



Discovery and Validation of Grain Shape Loci in U.S. Rice Germplasm Through Haplotype Characterization

Brijesh Angira^{1*}, Tommaso Cerioli^{2†} and Adam N. Famoso^{1*}

¹H. Rouse Caffey Rice Research Station, Louisiana State University Agricultural Center, Baton Rouge, LA, United States, ²School of Plant, Environmental and Soil Science, Louisiana State University, Baton Rouge, LA, United States

OPEN ACCESS

Edited by:

Jyothi Badri,
Indian Institute of Rice Research
(ICAR), India

Reviewed by:

Yingpeng Han,
Northeast Agricultural University,
China
Vanisri Satturu,
Professor Jayashankar Telangana
State Agricultural University, India

*Correspondence:

Brijesh Angira
bangira@agcenter.lsu.edu
Adam N. Famoso
AFamoso@agcenter.lsu.edu

†Present address:

Tommaso Cerioli,
Corteva Agriscience, Cremona,
Lombardy, Italy

Specialty section:

This article was submitted to
Plant Genomics,
a section of the journal
Frontiers in Genetics

Received: 18 April 2022

Accepted: 10 June 2022

Published: 12 September 2022

Citation:

Angira B, Cerioli T and Famoso AN
(2022) Discovery and Validation of
Grain Shape Loci in U.S. Rice
Germplasm Through
Haplotype Characterization.
Front. Genet. 13:923078.
doi: 10.3389/fgene.2022.923078

Rice grain shape is a major determinant of rice market value and the end-use. We mapped quantitative trait loci (QTL) for grain shape traits in a bi-parental recombinant inbred line population (Trenasse/Jupiter) and discovered two major grain length QTLs—*qGL3.1* and *qGL7.1*. Previously, a major grain shape gene *GS3* was reported in the *qGL3.1* region and grain length gene *GL7* was reported to be encompassing *qGL7.1* locus. The re-sequencing SNP data on the International Rice Research Institute (IRRI) 3K Rice Genome Project (RGP) panel were obtained from the IRRI SNP-Seek database for both genes and haplotype diversity was characterized for each gene in this diverse panel. United States rice germplasm was not well represented in the IRRI 3K RGP database. Therefore, a minimum SNP set was identified for each gene that could differentiate all the characterized haplotypes. These haplotypes in the 3K RGP panel were screened across 323 elite U.S. genotypes using the minimum SNP set. The screening of haplotypes and phenotype association confirmed the role of *GS3* under *qGL3.1*. However, screening of the *GL7* haplotypes in the U.S. germplasm panel showed that *GL7* did not play a role in *qGL7.1*, and in addition, *GL7.1* did not segregate in the Trenasse/Jupiter RIL population. This concluded that *qGL7.1* is a novel QTL discovered on chr7 for grain shape in the Trenasse/Jupiter RIL population. A high-throughput KASP-based SNP marker for each locus (*GS3* and *qGL7.1*) was identified and validated in elite U.S. rice germplasm to be used in an applied rice breeding program.

Keywords: rice, grain shape, grain length, QTL, haplotypes, high-throughput KASP marker, SNP

INTRODUCTION

Rice is a major source of calories for human diets, and the direct consumption of the grain makes grain quality a major component for rice industries. Grain shape is a major grain quality component of rice and directly influences market value. Consumers make decisions based on the cooking quality and presumed taste of rice depending on its grain length, grain width, aroma, and whiteness. Regions, their cultures, and rice cuisines also drive the preference for different types of rice. Rice is classified by grain types based on the length and width of the grains. However, there is no global standard for rice classification based on shape and size. In the United States, medium grain is predominantly grown in California, accounting for 70% of the U.S. medium-grain rice acres. Long grain is predominantly grown in the southern United States, accounting for 99% of the U.S. long-grain rice acres (USDA and National Agricultural Statistics Service, 2021).

Two primary subspecies or varietal groups—*Indica* and *Japonica*—represent two genetically distinct genetic pools of *O. sativa* (Oka, 1988; Wang and Tanksley, 1989; Sun et al., 2002; Childs, 2014). The varietal group *Indica* is adapted and predominantly grown in tropical latitudes; whereas, *Japonica* is grown in both temperate and tropical climates (Mackill and Lei, 1997). Each varietal group can be further divided into genetic subpopulations. The varietal group *Japonica* comprises *temperate japonica*, *tropical japonica*, and *aromatic* subpopulations, while *Indica* comprises *indica* and *aus* subpopulations (Second, 1985; Glaszmann, 1987; Garris et al., 2005; Caicedo et al., 2007; McCouch et al., 2016). Significant variation for grain length and shape exists between and within each group, and certain subpopulations are associated with specific grain classes. For example, the *temperate japonica* subpopulation is associated with short and round grains, *aus* with short and slender grains, *indica* with long, slender grains, and *aromatic* with both very short and extra-long, fragrant rice (Juliano and Villareal, 1993). Almost all cultivated varieties in the United States belong to the *Japonica* varietal group. United States long-grain rice is *tropical japonica*, while U.S. medium-grain rice is a mix of *temperate* and *tropical japonica* (Lu et al., 2005).

Grain shape traits are controlled by variation in naturally occurring genes that are often first identified as quantitative trait loci (QTLs). Some grain shape genes in rice are also responsible for grain weight and yield. Chen et al. (2022) reported *KRN2/OsKRN2* genes in rice and maize and their study showed that a complete loss of function allele of these genes increased grain numbers and yield without a negative impact on other agronomic traits. Many major genes and QTLs have been reported for grain shape traits in rice (Sun et al., 2018). The first grain size QTL to be cloned and characterized was *GS3* (Fan et al., 2006; Takano-Kai et al., 2009). *GS3* encodes a heterotrimeric G protein that interacts with other G-protein subunits as part of a signaling pathway that regulates grain size in rice (Botella, 2012; Sun et al., 2018). The wild-type *GS3* protein produces medium grain, a loss-of-function allele results in long grain, while truncated forms lacking the C-terminus produce very short grain (Mao et al., 2010; Takano-Kai et al., 2013). Other grain length QTLs have been identified on chromosome (chr) three using different mapping populations. These include *GW3.1* (Li et al., 2004), *GL3.1/GL3* (Qi et al., 2012), *qGL3* (Zhang et al., 2012), and *TGW3/GL3.3* (Hu et al., 2018; Xia et al., 2018). Several grain length QTLs have also been reported on chr7, including *qGL7* (Bai et al., 2010), *qGL7.2* (Shao et al., 2010), *GLW7/OsSPL13* (Si et al., 2016), *SLG7* (Zhou et al., 2015), and *GL7/GW7* (Wang S. et al., 2015). Copy number variation at the *GL7* locus contributes to grain size diversity in *O. sativa* (Wang Y. et al., 2015).

Numerous Grain width and grain weight loci are also found on different chromosomes, including *GW2* (Song et al., 2007), *GIF1* (Wang et al., 2008), *GW5/qGW5/SW5*, (Shomura et al., 2008; Weng et al., 2008), *GS5* (Li et al., 2011), *GW6a* (Song et al., 2015), *TGW6* (Ishimaru et al., 2013), *OsSPL16/GW7* (Wang S. et al., 2015), and *GW8/OsSPL16* (Wang et al., 2012).

Haplotype characterization approach is a powerful tool that help in understanding a particular germplasm and identify and

validate trait SNPs for applied breeding purposes. High-throughput SNP markers were identified and validated for semi-dwarf gene *sd1* (Angira et al., 2019), Cercospora disease locus *CRSP2.1* (Addison et al., 2021), and an aroma gene *BADH2* (Addison et al., 2020) of rice using haplotype characterization approach in the recent studies. Similar haplotype analysis approaches were reported in other crops (Maldonado et al., 2019; Brinton et al., 2020; Bruce et al., 2020; Marsh et al., 2022) and proven to be vital in understanding the trait genetic architecture in a particular germplasm. We developed KASP (Kompetitive allele specific polymerase) SNP markers using the haplotype characterization approach. KASP is an allele-specific oligo extension and fluorescence-based cost-effective genotyping technology (Kumputla et al., 2012; Semagn et al., 2014). In this study, we identified and validated SNPs for the identified loci and developed the SNPs as KASP assays to be used in an applied breeding program.

MATERIALS AND METHODS

United States Germplasm Panel

A United States rice germplasm panel (U.S. germplasm panel) was compiled to characterize grain shape phenotypes. Additionally, this panel was used to discover and validate grain shape loci and trait markers. The panel consists of 323 lines and represents the genetic diversity present in modern and elite southern U.S. rice germplasm. The panel consists of key historical varieties and germplasm, advanced breeding lines, and modern released varieties from the rice breeding programs of Louisiana, Arkansas, Texas, Mississippi, California, and Missouri.

“Trenasse” x ‘Jupiter’ Recombinant Inbred Line Population

A bi-parental recombinant inbred line (RIL) population consisting of 286 RILs was developed from a cross between the varieties Trenasse (T) and Jupiter (J) and is referred to as the ‘TJ RIL’ population. Trenasse is a typical U.S. long-grain variety (Linscombe et al., 2006), and Jupiter is a typical U.S. medium-grain variety (Sha et al., 2006). The TJ RIL population was developed using single seed decent breeding method and derived at the F8 generation. Jupiter is widely grown medium grain variety in the southern United State and Trenasse, released in 2006 and still exists in the pedigree of modern long grain varieties. The reported average yield of Trenasse and Jupiter was 8089 kg ha⁻¹ across 21 statewide trials during 2001–2003 (Linscombe et al., 2006) and 8734 kg ha⁻¹ across 37 statewide trials during 2002–2004 (Sha et al., 2006), respectively.

Public Database

For this study, we used the International Rice Informatics Consortium SNP-Seek (IRRI SNP-Seek) database available at snp-seek.irri.org (Alexandrov et al., 2014; Mansueto et al., 2016, 2017). The SNP-Seek portal hosts multiple genomic

datasets and formats including the sequencing data of the 3,000 rice genomes project (3K RGP) (Li et al., 2014) and the High-Density Rice Array (HDRA) dataset (McCouch et al., 2016). The HDRA dataset contains 700,000 SNPs (single nucleotide polymorphism) genotyped on 1,568 diverse rice accessions, including 92 U.S. rice varieties (McCouch et al., 2016). Of the five SNP sets available for panel—All (32M), All biallelic (29M), Base (18M), Filtered (4.8M), and Core (404k) (Mansueto et al., 2017)—this study utilized the 18M SNP set (3KBase) as well as the HDRA dataset to characterize haplotypes in the U.S. germplasm diversity panel.

