Short title: *GSW3* regulates grain size and weight

**Natural Allelic Variation in GRAIN SIZE AND WEIGHT** of

**Wild Rice Regulates the Grain Size and Weight**

Feng Bai† 1,2,3,4, Huijin Ma† 1,2,3,4, Yichang Cai 3, Muhammad Qasim Shahid1,2,3,4, Yuebin Zheng 3,5, Chuan Lang 3, Zhixiong Chen 1,2,3,4, Jinwen Wu1,2,3,4, Xiangdong Liu 1,2,3,4*, Lan Wang1,2,3,4*

1 State Key Laboratory for Conservation and Utilization of Subtropical Agro-Bioresources, College of Agriculture, South China Agricultural University, Guangzhou 510642, China;
2 Guangdong Laboratory for Lingnan Modern Agriculture, Guangzhou 510642, China;
3 College of Agriculture, South China Agricultural University, Guangzhou 510642, China
4 Guangdong Provincial Key Laboratory of Plant Molecular Breeding, College of Agriculture, South China Agricultural University, Guangzhou 510642, China;
5 State Key Laboratory of Vegetation and Environmental Change, Institute of Botany, Chinese Academy of Sciences, Beijing 100093, China.

*Corresponding authors: wanglan@scau.edu.cn; xdliu@scau.edu.cn
†Contributed equally

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (https://academic.oup.com/plphys/pages/General-Instructions) is: Lan Wang (wanglan@scau.edu.cn).
Abstract

Grain size is important for yield in rice (*Oryza sativa* L.). Although many genes involved in grain size have been isolated, few can be used in breeding due to their interactions and phenotypic effects. Here, we describe natural variation in the granule-type quantitative trait locus *GRAIN SIZE AND WEIGHT 3 (GSW3)* located on chromosome 3 in wild rice (*Oryza rufipogon* Griff.) that encodes a GTPase-regulated protein that negatively regulates grain length, grain width, and 1000-grain weight. The insertion of a 232-bp fragment of the genomic sequence in the wild rice, a natural allelic variant gene (*GSW3*), increased expression levels and reduced grain length and width and 1000-grain weight. Knockout of *GSW3* in the wild rice inbred line Huaye 3 increased grain length and width and 1000-grain weight. Introducing *GSW3*Huaye3 into cultivated rice line KJ01 and overexpressing *GSW3*Huaye3 in Huaye 3 resulted in reduced grain length and width and 1000-grain weight, and grain size and 1000-grain weight changes were closely related to *GSW3* expression levels. *GSW3* regulated grain length and width simultaneously by promoting grain glume cell division and longitudinal and transverse cell growth. *GSW3* was also involved in regulating the gibberellic acid signaling pathway and negatively regulated plant growth. Furthermore, a critical SNP in the *GSW3* coding region was obviously correlated with grain size variation in a core collection of cultivated rice. This SNP resulted in an amino acid substitution from Gln to Arg at position 161 in GSW3, which reduced grain size. Our study shows that *GSW3* negatively regulates grain shape, which could explain different grain shapes in modern cultivars and wild rice. *GSW3* may also be used for breeding rice varieties with improved grain shapes and higher yield.
Keywords: Rice; Natural variation; GSW3; Grain size; Molecular function

Introduction

Rice (Oryza sativa L.) is a staple food for at least half of the global population. Rice yield is mainly determined by grain shape and grain filling rate. Therefore, grain shape is a direct factor affecting rice yield. Although many quantitative trait loci (QTLs) related to grain size have been localized and dozens of QTLs have been isolated, few genes can be applied to production, and the molecular regulatory network remains unclear. Therefore, exploring more grain size-related genes and elucidating their functional mechanisms are of great importance for further improving rice yield and quality. To date, five major pathways have been reported to regulate rice grain size, including the ubiquitin-proteasome pathway, the G protein signaling pathway, the mitogen-activated protein kinase (MAPK) signaling pathway, the phytohormone signaling pathway, and the transcription regulatory factor pathway (Li et al., 2018). In addition, some other regulatory pathways, such as the miR167a-AUXIN RESPONSE FACTOR 6 (OsARF6)-AUXIN RESISTANT 3 (OsAUX3) module, regulate grain length and weight in rice (Qiao et al., 2021). OsARF6 acted as an upstream transcription factor regulating the expression of OsAUX3, and miR167a positively regulated grain length (GL) and grain width (GW) by directing OsARF6 mRNA silencing.

The genes that control grain length include GRAIN SIZE 3 (GS3) (Fan et al., 2006), GRAIN LENGTH 4 (GL4) (Wu et al., 2017), THOUSAND-GRAIN WEIGHT 3 (TGW3/GL3.3) (Ying et al., 2018; Xia et al., 2018), GRAIN LENGTH 7 (GL7) (Wang Y X et al., 2015) and GRAIN SHAPE 9 (GS9) (Zhao et al., 2018). Those that control grain width include GRAIN WIDTH 2 (GW2) (Song et al., 2007), GRAIN WIDTH 5 (GW5) (Weng et al., 2008), and GRAIN WIDTH 8 (GW8) (Wang et al., 2012). Those that control 1000-grain weight include GRAIN WIDTH 6a...
GW6α (Song et al., 2015) and THOUSAND-GRAIN WEIGHT 6 (TGW6) (Ishimaru et al., 2013).

GS3 is a major QTL cloned in rice that controls both grain length and weight (Fan et al., 2006; Fan et al., 2009). A single-nucleotide mutation in the GL4 gene resulted in smaller and less shattered seeds in African rice domestication, affecting 1000-grain weight and yield per plant (Wu et al., 2017). TGW3/GL3.3 is a negative regulator of grain length and weight that synergistically alters the size and number of cells in glumes (Ying et al., 2018) and has a genetically superior reciprocal effect with GS3, which results in larger seeds when combined with this gene (Xia et al., 2018).

GL7 can simultaneously regulate rice grain length and improve rice quality through tandem repeat copy number variation (Wang Y X et al., 2015). GW2 and GW5 regulate grain width and weight through the ubiquitin-proteasome pathway (Song et al., 2007; Weng et al., 2008). GW7 regulates cell division to produce longer grains; while GW8 simultaneously controls grain size, and rice quality (Wang et al., 2012), and GW8 can directly bind to the GW7 promoter and inhibit its expression (Wang S K et al., 2015).

In rice, G proteins have been shown to regulate various processes and traits, such as seed size, crop yield improvement, promotion of cell division and expansion, and response to plant hormones, which are important for plant growth and development. The rice genome contains only seven G proteins, namely, RICE G-PROTEIN α SUBUNIT 1 (RGA1), RICE G-PROTEIN β SUBUNIT 1 (RGB1), RICE G-PROTEIN γ SUBUNIT 1 (RGG1), RICE G-PROTEIN γ SUBUNIT 2 (RGG2), GS3, QTL PANICLE ERECTNESS 9-1/DENSE AND ERECT PANICLE 1(qPE9-1/DEP1) and G-PROTEIN γ SUBUNIT CYS-RICH 2 ON C-TERMINAL (GGC2) (Sun et al., 2018), all of which have been reported to regulate grain shape. D1/RGA1 regulates plant height and grain shape through the gibberellin and brassinolide pathways (Ashikari et al., 1999; Wang et al., 2006). RGG1 is involved in the cytokinin regulatory pathway that controls rice grain size (Tao et al., 2020). RGG2
negatively regulates the granulotype through the GA pathway (Miao et al., 2019). In addition, \textit{DEP1} and \textit{GGC2} are positive regulators of grain shape in rice and can interact with \textit{RGB1} to form dimers, inhibiting crop growth and leading to shorter grains. \textit{GS3} can compete with \textit{DEP1} and \textit{GGC2} for binding to \textit{RGB1}, resulting in shorter grains, while \textit{GS3} itself has no role in regulating grain size (Sun et al., 2018). In addition, three atypical, extra-large GTP-binding (XLG) protein subunits regulate panicle architecture, plant growth, development, grain weight, and disease resistance (Biswal et al., 2022). These findings collectively support the important role of G proteins in regulating grain size in rice.

