

RESEARCH ARTICLE

A novel rice grain size gene *OsSNB* was identified by genome-wide association study in natural population

Xiaosong Ma¹, Fangjun Feng¹, Yu Zhang¹, Ibrahim Eid Elesawi², Kai Xu¹, Tianfei Li¹, Hanwei Mei¹, Hongyan Liu¹, Ningning Gao¹, Chunli Chen², Lijun Luo^{1*}, Shunwu Yu^{1*}

1 Shanghai Agrobiological Gene Center, Shanghai, China, **2** College of Life Science and Technology, Huazhong Agricultural University, Wuhan, China

* lijun@sagc.org.cn (LL); ysw@sagc.org.cn (SW)



OPEN ACCESS

Citation: Ma X, Feng F, Zhang Y, Elesawi IE, Xu K, Li T, et al. (2019) A novel rice grain size gene *OsSNB* was identified by genome-wide association study in natural population. *PLoS Genet* 15(5): e1008191. <https://doi.org/10.1371/journal.pgen.1008191>

Editor: Zhiwu Zhang, Washington State University, UNITED STATES

Received: January 23, 2019

Accepted: May 13, 2019

Published: May 31, 2019

Copyright: © 2019 Ma et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The list of 270 rice accessions is available from S5 Table in Wu et al. (2015) [24]. All other relevant data are available from the manuscript and its Supporting Information files.

Funding: This work was sponsored by Natural Science Foundation of Shanghai (<http://stcsm.sh.gov.cn/>, 18ZR1433500 to XSM, 17ZR1425400 to SWY and 16ZR1431200 to KX), Shared Platform of Crop Germplasm Resources in Shanghai (<http://stcsm.sh.gov.cn/>, 18DZ2293700 to LJJL), Shanghai

Abstract

Increasing agricultural productivity is one of the most important goals of plant science research and imperative to meet the needs of a rapidly growing population. Rice (*Oryza sativa* L.) is one of the most important staple crops worldwide. Grain size is both a major determinant of grain yield in rice and a target trait for domestication and artificial breeding. Here, a genome-wide association study of grain length and grain width was performed using 996,722 SNP markers in 270 rice accessions. Five and four quantitative trait loci were identified for grain length and grain width, respectively. In particular, the novel grain size gene *OsSNB* was identified from *qGW7*, and further results showed that *OsSNB* negatively regulated grain size. Most notably, knockout mutant plants by CRISPR/Cas9 technology showed increased grain length, width, and weight, while overexpression of *OsSNB* yielded the opposite. Sequencing of this gene from the promoter to the 3'-untranslated region in 168 rice accessions from a wide geographic range identified eight haplotypes. Furthermore, Hap 3 has the highest grain width discovered in *japonica* subspecies. Compared to other haplotypes, Hap 3 has a 225 bp insertion in the promoter. Based on the difference between Hap 3 and other haplotypes, *OsSNB_Indel2* was designed as a functional marker for the improvement of rice grain width. This could be directly used to assist selection toward an improvement of grain width. These findings suggest *OsSNB* as useful for further improvements in yield characteristics in most cultivars.

Author summary

Grain weight, including grain length and grain width, is a complex trait, and hundreds of quantitative trait loci (QTLs) were detected in different genetic rice populations. However, only about 10 genes have been isolated and characterized until now. Nine QTLs for grain size were identified by genome-wide association study in a natural rice population. The novel grain size gene *OsSNB* was identified from *qGW7* based on the difference of expression levels between two different varieties with significantly different grain width. *OsSNB*

Agriculture Applied Technology Development Program (<http://nyncw.sh.gov.cn/nw/>, Grant No. 2017-02-08-00-08-F00071 to L.J.L.). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

is an AP2 transcription factor that is negatively regulated grain size. However, OsSNB was found to regulate the transition from the spikelet meristem to the floral meristem and the floral organ development in previous study. Compared to other haplotypes, Hap 3 has a 225 bp insertion in the promoter. Based on the difference between Hap 3 and other haplotypes, OsSNB_Indel2 was designed as a functional marker for the improvement of rice grain width. This can be directly used to assist selection for grain width improvement.

Introduction

Rice (*Oryza sativa* L.) is one of the most important staple food crops in the world. Grain yield in rice is determined by three components: number of panicles, number of grains per panicle, and grain weight, all of which are complex quantitative traits [1]. Among these traits, the most important trait is grain weight, which is measured as a 1,000-grain weight. The grain weight is largely determined by grain size, which, in turn, includes grain length, grain width, grain thickness, and the degree of filling [1, 2]. These four parameters are positively correlated with grain weight [2].

Over the past 30 years, fueled by the development of DNA markers and genomic sequencing technology, dramatic progress has been achieved in both the mapping and cloning of genes that control grain shape and grain weight in rice. To date, dozens of genes located in main effective quantitative trait loci that control grain shape and grain weight have been isolated by the map-based cloning strategy as well as functionally characterized. Prominent examples are: GRAIN SIZE 3 (GS3) [3, 4], *GL3.1/OsPPKL1* [5–7], *GW5/qSW5* [8, 9], *GS5* [10], *GW2* [11], *GW8/OsSPL16* [12], THOUSAND-GRAIN WEIGHT 6 (*TGW6*) [13], *GW6a* [14], *GL7/GW7* [15, 16], and GRAIN SIZE ON CHROMOSOME 2 (*GS2*) [17].

Among these QTLs/genes, *GS3* is a major QTL for both grain length and weight, and functions as a negative regulator for grain size [3, 4]. *TGW6* encodes a novel protein with indole-3-acetic acid (IAA)-glucose hydrolase activity that negatively regulates grain weight by limiting the number of cells [13]. *GW5/qSW5* encodes a calmodulin binding protein and acts as a negative regulator for both grain width and grain weight depended on the brassinosteroid (BR) signaling pathway [8, 9, 18]. *GW2*, encodes a RING-type E3 ubiquitin ligase, which also negatively regulates grain width, weight, and yield through negatively regulating cell division in the shell [11].

In addition to these genes that negatively regulate grain size, several genes that positively regulate grain size have also been identified. For example, *GL3.1*, encodes a protein phosphatase kelch (PPKL) family Ser/Thr phosphatase, that acts as a positive regulator for grain length [5, 6]. The major QTL *GS5*, which is a putative serine carboxypeptidase [10], and *GW8/OsSPL16*, which is a SBP-domain transcription factor [12], function as positive regulators of grain size, affecting both grain width and weight. *GW6a*, a major QTL for grain weight, encodes a new type of GNAT-like protein that harbors intrinsic histone acetyltransferase activity (*OsglHAT1*). Elevated expression of this gene enhances grain weight and yield by enlarging spikelet hulls *via* both increasing cell number and accelerating grain filling [14]. *GL7/GW7* has been identified as a major QTL for grain length and width, containing a tandem duplication of a 17.1-kb segment at the *GL7* locus. This leads to up-regulation of *GL7*, thus resulting in an increase in grain length [15]. The further dominant QTL *GS2* has been identified, which encodes Growth-Regulating Factor 4 (*OsGRF4*). Increase of *GS2* expression resulted in larger cells and increased numbers of cells, which thus enhances both grain weight and yield [17].

Although the above mentioned genes, that control rice grain size and weight, have been cloned and characterized, several hundred QTLs remain that have been detected by primary mapping to control rice grain shape and grain weight; however, these have not been cloned to date [2]. Isolating candidate genes *via* map-based cloning strategy would be very time consuming, because it needs a long time to develop near-isogenic lines that are required for fine mapping. With the rapid development of sequencing technologies and the continuing decrease in related costs, genome-wide association study (GWAS) has become a powerful tool for the detection of natural variations that account for complex traits [19]. Compared to map-based cloning, GWAS fully utilizes ancient recombination events to identify the genetic loci that underlie traits at a relatively high resolution. Moreover, GWAS requires less research time. GWAS has been successfully applied to the genetic dissection of complex traits in plants, e.g., *Arabidopsis thaliana* [20–23], *Oryza sativa* [24–27], and *Zea mays* [28–31]. Recently, *GLW7*, a major QTL for both grain length and weight, that encodes the plant-specific transcription factor *OsSPL13* was cloned based on a GWAS approach [32]. This example indicates that candidate genes that regulate rice grain size can be quickly discovered *via* the GWAS method.

In this study, GWAS of rice grain width and length was performed in 270 rice germplasms, and several new QTLs and genes associated with grain size in rice were detected. Furthermore, the novel grain width gene *OsSNB* was identified from *qGW7*. The grain width and grain weight of knock out mutant plants were significantly increased and the grain width and grain weight significantly decreased in over-expression transgenic plants. These findings enhance the understanding of the genetic mechanism underlying grain size in rice. These newly reported QTLs/genes of grain size may be useful in future molecular breeding programs aimed at improving grain yield in rice.

Results

Descriptive statistics of grain length and width

The natural population (collection 1) showed remarkable segregation for all studied traits (Table 1). Grain length ranged from 4.72 to 10.90 mm with a mean of 8.06 mm in 2011 and from 6.06 to 10.56 mm with a mean of 8.05 mm in 2012. Grain width ranged from 2.45 to 4.23 mm with a mean of 3.28 mm in 2011, and from 2.49 to 4.45 mm with a mean of 3.32 mm in 2012.

Genome-wide association study (GWAS) for grain length and grain width

Three associated loci (sf0316957236, sf0812791004, and sf1018115947) were detected for grain length at a significance level of $-\log_{10}(P) \geq 6.0$ in 2011, which explained 31.81%, 19.88%, and 17.58% of the phenotypic variation, respectively. Furthermore, three associated loci (sf0316957236, sf0608989271, and sf1201579378) were detected for grain length at a significance level of $-\log_{10}(P) \geq 6.0$ in 2012, which explained 37.49%, 11.91%, and 11.86% of the phenotypic variation, respectively (Fig 1, Table 2). The association sf0316957236 on chromosome 3 (QTL interval: 16.67–17.00 Mb with r^2 of LD > 0.4) was located downstream from the *GS3*, which was located at 16.67 Mb and has a length of about 230 kb. Thus, *GS3* belongs to this QTL interval.