Phenotyping Methods

The TJ RIL population ($n = 286$) was planted at the H. Rouse Caffey Rice Research Station (HRCRRS) near Crowley, LA in 2017 and 2018 and phenotyped for grain traits—grain length and grain width. The U.S. germplasm panel was planted at the HRCRRS in 2019 and phenotyped for grain traits. The panel was phenotyped only for 1 year because the grain dimension traits are highly heritable and the environment does not have a significant effect on the traits (Fan et al., 2006; Gao et al., 2016). Both the TJ RILs and the U.S. germplasm panels were planted with a Hege seed drill, and each row was 1.8 m in length. Standard agronomic practices were followed according to the Louisiana Rice Production handbook (LSU Agcenter, 2014). The SeedCount SC6000R - Reflectance Image Analysis System (nextinstruments.net) was used to phenotype grain traits: length, width, and length/width ratio (LWR). The instrument uses software and flatbed scanner technology to rapidly and accurately analyze the physical characteristics of a grain sample. The instrument reliably eliminates broken and double grains in a scanning plate.

Quantitative Trait Loci Mapping for Grain Dimension Traits

Quantitative trait loci mapping was conducted using the TJ RIL population. The population was genotyped using KASP SNP assays at the HRCRRS. Linkage map construction and QTL mapping of grain shape traits were executed using composite interval Mapping implemented in QTL IciMapping software V4.1 (Meng et al., 2015). The Kosambi mapping function was used for linkage map development, and the significant logarithm of the odds (LOD) threshold was calculated using 1000 permutations and Type I error rate. Additive QTLs were detected using a 1.0 cM stepwise regression with a 0.001 probability threshold.

Haplotype Characterization of Candidate Genes Across the 3K RGP Panel

Single nucleotide polymorphism data for the *GS3* and *GL7* genes were extracted from the IRRI SNP-Seek database using the 3K RGP dataset and 3kbase dataset filters (iric.irri.org; snp-seek.irri.org) (Alexandrov et al., 2014; Mansueto et al., 2017). The dataset consisted of 3,024 genotypes representing the global diversity of rice germplasm including *aus*, *indica*, *indica admix*, *japonica admix*, *tropical japonica*, *temperate japonica*, and *aromatic*

subpopulation groups. However, modern U.S. germplasm is not well represented in this dataset, which contains only 32 U.S. varieties. Therefore, the haplotypes identified in the 3K RGP diversity panel have to be surveyed across the U.S. rice germplasm panel.

SNP data for the *GS3* gene (*LOC_OS03G0407400*), located at 16,729,501–16,735,109 base pair (bp) (IRGSP v.1) on chr3, consisted of 169 SNPs identified in the 3,024 genotypes available in the SNP-Seek portal. Heterozygous SNP calls were considered as missing data. Single nucleotide polymorphisms having missing data >0.2 across the 3,024 germplasm samples were omitted from the analysis. Next, samples having >0.2 missing SNP calls were removed from further analysis, and then SNPs with minor allele frequency <0.001 were omitted. After these filtering steps, 2,426 samples remained, and these had genotyping data at 96 SNPs, which were organized into haplotype groups. A haplotype is defined as having identical alleles at all 96 SNP loci. A haplotype number was assigned to haplotypes found in more than 30 samples (>0.012); haplotypes found in less than 30 samples were considered rare haplotypes. This SNP haplotype characterization identified eight *GS3* haplotypes, which accounted for 92% of the 2,426 samples.

Similar haplotype characterization steps were followed for the *GL7* grain shape gene. Single nucleotide polymorphism data for the gene *GL7* (*LOC_Os07g41200*), located at 24,664,168–24,669,324 bp (IRGSP v.1) on chr7, consisted of 82 SNPs identified in the 3,024 genotypes available in the SNP-Seek portal. Heterozygous SNP calls were considered as missing data. Single nucleotide polymorphisms having >0.2 missing data across 3,024 germplasm samples were omitted. Next, samples having >0.2 missing data across the remaining SNPs were removed from further consideration. In the final filtering step, SNPs having minor allele frequency <0.001 were omitted. Following these filtering steps, 2,993 samples remained, and they had genotyping data at 71 SNPs which were then assigned to haplotype groups. Haplotypes present in less than 35 (<0.012) samples were classified as rare haplotypes. Samples that could not be classified into haplotype groups due to missing data were assigned to the Unclassified group. Nine haplotype groups accounted for 96% of the 2,993 samples. Similar haplotype characterization methods in the 3K RGP panel were previously reported by Angira et al. (2019) and Addison et al. (2021).

Haplotype Characterization of Identified Grain Length Genes Across U.S. Germplasm

Upon assigning haplotypes for the *GS3* and *GL7* genes in the 3K RGP panel, a set of minimum SNPs was developed that accurately differentiated all haplotypes. The minimum SNP sets consisted of seven and eight SNPs for *GS3* and *GL7* genes, respectively. Flanking sequences (100 bp each side) for each SNP were extracted using the Ricebase database (Edwards et al., 2016), and these SNPs were developed into KASP assays for screening across the U.S. germplasm panel. Kompetitive allele specific polymerase chain reaction genotyping was conducted on the

U.S. germplasm panel at the HRCRRS following protocols developed for the LGC SNPLINE system (biosearchtech.com/products/instruments-and-consumables/genotyping-instruments/snpline-genotyping-automation, accessed 22 October 2021). The minimum SNP set markers were screened across the U.S. germplasm panel for both genes to document the frequency of the haplotype groups characterized using the 3K RGP panel.

Haplotype Characterization of the *qGL7.1* QTL Region in HDRA Dataset

To explore haplotype variation across the more extensive *qGL7.1* QTL region, SNP data were extracted from the HDRA dataset using the SNP-Seek portal. The dataset consisted of 1,568 germplasm samples, and 763 SNPs were detected in the 475 Kb *qGL7.1* QTL region. To explore haplotype variation only in modern U.S. germplasm, we focused on the 34 U.S. genotypes in the HDRA dataset that were released after 1980. Single nucleotide polymorphisms having >0.15 missing data across the 34 U.S. samples, samples having >0.2 missing data across the SNPs, SNPs with minor allele frequency <0.02, and monomorphic SNPs were omitted from the further analysis. After these filtering steps, genotyping data for 358 SNPs on 34 U.S. lines remained for analysis. Haplotypes at a frequency of ≥ 0.05 were characterized in this dataset and a minimum SNP set was developed that could accurately classify all of the samples into the correct haplotype groups. This minimum SNP set was developed into KASP assays and screened across the U.S. germplasm panel.

Interaction of *GS3* and *qGL7.1* in the TJ RIL Population, Segregating Populations, and Advanced Breeding Lines

The TJ RIL population was used to examine the interaction between the *GS3* and *qGL7.1* genes. The RILs were classified into four classes—Class 1: *GS3*-Long/*qGL7.1*-Long; Class 2: *GS3*-Long/*qGL7.1*-Med; Class 3: *GS3*-Med/*qGL7.1*-Long; and Class 4: *GS3*-Med/*qGL7.1*-Med, based on the alleles at both loci. Grain length phenotypes were correlated with these classes to understand the effect of the combination of alleles at both loci. We also used 78 advanced breeding lines that were developed by crossing long- and medium-grain varieties. These varieties were phenotyped for grain length and grain width, in addition to yield, for 2 years.

RESULTS

Quantitative Trait Loci Mapping for Grain Dimension Traits

Quantitative trait loci (QTL) mapping was conducted for grain dimension traits using 286 RILs derived from a cross between Trenasse (long-grain) and Jupiter (medium-grain) based on phenotype data collected over 2 years. The population was genotyped with 139 KASP SNPs, and a genetic linkage map of

1547 cM was constructed. Two major QTLs were detected for grain length (*qGL3.1* and *qGL7.1*), grain width (*qGW3.1* and *qGW7.1*), and LWR (*qGLWR3.1* and *qGLWR7.1*). The long-grain parent, Trenasse, contributed the alleles for increased grain length and LWR at both QTLs. The medium-grain parent, Jupiter, contributed the alleles for increased grain width at both grain width QTLs. All QTLs were detected within and across both years. Major QTLs for grain length, *qGL3.1* and *qGL7.1*, explained about 53% and 22% of phenotypic variation, respectively (Table 1).

The QTLs detected for the three traits co-localized to the same two regions on chr3 and chr7, with both regions encompassing previously identified grain shape genes in rice. The grain dimension QTLs detected on chr3 were mapped between 16.73 and 24.00 Mb, a region that harbors the major grain size gene *GS3* (Fan et al., 2006, 2009; Takano-Kai et al., 2009). The QTLs detected on chr7 were mapped between 24.85 and 25.32 Mb, which is 176 kb from the closest flanking marker of the *GL7* locus (Wang Y. et al., 2015). Grain dimension traits (grain length, grain width, and LWR) were highly correlated and the QTLs for both the traits were co-localized. Thus, we focused on only grain length trait in this study. However, we explored all the reported grain shape traits in the QTLs. To investigate whether the genes *GS3* and *GL7* underlie the QTLs detected in this study, *qGL3.1* and *qGL7.1*, respectively, the haplotypes of each gene were characterized using the global rice diversity 3K RGP panel and evaluated across the U.S. germplasm panel that had been phenotyped for grain dimension traits.

Haplotype Characterization of Candidate Genes Across the 3K RGP Panel

The IRRI SNP-Seek database contains re-sequencing data for 3,024 diverse rice samples (3K RGP panel). However, the U.S. rice germplasm is not well represented in this panel. To leverage the 3K RGP re-sequencing data for use evaluating our specific germplasm of interest, we first characterized the haplotype diversity of the *GS3* and *GL7* genes across the 3K RGP panel and identified the minimum set of SNPs necessary to distinguish the haplotypes of each gene. These minimum sets of SNPs for both genes were subsequently used to develop KASP SNP assays to evaluate the TJ RIL mapping population and the U.S. rice germplasm diversity panel.

