighted in purple. The sequence context of each enzyme site is underlined. Allele 1 stands for the NIP-type sequence, while allele 2 stands for the non-NIP type.

Additional file 4: Table S1. Information on 14 grain size genes for marker development.

Additional file 5: Table S2. Origin and population structure of the Rice Diversity Panel 1 accessions included in the study based on McCouch et al. (2016).

Additional file 6: Table S3. Phenotypic data for the Rice Diversity Panel 1 accessions grown at the Songjiang Experiment Station near Shanghai and the Lingshui Experiment Station near Hainan, China.

Additional file 7: Table S4. Allele type of the Rice Diversity Panel 1 accessions genotyped by 20 functional markers.

Additional file 8: Table S5. Information on 123 markers used for QTL analysis of the YYP1/NIP F_2 population.

Additional file 9: Table S6. Effects and locations of the QTLs identified from the YYP1/NIP F_2 population.

Abbreviations

PCR: Polymerase chain reaction; FNPs: Functional nucleotide polymorphisms; QTL: quantitative trait locus; MAS: molecular assisted selection; GWAS: Genome-wide association studies; InDel: Insertion and deletion; SNP: single nucleotide polymorphism; UTR: untranslated region; NIP: Nipponbare

Acknowledgements

RDP1 accessions (*Oryza sativa*) used in this study were obtained from the USDA-ARS, Dale Bumpers National Rice Research Center, Stuttgart, Arkansas, Genetic Stocks Oryza Collection (www.ars.usda.gov/GSOR).

Authors' Contributions

LZ designed the research and wrote the paper; LZ and BM performed most of the experiments and analyzed the data; ZB, XL, CZ, JL, QL participated in the data preparation and analysis. QL and ZH oversaw and improved the manuscript. The author(s) read and approved the final manuscript.

Funding

This work was supported by the grants from the National Natural Science Foundation of China (31600990 and 31871217), Project of National Key Laboratory of Plant Molecular Genetics, the "Lv yang jin feng project" of Yangzhou City, the Key Research and Development Program of Jiangsu Province (BE2018357) and the Postgraduate Research & Practice Innovation Program of Jiangsu Province (SJCX19_0894).

Availability of Data and Materials

The datasets supporting the conclusions of this article are provided in the article and its additional files. The RDP1 accessions can be ordered from GSOR, and the remaining four rice cultivars are available from Dr. Lin Zhang upon reasonable request.

Ethics Approval and Consent to Participate

Not applicable.

Consent for Publication

Not applicable.

Competing Interests

The authors declare no competing financial interests.

Author details

¹Jiangsu Key Laboratory of Crop Genomics and Molecular Breeding /Joint International Research Laboratory of Agriculture and Agri-Product Safety of the Ministry of Education, Yangzhou University, Yangzhou 225009, China. ²Jiangsu Co-Innovation Center for Modern Production Technology of Grain Crops, Jiangsu Key Laboratory of Crop Genetics and Physiology, Yangzhou University, Yangzhou 225009, China. ³School of Life Science and Technology, Shanghai Tech University, Shanghai 201210, China. ⁴National Key Laboratory of Plant Molecular Genetics, CAS Center for Excellence in Molecular Plant Sciences, Shanghai Institute of Plant Physiology & Ecology, Chinese Academy of Sciences, Shanghai 200032, China.

