Indian J. Genet., 79 (2) 383-393 (2019)
DOI: 10.31742/IJGPB.79.2.1

Structure and expression analysis of *TaGW5* in common wheat

Puyang Ding[#], Jian Ma[#]*, Yujie Yang, Wei Luo, Yaya Zou, Ting Li, Yang Mu, Huaping Tang and Xiujin Lan[#]

Triticeae Research Institute, Sichuan Agricultural University, Chengdu, Sichuan 611130 and State Key Laboratory of Crop Gene Exploration and Utilization in Southwest China

(Received: June 2018; Revised: April 2019; Accepted: May 2019)

Abstract

***OsGW5* (*OsGSE*) regulates cell proliferation and is involved in regulating grain size and thousand-grain weight in rice. Our knowledge about its wheat ortholog *TaGW5* is limited. In the present study, we characterized the structure and expression of *TaGW5* at molecular level in wheat and predicted the *cis*-elements and transcription factor binding sites (TFBS) of its promoter. The *GW5* orthologs in barley (*HvGW5*), rice (*OsGW5*), *Triticum turgidum* L. (*TtGW5*) and *Brachypodium distachyon* (*BdGW5*) were also identified for comparative analyses. *TaGW5* was mapped onto the short arms of group 1 chromosomes (1AS, 1BS, and 1DS). Multiple alignments indicated *GW5* possesses three exons and two introns in all the analyzed species except for rice and the exon-intron junction composed of exon 2 and intron 2 was highly conserved. *GW5* has a conserved domain (DUF 4005) and two neighboring IQ domains and was mainly expressed in wheat young spikes, in barley immature inflorescences and in rice anthers. Drought, heat and biotic-stress treatments had no significant effects on *HvGW5* and *TaGW5* expression. Significant correlation between the expression patterns of predicted transcription factor *ABF2* and *OsGW5* was also detected. Taken together, these results broaden our understanding of *GW5* in wheat, barley and rice and will be helpful for further manipulating *GW5* and uncovering its roles in plants**

Key words : Wheat, *GW5*, gene structure, transcription factor, protein structure, expression pattern

Introduction

Common wheat (*Triticum aestivum* L., 2n=6x=42, AABBDD genome) is one of the most important food crops in the world. One of the most important goals in breeding program is to increase wheat yield. As the demand from population growth exceeds the global food supply, food shortages have once again become a serious problem (Takeda and Matsuoka 2008; Brown and Funk 2008). Therefore, identifying, understanding, and introducing genes or alleles to improve wheat yield is becoming essential. Wheat grain yield is determined by grain weight, grain number per panicle and panicles per unit area. Grain weight, a major component of yield, is mainly determined by grain size and the degree of grain filling (Brocklehurst 1977). Grain size/weight are important factors that determine grain yield and thus increasing grain size/weight may be an effective strategy in breeding for yield increase. So far, numerous quantitative trait loci (QTLs) associated with grain size/weight have been identified in wheat (Sun et al. 2009; Gegas et al. 2010; Cui et al. 2014; Li et al. 2015; Wu et al. 2015; Su et al. 2016). Given the important roles of yield related genes in regulating wheat yield, it's rather essential to clone these genes for functional analyses and further modification in wheat breeding. However, due to the huge and complex genome of wheat, a large number of repetitive nucleotide sequences, and lack of an annotated genome sequence it is difficult to directly isolate yield-related genes by map-based cloning strategies.

Rice functional genome has greatly facilitated the isolation and functional characterization of many genes associated with grain size/weight, such as *OsGW5* (Weng et al. 2008), *OsGS6* (Sun et al. 2013), *OsTGW6* (Ishimaru et al. 2013), *OsGS5* (Xu et al. 2015), *OsGL2* (Che et al. 2015), *TW3/GL3.3/TGW3* (Hu et al. 2018; Xia et al. 2018; Ying et al. 2018) and *GS9* (Zhao et al. 2018). The collinearity of genes in rice, wheat, barley, *Brachypodium*, and sorghum is generally well conserved (Choi et al. 2004; Vogel et al. 2010). The collinearity of the cereal genomes coupled with the availability of IWGSC (The

*Corresponding author's e-mail: plantgbmj@hotmail.com, lanxiujin@163.com. [#]These authors contributed equally to this work.
Published by the Indian Society of Genetics & Plant Breeding, A-Block, F2, First Floor, NASC Complex, IARI P.O., Pusa Campus, New Delhi 110 012; Online management by www.isgpb.org; indianjournals.com