Haplotype Characterization of the *GS3* Gene Across the 3K RGP Panel

A total of 169 SNPs were identified in the *GS3* gene (*LOC_OS03G0407400*) based on the re-sequencing data for the 3K RGP panel. After filtering SNPs and genotypes based on missing data (>0.2) and minor allele frequency (<0.001), a dataset was extracted consisting of genotypes for 2,426 samples evaluated at 96 SNP loci (Figure 1A, Supplementary Table S1). Based on these 96 SNPs, eight haplotype groups were identified with a frequency of >0.012. These eight haplotypes described the variation in *GS3* found in 2,222 (92%) of the samples. Haplotypes in 204 samples (8%) could either not be classified

TABLE 1 | Grain dimension Quantitative trait loci (QTL) detected in the Trenasse/Jupiter Recombinant Inbred Line population.

Trait	QTL name	chr	Pos (cM)	QTL interval start			QTL interval stop			LOD	PVE	Add	Donor ^a
				Marker name	Pos (cM)	IRGSP V.1	Marker name	Pos (cM)	IRGSP V.1				
Length	<i>qGL3.1</i>	3	89	LSU-963	88.58	16,733,441	LSU-92	111.08	23,999,431	74.91	53.34	-0.36	Trenasse
Length	<i>qGL5.1</i>	5	63	LSU-48	46.58	16,473,206	LSU-329	66.57	19,504,928	5.03	2.67	0.08	Jupiter
Length	<i>qGL7.1</i>	7	124	LSU-323	124.00	24,845,026	LSU-159	125.84	25,319,809	43.00	22.36	-0.23	Trenasse
Width	<i>qGW3.1</i>	3	49	LSU-1253	48.04	9,828,999	LSU-309	53.52	10,500,404	6.21	4.11	0.03	Jupiter
Width	<i>qGW3.2</i>	3	89	LSU-963	88.58	16,733,441	LSU-92	111.08	23,999,431	18.60	13.11	0.06	Jupiter
Width	<i>qGW3.3</i>	3	177	LSU-10	171.96	32,674,916	LSU-240	178.10	34,454,081	6.10	4.14	0.03	Jupiter
Width	<i>qGW7.1</i>	7	124	LSU-323	124.00	24,845,026	LSU-159	125.84	25,319,809	39.92	32.70	0.09	Jupiter
Width	<i>qGW8.1</i>	8	200	LSU-445	199.62	18,966,446	LSU-481	214.61	21,589,285	9.29	6.01	-0.04	Trenasse
Width	<i>qGW12.1</i>	12	81	LSU-1266	70.13	24,395,807	LSU-265	81.52	26,133,920	5.99	3.80	0.03	Jupiter
LWR	<i>qLWR3.1</i>	3	49	LSU-1253	48.04	9,828,999	LSU-309	53.52	10,500,404	5.99	2.28	-0.05	Trenasse
LWR	<i>qLWR3.2</i>	3	89	LSU-963	88.58	16,733,441	LSU-92	111.08	23,999,431	63.29	38.24	-0.19	Trenasse
LWR	<i>qLWR5.1</i>	5	14	LSU-587	12.11	3,341,107	LSU-340	28.11	6,607,675	6.42	2.63	0.05	Jupiter
LWR	<i>qLWR7.1</i>	7	124	LSU-323	124.00	24,845,026	LSU-159	125.84	25,319,809	58.98	33.12	-0.18	Trenasse
LWR	<i>qLWR8.1</i>	8	203	LSU-445	199.62	18,966,446	LSU-481	214.61	21,589,285	7.13	3.07	0.06	Jupiter

A total of 286 RILs, were phenotyped for grain length and grain width for 2 years and genotyped with 139 Kompetitive Allele-Specific PCR (KASP) single nucleotide polymorphism (SNP) markers. ICMMapping software was used to map the QTLs.

LOD, logarithm of the odds; PVE, percent variation explained; Add = additive effect; Pos = position; chr = chromosome.

^aDonor indicates the parent that contributed the allele for the increased value of the trait in the Jupiter/Trenasse RIL, population.

due to missing data or were rare; these samples were omitted from further analysis. Based on the haplotype assignments from the 96 SNPs, a subset of seven SNPs (minimum SNP set) was identified that could accurately differentiate the eight haplotype groups and correctly assign samples to each group (Figure 1A). This minimum SNP set was used to represent the haplotype diversity in the *GS3* gene.

For this study, we were most interested in haplotypes that could be used to differentiate *Japonica* samples, as U.S. rice germplasm is predominantly of *Japonica* ancestry. *GS3*-Hap1 was the most common SNP haplotype across the 3K RGP panel, found in 1,047 samples, of which about 58% were of *Indica* ancestry ($n = 601$), 40% were of *Japonica* ancestry ($n = 424$), and 2% were admixtures ($n = 22$) (Figure 1A, Supplementary Table S1). Within the *tropical japonica* subpopulation, *GS3*-Hap1 was the predominant haplotype, accounting for 76% ($n = 332$) of *tropical japonica* genotypes. *GS3*-Hap1 was distinguished from all other haplotypes by the previously reported casual C-A mutation allele in the second exon of *GS3* at 16,733,441 bp (*GS3*-SNP1), containing the long-grain “T” allele (Fan et al., 2009; Takano-Kai et al., 2009). *GS3*-Hap3 was present in 329 genotypes, of which 73% belonged to the *Japonica* varietal group ($n = 240$), 21% were *Indica* ($n = 68$), and 6% were admixed ($n = 21$). *GS3*-Hap3 was the third most common haplotype across the 3K RGP panel and the predominant haplotype within the *temperate japonica* subpopulation, accounting for 52% ($n = 139$) of all *temperate japonica* lines (Figure 1A, Supplementary Table S1). Based on the association of *GS3* haplotypes to specific subpopulations, we anticipated that the *GS3*-Hap1 allele would be predominant in U.S. long-grain germplasm and that *GS3*-Hap3 would be predominant in U.S. medium-grain germplasm. Other common haplotypes observed in the 3K RGP panel were predominant in

samples of *Indica* ancestry and were rare or not present in *Japonica* samples.

Haplotype Characterization of the *GL7* Gene Across 3K RGP Panel

A total of 78 SNPs were identified in the *GL7* gene (LOC_Os07g41200) based on the re-sequencing data for the 3K RGP panel. After filtering SNPs and genotypes based on missing data (>0.2) and minor allele frequency (<0.001), a dataset consisting of genotypes for 2,993 samples evaluated at 71 SNPs was extracted. Based on these 71 SNPs, nine haplotypes were identified with a frequency >0.012 . These nine haplotypes described the variation in *GL7* found in 2,873 (96%) samples (Figure 1B, Supplementary Table S2). Haplotypes in 120 samples (4%) could either not be classified due to missing data or were rare ($<1.5\%$), so these samples were omitted from further analysis. Based on the haplotype assignments from the 71 SNPs, a subset of eight SNPs (minimum SNP set) was identified that could accurately differentiate the nine haplotype groups. This minimum SNP set was used to represent haplotype variation in the *GL7* gene (Figure 1B).

GL7-Hap1 was the most common SNP haplotype in the 3K RGP panel and was found in 1,257 samples, of which 97% were of *Indica* ancestry and about 3% were admixtures. *GL7*-Hap2 was present in 595 samples, of which 90% belonged to the *Japonica* varietal group, 7% were *Indica*, and the remaining 3% were admixed (Figure 1B, Supplementary Table S2). *GL7*-Hap2 was the second most common haplotype group in the panel and the most common haplotype in *Japonica* samples. *GL7*-Hap2 was present in 93% of *tropical japonica* samples in the 3K RGP panel. *GL7*-Hap3 was present in 337 samples, of which 52% belonged to the *temperate japonica* subpopulation. *GL7*-Hap3 was the predominant haplotype within *temperate japonica*,

A

GS3 Haplotype ID	GS3-SNP1 (16,733,441)	GS3-SNP2 (16,734,715)	GS3-SNP3 (16,732,914)	GS3-SNP4 (16,731,092)	GS3-SNP5 (16,734,015)	GS3-SNP6 (16,730,370)	GS3-SNP7 (16,731,513)											Total
	LSU 963	LSU 964	LSU 967	LSU 969	LSU 977	LSU 1301	LSU 1302											
GS3-Hap1	T	T	C	G	A	A	C	1	412	188	22	32	332	45	15		1,047	
GS3-Hap2	G	C	C	A	A	A	C	1	350	188	12			1	1		553	
GS3-Hap3	G	T	C	G	A	A	C	40	24	4	21	139	44	19	38		329	
GS3-Hap4	G	C	C	G	A	G	C	3	66	15	1						85	
GS3-Hap5	G	C	C	G	A	A	C	58	12	2	2	1					75	
GS3-Hap6	G	T	T	G	A	A	T				2	60		4			66	
GS3-Hap7	G	T	T	G	A	A	C				2	20	8	3			33	
GS3-Hap8	G	C	C	A	G	A	C		23	10	1						34	
Rare								14	48	48	9	16	50	7	12		204	
Total								117	935	455	72	268	434	79	66		2,426	

B

GL7 Haplotype ID	GL7-SNP1 (24,666,092)	GL7-SNP2 (24,668,947)	GL7-SNP3 (24,669,200)	GL7-SNP4 (24,667,874)	GL7-SNP5 (24,668,019)	GL7-SNP6 (24,667,793)	GL7-SNP7 (24,664,520)	GL7-SNP8 (24,668,887)										Total
	LSU 902	LSU 908	LSU 911	LSU 921	LSU 916	LSU 914	LSU 1303	LSU 1305										
GL7-Hap1	T	C	A	T	G	T	T	C	6	807	405	36	2			1	1,257	
GL7-Hap2	C	A	C	T	G	T	T	C		26	14	18	39	449	47	2	595	
GL7-Hap3	T	A	C	T	G	T	T	C	72	35	18	12	176	2	16	6	337	
GL7-Hap4	T	C	C	T	G	T	T	C		170	83	2					255	
GL7-Hap5	T	C	A	T	G	T	T	A	100	3	13	12			1	34	163	
GL7-Hap6	T	A	C	C	G	T	T	C		2	28	1	62	1	12		106	
GL7-Hap7	T	C	A	T	G	C	T	C		65	5	1					71	
GL7-Hap8	T	C	A	T	A	T	T	C		40	5						45	
GL7-Hap9	T	A	C	T	G	T	C	C	8	4	12	4				16	44	
Rare									12	18	23	6	4	30	1	15	109	
Unclassified									2	1			4		4		11	
Total									200	1,171	606	92	287	482	81	74	2,993	

FIGURE 1 | SNP data was obtained from SNP-Seek online portal of the International Rice Research Institute for GS3 and GL7 genes and haplotypes were characterized based on specified criteria with each gene. **(A)** Haplotype characterization of the GS3 gene in IRRI SNP-Seek 3K RGP genotype data. **(B)** Haplotype characterization of the GL7 gene in IRRI SNP-Seek 3K RGP genotype data.

accounting for 61% of all *temperate japonica* samples. Based on these observations, we anticipated that the GL7-Hap2 would be the most common haplotype within U.S. long-grain germplasm and GL7-Hap3 would predominate within U.S. medium-grain germplasm.