Three associated loci (sf0313650794, sf0505372028, and sf0707318737) were detected for gain width in 2011, which explained 0.17%, 45.20%, and 15.23% of the phenotypic variation, respectively. Furthermore, three associated loci (sf0315013153, sf0505372028, and sf0707314881) were detected in 2012, which explained 46.03%, 44.79%, and 16.86% of the phenotypic variation, respectively (Fig 2, Table 2). The r^2 of LD = 0.02 between the sf0313650794 and the sf0315013153, thus these are two different association intervals. One associated SNP

Table 1. Descriptive statistics of grain length (GL) and grain width (GW) in collection 1.

Traits	Year	valid data	Mean±SD	Range	skewness	kurtosis
GL	2011	235	8.06±0.97	4.72–10.90	0.07	0.58
(mm)	2012	235	8.05±0.85	6.06–10.56	0.66	-0.24
GW	2011	235	3.28±0.41	2.45–4.23	0.06	-0.68
(mm)	2012	235	3.32±0.40	2.49–4.45	0.16	-0.62

Mean±SD, Mean ± Standard deviation

<https://doi.org/10.1371/journal.pgen.1008191.t001>

(sf0505364734 with $P = 2.32E-07$ or $1.37E-07$) was located in the promoter of *qSW5/GW5* and was detected for grain width in both 2011 and 2012. The r^2 of LD = 0.05 between the sf0315013153 (detected for grain width) and the sf0316957236 (detected for grain length); thus, these are two different association intervals. Additionally, the false discovery rate (FDR) values corresponding to the above nine QTL loci were below 0.05 (Table 2, Figs 1 and 2).

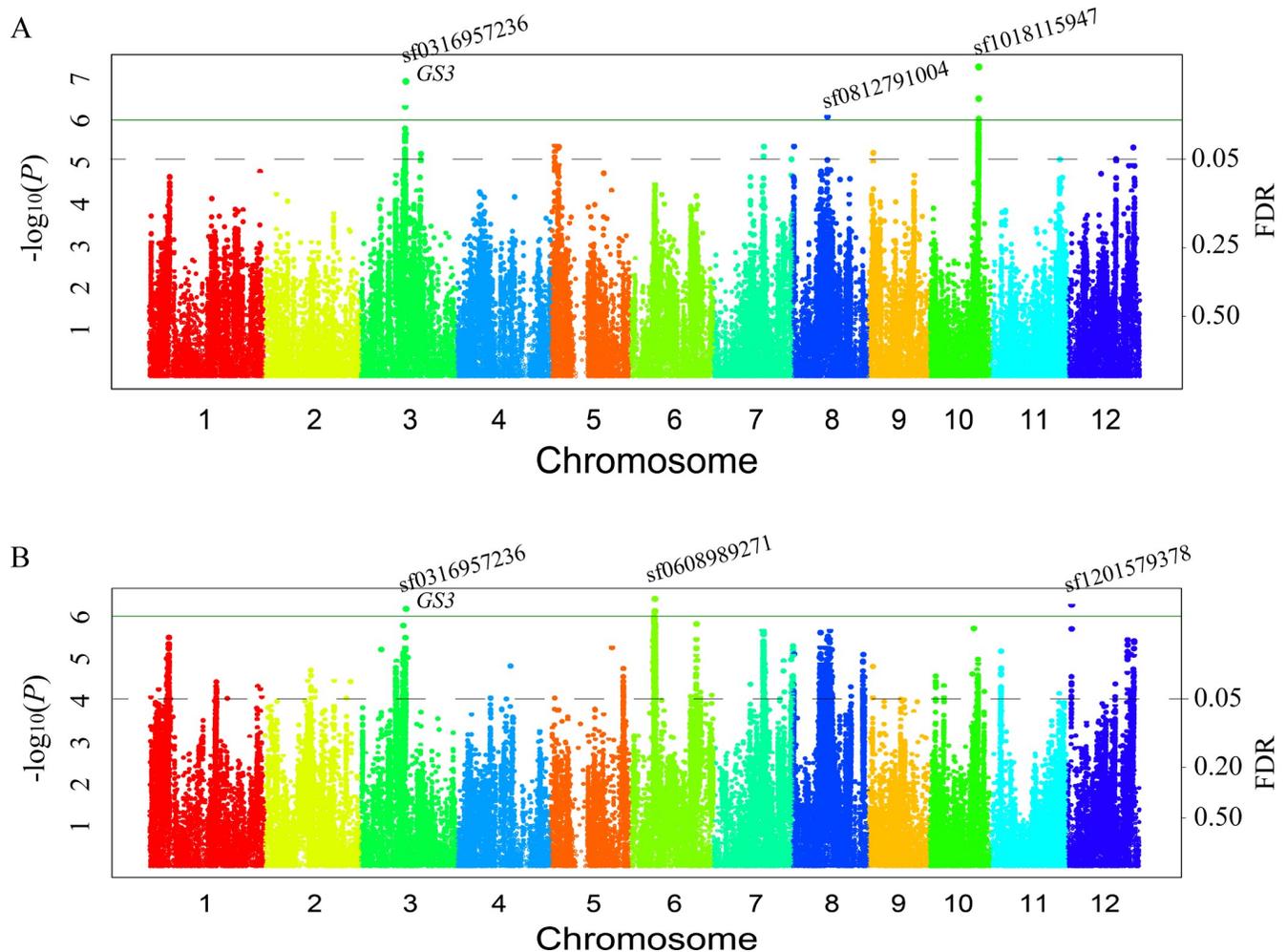


Fig 1. Genome-wide manhattan plots of association mapping for grain length. (A) Grain length in 2011. (B) Grain length in 2012. Note: the nomenclature of SNP includes the information of chromosome and physical position. sf indicates the name of the SNP, and the first two digits indicate the number of the chromosome and the following digits indicate the physical position. The solid green line represents $-\log_{10}(P) = 6.0$ and the black dotted line represents $FDR = 0.05$.

<https://doi.org/10.1371/journal.pgen.1008191.g001>

Table 2. Summary of GWAS loci of grain length (GL) and grain width (GW).

Loci	QTL	Trait	Year	SNP	Chr	MAF	P-value	FDR	PVE (%)
1	GS3	GL	2011	sf0316957236	3	0.3	1.25E-07	3.62E-02	31.81
2	GS3	GL	2012	sf0316957236	3	0.31	6.76E-07	3.28E-02	37.49
3	qGL6	GL	2012	sf0608989271	6	0.39	3.90E-07	3.28E-02	11.91
4	qGL8	GL	2011	sf0812791004	8	0.31	8.36E-07	3.62E-02	19.88
5	qGL10	GL	2011	sf1018115947	10	0.38	5.78E-08	3.62E-02	17.58
6	qGL12	GL	2012	sf1201579378	12	0.14	5.34E-07	3.28E-02	11.86
7	qGW3.1	GW	2011	sf0313650794	3	0.14	1.67E-07	2.59E-03	0.17
8	qGW3.2	GW	2012	sf0315013153	3	0.3	1.32E-07	7.32E-04	46.03
9	qSW5	GW	2011	sf0505372028	5	0.35	8.40E-15	2.08E-09	45.20
10	qSW5	GW	2012	sf0505372028	5	0.39	5.30E-14	1.31E-08	44.79
11	qGW7	GW	2011	sf0707318737	7	0.1	4.17E-07	4.81E-03	15.23
12	qGW7	GW	2012	sf0707314881	7	0.1	1.71E-08	5.58E-04	16.86

Chr, chromosome; MAF, minor allele frequency; FDR, false discovery rate; PVE, phenotype variation explained by each locus

<https://doi.org/10.1371/journal.pgen.1008191.t002>

Candidate gene for grain width was identified in associated loci

The SNP sf0707314881 for grain width was identified with a *p*-value of 1.71E-08 in 2012, and the SNP sf0707318737 for grain width was identified with a *p*-value of 4.17E-07 in 2011. The r^2 of LD = 0.91 between the sf0707314881 and the sf0707318737; thus, these belong to the same QTL interval (*qGW7*). 219 SNPs were found to be associated with grain width ($P < 5.00E-06$ as threshold line), and 33 putative genes were found in the region of *qGW7* (S1 Table). Two putative genes (LOC_Os07g13170 and LOC_Os07g12730) were differentially expressed (fold change = 1.5100, $P = 0.0077$; fold change = 0.5338, $P = 0.0171$, respectively) *via* microarray data in young panicles in EMATA YIN and *Nipponbare* (S1 Table), with grain widths of 3.06 mm and 3.43 mm, respectively. LOC_Os07g13170 (*OsSNB*), as a candidate drought resistance gene, has been detected in our previous study [33]. Thus, *OsSNB* was cloned from *Nipponbare*, and the over-expression lines were obtained under the 35S promoter *via* transformation. The results showed that the grain width of over-expression lines decreased significantly (Fig 3D, Fig 4). The results indicate *OsSNB* as the candidate gene for grain width in the *qGW7* locus.

OsSNB plays an important role for grain shape

Both sf077544175 and sf077544270 at the promoter of *OsSNB* (LOC_Os07g13170) were detected for grain width with a *p*-value of 1.48E-06 in 2012. *OsSNB* is an AP2 transcription factor and is involved in the development of rice reproductive growth [34, 35].

