plants

MDPI

Article

Genome-Wide Association Study of Rice Grain Shape and Chalkiness in a Worldwide Collection of Xian Accessions

Nansheng Wang [1,†], Huguang Chen [1,2,†], Yingzhi Qian [1,†], Zhaojie Liang [1], Guiqiang Zheng [1], Jun Xiang [1], Ting Feng [1], Min Li [1], Wei Zeng [1], Yaling Bao [1], Erbao Liu [1], Chaopu Zhang [1], Jianlong Xu [2,*] and Yingyao Shi [1,*]

[1] College of Agronomy, Anhui Agricultural University, Hefei 230036, China
[2] Institute of Crop Sciences, Chinese Academy of Agricultural Sciences, Beijing 100081, China
* Correspondence: xujlcaas@126.com (J.X.); shiyy123@163.com (Y.S.)
† These authors contributed equally to this work.

Abstract: Rice (*Oryza sativa* L.) appearance quality, which is mainly defined by grain shape and chalkiness, is an important target in rice breeding. In this study, we first re-sequenced 137 indica accessions and then conducted a genome-wide association study (GWAS) for six agronomic traits with the 2,998,034 derived single nucleotide polymorphisms (SNPs) by using the best linear unbiased prediction (BLUP) values for each trait. The results revealed that 195 SNPs had significant associations with the six agronomic traits. Based on the genome-wide linkage disequilibrium (LD) blocks, candidate genes for the target traits were detected within 100 kb upstream and downstream of the relevant SNP loci. Results indicate that six quantitative trait loci (QTLs) significantly associated with six traits (*qTGW4.1*, *qTGW4.2*, *qGL4.1*, *qGL12.1*, *qGL12.2*, *qGW2.1*, *qGW4.1*, *qGW6.1*, *qGW8.1*, *qGW8.2*, *qGW9.1*, *qGW11.1*, *qGLWR2.1*, *qGLWR2.2*, *qGLWR4.2*, *qPGWC5.1* and *qDEC6.1*) were identified for haplotype analysis. Among these QTLs, two (*qTGW4.2* and *qGW6.1*), were overlapped with *FLO19* and *OsbZIP47*, respectively, and the remaining four were novel QTLs. These candidate genes were further validated by haplotype block construction.

Keywords: rice; GWAS; QTLs; grain shape; chalkiness

check for updates

Citation: Wang, N.; Chen, H.; Qian, Y.; Liang, Z.; Zheng, G.; Xiang, J.; Feng, T.; Li, M.; Zeng, W.; Bao, Y.; et al. Genome-Wide Association Study of Rice Grain Shape and Chalkiness in a Worldwide Collection of Xian Accessions. *Plants* **2023**, *12*, 419. https://doi.org/10.3390/plants12030419

Academic Editors: Xiangqian Zhao, Jinsong Bao and Dali Zeng

Received: 19 December 2022
Revised: 11 January 2023
Accepted: 13 January 2023
Published: 17 January 2023

Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Rice (*Oryza sativa* L.) is a major staple crop in Asian countries, feeding more than half of the world's population. Higher yield and quality are two main objectives in rice breeding. Although different countries may have different preferences for rice quality, high-quality rice varieties are always favored by both consumers and producers. Rice consumers in India, southern China, Pakistan, Bangladesh, Sri Lanka, some southeast Asian countries and the USA prefer long and thin rice grains with a fluffy to firm texture and medium to high straight-chain starch contents. However, people in Japan, Korea and northern China prefer short, round, soft and sticky rice grains with low straight-chain starch content. The improvement of the quality of major crops through breeding of superior varieties with higher yield, nutrition and resistance is essential for adequate, reliable and sustainable food supply in the world [1,2]. Identification of genes associated with rice grain shape and chalkiness is important for the breeding of modern rice varieties with excellent rice quality.

With the rapid development of economy and the improvement of living standards, rice quality has become a major concern for many rice producers [3–7]. Rice grain quality comprises appearance, cooking, eating, and nutritional and milling quality, among which appearance quality is a key factor affecting its market acceptability [8]. Appearance quality is mainly represented by grain shape and chalkiness. Rice shape is generally described by grain length (GL), grain width (GW), grain thickness (GT) and grain length-to-width ratio (GLWR), and is closely related to grain weight [9,10]. Chalkiness is usually evaluated by the degree of endosperm chalkiness (DEC) and the percentage of grains with chalkiness (PGWC) [11]. Rice varieties with a DEC higher than 25% are generally unacceptable in

most world markets. Breeding for rice varieties with desirable appearance is a primary goal of rice breeders, which may be greatly facilitated by a better understanding of the genetic basis for these traits.

Rice grain size and chalkiness are determined by the interaction of genetic and environmental factors [12,13]. Traditional biparental mapping has localized many GL-, GW-, GT- and GLWR-related quantitative trait loci (QTLs) [14–16]. For example, to understand the genetic basis of rice appearance quality including grain size and chalkiness, Dai et al. bred a doubled-haploid population of two short-grain hybrid rice varieties, Chunjiang 06 (CJ06) and Rice Taichung Native One (TN1), in subtropical Hangzhou and tropical Hainan. Nineteen main-effect QTLs and nine epistatic interactions controlling grain chalkiness and grain shape were detected [17]. Mei et al. detected 40 QTLs associated with DEC, PGWC and endosperm transparency using three recombinant self-reproducing populations from a three-line cross combination [18]. Moreover, Bian et al. conducted a quantitative genetic analysis of a population consisting of 37 introgression lines (ILs) in two different environments. A total of 54 QTLs were detected on 11 chromosomes, 44 of which were associated with multiple traits [19]. Recently, a genome-wide association study (GWAS) has been employed to detect QTLs associated with quality traits in rice, which is much less time-consuming than traditional techniques and can locate more QTLs or alleles. For example, a previous study detected a total of 16 and 20 QTLs associated with grain appearance quality using single nucleotide polymorphism (SNP) and bin-GWAS methods, and identified dominant alleles for *GS3*, *GW5*, *GL3.1*, *GW7*, *Chalk5* and *qPGWC8.2* [20]. Zhong detected 152, 106 and 12 QTLs for three traits of yield, including GL, GW and GT, in 529 rice varieties using GWAS [21].

The genetic basis for grain shape and chalkiness in rice has been well-studied in the past decades [15,22], and many related genes have been identified and cloned, such as *GW2* [23], *GS3* [24], *GS5* [25], *qGL3* [26], *GW8* [27], *GL7* [28], *GW7* [29], *OsMAPK6* [30], *Chalk5* [31], *TGW6* [32], *GW5* [33], *qSW5* [34]. *GS3* [24] is the major gene controlling GL in rice. Mutation in the second exon will change the cysteine codon (TGC) to a stop codon (TGA) at the protein level, leading to diversity in rice GL. *GW5* [33] is an IQ calmodulin-binding motif family protein that regulates rice GW and thousand grain weight (TGW), and its mutation at the protein level can convert the cysteine codon (TGC) to a stop codon (TGA), resulting in diversity of GL in rice. The seeds of *GW5* loss-of-function mutant were wider than those of the wild type. *Chalk5* [31] is the major QTL regulating seed chalkiness, and also affects rice ears, rice yield and many other quality traits. It is an enzyme encoding a vesicular H++ translocating pyrophosphatase that combines inorganic pyrophosphatase hydrolysis with H+ translocating activity. Plants overexpressing *Chalk5* have higher chalkiness in the endosperm. Although a number of genes controlling quality traits have been identified, further research is needed to elucidate the molecular regulatory mechanisms of rice grain shape, yield and chalkiness.

Although many genes controlling rice grain shape and chalkiness have been cloned and identified, a number of QTLs controlling rice grain shape and grain weight were only detected by initial localization, and many more have not been cloned to date. The isolation of candidate genes by mapping-based cloning is very time-consuming and requires a long time to develop near-isogenic lines for precise localization. GWAS can decipher the relationships between traits and their causal genomic regions. The genetic basis for complex traits can be revealed by studying the diversity of phenotypes and genetic variations in a large number of unrelated relatives [35,36]. In this study, we re-sequenced 137 rice accessions worldwide and obtained approximately 2.89 million SNPs to analyze the genetic basis for rice appearance quality. Six agronomic traits (thousand grain weight (TGW), grain length (GL), grain width (GW), grain length to width ratio (GLWR), percentage of grains with chalkiness (PGWC), and degree of endosperm chalkiness (DEC)) of rice were investigated in three environments. To avoid false positives, we selected the best linear unbiased prediction (BLUP) values for the six agronomic traits within the three environments and conducted genetic diversity analysis for each trait. In addition, we

refined the number of candidate genes by combining haplotype block structure analysis with gene function annotation. The results provide an important genomic resource for the molecular breeding of rice and for studying the genetic basis for high yield and quality in rice.

2. Results

2.1. Phenotypic Variation and Correlation

In general, some traits appeared to be normally distributed, but other traits showed skewed distributions, especially GL, GLWR, PGWC and DEC (Figure S1). The panel showed large variations for all the measured traits. Significant variations between Hainan (HN) and Guangxi (GX) were observed for TGW, GL, GW, GLWR, PGWC and DEC; significant variations in TGW, GL, GW, PGWC and DEC were found between Hainan (HN) and Jiangxi (JX); and significant variations between GX and JX were observed for GL and GLWR (Figure 1a–e).

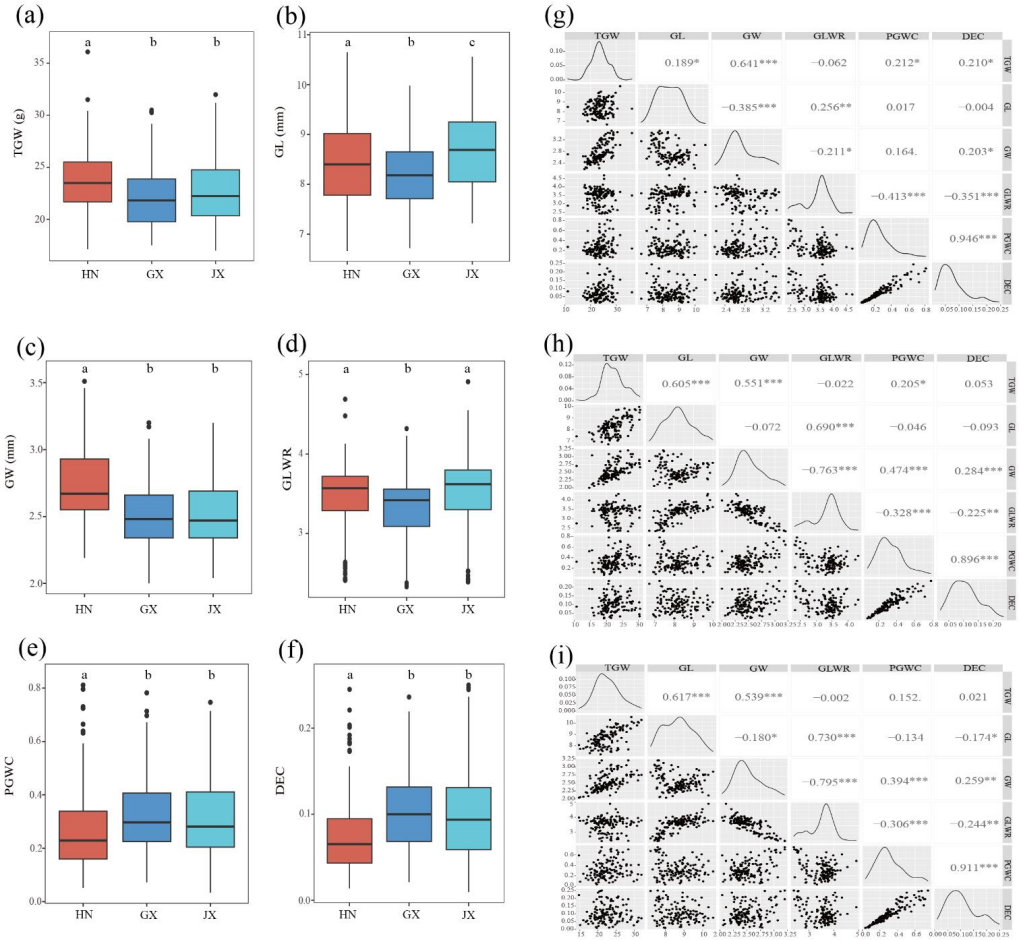


Figure 1. Box plots of six rice grain appearance and milling quality traits in three environments and phenotypic correlations of six traits in different environments. HN, Hainan; GX, Guangxi; JX, Jiangxi; (**a**) thousand grain weight; (**b**) grain length; (**c**) grain width; (**d**) grain length to width ratio; (**e**) percentage of grains with chalkiness; (**f**) degree of endosperm with chalkiness; (**g**) Hainan; (**h**) Guangxi; (**i**) Jiangxi; '*', '**', and '***' refer to significant correlations ($p < 0.05$, $p < 0.01$, and $p < 0.001$).