Received: 27 May 2020 Accepted: 2 September 2020

Published online: 09 September 2020

References

- Ali ML, McClung AM, Jia MH, Kimball JA, McCouch SR, Eizenga GC (2011) A rice diversity panel evaluated for genetic and agro-morphological diversity between subpopulations and its geographic distribution. *Crop Sci* 51:2021–2035
- Bagge M, Xia X, Lubberstedt T (2007) Functional markers in wheat. *Curr Opin Plant Biol* 10:211–216
- Calingacion M, Laborte A, Nelson A, Resurreccion A, Concepcion JC, Daygon VD, Mumm R, Reinke R, Dipti S, Bassinello PZ, Manful J, Sophany S, Lara KC, Bao JS, Xie LH, Loaiza K, El-hissewiy A, Gayin J, Sharma N, Rajeswari S, Manonmani S, Rani NS, Kota S, Indrasari SD, Habibi F, Hosseini M, Tavasoli F, Suzuki K, Umemoto T, Boualaphanh C, Lee HH, Hung YP, Ramli A, Aung PP, Ahmad R, Wattoo JI, Bandonill E, Romero M, Brites CM, Hafeel R, Lur HS, Cheapun K, Jongdee S, Blanco P, Bryant R, Lang NT, Hall RD, Fitzgerald M (2014) Diversity of global rice markets and the science required for consumer-targeted rice breeding. *PLoS One* 9:e85106. <https://doi.org/10.1371/journal.pone.0085106>
- Che RH, Tong HN, Shi BH, Liu YQ, Fang SR, Liu DP, Xiao YH, Hu B, Liu LC, Wang HR, Zhao MF, Chu CC (2016) Control of grain size and rice yield by *GL2*-mediated brassinosteroid responses (2015). *Nat Plants* 2:15195. <https://doi.org/10.1038/nplants.2015.195>
- Dong Q, Zhang ZH, Wang LL, Zhu YJ, Fan YY, Mou TM, Ma LY, Zhuang JY (2018) Dissection and fine-mapping of two QTL for grain size linked in a 460-kb region on chromosome 1 of rice. *Rice* 11:44. <https://doi.org/10.1186/s12284-018-0236-z>
- Drenkard E, Richter BG, Rozen S, Stutus LM, Angell NA, Mindrinos M, Cho RJ, Oefner PJ, Davis RW, Ausubel FM (2000) A simple procedure for the analysis of single nucleotide polymorphisms facilitates map-based cloning in *Arabidopsis*. *Plant Physiol* 124:1483–1492
- Duan PG, Ni S, Wang JM, Zhang BL, Xu R, Wang YX, Chen HQ, Zhu XD, Li YH (2016) Regulation of *OsGRF4* by *OsmiR396* controls grain size and yield in rice. *Nat Plants* 2:15203. <https://doi.org/10.1038/nplants.2015.203>
- Duan PG, Xu JS, Zeng DL, Zhang BL, Geng MF, Zhang GZ, Huang K, Huang LJ, Xu R, Ge S, Qian Q, Li YH (2017) Natural variation in the promoter of *GSE5* contributes to grain size diversity in rice. *Mol Plant* 10:685–694
- Eizenga GC, Ali ML, Bryant RJ, Yeater KM, McClung AM, McCouch SR (2014) Registration of the Rice diversity panel 1 for genomewide association studies. *J Plant Regist* 8:109–116
- Fan CH, Xing YZ, Mao HL, Lu TT, Han B, Xu CG, Li XH, Zhang QF (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
- Feng XM, Wang C, Nan JZ, Zhang XH, Wang RS, Jiang GQ, Yuan QB, Lin SY (2017) Updating the elite rice variety Kongyu 131 by improving the *Gn1a* locus. *Rice* 10:35. <https://doi.org/10.1186/s12284-017-0174-1>
- Hu J, Wang YX, Fang YX, Zeng LJ, Xu J, Yu HP, Shi ZY, Pan JJ, Zhang D, Kang SJ, Zhu L, Dong GJ, Guo LB, Zeng DL, Zhang GH, Xie LH, Xiong GS, Li JY, Qian Q (2015) A rare allele of *GS2* enhances grain size and grain yield in rice. *Mol Plant* 8:1455–1465