Haplotype Characterization of GS3 and GL7 in U.S. Germplasm

The haplotypes of GS3 and GL7 observed in the 3K RGP panel were investigated in the U.S. germplasm panel to determine whether they underlie the *qGS3.1* and *qGL7.1* QTLs detected

in the TJ RIL population. The minimum SNP sets that differentiated GS3 (7 SNPs) and GL7 (8 SNPs) haplotypes were converted into KASP assays and used to screen the U.S. germplasm panel.

Of the eight GS3 haplotypes observed in the 3K RGP panel, only two haplotypes were observed within the U.S. rice germplasm panel (Figure 2A, Supplementary Table S3). GS3-Hap1 was present in 281 genotypes, all of which were previously classified as long-grain germplasm. GS3-Hap3 was present in 42 genotypes, all were previously classified as medium or short grain germplasm. A comparison of grain length in samples carrying GS3-Hap1 or GS3-Hap3 was highly significant ($p < 0.0001$). The

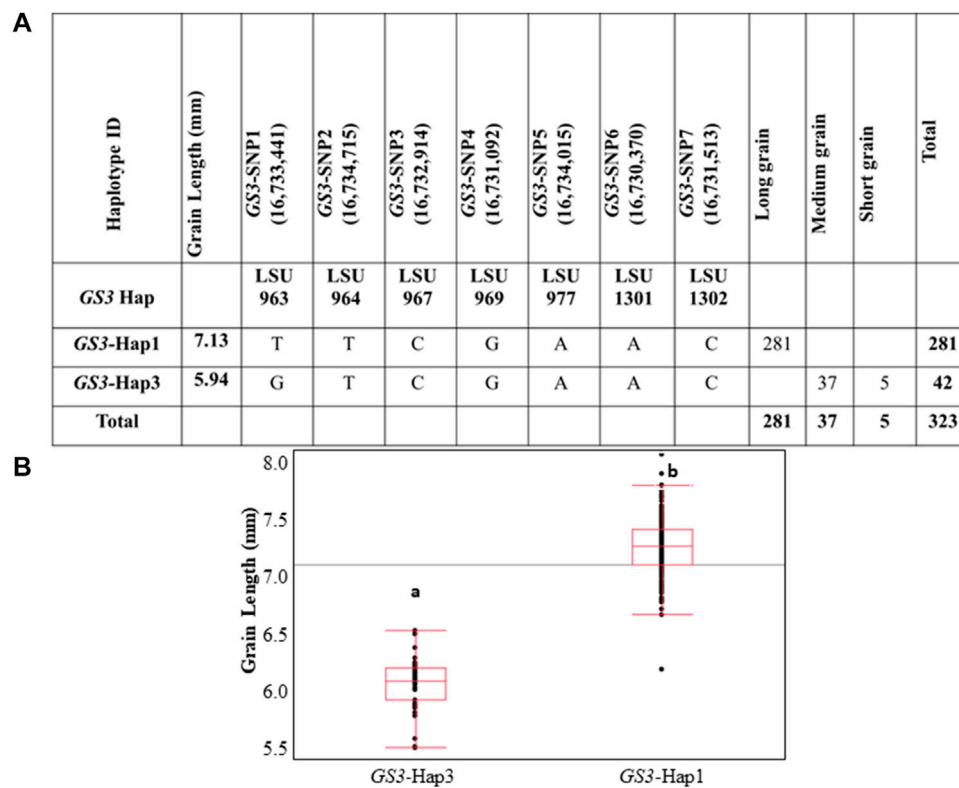


FIGURE 2 | (A) Haplotype characterization of GS3 in U.S. rice germplasm. Haplotypes that were characterized in the IRRI 3K RGP panel were screened in U.S. rice germplasm panel based on the minimum SNP set identified to differentiate the eight haplotype groups in IRRI 3K RGP dataset. **(B)** Phenotypic association of haplotype/SNP alleles and grain length across U.S. rice germplasm. Different alphabet letters on each class represent the statistically significant difference $p < 0.001$ within each graph.

average grain length of the entire panel was 6.98 mm, with a standard deviation of 0.47 mm, and a range of 5.37–7.95 mm. The GS3-Hap1 group had an average grain length of 7.13 mm, while the GS3-Hap3 group had an average length of 5.94 mm. GS3-Hap1 and GS3-Hap3 explained ~75% of grain length variation in the U.S. germplasm panel and accurately differentiated long- and medium-grain varieties (**Figures 2A,B**). Thus, we concluded that the GS3 gene underlies the *qGS3.1* QTL and that it is a major gene underlying grain length variation in U.S. germplasm, accurately distinguishing grain market classes. The SNP GS3-SNP1, which represents the known causal C-A point mutation in the GS3 gene, accurately differentiates long- and medium-grain varieties and is an ideal functional SNP for use as a trait marker.

Of the nine *GL7* haplotypes observed in the 3K RGP panel, only two haplotypes (*GL7*-Hap2 and *GL7*-Hap3) were present in U.S. rice germplasm (**Supplementary Table S3**). *GL7*-Hap2 was present in 98% of the U.S. panel, including 100% of long-grain varieties and 86% of medium/short-grain varieties. Only six U.S. varieties (S-201, S-102, M-206, Acadia, Pirogue, and Vista) contained the *GL7*-Hap3 allele, all of which were either California medium-grain varieties and/or historical varieties. No modern Southern U.S. medium-grain variety contained *GL7*-Hap3. The parents of the TJ RIL population contained *GL7*-Hap2 at the *GL7* gene,

as did all the individual RILs, demonstrating that *GL7* is not the gene underlying the QTL *qGL7.1* detected in this study.

Haplotype Characterization of the *qGL7.1* QTL Region Using the HDRA Dataset

Upon concluding that the candidate gene *GL7* did not underlie the QTL *qGL7.1*, haplotype analysis was conducted across the *qGL7.1* QTL region. The QTL was mapped to a region of 474,783 bp (1.84 cM) on chr7, which was too large to conduct haplotype analysis based on re-sequencing data. We therefore utilized the HDRA dataset (McCouch et al., 2016) and identified 763 SNPs across the QTL region based on genotypes from 1,568 samples. To prioritize the identification and characterization of the haplotypes most likely to be present in modern U.S. rice germplasm, we focused on the 34 U.S. varieties in the HDRA dataset that were released after 1980. Single nucleotide polymorphisms having >0.15 missing data across 34 U.S. varieties, samples with >0.2 missing data across the SNPs, SNPs with minor allele frequency <0.02, and monomorphic SNPs were excluded from the analysis. Upon filtering of SNPs, 358 SNPs remained for haplotype characterization across 34 U.S. lines. Based on these 358 SNPs, four haplotype groups were identified at a frequency of ≥ 0.05 . A minimum set of three SNPs

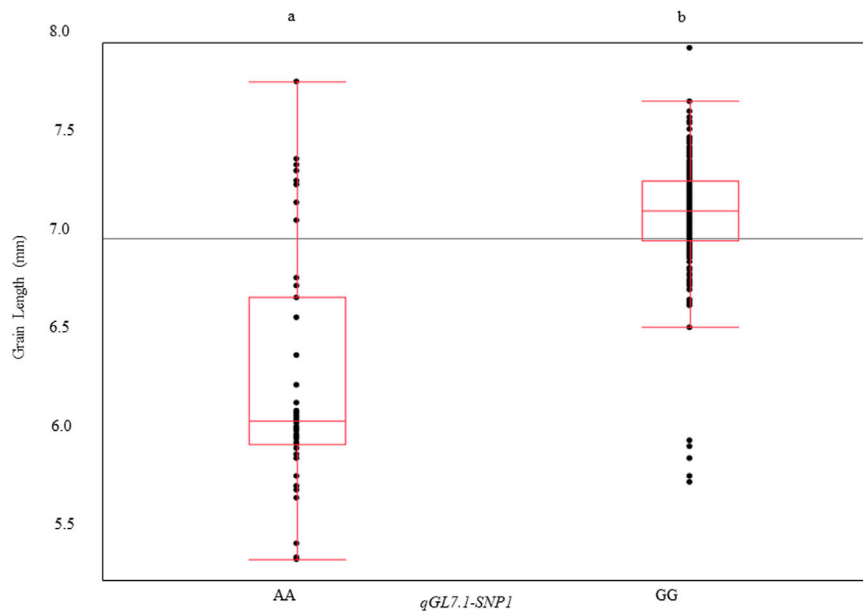


FIGURE 3 | Grain length variation explained by *qGL7.1-SNP1* (LSU-1480) across a U.S. rice germplasm panel. *qGL7.1-SNP1* was identified in the HDRA dataset that was obtained from SNP-Seek online portal within the *qGL7.1* QTL region. QTL *qGL7.1* was detected on Chromosome seven in Trenasse/Jupiter RIL population. *qGL7.1-SNP1* differentiated long-grain haplotypes from medium-grain haplotypes within U.S. lines of HDRA dataset. Allele A is associated with the medium-grain genotypes and allele G is associated with long-grain genotypes. Different alphabet letters on each class represent the statistically significant difference $p < 0.001$ within each graph.

was identified that accurately differentiated the four haplotypes. The four haplotypes were perfectly associated with the previously defined grain classes. *qGL7.1-Hap1* ($n = 13$; 38%) and *qGL7.1-Hap2* ($n = 2$; 6%) were present in samples classified as long grain, where *qGL7.1-Hap3* ($n = 7$; 21%) and *qGL7.1-Hap4* ($n = 2$; 0.1%) were present in samples classified as medium/short grain (Supplementary Figure S1A; Supplementary Table S4).