The TGW value ranged from 17.01 to 36.08 g, with an average value of 23.67, 22.37 and 22.68 g in HN, GX and JX, respectively. GL was both the highest (10.65 mm) and lowest (6.66 mm) in HN, and its average value in the 137 accessions was 8.43, 8.23 and 8.69 mm in HN, GX and JX, respectively. The GW value ranged from 2.00 to 3.51 mm, and was

averagely 2.75, 2.32 and 2.52 mm in HN, GX and JX, respectively. The mean of GLWR in the 137 accessions was 3.45, 3.31 and 3.51 in HN, GX and JX, respectively. The GLWR value was the lowest in GX (2.33) and the highest in JX (4.91). PGWC was averagely 0.27, 0.32 and 0.32 in HN, GX and JX, respectively, and ranged from 0.03 to 0.81.

The mean DEC in the 137 accessions was 0.08, 0.10 and 0.10 in HN, GX and JX, respectively. In addition, the minimum DEC value was found in GX and JX (0.01), and the maximum value was observed in GX and JX (0.25). The broad-sense heritability (H_B^2) averaged across the three environments was 0.41, 0.52, 0.43, 0.98, 0.78 and 0.67 for TGW, GL, GW, GLWR, PGWC and DEC, respectively (Table 1). ANOVA analysis indicated that the effects of accession, environment and their interaction were highly significant ($p < 0.001$) except for the effect of environment on TGW ($p = 0.0012$) (Table S2). The pairwise correlations between the measured traits were similar in the three environments. TGW was positively correlated with GL, GW, PGWC and DEC, but negatively correlated with GLWR. Positive correlations were observed between DEC, PGWC and GW. GL was negatively correlated with GW. These results demonstrated that rice grain traits are highly related to each other, providing important information for rice grain shape modification (Figure 1g–i).

Table 1. Statistics of TGW, GL, GW, GLWR, PGWC and DEC in different environments.

Phenotype	Env.	Mean ± SD	Max	Min	CV (%)	H_B^2
TGW (g)	HN	23.67 ± 3.22	36.08	17.15	13.62%	0.41
	GX	22.37 ± 3.16	30.50	17.53	14.13%	
	JX	22.68 ± 3.33	31.98	17.01	14.69%	
GL (mm)	HN	8.43 ± 0.77	10.65	6.66	9.10%	0.52
	GX	8.23 ± 0.72	9.98	6.72	8.77%	
	JX	8.69 ± 0.77	10.56	7.22	8.90%	
GW (mm)	HN	2.75 ± 0.31	3.51	2.19	11.37%	0.43
	GX	2.52 ± 0.25	3.20	2.00	9.85%	
	JX	2.52 ± 0.25	3.20	2.04	10.10%	
GLWR	HN	3.45 ± 0.45	4.69	2.41	13.15%	0.98
	GX	3.31 ± 0.42	4.32	2.33	12.65%	
	JX	3.51 ± 0.48	4.91	2.39	13.65%	
PGWC	HN	0.27 ± 0.16	0.81	0.05	59.11%	0.78
	GX	0.32 ± 0.14	0.78	0.07	43.07%	
	JX	0.32 ± 0.16	0.75	0.03	51.09%	
DEC	HN	0.08 ± 0.05	0.25	0.01	62.36%	0.67
	GX	0.10 ± 0.05	0.24	0.02	44.21%	
	JX	0.10 ± 0.06	0.25	0.01	55.31%	

2.2. Phylogenetic and Population Structure Analysis

The phylogenetic tree (Figure 2a) shows that the structure of the population used in this experiment is uniform without strong population stratification. Based on the SNPs, the Admixture [37] software was used to analyze the population structure. Cross-validation error analysis revealed that the error peak was the lowest at K = 4, indicating that the grouping was optimal (Figure 2c). Population structure analysis showed no obvious pedigree differentiation in the selected plant materials, confirming that they were suitable for subsequent GWAS analysis. The phylogenetic tree shows that the selected population could be divided into four subgroups, which verified the conclusion that K = 4 is the optimal result in the population structure analysis. Therefore, the Q matrix with K = 4 was selected for subsequent association analysis (Figure 2d). Based on the SNP data, R software was used to perform Principal Component Analysis (PCA) [38] to cluster the sample (Figure 2b). The results showed that the 137 materials were not clustered together and were scattered all over the place. The PCA results supported the evolutionary analysis, further confirming that there was a low degree of discreteness of the individual kinship in the population. Kinship, PCA and phylogenetic trees together assess population structure.

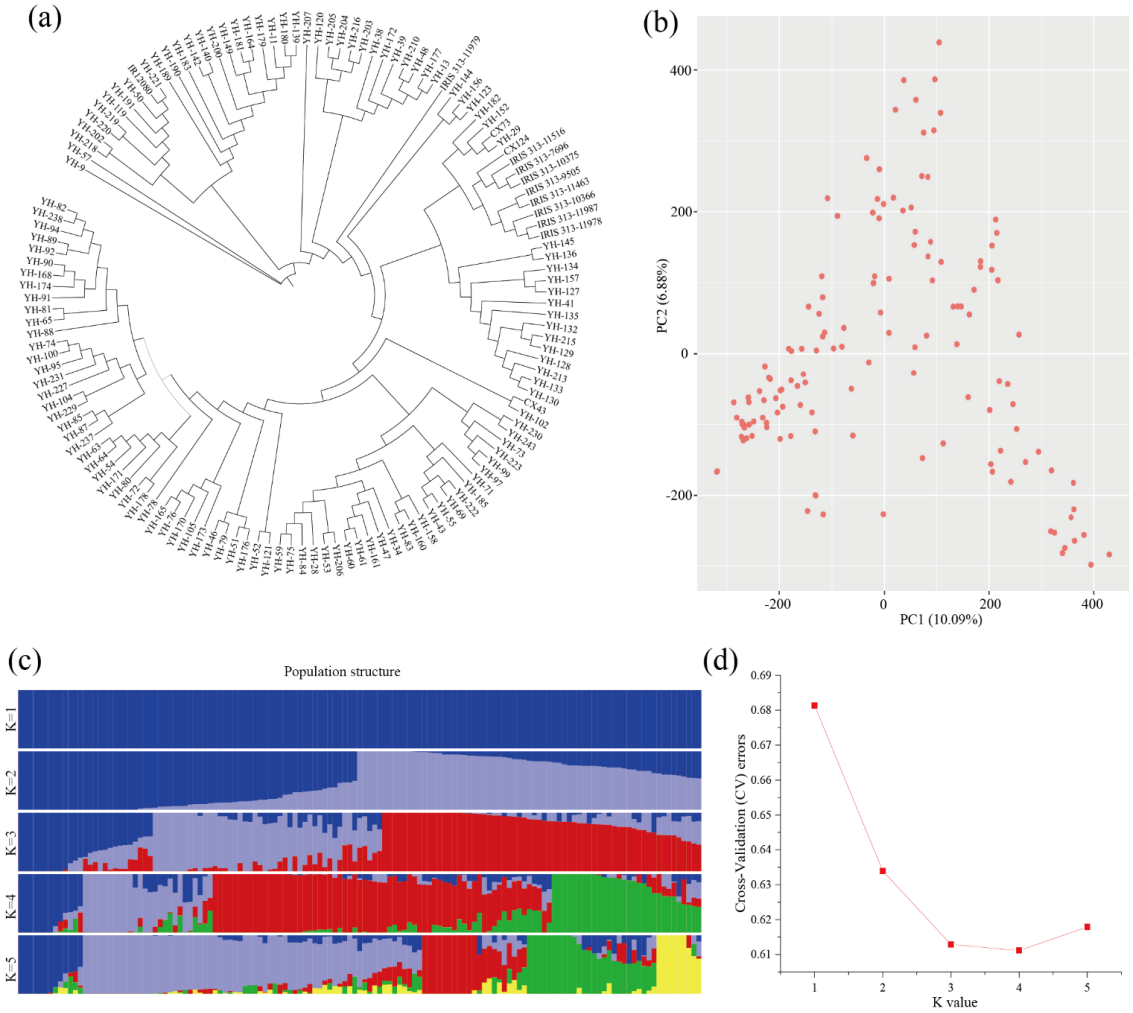


Figure 2. Genetic evolution of natural populations of indica rice. (**a**) Phylogenetic tree; each branch is a rice accession. (**b**) Principal component analysis on 2.89 million SNPs of 137 rice accessions. PC1 and PC2 refer to the first and second principal components, respectively. The numbers in parentheses refer to the proportion of variance explained by the corresponding axes. Red points represent the 137 rice accessions, with each point representing one rice accession. A shorter distance between the points indicates a closer relationship. (**c**) Cluster analysis results of population genotypes, in which each color represents a group and each row indicates a group value. (**d**) Cross validation error rate for each k value. Among them, K is the lowest when k is 4.

2.3. Genome-Wide LD Patterns and QTL Detection by GWAS

The maximum LD was 0.30 in the whole population. LD reached half of its initial value at around 92 kb (Figure S2).