- Hu W, Zhou T, Wang P, Wang B, Song J, Han Z, Chen L, Liu K, Xing Y (2020) Development of whole-genome agarose-resolvable LInDel markers in rice. *Rice (N Y)* 13:1. <https://doi.org/10.1186/s12284-019-0361-3>
- Hu ZJ, Lu SJ, Wang MJ, He HH, Sun L, Wang HR, Liu XH, Jiang L, Sun JL, Xin XY, Kong W, Chu CC, Xue HW, Yang JS, Luo XJ, Liu JX (2018) A novel QTL *qTGW3* encodes the GSK3/SHAGGY-like kinase *OsGSK5/OsSK41* that interacts with *OsARF4* to negatively regulate grain size and weight in rice. *Mol Plant* 11:736–749
- Huang XH, Wei XH, Sang T, Zhao QA, Feng Q, Zhao Y, Li CY, Zhu CR, Lu TT, Zhang ZW, Li M, Fan DL, Guo YL, Wang A, Wang L, Deng LW, Li WJ, Lu YQ, Weng QJ, Liu KY, Huang T, Zhou TY, Jing YF, Li W, Lin Z, Buckler ES, Qian QA, Zhang QF, Li JY, Han B (2010) Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat Genet* 42:961–976
- Ishimaru K, Hirotsu N, Madoka Y, Murakami N, Hara N, Onodera H, Kashiwagi T, Ujicie K, Shimizu B, Onishi A, Miyagawa H, Katoh E (2013) Loss of function of the IAA-glucose hydrolase gene *TGW6* enhances rice grain weight and increases yield. *Nat Genet* 45:707–711
- Jiang HC, Li Z, Liu J, Shen ZK, Gao GJ, Zhang QL, He YQ (2019) Development and evaluation of improved lines with broad-spectrum resistance to rice blast using nine resistance genes. *Rice* 12:29. <https://doi.org/10.1186/s12284-019-0292-z>
- Jiang JF, Mou TM, Yu HH, Zhou FS (2015) Molecular breeding of thermo-sensitive genic male sterile (TGMS) lines of rice for blast resistance using *Pi2* gene. *Rice* 8:11. <https://doi.org/10.1186/s12284-015-0048-3>
- Kim SR, Ramos J, Ashikari M, Virk PS, Torres EA, Nissila E, Hechanova SL, Mauleon R, Jena KK (2016) Development and validation of allele-specific SNP/indel markers for eight yield-enhancing genes using whole-genome sequencing strategy to increase yield potential of rice, *Oryza sativa* L. *Rice* 9:12. <https://doi.org/10.1186/s12284-016-0084-7>
- Li YB, Fan CC, Xing YZ, Jiang YH, Luo LJ, Sun L, Shao D, Xu CJ, Li XH, Xiao JH, He YQ, Zhang QF (2011) Natural variation in *GS5* plays an important role in regulating grain size and yield in rice. *Nat Genet* 43:1266–1269
- Liu J, Huang S, Sun M, Liu S, Liu Y, Wang W, Zhang X, Wang H, Hua W (2012a) An improved allele-specific PCR primer design method for SNP marker analysis and its application. *Plant Methods* 8:34. <https://doi.org/10.1186/1746-4811-8-34>
- Liu JF, Chen J, Zheng XM, Wu FQ, Lin QB, Heng YQ, Tian P, Cheng ZJ, Yu XW, Zhou KN, Zhang X, Guo XP, Wang JL, Wang HY, Wan JM (2017) *GW5* acts in the brassinosteroid signalling pathway to regulate grain width and weight in rice. *Nat Plants* 3:17043. <https://doi.org/10.1038/nplants.2017.43>
- Liu Q, Han RX, Wu K, Zhang JQ, Ye YF, Wang SS, Chen JF, Pan YJ, Li Q, Xu XP, Zhou JW, Tao DY, Wu YJ, Fu XD (2018) G-protein beta gamma subunits determine grain size through interaction with MADS-domain transcription factors in rice. *Nat Commun* 9:852. <https://doi.org/10.1038/s41467-018-03047-9>