The three SNPs that differentiated the *qGL7.1* haplotypes were developed into KASP assays and surveyed across the entire U.S. germplasm panel. Three of the four haplotypes (*qGL7.1-Hap1*, *qGL7.1-Hap3*, and *qGL7.1-Hap4*) were detected in modern U.S. germplasm. The most frequent haplotype was *qGL7.1-Hap1* ($n = 271$; 83.9%), found in 266 long-grain and five medium-grain varieties. Haplotype *qGL7.1-Hap3* ($n = 43$; 13.3%) was observed in 14 long-grain, 27 medium-grain, and two short-grain samples, and was the predominant haplotype in medium grains. *qGL7.1-Hap4* ($n = 5$; 1.5%) was observed in two long-grain, one medium-grain, one short-grain samples, and was rare in southern U.S. rice germplasm. Four samples (1.2%) had unclassified haplotypes due to heterozygous SNP calls (Supplementary Figure S1B; Supplementary Table S5). Together, *qGL7.1* haplotypes explained 46% of the variation for grain length across the U.S. germplasm panel. *qGL7.1-Hap1* was associated with long-grain length (7.1 mm) that was significantly longer than either *qGL7.1-Hap3* (6.3 mm) or *qGL7.1-Hap4* (5.6 mm).

To identify a single SNP that could reliably distinguish long- and medium-grain varieties for routine use in southern U.S. breeding programs, the three haplotype SNPs were compared.

The SNP *qGL7.1-SNP1* (25,211,630 bp) was able to differentiate *qGL7.1-Hap1* (predominantly found in long-grain varieties) from *qGL7.1-Hap3* (predominantly found in medium-grain varieties), but it does not differentiate *qGL7.1-Hap3* from *qGL7.1-Hap4*. If used together, *GL7.1-SNP1* and *qGL7.1-SNP3* can differentiate all four haplotypes found in the U.S. germplasm, but a single SNP capable of differentiating long-grain *qGL7.1-Hap1* from medium-grain *qGL7.1-Hap3* would be sufficient for applied breeding programs focusing on the southern United States. When used alone, SNP *qGL7.1-SNP1* explained 43% (Figure 3) of the phenotypic variation for grain length. We recommend that it be used as a diagnostic marker to identify the long-grain allele at *qGL7.1* in applied breeding applications in the United States.

Interaction of GS3 and *qGL7.1* Within a RIL Population, Segregating Populations, and Advanced Breeding Lines

The allelic states of *GS3* and *qGL7.1* were highly correlated across the U.S. germplasm panel, with 96% of long-grain varieties containing the long-grain allele at both loci and 95% of medium-grain varieties containing the medium-grain allele at both loci. To understand the interaction and effect of each locus in a controlled genetic background, the TJ RILs were grouped into four classes based on alleles at *GS3-SNP1* and *qGL7.1-SNP1*, and these classes were compared for grain length. The classes were defined as Class1: *GS3-Long/qGL7.1-Long*; Class2: *GS3-Long/*

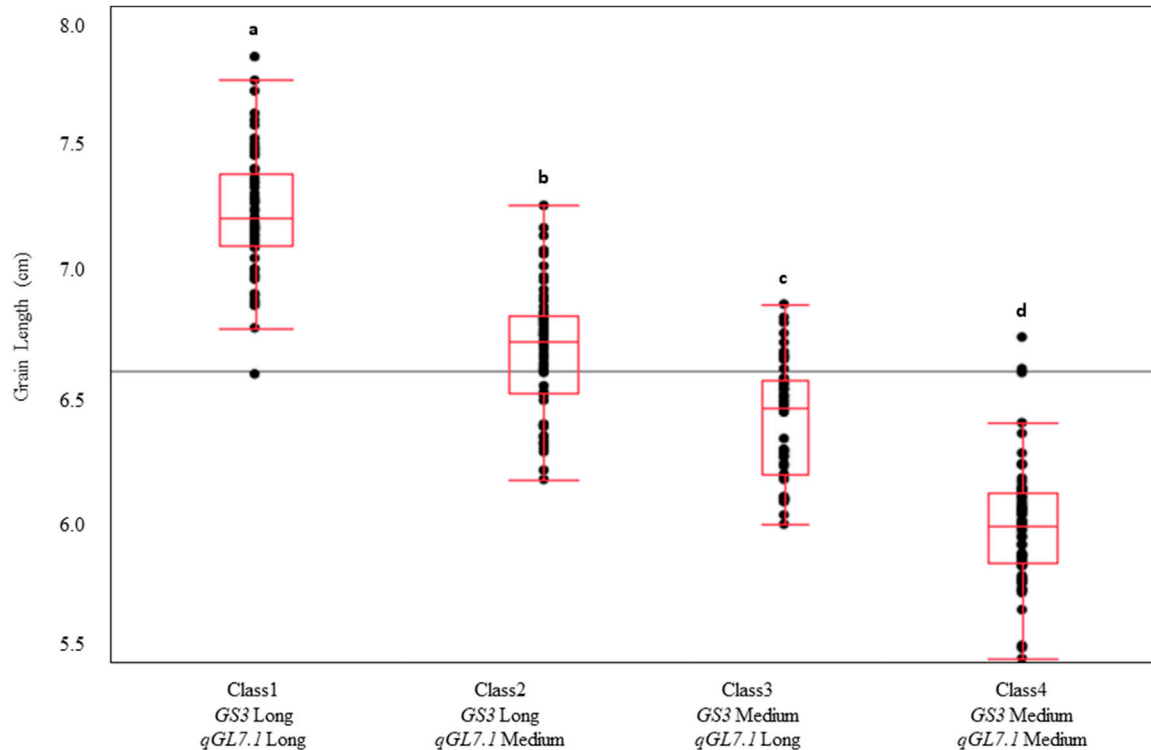


FIGURE 4 | Genotypic classes of Trenasse/Jupiter RILs based on the alleles at *GS3* and *qGL7.1* in the RIL population. Four genotypic classes were identified based on the allele state at *GS3* and *qGL7.1*. Classes denoted by a different letter indicate significant differences between classes ($p < 0.001$).

qGL7.1-Med; Class3: *GS3*-Med/*qGL7.1*-Long; Class4: *GS3*-Med/*qGL7.1*-Med. Significant differences in grain length were observed among classes ($p < 0.001$). Class1 (7.2 mm) had the longest grain length and Class4 had the shortest (6.0 mm). Class2 exhibited a longer grain length than Class3, demonstrating that *GS3* has a larger effect than *qGL7.1* in this population (Figure 4). Selecting long-grain alleles at both loci would help develop a longer grain rice variety using U.S. germplasm.

To further validate the effect of *GS3* and *qGL7.1* in different genetic backgrounds, four F_3 populations (CL272/Mermentau, CL272/Presidio, CRPS/Titan, and Mermentau/Titan) that segregated for both *GS3* and *qGL7.1* loci were examined. The populations were divided into nine genotypic classes based on the allele configurations [homozygous (homo) long (L) allele, heterozygous (het), or homozygous medium (M) allele] at both loci. The genotypic classes were defined as *GS3*-homo L/*qGL7.1*-homo L; *GS3*-homo L/*qGL7.1*-homo M; *GS3*-homo L/*qGL7.1*-het; *GS3*-homo M/*qGL7.1*-homo L; *GS3*-homo M/*qGL7.1*-homo M; *GS3*-homo M/*qGL7.1*-het; *GS3*-het/*qGL7.1*-homo L; *GS3*-het/*qGL7.1*-homo M; and *GS3*-het/*qGL7.1*-het). The average grain length of the genotypic class that was homozygous long at both loci was 7.2 mm, compared to 6.4 mm for the genotypic class that was homozygous medium at both loci. The genotypic class that was heterozygous at both loci was 6.6 mm. The genotypic class containing the homozygous long-grain allele at *GS3* and homozygous medium-grain allele at

qGL7.1 had a longer grain length (7.1 mm) compared to the class containing the homozygous medium-grain allele at *GS3* and homozygous long-grain (6.7 mm) allele at *qGL7.1* (Supplementary Figure S2). Thus, *GS3* again appears to have a larger effect than *qGL7.1* in these additional populations, consistent with observations in the TJ RIL population. Among the lines homozygous for the long-grain *GS3* allele, the allelic state (homozygous long, homozygous medium, heterozygous) at *qGL7.1* did not significantly contribute to grain length. However, among the lines that were homozygous for the medium-grain allele at the *GS3* gene, significantly different grain lengths were observed for the *qGL7.1* allele classes (Supplementary Figure S2).

In addition to the RIL population and segregating populations, we also investigated the effect of these loci on grain length in advanced breeding lines that were developed by crossing long- and medium-grain varieties. Seventy eight breeding lines were genotyped at both loci and phenotyped for grain length across 2 years. These lines were classified into four classes based on allele state at *GS3* and *qGL7.1*. Lines that had the long-grain allele at both loci had an average grain length of 6.9 mm, compared to a grain length of 6.1 mm for lines with the medium-grain allele at both loci. Lines with a long-grain allele at *GS3* and medium-grain allele at *qGL7.1* had an average grain length of 6.5 mm, compared to 6.3 mm for lines that had the medium-grain allele at *GS3* and the long-grain allele at *qGL7.1*. These results confirm that both loci have a significant effect across multiple populations and

genetic backgrounds and that *GS3* has a larger effect on grain length than *qGL7.1*.