In this study, the association analysis was performed using the BLUP method for each accession to reduce the environment effects and simplify the calculation. The general linear model (GLM) was used to conduct GWAS on grain weight, grain shape and chalkiness (TGW, GL, GW, GLWR, PGWC and DEC). Considering the decay distance of LD in rice, adjacent SNPs with spans less than 200 kb were defined as single QTLs, and the SNP with the lowest p value was taken as the lead SNP to reduce redundant association signals of different traits. Stringent criterion of $-\log10\ (p) > 5.4$ in the three environments was used to determine the association significance of grain weight, grain shape and chalkiness (Table S3). The results showed that 195 SNPs had significant associations with six agronomic traits. We successfully identified both known genes and previously reported QTLs from rice as well as some novel candidate loci in rice genome. The results revealed that two,

three, seven, five, six and four QTLs detected by GLM were significantly correlated with TGW, GL, GW, GLWR, PGWC and DEC, respectively (Table 2). A total of 379 candidate genes were obtained (Table S4). In general, two QTL regions (*qTGW4.1* and *qTGW4.2* on chromosome 4) were determined to be significantly correlated with TGW, which accounted for 17.7% and 20.4% of the phenotypic variance, respectively (Table 2, Figure 3a). The three QTLs associated with GL were *qGL4.1* on chromosome 4 and *qGL12.1* and *qGL12.2* on chromosome 12, which accounted for 18.0%, 15.7% and 15.7% of the phenotypic variance, respectively (Table 2, Figure 3b). Seven QTLs were significantly correlated with GW, including *qGW2.1* on chromosome 2, *qGW4.1* on chromosome 4, *qGW6.1* on chromosome 6, *qGW8.1* and *qGW8.2* on chromosome 8, *qGW9.1* on chromosome 9 and *qGW11.1* on chromosome 11. These QTLs explained 13.9–20.2% of the total phenotypic variance. (Table 2, Figure 3c). Five QTLs were significantly associated with GLWR, including *qGLWR2.1* and *qGLWR2.2* on chromosome 2, *qGLWR4.1* and *qGLWR4.2* on chromosome 4 and *qGLWR10.1* on chromosome 10, which accounted for 14.0–19.3% of the phenotypic variance (Table 2, Figure 3d). Six QTLs were significantly correlated with PGWC, including *qPGWC3.1* on chromosome 3, *qPGWC4.1* on chromosome 4, *qPGWC5.1* and *qPGWC5.2* on chromosome 5, *qPGWC9.1* on chromosome 9 and *qPGWC11.1* on chromosome 11, which accounted for 21.8%, 20.1%, 19.8%, 20.3%, 21.1% and 18.7% of the phenotypic variance, respectively (Table 2, Figure 3e). Four QTLs had significant correlations with DEC, including *qDEC1.1* on chromosome 1, *qDEC6.1* on chromosome 6, *qDEC11.1* on chromosome 11 and *qDEC12.1* on chromosome 12. They accounted for 16.3%, 17.0%, 17.3% and 16.6% of the total phenotypic variance, respectively (Table 2, Figure 3f).

Table 2. Twenty-seven QTLs with significant associations with TGW, GL, GW, GLWR, PGWC and DEC.

Trait	QTL	Chr	Lead SNP (bp)	R^2 (%)	*p* Value	Known Genes/QTLs
TGW	*qTGW4.1*	4	608,507	17.7%	1.22×10^{-6}	
	qTGW4.2	4	1,225,773	20.4%	1.64×10^{-6}	*FLO19* [39]
GL	*qGL4.1*	4	16,205,291	18.0%	7.28×10^{-7}	
	qGL12.1	12	225,117	15.7%	3.24×10^{-6}	
	qGL12.2	12	1,188,996	15.7%	3.47×10^{-6}	
GW	*qGW2.1*	2	26,502,715	16.4%	2.03×10^{-6}	
	qGW4.1	4	502,118	15.7%	3.70×10^{-6}	
	qGW6.1	6	8,695,272	13.9%	2.88×10^{-6}	*OsbZIP47* [40]
	qGW8.1	8	19,551,373	20.2%	2.32×10^{-6}	
	qGW8.2	8	28,472,120	17.8%	1.98×10^{-6}	
	qGW9.1	9	13,050,866	17.1%	1.15×10^{-6}	
	qGW11.1	11	8,695,392	16.8%	1.99×10^{-6}	
GLWR	*qGLWR2.1*	2	26,502,715	17.4%	1.07×10^{-6}	
	qGLWR2.2	2	35,438,743	17.5%	1.73×10^{-6}	
	qGLWR4.1	4	18,391,272	16.1%	3.43×10^{-6}	
	qGLWR4.2	4	29,199,879	14.0%	3.23×10^{-6}	
	qGLWR10.1	10	24,820,313	19.3%	2.36×10^{-7}	
PGWC	*qPGWC3.1*	3	5,812,574	21.8%	1.35×10^{-8}	
	qPGWC4.1	4	20,153,796	20.1%	3.73×10^{-7}	
	qPGWC5.1	5	267,911	19.8%	1.75×10^{-6}	
	qPGWC5.2	5	14,313,526	20.3%	4.10×10^{-8}	
	qPGWC9.1	9	12,763,755	21.1%	2.29×10^{-6}	
	qPGWC11.1	11	23,382,869	18.7%	1.13×10^{-6}	
DEC	*qDEC1.1*	1	42,431,688	16.3%	2.29×10^{-6}	
	qDEC6.1	6	27,376,022	17.0%	3.19×10^{-6}	
	qDEC11.1	11	7,169,791	17.3%	3.54×10^{-6}	
	qDEC12.1	12	17,698,258	16.6%	2.80×10^{-6}	

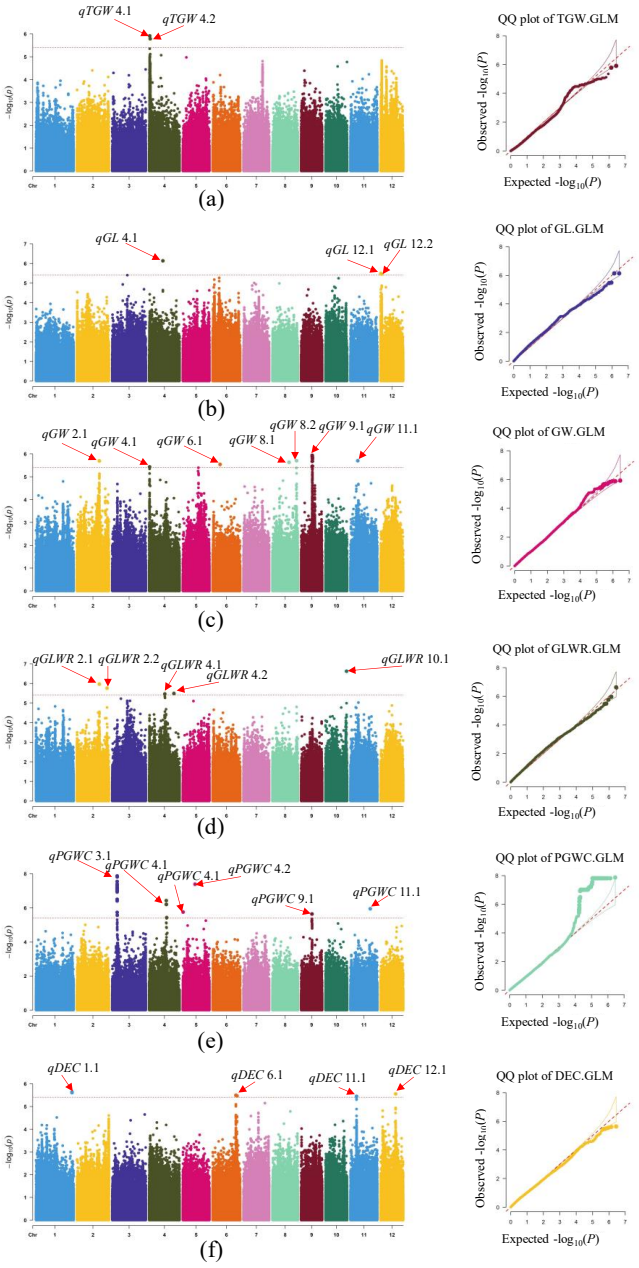


Figure 3. Genome-wide association plots of TGW, GL, GW, GLWR, PGWC and DEC in rice population plotted using the generalized linear model. The Manhattan map of genome-wide scans shows the −log10(p) values corresponding to the location of each of 12 chromosomes. Red solid lines represent the whole genome significant threshold $p = 3.9 \times 10^{-6}$. Red arrows indicate QTLs *qTGW4.1*(rs4_608507), *qTGW4.2*(rs4_1225773), *qGL4.1*(rs4_16205291), *qGL12.1*(rs12_225117), *qGL12.2*(rs12_1188996), *qGW2.1*(rs2_26502715), *qGW4.1*(rs4_502118), *qGW6.1*(rs6_8695272), *qGW8.1* (rs8_19551373), *qGW8.2*(rs8_28472120), *qGW9.1*(rs9_13050866), *qGW11.1*(rs11_8695392), *qGLWR2.1* (rs2_26502715), *qGLWR2.2*(rs2_35438743), *qGLWR4.1*(rs4_18391272), *qGLWR4.2*(rs4_29199879), *qGLWR10.1* (rs10_24820313), *qPGWC3.1*(rs3_5812574), *qPGWC4.1*(rs4_20153796), *qPGWC5.1*(rs5_267911), *qPGWC5.2* (rs5_14313526), *qPGWC9.1*(rs9_12763755), *qPGWC11.1*(rs11_23382869), *qDEC1.1*(rs1_42431688), *qDEC6.1*(rs6_27376022), *qDEC11.1*(rs11_7169791) and *qDEC12.1*(rs12_17698258) colocalized by GLM. The horizontal axis in the quantile–quantile (QQ) plot represents the expected value of the −log10 transformation, whereas the vertical axis indicates the observed value of the −log10 transformation. Manhattan plot and QQ plot of TGW (**a**), GL (**b**), GW (**c**), GLWR (**d**), PGWC (**e**) and DEC (**f**) in GLM.

2.4. Candidate Gene Identification and Haplotype Analysis

The twenty-seven identified QTLs were used for high-density association and gene-based haplotype analysis to identify the candidate genes. In the region of *qTGW4.2* (1.12–1.32 Mb on chromosome 4), the 2795 SNPs of ten genes were used for high-density association analysis. The annotated gene with the most significant hit was LOC_Os04g02900 (Figure 4b,c). Four major haplotypes were detected among the 137 accessions based on eight SNPs in the LOC_Os04g02900 promoter, eight SNPs in the exon, and ten SNPs in the intron. The mean TGW was 22.85, 22.86, 21.15 and 22.98 g for HapA, HapB, HapC and HapD, respectively (Figure 4a). Haplotype analysis of the whole population revealed that HapD had a significantly higher TGW than other three haplotypes. Significant differences in TGW were observed among the four haplotypes in the population (Figure 4d).