To test whether *OsSNB* regulates grain width, two gRNA constructs were generated and introduced into the *Nipponbare* to knock out the *OsSNB* gene *via* a CRISPR/Cas 9 strategy [36]. Several homozygous mutant plants were obtained. The mutant plants were identified *via* sequencing of PCR products. The two targets of this gene were knocked out (Fig 3A), and three types of knocked out mutants were identified: KO1-1 (-13 nucleotides), KO1-2 (-1 nucleotide), and KO2 (+1 nucleotide) (Fig 3B). Furthermore, overexpression lines were obtained under the 35S promoter *via* transformation. The transcription levels of overexpression lines increased significantly, showing an approximate 49-fold change compared to wild type (Fig 3C). Compared to wild type, the grain width and 1,000-grain weight of the KO1 mutant lines through first sgRNA vector had increased significantly (Fig 4). Moreover, the grain width and grain length, as well as the 1,000-grain weight of the over-expression lines decreased significantly (Fig 4). The KO2 mutant lines from the second sgRNA resulted in indeterminacy of the

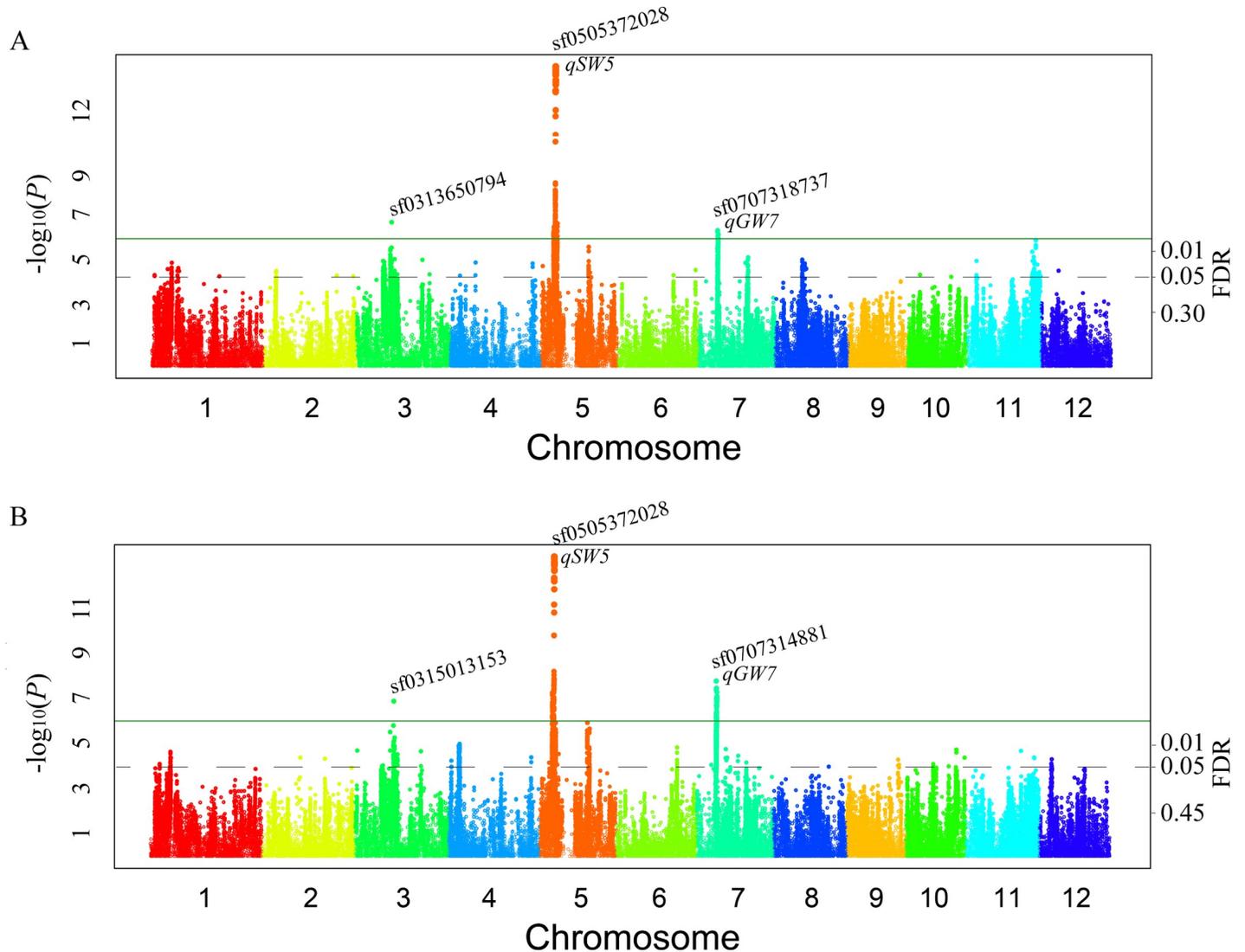


Fig 2. Genome-wide manhattan plots of association mapping for grain width. (A) Grain width in 2011. (B) Grain width in 2012.

<https://doi.org/10.1371/journal.pgen.1008191.g002>

floret, which were similar to the *snb* T-DNA knockout mutants reported in a previous study [37]. To obtain the glume cell size of transgenic plants, histological spikelet analysis was conducted. Cross-sections of central parts of the spikelet showed that the glume cells in transgenic KO-1 plants were longer and larger than the corresponding layer in both OE-2 and wild type plants (Fig 4G). These results indicate the involvement of *OsSNB* in cell size regulation during organ development.

A number of genes that control grain size have been isolated *via* the map-based cloning strategy and have been functionally characterized. To further analyze the regulatory relationship between *OsSNB* and other known functional genes that control grain shape and grain weight, the transcription level of eight known functional genes was analyzed using RT-qPCR (Fig 5). Compared to wild type, the transcription level of *GS5* as positive regulator of grain size was increased in KO1 mutant lines with big grain size, and decreased in over-expression lines with small grain size. The transcription level of *TGW6* acts as a negative regulator and decreased significantly in the KO1 mutant lines with big grain size. These results imply that

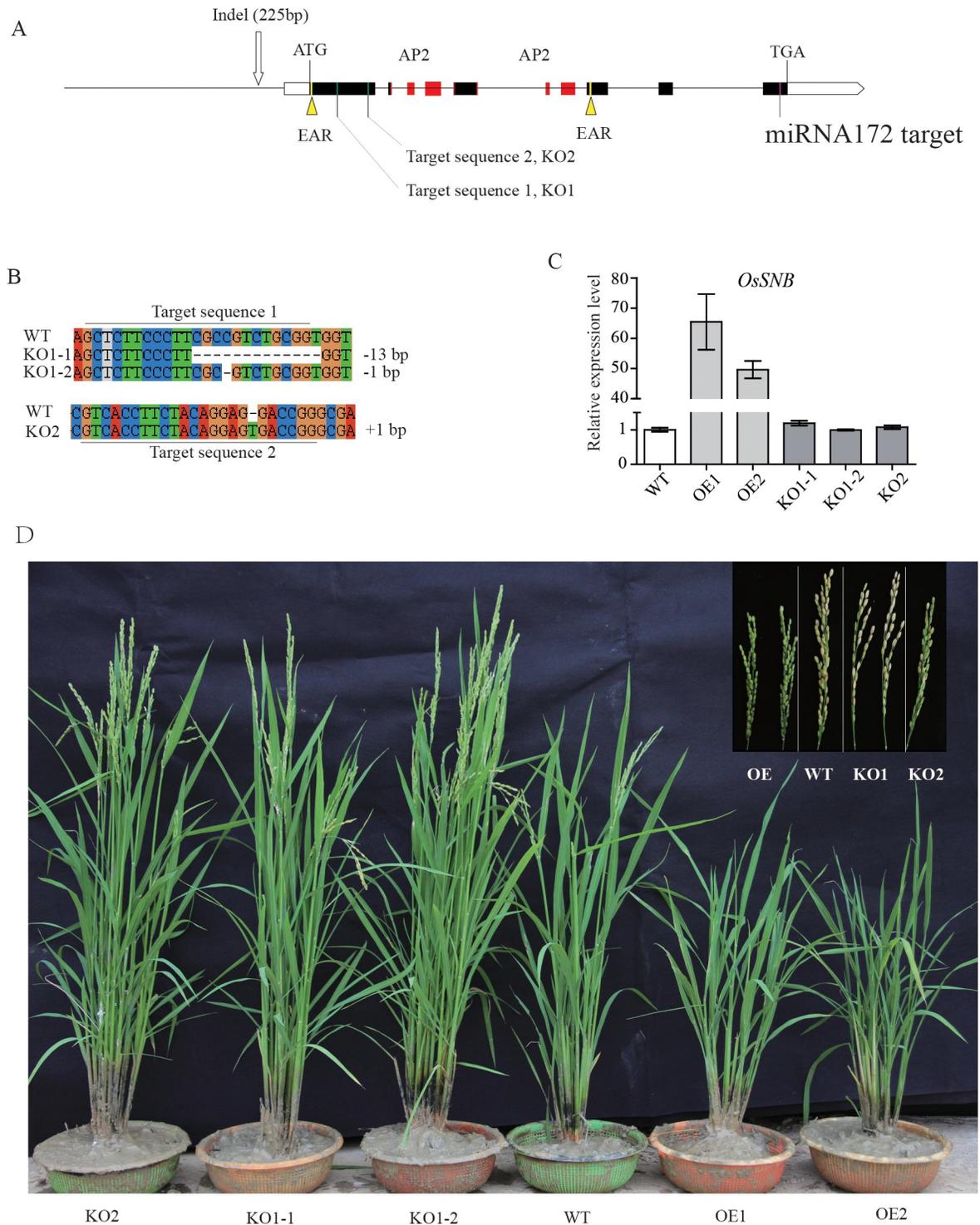


Fig 3. Molecular identification of *OsSNB* transgenic plants. (A) The structure of *OsSNB*. (B) The knockout mutant plants were obtained by CRISPR/Cas 9 technology. (C) Relative expression levels of *OsSNB* in transgenic plants. WT: wildtype; OE1, OE2: overexpression lines; KO1-1, KO1-2, KO2: knockout mutant lines. (D) Phenotypes of transgenic rice plants and WT plants.