Validation and Technical Performance of the Identified Grain Length Trait SNPs

Haplotype characterization and phenotypic association identified a single candidate SNP/locus for use as a trait marker for the major grain size QTLs, *qGS3.1* and *qGL7.1*. For a trait marker to have broad utility in an applied breeding program, it should be informative across the target breeding germplasm and provide high-quality, cost-effective, and reproducible genotype data. *GS3*-SNP1 and *qGL7.1*-SNP1 were validated across the U.S. germplasm panel and shown to be effective at differentiating long- and medium-grain cultivars based on alleles at *GS3* and *qGL7.1*. Thus, these SNPs have potential utility for use in U.S. rice breeding programs. However, for a SNP marker to be effectively deployed in an applied breeding program, it is also essential that the marker provide reproducible results and that it is amenable to high-throughput screening and scoring. To ensure that each of the selected SNPs provides reliable genotyping results for high-throughput molecular breeding activities, the technical performance of the KASP assay was tested. Each assay was tested with three replicates across the U.S. germplasm panel, and a plate of 96 samples from the F_1 population that included all three genotypic classes was used to evaluate SNP clustering in the presence of heterozygous plants. Each assay provided >99% data return, consistent allele calls across the replicates, and exhibited clear and distinct clusters that could be accurately assessed using auto-scoring (Supplementary Figure S3).

Investigation Into the Gene Action and Inheritance of *GS3* and *qGL7.1*

The four F_3 populations (CL272/Mermentau, CL272/Presidio, CRPS/Titan, and Mermentau/Titan) segregating for both *GS3* and *qGL7.1* loci were used to investigate the gene action and inheritance of the two loci. Grain length of the lines that were heterozygous at *GS3* and homozygous for the long allele at *qGL7.1* had grain lengths similar to the lines that were homozygous for the medium-grain allele at *GS3* and for the long-grain allele at *qGL7.1*. Similarly, lines that were heterozygous at *GS3* and homozygous for the medium-grain allele at *qGL7.1* had grain lengths similar to lines that were homozygous for the medium alleles at *GS3* and *qGL7.1*. However, lines that were homozygous for the long alleles at *GS3* and *qGL7.1* were significantly longer than lines that were heterozygous at the *GS3* gene and homozygous for the long allele at *qGL7.1* (Supplementary Figure S2). These results demonstrate that the long-grain allele of *GS3* is recessive, consistent with previous reports.

Compared to *GS3*, the gene action of the *qGL7.1* locus was less clear. Among the $F_2:F_3$ lines that were homozygous for the long-grain *GS3* allele, no significant differences in grain length were observed among the three *qGL7.1* genetic classes (homozygous long, heterozygous, and homozygous medium). However, among $F_2:F_3$ lines homozygous for the medium-grain alleles at *GS3*, the heterozygous class at *qGL7.1* was not significantly different from

the class that had long-grain allele but was significantly longer from the class that had medium-grain allele at *qGL7.1* (Supplementary Figure S2). These results do not provide sufficient evidence to conclude dominance/recessive nature of *qGL7.1*. Further studies using true F_1 plants are required to conclude the gene action of *qGL7.1*.

DISCUSSION

Many genes conferring grain shape have been reported in previous studies. Here, a novel QTL, *qGL7.1*, for grain shape was mapped to a 475 kb region and explained 22% of the phenotypic variation in the TJ RIL population. We explored candidate genes in this region using data available on SNP-Seek platform (Mansueto et al., 2016, 2017). This region harbors 80 genes including 25 expressed proteins, 13 retrotransposons, seven DC1 domain-containing proteins, three C1-like domain-containing proteins, three transposon proteins, two sulfotransferase domain-containing proteins, and two stress responsive A/B barrel domain-containing proteins (Supplementary Table S6). The most promising candidate gene is LOC_Os07g41590, located between 24,931,739–24,933,197 bp. It is a gibberellin (GA) receptor gene, and GAs are reported to be crucial in many developmental processes, including stem elongation, leaf expansion, trichome development, seed germination, pollen maturation, and flowering initiation (Achard and Genschik, 2009). Vaughan et al. (2011) documented the role of GA in determining wheat grain size. They reported that an increase in GA production resulted in up to a 15% increase in grain weight. Therefore, LOC_Os07g41590 may be a grain shape candidate gene underlying the *qGL7.1* QTL. We also explored mutants in the *qGL7.1* region and identified the Giant Embryo (*GE*) gene (Koh et al., 1996; Nagasawa et al., 2013) located at 25,373,984–25,375,652 bp and the *BUI1* (*BENT UPPERMOST INTERNODE1*) gene (Yang et al., 2011) at 25,373,984–25,375,652 bp. *GE* controls cell size in the embryo and cell death in the endosperm. Endogenous *GE* expression is upregulated in mutants, and *GE* overexpression causes a small embryo and enlarged endosperm. Thus, *GE* might be another candidate gene underlying grain length in the *qGL7.1* region. Further investigation is needed to identify other candidate genes in this region that may play a role in grain shape.

The effects of *GS3* and *qGL7.1* on grain length were studied in a wide range of materials including a U.S. germplasm panel ($n = 323$), a RIL (TJ) population ($n = 286$), four $F_2:F_3$ breeding populations, and 78 advanced breeding lines, developed by crossing long and medium-grain parents segregating at both *GS3* and *qGL7.1* loci. In all these materials, both loci have a significant effect on grain length, with *GS3* having a larger effect than *qGL7.1*. Lines carrying the long-grain allele at both loci had the longest grains, and lines carrying medium-grain alleles at both loci exhibited the shortest grains, with other combinations of alleles conferring intermediate grain lengths. These results across a wide range of plant material clearly showed the benefit of using allele-specific markers at both loci to select for grain length in U.S. rice germplasm.

United States long-grain and medium-grain rice varieties differ in their cooking qualities in addition to their grain length. Long-grain rice has an intermediate amylose content of 21–24% and an intermediate gel temperature of 71–74°C as indicated by intermediate alkali spreading value of three–5 (Webb, 1985; Gravois and Webb, 1997; Linscombe, 2017). In contrast, short-to medium-grain rice has a lower amylose content of 15–19% and a low gel temperature of 65–68°C as indicated by alkali spreading value of 6–7 (Webb, 1985; Gravois and Webb, 1997; Linscombe, 2017). Due to these differences in both cooking quality and grain length, it is challenging to cross long- and medium-grain varieties and to select a long- or medium-grain quality line from the segregating population. In this study, we reported high-throughput markers for a major grain length locus *GS3* and a novel QTL *qGL7.1*. Within each grain class (long or medium), these loci do not segregate, and the reported markers may not play any role. However, when crosses are made between long- and medium-grain parents, validated markers *GS3*-SNP1 and *qGL7.1*-SNP1 can play an important role in selecting either long-grain or medium-grain lines. Major waxy alleles are characterized and reported in previous studies (Yang et al., 2019; Shao et al., 2020). Markers are also available for amylose and gel temperature traits. These cooking quality SNP markers along with the validated *GS3*-SNP1 and *qGL7.1*-SNP1 markers for grain length will facilitate breeders' efforts to select the desired grain class and cooking quality when crossing long- and medium-grain varieties.

The results of this study confirmed that *GS3* is a major grain length gene, and *qGL7.1* is a novel locus on chr7 that significantly affects grain shape in the U.S. germplasm. We identified and validated *GS3*-SNP1 and *qGL7.1*-SNP1 markers which accurately identify long- and medium-grain alleles at *GS3* and *qGL7.1*. Alleles at *GS3* and *qGL7.1* do not segregate within the long- or the medium-grain class, but they do segregate when long- and medium-grain varieties are crossed. Selecting for long-grain alleles at both loci will help increase grain length, and selecting for medium-grain alleles at both loci will facilitate the development of shorter grain varieties. Further studies are needed to investigate the gene action of *qGL7.1* and to discover the genes in the locus.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/**Supplementary Material**, further inquiries can be directed to the corresponding authors.

REFERENCES

- Achard, P., and Genschik, P. (2009). Releasing the Brakes of Plant Growth: How GAs Shutdown DELLA Proteins. *J. Exp. Bot.* 60, 1085–1092. doi:10.1093/JXB/ERN301
- Addison, C. K., Angira, B., Cerioli, T., Groth, D. E., Richards, J. K., Linscombe, S. D., et al. (2021). Identification and Mapping of a Novel Resistance Gene to the

AUTHOR CONTRIBUTIONS

BA mapped QTLs for grain shape traits using KASP markers, analyzed the data, executed haplotype characterization in IRRI 3K RGP dataset and the U.S. germplasm, identified and validated markers for use in applied breeding programs, and lead manuscript writing. AF assisted in designing the project, troubleshooting research problems, field planting and design, and contributed in writing the manuscript. TC developed advanced breeding lines crossing long- and medium-grain varieties. These advanced breeding lines were used to validate the effect of the novel locus and to validate the markers.

FUNDING

This research study was supported by Louisiana Rice Research Board (grant GR- 59400008264, PI: AF and grant GR-00006303 and GR-00008230, PI: BA), Horizon Ag. (grant GR-00004555, PI: AF), National Science Foundation (grant number (FAIN) 1826836, PI: A. Pereira), Hatch project LAB94468 (PI: BA), and Hatch project LAB94342 (PI: AF).

ACKNOWLEDGMENTS

The authors thank Valerie Dartez for contribution in manuscript preparation and proofreading.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2022.923078/full#supplementary-material>

Supplementary Figure 1 | (A) Haplotype characterization within QTL *qGL7.1* region using HDRA SNP-Seek genotype dataset to study the diversity in the QTL region. The haplotype characterization was conducted using only US varieties that were released after 1980. **(B)** Haplotype characterization of the QTL *qGL7.1* across US germplasm using minimum SNP set developed from HDRA data within the QTL region.

Supplementary Figure 2 | Four segregating F3 populations - CL272/Mermentau, CL272/Presidio, CRPS/Titan, and Mermentau/Titan, that segregated for both *GS3* and *qGL7.1* genes were used to study the effect of the genes on grain length. Each population had about 70 panicles. A total of 276 panicles were genotyped for *GS3* and *qGL7.1*, and all the seeds of each panicle were phenotyped for grain length. First letter in class name is allele at *GS3* gene, and the second letter followed by forward slash is allele state at *qGL7.1* gene. L stands for long-grain allele, M stands for medium-grain allele, and H stands for heterozygote state at the gene. **(A)** Average grain length of the defined classes. Different alphabet letters on each class represent the statistically significant difference $p < 0.001$ within each graph.