International Wheat Genome Sequencing Consortium, 2014) Ref Seq v1.0 provides a possibility for cloning yield-related genes in wheat. For example, *TaGW2*, the orthologous gene of *OsGW2* was isolated in wheat and was reported to regulate grain width and weight (Yang et al. 2012; Simmonds et al. 2016). A single T-base insertion was detected in a conserved exon of wheat *TaGW2*. The association analysis based on a SNP marker developed from this insertion mutation indicated that the mutated *TaGW2* allele significantly increased grain width and thousand-grain weight, and slightly increased grain length (Yang et al. 2012). Additionally, *TaGS5* was successfully obtained from bread wheat by in silico cloning and correlation analysis with agronomic traits showed that wheat cultivars with *TaGS5-A1b* allele usually possessed favored agronomic traits (Wang et al. 2015). In rice, *GW5*/*GSE5* encodes a calmodulin binding protein and regulates grain size predominantly by influencing cell proliferation in spikelet hulls (Duan et al. 2017; Liu et al. 2017). Overexpression of *GW5* resulted in narrow grains, while loss of *GW5* function caused wide and heavy grains (Liu et al. 2017). Here, we describe our efforts in identification of the *GW5* gene in wheat and its characterization.

Materials and methods

Sequence data and identification of GW5 genes

The collection of sequence data was according to previous study reported by Zhang et al. (2018). The reported *GW5* (*LOC_Os05g09520*) (Duan et al. 2017; Liu et al. 2017) in rice wasused to BLAST (Basic Local Alignment Search Tool) against the CDS regions for wheat, barley and *B. distachyon*. We used the BLAST++ BLASTN algorithm with an E value cut-off of 10^{-5} to determine full-length genomic and promoter sequences (Ma et al. 2013) based on the above said genome sequences. The accessions and locations of promoters, genomic sequences, and CDS regions are listed in Supplementary (Suppl.) Table S1.

Isolation of the GW5 gene in wheat

To confirm the retrieved *GW5* orthologous sequences in wheat, we isolated the ortholog from 'Chinese Spring' variety. We used the cetyltrimethyl ammonium bromide (CTAB) method (Murray and Thompson 1980) to extract genomic DNA from 20-day-old seedlings. Three sets of primers (Table 1; Fig. 1a) were designed to amplify the full-length of *GW5* sequence based on the genomic sequences of 'Chinese Spring', barley, *Aegilops tauschii*, and *Triticum turgidum* (Suppl. Table S1). PCR amplification was performed with the Gene Amp® PCR System 9700 (Applied Biosystems). PCR reaction mixtures (25 µL) contained 50 ng of genomic DNA, 200 µM of each dNTP, 0.2 µM of each primer, and 0.625 units of LA Taq polymerase (TaKaRa, Shiga, Japan) with 1 GC buffer for GC-rich templates to avoid introducing errors into the sequence. The cycling parameters were 94°C for 5min for pre-denaturation, followed by 40 cycles of 94°C for 45s, 60°C for 40s, 72°C for 2min in genomic DNA, and a final extension at 72°C for 10min. Amplification products were separated on 1.5 % agarose gels, and fragments of interest were purified using Gel Extraction Kit D2500 (Omega Bio-Tek, American). The recovered PCR products were cloned and sent to Tsingke Biotech Co., Ltd. for sequencing. At least three independent clones for each amplicon were sequenced in both directions. From the sequenced fragments, we identified *GW5* homeologs from A, B and D sub-genomes. Thus, these three homeologs were used for the following analysis. The CDS of the *GW5* homeolog from the A, B and D sub-genome was deduced by the isolated genomic sequences.