<https://doi.org/10.1371/journal.pgen.1008191.g003>

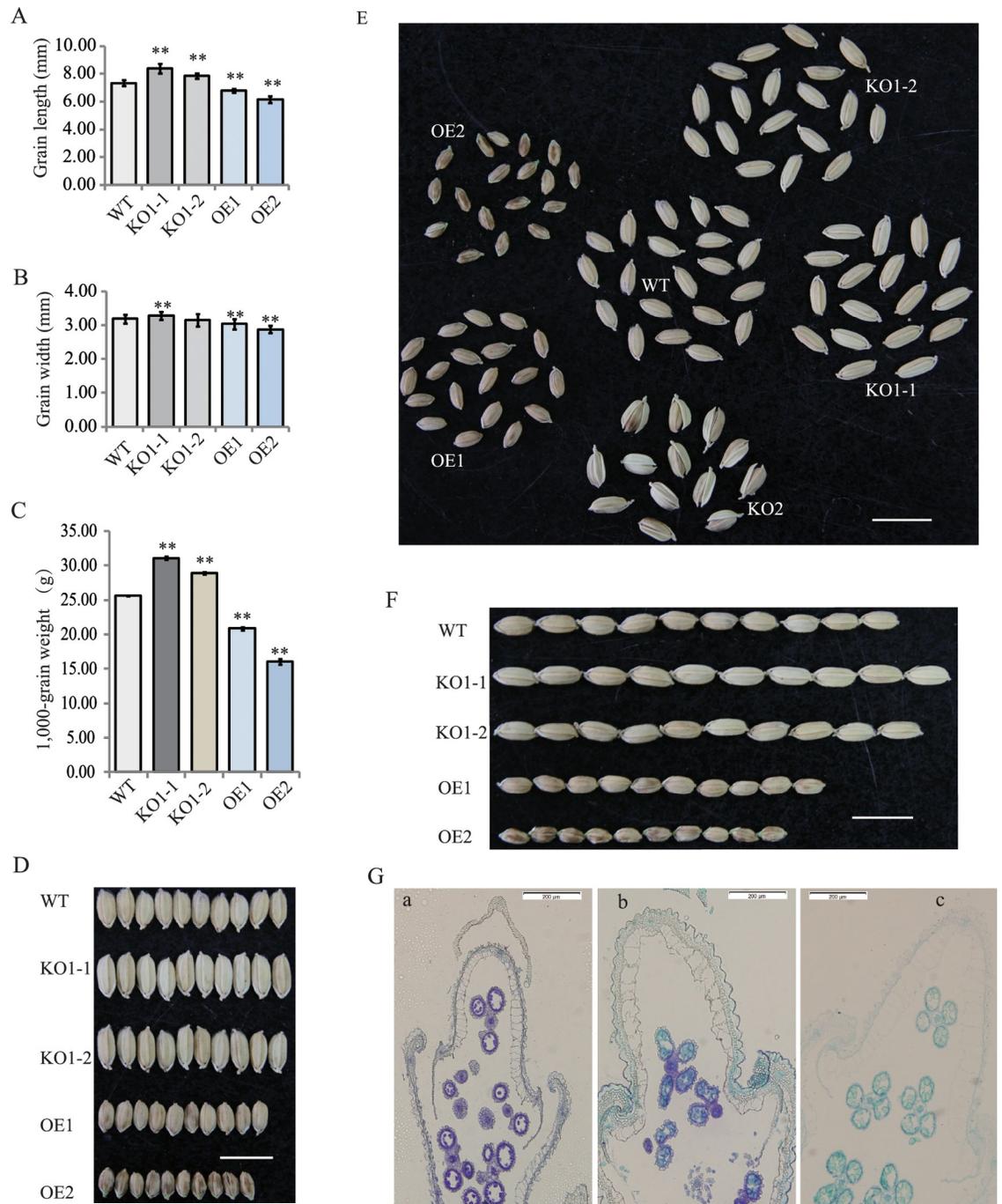


Fig 4. Grain shape was regulated by *OsSNB*. (A) Grain length of transgenic plants. (B) Grain width of transgenic plants. (C) 1,000-grain weight. (D-F) Phenotypes of transgenic plants and wild-type; Bar = 10 mm. (G) Histological analysis of young spikelet in transgenic plants and wild type; OE2 (a), WT (b), KO1-1 (c); Bar = 200 μ m.

<https://doi.org/10.1371/journal.pgen.1008191.g004>

OsSNB may either directly or indirectly regulate the expression of *GS5* and *TGW6*. Additionally, the transcription levels of *OsWAK11* were greatly elevated in the KO1 mutant lines with big grain size. As a cell wall-associated receptor kinase, *OsWAK11* is involved in cell expansion, and its expression is regulated by aluminum, sodium, and copper [38]. However, the

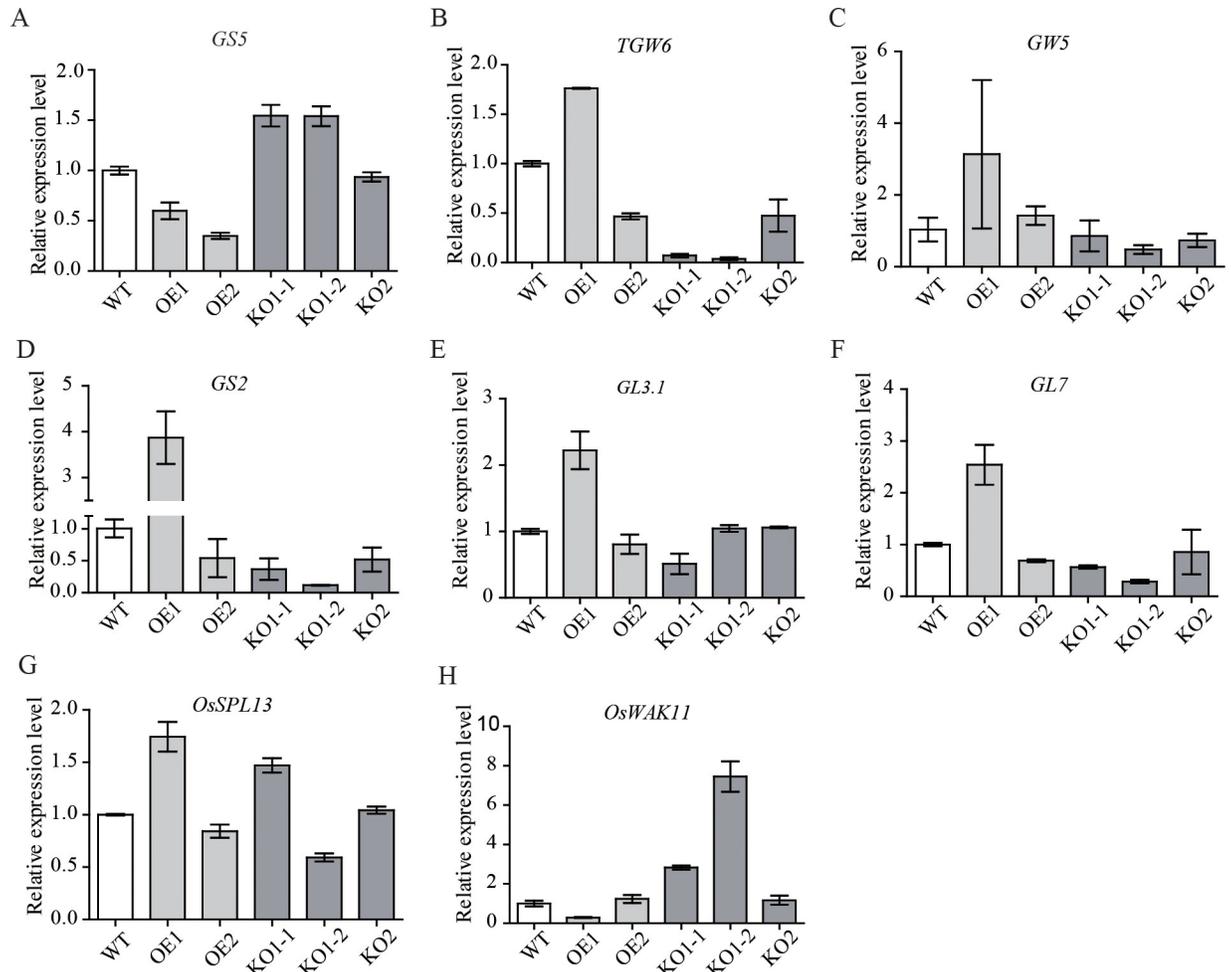


Fig 5. Relative expression levels of known functional genes controlling grain size and grain weight. (A-G) Known functional genes controlling grain size. (H) *OsWAK11*, a cell wall-associated receptor kinase. WT: wild type; OE1, OE2: overexpression lines; KO1-1, KO1-2, KO2: knockout mutant lines.

<https://doi.org/10.1371/journal.pgen.1008191.g005>

Table 3. Summary of parameters for the analysis of polymorphism loci in *OsSNB*.

Parameters	Promoter	Coding region	Downstream	Entire region
Total length of amplicons(bp)	2014	3679	1169	6862
Number of all sequence variants (SNPs and indels)	33	25	7	65
Number of SNPs	25	23	7	55
Number of indels	8	2	0	10
Average indels length (bp)	30.875	4.5	-	25.6
Pi	0.00368	0.00257	0.00173	0.00275
Tajima's D	2.28348*	3.72606***	1.40454	3.09934**
Fu and Li's D*	1.83376**	1.83376**	1.15433	2.32478**
Fu and Li's F*	2.41724**	3.10554**	1.47627	3.19789**

*Indicates statistical significance at $P < 0.05$

<https://doi.org/10.1371/journal.pgen.1008191.t003>

expression of these genes remained unaffected in KO2 mutant lines (Fig 5). This suggests that the *OsSNB* fragment with its nuclear localizing signal also has a specific function. Additionally, *OsSNB* can affect both plant growth and development. For example, compared to wild type plants, the over-expression lines of *OsSNB* have delayed flowering (Fig 3D).

Polymorphism of the *OsSNB* gene

OsSNB was sequenced from the promoter to the 3'-UTR in 168 accessions of the natural population and 65 polymorphic loci were identified (Table 3, S2 Table). The aligned length, including both promoter and open reading frame region of the *OsSNB* gene was 6862 nucleotides. A total of 10 indels and 55 SNPs were detected among the sequenced regions. The number of SNPs exceeded the number of indels. 33 polymorphism loci were located at the promoter region, four and three SNPs were located at exon 1 and exon 10, respectively, 16 SNPs and two indels were located at intron regions, and seven SNPs were located at 3'-UTR. Seven SNPs that were located at the exon region caused non-synonymous mutation. +148 T/C and +149 C/T, result TCG (Ser) to CCG (Pro), TCG (Ser) to CTG (Leu); +230 T/C result CTG (Leu) to CCG (Pro); +242 T/C result ATG (Met) to ACG (Thr); +3612 G/A result CGC (Arg) to CAC (His); +3614 T/G result TCG (Ser) to GCG (Ala); +3674 G/A result GCC (Ala) to ACC (Thr). No mutation is located at the target loci of miR172. Haplotype analyses were performed based on 65 polymorphism loci of *OsSNB* and the genotype was classified into eight haplotypes (Hap) (Fig 6, S2 Table). Hap 1 and Hap 8 mainly contained the *indica* subpopulation, while Hap 2, 3, 4, 6 and 7 mainly contained the *japonica* subpopulation. Hap 3 has the highest grain width and Hap 1 has the highest panicle length.