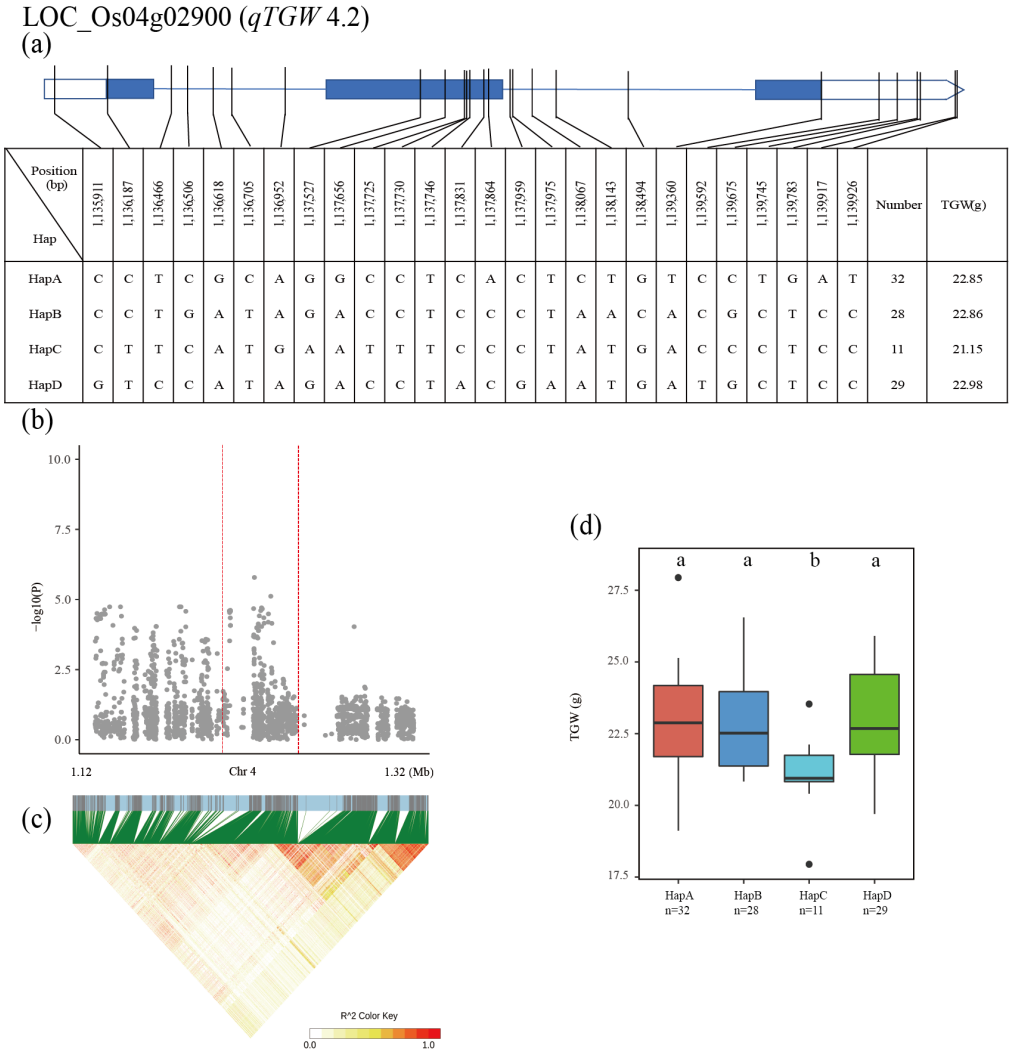


Figure 4. Identification of candidate genes for TGW. (**a**) Based on 26 SNPs in all evaluated rice accessions, four haplotypes of LOC_Os04g02900 were identified. In the gene structure diagram of LOC_Os04g02900 (http://rice.plantbiology.msu.edu, accessed on 15 May 2022), the promoter is indicated by white frame; the exon is represented by blue frame; and the intron and intergenic region are marked by blue lines. A thin black line represents the genomic location of each SNP. Haplotypes with fewer than 10 accessions are not shown. TGW (**b**) based on single polymorphism and LD heat map of local Manhattan map (**c**), around the peak on chromosome 4. Red dotted lines represent candidate regions for associated SNPs. Based on TGW (**d**) of LOC_Os04g02900 haplotype, differences between haplotypes were statistically analyzed using Tukey's test.

The QTL *qGL12.2* was identified in a 1.08–1.28 Mb region on chromosome 12, and the 1660 SNPs of 21 genes were used for high-density association analysis. The most significant hit was located in LOC_Os12g03040 (Figure 5b,c). Three major haplotypes were detected among the 137 accessions based on six SNPs in the LOC_Os12g03040 promoter, four SNPs in the exon, and two SNPs in the intron. The mean GL was 8.54, 8.57 and 8.21 mm for HapA, HapB and HapC, respectively (Figure 5a). HapB had a significantly higher GL than the other two haplotypes, and there was a significant difference between HapA and HapC in GL (Figure 5d).

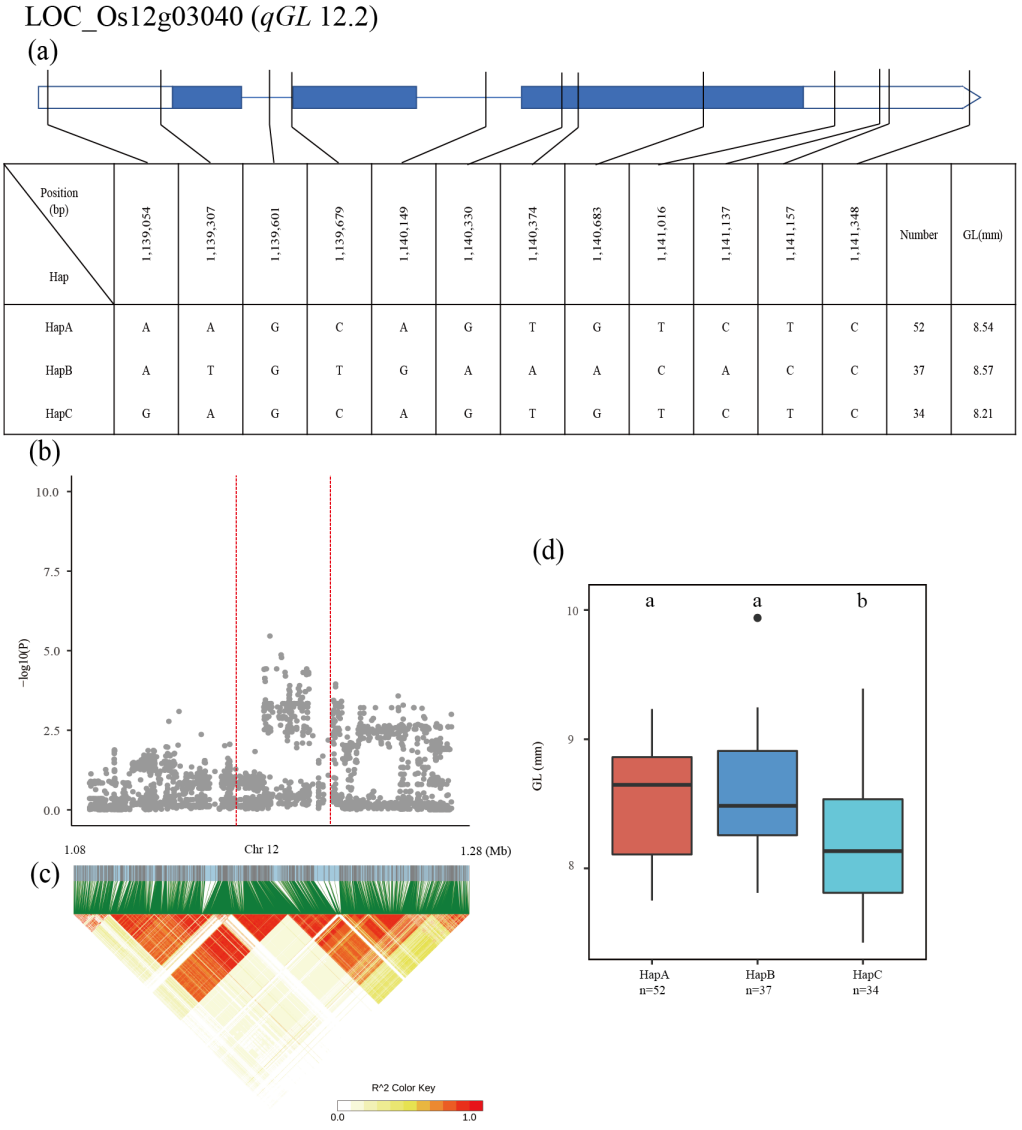


Figure 5. Identification of candidate genes for GL. (**a**) Based on 12 SNPs in all evaluated rice accessions, three haplotypes of LOC_Os12g03040 were identified. In the gene structure diagram of LOC_Os12g03040, the promoter is indicated by white frame; the exon is represented by blue frame; and the intron and intergenic region are marked by blue lines. A thin black line represents the genomic location of each SNP. Haplotypes with fewer than 10 accessions are not shown. GL (**b**) based on single polymorphism and LD heat map of local Manhattan map (**c**), around the peak on chromosome 12. Red dotted lines represent candidate regions for associated SNPs. Based on GL (**d**) of LOC_Os12g03040 haplotype, differences between haplotypes were statistically analyzed using Tukey's test.

The QTL *qGW6.1* was detected in the region from 8.59 Mb to 8.79 Mb on chromosome 6, and the 2313 SNPs of ten genes were used for association analysis. LOC_Os06g15480

was subsequently screened as the candidate gene for *qGW6.1* (Figure 6b,c). Two major haplotypes were detected among the 137 accessions based on six SNPs in the LOC_Os06g15480 promoter, nine SNPs in the exon and six SNPs in the intron. The average GW was 2.64 mm for HapA and 2.59 mm for HapB, showing significant differences from each other (Figure 6a,d).

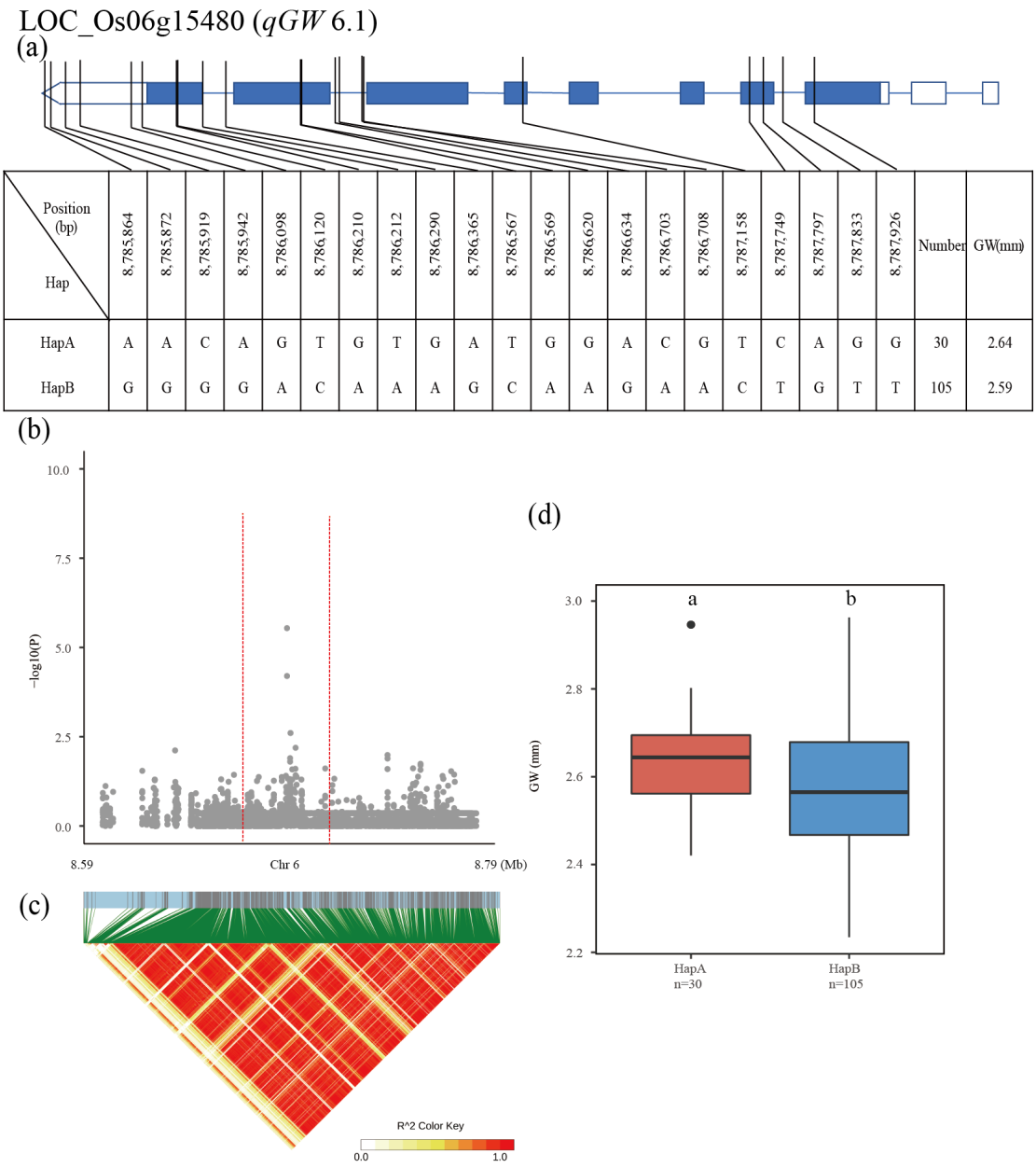


Figure 6. Identification of candidate genes for GW. (**a**) Based on 21 SNPs in all evaluated rice accessions, two haplotypes of LOC_Os06g15480 were identified. In the gene structure diagram of LOC_Os06g15480, the promoter is indicated by white frame; the exon is represented by blue frame; and the intron and intergenic region are marked by blue lines. A thin black line represents the genomic location of each SNP. Haplotypes with fewer than 10 accessions are not shown. GW (**b**) based on single polymorphism and LD heat map of local Manhattan map (**c**), around the peak on chromosome 6. Red dotted lines represent candidate regions for associated SNPs. Based on GW (**d**) of LOC_Os06g15480 haplotype, differences between haplotypes were statistically analyzed using Tukey's test.