In our previous research, the wild rice (\textit{Oryza rufipogon} Griff.) inbred line Huaye 3 with short grains of 7.19 mm and cultivated rice (\textit{Oryza sativa} L.) line KJ01 with long grains of 13.91 mm were used to construct a segregating population. A QTL was localized between the flanking molecular markers PSM379 and RID24455 on chromosome 3 through QTL linkage analysis (Zheng et al., 2020). Then, the region was narrowed using bulked segregant analysis (BSA) to an interval of 340 kb, which was colocalized to 37 trait-associated genes, and finally mapped to \textit{qGL3.5} by 2786 BC\textsubscript{4}F\textsubscript{2} individuals (Wang et al., 2022). Interestingly, 37 trait-associated genes were identified with BSA-seq, including \textit{ORF25}, which encodes a GTPase regulatory protein described in the Rice Annotation Project. Here, we reported the characterization and biological function of \textit{ORF25}, a natural allelic variant in wild rice, hereafter referred to as \textit{GRAIN SIZE AND WEIGHT 3} (\textit{GSW3}) which regulates rice seed size and weight. The allele of \textit{GSW3} in the wild rice inbred line Huaye 3, characterized by a 232 bp insertion, increases its transcript level. As a result, plants containing such alleles exhibit fewer, shorter and narrower cells in the glumes, which ultimately shorten the seeds. This study broadens our understanding of the role of genetic factors and their concomitant mechanisms in determining rice grain shape, thus providing a resource for the targeted breeding of specific varieties.
Results

Phylogenetic analysis of G protein in rice and genomic sequence analysis of GSW3 between the parents

G proteins, such as RGA1, RGB1, RGG1, RGG2, GS3, qPE9-1/DEP1 and GGC2, have been reported to regulate grain shape in rice (Sun et al., 2018). GSW3 encodes a GTPase regulatory protein but has no typical G-protein domain. An unrooted phylogenetic tree was constructed from the full-length protein sequence of GSW3 and the seven G proteins with MEGA7 (Figure 1A). The eight genes were classified into two distinct groups, and GSW3 and RGA1 were grouped into the same clade. GSW3 was speculated to have the same function as RGA1. In addition, according to linkage analysis performed in our previous work, the contribution of the interval PSM379 to RID24455 to the phenotype reached 54.85 % (Zheng et al., 2020), and we inferred that more than one gene-regulated grain length in the identified region. Therefore, GSW3 was subjected to functional studies using the CRISPR/Cas9 technique.

Comparing the genomic coding sequences of GSW3 between the parents Huaye 3 and KJ01, a total of 12 SNPs were generated; among them, seven differences in SNPs led to amino acid substitution (Table S1). More importantly, there is a 232 bp insertion at position 289 of GSW3 in Huaye 3, which directly causes the premature termination of amino acid translation (Figure 1B). The promoter region of GSW3 between Huaye3 and KJ01 also produced 20 SNPs in the 1000 bp range upstream of the initiation codon (Table S2).

Expression patterns of GSW3

To explore the spatiotemporal expression difference of GSW3 between KJ01 and Huaye 3, RNA was extracted from the roots, stems, leaves and panicles of Huaye 3 and KJ01 in the same period, and the expression of GSW3 was detected by RT-qPCR. The results showed that GSW3 was a constitutively expressed gene in all tissues, including roots, stems, leaves, and panicles, of Huaye 3 and KJ01. The expression level of GSW3 in each tissue of Huaye 3 was higher than that in KJ01 and was high in
the root and panicle (Figure 1C). With the development of young spikelets, the GSW3 expression level in Huaye 3 spikelets gradually decreased (Figure 1D). During the developing of Huaye 3 grains, the mRNA expression level of GSW3 gradually increased from 3 to 7 days after flowering (DAF), decreased at 10 DAF, and then gradually increased from 15 to 25 DAF (Figure 1E). To further characterize GSW3 expression patterns, the activity of the GSW3 promoter was detected using β-galactosidase (GUS) in the Nipponbare background. We detected GUS activity in roots, stems, leaves, and panicles, which was consistent with our RT-qPCR results (Figure 1F). Younger spikelets showed higher GUS activity than older spikelets.

**GSW3 encodes a nuclear protein**

To explore the localization of the GSW3 protein, a pBin19-eGFP-GSW3 vector was constructed, and the localization of the GSW3 protein was observed by infiltrating the *Nicotiana benthamiana* leaf epidermis. The results of transient expression in epidermal cells showed that GSW3 was localized in the cell nucleus (Figure 2A). In addition, we also constructed pOX-GFP-GSW3 and PAN580-mKATE fusions driven by the 35S promoter. The two fusions were transferred together into rice protoplasts, and we observed that the pOX-GFP-GSW3 fusion protein was localized in the nucleus of rice protoplasts (Figure 2B). These results were similar to those of the experiment involving *Nicotiana benthamiana* leaf epidermis infestation.

**GSW3 negatively regulates grain size**

To study the molecular function of GSW3, we first knocked out GSW3 in the Huaye 3 background using CRISPR/Cas9 gene-editing technology. A total of 19 transformed plants were obtained, among which five plants (named #1, #2, #3, #4 and #5) generated phenotypic variation including grain size, spike shape, and plant height. Then, the five phenotypically variable plants were grown into T1 lines, each line containing 50 individuals. In the T1 lines, these variant traits remained stably inherited. Compared to wild-type Huaye 3, T1 lines of the five phenotypically variable
plants (#1~#5) showed significant increases in plant height, panicle length, grain length and grain width (Table S3). Next, the two target sites of GSW3 from the five phenotypically variable plants were sequenced and checked for missing targets. Surprisingly, lines #3, #4, and #5 appeared to be missing the targets. GSW3 exhibited five editing types in the two target regions of lines #1 and #2, and the five editing types were named KO-1, KO-2, KO-3, KO-4, and KO-5 (Figure 3A). These editing types were all base insertion or deletion mutants and resulted in frameshift mutations of amino acids. Grain length and width of KO mutants significantly increased in different editing types relative to wild-type Huaye 3 (Figure 3B, 3C). In the T2 generation of the transgenic lines (50 individuals of each line), two mutant types (Figure 3D) with T-DNA free and no missing target of KO-1, with a 1-nt (G) insertion in the first target site and 2-nt (TT) deletion in the second target site, and KO-2, with a 1-nt (G) insertion in the first target site and 1-nt (T) deletion in the second target site, were further analyzed for economic characteristics. Compared with Huaye 3, the grain length of the KO-1 and KO-2 lines increased by 20.16% and 14.05%, grain width by 6.7% and 4.6%, and 1000-grain weight by 69.01% and 52.64%, respectively (Table 1, Figure 3E). These directly increased the yield of each plant of KO-1 and KO-2 by 167.67% and 38.07%, respectively (Table 1, Figure 3F). Compared with wild-type Huaye 3, the mRNA expression levels of GSW3 in KO-1 and KO-2 lines were greatly decreased (Figure 3G).

Then, GSW3 derived from Huaye 3 was transformed into KJ01 to generate a functional complementary transgenic line. A total of twenty-one transgenic plants were obtained, of which the grain length and grain width of four plants showed significant differences (Table S4), and the grain length became shorter than that of wild-type KJ01. Two functional complementary transgenic plants (CF-1 and CF-15) with significantly different phenotypes were grown into T1 lines and T2 lines, with each line including 50 individuals. The various phenotypes remained stably inherited.
in the homozygous mutants with $GW3^{Huaye^3}$, mainly showing reduced grain length and grain width (Figure 3E) and slightly shortened plant height (Figure 3H). The mRNA expression levels of $GW3$ in the CF-1 and CF-15 lines also substantially increased compared with those in wild-type KJ 01 (Figure 3H).