Nucleotide diversity and selection of *OsSNB*

A total of 6862 bp of genomic region covering *OsSNB*, including a 2014 bp upstream region, 3679 bp of the coding region, and a 1169 bp downstream region, were sequenced in 168 rice accessions. Nucleotide substitutions (SNP) and both insertion and deletion (InDel) variations at the *OsSNB* locus are summarized in Table 3 and S2 Table. The level of polymorphism was highest upstream and a 225 bp InDel at -357 bp was identified upstream. A sliding window of 100 bp at a step size of 25 bp was used to calculate the overall nucleotide diversity (Pi). Pi of *OsSNB* was 0.00275 for all 168 rice accessions, and the coding regions were less diverse than upstream regions. The lowest nucleotide diversity was downstream. The results showed that upstream regions were more diverse than other regions.

Tajima's D statistic was calculated for upstream, coding region, and downstream, as well as the entire region. Tajima's D statistic was significant when the entire region, as well as upstream and coding regions were considered. Both Fu and Li's D* and F* statistics were significant for upstream and coding regions, as well as for the entire region, and not significant for downstream (Table 3). These results indicate that upstream and coding regions may be currently undergoing selection.

An indel in the promoter region of *OsSNB* is highly associated with rice grain width and was validated as a functional marker

Compared to other haplotypes, Hap 3 has a 225 bp insertion located at -357 bp in the promoter region (S2 Table). A significant difference in grain width between Hap 3 and other haplotypes (including Hap 1, Hap 7, and Hap 8) was observed (S2 Table). Thus, an indel marker (*OsSNB_Indel2*) was developed across a 225 bp insertion/deletion. To test whether *OsSNB_Indel2* was a reliable functional marker, the PCR products of the natural population were used for validation by electrophoresis. The results showed that *OsSNB_Indel2* can distinguish between

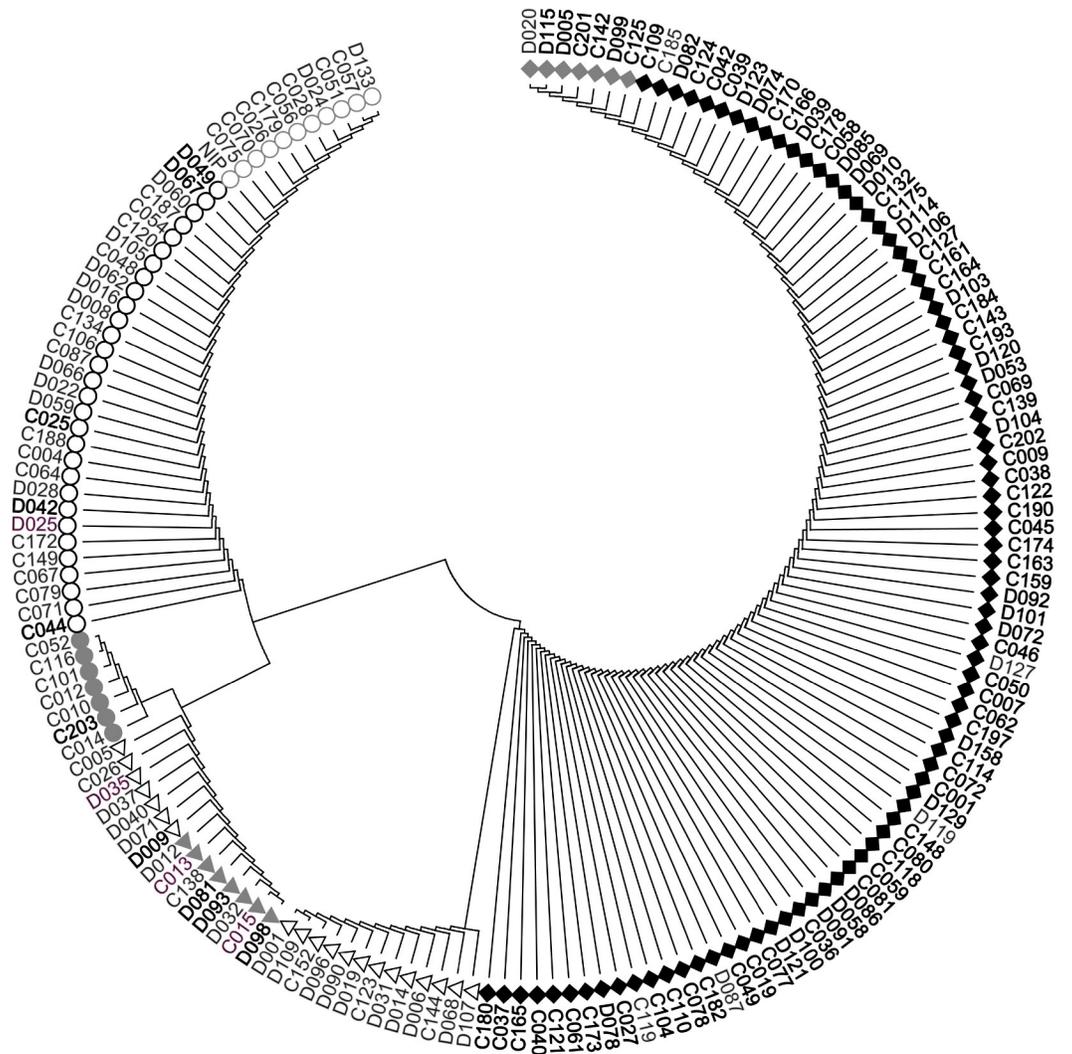


Fig 6. Cluster analysis of OsSNB. Black font represents *indica*, gray font represents *japonica*, purple font represents *aus*. NIP = *Nipponbare*.

<https://doi.org/10.1371/journal.pgen.1008191.g006>

different genotypes (Fig 7). The grain widths of 22 rice accessions with 225 bp insertion were 3.82 ± 0.39 mm, the grain widths of 204 rice accessions with 225 bp deletion were 3.28 ± 0.37 mm, the *p* value of Student's *t*-test was $1.48E-06$, and the *p* value of Mann-Whitney *U* test was $3.38E-08$. The null hypothesis for this test was that the grain widths are not different between Hap 3 and others. OsSNB_Indel2 can also be used to distinguish the genotypes of recombinant inbred lines (RILs), which were developed from both Zhenshan97B (*indica*) and IRAT109 (*japonica*). The grain widths of 71 recombinant inbred lines with 225 bp insertion were 3.72 ± 0.18 mm. The grain widths of 117 recombinant inbred lines with 225 bp deletion were 3.61 ± 0.30 mm, the *p* value of Student's *t*-test was $2.58E-03$ (Fig 7). These results clearly confirm that this mutation (225 bp insertion or deletion) was highly associated with grain width. Hap 3 was only discovered in *japonica* subspecies, and Hap 3 can be used for marker assisted selection (MAS) for increasing grain width in *indica* subspecies. OsSNB_Indel2 is likely a functional marker for the improvement of rice grain width.

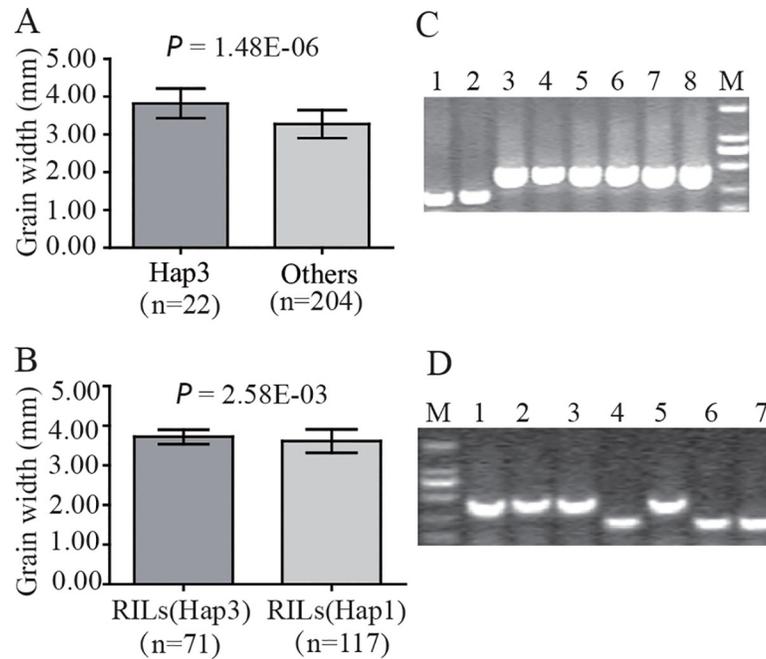


Fig 7. Genotype assay of GWAS population and RILs by the indel marker OsSNB_Indel2. (A) GWAS population. (B) Recombinant inbred lines (RILs). (C-D) Marker assay of OsSNB_Indel2 in natural population and RILs, respectively. Marker (M) indicates the DNA length of 2000 bp, 1000 bp, 750 bp, 500 bp, 250 bp and 100 bp.

<https://doi.org/10.1371/journal.pgen.1008191.g007>

Discussion

Novel QTLs were identified by GWAS

Grain shape is controlled by several hundred QTLs and is an important grain yield characteristic in rice. This study detected five significant associated intervals for grain length *via* GWAS in 270 rice accessions, and showed that GS3 was very closely positioned to the significantly associated interval on chromosome 3 [3, 4]. Four significantly associated intervals for grain width were detected by GWAS in the current study. One associated SNP (sf0505364734) was located on the promoter region of *qSW5* [9]. Two associated loci (sf0313650794 and sf0315013153) were located on chromosome 3, where no grain width QTL/gene was reported in previous studies; thus, these are two novel QTLs. Furthermore, another associated region on chromosome 7 was a novel locus for grain width, which was far away from the functional genes *GL7/GW7* [15, 16] and *OsSPL13* [32], located on chromosome 7. Compared to previous studies, many QTLs for grain size were quickly mapped to smaller intervals in this study. These QTL intervals were more conducive to marker-assisted selection in breeding. Furthermore, smaller QTL intervals enable easier identification of candidate genes. Additionally, gene expression profiling data of predicted genes in the QTL interval can be used to assist in identifying the most promising candidates.