Supplementary Figure 3 | Technical validation of *qGS3*-SNP1 (left) and *qGL7.1* SNP1 (right) in breeding populations segregating at *GS3* and *qGL7.1*. Both clusters for each SNP are distinct and suitable for automated computer scoring.

- Rice Pathogen, *Cercospora janseana*. *Theor. Appl. Genet.* 134, 2221–2234. doi:10.1007/S00122-021-03821-2
- Addison, C. K., Angira, B., Kongchum, M., Harrell, D. L., Baisakh, N., Linscombe, S. D., et al. (2020). Characterization of Haplotype Diversity in the *BADH2* Aroma Gene and Development of a KASP SNP Assay for Predicting Aroma in U.S. Rice. *Rice* 13, 1–9. doi:10.1186/S12284-020-00410-7/TABLES/2

- Alexandrov, N., Tai, S., Wang, W., Mansueto, L., Palis, K., Fuentes, R. R., et al. (2014). SNP-Seek Database of SNPs Derived from 3000 Rice Genomes. *Nucleic Acids Res.* 43, D1023–D1027. doi:10.1093/nar/gku1039
- Angira, B., Addison, C. K., Cerioli, T., Rebong, D. B., Wang, D. R., Pumphin, N., et al. (2019). Haplotype Characterization of the *Sd1* Semidwarf Gene in United States Rice. *Plant Genome* 12, 190010. doi:10.3835/plantgenome2019.02.0010
- Bai, X., Luo, L., Yan, W., Kovi, M. R., Zhan, W., and Xing, Y. (2010). Genetic Dissection of Rice Grain Shape Using a Recombinant Inbred Line Population Derived from Two Contrasting Parents and Fine Mapping a Pleiotropic Quantitative Trait Locus *qGL7*. *BMC Genet.* 11, 1. doi:10.1186/1471-2156-11-16
- Botella, J. R. (2012). Can Heterotrimeric G Proteins Help to Feed the World? *Trends Plant Sci.* 17, 563–568. doi:10.1016/J.TPLANTS.2012.06.002
- Brinton, J., Ramirez-Gonzalez, R. H., Ramirez-Gonzalez, R. H., Simmonds, J., Wingen, L., Orford, S., et al. (2020). A Haplotype-Led Approach to Increase the Precision of Wheat Breeding. *Commun. Biol.* 3 (3), 1–11. doi:10.1038/s42003-020-01413-2
- Bruce, R. W., Torkamaneh, D., Grainger, C. M., Belzile, F., Eskandari, M., and Rajcan, I. (2020). Haplotype Diversity Underlying Quantitative Traits in Canadian Soybean Breeding Germplasm. *Theor. Appl. Genet.* 133, 1967–1976. doi:10.1007/S00122-020-03569-1
- Caicedo, A. L., Williamson, S. H., Hernandez, R. D., Boyko, A., Fledel-Alon, A., York, T. L., et al. (2007). Genome-wide Patterns of Nucleotide Polymorphism in Domesticated Rice. *PLoS Genet.* 3, e163–1756. doi:10.1371/journal.pgen.0030163
- Chen, W., Chen, L., Zhang, X., Yang, N., Guo, J., Wang, M., et al. (2022). Convergent Selection of a WD40 Protein that Enhances Grain Yield in Maize and Rice. *Science* 375, eabg7985. doi:10.1126/science.abg7985
- Childs, N. W. (2014). “Chapter 1: Production and Utilization of Rice,” in *Rice Chemistry and Technology*. 3rd ed. (American Association of Cereal Chemist), 1–23. doi:10.1094/1891127349.001
- Edwards, J. D., Baldo, A. M., and Mueller, L. A. (2016). Ricebase: a Breeding and Genetics Platform for Rice, Integrating Individual Molecular Markers, Pedigrees and Whole-Genome-Based Data. *Database* 2016, baw107. doi:10.1093/database/baw107
- Fan, C., Xing, Y., Mao, H., Lu, T., Han, B., Xu, C., et al. (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. doi:10.1007/s00122-006-0218-1
- Fan, C., Yu, S., Wang, C., and Xing, Y. (2009). A Causal C-A Mutation in the Second Exon of GS3 Highly Associated with Rice Grain Length and Validated as a Functional Marker. *Theor. Appl. Genet.* 118, 465–472. doi:10.1007/s00122-008-0913-1
- Gao, F.-y., Zeng, L.-h., Qiu, L., Lu, X.-j., Ren, J.-s., Wu, X.-t., Su, X.-w., et al. (2016). QTL Mapping of Grain Appearance Quality Traits and Grain Weight Using a Recombinant Inbred Population in Rice (*Oryza Sativa* L.). *J. Integr. Agric.* 15, 1693–1702. doi:10.1016/S2095-3119(15)61259-X
- Garris, A. J., Tai, T. H., Coburn, J., Kresovich, S., and McCouch, S. (2005). Genetic Structure and Diversity in *Oryza Sativa* L. *Genetics* 169, 1631–1638. doi:10.1534/genetics.104.035642
- Glazmann, J. C. (1987). Isozymes and Classification of Asian Rice Varieties. *Theor. Appl. Genet.* 74, 21–30. doi:10.1007/BF00290078
- Gravois, K. A., and Webb, B. D. (1997). Inheritance of Long Grain Rice Amylograph Viscosity Characteristics. *Euphytica* 97, 25–29. doi:10.1023/A:1003082930147
- Hu, Z., Lu, S.-J., Wang, M.-J., He, H., Sun, L., Wang, H., et al. (2018). A Novel QTL *qTGW3* Encodes the GSK3/SHAGGY-Like Kinase OsGSK5/OsSK41 that Interacts with OsARF4 to Negatively Regulate Grain Size and Weight in Rice. *Mol. Plant* 11, 736–749. doi:10.1016/j.molp.2018.03.005
- Ishimaru, K., Hirotsu, N., Madoka, Y., Murakami, N., Hara, N., Onodera, H., et al. (2013). Loss of Function of the IAA-Glucose Hydrolase Gene *TGW6* Enhances Rice Grain Weight and Increases Yield. *Nat. Genet.* 45, 707–711. doi:10.1038/ng.2612
- Juliano, B. O., and Villareal, C. (1993). *Grain Quality Evaluation of World Rices*. Manila, Philippines: International Rice Research Institute.
- Koh, H.-J., Heu, M.-H., and McCouch, S. R. (1996). Molecular Mapping of the *Ge^S* Gene Controlling the Super-giant Embryo Character in Rice (*Oryza Sativa* L.). *Theor. Appl. Genet.* 93–93, 257–261. doi:10.1007/BF00225754
- Kumpatla, S., Buyyarapu, R., Mammadov, J., and Abdurakhmonov, I. (2012). Genomics-Assisted Plant Breeding in the 21st Century: Technological Advances and Progress. *Sci. Acad. Pap. Collect.* Available at: <https://scienceweb.uz/publication/276> (Accessed May 5, 2022). doi:10.5772/37458
- Li, J., Thomson, M., and McCouch, S. R. (2004). Fine Mapping of a Grain-Weight Quantitative Trait Locus in the Pericentromeric Region of Rice Chromosome 3. *Genetics* 168, 2187–2195. doi:10.1534/GENETICS.104.034165
- Li, Y., Fan, C., Xing, Y., Jiang, Y., Luo, L., Sun, L., et al. (2011). Natural Variation in *GS5* Plays an Important Role in Regulating Grain Size and Yield in Rice. *Nat. Genet.* 43, 1266–1269. doi:10.1038/ng.977
- Li, Z., Fu, B. Y., Gao, Y. M., Wang, W. S., Xu, J. L., Zhang, F., et al. (2014). The 3,000 Rice Genomes Project: New Opportunities and Challenges for Future Rice Research. *GigaSci* 3, 3–8. doi:10.1186/2047-217X-3-7
- Linscombe, S. D., Sha, X., Bond, J. A., Bearb, K., Rush, M. C., Chu, Q. R., et al. (2006). Registration of ‘Trenasse’ Rice. *Crop Sci.* 46, 2318–2319. doi:10.2135/CROPSCI2006.03.0208
- Linscombe, S. (2017). *Many Types of Rice to Choose*. LSU AgCenter. Available at: <https://www.lsuagcenter.com/profiles/vdartz/articles/page1498155996996> (Accessed June 3, 2022).
- LSU Agcenter (2014). *Louisiana Rice Production Handbook*. Available at: https://www.lsuagcenter.com/portals/communications/publications/publications_catalog/crops_livestock/rice/rice-production-handbook1 (Accessed May 20, 2022).
- Lu, H., Redus, M. A., Coburn, J. R., Rutger, J. N., McCouch, S. R., and Tai, T. H. (2005). Population Structure and Breeding Patterns of 145 U.S. Rice Cultivars Based on SSR Marker Analysis. *Crop Sci.* 45, 66–76. doi:10.2135/cropsci2005.0066
- Mackill, D. J., and Lei, X. (1997). Genetic Variation for Traits Related to Temperate Adaptation of Rice Cultivars. *Crop Sci.* 37, 1340–1346. doi:10.2135/cropsci1997.0011183X003700040051x
- Maldonado, C., Mora, F., Scapim, C. A., and Coan, M. (2019). Genome-Wide Haplotype-Based Association Analysis of Key Traits of Plant Lodging and Architecture of Maize Identifies Major Determinants for Leaf Angle: *hapLA4*. *PLoS One* 14, e0212925. doi:10.1371/JOURNAL.PONE.0212925
- Mansueto, L., Fuentes, R. R., Chebotarov, D., Borja, F. N., Detras, J., Abriol-santos, J. M., et al. (2016). SNP-Seek II: A Resource for Allele Mining and Analysis of Big Genomic Data in *Oryza Sativa*. *Curr. Plant Biol.* 7–8, 16–25. doi:10.1016/j.cpb.2016.12.003
- Mansueto, L., Fuentes, R. R., Borja, F. N., Detras, J., Abriol-Santos, J. M., Chebotarov, D., et al. (2017). Rice SNP-Seek Database Update: New SNPs, Indels, and Queries. *Nucleic Acids Res.* 45, D1075–D1081. doi:10.1093/nar/gkw1135
- Mao, H., Sun, S., Yao, J., Wang, C., Yu, S., Xu, C., et al. (2010). Linking Differential Domain Functions of the GS3 Protein to Natural Variation of Grain Size in Rice. *Proc. Natl. Acad. Sci. U. S. A.* 107, 19579–19584. doi:10.1073/pnas.1014419107
- Marsh, J. I., Hu, H., Peterreit, J., Bayer, P. E., Valliyodan, B., Batley, J., et al. (2022). Haplotype Mapping Uncovers Unexplored Variation in Wild and Domesticated Soybean at the Major Protein Locus *cqProt-003*. *Theor. Appl. Genet.* 135, 1443–1455. doi:10.1007/S00122-022-04045-8/FIGURES/5
- McCouch, S. R., Wright, M. H., Tung, C.-W., Maron, L. G., McNally, K. L., Fitzgerald, M., et al. (2016). Open Access Resources for Genome-Wide Association Mapping in Rice. *Nat. Commun.* 7, 10532. doi:10.1038/ncomms10532
- Meng, L., Li, H., Zhang, L., and Wang, J. (2015). QTL IciMapping: Integrated Software for Genetic Linkage Map Construction and Quantitative Trait Locus Mapping in Biparental Populations. *Crop J.* 3, 269–283. doi:10.1016/J.CJ.2015.01.001
- Nagasawa, N., Hibara, K.-I., Heppard, E. P., Vander Velden, K. A., Luck, S., Beatty, M., et al. (2013). *GIANT EMBRYO* Encodes CYP78A13, Required for Proper Size Balance Between Embryo and Endosperm in Rice. *Plant J.* 75, 592–605. doi:10.1111/TPJ.12223
- Oka, H. (1988). *Origin of Cultivated Rice*. Tokyo: Elsevier Science/Japan Scientific Societies Press.
- Qi, P., Lin, Y.-S., Song, X.-J., Shen, J.-B., Huang, W., Shan, J.-X., et al. (2012). The Novel Quantitative Trait Locus *GL3.1* Controls Rice Grain Size and Yield by Regulating Cyclin-T1;3. *Cell Res.* 22, 1666–1680. doi:10.1038/CR.2012.151