Table 1. Details of primers used for amplification and RT-qPCR

Primer	Sequence (5'-3')	Description
GW5-1AF	TCTTTGTTCACTAGC AGGC	gene isolation
GW5-1BF	CACTGCTCTCTTGG TTCACT	gene isolation
GW5-1BR	GCGAGTCGTAATGG AAGAACA	gene isolation
GW5-1DF	CGCGCTAATGGCTG CCCTAATA	gene isolation
GW5-t1	ACGCAGAGCTCTCG TCGTCT	gene isolation
qGW5-AF2	CATGTCCAGCACGCA GTCGTT	expression analysis
qGW5-AR2	AACGGCCGTCTTGAA GTTG	expression analysis
qGW5-DF1	AGCTTCGGCAAGTCG TCGCGC	expression analysis
qGW5-DR1	TGCTTGGGTGGCCG CCGCCA	expression analysis
β-actin-F1	ACCTTCAGTTGCCC AGCAAT	expression analysis
β-actin-R1	CAGAGTCGAGCACAA TACCAGTTG	expression analysis

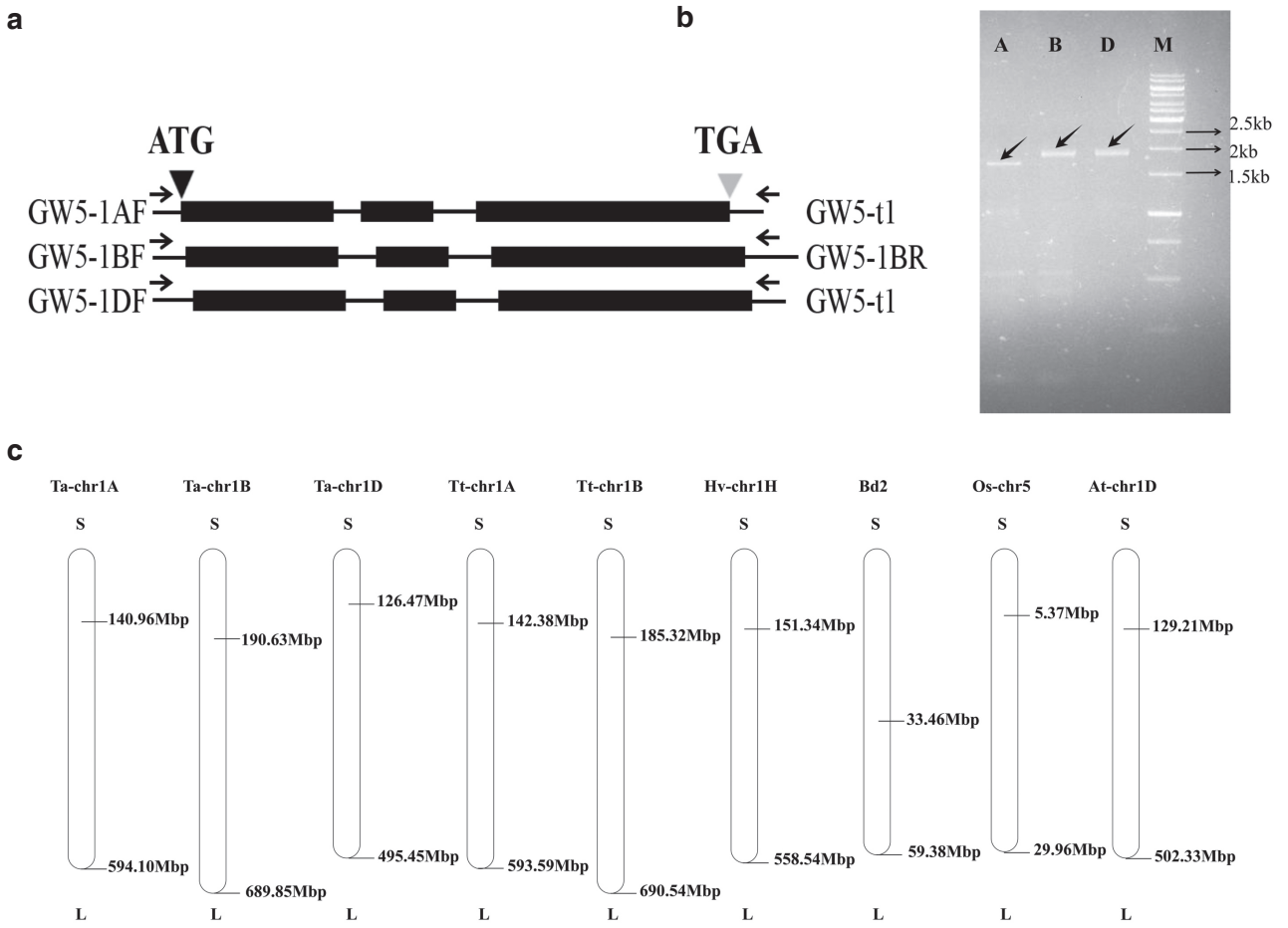


Fig. 1. Characterization of *GW5* in various species. a, location of primers for amplification of *GW5* in wheat. b, isolation of *GW5* from wheat M: markers for DNA Ladder, A: the band amplified by GW5-1AF/GW5-t1, B: the band amplified by GW5-1BF/GW5-1BR, D: the band amplified by GW5-1DF/GW5-t1.c, Chromosomal location of *GW5* in different species S and L represent short and long arms in a given chromosome, respectively. Ta, Tt, Hv, Bd, At, Os represent wheat,*T. turgidum*, barley, *B.distachyon*, *A. tauschii*, and rice, respectively

Chromosomal locations of GW5 genes and identification of nucleotide and protein structures

In order to map the locations of *GW5* genes in different cereal species, the CDS regions of *GW5* sequences from each species were used to BLAST against chromosome sequences (E value cut-off of 10^{-5}) (Ma et al. 2013).

The structural analysis of intron-exon and the encoded protein motifs of *GW5* in different species were conducted as per the procedure described in Zhang et al. (2018). Genomic and coding sequences are listed in Table S1. The accession numbers of amino acid sequences for protein structure analysis are the same as those used to analyze the evolutionary relationship described below.

Multiple alignments and phylogenetic analysis

Multiple alignments of coding sequences for various species were performed by Clustal W (V1.83) (Thompson et al. 1994). We used sequences of *GW5* sequences from *T. aestivum* (*TaGW5-1AS*, *TaGW5-1BS*, and *TaGW5-1DS*), *Hordium vulgare* (*HvGW5*), *T. turgidum* (*T. turgidum-1A* and *T. turgidum-1B*), *B. distachyon* (*BdGW5*), and *O. sativa* (*OsGW5*) to construct the phylogenetic trees. We used MEGA 7 to reveal the evolutionary relationship of the *GW5* genes by generating a dendrogram with the neighbor joining method (Kumar et al. 2016).

Identification of cis acting elements in the promoters and prediction of transcription factor binding sites

We analyzed the 1500 bp sequence upstream of *GW5* from various species to analyze the main promoter elements (Hackett et al. 1996). Thus, ~1500 bp fragments upstream of GW5 from rice, barley and wheat (Suppl. Table S1) were retrieved for further analysis. These sequences were used to predict the cis-acting elements of *GW5* promoters in the PlantCARE database (Lescot et al. 2002). The common elements were identified from http://bioinformatics.psb.ugent.be/webtools/Venn/. Transcription factor binding sites, transcription factors and correlation analysis of expression patterns between predicted transcription factors and *GW5* in wheat, barley and rice were carried out based on previous study reported by Zhang et al. (2018).

