Identification and Allele Mining of New Candidate Genes Underlying Rice Grain Weight and Grain Shape by Genome-Wide Association Study

Yanan Niu

Chinese Academy of Agricultural Sciences Institute of Crop Sciences

Tianxiao Chen

Chinese Academy of Agricultural Sciences Institute of Crop Sciences

Kai Chen

Chinese Academy of Agricultural Sciences Agricultural Genomes Institute at Shenzhen

Congcong Shen

Chinese Academy of Agricultural Sciences Agricultural Genomes Institute at Shenzhen

Huizhen Chen

Pngxiang Institute of Agricultural Sciences

Zhichao Wu

Chinese Academy of Agricultural Sciences Institute of Crop Sciences

Chunchao Wang

Chinese Academy of Agricultural Sciences Institute of Crop Sciences

Tianqing Zheng

Chinese Academy of Agricultural Sciences Institute of Crop Sciences

Jianlong Xu (✉ xujlcaas@126.com)

Chinese Academy of Agricultural Sciences Institute of Crop Sciences <https://orcid.org/0000-0001-6623-1256>

Original article

Keywords: Grain shape, genome-wide association study (GWAS), quantitative trait locus/loci (QTL), candidate gene analysis, favorable haplotype

Posted Date: May 13th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-28490/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Version of Record: A version of this preprint was published at BMC Genomics on August 6th, 2021. See the published version at <https://doi.org/10.1186/s12864-021-07901-x>.

Abstract

Background: Grain weight and grain shape are crucial agronomic traits that affect grain yield potential and grain quality of rice. Both Grain weight and grain shape are controlled by multiple genes known as quantitative trait locus/loci (QTL). The 3,000 Rice Genomes Project enables the discovery of valuable genetic variants and resources for grain weight and grain shape.

Results: 2445 rice accessions were classified into 12 sub-populations which showed diverse variations in grain weight and grain shape. Genome-wide association study (GWAS) revealed three novel QTL (*qTGW3.1*, *qTGW9* and *qTGW11*) for grain weight and four novel QTL (*qGL4/qRLW4*, *qGL10*, *qGL11* and *qRLW1*) for grain shape. We finally predicted eight candidate genes (*Os03g0186600*, *Os09g0544400*, *Os11g0163600*, *Os04g0580700*, *Os10g0399700*, *Os10g0400100*, *Os11g0143700* and *Os01g0171000*) for the seven novel QTL by gene-based association, haplotype and bioinformatics analyses. The accessions with excellent haplotypes at the eight candidate genes and five previously cloned genes were also identified.

Conclusion: The identified candidate genes and accessions provided valuable sources for future functional characterization and genetic improvement of rice grain yield and grain quality.

Background

Rice (*Oryza sativa* L.) is one of the most widely grown crops in the world and is a staple diet for more than 3.5 billion people around the world, particularly in Asia. It was predicted that rice production must increase by ~42% to keep pace with the increasing global food demand by 2050 (Ray et al. 2013). As the main component of grain yield, grain weight or thousand-grain weight (TGW) has always been a dominating trait for breeders. Meanwhile, grain shape is an important factor affecting physical or appearance quality, where slender grains with grain length-to-width ratio > 3 are preferred by most rice consumers (Gong et al. 2017). Grain shape is specified by grain length (GL), grain width (GW), the ratio of length-to-width (RLW) and grain thickness, and grain weight shows positive correlations with these four parameters (Tan et al. 2000).

To date, more than 60 genes associated with grain weight and grain shape have been identified by linkage mapping and association studies with some genes being cloned by map-based cloning. Among them, *Dwarf1* (*D1*), the first cloned gene regulates grain shape and causes smaller grains, which is indirectly caused by the inability to form the heterotrimeric G protein alpha subunit (Ashikari et al. 1999). Predominant genes like *GW2* (Song et al. 2007), *GS3* (Fan et al. 2006) and *GW5* (Liu et al. 2017) are negative regulators of grain shape, while *GS5* (Li et al. 2011), *GL7/GW7* (Wang et al. 2015a; Wang et al. 2015b), *GW8* (Wang et al. 2012), *GLW7* (Si et al. 2016), and *OsLG3* (Yu et al. 2017) are positive regulators. More QTL have been revealed recently. These include *qGRL7.1* (Singh et al. 2012), *qGL4b* (Kato et al. 2011), *qSS7* (Qiu et al. 2012), *tgw11* (Oh et al. 2010).

Next-generation sequencing (NGS) technologies have made genotyping more efficient and the availability of substantially increased SNP markers enables the exploration of QTL or related genes for target traits with no need of constructing a biparental segregating population (Chen et al. 2014; Yano et al. 2016). Genome-wide association study (GWAS) is not only conducted in human genetic but also in rice agronomic traits from 2010 (Huang et al. 2010). A number of genes related to grain weight and grain shape were identified by GWAS in rice, such as well-known *GW5* for TGW and GW (Liu et al. 2017), *OsLG3* for TGW and GL (Yu et al. 2017), and *GLW7* for GL and TGW (Si et al. 2016). GWAS is also becoming a dominant method to identify favorable alleles underlying agronomic traits using natural germplasm populations (McCouch et al. 2016; Zhai et al. 2018).

The 3,000 Rice Genomes Project (3 K RGP) contributed 29 million single nucleotide polymorphisms (SNPs), 2.4 million small InDels and over 90,000 structural variations. Its pan-genome analyses have exhibited at least 10,000 novel full-length protein-coding genes and a large amount of presence-absence variations (The-3K-rice-genomes-project 2014; Wang

et al. 2018). These available molecular genome data enabled the discovery of valuable genetic resources and high-throughput data for scientific research and practical molecular rice breeding (Wang et al. 2020; Zheng et al. 2015).

This project aimed to identify more candidate genes underlying rice grain weight and grain shape using a high-density association analysis followed by haplotype analysis of over 2,000 germplasms. Our results will enhance the knowledge of the genetic basis of grain weight and grain shape and provide valuable information for improving grain shape and grain yield in rice breeding program.

Materials And Methods

Plant materials

A total of 2445 accessions from 88 countries or regions were selected from the 3 K RGP (Additional file 2: Fig. S1). This panel consisted of seven types, including 1417 *xian* (*indica*) accessions (206 *xian*-1A, 179 *xian*-1B, 246 *xian*-2, 259 *xian*-3 and 527 intermediate *xian*), 261 temperate *geng* (*japonica*) accessions, 284 tropical *geng* (*japonica*) accessions, 82 subtropical *geng* (*japonica*) accessions, 77 intermediate *geng* (*japonica*) accessions, 81 accessions of intermediate type, 182 aus/boro accessions, and 61 basmati/sadri accessions.

Field trials and trait measurements

All accessions were grown in winter seasons during Nov 20–April 25 at Sanya (18.3°N, 109.3°E) in Hainan Province, China for three years (2015–2017). Each accession was planted in a three-row plot with 10 individuals in each row at a spacing of 20 cm × 25 cm with two replications. The field management followed the local farmers' standard practices. At the full-ripe stage (about 40 days after heading), 10 unanimous plants in the middle of each plot were harvested and air-dried in the screen house before testing. Around 300 seeds of each accession were then used to measure thousand grain weight (TGW, g), grain length (GL, mm), grain width (GW, mm) and the ratio of GL to GW (RLW) using the automatic seed scanning system (SC-G, Wanshen Technology Company, Hangzhou, China). The average trait values of each accession were used for data analyses.