- Liu Y, He Z, Appels R, Xia X (2012b) Functional markers in wheat: current status and future prospects. *Theor Appl Genet* 125:1–10
- Liu YL, Chen LM, Liu YQ, Dai HM, He J, Kang HY, Pan G, Huang J, Qiu ZY, Wang Q, Hu JL, Liu LL, Chen YZ, Cheng XN, Jiang L, Wan JM (2016) Marker assisted pyramiding of two brown planthopper resistance genes, *Bph3* and *Bph27* (*t*), into elite rice cultivars. *Rice* 9:27. <https://doi.org/10.1186/s12284-016-0096-3>
- Lu L, Shao D, Qiu XJ, Sun L, Yan WH, Zhou XC, Yang L, He YQ, Yu SB, Xing YZ (2013) Natural variation and artificial selection in four genes determine grain shape in rice. *New Phytol* 200:1269–1280
- Mao HL, Sun SY, Yao JL, Wang CR, Yu SB, Xu CG, Li XH, Zhang QF (2010) Linking differential domain functions of the *GS3* protein to natural variation of grain size in rice. *P Natl Acad Sci USA* 107:19579–19584
- McClung AM, Edwards JD, Jia MH, Huggins TD, Bockelman HE, Ali ML, Eizenga GC (2020) Enhancing the searchability, breeding utility and efficient management of germplasm accessions in the USDA-ARS rice collection. *Crop Sci*. <https://doi.org/10.1002/csc2.20256>
- McCouch SR, Wright MH, Tung CW, Maron LG, McNally KL, Fitzgerald M, Singh N, DeClerck G, Agosto-Perez F, Korniliev P, Greenberg AJ, Naredo ME, Mercado SM, Harrington SE, Shi Y, Branchini DA, Kuser-Falcao PR, Leung H, Ebana K, Yano M, Eizenga G, McClung A, Mezey J (2016) Open access resources for genome-wide association mapping in rice. *Nat Commun* 7:10532. <https://doi.org/10.1038/ncomms10532>
- Meng L, Li HH, Zhang LY, Wang JK (2015) QTL IciMapping: integrated software for genetic linkage map construction and quantitative trait locus mapping in biparental populations. *Crop J* 3:269–283
- Nan JZ, Feng XM, Wang C, Zhang XH, Wang RS, Liu JX, Yuan QB, Jiang GQ, Lin SY (2018) Improving rice grain length through updating the *GS3* locus of an elite variety Kongyu 131. *Rice* 11:21. <https://doi.org/10.1186/s12284-018-0217-2>
- Nelson JC, McClung AM, Fjellstrom RG, Moldenhauer KAK, Boza E, Jodari F, Oard JH, Linscombe S, Scheffler BE, Yeater KM (2011) Mapping QTL main and interaction influences on milling quality in elite US rice germplasm. *Theor Appl Genet* 122:291–309
- Qi P, Lin YS, Song XJ, Shen JB, Huang W, Shan JX, Zhu MZ, Jiang LW, Gao JP, Lin HX (2012) The novel quantitative trait locus *GL3.1* controls rice grain size and yield by regulating Cyclin-T1;3. *Cell Res* 22:1666–1680
- Qiu XJ, Gong R, Tan YB, Yu SB (2012) Mapping and characterization of the major quantitative trait locus *qSS7* associated with increased length and decreased width of rice seeds. *Theor Appl Genet* 125:1717–1726
- Segami S, Yamamoto T, Oki K, Noda T, Kanamori H, Sasaki H, Mori S, Ashikari M, Kitano H, Katayose Y, Iwasaki Y, Miura K (2016) Detection of novel QTLs regulating grain size in extra-large grain rice (*Oryza sativa* L.) lines. *Rice* 9:34. doi: <https://doi.org/10.1186/s12284-016-0109-2>
- 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