The QTL *qGLWR4.2* was identified in a 29.0–29.2Mb region on chromosome 4, and the 345 SNPs of 15 genes were used for high-density association analysis. The most significant hit was located in LOC_Os04g49130 (Figure 7b,c). Two major haplotypes were detected among the 137 accessions based on one SNP in the LOC_Os04g49130 promoter. The average GLWR was 3.48 for HapA and 3.21 for HapB with significant differences from each other (Figure 7a,d).

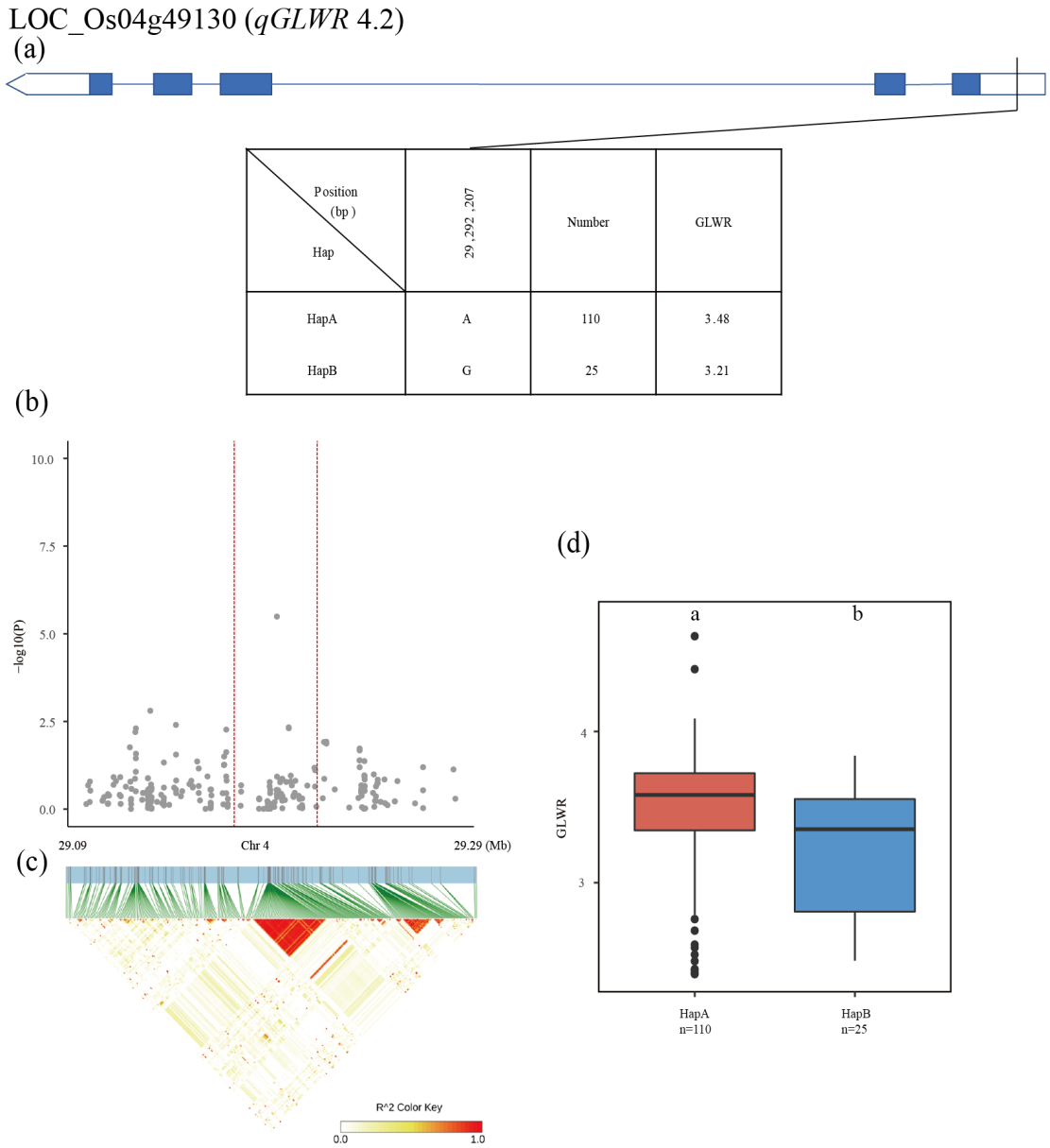


Figure 7. Identification of candidate genes for GLWR. (**a**) Based on one SNP in all evaluated rice accessions, two haplotypes of LOC_Os04g49130 were identified. In the gene structure diagram of LOC_Os04g49130, the promoter is indicated by white frame; the exon is represented by blue frame; and the intron and intergenic region are marked by blue lines. A thin black line represents the genomic location of each SNP. Haplotypes with fewer than 10 accessions are not shown. GLWR (**b**) based on single polymorphism and LD heat map of local Manhattan map (**c**), around the peak on chromosome 4. Red dotted lines represent candidate regions for associated SNPs. Based on GLWR (**d**) of LOC_Os04g49130 haplotype, differences between haplotypes were statistically analyzed using Tukey's test.

The QTL *qPGWC5.1* was detected in the region from 0.16 Mb to 0.36 Mb on chromosome 5, and the 952 SNPs of 28 genes were used for the association analysis. LOC_Os05g01430 was subsequently screened as the candidate gene for *qPGWC5.1* (Figure 8b,c). Three major haplotypes were detected among the 137 accessions based on five SNPs in the LOC_Os05g01430 exon. The mean PGWC was 0.26, 0.32 and 0.32 for HapA, HapB, and HapC, respectively (Figure 8a). The PGWC of HapA was significantly lower than that of other two haplotypes (Figure 8d).

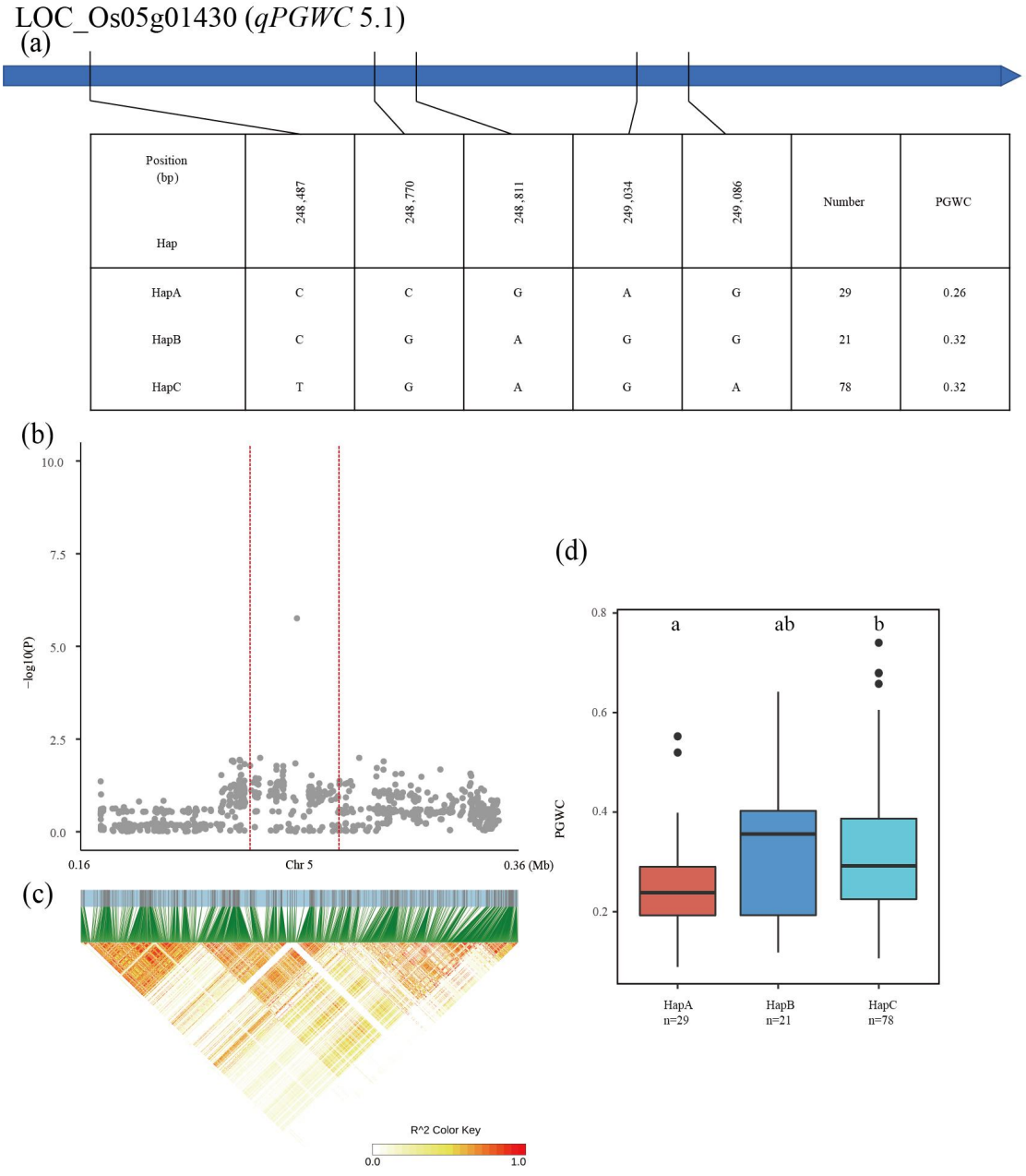


Figure 8. Identification of candidate genes for PGWC. (**a**) Based on five SNPs in all evaluated rice accessions, three haplotypes of LOC_Os05g01430 were identified. In the gene structure diagram of LOC_Os05g01430, the exon is represented by blue frame. A thin black line represents the genomic location of each SNP. Haplotypes with fewer than 10 accessions are not shown. PGWC (**b**) based on single polymorphism and LD heat map of local Manhattan map (**c**), around the peak on chromosome 5. Red dotted lines represent candidate regions for associated SNPs. Based on PGWC (**d**) of LOC_Os05g01430 haplotype, differences between haplotypes were statistically analyzed using Tukey's test.

The QTL *qDEC6.1* was identified in a 27.27–27.47 Mb region on chromosome 6, and the 1056 SNPs of 28 genes were used for high-density association analysis. The most significant hit was located in LOC_Os06g45300 (Figure 9b,c). Three major haplotypes were detected among the 137 accessions based on ten SNPs in the LOC_Os06g45300 promoter, three SNPs in the exon, and three SNPs in the intron. The mean DEC was 0.13, 0.09 and 0.09 for HapA, HapB and HapC, respectively (Figure 9a). The DEC of HapA was significantly higher than that of other two haplotypes (Figure 9d).