LD decay and genome-wide association mapping

The 3 K RGP 4.8 M SNP dataset was downloaded from the Rice SNP-Seek Database (<http://www.ory-zasnp.org/>) (Alexandrov et al. 2015). SNPs with loci missing rate over 20% and minor allele frequency (MAF) less than 5% were filtered using plink software with the parameter '–geno 0.2', and '–maf 0.05' (Purcell et al. 2007). Finally, a total of 3,343,302 SNPs was used for the whole panel GWAS, and 2,240,362, 865,777, 1,548,277, 1,884,822 and 1,728,815 SNPs for *xian*, temperate *geng*, tropical *geng*, aus/boro and basmati/sadri GWAS, respectively. The effective number of independent markers (N) was calculated using the GEC software (Li et al. 2012) and suggestive *P*-value thresholds of association (1/N) were 1.76E-6, 2.40E-6, 1.03E-5, 5.63E-6, 4.08E-6 and 5.93E-6 for the whole panel, *xian*, temperate *geng*, tropical *geng*, aus/boro and basmati/sadri, respectively. The significance thresholds we adopted were $P = 1 \times 10^{-6}$ for the whole panel and $P = 1 \times 10^{-5}$ for subpopulations. We calculated r^2 , an estimation of linkage disequilibrium (LD), using plink software with the parameter '–r2 –ld-window-kb 1000 –ld-window 99999 –ld-window-r2 0' to deal with the SNPs. The LD decay rate was measured as the chromosomal distance at which the average r^2 dropped to half of its maximum value.

GWAS was performed with a linear mixed-effects model to determine the association between each SNP and the measured phenotypes using an efficient mixed-model analysis with EMMA eXpedited (EMMAX) software (Kang et al. 2010). An identical-by-state matrix based on genome-wide SNP data was used to create the kinship matrix that measured the genetic similarities between individuals. According to the previous significance thresholds for the whole panel and subgroups, we defined the SNP with the maximum *P*-value in a cluster that was considered as the lead SNP. Given the estimated LD decay rate of about 200 to 300 kb in rice (*O. sativa* L.) (Li et al. 2017), we determined the candidate regions

for each important QTL. Manhattan and Q–Q plots were created by the R package ‘qqman’ (Turner 2014) using the GWAS results. The R-ggplot2 package was used for creating the box plot.

Candidate gene identification

High-density SNP association analysis was carried out to detect candidate genes for important QTL. The following four criteria were conducted to identify important QTL: (1) coincidentally identified in at least two years; (2) affecting more than one trait; (3) coincidentally identified in different subpopulation; (4) located close to reported cloned genes or fine-mapped QTL. The following five steps were conducted to identify candidate genes for each novel QTL: 1) all available SNPs located within the QTL were acquired from 29 M SNPs data generated from 3 K RGP in the Rice SNP-Seek Database (Alexandrov et al. 2015); 2) the SNPs were filtered with the fore-mentioned parameters and the remaining high-quality SNPs were used to perform high-density association analyses through EMMA eXpedited (EMMAX) software; 3) for each QTL region, the SNPs with $-\log_{10}(P) > 6$ for the whole panel and > 5 for the subgroups were regarded as significant; 4) for the genes that harbor SNPs with $-\log_{10}(p)$ above the threshold or locate near the peak SNP were extracted from Rice Genome Annotation Project Database (RAP-DB) (Kawahara et al. 2013); 5) haplotype analysis was carried out for each candidate gene in each important QTL region using all non-synonymous SNPs located in the coding sequence (CDS) regions. Candidate genes were finally determined by testing the significant differences among major haplotypes (containing more than 1% panel samples number) through ANOVA.

Results

Trait variations and correlations

All four measured traits showed tremendous variations in the given subpopulations across the three years (Table 1 and Additional file 1: Table S1). Among 2445 accessions, subtropical *geng* presented the highest TGW among all the groups primarily due to significantly longer GL and wider GW than other groups, followed by *Xian-3* and intermediate *geng*. Although *xian-1B*, *xian-3*, subtropical *geng* and intermediate *geng* had similar GL, two *xian* groups (-1B and -3) showed significantly larger RLW than two *geng* groups (subtropical and intermediate). Each group also showed continuous distribution in above four traits, with IRIS_313-12340 having the highest TGW (46.26 g) and IRIS_313-12336 having the widest GW (4.20 mm) in subtropical *geng* group; IRIS_313-9817 having the longest GL (11.73 mm) in *xian-2* group; IRIS_313-15909 having the largest RLW (4.66) in basmati/sadri group;; and IRIS_313-8118 having the highest TGW (40.15 g) and longest GL (9.90 mm) in the temperate *geng* group.

Table 1
Performance of grain weight and grain shape in different subpopulations in 2015–2017

Trait ^a	Group	Mean ^b	Range	CV (%)	Elite accession of each trait
TGW (g)	<i>Xian-1A</i>	24.73 de	16.69–35.00	10.38	B149
	<i>Xian-1B</i>	24.16 ef	18.39–32.22	11.07	IRIS_313-11517
	<i>Xian-2</i>	24.55 e	12.20–36.06	15.82	IRIS_313-9817
	<i>Xian-3</i>	26.32 bc	13.00–38.41	19.64	IRIS_313-11081
	Intermediate <i>xian</i>	24.65 e	12.99–35.76	13.74	CX182
	Temperate <i>geng</i>	25.86 cd	17.47–40.15	12.39	IRIS_313-8118
	Subtropical <i>geng</i>	31.44 a	12.43–46.26	20.76	IRIS_313-12340
	Tropical <i>geng</i>	25.66 cd	8.20–36.31	16.52	IRIS_313-7959
	Intermediate <i>geng</i>	27.86 b	15.84–38.48	16.86	IRIS_313-8121
	Intermediate type	25.00 cde	15.50–35.09	14.78	IRIS_313-11189
	Aus/boro	22.78 f	11.73–32.48	16.92	IRIS_313-11232
	Basmati/sadri	19.78 g	11.54–33.75	22.55	CX110
GL (mm)	<i>Xian-1A</i>	7.75 d	5.94–9.65	7.66	B149
	<i>Xian-1B</i>	8.77 a	6.56–10.38	7.56	CX250
	<i>Xian-2</i>	8.02 c	5.17–11.73	10.77	IRIS_313-9817
	<i>Xian-3</i>	8.68 a	5.70–10.72	10.61	IRIS_313-11180
	Intermediate <i>xian</i>	8.33 b	6.11–10.89	9.61	IRIS_313-9841
	Temperate <i>geng</i>	7.5 de	6.34–9.90	10.44	IRIS_313-8118
	Subtropical <i>geng</i>	8.77 a	5.90–11.45	13.08	IRIS_313-11923
	Tropical <i>geng</i>	8.41 b	5.30–11.18	8.98	CX243
	Intermediate <i>geng</i>	8.52 ab	6.54–11.67	13.21	IRIS_313-10061
	Intermediate type	8.25 bc	6.35–10.68	12.89	IRIS_313-9818
	Aus/boro	7.67 de	5.54–9.24	9.15	IRIS_313-11481
	Basmati/sadri	7.28 e	5.04–10.90	18.61	CX110
GW (mm)	<i>Xian-1A</i>	3.10 cd	2.44–3.66	6.73	IRIS_313-11986
	<i>Xian-1B</i>	2.65 g	2.13–3.60	9.29	B058
	<i>Xian-2</i>	2.90 f	1.99–3.57	9.89	IRIS_313-10628
	<i>Xian-3</i>	2.94 f	2.14–3.92	11.82	IRIS_313-11898
	Intermediate <i>xian</i>	2.89f	2.02–3.89	9.57	IRIS_313-10739
^a TGW: thousand grain weight, GL: grain length, GW: grain width, RLW: ratio of grain length to grain width					
^b The letters indicate multiple comparisons result at the significant level 0.05					