Analysis of expression patterns and quantitative RT-PCR analysis

We searched PLEXdb (http://www.plexdb.org/index.php) to obtain expression sequence tags (ESTs) for barley and rice of the *GW5* sequences (Dash et al. 2012), and retrieved microarray data for barley (Contig15603_at) and rice (Os.26803.1.S1_x_at) from NCBI, EBI (http://www.ebi.ac.uk/), and PLEXdb (http://www.plexdb.org/index.php). The retrieved data contained 461 samples from nine experiment series (Ma et al. 2017). Previous methods (Jiang et al. 2011; Long et al. 2010; Ma et al. 2017) were used to analyze microarray data. The data were retrieved for either development-specific expression or expression under a given stress treatment in a given species from a single accession and within the same laboratory, growth condition and expression analysis to improve the reliability of the results. Further, for a given species, all the data were collected from the same common microarray platform. In addition, the transcriptome data of rice (Xia et al. 2017) was also chosen for expression analysis. For barley, the transcriptome data for *GW5* was obtained by utilizing the method reported by Zhang et al. (2018). The expression patterns of the three *GW5* homeologs in wheat (*Traes_1AS_094FB4F451.1* for *TaGW5-1AS, Traes_1BS_AD504889B.2* for *TaGW5-1BS*, and *Traes_1DS_FBF57DBCD.2* for *TaGW5-1DS*) were obtainedby retrieving the processed expression value as transcripts per million (tpm) from expVIP (http://www.wheat-expression.com/) (Borrill et al. 2016). For further identificationof the expression pattern of *GW5* in wheat, the expression values as FPKM (Fragments Per Kilobase of transcript per Million mapped reads) for the three homeologs (*Traes_1AS_094FB4F451.1* for *TaGW5-1AS, Traes_1BS_AD504889B.2* for *TaGW5-1BS*, and *Traes_1DS_FBF57DBCD.2* for *TaGW5-1DS*) were also retrieved from WheatExp (http://wheat.pw.usda.gov/WheatExp/) (Pearce et al. 2015). The Zadoks scale (Zadoks et al. 2010) for wheat growth stage was adopted and detailed information for each stage is provided in Suppl. Table S2. RT-qPCR was adopted to validate the expression patterns of *TaGW5*. The different tissues of the wheat variety Chinese Spring were collected and used for RNA extraction and further cDNA synthesis. The crop was raised according to common agricultural practice (Ma et al. 2019; Yu et al. 2018) and each sample had three biological replicates. We used SYBR Premix Ex TaqTM II (TaKaRa, Shiga, Japan) in the amplification system. 5 µL SYBR Green II Mix, 0.5 µL forward and backward primers, 100 ng cDNA, 3 µL ddH_2O constituted the ten microliter reaction mixtures. The detailed samples collection and RT-PCR procedures were as per the report of Zhang et al. (2018). The primer sequences are given in Table 1.