To further verify the functions of $GW3$ in rice seed development, $GW3^{Huaye^3}$ was overexpressed in the Huaye 3 background. Four of sixteen transgenic plants showed highly significant grain length and width reductions compared with wild-type Huaye 3 (Table S5). Then, two variant plants (OE-6 and OE-11) were further analyzed. In the T$_1$ generation (each line included 50 individuals), the variant traits remained stably inherited (Figure 3E), and the mRNA expression levels of $GW3$ in the two OE mutants increased relative to wild-type Huaye 3 (Figure 3E).

These data suggested that $GW3$ is a negative regulator of grain size.

**$GW3$ regulates the cell division and cell expansion of spikelet hulls**

The grain size of rice is mainly limited by the spikelet hulls, which are regulated by cell proliferation and cell expansion of grains (Qi et al., 2012). To explain the reason for grain size variation in the $GW3$ KO mutants from a cytological perspective, a transverse section at the central part of the spikelet in the $GW3$ KO-2 mutant was microscopically observed (Figure 4A). The average single-cell area of the KO-2 mutant was significantly increased, and the cell numbers of the transverse section were lower than those in Huaye 3 (Figure 4B).

Furthermore, the outer epidermal cells of the mature rice glumes of the Huaye 3, KO (KO-1 and KO-2) and OE (OE-6 and OE-15) mutant lines were observed by scanning electron microscopy (SEM) (Figure 4C). SEM analysis showed that the length, numbers in the longitudinal direction, width and numbers in the transverse direction of outer epidermal cells showed significant changes compared with those in Huaye 3. In the $GW3$ KO mutant lines, the outer epidermal cells became longer and narrower, and the outer epidermal cell numbers decreased along the longitudinal axis.
and increased along the transverse axis (Figure 4D). In the GSW3 OE mutant lines, the outer epidermal cells became shorter and narrower, and the outer epidermal cell numbers increased along the longitudinal and transverse axes (Figure 4D). These results were consistent with those of the plastic semithin sections. In conclusion, GSW3 regulates seed grain size by affecting both cell expansion and cell division.

Transcriptome analysis of GSW3
To further explore the regulatory mechanism by which GSW3 affects grain length, we performed transcriptome sequencing analysis (RNA-seq) of young spikelet hulls with different lengths from Huaye 3 and KO-2 (young spikelet hulls were 25% and 50% as long as the full-length spikelet hulls, respectively). There were 4720 differentially expressed genes (DEGs) in 25% of the full-length glume stage (Table S6). Among them, 2479 genes were significantly upregulated, and 2241 genes were significantly downregulated. There were only 573 DEGs in 50% of the full-length glume stage, of which 344 DEGs were upregulated and 229 DEGs were downregulated (Table S7). A total of 409 common genes were differentially expressed in the two periods. With the development of grain glumes, the number of DEGs gradually decreased.

The quantity of transcriptome data was adequate for the necessary level of quality for the following experimental analysis (Figure 5A, B). Because there were more DEGs in the early stage of glumes than in later stages, we further analyzed the DEGs for 25% full-length glumes. The gene ontology (GO) terms for these DEGs were significantly enriched in molecular transport activities, signal transduction activities, transcription factor activities, protein binding and other processes (Figure 5C). Furthermore, we also analyzed the kyoto encyclopedia of genes and genomes (KEGG) pathways of DEGs, indicating that many genes were enriched in plant-pathogen interaction, plant hormone signal transduction, starch and sucrose metabolism (Figure 5D). These activities were closely related to the growth of rice glumes and endosperm development. Further analysis of DEGs revealed several genes...
reported to regulate grain shape and the grain filling, such as RGG2, OsMADS1, SDT, OsGS1, GS5, and OsRac1, and RT-qPCR analysis verified the differential expression of these genes (Figure 5E). Overall, these results all indicated that GSW3 could regulate the grain shape of rice by regulating the transcription levels of genes involved in grain shape and the grain filling rate during the growth and endosperm development of rice.

**GSW3 functions as a negative regulator in the gibberellin pathway**

Multiple phytohormones were reported to be involved in controlling seed development and plant growth (Cao et al., 2020; Su et al., 2021; Tian et al., 2022). In the present study, compared to the wild-type Huaye 3, KO mutants showed increased plant height. To investigate whether GSW3 is involved in the gibberellin acid (GA) signaling pathway, we analyzed the difference in GA sensitivity between Huaye 3, GSW3 KO-1 and KO-2 mutants by measuring the length of the second leaf sheath after GA3 treatment (Figure 6A). Before GA3 treatment, seeds were soaked in 5 μM paclobutrazol for 2 d to eliminate the background. Under 0.01 μM and 1 μM exogenous GA3 treatment, the second leaf sheath of GSW3 KO-1 and KO-2 mutant plants was significantly elongated compared with that of Huaye 3, but the average growth rate was the highest under 0.01 μM (Figure 6B, C), indicating that GSW3 could negatively regulate the response to GA and that 0.01 μM was the most suitable treatment concentration. Then, the expression levels of genes involved in GA biosynthesis and signal transduction in the GSW3 KO-2 mutant line were examined by RT-qPCR under treatment with 0.01 μM exogenous GA3. The results showed significant upregulation of the GA positive synthesis gene, GA3ox2. In contrast, negative GA synthesis pathway genes, including GA20ox1, GA20ox2, GA20ox3, GA2ox1, GA2ox3, KO2, and KAO, were significantly downregulated, and the GA signal transduction genes, GID2 and SLR1, were also significantly downregulated (Figure 6D). In conclusion, this information suggests that GSW3 mediates internal
GA biosynthesis and participates in the GA signaling pathway to negatively regulate plant growth.

Physically interacting proteins screened by yeast two-hybrid

From the RNA-seq data, seventeen upregulated and downregulated genes, related with grain size, grain filling, and GA signaling pathway, such as GS3, OsMADS1, GS5, OsRac1, and so on, were constructed BD vector. Their protein interactions with GSW3-AD were tested by yeast two-hybrid with point-to-point experiment. However, regrettfully, no interacting protein was found. And then, two yeast libraries constructed using the whole development stage and seedings grown for 20 days were screened to search the interacting protein. In the preliminary screening, more than two hundred genes were found. Then further screened on medium added 3-AT (3-amino-1,2,4-triazole) and PCR identified, 44 genes were selected out (Figure S1). Finally, five genes were screened out by sequencing, respectively for Os10g0577700, Os03g0265600, Os04g0445200, Os01g0282800, and Os05g0108800 (Table S8). Among them, Os03g0265600 and Os01g0282800 were respectively upregulated and downregulated in the RNA-seq data for 50% of the full-length glume stage. Os01g0282800 (OsTUB1) encoded a tubulin family protein with GTPase domain. In the previous report, OsTUB1 involved into grain development. OsTUB1 was overexpressed in the Nipponbare background and increased grain size (Chen et al., 2022).