Mechanism of OsSNB controlling grain size

OsSNB is a nuclear protein, that contains two highly conserved AP2 domains [37] and two ETHYLENE-RESPONSE FACTOR Amphiphilic Repression (EAR) motifs (Fig 3A). The transcription level of *OsSNB* is most strongly expressed in the newly emerging spikelet meristems, as well as expressed in other tissues, including shoot, root, sheath, leaf, and seed [37]. In previous studies, *SUPERNUMERARY BRACT (SNB)* was reported to regulate the transition from

the spikelet meristem to the floral meristem and to regulate the floral organ development in rice [35, 37], while inhibiting the flowering time in rice by suppressing the expression of *Ehd1* [39]. Overexpression of *rSNB* without the miR172 target site created branches and spikelets in panicle and RNAi inhibitor of *OsSNB* was the opposite [34], which was similar to the KO1 knockout mutants that had decreased branches and spikelets in this study (Fig 3D). The *snb* T-DNA knockout mutants showed that the transition from spikelet meristems to floral meristems was delayed, which resulted in indeterminacy of the floret [37]. This phenotype was also found in KO2 mutant plants in the present study. *OsSNB* is one of the target genes of miR172, over-expression of miR172 also caused loss of spikelet determinacy and floral organ abnormalities, similar to *snb* T-DNA knockout mutants in rice [40]. Furthermore, no significant difference was found in grain size between wild type and KO2 mutant plants. This indicates that the peptide with 130 amino acid length at the N-terminal is possible playing a competitive role to normal *OsSNB* protein like miR172, because the KO2 mutant still retains a fragment of the nuclear localizing signal. However, further investigations are required to understand other possible roles of *OsSNB*.

It also remains unclear whether *OsSNB* regulates grain size. It has been shown to regulate the transition from spikelet meristem to floral meristem and to regulate the floral organ development [37]. This study confirmed that *OsSNB* has a novel function, and can regulate grain size. Compared to wild type, KO1 mutant plants have increased grain size; however, over-expression of *OsSNB* plants decreased grain size. Histological analysis of spikelets also showed that *OsSNB* knockout could promote cell enlargement in glumes (Fig 4). These results confirm that the novel grain size gene, *OsSNB* was identified in *qGW7* via GWAS. Moreover, *OsSNB* can regulate the transcription levels of *GS5* and *TGW6*, two important known functional genes controlling grain size. As a positive regulator, *GS5* controls grain size by BR signaling through regulating grain width, filling, and weight [10, 41]. *GS5* functions putatively as a positive modulator upstream of cell cycle genes; furthermore, its over-expression may result in an increase in cell numbers by promoting mitotic division [10]. The Nipponbare *tgw6* allele affected the timing of the transition from syncytial to cellular phase by controlling IAA supply and limiting both cell number and grain length. Loss of function of the Kasalath allele enhanced grain weight via pleiotropic effects on source organs and leads to significant yield increases [13]. These results implied that *OsSNB* may regulate cell number and size and thus affect grain shape through simultaneous BRs signaling and IAA signaling simultaneously (Fig 8). *OsSNB* has been reported to control inflorescence architecture and the establishment of floral meristem in rice [35, 37], and inhibited flowering time in rice by suppressing expression of *Ehd1* [39]. Simultaneously, *OsSNB* was reported to be involved in abiotic stress signaling and was induced by both NaCl and drought [33]. This indicated that *OsSNB* plays a multifunctional role in the response to abiotic stresses and the development of component traits of grain yield.

Polymorphism analysis of *OsSNB* and development of a functional marker for grain size improvement

In this study, *OsSNB* was sequenced from the promoter to the 3'-UTR in 168 rice accessions. A total of 65 polymorphism sites including 10 indels and 55 SNPs were identified. Rice, as a self-pollinating crop, has a lower nucleotide diversity than maize, which is a typical outcrossing crop [42]. The nucleotide diversity of *OsSNB* exceeds that of *OsCl*, and is lower than that of *OsWx* in rice [43]. Neutrality analysis showed that the upstream and coding regions of *OsSNB* may be undergoing selection; however, the values of Tajima's D and Fu and Li's D* and F* based on the *Wx* and *OsCl* locus did not significantly differ from neutral expectations. Eight haplotypes at the *OsSNB* locus were identified by haplotype analysis based on polymorphism

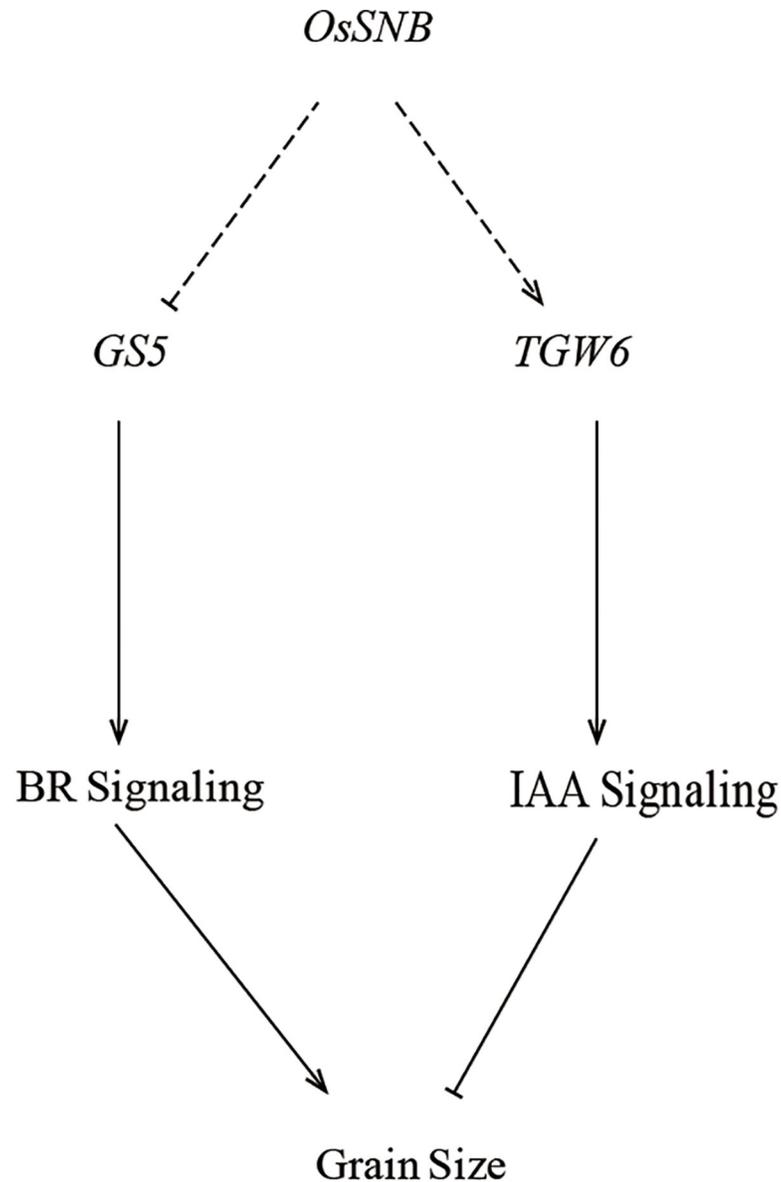


Fig 8. The hypothetical molecular mechanism of *OsSNB* regulation of grain size.

<https://doi.org/10.1371/journal.pgen.1008191.g008>

sites. The rice accessions with Hap 3 have the highest grain width and Hap 3 is only discovered in *japonica* subspecies. Additionally, the grain width of the lines with Hap 3 was confirmed to be longer than that of lines with Hap 1, using *OsSNB_Indel2* in RILs population. Thus, Hap 3 can be potentially used to increase grain width in *indica* subspecies. *OsSNB_Indel2* was designed as a functional marker for the improvement of rice grain width, which could be directly used to assist selection for grain width improvement. This marker has practical utility in a variety of tests based on a 225 bp insertion or deletion of specific alleles for rice breeders. This simple, robust, and economic marker genotyping method will both simplify and streamline MAS for grain width in rice breeding.

Conclusion

The associated loci and genes identified in this study highlight the important underlying mechanisms of grain size. Furthermore, the novel grain size gene *OsSNB* was identified *via* GWAS and was functionally characterized. This locus has eight haplotypes, and Hap 3 has important application value in MAS for grain width in rice breeding. These results provide information for dissecting the genetic and molecular basis of grain size and for improving the grain size of rice through molecular breeding in the near future.

Materials and methods

Plant materials

Two rice collections were used in this study. Collection 1 comprised of 270 rice germplasms, including four *aus* accessions, 111 *japonica* accessions, and 155 *indica* accessions. This population has previously been used for GWAS of mesocotyl elongation and drought resistance [24, 44]. This collection was used for GWAS analysis. Collection 2 comprised of 200 F₁₀ recombinant inbred lines (RILs), which were developed from Zhenshan97B (*indica*) and IRAT109 (*japonica*). This collection was used for molecular marker verification.

Phenotyping

To measure grain size, Collection 1 was grown in the field at the Baihe experimental station of the Shanghai Academy of Agricultural Sciences (Shanghai, China) in 2011 and 2012. Collection 2 and two parents (IRAT109 and Zhenshan 97B) were grown in the field at the Baihe experimental station of the Shanghai Academy of Agricultural Sciences in 2014, and grown at the Hainan Experimental Station of the Shanghai Academy of Agricultural Sciences (Hainan, China) in 2015.

Dry seeds were evaluated for grain width and length *via* morphology analyzer (<http://www.jsjeda.com>). Descriptive statistics analysis of the phenotype data was performed by SPSS version 19.0 (IBM).

GWAS analysis

Genotype data of the natural population were re-sequenced by Hiseq 2000. Paired-end sequence reads were mapped to a rice reference genome sequence of *japonica* cv. *Nipponbare* (MSU v6.1), and were used for SNP identification, following the procedures described by Wu et al. [24].

Genome-wide association mapping was conducted by the Efficient mixed model Association (EMMA) method using the Genomic Association and Prediction Integrated Tool (GAPIT) software package in R [45]. A total of 996,722 SNP markers (minimum allelic frequency (MAF) ≥ 0.05), distributed on 12 chromosomes, were used for GWAS in this study. 144,995 SNPs, with missing data below 10% in this natural population, were used to calculate kinship among individuals. Principal component analysis (PCA) was used to adjust the population structure. The genome-wide threshold of $-\log_{10}(P) = 6.0$ was calculated from the formula of “ $-\log_{10}(1 / \text{effective number of SNPs})$ ”. r^2 of LD was calculated by PLINK ver1.07. The variance explained the phenotype by the lead SNP (with the lowest P value) and was calculated by comparing the sum of squares of the variance between groups and the sum of squares of the variance of the full model. These procedures have been described by Ma et al. [44]. The original phenotype data for GWAS are provided in Supplemental S3 Table.