Trait ^a	Group	Mean ^b	Range	CV (%)	Elite accession of each trait
	Temperate <i>geng</i>	3.30 b	2.37–3.88	7.09	B111
	Subtropical <i>geng</i>	3.51 a	2.44–4.20	10.77	IRIS_313-12336
	Tropical <i>geng</i>	3.04 de	2.23–4.15	12.7	IRIS_313-10645
	Intermediate <i>geng</i>	3.19 bc	2.35–3.82	11.85	B144
	Intermediate type	3.00 def	2.17–3.67	11.01	IRIS_313-11189
	Aus/boro	2.95 ef	2.11–3.77	10.2	IRIS_313-11483
	Basmati/sadri	2.71 g	2.12–3.43	13.84	IRIS_313-11289
RLW	<i>Xian</i> -1A	2.52 g	1.77–3.66	12.29	B239
	<i>Xian</i> -1B	3.36 a	2.24–4.32	13.24	IRIS_313–7808
	<i>Xian</i> -2	2.81 de	1.73–4.35	17.15	IRIS_313-12082
	<i>Xian</i> -3	3.00 b	1.99–4.19	15.03	IRIS_313-11683
	Intermediate <i>xian</i>	2.93 bc	1.95–4.30	15.16	CX363
	Temperate <i>geng</i>	2.31 h	1.81–3.75	15.67	IRIS_313–8160
	Subtropical <i>geng</i>	2.52 fg	1.74–3.63	14.18	IRIS_313-11923
	Tropical <i>geng</i>	2.84 cd	1.80–4.20	17.85	IRIS_313–8062
	Intermediate <i>geng</i>	2.75 def	1.75–4.48	22.17	IRIS_313-10061
	Intermediate type	2.83 cde	2.00–4.56	21.69	CX46
	Aus/boro	2.64 efg	1.73–4.25	14.95	IRIS_313–8390
	Basmati/sadri	2.77 def	1.88–4.66	26.16	IRIS_313-15909
^a TGW: thousand grain weight, GL: grain length, GW: grain width, RLW: ratio of grain length to grain width					
^b The letters indicate multiple comparisons result at the significant level 0.05					

The phenotypic pairwise correlations between the measured traits were similar in three years (Additional file1: Table S2). TGW was significantly positively correlated with GL and GW, while it showed slightly negative correlations with RLW.

Basic statistics of markers

For the 4.8 M high-quality SNP data, the number of markers per chromosome ranged from 197,757 on chromosome 9 to 422,795 on chromosome 1. The length of the chromosome varied from 23.01 Mb for chromosome 9 to 43.27 Mb for chromosome 1. The whole-genome size was 373.24 Mb with an average marker density of 9085 SNPs/Mb (Additional file 2: Fig. S2; Additional file 1: Table S3).

Population structure and linkage disequilibrium (LD) decay

The 3D plot showed that five principal components (PC) can be retained as effective, exhibiting the distribution of each accession in the rice diversity panel (Additional file 2: Fig. S3). The LD decay in the whole panel was much faster than any other panels. The maximum LD was 0.47, 0.64, 0.69, 0.53, 0.68 and 0.67 in the whole panel, aus/boro, basmati/sadri, *xian*, temperate *geng*, and tropical *geng*, respectively. LD reached half of its initial value at around 300 kb in the basmati/sadri,

temperate *geng* and the whole panel, 250 kb in the aus/boro panel, 130 kb in the *xian* panel and 260 kb in the tropical *geng* panel (Additional file 2: Fig. S4).

QTL detection by GWAS

A total of 22 important QTL were consistently identified for the four traits in at least two years (Table 2; Additional file 2: Fig. S5A–S5F). For TGW, six QTL were detected on chromosomes 3, 5, 7, 9 and 11. Among them, *qTGW3.2*, *qTGW5*, and *qTGW7* were stably detected in more than one panel in the three years. Two QTL, *qTGW3.1* and *qTGW9*, were detected only in the whole panel while *qTGW11* were just detected in the *xian* panel.

Table 2
Important QTL consistently identified for grain weight and grain shape at least two years by GWAS

Trait ^a	QTL	Chr.	Peak SNP ^a	QTL region (Mb)	Allele ^b	P	Panel	Year of QTL detected	Previously cloned gene/QTL
TGW (g)	<i>qTGW3.1</i>	3	rs3_4504988	4.30–4.70	C/T	2.59E-08	Whole	2015, 2016	
	<i>qTGW3.2</i>	3	rs3_16733441	16.53–16.93	T/G	2.25E-08	Whole	2015–2017	GS3 (Fan et al., 2006)
		3	rs3_16733441	16.60–16.86	T/G	1.78E-11	<i>Xian</i>	2015–2017	
	<i>qTGW5</i>	5	rs5_5371529	5.17–5.57	C/A	7.22E-10	Whole	2015–2017	<i>qSW5/GW5</i> (Liu et al., 2017)
		5	rs5_5361877	5.23–5.49	T/A	1.38E-10	<i>Xian</i>	2015–2017	
	<i>qTGW7</i>	7	rs7_28305040	28.10–28.50	A/G	8.12E-18	Whole	2015–2017	<i>FZP/BFL1</i> (Bai et al., 2017)
		7	rs7_28762557	28.63–28.89	G/A	7.37E-08	<i>Xian</i>	2015–2017	
		7	rs7_29143304	28.94–29.34	T/G	6.23E-06	Tropical <i>geng</i>	2016, 2017	
		7	rs7_28322588	28.12–28.52	T/C	1.04E-09	Basmati/sadri	2015–2017	
		7	rs7_28287280	28.08–28.48	A/G	9.14E-12	Aus/boro	2015–2017	
	<i>qTGW9</i>	9	rs9_21404841	21.20–21.60	C/T	2.47E-08	Whole	2015, 2017	
	<i>qTGW11</i>	11	rs11_3019935	2.88–3.14	T/C	5.17E-07	<i>Xian</i>	2016, 2017	
GL (mm)	<i>qGL3</i>	3	rs3_16733441	16.53–16.93	T/G	1.97E-73	Whole	2015–2017	GS3 (Fan et al., 2006)
		3	rs3_16733441	16.60–16.86	T/G	8.65E-59	<i>Xian</i>	2015–2017	
		3	rs3_16733441	16.53–16.93	T/G	2.54E-15	Temperate <i>geng</i>	2016, 2017	
		3	rs3_16733441	16.53–16.93	T/G	5.16E-09	Tropical <i>geng</i>	2015–2017	
		3	rs3_16728308	16.52–16.92	C/T	4.75E-09	Aus/boro	2016, 2017	

Note: For trait abbreviations, see Table 1.

^a The peak region in the chromosome.

^b Allele effect with respect to the minor allele.