SNPs representing natural variation in the GSW3 coding region affect grain length

In the coding region of GSW3, a total of 12 SNPs was generated between KJ01 and Huaye3. Among them, seven SNPs led to amino acid substitutions (Table S1). To investigate the possible effect of SNPs in the coding region of GSW3 on grain size and weight, the sequences of 1025 accessions of cultivated rice from the 2013 Rice Genome Project were analyzed (https://www. rmbreeding. cn. Index/). Using the Nipponbare sequence as the reference genome, seven nonsynonymous SNPs were
identified in the GSW3 coding region. Using these SNPs, the accessions could be divided into five haplotypes. Four haplotypes (Hap1, Hap2, Hap3, and Hap4) were found in indica accessions, and only one was found in the other accessions; specifically, Hap1 in japonica accessions, Admix and Bas, and Hap5 in Aus; however, Hap1 with 648 Japonica accessions was far higher than that of other haplotypes (Figure 7A). The variable bases of GSW3 in KJ01 were classified as Hap1, and those in Huaye 3 were classified as Hap2. Compared with Hap2, only one SNP differed in the open reading frame of GSW3 caused significant variation in grain size and weight of Hap3, Hap4, and Hap5. The SNP of Hap3, Hap4, and Hap5 caused significant variation in grain length and weight, and the SNP of Hap3 and Hap5 caused significant variation in the ratio of grain length to width. The SNP of Hap4 caused significant variation in grain width (Figure 7B). Degenerate base R (A/G) (+80) encoded Lys$^{26}$/Arg$^{26}$ in Hap4, while in Hap2, base C encoded Thr$^{26}$. Base G (+161) encoded Arg$^{53}$ in Hap5, while in Hap2, base A encoded Gln$^{53}$. Degenerate base Y (C/T) (+373) encoded Ala$^{125}$/Val$^{125}$ in Hap3, while base G encoded Gly$^{125}$ in Hap2.

Accessions with Hap4 had increased grain length and width, and those with Hap5 had apparently shortened grain length. Compared with Hap2 (Huaye3), Hap5 showed a SNP difference, and the substitution base G (+161) was just the same base as that of GSW3-KJ01. This SNP results in an amino acid substitution from Gln to Arg at position 161 of GSW3 that decreased the GTPase activity of GSW3 and thus reduced rice grain length. These findings suggested that grain size and 1000-grain weight regulation between KJ01 and Huaye3 is influenced by the nonsynonymous SNP at position 161 of GSW3.

Discussion

The allele of GSW3 derived from a wild rice inbred line negatively regulates grain size, and enhances rice yield.
A major challenge for rice breeders is finding stable and sustainable ways to increase yield potential and improve rice quality; grain size breeding has been a research hotspot in recent years. In recent decades, large numbers of genes and their functional roles in regulating of grain size have been identified and characterized (Ikeda et al., 2013; Jiang et al., 2022). However, most of them were derived from cultivated rice. 

GS3 was developed by the cross of two indica cultivars, Minghui 63 and Chuan 7 (Fan et al., 2006). GS3 was found to contain four putative domains with an organ size regulation (OSR) domain and to negatively regulate grain size (Mao et al., 2010). GS2 was initially mapped from a BC\textsubscript{4}F\textsubscript{2} population of landraces Baodali and Zhonghua 11 (a japonica cultivar). GS2 encodes growth-regulating factor 4 (OsGF4) and regulates grain size by promoting cell division and expansion (Hu et al., 2015). SMG3 was identified from the cross of japonica cultivar M494 and indica cultivar Zhong 9 B. SMG3 encoded a MYB-like protein and coordinately regulated grain size, grain number per panicle, and grain weight (Li et al., 2022). With advances in gene function research, the genetic resources derived from cultivated rice have become increasingly narrow. This study identified a natural allele, GSW3, derived from a wild rice inbred line. GSW3 negatively regulated several agronomic traits, including grain length and width and 1000-grain weight. By controlling cell division and spikelet hull cell expansion, GSW3 encodes a GTPase regulatory protein that controls rice size.

Although many genes controlling grain size have been isolated, only a few have been applied in rice production due to their phenotypic defects. For example, GS2 caused a simultaneous increase in grain length, width and 1000-grain weight, but this resulted in poor grain appearance with a significant increase in grain chalkiness (Hu et al., 2015). GL7 increased grain length but decreased grain width without changing grain weight (Wang Y X et al., 2015). In our study, GSW3 was located between molecular markers PSM379 and RID24455, with close linkage with qGL3.5, and this region contributed 54.85\% to the phenotypic variation in the F\textsubscript{2} population (Zheng et al., 2020). Using CRISPR/CAS9 gene editing technology to target mutations of
In rice, two homozygous mutants were obtained in the Huaye 3 background, which produced significantly longer and wider grains (Table 1). Compared with wild-type Huaye 3, the grain lengths of the KO-1 and KO-2 lines increased by 20.16% and 14.05%, and the grain widths increased by 6.7% and 4.6%, respectively. In addition, the two mutants also had an exponential increase in the number of secondary branching peduncles, which significantly increased rice yield per plant. According to these results, the GSW3 allele has large utilization potential in rice production.

G protein signaling has been reported to regulate grain size and grain filling in rice (Xu et al., 2019; Zhang et al., 2019). The rice genome encodes one Ga (RGA1), one Gβ (RGB1), and five Gγs (RGG1, RGG2, GS3, DEP1 and GGC2) (Ishiawa et al., 1995; Kato et al., 2004). Miao et al. (2019) reported that RGG2 encoded a type B heterotrimeric G protein γ subunit and increased grain size. Two knockout mutants, zrgg2-1 and zrgg2-2, in the Zhenshan 97 background enhanced grain yield per plant by 11.8% and 16.0%, respectively. Three atypical extra-large GTP-binding protein (XLG) subunits regulate panicle architecture and grain weight (Biswal et al., 2022). The three XLG proteins work together to control seed filling; the loss of one XLG reduced grain weight by 14%, and the loss of all three increased grain weight by 32%.

Here, GSW3 encoded a GTPase regulatory protein without a typical G protein domain. Between the parents of Huaye 3 and KJ01, a 232-bp large fragment was inserted into the GSW3 genome of Huaye 3, which directly led to the premature termination of amino acid translation. In addition, seven non-synonymous SNPs in the GSW3 coding region between the two parents resulted in amino acid substitutions. Haplotype analysis of the seven nonsynonymous SNPs revealed that base A to G (+161) in the GSW3 coding region caused significant variation in grain size and weight, and it was the critical SNP affecting grain size. In the Huaye3 background, the GSW3 mutant lines, KO-1 and KO-2, appeared to have increased grain length and width, directly resulting in 1000-grain weight increases of 69.01% and 52.64%, respectively. Yield evaluation of each plant indicated that KO-1 and KO-2 increased...
grain yield by 167.67% and 38.07%, respectively. Therefore, the allele of GSW3 can
be used to breed rice varieties with higher yields and optimized grain shapes.

RNA-seq and RT-qPCR analysis revealed that some genes regulating grain shape
and the grain filling rate, such as OsMADS1, GS5, and OsRac1, were downregulated
by GSW3; while RGG2, SDT, and OsGS1 were upregulated by GSW3. Five genes
(Os10g0577700, Os03g0265600, Os04g0445200, Os01g0282800, and
Os05g0108800) were searched to interact with GSW3 using yeast two-hybrid
screening. These interacting proteins need to be further confirmed by BiFC and Pull-
down analysis. It is worth further studying how these genes interact to regulate grain
development.