Results

Characterization and chromosomal mapping of GW5 genes

Three sets of A, B and D chromosomally specific primers were used to get the expected wheat *GW5* gene fragments (~1.7kb, 1.9kb and 1.8kb, Fig. 1b). BLAST results demonstrated that the wheat *GW5* (*TaGW5*) gene is a single-copy gene which has three homeologs on A (*TaGW5-1A*), B (*TaGW5-1B*), and D (*TaGW5-1D*) sub-genomes, respectively. The multiple alignment showed 98.24% similarity on three homeologs, and their full-length genomic sequences were 1563 bp (*TaGW5-1A*), 1593 bp (*TaGW5-1B*), and 1592 bp (*TaGW5-1D*) in length, respectively. The coding sequences of three homeologs were all 1365 bp (*TaGW5-1A, TaGW5-1B* and *TaGW5-1D*), and these sequences shared 99.05% similarity based on multiple-alignment. The sequences of these isolated *GW5* homeologs on A (*TaGW5-1A*), B (*TaGW5-1B*) and D (*TaGW5-1D*) sub-genome are listed in Suppl. Table S1. BLAST results showed that *GW5* in rice, barley, *T. turgidum* and *A. tauschii* is also a single copy gene. Physical mapping demonstrated that these homeologs of *GW5* in wheat were located on group 1 chromosomes i.e., 1AS, 1BS, and 1DS (Fig. 1c), respectively. The *HvGW5* and *TtGW5* of *GW5* gene in barley were physically mapped to the short arm of

chromosome 1 while *BdGW5* and *OsGW5* of *GW5* sequence in *B. distachyon* and rice mapped on chromosome 2 and 5 respectively (Fig. 1c).

Gene structure of GW5

Among the species analyzed, the full-length genomic sequences of *GW5* genes range from 1563 bp (*T. aestivum, 1AS*) to 1827 bp (*H. vulgare*) (Suppl. Table S3), and CDS lengths range from 1365 bp (*T. aestivum, 1AS, 1BS, 1DS* and *A. tauschii*) to 1584 bp (*H. vulgare*; Table S3). The lengths of introns range from 77 bp (intron 1 of *T. aestivum* and *T. turgidum* 1AS) to 140 bp (intron 1 of *B. distachyon*), and those of exons range from 201 bp (exon 2 of *B. distachyon*) to 772 bp (exon 3 of *B. distachyon*; Table S3). Sequence alignments and gene structure analysis demonstrated that *GW5* genes from barley (*HvGW5*), wheat (*TaGW5-1AS*, *TaGW5-1BS* and *TaGW5-1DS*), *B. distachyon* (*BdGW5*), *A. tauschii* (*AtGW5*) and *T. turgidum* (*TtGW5-1A, TtGW5-1B*) contained three exons and two introns (Fig. 2). *GW5* in rice (*OsGW5)* consisted of two exons and one intron. Alignment of the exons and introns of the *GW5* genes indicated that *OsGW5* lacks exons 3 and introns 2 compared with the orthologs in investigated species. In addition, exon 2 of *OsGW5* was in fact homologous to exon 2, intron 2 and exon 3 in wheat and *T. turgidum*. Furthermore, the barley and *B. distachyon* has longer exon1 and exon3 than the other species (Fig. 2 and Suppl. Table S3).