Trait ^a	QTL	Chr.	Peak SNP ^a	QTL region (Mb)	Allele ^b	P	Panel	Year of QTL detected	Previously cloned gene/QTL
	<i>qGL4</i>	4	rs4_29308534	29.15–29.42	C/T	1.79E-06	Whole	2015, 2017	<i>qSW5/GW5</i> (Liu et al., 2017)
	<i>qGL5</i>	5	rs5_5373491	5.17–5.57	A/G	2.49E-21	Whole	2015–2017	
		5	rs5_5361894	5.23–5.49	A/G	2.09E-18	<i>Xian</i>	2015–2017	
		5	rs5_5351264	5.15–5.55	T/A	3.44E-08	Tropical <i>geng</i>	2016, 2017	
		5	rs5_5288638	5.08–5.48	G/A	2.58E-07	Aus/boro	2015–2017	
	<i>qGL7</i>	7	rs7_28289869	28.08–28.48	T/C	4.62E-19	Whole	2015–2017	<i>FZP/BFL1</i> (Bai et al., 2017)
		7	rs7_28787400	28.65–28.91	C/G	7.12E-07	<i>Xian</i>	2016, 2017	
		7	rs7_28290297	28.09–28.49	G/A	1.40E-10	Tropical <i>geng</i>	2015–2017	
		7	rs7_28287280	28.08–28.48	A/G	7.82E-08	Basmati/sadri	2016, 2017	
	<i>qGL10</i>	10	rs10_13616240	13.45–13.65	A/C	1.19E-07	Whole	2015–2017	
	<i>qGL11</i>	11	rs11_2044476	19.00–21.50	G/A	1.22E-06	Aus/boro	2016, 2017	
GW (mm)	<i>qGW3</i>	3	rs3_16733441	16.53–16.93	T/G	6.82E-09	Whole	2015–2017	<i>GS3</i> (Fan et al., 2006)
		3	rs3_16717406	16.58–16.84	G/A	2.21E-08	<i>Xian</i>	2015–2017	
	<i>qGW5</i>	5	rs5_5371772	5.17–5.57	G/A	1.84E-77	Whole	2015–2017	<i>qSW5/GW5</i> (Liu et al., 2017)
		5	rs5_5361894	5.23–5.49	A/G	1.96E-71	<i>Xian</i>	2015–2017	
		5	rs5_5364791	5.16–5.56	T/C	3.73E-10	Temperate <i>geng</i>	2015–2017	
		5	rs5_5375786	5.17–5.57	G/A	2.07E-25	Tropical <i>geng</i>	2015–2017	
		5	rs5_5346606	5.14–5.54	T/A	2.16E-19	Aus/boro	2015–2017	

Note: For trait abbreviations, see Table 1.

^a The peak region in the chromosome.

^b Allele effect with respect to the minor allele.

Trait ^a	QTL	Chr.	Peak SNP ^a	QTL region (Mb)	Allele ^b	<i>P</i>	Panel	Year of QTL detected	Previously cloned gene/QTL
	<i>qGW7</i>	7	rs7_24898274	24.69–25.09	A/T	2.59E-16	Whole	2015–2017	<i>GL7/GW7</i> (Wang et al. 2015a; Wang et al. 2015b)
		7	rs7_24902815	24.70–25.10	A/G	3.17E-07	Tropical <i>geng</i>	2015–2017	
	<i>qGW8</i>	8	rs8_26504638	26.30–26.70	A/G	3.58E-10	Aus/boro	2016, 2017	<i>GW8</i> (Wang et al., 2012)
RLW	<i>qRLW1</i>	1	rs1_3657795	3.45–3.85	A/G	1.42E-07	Aus/boro	2015, 2017	
	<i>qRLW3</i>	3	rs3_16733441	16.53–16.93	T/G	2.50E-55	Whole	2015–2017	<i>GS3</i> (Fan et al., 2006)
		3	rs3_16733441	16.60–16.86	T/G	4.00E-43	<i>Xian</i>	2015–2017	
		3	rs3_16733441	16.53–16.93	T/G	4.11E-07	Temperate <i>geng</i>	2015–2017	
		3	rs3_16728485	16.52–16.92	G/C	6.65E-11	Tropical <i>geng</i>	2016, 2017	
		3	rs3_16686074	16.48–16.88	G/A	7.33E-08	Basmati/sadri	2015–2017	
	<i>qRLW4</i>	4	rs4_29317460	29.15–29.42	C/A	3.68E-09	Whole	2015–2017	
		4	rs4_29309086	29.10–29.50	C/G	1.14E-08	Aus/boro	2015–2017	
	<i>qRLW5</i>	5	rs5_5371609	5.17–5.57	A/G	3.59E-69	Whole	2015–2017	<i>qSW5/GW5</i> (Liu et al., 2017)
		5	rs5_5361894	5.23–5.49	A/G	1.25E-67	<i>Xian</i>	2015–2017	
		5	rs5_5364791	5.16–5.56	T/C	9.49E-12	Temperate <i>geng</i>	2015–2017	
		5	rs5_5375793	5.17–5.57	G/A	5.48E-25	Tropical <i>geng</i>	2016, 2017	
		5	rs5_5375764	5.17–5.57	A/G	5.56E-15	Aus/boro	2015–2017	

Note: For trait abbreviations, see Table 1.

^a The peak region in the chromosome.

^b Allele effect with respect to the minor allele.

Trait ^a	QTL	Chr.	Peak SNP ^a	QTL region (Mb)	Allele ^b	P	Panel	Year of QTL detected	Previously cloned gene/QTL
	<i>qRLW7</i>	7	rs7_24533303	24.33–24.73	T/C	4.89E-13	Whole	2015–2017	<i>GL7/GW7</i> (Wang et al. 2015a; Wang et al. 2015b)
		7	rs7_24537476	24.40–24.66	C/T	3.67E-07	<i>Xian</i>	2015–2017	
		7	rs7_24902815	24.70–25.10	A/G	8.02E-07	Tropical <i>geng</i>	2016, 2017	
	<i>qRLW8</i>	8	rs8_26504638	26.30–26.70	A/G	1.17E-14	Aus/boro	2015–2017	<i>GW8</i> (Wang et al., 2012)

Note: For trait abbreviations, see Table 1.

^a The peak region in the chromosome.

^b Allele effect with respect to the minor allele.

Six QTL for GL were mapped on chromosomes 3–5, 7, 10 and 11. The *qGL3*, *qGL5* and *qGL7* were consistently detected in more than one panel in the three years. The *qGL4* and *qGL10* were just detected in the whole panel whereas *qGL11* was only detected in the aus/boro panel.

Four QTL were detected for GW on chromosomes 3, 5, 7 and 8. Three QTL, *qGW3*, *qGW5*, and *qGW7* were stably detected in more than one panel in the three years. One QTL, *qGW8* was detected only in aus/boro panel over the three years.

Six QTL affecting RLW were identified on chromosomes 1, 3–5, 7 and 8. Four QTL, *qRLW3*, *qRLW4*, *qRLW5*, and *qRLW7* were consistently detected in more than one panel in the three years. Two QTL, *qRLW1* and *qRLW8* were detected only in the aus/boro panel.

We specified QTL with overlapping QTL regions identified for different traits or in different years as the same QTL, which lead to the identification of 12 different QTL (*qTGW3.1*, *qTGW3.2/qGL3/qGW3/qRLW3*, *qTGW5/qGL5/qGW5/qRLW5*, *qTGW7/qGL7*, *qTGW9*, *qTGW11*, *qGL4/qRLW4*, *qGL10*, *qGL11*, *qGW7/qRLW*, *qGW8/qRLW8* and *qRLW1*).

Candidate genes for important QTL

The seven newly identified QTL regions were used for gene-based association analysis and haplotype analysis to identify candidate genes. In the region of 4.30–4.70 Mb on chromosome 3 harboring *qTGW3.1*, 9,177 non-synonymous SNPs of 38 genes were used for gene-based association analysis in the whole panel. SNPs with $-\log_{10}(p)$ above threshold 6.0 centered around four candidate genes, *Os03g0184400*, *Os03g0185200*, *Os03g0186600* and *Os03g0189400* (Fig. 1a). Haplotype analyses indicated that *Os03g0189400* showed no significant differences between various haplotypes (Additional file2: Fig. S8). Significant differences for TGW were observed among different haplotypes at *Os03g0186600*, *Os03g0184400* and *Os03g0185200* in the whole population. Of them, Hap3 with average TGW of 26.81 g, 26.73 g and 26.81 g for *Os03g0184400*, Hap5 with average TGW of 28.19 g, 28.86 g and 28.90 g for *Os03g0185200*, and Hap3 with average TGW of 26.03 g, 26.13 g and 26.51 g for *Os03g0186600* were favorable alleles across the three years (Fig. 1b–d).