- Si LZ, Chen JY, Huang XH, Gong H, Luo JH, Hou QQ, Zhou TY, Lu TT, Zhu JJ, Shangguan YY, Chen EW, Gong CX, Zhao Q, Jing YF, Zhao Y, Li Y, Cui LL, Fan DL, Lu YQ, Weng QJ, Wang YC, Zhan QL, Liu KY, Wei XH, An K, An G, Han B (2016) *OsSPL13* controls grain size in cultivated rice. *Nat Genet* 48:447–456
- Song XJ, Huang W, Shi M, Zhu MZ, Liu HX (2007) A QTL for rice grain width and weight encodes a previously unknown RING-type E3 ubiquitin ligase. *Nat Genet* 39:623–630
- Song XJ, Kuroha T, Ayano M, Furuta T, Nagai K, Komeda N, Segami S, Miura K, Ogawa D, Kamura T, Suzuki T, Higashiyama T, Yamasaki M, Mori H, Inukai Y, Wu JZ, Kitano H, Sakakibara H, Jacobsen SE, Ashikari M (2015) Rare allele of a previously unidentified histone H4 acetyltransferase enhances grain weight, yield, and plant biomass in rice. *P Natl Acad Sci USA* 112:76–81
- Sun LJ, Li XJ, Fu YC, Zhu ZF, Tan LB, Liu FX, Sun XY, Sun XW, Sun CQ (2013) *GS6*, a member of the GRAS gene family, negatively regulates grain size in rice. *J Integr Plant Biol* 55:938–949
- Takano-Kai N, Jiang H, Powell A, McCouch S, Takamura I, Furuya N, doi K, Yoshimura A (2013) Multiple and independent origins of short seeded alleles of *GS3* in rice. *Breeding Sci* 63:77–85
- Tan YF, Xing YZ, Li JX, Yu SB, Xu CG, Zhang QF (2000) Genetic bases of appearance quality of rice grains in Shanyou 63, an elite rice hybrid. *Theor Appl Genet* 101:823–829
- Wang CR, Chen S, Yu SB (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 HB, Ye ST, Mou TM (2016) Molecular breeding of rice restorer lines and hybrids for brown planthopper (BPH) resistance using the *Bph14* and *Bph15* genes. *Rice* 9:53. <https://doi.org/10.1186/s12284-016-0126-1>
- Wang RS, Jiang GQ, Feng XM, Nan JZ, Zhang XH, Yuan QB, Lin SY (2019) Updating the genome of the elite rice variety Kongyu131 to expand its ecological adaptation region. *Front Plant Sci* 10:288. <https://doi.org/10.3389/fpls.2019.00288>
- Wang SK, Li S, Liu Q, Wu K, Zhang JQ, Wang SS, Wang Y, Chen XB, Zhang Y, Gao CX, Wang F, Huang HX, Fu XD (2015a) The *OsSPL16-GW7* regulatory module determines grain shape and simultaneously improves rice yield and grain quality. *Nat Genet* 47:949–954
- Wang SK, Wu K, Yuan QB, Liu XY, Liu ZB, Lin XY, Zeng RZ, Zhu HT, Dong GJ, Qian Q, Zhang GQ, Fu XD (2012) Control of grain size, shape and quality by *OsSPL16* in rice. *Nat Genet* 44:950–954
- Wang Y, Jiang WH, Liu HM, Zeng Y, Du B, Zhu LL, He GC, Chen RZ (2017) Marker assisted pyramiding of *Bph6* and *Bph9* into elite restorer line 93-11 and development of functional marker for *Bph9*. *Rice* 10:51. <https://doi.org/10.1186/s12284-017-0194-x>
- Wang YX, Xiong GS, Hu J, Jiang L, Yu H, Xu J, Fang YX, Zeng LJ, Xu EB, Xu J, Ye WJ, Meng XB, Liu RF, Chen HQ, Jing YH, Wang YH, Zhu XD, Li JY, Qian Q (2015b) Copy number variation at the *GL7* locus contributes to grain size diversity in rice. *Nat Genet* 47:944–948