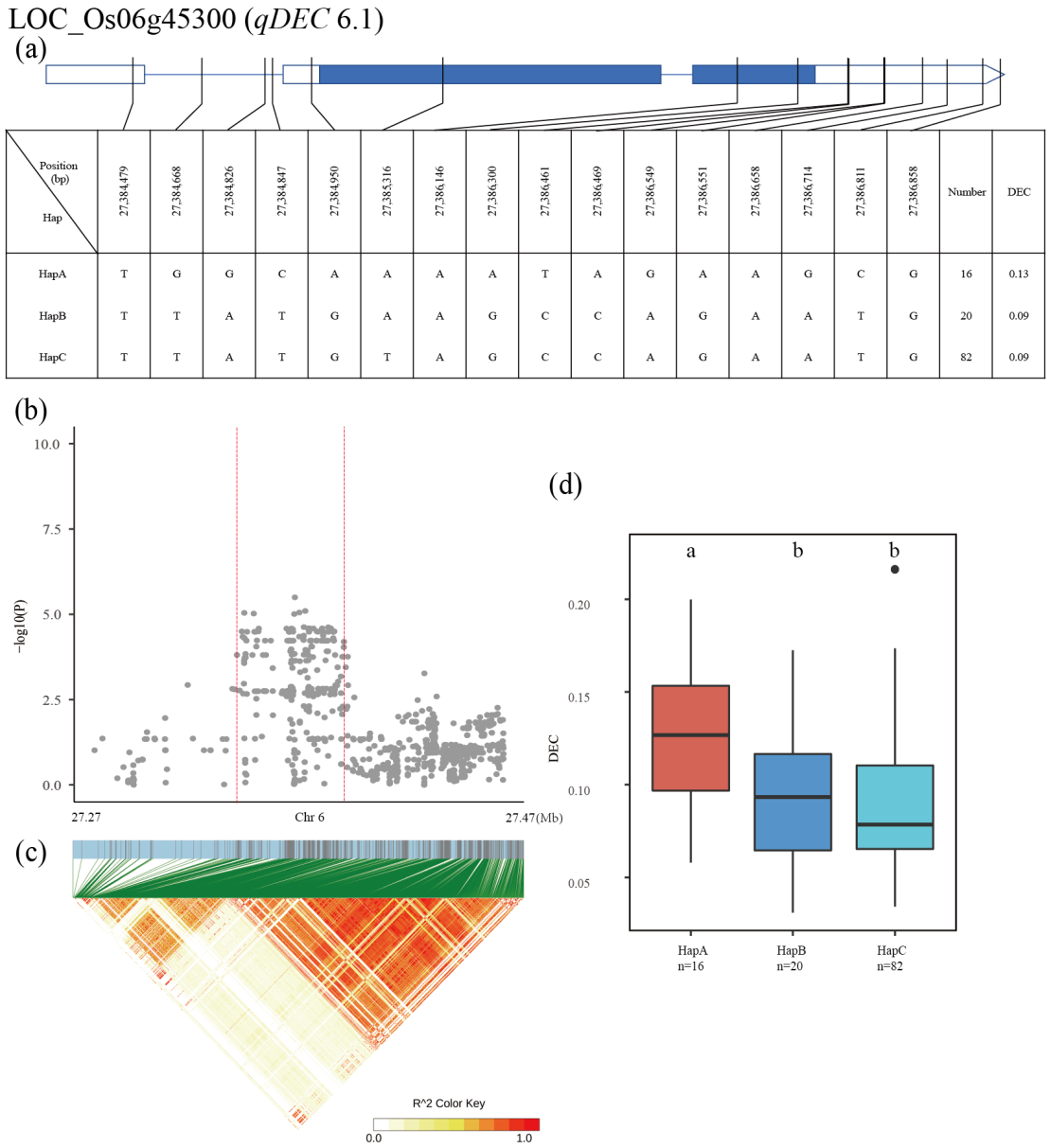


Figure 9. Identification of candidate genes for DEC. (**a**) Based on 16 SNPs in all evaluated rice accessions, three haplotypes of LOC_Os06g45300 were identified. In the gene structure diagram of LOC_Os06g45300, the promoter is indicated by white frame; the exon is represented by blue frame; and the intron and intergenic region are marked by blue lines. A thin black line represents the genomic location of each SNP. Haplotypes with fewer than 10 accessions are not shown. DEC (**b**) based on single polymorphism and LD heat map of local Manhattan map (**c**), around the peak on chromosome 6. Red dotted lines represent candidate regions for associated SNPs. Based on DEC (**d**) of LOC_Os06g45300 haplotype, differences between haplotypes were statistically analyzed using Tukey's test.

3. Discussion

The availability of high-density genotype data provided by next generation sequencing offers great opportunities to re-analyze previously collected panels of phenotypes. High-density SNP datasets can provide higher genomic coverage and resolution, and GWAS can identify segregating loci in populations [41], loci with reduced genetic background and QTL–environment interactions [42–44]. The use of larger structured populations may improve the mapping resolution for detecting global QTLs. Regression models are often constructed to test the correlations between markers and phenotypes. Population structure is usually represented by the proportion of subpopulations to which individuals belong, which is also known as the Q-matrix. Since the subsets in the Q matrix have the fitting of fixed effects, the general linear model (GLM) can be used to test for genetic markers [38,45]. This model can be conceptually expressed as y = Q + S + e, where y and e are the phenotype and residue, respectively.

In this study, we genotyped 137 indica rice accessions using 2,998,034 SNPs (Figure S3). We found sufficient diversity to map QTLs associated with rice grain weight, grain size and chalkiness, and identified haplotypes with significant differences in grain weight, grain size and chalkiness. Although the natural population consisting of 137 accessions was not large enough, there were significant phenotypic variations in grain weight, grain shape and chalkiness. In this study, the variation coefficients of TGW, GL, GW, GLWR, PGWC and DEC were 13.62–14.69%, 8.77–9.10%, 9.85–11.37%, 12.65–13.65%, 43.07–59.11% and 44.21–62.36% in the whole population, respectively. Particularly, HN had the highest TGW (36.08 g), GL (10.65 mm), GW (3.51 mm) and PGWC (0.81). These significant phenotypic variations may be associated with high genetic diversity.

By using GWAS and gene-based association analysis, combined with haplotype analysis of candidate genes, we screened six candidate genes for six important QTLs controlling the measured traits. Of the three major components (panicle number per plant, number of grains per panicle and grain weight) of rice yield, grain weight measured as TGW is the most stable and heritable trait. In this study, one promising gene located in the QTL for grain weight was also identified. For *qTGW4.2*, LOC_Os04g02900 (*FLO19*) [39] encodes a pyruvate dehydrogenase complex E1-alpha subunit involved in the biosynthesis of galactolipids, which is required for the development of amyloplasts. The mutation of this gene significantly decreased the pyruvate dehydrogenase complex enzyme activity, accompanied by a significant decrease in total galactolipid content, which led to the abnormal development of amyloplasts, impaired starch synthesis and ultimately seriously affected rice yield and quality.

Rice grains grow inside the spikelet hull with a limited caryopsis space. Therefore, rice grain shape and size are strictly determined by the maternal genotype that controls the cell number and size of glumes. The genes cloned so far provide insights into the regulatory pathway of grain shape and size, including the ubiquitin-proteasome, G-protein signaling, and mitogen-activated protein kinase (MAPK) signaling pathways, as well as plant hormones and transcriptional regulators [46,47]. In the present study, three promising genes (one known gene and two novel genes) located in the three new QTLs for grain weight were identified using a large natural population with 137 accessions. The first one was *qGL12.2*, in which LOC_Os12g03040 is annotated as a NAC (NAM/ATAF/CUC) transcription factor playing important roles in regulating plant growth and development as well as biotic and abiotic stress responses. The second one was *qGW6.1*, in which LOC_Os06g15480 is annotated as a basic region-leucine zipper (bZIP) transcription factor. This gene has been identified in a previous study and designated as *OsbZIP47* [40] with transcriptional activation activity, but wg1 can directly interact with *OsbZIP47* and recruit the transcriptional co-repressor aberrant spikelet and panicle1 (*ASP1*) to repress its transcriptional activity. *OsbZIP47* overexpression lines showed narrower seeds, which is similar to the *wg1* mutant. The *wg1-1 osbzip47-c2* double mutant had wider and heavier seeds than the *wg1-1* single mutant, and *OsbZIP47* was found to negatively regulate grain width by limiting cell proliferation. The third one was *qGLWR4.2*, in which LOC_Os04g49130 is annotated as a

small ubiquitin-like modifier (SUMO)-conjugating enzyme E2. SUMOylation modification is an important eukaryotic post-translational modification regulating many cellular processes in plants, from seed development to stress response. The transcript levels of SUMOylation target genes, including Abscisic Acid (ABA) and gibberellins (GA) associated with cystatin-related epididymal spermatogenic (CREs), are responsive to treatment with these hormones.

Chalkiness is caused by the deposition of starch and storage proteins in the endosperm, which is tightly related to grain filling. Grain filling is a dynamic process related to the source–sink balance and controlled by a complex genetic mechanism and sensitive to environmental conditions [48–50]. In this study, two promising genes located in the two new QTLs for grain weight were identified using a large natural population with 137 accessions. The first one was *qPGWC5.1*, from which LOC_Os05g01430 was screened as a candidate gene and annotated as polygalacturonase inhibiting protein 3. LOC_Os05g01430 is a plant cell wall glycoprotein inhibiting fungal endopolygalacturonases and modulating their activity to promote the accumulation of elicitor-active oligogalacturonides. The second one was *qDEC6.1*, in which LOC_Os06g45300 was identified as the candidate gene and annotated as a rice dual-specificity protein kinase. Dual-specificity protein kinases are a group of protein kinases able to phosphorylate both tyrosine and serine/threonine residues. These results improve our understanding of the genetic basis for grain weight, grain shape and chalkiness in rice and provide valuable information for elucidating the molecular mechanisms underlying these traits.

Overall, we obtained a set of QTLs significantly associated with grain weight, grain shape and chalkiness in rice through the GWAS analysis of 137 rice accessions. Candidate genes significantly associated with agronomic traits and functional annotation of each gene were further screened by haplotype block structure analysis. As expected, although a large LD contains many SNPs in a candidate region, our results suggest that the number of candidate genes can be significantly reduced by combining haplotype block structure analysis and gene function annotation. In conclusion, our findings will contribute to future gene functional analysis and provide valuable information for rice gene cloning.

4. Materials and Methods

4.1. Plant Materials, Field Trials and Trait Measurements

A total of 137 accessions (Table S1) from the world were used to test the association between the SNP genotype and the phenotype of grain weight, grain shape and chalkiness. All of these accessions were grown in three environments, including Hainan (18.3 N, 109.3 E; HN), Guangxi (20.54 N, 104.29 E; GX) and Jiangxi (24.29 N, 118.28 E; JX) in 2018. In all the three environments, each accession was planted in a two-row plot with 10 individuals in each row at a spacing of 20 cm × 25 cm with two replicates for each accession. Field management included irrigation, fertilizer application and pest control, following normal agricultural practices. At maturity (about 40 days after flowering), seeds of ten plants in the middle of each plot were harvested, air-dried and stored at room temperature for at least three months before testing [51]. Then, all full head milled rice kernels of each accession were used to measure the grain length (GL, mm), grain width (GW, mm), grain length-width ratio (GLWR), degree of endosperm chalkiness (DEC) and percentage of grains with chalkiness (PGWC) using a rice grain appearance quality scanning machine ((Model SC-G, Hangzhou, China, http://www.wseen.com/, accessed on 20 April 2022). Subsequently, the weight of these seeds was measured using a high precision electronic balance (1/1000, APTP456 series) and the TGW in grams was subsequently calculated. The scanner was calibrated with a calibration target before each measurement.