Sequence analysis of *OsSNB*

OsSNB was sequenced from the promoter to the 3'-UTR in 168 rice accessions. Sequence alignment was conducted with ClustalX1.83. Cluster analysis was conducted *via* MEGA 6.0 [46]. Allelic diversities were determined by DnaSP5.0 [47], Pi parameters of nucleotide diversity were estimated by the average number of nucleotide differences per site between any two DNA sequences. Tajima's D [48] and Fu and Li's D* and F* statistical tests [49] were used to test the evidence of neutral evolution, which were calculated by DnaSP5.0.

Molecular cloning and transformation of rice

To construct the overexpression vector, the full-length cDNA of *OsSNB* with CAMV35S promoter was digested with BamHI and BstEII, and then ligated to pCAMBIA1300. For the construction of the Crispr/Cas9 system knockout vector, two *OsSNB* site-specific single guide RNAs (sgRNA1, 5'-gctcttccttcgcccgtctgccc-3'; sgRNA2, 5'-gtcaccttctacaggaggaccgg-3') were introduced into pYLCRISPRCas9Pubi-H according to previously described methods [36]. The three vectors were introduced into the *japonica* rice cultivar Nipponbare *via* the *Agrobacterium tumefaciens*-mediated transformation method [50].

RNA isolation and RT-qPCR

Total RNA was isolated from rice leaves using the TRNzol reagent (TIANGEN, China). cDNA templates were synthesized using Superscript II reverse transcriptase (TaKaRa, Japan) according to the manufacturer's instructions. Real-time quantitative RT-qPCR was performed on a CFX96 Real-Time PCR system (Bio-Rad, USA) using SYBR Premix Ex Taq (TaKaRa, Japan) according to the manufacturer's instructions. The rice Actin1 gene was used as endogenous control. All primers are listed in Supplemental S4 Table.

Histological analysis

Fresh spikelet samples were collected and fixed in FAA solution for 24 h. After dehydrated in a gradient series of ethanol (70% alcohol, 85% alcohol, and 95% alcohol contain 1% eosin Y) once and pure alcohol twice, each step for 1 h, Spikelet samples were transparentized by a different gradient concentration of chloroform; the utilized volume ratio of chloroform and alcohol was 1/5, 2/5, 3/5, 4/5, and pure chloroform. Then, the samples were infiltrated *via* tiny paraffin in chloroform (at least two days) and embedded in paraffin. Furthermore, microtome sections of 10 μ m thickness were put on glass slides. Finally, the slides were deparaffinized in 100% xylene prior to staining by toluidine blue 0.05% and observation *via* light microscope.

Supporting information

S1 Table. Selection of putative genes in *qGW7*.

(XLSX)

S2 Table. Haplotype analysis of *OsSNB*.

(XLSX)

S3 Table. Phenotype data of the GWAS population.

(XLSX)

S4 Table. Primers used in this study.

(XLSX)

Acknowledgments

We would like to thank the National Key Laboratory of Crop Genetic Improvement at the Huazhong Agricultural University for providing the seeds of the rice germplasm collection.

Author Contributions

Conceptualization: Xiaosong Ma, Shunwu Yu.

Data curation: Xiaosong Ma, Fangjun Feng, Hanwei Mei, Hongyan Liu, Shunwu Yu.

Formal analysis: Xiaosong Ma, Fangjun Feng, Hanwei Mei, Hongyan Liu, Shunwu Yu.

Funding acquisition: Xiaosong Ma, Kai Xu, Lijun Luo, Shunwu Yu.

Investigation: Xiaosong Ma, Yu Zhang, Ibrahim Eid Elesawi, Kai Xu, Tianfei Li, Hongyan Liu, Ningning Gao, Chunli Chen, Shunwu Yu.

Methodology: Xiaosong Ma, Shunwu Yu.

Project administration: Lijun Luo, Shunwu Yu.

Resources: Xiaosong Ma, Hanwei Mei.

Supervision: Lijun Luo, Shunwu Yu.

Validation: Xiaosong Ma, Hongyan Liu, Lijun Luo, Shunwu Yu.

Visualization: Xiaosong Ma, Hongyan Liu, Lijun Luo, Shunwu Yu.

Writing – original draft: Xiaosong Ma, Hongyan Liu, Lijun Luo, Shunwu Yu.

Writing – review & editing: Lijun Luo, Shunwu Yu.

References

1. Xing Y, Zhang Q. Genetic and molecular bases of rice yield. *Annu Rev Plant Biol.* 2010; 61:421–42. Epub 2010/03/03. <https://doi.org/10.1146/annurev-arplant-042809-112209> PMID: 20192739.
2. Huang R, Jiang L, Zheng J, Wang T, Wang H, Huang Y, et al. Genetic bases of rice grain shape: so many genes, so little known. *Trends in plant science.* 2013; 18(4):218–26. Epub 2012/12/12. <https://doi.org/10.1016/j.tplants.2012.11.001> PMID: 23218902.
3. Fan C, Xing Y, Mao H, Lu T, Han B, Xu C, et al. *GS3*, a major QTL for grain length and weight and minor QTL for grain width and thickness in rice, encodes a putative transmembrane protein. *Theor Appl Genet.* 2006; 112(6):1164–71. Epub 2006/02/03. <https://doi.org/10.1007/s00122-006-0218-1> PMID: 16453132.
4. Mao H, Sun S, Yao J, Wang C, Yu S, Xu C, et al. Linking differential domain functions of the *GS3* protein to natural variation of grain size in rice. *Proc Natl Acad Sci U S A.* 2010; 107(45):19579–84. Epub 2010/10/27. <https://doi.org/10.1073/pnas.1014419107> PMID: 20974950; PubMed Central PMCID: PMC2984220.
5. Qi P, Lin YS, Song XJ, Shen JB, Huang W, Shan JX, et al. The novel quantitative trait locus *GL3.1* controls rice grain size and yield by regulating Cyclin-T1;3. *Cell Res.* 2012; 22(12):1666–80. Epub 2012/11/14. <https://doi.org/10.1038/cr.2012.151> PMID: 23147796; PubMed Central PMCID: PMC3515756.
6. Zhang X, Wang J, Huang J, Lan H, Wang C, Yin C, et al. Rare allele of *OsPPKL1* associated with grain length causes extra-large grain and a significant yield increase in rice. *Proc Natl Acad Sci U S A.* 2012; 109(52):21534–9. Epub 2012/12/14. <https://doi.org/10.1073/pnas.1219776110> PMID: 23236132.
7. Hu Z, He H, Zhang S, Sun F, Xin X, Wang W, et al. A Kelch motif-containing serine/threonine protein phosphatase determines the large grain QTL trait in rice. *J Integr Plant Biol.* 2012; 54(12):979–90. Epub 2012/11/10. <https://doi.org/10.1111/jipb.12008> PMID: 23137285.
8. Weng J, Gu S, Wan X, Gao H, Guo T, Su N, et al. Isolation and initial characterization of *GW5*, a major QTL associated with rice grain width and weight. *Cell Res.* 2008; 18(12):1199–209. Epub 2008/11/19. <https://doi.org/10.1038/cr.2008.307> PMID: 19015668.

9. Shomura A, Izawa T, Ebana K, Ebitani T, Kanegae H, Konishi S, et al. Deletion in a gene associated with grain size increased yields during rice domestication. *Nat Genet.* 2008; 40(8):1023–8. Epub 2008/07/08. <https://doi.org/10.1038/ng.169> PMID: 18604208.
10. Li Y, Fan C, Xing Y, Jiang Y, Luo L, Sun L, et al. Natural variation in *GS5* plays an important role in regulating grain size and yield in rice. *Nat Genet.* 2011; 43(12):1266–9. Epub 2011/10/25. <https://doi.org/10.1038/ng.977> PMID: 22019783.
11. Song XJ, Huang W, Shi M, Zhu MZ, Lin HX. A QTL for rice grain width and weight encodes a previously unknown RING-type E3 ubiquitin ligase. *Nat Genet.* 2007; 39(5):623–30. Epub 2007/04/10. <https://doi.org/10.1038/ng2014> PMID: 17417637.
12. Wang S, Wu K, Yuan Q, Liu X, Liu Z, Lin X, et al. Control of grain size, shape and quality by *OsSPL16* in rice. *Nat Genet.* 2012. Epub 2012/06/26. <https://doi.org/10.1038/ng.2327> PMID: 22729225.
13. Ishimaru K, Hirotsu N, Madoka Y, Murakami N, Hara N, Onodera H, et al. Loss of function of the IAA-glucose hydrolase gene *TGW6* enhances rice grain weight and increases yield. *Nat Genet.* 2013; 45(6):707–11. Epub 2013/04/16. <https://doi.org/10.1038/ng.2612> PMID: 23583977.
14. Song XJ, Kuroha T, Ayano M, Furuta T, Nagai K, Komeda N, et al. 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.* 2015; 112(1):76–81. Epub 2014/12/24. <https://doi.org/10.1073/pnas.1421127112> PMID: 25535376; PubMed Central PMCID: PMC4291654.
15. Wang Y, Xiong G, Hu J, Jiang L, Yu H, Xu J, et al. Copy number variation at the *GL7* locus contributes to grain size diversity in rice. *Nat Genet.* 2015. Epub 2015/07/07. <https://doi.org/10.1038/ng.3346> PMID: 26147619.
16. Wang S, Li S, Liu Q, Wu K, Zhang J, Wang Y, et al. The *OsSPL16-GW7* regulatory module determines grain shape and simultaneously improves rice yield and grain quality. *Nat Genet.* 2015. Epub 2015/07/07. <https://doi.org/10.1038/ng.3352> PMID: 26147620.
17. Hu J, Wang Y, Fang Y, Zeng L, Xu J, Yu H, et al. A rare allele of *GS2* enhances grain size and grain yield in rice. *Mol Plant.* 2015. <https://doi.org/10.1016/j.molp.2015.07.002> PMID: 26187814.
18. Liu J, Chen J, Zheng X, Wu F, Lin Q, Heng Y, et al. *GW5* acts in the brassinosteroid signalling pathway to regulate grain width and weight in rice. *Nat Plants.* 2017; 3:17043. Epub 2017/04/11. <https://doi.org/10.1038/nplants.2017.43> PMID: 28394310.
19. Yu J, Buckler ES. Genetic association mapping and genome organization of maize. *Curr Opin Biotechnol.* 2006; 17(2):155–60. Epub 2006/03/01. <https://doi.org/10.1016/j.copbio.2006.02.003> PMID: 16504497.
20. Atwell S, Huang YS, Vilhjalmsson BJ, Willems G, Horton M, Li Y, et al. Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature.* 2010; 465(7298):627–31. Epub 2010/03/26. <https://doi.org/10.1038/nature08800> PMID: 20336072.
21. Li Y, Huang Y, Bergelson J, Nordborg M, Borevitz JO. Association mapping of local climate-sensitive quantitative trait loci in *Arabidopsis thaliana*. *Proc Natl Acad Sci U S A.* 2010; 107(49):21199–204. Epub 2010/11/17. <https://doi.org/10.1073/pnas.1007431107> PMID: 21078970; PubMed Central PMCID: PMC3000268.
22. Brachi B, Faure N, Horton M, Flahauw E, Vazquez A, Nordborg M, et al. Linkage and association mapping of *Arabidopsis thaliana* flowering time in nature. *PLoS Genet.* 2010; 6(5):e1000940. Epub 2010/05/14. <https://doi.org/10.1371/journal.pgen.1000940> PMID: 20463887; PubMed Central PMCID: PMC2865524.
23. Chao DY, Silva A, Baxter I, Huang YS, Nordborg M, Danku J, et al. Genome-wide association studies identify heavy metal ATPase3 as the primary determinant of natural variation in leaf cadmium in *Arabidopsis thaliana*. *PLoS Genet.* 2012; 8(9):e1002923. Epub 2012/09/13. <https://doi.org/10.1371/journal.pgen.1002923> PMID: 22969436; PubMed Central PMCID: PMC3435251.
24. Wu J, Feng F, Lian X, Teng X, Wei H, Yu H, et al. Genome-wide Association Study (GWAS) of mesocotyl elongation based on re-sequencing approach in rice. *BMC Plant Biol.* 2015; 15(1):218. Epub 2015/09/13. <https://doi.org/10.1186/s12870-015-0608-0> PMID: 26362270; PubMed Central PMCID: PMC4566844.
25. Huang X, Wei X, Sang T, Zhao Q, Feng Q, Zhao Y, et al. Genome-wide association studies of 14 agronomic traits in rice landraces. *Nat Genet.* 2010; 42(11):961–7. Epub 2010/10/26. <https://doi.org/10.1038/ng.695> PMID: 20972439.
26. Huang X, Zhao Y, Wei X, Li C, Wang A, Zhao Q, et al. Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. *Nat Genet.* 2012; 44(1):32–9. Epub 2011/12/06. <https://doi.org/10.1038/ng.1018> PMID: 22138690.
27. Chen W, Gao Y, Xie W, Gong L, Lu K, Wang W, et al. Genome-wide association analyses provide genetic and biochemical insights into natural variation in rice metabolism. *Nat Genet.* 2014; 46(7):714–21. Epub 2014/06/09. <https://doi.org/10.1038/ng.3007> PMID: 24908251.