The *qTGW9* was identified for TGW in the region of 21.20–21.60 Mb on chromosome 9, 9,593 non-synonymous SNPs of 35 genes were used for gene-based association analysis in the whole panel. SNPs with $-\log_{10}(p)$ above threshold 6.0 centered around three candidate genes, *Os09g0544300*, *Os09g0544400*, and *Os09g0544800* (Fig. 2a). No haplotype

differences were found in CDS regions of *Os09g0544300*, and *Os09g0544800* (Additional file2: Fig. S9). Three major haplotypes within *Os09g0544400* in the CDS region were detected in the whole panel with Hap1 associated with the highest TGW of 24.89 g, 25.13 g, and 25.47 g across the three years (Fig. 2b).

The *qTGW11* was detected for TGW in the region from 2.88 Mb to 3.14 Mb on chromosome 11 in the *xian* panel, harboring 9,948 non-synonymous SNPs of 26 genes. SNPs with $-\log_{10}(p)$ above the threshold 5.0 hits only one candidate genes, *Os11g0163600* (Fig. 3a). However, significant differences in *Os11g0163600* in the CDS region were detected only in 2016 and 2017 (Additional file2: Fig. S10). Hap2 of the three major haplotypes was the favorable allele with the TGW of 24.78 g, 25.29 g, and 25.40 g over the three years (Fig. 3b).

For *qGL4/qRLW4* in the region of 29.15–29.42 Mb on chromosome 4, 9,638 non-synonymous SNPs of 23 genes were used for gene-based association analysis in the whole panel. *Os04g0580500*, *Os04g0580700*, and *Os04g0580800* were presumed to be candidate genes (Fig. 4a–b). No haplotype differences for GL or RLW were found in the CDS region of *Os04g0580500* and *Os04g0580800* (Additional file2: Fig. S11). However, significant differences in GL and RLW were found in the CDS regions of *Os04g0580700*. Six haplotypes at *Os04g0580700* were identified with Hap5 having the highest GL and Hap1 having the highest RLW (Fig. 4c).

qGL10 was predicted in the region of 13.45–13.65 Mb on chromosome 10 covering 15,535 non-synonymous SNPs of 13 genes. These SNPs were used for gene-based association analysis in the whole panel. Two genes (*Os10g0399700* and *Os10g0400100*) harboring SNPs with $-\log_{10}(P)$ larger than the threshold 6.0 were identified (Fig. 5a). Highly significant differences for *qGL10* among different haplotypes in the CDS regions were observed at the two genes in the whole population (Additional file2: Fig. S12). Three major haplotypes for *Os10g0399700* and four major haplotypes for *Os10g0400100* were detected, and Hap3 of *Os10g0399700* (8.39 mm, 8.42 mm and 8.70 mm) and *Os10g0400100* (8.39 mm, 8.43 mm and 8.71 mm) were associated with significantly longest GL in the three years (Fig. 5b–c).

In the region from 19.00 to 21.50 Mb on chromosome 11 harboring *qGL11*, 7,866 non-synonymous SNPs of 32 genes were used for gene-based association analyses in the aus/boro panel. SNPs with $-\log_{10}(p)$ above the threshold 5.0 were found to locate in two candidate genes, *Os11g0143700* and *Os11g0144500* (Fig. 6a). Haplotype differences were found in the CDS region only for *Os11g0143700* (Additional file2: Fig. S13). Two major haplotypes of *Os11g0143700* were found. Of them, Hap2 for *Os11g0143700* (7.78 mm, 7.80 mm and 8.21 mm) was associated with significantly longer GL in the three years (Fig. 6b).

On chromosome 1, a high peak of *qRLW1* was found, 8,116 non-synonymous SNPs of 23 genes were used for gene-based association analysis in the predicted candidate region from 3.45–3.85 Mb in the aus/boro panel. SNPs with $-\log_{10}(p)$ above threshold 5.0 centered around two candidate genes, *Os01g0171000* and *Os01g0171100* (Fig. 7a). Haplotype differences were found in the CDS regions of both *Os01g0171000* and *Os01g0171100* (Additional file2: Fig. S14). Hap2 for *Os01g0171000* and Hap2 for *Os01g0171100* were found to harbor the highest RLW (Fig. 7b–c).

Discussion

Variations of grain weight and grain shape in the germplasms

TGW, GL and RLW are three components that can vary dramatically depending on the rice variety while the grain width (GW) and grain thickness (GT) are much less variable (Huang et al. 2013). Genetically, most genes underlying TGW, GL, GW and RLW are additive, although dominant effects have also been detected in some studies (Tan et al. 2000). Grain length has been shown to contribute more to grain weight than other grain shape traits (Shi and Shen 1996). As shown in this study, major QTL/genes also showed pleiotropic effects on grain shape and grain weight. For instance, two loci

(*qTGW3.2/qGL3/qGW3/qRLW3* and *qTGW5/qGL5/qGW5/qRLW5*) had pleiotropic effects on all the four traits, and *qTGW7/qGL7* showed pleiotropic effects on grain weight and grain length.

Each of the 12 sub-populations (*xian-1A*, *xian-1B*, *xian-2*, *xian-3*, intermediate *xian*, temperate *geng*, subtropical *geng*, tropical *geng*, intermediate *geng*, intermediate type, aus/boro, and basmati/sadri) carried different combinations of grain size and shape alleles, and some sub-populations contained significantly more genetic variations for grain size and shape than others. Overall, subtropical *geng* group had the highest average TGW (31.44 g), GL (8.77 mm) and GW (3.51 mm) while *xian-1B* group has the highest average RLW (3.36) (Table 1 and Additional file 1: Table S1). The overall lowest average TGW (19.78 g), GL (7.28 mm) and GW (2.71 mm) were detected in basmati/sadri group, and the overall lowest average RLW was found in *xian-1A* group, which is equivalent to *Ind1* subpopulation originated from South China (Xie et al. 2015). Basmati/sadri group showed a higher variation in both grain weight and grain shape, whereas *geng* group showed less variation than *xian* group (Additional file 1: Supplementary Table S1). The power to obtain different combinations of alleles conferring particular grain shapes and sizes from almost each sub-population has implications for improvement of yield and grain quality, potentially enabling breeders to develop high-yielding varieties with specific grain quality to satisfy diverse quality requirements.

Comparison of the QTL detected in this study with the previously cloned genes

Comparisons of QTL detected in this study with the previously cloned genes were performed within 1 Mb physical distance. Of the 12 different important QTL, five QTL intervals were found to locate in the same regions with the previously cloned genes in rice (Table 2). *qTGW3.2/qGL3/qGW3/qRLW3* on chromosome 3 was mapped close to *GS3*, a gene for grain size previously identified in the *xian* varieties (Fan et al. 2006). *qTGW5/qGL5/qGW5/qRLW5* was mapped to the *qSW5/GW5* gene which regulates grain width and weight (Liu et al. 2017). *qGW7/qLWR7* was co-located with the gene *GL7/GW7* which controls GL and GW (Wang et al. 2015a; Wang et al. 2015b). *qTGW7/qGL7* was at the same position to the gene *FZP* which regulates the number of spikelets per panicle, GL and TGW (Bai et al. 2017). *qGW8/qLWR8* was mapped close to *GW8* which controls grain size and quality (Wang et al. 2012). The allelisms of above five loci identified in this study with the previously reported QTL or genes need to be further tested by fine-mapping and cloning.