4.2. Statistical Analysis

Excel 2018 was employed for data compilation, and the mean, standard deviation and coefficient of variation of each trait were calculated. The correlation analysis and frequency analysis of six grain quality-related traits were carried out by R (TGW, GL, GW, GLWR,

PGWC and DEC). The R package 'lme4' [52] was used to obtain the best linear unbiased estimate (BLUP) for each genotype–environment combination and variance components using generalized linear models.

4.3. DNA Extraction and SNP Genotyping

For each of the 137 accessions to be sequenced, two leaves were collected from a single plant at the tillering stage (one month after seedling transplantation), and genomic DNA was extracted using a standard cetyltrimethylammonium bromide protocol [53]. According to the manufacturer's instructions (Illumina, https://www.illlighta.com/, accessed on 26 April 2019), paired-end sequencing libraries were constructed using 5 μg of genomic DNA, with inserted fragments of approximately 350 bp. The Illumina HiSeq X10 platform was used to obtain the pair-ends of 150 bp reads, and the original sequence was further processed to remove adaptor-containing and low-quality reads. Library construction, sequencing and sequence cleaning were all performed by BGI Shenzhen Company. The reference genome was Shuhui 498 (R498). GATK was used to call SNPs [54]. The mapping results were converted to VCF format using SAMtools (version 0.1.18) [55]. SNPs with MAF $\geq$ 5% and missing rate $\leq$ 20% were retained. IMPUTE2 [56] was used to impute missing genotypes, and 2,898,034 high-quality SNPs were finally obtained.

4.4. Population Structure and Kinship Analysis

The TASSEL [57] was used to calculate the population structure (Q) and kinship (K). All SNPs were used in the calculation. Principal component analysis (PCA) was used to evaluate the population structure. The PCA score and relationship matrix would be used in the generalized linear model (GLM) [38] below.

SNP data of candidate genes were extracted based on genotype of SNP with MAF > 0.05. This dataset only contained double-allele SNPs. Further haplotype analysis excluded heterozygotes and missing alleles. Haplogroups consisting of fewer than 10 accessions were deleted. For the genes found in QTLs, haplotype analysis only used the non-synonymous SNPs in the coding region of these genes for haplotype analysis of R.

4.5. Linkage Disequilibrium Analysis

The software "PopLDdecay" was used to calculate the linkage disequilibrium (LD) between pairs of markers [58]. "R2" was the square of the Pearson's correlation coefficient. When the correlation coefficient (R2) dropped to half of its maximum value, the distance across the chromosome is called the LD decay distance [59].

4.6. Genome-Wide Association Mapping

In this study, we obtained 2,898,034 SNPs (MAF > 0.05) and six sets of phenotypic data. The SNPs and phenotypic data were used to conduct GWAS in the TASSEL (version 5.2.40) [57] software using a GLM. For the GLM, the suggested p value of dominance was 3.9×10^{-6} to control the genetic false positive error rate of the population. The SNPs in the same LD region were regarded as one QTL. Here, by referring to the previous report, we used the LD decay distance of 100 kb [60], and the SNP with the lowest p-value was regarded as the lead SNP. The Manhattan plot was drawn using the R package "CMplot".

4.7. Identification of Candidate Genes and Haplotype Analysis

In order to identify candidate genes related to TGW, GL, GW, GLWR, PGWC and DEC, the rice genome annotation project (http://rice.plantbiology.msu.edu, accessed on 26 April 2022) was used to search for candidate genes in the 200 kb genomic region of the selected SNPs. Among all candidate genes, four types of genes, including expression proteins, hypothetical proteins, retrotransposons and transposons, were excluded. The SNP data of candidate genes were extracted based on genotype of SNP with MAF > 0.05. This dataset only contained double-allele SNPs. Further haplotype analysis excluded heterozygotes and missing alleles. Haplogroups consisting of fewer than 10 accessions were deleted. For

the genes found in QTLs, only non-synonymous SNPs in the coding region of these genes were used for the haplotype analysis of R, and Student's t-test was performed to determine whether this locus could cause changes in rice rain weight, grain shape and chalkiness. R was used to visualize the results.

5. Conclusions

There are considerable genetic variations for six grain quality traits in the panel consisting of 137 rice accessions. Twenty-seven QTLs were identified using GWAS. Six candidate genes were also screened by high-density association and gene-based haplotype analysis. The findings improve our understanding of the genetic basis for grain weight, grain shape and chalkiness in rice and provide valuable information for elucidating the molecular mechanisms underlying these traits. This study also provides a reference for future marker-assisted breeding with QTL or gene pyramiding to stabilize and improve rice quality.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/plants12030419/s1, Figure S1: Histogram of the phenotypic frequency distribution of rice grain weight, grain shape and grain chalkiness in 137 rice accessions. Table S1: Information of 137 rice accessions. Figure S2: LD decay distance estimated for 220 rice accessions. Table S2: Analysis of variance associated with grain weight, grain shape and chalkiness traits. Figure S3: Distribution of single nucleotide polymorphisms (SNPs) and nucleotide diversity across the rice Nipponbare genome in the rice association panel. Table S3: BLUP values consistently determined QTN of grain weight, grain shape and chalkiness traits by TASSEL in the three environments. Table S4: List of genes associated with grain weight, grain shape and chalkiness traits.

Author Contributions: H.C. and N.W. conducted experiments and collected data; Y.Q., G.Z., W.Z., Y.B., T.F., M.L., Z.L., E.L., C.Z., J.X. (Jianlong Xu) and N.W. data collation and statistical analysis; N.W., Y.Q., J.X. (Jun Xiang) and Z.L. graphic finishing; N.W. wrote the paper; Y.S. designed the experiment, provided intellectual guidance and wrote and reviewed the paper. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Scientific Research Plan Major Project of Anhui Province (grant number 2022AH040126), the Science and Technology Major Project of Anhui Province (grant number 2021d06050002), the Improved Varieties Joint Research (Rice) Project of Anhui Province (the 14th five-year plan) and the National Natural Science Foundation of China [grant number U21A20214].

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Chen, W.; Gao, Y.; Xie, W.; Gong, L.; Lu, K.; Wang, W.; Li, Y.; Liu, X.; Zhang, H.; Dong, H.; et al. Genome-wide association analyses provide genetic and biochemical insights into natural variation in rice metabolism. *Sci. Found. China* **2014**, *46*, 714–721. [CrossRef] [PubMed]
2. Gong, L.; Chen, W.; Gao, Y.; Liu, X.; Zhang, H.; Xu, C.; Yu, S.; Zhang, Q.; Luo, J. Genetic analysis of the metabolome exemplified using a rice population. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 20320–20325. [CrossRef] [PubMed]
3. Fao, R.I.E. *Increasing Crop Production Sustainably: The Perspective of Biological of Processes*; Food and Agriculture Organization: Rome, Italy, 2009.
4. Godfray, H.C.J. The challenge of feeding 9-10 billion people equitably and sustainably. *J. Agric. Sci.* **2014**, *152*, S2–S8. [CrossRef]
5. Seck, P.A.; Diagne, A.; Mohanty, S.; Wopereis, M. Crops that feed the world 7: Rice. *Food Secur.* **2012**, *4*, 7–24. [CrossRef]
6. Heong, K.L.; Toriyama, K.; Hardy, B. Rice is life: Scientific perspectives for the 21st century. In Proceedings of the World Rice Reseach Conference, Tsukuba, Japan, 5–7 November 2004.
7. Khush, G.S. What it will take to Feed 5.0 Billion Rice consumers in 2030. *Plant Mol. Biol.* **2005**, *59*, 1–6. [CrossRef]
8. Zheng, T.Q.; Xu, J.L.; Li, Z.K.; Zhai, H.Q.; Wan, J.M. Genomic regions associated with milling quality and grain shape identified in a set of random introgression lines of rice (*Oryza sativa* L.). *Plant Breed.* **2007**, *126*, 158–163. [CrossRef]
9. Fan, C.; Xing, Y.; Mao, H.; Lu, T.; Han, B.; Xu, C.; Li, X.; Zhang, Q. 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.* **2006**, *112*, 1164–1171. [CrossRef]