28. Lu Y, Zhang S, Shah T, Xie C, Hao Z, Li X, et al. Joint linkage-linkage disequilibrium mapping is a powerful approach to detecting quantitative trait loci underlying drought tolerance in maize. *Proc Natl Acad Sci U S A*. 2010; 107(45):19585–90. Epub 2010/10/27. <https://doi.org/10.1073/pnas.1006105107> PMID: 20974948; PubMed Central PMCID: PMC2984198.
29. Tian F, Bradbury PJ, Brown PJ, Hung H, Sun Q, Flint-Garcia S, et al. Genome-wide association study of leaf architecture in the maize nested association mapping population. *Nat Genet*. 2011; 43(2):159–62. Epub 2011/01/11. <https://doi.org/10.1038/ng.746> PMID: 21217756.
30. Li H, Peng Z, Yang X, Wang W, Fu J, Wang J, et al. Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels. *Nat Genet*. 2012; 45(1):43–50. Epub 2012/12/18. <https://doi.org/10.1038/ng.2484> PMID: 23242369.
31. Wen W, Li D, Li X, Gao Y, Li W, Li H, et al. Metabolome-based genome-wide association study of maize kernel leads to novel biochemical insights. *Nat Commun*. 2014; 5:3438. Epub 2014/03/19. <https://doi.org/10.1038/ncomms4438> PMID: 24633423; PubMed Central PMCID: PMC3959190.
32. Si L, Chen J, Huang X, Gong H, Luo J, Hou Q, et al. *OsSPL13* controls grain size in cultivated rice. *NATURE GENETICS*. 2016; 48(4):447–56. Epub 2016/03/08. <https://doi.org/10.1038/ng.3518> PMID: 26950093.
33. Yu S, Liao F, Wang F, Wen W, Li J, Mei H, et al. Identification of rice transcription factors associated with drought tolerance using the Ecotilling method. *PLoS One*. 2012; 7(2):e30765. Epub 2012/02/22. <https://doi.org/10.1371/journal.pone.0030765> PMID: 22348023; PubMed Central PMCID: PMC3278407.
34. Wang L, Sun S, Jin J, Fu D, Yang X, Weng X, et al. Coordinated regulation of vegetative and reproductive branching in rice. *Proc Natl Acad Sci U S A*. 2015; 112(50):15504–9. Epub 2015/12/04. <https://doi.org/10.1073/pnas.1521949112> PMID: 26631749; PubMed Central PMCID: PMC4687603.
35. Lee DY, An G. Two AP2 family genes, supernumerary bract (*SNB*) and Osindeterminate spikelet 1 (*OsIDS1*), synergistically control inflorescence architecture and floral meristem establishment in rice. *Plant J*. 2012; 69(3):445–61. Epub 2011/10/19. <https://doi.org/10.1111/j.1365-313X.2011.04804.x> PMID: 22003982.
36. Ma X, Zhang Q, Zhu Q, Liu W, Chen Y, Qiu R, et al. A Robust CRISPR/Cas9 System for Convenient, High-Efficiency Multiplex Genome Editing in Monocot and Dicot Plants. *Mol Plant*. 2015; 8(8):1274–84. Epub 2015/04/29. <https://doi.org/10.1016/j.molp.2015.04.007> PMID: 25917172.
37. Lee DY, Lee J, Moon S, Park SY, An G. The rice heterochronic gene SUPERNUMERARY BRACT regulates the transition from spikelet meristem to floral meristem. *Plant J*. 2007; 49(1):64–78. Epub 2006/12/06. <https://doi.org/10.1111/j.1365-313X.2006.02941.x> PMID: 17144896.
38. Wei Hu YL, Wenrui Lei, Xiang, Yahua Chen, Luqing Zheng, Yan Xia, Zhenguo Shen. Cloning and characterization of the *Oryza sativa* wall-associated kinase gene *OsWAK11* and its transcriptional response to abiotic stresses. *Plant and soil*. 2014; 384:335–46. <https://doi.org/10.1007/s11104-014-2204-8>
39. Lee YS, Lee DY, Cho LH, An G. Rice miR172 induces flowering by suppressing *OsIDS1* and *SNB*, two AP2 genes that negatively regulate expression of *Ehd1* and florigens. *Rice (N Y)*. 2014; 7(1):31. Epub 2015/08/01. <https://doi.org/10.1186/s12284-014-0031-4> PMID: 26224560; PubMed Central PMCID: PMC4884018.
40. Zhu QH, Upadhyaya NM, Gubler F, Helliwell CA. Over-expression of miR172 causes loss of spikelet determinacy and floral organ abnormalities in rice (*Oryza sativa*). *BMC Plant Biol*. 2009; 9:149. Epub 2009/12/19. <https://doi.org/10.1186/1471-2229-9-149> PMID: 20017947; PubMed Central PMCID: PMC2803185.
41. Xu C, Liu Y, Li Y, Xu X, Li X, Xiao J, et al. Differential expression of *GS5* regulates grain size in rice. *Journal of experimental botany*. 2015; 66(9):2611–23. Epub 2015/02/26. <https://doi.org/10.1093/jxb/erv058> PMID: 25711711.
42. Li P, Pan T, Wang H, Wei J, Chen M, Hu X, et al. Natural variation of *ZmHKT1* affects root morphology in maize at the seedling stage. *Planta*. 2018. Epub 2018/11/22. <https://doi.org/10.1007/s00425-018-3043-2> PMID: 30460404.
43. Choudhury BI, Khan ML, Dayanandan S. Patterns of nucleotide diversity and phenotypes of two domestication related genes (*OsC1* and *Wx*) in indigenous rice varieties in Northeast India. *BMC Genet*. 2014; 15:71. Epub 2014/06/18. <https://doi.org/10.1186/1471-2156-15-71> PMID: 24935343; PubMed Central PMCID: PMC4070345.
44. Ma X, Feng F, Wei H, Mei H, Xu K, Chen S, et al. Genome-Wide Association Study for Plant Height and Grain Yield in Rice under Contrasting Moisture Regimes. *Frontiers in plant science*. 2016; 7:1801. Epub 2016/12/15. <https://doi.org/10.3389/fpls.2016.01801> PMID: 27965699; PubMed Central PMCID: PMC5126757.

45. Lipka AE TF, Wang Q, Peiffer J, Li M, Bradbury PJ, Gore MA, Buckler ES, Zhang ZW. GAPIT: genome association and prediction integrated tool. *Bioinformatics*. 2012;23:97–9. <https://doi.org/10.1093/bioinformatics/bts444> PMID: 22796960
46. Tamura K, Dudley J, Nei M, Kumar S. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol*. 2007; 24(8):1596–9. Epub 2007/05/10. <https://doi.org/10.1093/molbev/msm092> PMID: 17488738.
47. Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*. 2009; 25(11):1451–2. Epub 2009/04/07. <https://doi.org/10.1093/bioinformatics/btp187> PMID: 19346325.
48. Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*. 1989; 123(3):585–95. Epub 1989/11/01. PMID: 2513255; PubMed Central PMCID: PMC1203831.
49. Fu YX, Li WH. Statistical tests of neutrality of mutations. *Genetics*. 1993; 133(3):693–709. Epub 1993/03/01. PMID: 8454210; PubMed Central PMCID: PMC1205353.
50. Lin YJ, Zhang Q. Optimising the tissue culture conditions for high efficiency transformation of indica rice. *Plant cell reports*. 2005; 23(8):540–7. Epub 2004/08/17. <https://doi.org/10.1007/s00299-004-0843-6> PMID: 15309499.