- Weng JF, Gu SH, Wan XY, Gao H, Guo T, Su N, Lei CL, Zhang X, Cheng ZJ, Guo XP, Wang JL, Jiang L, Zhai HQ, Wan JM (2008) Isolation and initial characterization of *GW5*, a major QTL associated with rice grain width and weight. *Cell Res* 18:1199–1209
- Xia D, Zhou H, Liu RJ, Dan WH, Li PB, Wu B, Chen JX, Wang LQ, Gao GJ, Zhang QL, He YQ (2018) *GL3.3*, a novel QTL encoding a GSK3/SHAGGY-like kinase, epistatically interacts with *GS3* to produce extra-long grains in rice. *Mol Plant* 11:754–756
- Xia D, Zhou H, Qiu L, Jiang HC, Zhang QL, Gao GJ, He YQ (2017) Mapping and verification of grain shape QTLs based on an advanced backcross population in rice. *PLoS One* 12:e0187553. <https://doi.org/10.1371/journal.pone.0187553>
- Xing YZ, Zhang QF (2010) Genetic and molecular bases of rice yield. *Annu Rev Plant Biol* 61:421–442
- Xu YH, Zhu YY, Zhou HC, Li Q, Sun ZX, Liu YG, Lin HX, He ZH (2004) Identification of a 98-kb DNA segment containing the rice *Eui* gene controlling uppermost internode elongation, and construction of a TAC transgene sublibrary. *Mol Gen Genomics* 272:149–155
- Yamaki S, Ohyanagi H, Yamasaki M, Eiguchi M, Miyabayashi T, Kubo T, Kurata N, Nonomura KI (2013) Development of INDEL markers to discriminate all genome types rapidly in the genus *Oryza*. *Breeding Sci* 63:246–254
- Ying JZ, Ma M, Bai C, Huang XH, Liu JL, Fan YY, Song XJ (2018) *TGW3*, a major QTL that negatively modulates grain length and weight in rice. *Mol Plant* 11:750–753
- Yu JP, Miao JL, Zhang ZY, Xiong HY, Zhu XY, Sun XM, Pan YH, Liang YT, Zhang Q, Rehman RMA, Li JJ, Zhang HL, Li ZC (2018) Alternative splicing of *OsLG3b* controls grain length and yield in japonica rice. *Plant Biotechnol J* 16:1667–1678
- Zeng DL, Tian ZX, Rao YC, Dong GJ, Yang YL, Huang LC, Leng YJ, Xu J, Sun C, Zhang GH, Hu J, Zhu L, Gao ZY, Hu XM, Guo LB, Xiong GS, Wang YH, Li JY, Qian Q (2017) Rational design of high-yield and superior-quality rice. *Nat Plants* 3:17031. <https://doi.org/10.1038/nplants.2017.31>
- Zhang XJ, Wang JF, Huang J, Lan HX, Wang CL, Yin CF, Wu YY, Tang HJ, Qian Q, Li JY, Zhang HS (2012) Rare allele of *OsPPKL1* associated with grain length causes extra-large grain and a significant yield increase in rice. *P Natl Acad Sci USA* 109:21534–21539
- Zhao DS, Li QF, Zhang CQ, Zhang C, Yang QQ, Pan LX, Ren XY, Lu J, Gu MH, Liu QQ (2018) *GS9* acts as a transcriptional activator to regulate rice grain shape and appearance quality. *Nat Commun* 9:1240. <https://doi.org/10.1038/s41467-018-03616-y>
- Zhao K, Tung CW, Eizenga GC, Wright MH, Ali ML, Price AH, Norton GJ, Islam MR, Reynolds A, Mezey J, McClung AM, Bustamante CD, McCouch SR (2011) Genome-wide association mapping reveals a rich genetic architecture of complex traits in *Oryza sativa*. *Nat Commun* 2:467. <https://doi.org/10.1038/ncomms1467>

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen® journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)