It is notable that all the five cloned genes except *GW8* which was only detected in aus/boro, were detected in diverse subgroups/panels in the current study, suggesting that abundant variations may exist in these genes. Several important structural and functional features have been revealed for the *GS3* protein in grain size regulation (Mao et al. 2010). Almost all the cloned genes mentioned above are widely used in rice breeding programs, so it is valuable to mine excellent alleles of these loci from diverse germplasms for rice breeding of high yield and grain quality. Haplotype analyses showed that Hap1 (GT) of *GS3*, Hap3 (GCGG) of *GW5*, Hap2 (A) of *FZP*, Hap2 (CGACAG) of *GL7/GW7* and Hap5 (TCCAGTGG) of *GW8* are the favorable alleles (Additional file 2: Figure S15). More than 60% accessions with Hap1 (GT) of *GS3* and Hap2 (A) of *FZP* were come from *xian* subpopulation, while interestingly, tropical *geng* occupied over 70% of Hap3 (GCGG) of *GW5*, Hap2 (CGACAG) for *GL7/GW7* and Hap5 (TCCAGTGG) for *GW8* (Fig. 8).

Candidate gene identification of the important QTL

Map-based cloning of QTL is still complex and time-consuming nowadays. In this study, the strategy of GWAS and gene-based association analysis combining with haplotype analysis of candidate genes was employed. Eleven candidate genes governing seven new QTL were shortlisted.

Os03g0186600 is the most likely candidate genes for *qTGW3.1*, which encodes a rice AG-like MADS-box protein, *OsMDP1*, and involves in brassinosteroid signaling and seed development (Duan et al. 2006). The candidate gene *Os09g0544400* of *qTGW9* encodes a glutathione S-transferase which is a C-terminal domain-containing protein. Due to a new-type GNAT-

like protein that harbors intrinsic histone acetyltransferase activity (*OsglHAT1*) for *GW6a* which has a crucial impact on grain weight and yield (Song et al. 2015), Glutathione S-transferase expression regulated by *Os09g0544400* may have latent impacts on grain shape and weight. Only one candidate gene *Os11g0163600* was identified in *qTGW11* region, and it encodes a F-box and tubby domain containing protein. The candidate gene *Os04g0580700* (*OsMADS17*) for *qGL4/qRLW4* encodes a rice MADS-box protein and plays a critical role in rice floral organ identity specification and floral meristem determinacy (Hu et al. 2015). Of the two candidate genes for *qGL10*, *Os10g0399700* encodes a cystathionine gamma-synthase while *Os10g0400100* encodes a methionyl-tRNA synthetase. More evidences are needed to prove the causal relationship between these two gene products and grain shape traits. The candidate gene *Os11g0143700* for *qGL11* is an unknown function gene without functional annotation. Of the two candidate genes for *qRLW1*, *Os01g0171000* and *Os01g0171100*, *Os01g0171000* is more probable as it encodes *BRASSINOSTEROID INSENSITIVE 1*-associated receptor kinase 1 precursor. Previous research suggested that loss of function of a rice *BRASSINOSTEROID INSENSITIVE1* ortholog, *OsBRI1*, leading to small grain size but the same grain yield as that of wild type (Morinaka et al. 2006). Therefore, *Os03g0186600*, *Os09g0544400*, *Os11g0163600*, *Os04g0580700*, *Os10g0399700*, *Os10g0400100*, *Os11g0143700* and *Os01g0171000* were ultimately presumed as the most likely candidate genes for grain weight and grain shape (Table 3). Transgenic experiments are under way to verify the functionalities and evidence of their genetic pathways underlying grain weight and grain shape is needed.

Table 3
List of 8 candidate genes for 7 important QTL associated with grain weight and grain shape

QTL	Candidate gene ^a	Function
<i>qTGW3.1</i>	<i>Os03g0186600</i>	<i>OsMDP1</i> , MADS-box transcription factor gene
<i>qTGW9</i>	<i>Os09g0544400</i>	Glutathione S-transferase, C-terminal domain containing protein
<i>qTGW11</i>	<i>Os11g0163600</i>	OsFBT13-F-box and tubby domain containing protein
<i>qGL4/qRLW4</i>	<i>Os04g0580700</i>	<i>OsMADS17</i> , MADS-box family gene
<i>qGL10</i>	<i>Os10g0399700</i>	Cystathionine gamma-synthase
	<i>Os10g0400100</i>	Methionyl-tRNA synthetase
<i>qGL11</i>	<i>Os11g0143700</i>	Unknown
<i>qRLW1</i>	<i>Os01g0171000</i>	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase 1 precursor
^a Gene names from the MSU Rice Genome Annotation Project Database		

Implication in rice breeding for grain yield and quality

Among three major components (panicle number per plant, number of grains per panicle and grain weight) of grain yield, the most reliable trait is grain weight, which is measured as the TGW. Grain shape is mainly determined by a combination of GL, GW, and RLW. These three parameters are positively correlated with grain weight (Tan et al. 2000). Besides its contribution to grain yield, grain shape is an important quality trait that has a major impact on the market values of rice grain products. In the current study, eight promising genes were identified to be significantly associated with grain weight or grain shape and the linked SNP variations in CDS regions were identified. Information on the distributions of different haplotypes for grain weight and grain shape in various subpopulations may offer more options for various breeding strategies. The SNP information combined with Kompetitive Allele Specific PCR (KASP) technique (Semagn et al. 2014) could be an excellent choice for pyramiding multiple genes/QTL controlling grain weight and grain shape. Of course, precise genome-editing techniques in plants are an alternative important tool for molecular plant breeding (Zhang et al. 2018; Zhang et al. 2020). The application of the CRISPR/Cas9-derived system on *GW5* has proven that genome editing can verify the result of GWAS precisely in such short time (Liu et al. 2017). The genetic information of the favorable

haplotypes in our study and their further testified functional SNPs will provide useful resources for precise gene editing breeding. Additionally, dozens of elite rice accessions carrying combinations of favorable haplotypes of the eight new candidate and five cloned genes were identified (Additional file 1: Table S4). These accessions showing high TGW (> 30 g) and RLW (> 3.0) could be used as donor parents for breeding and genetic materials for functional investigation.

Conclusion

Abundant variations and excellent haplotypes for grain weight and grain shape exist among 12 sub-populations. Seven new QTL (*qTGW3.1*, *qTGW9*, *qTGW11*, *qGL4/qRLW4*, *qGL10*, *qGL11* and *qRLW1*) were identified for grain weight and grain shape by GWAS. The most likely eight candidate genes were identified via gene-based association analysis, haplotype analysis, and functional annotation. The accessions with excellent haplotypes at the eight candidate genes and five previously cloned genes were also identified. The results will enhance our knowledge of the genetic basis of grain weight and grain shape in rice and facilitate exploring the molecular mechanisms underlying the traits. The exploitation of glorious haplotypes and germplasm resources provide valuable information for improving rice grain yield and grain quality by MAS breeding or genome-editing.

Abbreviations

GL: grain length; GW:grain width; TGW:thousand grain weight; RLW:ratio of grain length to grain width; 3 K RGP:3,000 Rice Genome Project; ANOVA:analysis of variance; GWAS:genome-wide association study; QTL:quantitative trait locus/loci; SNP:single nucleotide polymorphism; MAF:minor allele frequency; kb:kilobyte; Mb:megabyte; LD:linkage disequilibrium; PC:principal components; CV:coefficient of variation; CDS:coding DNA sequence; SD:standard deviation.

Declarations

Acknowledgements

We thank Guoliang Cao and Shuangbing Zhu for field management and Laiyuan Zhai for seed collection.