**GSW3 regulated rice size at the mRNA expression level**

Gene molecular function was carried out by mRNA expression level. Some gene
expression negatively regulated the phenotype, and some positively regulated the
phenotype. DEP1 overexpression produced the expected longer grains in
daohuaxiang2; nevertheless, overexpressing the predicted OSR domains of AGG3
and ZmGS3 resulted in reduced grain length (Sun et al., 2018). The NIL-GL7
expression level in young panicles was higher than that in NPB, and GL7s1 was
transformed into NPB, which resulted in an ~10% longer grain length in the
transgenic plants (Wang et al., 2015). Mehra et al. (2022) reported that overexpression
of OsJAZ11 enhanced seed width and weight in rice. In our study, GSW3 regulated
grain size at the mRNA expression level. The expression level of GSW3 increased
dramatically when a 232-bp-long fragment was inserted into Huaye 3, and its grains
became shorter and narrower than those of KJ01. When GSW3Huaye3 was knocked out,
the grains became longer and wider, and the mRNA expression level decreased
dramatically again (Figure 3E, 3G). GSW3Huaye3 was induced in KJ01, and then the
expression level increased, which resulted in its grains becoming shorter and narrower
(Figure 3E, 3F). In the background of KJ01, GSW3Huaye3 can complete the function
of GSW3KJ01 by regulating the mRNA expression level.
GSW3 is involved in the GA signaling pathway in rice

Recently, many genes were reported to regulate grain size by participating in multiple phytohormone signaling pathways, such as GA, Brassinosteroid (BR), and auxin (Gao et al., 2018, 2021; Duan et al., 2020; Jin et al., 2022). *GNP1* encodes GA 20-oxidase 1, which promotes cytokinin activity during the dynamic regulation of the KNOX protein, increasing grain number and yield in rice (Wu et al., 2016). *SD1* encodes GA 20-oxidase 2 and regulates panicle development through DELLA-KNOX signaling, and *GNP1* and *SD1* have nonredundant functions in panicle development (Su et al., 2021). *OsGASR9* is a positive GA response regulator that controls rice grain size and yield (Li et al., 2019). *OsGF14b* acts as a negative regulator of BR signaling and inhibits *OsBZR1*’s transcriptional activation activity. *qGL3* causes *OsGF14b* to become phosphorylated, which modulates *OsBZR1*’s nucleocytoplasmic shuttling and transcriptional activation activity. This negatively regulates rice grain length and BR signaling (Gao et al., 2021). *GLW7.1* could increase endogenous GA content by upregulating the expression of GA biosynthesis genes (Liu et al., 2022). Rice G proteins have also been shown to be involved in several phytohormone signaling pathways (Wang et al., 2006; Hu et al., 2013; Zhang et al., 2015). G-protein and the hormone signaling pathway regulate panicle development and affect kernel size (Liang et al, 2022). Studies in our laboratory have shown that GSW3 regulates rice plant height and seed size and is involved in phytohormones. KEGG enrichment indicated that many DEGs were significantly enriched in plant hormone signal transduction pathways.

The *dwarf 1/RGA1 (D1)* gene encodes a GTP-binding protein and regulates plant growth and development by associating with gibberellin signal transduction (Ashikari et al., 1999). Unrooted phylogenetic tree analysis revealed that GSW3 and RGA1 were grouped in the same clade (*Figure 1A*). GA is one of the most important hormones affecting rice plant height and seed size. By comparing the second leaf sheath length of Huaye 3 and GSW3 KO mutants under exogenous GA3 treatment, we discovered
that the second leaf sheath length increased in the mutants. Additionally, the expression levels of genes related to GA synthesis and signal transduction changed significantly in the mutants, which suggests that GSW3 probably affects rice plant height and seed size through the GA pathway. However, its molecular mechanism needs to be further investigated.

**Conclusion**

**GSW3**, a gene closely linked with qGL3.5, was described to regulate grain length, grain width, and grain weight in rice. GSW3 was knocked out by CRISPR/Cas9 gene-editing technology in the Huaye3 background, and two homozygous mutants were obtained. The grain length and width and 1000-grain weight were all significantly increased in the GSW3 knockout mutants. When GSW3 derived from Huaye3 was induced into KJ01 or overexpressed in the background of Huaye 3, the grains dramatically became narrower and shorter. GSW3 negatively regulated grain size by cell expansion and division. GSW3 is a constitutively expressed gene located in the cell nucleus. GSW3 had a high expression level in the young panicle, which gradually decreased with young spikelet development. GSW3 is involved in the GA signaling pathway and negatively regulates plant growth.

**Materials and Methods**

**Plant material**

The wild rice (*Oryza rufipogon* Griff.) inbred line, Huaye 3, was bred by our research group (Yu et al., 2018); the cultivar (*Oryza sativa* L.) KJ01 was provided by Professor Kangjing Liang at Fujian Agriculture and Forestry University, Nipponbare, as were all other transgenic lines used for this experimental study. These materials were planted in the field at South China Agricultural University from 2019 to 2022. Each trial plot was planted in 10 rows spaced 18 cm × 21 cm apart. All the T₁ and T₂ lines
were planted with 50 individuals, and their agronomic traits were statistically analyzed.

The grains were labeled at the time of flowering, collected 3, 5, 7, 10, 15, 20, and 25 days after flowering (DAF) and stored at -80 °C in a freezer.

**Agronomic trait analysis**

Rice grains were harvested at the mature stage by collecting three major panicles of each plant, after which the materials were put into labeled envelopes. Samples of 30 randomly selected filled grains of each plant were assessed for grain length and grain width. The length of 10 grains placed end to end was measured using a ruler with a precision of ±0.01 mm, and the width of 10 seeds placed shoulder to shoulder was measured using a ruler with a precision of ±0.01 mm. The grains of each plant were measured three times, and the average length and width of a single rice grain were then calculated. The grain weight was determined by weighing 200 filled dry grains using an electronic analytical balance and finally converted into the weight of 1000 seeds. The 200-grain weight of each plot was measured for three times, and the average weight was used as the 200-grain weight, and then was changed into 1000-grain grain weight. Panicle length was calculated from the tip of each panicle to the node of the panicle neck, and the number of primary branching peduncles was counted for the branches growing on the panicle axis. The number of secondary branching peduncles was counted for the branches growing on the primary peduncles.

Statistical analyses of the data using sample t-tests were performed using Prism 8 software.

**Phylogenetic analysis**

Full-length G proteins used to construct phylogenetic trees were selected with the NCBI program (https://www.ncbi.nlm.nih.gov). ClustalW was used to align the amino acid sequences of selected proteins. Unrooted phylogenetic trees were generated using the neighbor-joining method in MEGA software version 7.0 (Kumar et al., 2016). Bootstrap values were calculated from 1 000 replicates.
Vector construction and plant transformation

A CRISPR/Cas9 gene editing construct of $GW3$ was designed as previously described (Zeng et al., 2018). Two targets were designed for $GW3^{Huaye}$ via the CRISPR-GE website (http://skl.scau.edu.cn/). Multiple rounds of PCR were performed to construct the expression cassette vector by using primes Cas9-$GW3$-T1 and Cas9-$GW3$-T2 (Table S9). The target sequences were introduced into sgRNA expression cassettes by overlapping PCR. The purified PCR products (sgRNA expression cassettes) were inserted into a pYLCRISPR/Cas9 vector based on the Golden Gate system by using primes ps-GGL/gs-GG2 and ps-GG2/gs-GGR (Table S9). The ligated products harboring the sgRNA expression cassettes were directly used to transform Escherichia coli DH5a competent cells. The CRISPR/Cas9 constructs that were successfully generated were introduced into Agrobacterium tumefaciens strain EHA105 and then reduced into Huaye 3.

To generate the complementary function construct, a 2800-bp fragment (including the 2000-bp promoter sequence and 325-bp terminator sequence) was amplified by the primer CF- $GW3$ from Huaye 3 (Table S9) and then cloned into the plant binary vector pCAMBIA1300S to generate a complementary function construct vector. The successfully generated plasmid was introduced into the callus of KJ01 by Agrobacterium-mediated transformation.

To generate the overexpression construct, the full-length coding sequence (CDS) of $GW3$ was amplified by the primer OE-$GW3$ from Huaye 3 and cloned into the plant binary vector pCAMBIA1300S (Table S9). The successfully generated plasmid was introduced into the callus of Huaye 3 by Agrobacterium-mediated transformation.