10. Qiu, X.; Pang, Y.; Yuan, Z.; Xing, D.; Xu, J.; Dingkuhn, M.; Li, Z.; Ye, G. Genome-Wide Association Study of Grain Appearance and Milling Quality in a Worldwide Collection of Indica Rice Germplasm. *PLoS ONE* **2015**, *10*, e0145577. [CrossRef]
11. Liu, X.; Hua, Z.T.; Wang, Y. Quantitative trait locus (QTL) analysis of percentage grains chalkiness using AFLP in rice (*Oryza sativa* L.). *Afr. J. Biotechnol.* **2011**, *10*, 2399–2405.
12. Tan, Y.F.; Xing, Y.Z.; Li, J.X.; Yu, S.B.; Xu, C.G.; Zhang, Q.F. Genetic bases of appearance quality of rice grains in Shanyou 63, an elite rice hybrid. *Theor. Appl. Genet.* **2000**, *101*, 823–829. [CrossRef]
13. Zhao, X.Q.; Zhou, L.J.; Ponce, K.; Ye, G.Y. The Usefulness of Known Genes/Qtls for Grain Quality Traits in an Indica Population of Diverse Breeding Lines Tested using Association Analysis. *Rice* **2015**, *8*, 13. [CrossRef] [PubMed]
14. Li, Z.; Wan, J.; Xia, J.; Yano, M. Mapping of Quantitative Trait Loci Controlling Physico-chemical Properties of Rice Grains (*Oryza sativa* L.). *Breed. Sci.* **2003**, *53*, 209–215. [CrossRef]
15. Jiang, G.H.; Hong, X.Y.; Cai-Guo, X.U.; Xiang-Hua, L.I.; Yu-Qing, H.E. Identification of Quantitative Trait Loci for Grain Appearance and Milling Quality Using a Doubled-Haploid Rice Population. *J. Integr. Plant Biol.* **2005**, *47*, 1391–1403. [CrossRef]
16. Xie, X.; Song, M.H.; Jin, F.; Ahn, S.N.; Suh, J.P.; Hwang, H.G.; Mccouch, S.R. Fine mapping of a grain weight quantitative trait locus on rice chromosome 8 using near-isogenic lines derived from a cross between *Oryza sativa* and *Oryza rufipogon*. *Theor. Appl. Genet.* **2006**, *113*, 885–894. [CrossRef] [PubMed]
17. Dai, L.; Lan, W.; Leng, Y.; Yang, Y.; Zeng, D. Quantitative Trait Loci Mapping for Appearance Quality in Short-Grain Rice. *Crop Sci.* **2016**, *56*, 1484–1492. [CrossRef]
18. Mei, D.Y.; Zhu, Y.J.; Yong-Hong, Y.U.; Fan, Y.Y.; Huang, D.R.; Zhuang, J.Y. Quantitative Trait Loci for Grain Chalkiness and Endosperm Transparency Detected in Three Recombinant Inbred Line Populations of Indica Rice. *J. Integr. Agric.* **2013**, *12*, 1–13. [CrossRef]
19. Bian, J.M.; He, H.H.; Li, C.J.; Shi, H.; Yan, S. Identification and analysis of QTLs for grain quality traits in rice using an introgression lines population. *Euphytica Int. J. Plant Breed.* **2014**, *195*, 83–93. [CrossRef]
20. Ayaad, M.; Han, Z.; Zheng, K.; Hu, G.; Xing, Y. Bin-based genome-wide association studies reveal superior alleles for improvement of appearance quality using a 4-way MAGIC population in rice. *J. Adv. Res.* **2021**, *28*, 183–194. [CrossRef]
21. Zhong, H.; Liu, S.; Sun, T.; Kong, W.; Deng, X.; Peng, Z.; Li, Y. Multi-locus genome-wide association studies for five yield-related traits in rice. *BMC Plant Biol.* **2021**, *21*, 364. [CrossRef]
22. Xiao, J.; Li, J.; Grandillo, S.; Ahn, S.N.; Yuan, L.; Tanksley, S.D.; Mccouch, S.R. Identification of trait-improving quantitative trait loci alleles from a wild rice relative, *Oryza rufipogon*. *Genetics* **1998**, *150*, 899–909. [CrossRef]
23. Song, X.J.; Huang, W.; Shi, M.; Zhu, M.Z.; Lin, H.X. A QTL for rice grain width and weight encodes a previously unknown RING-type E3 ubiquitin ligase. *Nat. Genet.* **2007**, *39*, 623–630. [CrossRef] [PubMed]
24. Mao, H.L.; Sun, S.Y.; Yao, J.L.; Wang, C.R.; Yu, S.B.; Xu, C.G.; Li, X.H.; Zhang, Q.F. Linking differential domain functions of the GS3 protein to natural variation of grain size in rice. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 19579–19584. [CrossRef]
25. Li, Y.B.; Fan, C.C.; Xing, Y.Z.; Jiang, Y.H.; Luo, L.J.; Sun, L.; Shao, D.; Xu, C.J.; Li, X.H.; Xiao, J.H.; et al. Natural variation in GS5 plays an important role in regulating grain size and yield in rice. *Nat. Genet.* **2011**, *43*, 1266–1269. [CrossRef] [PubMed]
26. Zhang, X.J.; Wang, J.F.; Huang, J.; Lan, H.X.; Wang, C.L.; Yin, C.F.; Wu, Y.Y.; Tang, H.J.; Qian, Q.; Li, J.Y.; et al. Rare allele of OsPPKL1 associated with grain length causes extra-large grain and a significant yield increase in rice. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 21534–21539. [CrossRef] [PubMed]
27. Wang, S.K.; Wu, K.; Yuan, Q.B.; Liu, X.Y.; Liu, Z.B.; Lin, X.Y.; Zeng, R.Z.; Zhu, H.T.; Dong, G.J.; Qian, Q.; et al. Control of grain size, shape and quality by OsSPL16 in rice. *Nat. Genet.* **2012**, *44*, 950–954. [CrossRef]
28. Wang, S.K.; Li, S.; Liu, Q.; Wu, K.; Zhang, J.Q.; Wang, S.S.; Wang, Y.; Chen, X.B.; Zhang, Y.; Gao, C.X.; et al. The OsSPL16-GW7 regulatory module determines grain shape and simultaneously improves rice yield and grain quality. *Nat. Genet.* **2015**, *47*, 949–954. [CrossRef]
29. Wang, Y.X.; Xiong, G.S.; Hu, J.; Jiang, L.; Yu, H.; Xu, J.; Fang, Y.X.; Zeng, L.J.; Xu, E.B.; Xu, J.; et al. Copy number variation at the GL7 locus contributes to grain size diversity in rice. *Nat. Genet.* **2015**, *47*, 944–948. [CrossRef]
30. Liu, S.Y.; Hua, L.; Dong, S.J.; Chen, H.Q.; Zhu, X.D.; Jiang, J.E.; Zhang, F.; Li, Y.H.; Fang, X.H.; Chen, F. OsMAPK6, a mitogen-activated protein kinase, influences rice grain size and biomass production. *Plant J.* **2015**, *84*, 672–681. [CrossRef]
31. Li, Y.; Fan, C.; Xing, Y.; Yun, P.; Luo, L.; Yan, B.; Peng, B.; Xie, W.; Wang, G.; Li, X. Chalk5 encodes a vacuolar H^+-translocating pyrophosphatase influencing grain chalkiness in rice. *Nat. Genet.* **2014**, *46*, 398–404. [CrossRef]
32. Ishimaru, K.; Hirotsu, N.; Madoka, Y.; Murakami, N.; Hara, N.; Onodera, H.; Kashiwagi, T.; Ujiie, K.; Shimizu, B.; Onishi, A.; et al. Loss of function of the IAA-glucose hydrolase gene TGW6 enhances rice grain weight and increases yield. *Nat. Genet.* **2013**, *45*, 707. [CrossRef]
33. Weng, J.F.; Gu, S.H.; Wan, X.Y.; Gao, H.; Guo, T.; Su, N.; Lei, C.L.; Zhang, X.; Cheng, Z.J.; Guo, X.P.; et al. Isolation and initial characterization of GW5, a major QTL associated with rice grain width and weight. *Cell Res.* **2008**, *18*, 1199–1209. [CrossRef] [PubMed]
34. Shomura, A.; Izawa, T.; Ebana, K.; Ebitani, T.; Kanegae, H.; Konishi, S.; Yano, M. Deletion in a gene associated with grain size increased yields during rice domestication. *Nat. Genet.* **2008**, *40*, 1023–1028. [CrossRef] [PubMed]
35. McCouch, S.R.; Wright, M.H.; Tung, C.W.; Maron, L.G.; McNally, K.L.; Fitzgerald, M.; Singh, N.; DeClerck, G.; Perez, F.A.; Korniliev, P.; et al. Open access resources for genome-wide association mapping in rice. *Nat. Commun.* **2016**, *7*, 10532. [CrossRef] [PubMed]

36. Zhu, C.S.; Gore, M.; Buckler, E.S.; Yu, J.M. Status and Prospects of Association Mapping in Plants. *Plant Genome* **2008**, *1*, 5–20. [CrossRef]
37. Alexander, D.H.; Novembre, J.; Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **2009**, *19*, 1655–1664. [CrossRef]
38. Price, A.L.; Patterson, N.J.; Plenge, R.M.; Weinblatt, M.E.; Shadick, N.A.; Reich, D. Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **2006**, *38*, 904–909. [CrossRef]
39. Lei, J.; Teng, X.; Wang, Y.; Jiang, X.; Zhao, H.; Zheng, X.; Ren, Y.; Dong, H.; Wang, Y.; Duan, E. Plastidic pyruvate dehydrogenase complex E1 component subunit Alpha1 is involved in galactolipid biosynthesis required for amyloplast development in rice. *Plant Biotechnol. J.* **2022**, *20*, 437–453. [CrossRef]
40. Hao, J.; Wang, D.; Wu, Y.; Huang, K.; Duan, P.; Li, N.; Xu, R.; Zeng, D.; Dong, G.; Zhang, B. The GW2-WG1-OsbZIP47 pathway controls grain size and weight in rice. *Mol. Plant* **2021**, *14*, 1266–1280. [CrossRef]
41. Langridge, P.; Lagudah, E.; Holton, T.; Appels, R.; Sharp, P.; Chalmers, K. Trends in genetic and genome analyses in wheat: A review. *Aust. J. Agric. Res.* **2001**, *52*, 1043–1077. [CrossRef]
42. Malosetti, M.; Voltas, J.; Romagosa, I.; Ullrich, S.; Van Eeuwijk, F. Mixed models including environmental covariables for studying QTL by environment interaction. *Euphytica* **2004**, *137*, 139–145. [CrossRef]
43. Shorter, R.; Van Eeuwijk, F. Multi-environment QTL mixed models for drought stress adaptation in wheat. *Theor. Appl. Genet.* **2008**, *117*, 1077–1091.
44. Gutiérrez, L.; Germán, S.; Pereyra, S.; Hayes, P.M.; Pérez, C.A.; Capettini, F.; Locatelli, A.; Berberian, N.M.; Falconi, E.E.; Estrada, R. Multi-environment multi-QTL association mapping identifies disease resistance QTL in barley germplasm from Latin America. *Theor. Appl. Genet.* **2015**, *128*, 501–516. [CrossRef] [PubMed]
45. Liu, X.; Huang, M.; Fan, B.; Buckler, E.S.; Zhang, Z. Iterative usage of fixed and random effect models for powerful and efficient genome-wide association studies. *PLoS Genet.* **2016**, *12*, e1005767. [CrossRef] [PubMed]
46. Li, N.; Xu, R.; Li, Y. Molecular networks of seed size control in plants. *Annu. Rev. Plant Biol.* **2019**, *70*, 435–463. [CrossRef]
47. Li, X.; Tao, Q.; Miao, J.; Yang, Z.; Gu, M.; Liang, G.; Zhou, Y. Evaluation of differential qPE9-1/DEP1 protein domains in rice grain length and weight variation. *Rice* **2019**, *12*, 1–10. [CrossRef]
48. Liu, J.; Wu, M.-W.; Liu, C.-M. Cereal Endosperms: Development and Storage Product Accumulation. *Annu. Rev. Plant Biol.* **2022**, *73*, 255–291. [CrossRef]
49. Ishimaru, T.; Miyazaki, M.; Shigemitsu, T.; Nakata, M.; Kuroda, M.; Kondo, M.; Masumura, T. Effect of high temperature stress during ripening on the accumulation of key storage compounds among Japanese highly palatable rice cultivars. *J. Cereal Sci.* **2020**, *95*, 103018. [CrossRef]
50. Yamakawa, H.; Ebitani, T.; Terao, T. Comparison between locations of QTLs for grain chalkiness and genes responsive to high temperature during grain filling on the rice chromosome map. *Breed. Sci.* **2008**, *58*, 337–343. [CrossRef]
51. Wang, X.; Pang, Y.; Wang, C.; Chen, K.; Zhu, Y.; Shen, C.; Ali, J.; Xu, J.; Li, Z. New Candidate Genes Affecting Rice Grain Appearance and Milling Quality Detected by Genome-Wide and Gene-Based Association Analyses. *Front. Plant Sci.* **2017**, *7*, 1998. [CrossRef]
52. Gong, J.Y.; Miao, J.S.; Zhao, Y.; Zhao, Q.; Feng, Q.; Zhan, Q.L.; Cheng, B.Y.; Xia, J.H.; Huang, X.H.; Yang, S.H.; et al. Dissecting the Genetic Basis of Grain Shape and Chalkiness Traits in Hybrid Rice Using Multiple Collaborative Populations. *Mol. Plant* **2017**, *10*, 1353–1356. [CrossRef]
53. Murray, M.; Thompson, W. Rapid isolation of high molecular weight plant DNA. *Nucleic Acids Res.* **1980**, *8*, 4321–4326. [CrossRef] [PubMed]
54. 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. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **2011**, *43*, 491–498. [CrossRef] [PubMed]
55. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. The sequence alignment/map format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [CrossRef] [PubMed]
56. Howie, B.N.; Donnelly, P.; Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. *PLoS Genet.* **2009**, *5*, e1000529. [CrossRef]
57. Shakiba, E.; Eizenga, G.; Maccouch, S. Using GWAS to identify SNPs associated with rice seedling cold tolerance. In Proceedings of the Rice Technical Workshop Group, Hangzhou, China, 3–6 November 2014.
58. Zhang, C.; Dong, S.-S.; Xu, J.-Y.; He, W.-M.; Yang, T.-L. PopLDdecay: A fast and effective tool for linkage disequilibrium decay analysis based on variant call format files. *Bioinformatics* **2019**, *35*, 1786–1788. [CrossRef]
59. Huang, X.; Sang, T.; Zhao, Q.; Feng, Q.; Zhao, Y.; Li, C.; Zhu, C.; Lu, T.; Zhang, Z.; Li, M. Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat. Genet.* **2010**, *42*, 961–967. [CrossRef] [PubMed]
60. Yang, L.; Yueying, W.; Jahan, N.; Haitao, H.; Ping, C.; Lianguang, S.; Haiyan, L.; Guojun, D.; Jiang, H.; Zhenyu, G. Genome-wide association analysis and allelic mining of grain shape-related traits in rice. *Rice Sci.* **2019**, *26*, 384–392. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.