Authors' contributions

JX conceived and supervised the study. YN, KC, CS, HC collected data, CW, ZW and TZ conducted analyses, YN and TC wrote the manuscript. TC and JX drew and revised the manuscript. All authors read and approved the final manuscript.

Funding

This work was funded by the National Key R&D Program of China (2016YFD0100301, 2017YFD0100100) and the 863 Key Project from the Chinese Ministry of Science & Technology (Grant No. 2014AA10A601), and the Agricultural Science and Technology Innovation Program and the Cooperation and Innovation Mission (Grant No. CAAS-ZDXT202001)

Availability of data and material

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request. Seeds of 3 K can be requested from JX.

Ethics approval and consent to participate

Not applicable.

Consent for publication

All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no conflict interests.

Author details

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

²Guangdong Laboratory of Lingnan Modern Agriculture, Genome Analysis Laboratory of the Ministry of Agriculture, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, China

³Pingxiang Institute of Agricultural Sciences, Pingxiang 337000, China

References

1. Alexandrov N, Tai S, Wang W, Mansueto L, Palis K, Fuentes RR, Ulat VJ, Chebotarov D, Zhang G, Li Z, Mauleon R, Hamilton RS, McNally KL (2015) SNP-Seek database of SNPs derived from 3000 rice genomes. *Nucleic Acids Res* 43:D1023–D1027
2. Ashikari M, Wu J, Yano M, Sasaki T, Yoshimura A (1999) Rice gibberellin-insensitive dwarf mutant gene Dwarf 1 encodes the alpha-subunit of GTP-binding protein. *Proc Natl Acad Sci USA* 96:10284–10289
3. Bai X, Huang Y, Hu Y, Liu H, Zhang B, Smaczniak C, Hu G, Han Z, Xing Y (2017) Duplication of an upstream silencer of FZP increases grain yield in rice. *Nature Plants* 3:885–893
4. Chen H, Xie W, He H, Yu H, Chen W, Li J, Yu R, Yao Y, Zhang W, He Y, Tang X, Zhou F, Deng XW, Zhang Q (2014) A high-density SNP genotyping array for rice biology and molecular breeding. *Mol Plant* 7:541–553
5. Duan K, Li L, Hu P, Xu S, Xu Z, Xue H (2006) A brassinolide-suppressed rice MADS-box transcription factor, *OsMDP1*, has a negative regulatory role in BR signaling. *Plant J* 47:519–531
6. Fan C, Xing Y, Mao H, Lu T, Han B, Xu C, Li X, Zhang Q (2006) *GS3*, a major QTL for grain length and weight and minor QTL for grain width and thickness in rice, encodes a putative transmembrane protein. *Theor Appl Genet* 112:1164–1171
7. Gong J, Miao J, Zhao Y, Zhao Q, Feng Q, Zhan Q, Cheng B, Xia J, Huang X, Yang S, Han B (2017) Dissecting the genetic basis of grain shape and chalkiness traits in hybrid rice using multiple collaborative populations. *Mol Plant* 10:1353–1356
8. Hu Y, Liang W, Yin C, Yang X, Ping B, Li A, Jia R, Chen M, Luo Z, Cai Q, Zhao X, Zhang D, Yuan Z (2015) Interactions of *OsMADS1* with floral homeotic genes in rice flower development. *Mol Plant* 8:1366–1384
9. Huang R, Jiang L, Zheng J, Wang T, Wang H, Huang Y, Hong Z (2013) Genetic bases of rice grain shape: so many genes, so little known. *Trends Plant Sci* 18:218–226
10. Huang X, Wei X, Sang T, Zhao Q, Feng Q, Zhao Y, Li C, Zhu C, Lu T, Zhang Z, Li M, Fan D, Guo Y, Wang A, Wang L, Deng L, Li W, Lu Y, Weng Q, Liu K, Huang T, Zhou T, Jing Y, Li W, Lin Z, Buckler ES, Qian Q, Zhang Q, Li J, Han B (2010)

Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat Genet* 42:961–976

11. Kang HM, Sul JH, Service SK, Zaitlen NA, Kong S-y, Freimer NB, Sabatti C, Eskin E (2010) Variance component model to account for sample structure in genome-wide association studies. *Nat Genet* 42:348–354
12. Kawahara Y, de la Bastide M, Hamilton JP, Kanamori H, McCombie WR, Ouyang S, Schwartz DC, Tanaka T, Wu J, Zhou S, Childs KL, Davidson RM, Lin H, Quesada-Ocampo L, Vaillancourt B, Sakai H, Lee SS, Kim J, Numa H, Itoh T, Buell CR, Matsumoto T (2013) Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. *Rice* 6:4
13. Li M-X, Yeung JMY, Cherny SS, Sham PC (2012) Evaluating the effective numbers of independent tests and significant p -value thresholds in commercial genotyping arrays and public imputation reference datasets. *Hum Genet* 131:747–756
14. Li X, Guo Z, Lv Y, Cen X, Ding X, Wu H, Li X, Huang J, Xiong L (2017) Genetic control of the root system in rice under normal and drought stress conditions by genome-wide association study. *Plos Genetics* 13:e1006889
15. Li Y, Fan C, Xing Y, Jiang Y, Luo L, Sun L, Shao D, Xu C, Li X, Xiao J, He Y, Zhang Q (2011) Natural variation in *GS5* plays an important role in regulating grain size and yield in rice. *Nat Genet* 43:1266–1269
16. Liu J, Chen J, Zheng X, Wu F, Lin Q, Heng Y, Tian P, Cheng Z, Yu X, Zhou K, Zhang X, Guo X, Wang J, Wang H, Wan J (2017) *GW5* acts in the brassinosteroid signalling pathway to regulate grain width and weight in rice. *Nature Plants* 3:17043
17. Mao H, Sun S, Yao J, Wang C, Yu S, Xu C, Li X, Zhang Q (2010) Linking differential domain functions of the *GS3* protein to natural variation of grain size in rice. *Proc Natl Acad Sci USA* 107:19579–19584
18. McCouch SR, Wright MH, Tung C-W, Maron LG, McNally KL, Fitzgerald M, Singh N, DeClerck G, Agosto-Perez F, Korniliev P, Greenberg AJ, Naredo MEB, Mercado SMQ, Harrington SE, Shi Y, Branchini DA, Kuser-Falcão 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
19. Morinaka Y, Sakamoto T, Inukai Y, Agetsuma M, Kitano H, Ashikari M, Matsuoka M (2006) Morphological alteration caused by brassinosteroid insensitivity increases the biomass and grain production of rice. *Plant Physiol* 141:924–931
20. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, Sham PC (2007) PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81:559–575
21. Ray DK, Mueller ND, West PC, Foley JA (2013) Yield trends are insufficient to double global crop production by 2050. *PLOS ONE* 8:e66428
22. Semagn K, Babu R, Hearne S, Olsen M (2014) Single nucleotide polymorphism genotyping using Kompetitive Allele Specific PCR (KASP): overview of the technology and its application in crop improvement. *Mol Breeding* 33:1–14
23. Shi C, Shen Z (1996) Additive and dominance correlation analysis of grain shape and yield traits in Indica rice. *Acta Agronomica Sinica* 22:36–42
24. Si L, Chen J, Huang X, Gong H, Luo J, Hou Q, Zhou T, Lu T, Zhu J, Shangguan Y, Chen E, Gong C, Zhao Q, Jing Y, Zhao Y, Li Y, Cui L, Fan D, Lu Y, Weng Q, Wang Y, Zhan Q, Liu K, Wei X, An K, An G, Han B (2016) *OsSPL13* controls grain size in cultivated rice. *Nat Genet* 48:447–456
25. Song X, Huang W, Shi M, Zhu M, Lin H (2007) A QTL for rice grain width and weight encodes a previously unknown RING-type E3 ubiquitin ligase. *Nat Genet* 39:623–630
26. Song X, 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 J, 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. *Proc Natl Acad Sci USA* 112:76–81