Multiple-alignments of exons and introns showed that the similarity of exons among the five species ranges from 65.29% (exon 1) to 88.70% (exon 3) with a mean of 79.47%, and the similarity of their introns range from 64.64% (intron 1) to 78.34% (intron 2, Suppl. Table S3). Additionally, except rice, a specific exon-intron junction of *GW5* was found to be highly conserved among these different species. It is composed of intron 2 and exon 2 (Fig. 2).

Protein structure and phylogenetic analysis

Two highly conserved IQ domains and a highly conserved domain of unknown function (DUF 4005) were detected at GW5 protein. Further, except rice, we identified a conserved neighboring low complexity region upstream of the DUF 4005 domain among the investigated species. Three low complexity regions were detected in rice. Moreover, we also identified some other low complexity regions. In addition to those homologous regions on chromosome A of *T. turgidum* and wheat, the low complexity regions in barley, *B. distachyon*, rice, the chromosome B of *T. turgidum* and the chromosome B, D of wheat were similar along the amino acid sequences (Fig. 3a and Suppl. Table S4). The phylogenetic analysis demonstrated that *GW5* genes in *Triticum* and related genera (i.e., *A. tauschii*, *T. turgidum* and *T. aestivum*) were gathered into a single group. The *GW5* of *Triticum* exhibited a closer relationship with that of barley compared to those of *B. distachyon* and rice (Fig. 3b).

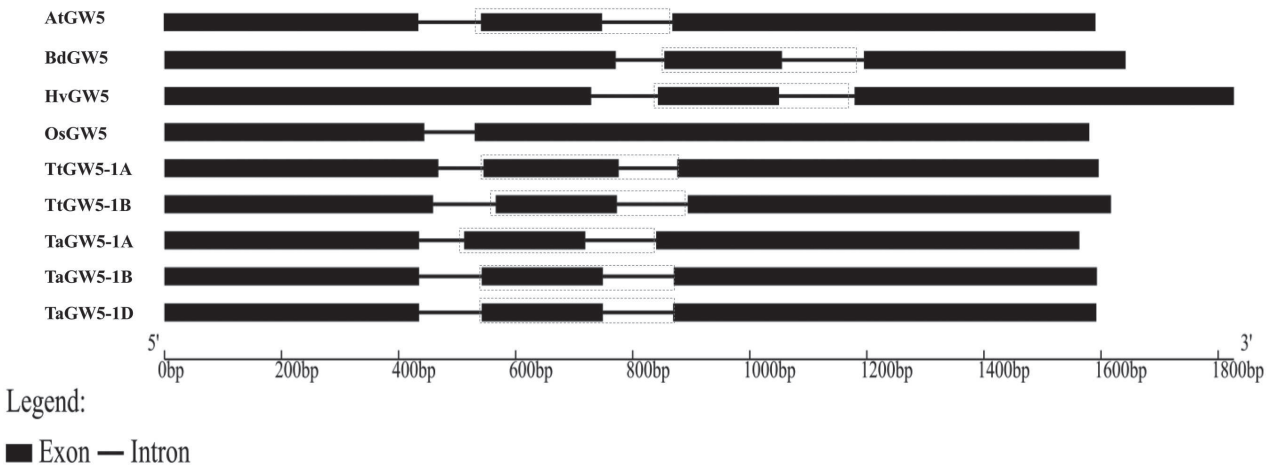


Fig. 2. Gene structures of *GW5* from a range of species. Symbols include: At for *A. tauschii*, Bd for *B. distachyon*, Hv for *H. vulgare*, Os for *O. sativa*, Tt for *T. turgidum* and Ta for *T. aestivum*. Red box represents the conservative exon-intron junctions in addition to rice

Analyses of promoter

We compared *cis*-acting elements and identified 30, 27, 23, 29 and 31 elements in *HvGW5, OsGW5,*