- Second, G. (1985). Evolutionary Relationships in the *Sativa* Group of *Oryza* Based on Isozyme Data. *Genet. Sel. Evol.* 17, 89. doi:10.1186/1297-9686-17-1-89
- Semagn, K., Babu, R., Hearne, S., and Olsen, M. (2014). Single Nucleotide Polymorphism Genotyping Using Competitive Allele Specific PCR (KASP): Overview of the Technology and its Application in Crop Improvement. *Mol. Breed.* 33, 1–14. doi:10.1007/s11032-013-9917-x
- Sha, X., Linscombe, S. D., Groth, D. E., Bond, J. A., White, L. M., Chu, Q. R., et al. (2006). Registration of ‘Jupiter’ Rice. *Crop Sci.* 46, 1811–1812. doi:10.2135/CROPSCI2005.10-0393
- Shao, G., Tang, S., Luo, J., Jiao, G., Wei, X., Tang, A., et al. (2010). Mapping of *qGL7-2*, a Grain Length QTL on Chromosome 7 of Rice. *J. Genet. Genomics* 37, 523–531. doi:10.1016/S1673-8527(09)60072-2
- Shao, Y., Peng, Y., Mao, B., Lv, Q., Yuan, D., Liu, X., et al. (2020). Allelic Variations of the *Wx* Locus in Cultivated Rice and Their Use in the Development of Hybrid Rice in China. *PLoS One* 15, e0232279. doi:10.1371/JOURNAL.PONE.0232279
- Shomura, A., Izawa, T., Ebana, K., Ebitani, T., Kanegae, H., Konishi, S., et al. (2008). Deletion in a Gene Associated with Grain Size Increased Yields During Rice Domestication. *Nat. Genet.* 40, 1023–1028. doi:10.1038/ng.169
- Si, L., Chen, J., Huang, X., Gong, H., Luo, J., Hou, Q., et al. (2016). *OsSPL13* Controls Grain Size in Cultivated Rice. *Nat. Genet.* 48, 447–456. doi:10.1038/NG.3518
- Song, X.-J., Huang, W., Shi, M., Zhu, M.-Z., and Lin, H.-X. (2007). A QTL for Rice Grain Width and Weight Encodes a Previously Unknown RING-Type E3 Ubiquitin Ligase. *Nat. Genet.* 39, 623–630. doi:10.1038/ng2014
- Song, X. J., Kuroha, T., Ayano, M., Furuta, T., Nagai, K., Komeda, N., et al. (2015). Rare Allele of a Previously Unidentified Histone H4 Acetyltransferase Enhances Grain Weight, Yield, and Plant Biomass in Rice. *Proc. Natl. Acad. Sci. U.S.A.* 112, 76–81. doi:10.1073/PNAS.1421127112
- Sun, C., Wang, X., Yoshimura, A., and Doi, K. (2002). Genetic Differentiation for Nuclear, Mitochondrial and Chloroplast Genomes in Common Wild Rice (*Oryza Rufipogon* Griff.) and Cultivated Rice (*Oryza Sativa* L.). *Theor. Appl. Genet.* 104, 1335–1345. doi:10.1007/s00122-002-0878-4
- Sun, S., Wang, L., Mao, H., Shao, L., Li, X., Xiao, J., et al. (2018). A G-Protein Pathway Determines Grain Size in Rice. *Nat. Commun.* 9 (9), 1–11. doi:10.1038/s41467-018-03141-y
- Takano-Kai, N., Jiang, H., Kubo, T., Sweeney, M., Matsumoto, T., Kanamori, H., et al. (2009). Evolutionary History of *GS3*, a Gene Conferring Grain Length in Rice. *Genetics* 182, 1323–1334. doi:10.1534/genetics.109.103002
- Takano-Kai, N., Jiang, H., Powell, A., McCouch, S., Takamura, I., Furuya, N., et al. (2013). Multiple and Independent Origins of Short Seeded Alleles of *GS3* in Rice. *Breed. Sci.* 63, 77–85. doi:10.1270/JSBBS.63.77
- USDA; National Agricultural Statistics Service (2021). Available at: https://www.nass.usda.gov/Publications/Todays_Reports/reports/acrg0621.pdf (Accessed May 20, 2022).
- Vaughan, S., Wanchoo, A., Patil, A., Huttly, A., Phillips, A., and Hedden, P. (2011). The Role of Gibberellin in Determining Wheat Grain Size and Quality. *Asp. Appl. Biol.* 110. Available at: <https://www.cabdirect.org/cabdirect/abstract/20133176502> (Accessed March 6, 2022).
- Wang, Z. Y., and Tanksley, S. D. (1989). Restriction Fragment Length Polymorphism in *Oryza Sativa* L. *Genome* 32, 1113–1118. doi:10.1139/g89-563
- Wang, E., Wang, J., Zhu, X., Hao, W., Wang, L., Li, Q., et al. (2008). Control of Rice Grain-Filling and Yield by a Gene with a Potential Signature of Domestication. *Nat. Genet.* 40, 1370–1374. doi:10.1038/ng.220
- Wang, S., Wu, K., Yuan, Q., Liu, X., Liu, Z., Lin, X., et al. (2012). Control of Grain Size, Shape and Quality by *OsSPL16* in Rice. *Nat. Genet.* 44, 950–954. doi:10.1038/ng.2327
- Wang, S., Li, S., Liu, Q., Wu, K., Zhang, J., Wang, S., et al. (2015a). The *OsSPL16-GW7* Regulatory Module Determines Grain Shape and Simultaneously Improves Rice Yield and Grain Quality. *Nat. Genet.* 47, 949–954. doi:10.1038/ng.3352
- Wang, Y., Xiong, G., Hu, J., Jiang, L., Yu, H., Xu, J., et al. (2015b). Copy Number Variation at the *GL7* Locus Contributes to Grain Size Diversity in Rice. *Nat. Genet.* 47, 944–948. doi:10.1038/ng.3346
- Webb, B. D. (1985). “Criteria of Rice Quality in the United States,” in *Rice Chemistry and Technology*. Editor B. O. Juliano (St. Paul, Minnesota, USA: American Association of Cereal Chemists AACC), 403–442.
- Weng, J., Gu, S., Wan, X., Gao, H., Guo, T., Su, N., et al. (2008). Isolation and Initial Characterization of *GW5*, a Major QTL Associated with Rice Grain Width and Weight. *Cell Res.* 18, 1199–1209. doi:10.1038/cr.2008.307
- Xia, D., Zhou, H., Liu, R., Dan, W., Li, P., Wu, B., et al. (2018). *GL3.3*, a Novel QTL Encoding a GSK3/SHAGGY-Like Kinase, Epistatically Interacts with *GS3* to Produce Extra-Long Grains in Rice. *Mol. Plant* 11, 754–756. doi:10.1016/j.molp.2018.03.006
- Yang, W., Ren, S., Zhang, X., Gao, M., Ye, S., Qi, Y., et al. (2011). *BENT UPPERMOST INTERNODE1* Encodes the Class II Formin FH5 Crucial for Actin Organization and Rice Development. *Plant Cell* 23, 661–680. doi:10.1105/TPC.110.081802
- Yang, G., Chen, S., Chen, L., Gao, W., Huang, Y., Huang, C., et al. (2019). Development and Utilization of Functional KASP Markers to Improve Rice Eating and Cooking Quality Through MAS Breeding. *Euphytica* 215, 1–12. doi:10.1007/S10681-019-2392-7
- Zhang, X., Wang, J., Huang, J., Lan, H., Wang, C., Yin, C., et al. (2012). Rare Allele of *OsPPKL1* Associated with Grain Length Causes Extra-Large Grain and a Significant Yield Increase in Rice. *Proc. Natl. Acad. Sci. U.S.A.* 109, 21534–21539. doi:10.1073/PNAS.1219776110
- Zhou, Y., Miao, J., Gu, H., Peng, X., Leburu, M., Yuan, F., et al. (2015). Natural Variations in *SLG7* Regulate Grain Shape in Rice. *Genetics* 201, 1591–1599. doi:10.1534/genetics.115.181115

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher’s Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Angira, Cerioli and Famoso. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.