To analyze the expression pattern of $GW3$, a 2000-bp promoter fragment was cloned by the primer GUS-$GW3$ from Huaye 3-$GW3$ and introduced into the pCAMBIA13001 vector (Table S9). The Agrobacterium-mediated method was used to transform the constructed plasmid into the rice variety Nipponbare.
To generate the vectors pBin19:eGFP-GSW3 and pOX:eGFP-GSW3 using the primers pBin19-GSW3 and pOX-GSW3, respectively (Table S9), a 444-bp fragment of GSW3 cDNA was cloned into the pBin19:eGFP and pOX:eGFP vectors, respectively.

**RNA extraction and RT-qPCR analysis**

Total RNA was extracted from different tissues of Huaye 3, KJ01, the transgenic lines, and Huaye3 panicles of different lengths and grains of different development stages using the TRIzol method (Yeasen). Total RNA was treated with DNase I (Servicebio), and then a reverse transcription kit (Yeasen Hifair™ II 1st Strand cDNA Synthesis Super Mix for qPCR (gDNA digester plus)) was used to perform the first-strand RT‒qPCR with GDSBio (SYBR Green qPCR Mix) and a Light Cycler 480 RT‒qPCR instrument system. The rice ACTIN gene was used as an internal control.

Three biological replicates and three technical replicates were included for each sample. The primer sequences used for RT-qPCR are listed in Table S9.

**GUS Assays**

GUS activity analysis of pGSW3::GUS rice transgenic lines was performed as described previously (Qi et al., 2012). The samples were stained in GUS staining buffer (Leagene Biotechnology, β-Galactosidase Reporter Gene Staining Kit) for 1-24 h at 37 °C. Blue gradually appeared, and the chlorophyll of the sample was removed with 70% (V/V) ethanol until the negative control turned white. The sample was stored in ethanol and observed under a stereoscope. The blue on the white background is the GUS expression site.

**Rice glume plastic semithin slices**

Spikelet shells were collected after 3 days of flowering from wild-type Huaye 3 and GSW3 KO-2 mutants and fixed in FAA fixative (5% formaldehyde, 6% glacial acetic acid, and 89% ethanol(70%); V/V ) for 48 h. Then, plastic semi-packets of rice glumes were prepared using a previously described method (Lu et al., 2020).

**Scanning electron microscopy observation of rice glumes**
Spikelet shells of the mature Huaye 3, GSW3 KO and OE mutants were selected and dried completely, and then the samples were processed and analyzed according to a previously described method (Wang et al., 2022).

**Subcellular localization**

*Agrobacterium* pBin19:eGFP-GSW3, pBin:eGFP and pBin19 were activated separately and cultured in YEP liquid medium containing the corresponding antibiotics (50 μg/mL kan^+^, 20 μg/mL rif^+^) until the OD_{600} value was approximately 0.6-1. The bacteria were collected by centrifugation at 4000 rpm for 10 min. The bacteria were resuspended in resuspension solution (10 mM MES-KOH pH=5.6, 1 mL; 10 mM MgCl_2, 1 mL; 100 μM acetosyringone, 10 μL; ddH_2O, added to 100 mL) to reach a concentration of OD_{600}=1. The pBin19:eGFP-GSW3, pBin:eGFP and pBin19 resuspensions were mixed at a volume ratio of 1:1:1 and incubated for 3-4 h at room temperature in the dark, and the mixture was aspirated using a 2 mL syringe. The suspension was injected into the lower epidermis of *Nicotiana benthamiana* leaves. After 1-2 days of incubation at room temperature, the injected *Nicotiana benthamiana* lower epidermis was removed and observed under Leica SPE laser scanning confocal microscope (Excitation spectrum: 488nm; Emission spectrum: 510nm).

The coding sequence (from ATG to TGA) with deletion of the termination codon TGA of GSW3 from Huaye 3 plants was fused to pOX-GFP. The fusion protein with the insertion was cotransfected into rice protoplasts with NLS-mKate as a marker according to a previously described method with minor modifications (Yoo et al., 2007). Transformation mixtures (10 μg target plasmid or 10 μg marker plasmid for cotransformation with 200 μl protoplasts in 250 μl PEG solution) were agitated gently. After 30 minutes at room temperature in the dark, protoplasts were harvested and washed. They were then centrifuged, resuspended and cultured in the dark at room temperature, generally for 16-24 h. Then, the fluorescence was observed under a

**RNA-seq analysis**

RNA-seq analysis was conducted with 25% and 50% full-length glumes of wild-type Huaye 3 and GSW3-KO-2 transgenic lines, respectively, with 3 biological replicates. All the samples were stored on dry ice and sent to BioMarker Biotechnology Company for RNA extraction and sequencing. Reads were aligned to the rice genome (MSU7.0) using TopHat. Differentially expressed genes (DEGs) were identified based on $|\log_2 \text{(fold change)}| > 2$ and FDR < 0.01. The Actin gene was used as an internal control, and the DEGs were validated by RT-qPCR, as described previously (Wu et al., 2015).

**Gibberellin treatment**

Seeds of uniform size from wild-type Huaye 3 and GSW3 KO mutants were selected, treated with 75% alcohol (V/V) for 1-2 min and rinsed twice with sterilized water. The seeds were then soaked for 2 d with 10 μM paclobutrazol (PAC) to eliminate the background, and in sterile distilled water for an additional 2 d after washing off the paclobutrazol. And those with consistent germination were selected after seed exposure and incubated in aqueous solutions of GA$_3$ at different concentrations (0 μM, 10$^{-2}$ μM, 10$^{-4}$ μM, 10$^{-6}$ μM, and 1 μM). After 15 d of incubation, the lengths of the second leaf sheaths of wild-type and mutant seedlings were determined. The method was based on Duan et al. (2020) with some modifications.

**Yeast two-hybrid screening**

The full-length coding region of Huaye3-GSW3 was cloned into the pGBK7 vector, containing a binding domain (BD), and then induced into Y2H Gold. The two yeast libraries, containing an activating domain (AD), were constructed using the whole development stage and seedings grown for 20 days. Yeast Two-Hybrid System User ManualInteracting partners were isolated using the BD Matchmaker Library Construction and Screening Kit, according to the manufacturer’s instructions. The AD
plasmids of yeast library were combined with Y2H Gold cells carrying BD-GSW3 in 2× YPDA liquid medium using yeast mating method followed by incubation at 30°C for 1.5 h. The mating culture was plated on SD/-Leu/-Trp agar plates for 2-3 d, to test transformation efficiency. And all the colonies were patched out onto SD/-Leu/-Trp/-His/-Ade agar plates for 5-7 d, to screen the interacting protein. The second screening was conducted on SD/-Leu-Trp-His-Ade medium added 3-AT (3-amino-1,2,4-triazole). Finally, the AD-Prey plasmids were rescued from yeast strain and sequenced. The method was based on Kong et al. (2019) with some modifications.

**Haplotype analysis across the GSW3 coding region**

The haplotypes were constructed from seven nonsynonymous SNPs identified in the coding region of GSW3 from 1025 accessions of *O. sativa* (https://www.rmbreeding.cn. Index/). Boxplots for grain length, width, ratio of length to width, and weight were generated using data from indica and japonica rice cultivars of Asian origin (1025 accessions from the 2013 Rice Genome Project). The haplotypes of more than three accessions were selected, and seven variation sites were analyzed (Wang et al., 2020). The violin plot was mapped in Graphpad.