27. Tan Y, Xing Y, Li J, Yu S, Xu C, Zhang Q (2000) Genetic bases of appearance quality of rice grains in Shanyou 63, an elite rice hybrid. *Theor Appl Genet* 101:823–829
28. The-3K-rice-genomes-project (2014) The 3,000 rice genomes project. *GigaScience* 3:2047–2217X
29. Turner SD (2014) qqman: an R package for visualizing GWAS results using QQ and manhattan plots. *Biorxiv*:005165
30. Wang C, Yu H, Huang J, Wang W, Faruquee M, Zhang F, Zhao X, Fu B, Chen K, Zhang H, Tai S, Wei C, McNally KL, Alexandrov N, Gao X, Li J, Li Z, Xu J, Zheng T (2020) Towards a deeper haplotype mining of complex traits in rice with RFGB v2.0. *Plant Biotechnol J* 18:14–16
31. Wang S, Li S, Liu Q, Wu K, Zhang J, Wang S, Wang Y, Chen X, Zhang Y, Gao C, Wang F, Huang H, Fu X (2015a) The *OsSPL16-GW7* regulatory module determines grain shape and simultaneously improves rice yield and grain quality. *Nat Genet* 47:949–954
32. Wang S, Wu K, Yuan Q, Liu X, Liu Z, Lin X, Zeng R, Zhu H, Dong G, Qian Q, Zhang G, Fu X (2012) Control of grain size, shape and quality by *OsSPL16* in rice. *Nat Genet* 44:950–954
33. Wang W, Mauleon R, Hu Z, Chebotarov D, Tai S, Wu Z, Li M, Zheng T, Fuentes RR, Zhang F, Mansueto L, Copetti D, Sanciangco M, Palis KC, Xu J, Sun C, Fu B, Zhang H, Gao Y, Zhao X, Shen F, Cui X, Yu H, Li Z, Chen M, Detras J, Zhou Y, Zhang X, Zhao Y, Kudrna D, Wang C, Li R, Jia B, Lu J, He X, Dong Z, Xu J, Li Y, Wang M, Shi J, Li J, Zhang D, Lee S, Hu W, Poliakov A, Dubchak I, Ulat VJ, Borja FN, Mendoza JR, Ali J, Li J, Gao Q, Niu Y, Yue Z, Naredo MEB, Talag J, Wang X, Li JJ, Fang X, Yin Y, Glaszmann JC, Zhang J, Li J, Hamilton RS, Wing RA, Ruan J, Zhang G, Wei C, Alexandrov N, McNally KL, Li Z, Leung H (2018) Genomic variation in 3,010 diverse accessions of Asian cultivated rice. *Nature* 557:43–49
34. Wang Y, Xiong G, Hu J, Jiang L, Yu H, Xu J, Fang Y, Zeng L, Xu E, Xu J, Ye W, Meng X, Liu R, Chen H, Jing Y, Wang Y, Zhu X, Li J, Qian Q (2015b) Copy number variation at the *GL7* locus contributes to grain size diversity in rice. *Nat Genet* 47:944–948
35. Xie W, Wang G, Yuan M, Yao W, Lyu K, Zhao H, Yang M, Li P, Zhang X, Yuan J, Wang Q, Liu F, Dong H, Zhang L, Li X, Meng X, Zhang W, Xiong L, He Y, Wang S, Yu S, Xu C, Luo J, Li X, Xiao J, Lian X, Zhang Q (2015) Breeding signatures of rice improvement revealed by a genomic variation map from a large germplasm collection. *Proceedings of the National Academy of Sciences* 112:E5411
36. Yano K, Yamamoto E, Aya K, Takeuchi H, Lo P-c, Hu L, Yamasaki M, Yoshida S, Kitano H, Hirano K, Matsuoka M (2016) Genome-wide association study using whole-genome sequencing rapidly identifies new genes influencing agronomic traits in rice. *Nat Genet* 48:927–934
37. Yu J, Xiong H, Zhu X, Zhang H, Li H, Miao J, Wang W, Tang Z, Zhang Z, Yao G, Zhang Q, Pan Y, Wang X, Rashid MAR, Li J, Gao Y, Li Z, Yang W, Fu X, Li Z (2017) *OsLG3* contributing to rice grain length and yield was mined by Ho-LAMap. *Bmc Biology* 15:28
38. Zhai L, Zheng T, Wang X, Wang Y, Chen K, Wang S, Wang Y, Xu J, Li Z (2018) QTL mapping and candidate gene analysis of peduncle vascular bundle related traits in rice by genome-wide association study. *Rice* 11:13
39. Zhang Y, Massel K, Godwin ID, Gao C (2018) Applications and potential of genome editing in crop improvement. *Genome Biol* 19:210
40. Zhang Y, Pribil M, Palmgren M, Gao C (2020) A CRISPR way for accelerating improvement of food crops. *Nature Food* 1:200–205
41. Zheng T, Yu H, Zhang H, Wu Z, Wang W, Tai S, Chi L, Ruan J, Wei C, Shi J (2015) Rice functional genomics and breeding database (RFGB)-3K-rice SNP and InDel sub-database. *Chin Sci Bull* 60:367–371

Figures

Figure 1

(a) Gene-based association analysis of *qTGW3.1* in 2015, 2016 and 2017. Solid line showed the threshold to determine significant SNP. (b–d) Gene structural and haplotypes analysis of targeted genes *Os03g0184400*, *Os03g0185200* and *Os03g0186600* for TGW in the whole panel

Figure 2

(a) Gene-based association analysis of *qTGW9* in 2015, 2016 and 2017. Solid line showed the threshold to determine significant SNP. (b) Gene structural and haplotypes analysis of targeted gene *Os09g0544400* for TGW in the whole panel

Figure 3

(a) Gene-based association analysis of *qTGW11* in 2015, 2016 and 2017. Solid line showed the threshold to determine significant SNP. (b) Gene structural and haplotypes analysis of targeted gene *Os11g0163600* for TGW in the xian panel

Figure 4

(a–b) Gene-based association analysis of *qGL4*/*qRLW4* in 2015, 2016 and 2017. Solid line showed the threshold to determine significant SNP. (c) Gene structural and haplotypes analysis of targeted gene *Os04g0580700* for GL and RLW in the whole panel

Figure 5

(a) Gene-based association analysis of *qGL10* in 2015, 2016 and 2017. Solid line showed the threshold to determine significant SNP. (b–c) Gene structural and haplotypes analysis of targeted genes *Os10g0399700* and *Os10g0400100* for GL in the whole panel

Figure 6

(a) Gene-based association analysis of *qGL11* in 2015, 2016 and 2017. Solid line showed the threshold to determine significant SNP. (b) Gene structural and haplotypes analysis of targeted gene *Os11g0143700* for GL in the aus/boro panel

Figure 7

(a) Gene-based association analysis of *qRLW1* in 2015, 2016 and 2017. Solid line showed the threshold to determine significant SNP. (b) Gene structural and haplotypes analysis of targeted genes *Os01g0171000* and *Os01g0171100* for RLW in the aus/boro panel

Figure 8

Subpopulation distribution of the accessions assigned to different haplotype for GS3 (a), GW5(b)FZP(c), GL7 (d), and GW8 (e)

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Additionalfile2.pptx](#)
- [Additionalfile1.xlsx](#)