**Statistical analyses**

Differences between samples were analyzed using the two-tailed, unpaired Student *t* test available in the GraphPad Prism 5 software package. *P* values of less than 0.05 (indicated by asterisks) were considered statistically significant. The following categories were used: ***, *P* < 0.01; *, *P* < 0.05. The traits analysis of the violin plot was determined by Tukey’s honestly significant difference test in the GraphPad Prism 5, and different letters indicate significant differences between groups (*p*<0.01)

**Accession Numbers**

Sequence data from this article can be found in the GenBank/EMBL data libraries under accession numbers (Table S10).
This work was supported by the Laboratory of Lingnan Modern Agriculture Project (NT2021001), the Guangdong Provincial Natural Science Foundation Program (No. 2019A1515011826), Guangdong Provincial Special Projects in Key Field of Rural Revitalization (No. 2020ZDZX1038).

The authors are thankful to Professor Kangjing Liang to donate the rice germplasm with long grain, KJ01. The authors also thank Professor Yaoguang Liu’s and Jian Wu’s laboratory teams for their valuable technical assistance.
AUTHOR CONTRIBUTIONS

L.W. and X.L. designed the research, revised the paper, and agree to serve as the author responsible for contact and ensure communication. F.B. and H.M. performed most of the experiments, F.B. prepared article draft; Y.Z. constructed the vector of CRISPR/Cas9 gene editing; C.L. measured some traits of transgenic seeds; Y.C., M. Q. S., and Z.C. revised paper; J.W. managed the field; X.L. designed the research and revised paper, and developed the material of Huaye 3; L.W. designed the research, analyzed data, wrote and revised the paper.

Tables

Table 1 Grain size analysis of KO-1 and KO-2 in T3 lines.

<table>
<thead>
<tr>
<th>Variety</th>
<th>Grain length (mm)</th>
<th>Grain width (mm)</th>
<th>1000-grain weight (g)</th>
<th>No. secondary branching peduncles</th>
<th>Grain yield per plant (g)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Huaye 3</td>
<td>7.19±0.20</td>
<td>2.47±0.07</td>
<td>10.24±1.18</td>
<td>12.95±2.57</td>
<td>6.96±0.13</td>
</tr>
<tr>
<td>KO-1</td>
<td>8.64±0.3**</td>
<td>2.56±0.09**</td>
<td>17.3±1.26**</td>
<td>29.00±5.02**</td>
<td>18.63±0.83**</td>
</tr>
<tr>
<td>KO-2</td>
<td>8.20±0.14**</td>
<td>2.51±0.06**</td>
<td>15.56±1.23**</td>
<td>29.67±3.67**</td>
<td>9.61±0.11**</td>
</tr>
</tbody>
</table>

Notes: *p < 0.05 and **p < 0.01 indicate significant differences of means ± SD, as determined by Student’s t-test. 50 individuals were planted for each line, and three-sampling replicates were conducted for each plant trait.

Figure Legends

Figure 1 The genomic sequence difference and expression pattern of GSW3 between the parents Huaye 3 and KJ01.

A. Unrooted phylogenetic tree of GSW3 and seven G proteins in rice was constructed using the MEGA7.0 software, where the star indicates GSW3. 

B. The genomic sequence difference of GSW3. There are twelve SNPs in the GSW3 coding region between Huaye 3 and KJ01, and the bases marked with red generate variation in amino acids. The triangle indicates the large-fragment insertion position.
C. Expression pattern of GSW3 in Huaye 3 and KJ01 cells determined by RT-qPCR (n=3).

D. Expression difference of GSW3 among different developmental panicles of Huaye 3 determined by RT-qPCR (n=3). DAF, days after flowering.

E. Expression difference of GSW3 among different developmental grains of Huaye 3 determined by RT-qPCR (n=3).

F. Histochemical analysis of GUS activity in different tissues and different developmental stage of grains.

Data are given as means ± SD.

**Figure 2 Subcellular localization of GSW3.**

A. pBin19-EGFP-GSW3 and pBin19-EGFP were transformed into N. benthamiana leaves and transiently coexpressed in N. benthamiana leaf epidermal cells.

B. pOX-GFP-GSW3 (GSW3 and green fluorescent protein fusion protein) and NLS-mKATE (nuclear localization sequence-fused mKATE protein) were cotransformed into rice protoplasts. Fluorescence was visualized using confocal microscopy.

**Figure 3 Analysis of the generations of transformed plants.**

A. Five editing types generated in the two target sites of GSW3 of lines #1 and #2, where “-” indicates the base deletion.

B. The grain length, grain width of five KO mutants increases obvious relative to wild-type Huaye 3 in T1 transgenic lines. Three sample repetitions for each plant trait were conducted.

C. Compared with wild-type Huaye 3, the difference of grain length and width for the five KO mutants was very significant in T1 transgenic lines (n=50). Student’s t-test was used to generate the p values, and “***” indicates a highly significant difference at the p<0.01 level.

D. The plant type of KO-1 and KO-2 in T2 transgenic lines (n=50).

E. The grain length and grain width analysis for KO mutants (KO-1 and KO-2 in T3 lines), and OE mutants (OE-6 and OE-11 in T1 lines), and CF mutants (CF-1 and CF-
15 in $T_2$ lines). 50 individuals of each line were grown, three sample repetitions for each plant trait were conducted.

F. Yield analysis for each plant of KO mutants (KO-1 and KO-2) in $T_3$ lines. 50 individuals of each line were grown, three sample repetitions for each plant were conducted.

G. The mRNA expression levels of $GSW3$ in KO mutants (KO-1 and KO-2) in $T_3$ lines and OE mutants (OE-6 and OE-11) in $T_1$ lines (n=3).

H. Compared with wild type KJ01, the mRNA expression of $GSW3$ and plant type analysis in complementary function $T_2$ lines (CF-1 and CF-15) (n=3).

Data are given as means ± SD.

**Figure 4** Cytological analysis of the glumes of KO and OE mutants.

A. Fluorescence microscope observation of the grain outer surfaces for Huaye3 and KO-2 performed by using plastic semithin-section technology. The regions marked with a white box were enlarged (middle panel).

B. Significant differences in the cell area and epidermal cell numbers of the lemma were analyzed in Huaye3 and KO-2. 30 grains of each sample were conducted by semithin-section, and three views of each grain’s central-part section were analyzed.

C. Scanning electron microscopy observation of the central region of the grain outer surfaces for Huaye3, KO mutants (KO-1 and KO-2), and OE mutants (OE-6 and OE-15).

D. Significant difference analysis of the cell length, longitudinal cell numbers, cell width, and the transverse cell numbers of the grain outer glumes between the wild-type Huaye3, KO mutants (KO-1 and KO-2), and OE-6 mutants (OE-6 and OE-15). Three grains of each sample were analyzed.

Data are given as means ± SD, and Student’s t-test was used to generate the $p$ values, **$p<0.01$.

**Figure 5** Transcriptome analysis for the different developmental glumes of KO-2 and Huaye3.
A. Distributions of the position and coverage depth for mapped reads in the reference genome. Blue is the positive strand, and green is the negative strand.

B. Read distribution map of different regions in the genome.

C. Statistical chart of GO annotation classification results of differentially expressed genes for 25% full-length glumes. The abscissa is the GO classification, the left side of the ordinate is the percentage of the gene numbers, and its right side is the gene numbers.

D. KEGG pathway enrichment analysis of differentially expressed genes for 25% full-length glumes. Each circle in the figure represents a KEGG pathway, the ordinate represents the pathway name, and the abscissa represents the enrichment factor. The larger the enrichment factor is, the more significant the enrichment level of differentially expressed genes in this pathway. The color of the circle represents the Q value, and the Q value is the p value after correction by a multiple hypothesis test. The smaller the Q value is, the more reliable the enrichment significance of differentially expressed genes in this pathway. The size of the circle indicates the number of genes enriched in the pathway, and the larger the circle is, the more genes.

E. RT-qPCR for six genes associated with grain shape and the grain filling rate (n=3). Data are given as means ± SD.

Figure 6 Second leaf sheath length and relative mRNA expression analysis of GSW3 in Huaye 3 and GSW3 KO mutants under exogenous GA3 treatment.

A. Phenotypic analysis under different concentrations of exogenous GA3.

B. Length analysis of the second leaf sheath under different concentrations of exogenous GA3 (n=30).

C. Average growth rate of the second leaf sheath under different concentrations of exogenous GA3.

D. Expression analysis of GA biosynthesis-related and GA signaling genes by RT-qPCR in KO-2 mutant (n=3).
Data are given as means ± SD. Student’s t-test was used to generate the \( p \) values.

**Figure 7 Seven nonsynonymous SNPs in the *GSW3* coding region and their effect on grain size and grain weight.**

A. Haplotype analysis of seven nonsynonymous SNPs in the *GSW3* coding region, where \( Y \) stands for C/T, and \( R \) stands for A/G. Num., number of accessions in each category. B. Significant difference analysis of grain size and grain weight for five haplotypes. 1025 accessions of cultivated rice from the 2013 Rice Genome Project were analyzed, and the violin plot was constructed in Graphpad. Different letters indicate statistically significant differences between groups (\( p < 0.01 \)), as determined by Tukey’s honestly significant difference test. The longitudinal dots of each violin shows the trait average value of each haplotype group, and the transverse dotted line of each violin shows respectively the quartile and median of each haplotype.

**References**


**Figure 1** The genomic sequence difference and expression pattern of *GSW3* between the parents of Huaye 3 and KJ01.

A. Unrooted phylogenetic tree of GSW3 and seven G proteins in rice was constructed using the MEGA7.0 software, where the star indicates GSW3.

B. The genomic sequence difference of *GSW3*. There are twelve SNPs in the *GSW3* coding region between Huaye 3 and KJ01, and the bases marked with red generate variation in amino acids. The triangle indicates the large-fragment insertion position.

C. Expression pattern of *GSW3* in Huaye 3 and KJ01 cells determined by RT-qPCR (n=3).

D. Expression difference of *GSW3* among different developmental panicles of Huaye 3 determined by RT-qPCR (n=3).

E. Expression difference of *GSW3* among different developmental grains of Huaye 3 determined by RT-qPCR (n=3). DAF, days after flowering.

F. Histochemical analysis of GUS activity in different tissues and different developmental stage of grains.

Data are given as means ± SD.
**Figure 2** Subcellular localization of GSW3.

A. pBin19-EGFP-GSW3 and pBin19-EGFP were transformed into *N. benthamiana* leaves and transiently coexpressed in *N. benthamiana* leaf epidermal cells.

B. pOX-GFP-GSW3 (GSW3 and green fluorescent protein fusion protein) and NLS-mKATE (nuclear localization sequence-fused mKATE protein) were cotransformed into rice protoplasts. Fluorescence was visualized using confocal microscopy.
Figure 3 Analysis of the generations of transformed plants.
A. Five editing types generated in the two target sites of GSW3 of lines #1 and #2, where “−” indicates the base deletion. B. The grain length, grain width of five KO mutants increases obvious relative to wild-type Huaye 3 in T1 transgenic lines. Three sample repetitions for each plant trait were conducted. C. Compared with wild-type Huaye 3, the difference of grain length and width for the five KO mutants was very significant in T1 transgenic lines (n=50). Student's t-test was used to generate the p values, and “***” indicates a highly significant difference at the p<0.01 level. D. The plant type of KO-1 and KO-2 in T2 transgenic lines (n=50).
E. The grain length and grain width analysis for KO mutants (KO-1 and KO-2 in T1 lines), and OE mutants (OE-6 and OE-11 in T1 lines), and CF mutants (CF-1 and CF-15 in T1 lines). 50 individuals of each line were grown, three sample repetitions for each plant trait were conducted. F. Yield analysis for each plant of KO mutants (KO-1 and KO-2) in T1 lines. 50 individuals of each line were grown, three sample repetitions for each plant were conducted. G. The mRNA expression levels of GSW3 in KO mutants (KO-1 and KO-2) in T1 lines and OE mutants (OE-6 and OE-11) in T1 lines (n=3). H. Compared with wild type KJ01, the mRNA expression of GSW3 and plant type analysis in complementary function T1 lines (CF-1 and CF-15) (n=3).
Data are given as means ± SD.
Figure 4  Cytological analysis of the glumes of KO and OE mutants.
A. Fluorescence microscope observation of the grain outer surfaces for Huaye3 and KO-2 performed by using plastic semithin-section technology. The regions marked with a white box were enlarged (middle panel).
B. Significant differences in the cell area and epidermal cell numbers of the lemma were analyzed in Huaye3 and KO-2. 30 grains of each sample were conducted by semithin-section, and three views of each grain’s central-part section were analyzed.
C. Scanning electron microscopy observation of the central region of the grain outer surfaces for Huaye 3, KO mutants (KO-1 and KO-2), and OE mutants (OE-6 and OE-15).
D. Significant difference analysis of the cell length, longitudinal cell numbers, cell width, and the transverse cell numbers of the grain outer glumes between the wild-type Huaye 3, KO mutants (KO-1 and KO-2), and OE-6 mutants (OE-6 and OE-15). Three grains of each sample were analyzed.
Data are given as means ± SD, and Student’s t-test was used to generate the p values, **p<0.01.
Figure 5 Transcriptome analysis for the different developmental glumes of KO-2 and Huaye 3.
A. Distributions of the position and coverage depth for mapped reads in the reference genome. Blue is the positive strand, and green is the negative strand.
B. Read distribution map of different regions in the genome.
C. Statistical chart of GO annotation classification results of differentially expressed genes for 25% full-length glumes. The abscissa is the GO classification, the left side of the ordinate is the percentage of the gene numbers, and its right side is the gene numbers.
D. KEGG pathway enrichment analysis of differentially expressed genes for 25% full-length glumes. Each circle in the figure represents a KEGG pathway, the ordinate represents the pathway name, and the abscissa represents the enrichment factor. The larger the enrichment factor is, the more significant the enrichment level of differentially expressed genes in this pathway. The color of the circle represents the Q value, and the Q value is the p value after correction by a multiple hypothesis test. The smaller the Q value is, the more reliable the enrichment significance of differentially expressed genes in this pathway. The size of the circle indicates the number of genes enriched in the pathway, and the larger the circle is, the more genes.
E. RT-qPCR for six genes associated with grain shape and the grain filling rate (n=3). Data are given as means ± SD.
Figure 6 Second leaf sheath length and relative mRNA expression analysis of GSW3 in Huaye 3 and GSW3 KO mutants under exogenous GA3 treatment.

A. Phenotypic analysis under different concentrations of exogenous GA3.

B. Length analysis of the second leaf sheath under different concentrations of exogenous GA3 (n=30).

C. Average growth rate of the second leaf sheath under different concentrations of exogenous GA3.

D. Expression analysis of GA biosynthesis-related and GA signaling genes by RT-qPCR in KO-2 mutant (n=3).

Data are given as means ± SD. Student’s t-test was used to generate the p values.
Figure 7 Seven nonsynonymous SNPs in the GSW3 coding region and their effect on grain size and grain weight.

A. Haplotype analysis of seven nonsynonymous SNPs in the GSW3 coding region, where Y stands for C/T, and R stands for A/G. Num., number of accessions in each category. B. Significant difference analysis of grain size and grain weight for five haplotypes. 1025 accessions of cultivated rice from the 2013 Rice Genome Project were analyzed, and the violin plot was constructed in Graphpad. Different letters indicate statistically significant differences between groups ($p<0.01$), as determined by Tukey’s honestly significant difference test. The longitudinal dots of each violin shows the trait average value of each haplotype group, and the transverse dotted line of each violin shows respectively the quartile and median of each haplotype.