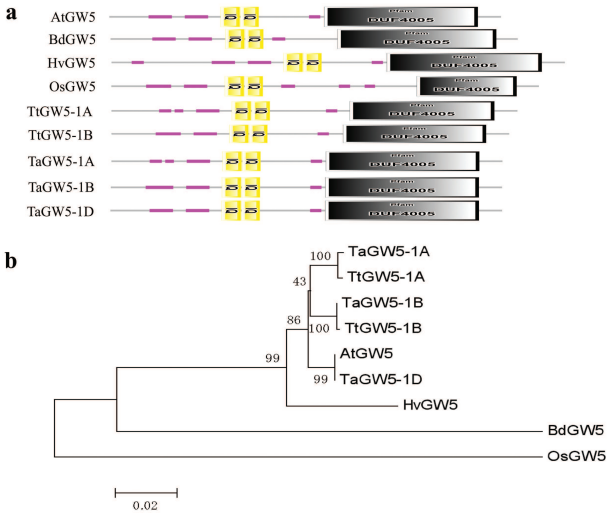


Fig. 3. Protein structures of GW5. a, and phylogenetic relationship. b, from a range of species based on *GW5* coding sequences. Symbols include Bd for *B. distachyon*, Hv for *H. vulgare*, Os for *O. sativa*, Tt for *T. turgidum* and Ta for *T. aestivum*. DUF4005 is domain of the deduced proteins. Purple box represents different low complexity regions. The detailed domain sequences and their locations were listed in Table S4

TaGW5-1AS, *TaGW5-1BS* and *TaGW5-1DS*, respectively (Suppl. Table S5). In total, twelve common elements shared by *OsGW5*, *HvGW5*, *TaGW5-1AS*, *TaGW5-1BS* and *TaGW5-1DS* (Table 2) were detected. The common elements contain core or common promoter elements such as ABRE, regulatory element involved in the abscisic acid responsiveness (Table 2). Comparison of *cis*-acting elements showed that three and five elements were from barley and wheat only, respectively (Suppl. Table S5). Prediction of transcription factor binding sites (TFBS) showed that the three species displayed five common matrixes including AG, ABI5, ABI3, ABF2 and abi4, respectively (Table 3). In addition, we found that MA1244.1 which binds ABR1 transcription factor shared in wheat and rice, was absent in barley (Suppl. Table S6). All detected motifs were further validated on footprint DB (http://floresta.eead.csic.es/footprintdb) (Zhang et al. 2018). The orthologs of predicated transcription factors in wheat, barley and rice based on the binding sites are listed in Suppl. Table S7.

Tissue-specificity of GW5 expression and response of GW5 to different stresses

In wheat, we detected the highest expression level of three *GW5* homeologs in spikes when second

Table 2. Putative common motifs of GW5 promoters in barley, rice and wheat

Motifs name	Function
CGTCA-motif	*cis*-acting regulatory element involved in the MeJA-responsiveness
CCGTCC-box	*cis*-acting regulatory element related to meristem specific activation
A-box	*cis*-acting regulatory element
AAGAA-motif	unknown
TGACG-motif	*cis*-acting regulatory element involved in the MeJA-responsiveness
Sp1	light responsive element
GAG-motif	part of a light responsive element
CAAT-box	common *cis*-acting element in promoter and enhancer regions
G-Box	*cis*-acting regulatory element involved in light responsiveness
Skn-1_motif	*cis*-acting regulatory element required for endosperm expression
TATA-box	core promoter element around –30 of transcription start
ABRE	*cis*-acting element involved in the abscisic acid responsiveness

Table 3. Predicted consensus motifs in wheat, barley and rice

Common matrix ID	Name	Predicted sequence
MA0005.1	AG	CCTATTTAAGC (wheat, rice, barley)
MA0931.1	ABI5	GTACACGTGC (wheat-1AS, wheat-1DS) ACACACGTAG (wheat-1BS) GCACACGTGC (barley) TACCACGTAA (rice)
MA0564.1	ABI3	TCGCATGCT (wheat-1AS, wheat-1DS) TTGCATGCT (wheat-1BS) CTGCATGCC (barley) GTGCATGCG (rice)
MA0941.1	ABF2	GGGTACACGTGCA (wheat-1AS, wheat-1DS) TAACACACGTAGA (wheat-1BS) GGGCACACGTGCA (barley) AGTACCACGTAAG (rice)
MA0123.1	abi4	CGCTGCCCCC (wheat) CGCTGCCCCT (barley) CGGCGCGCTC (rice)

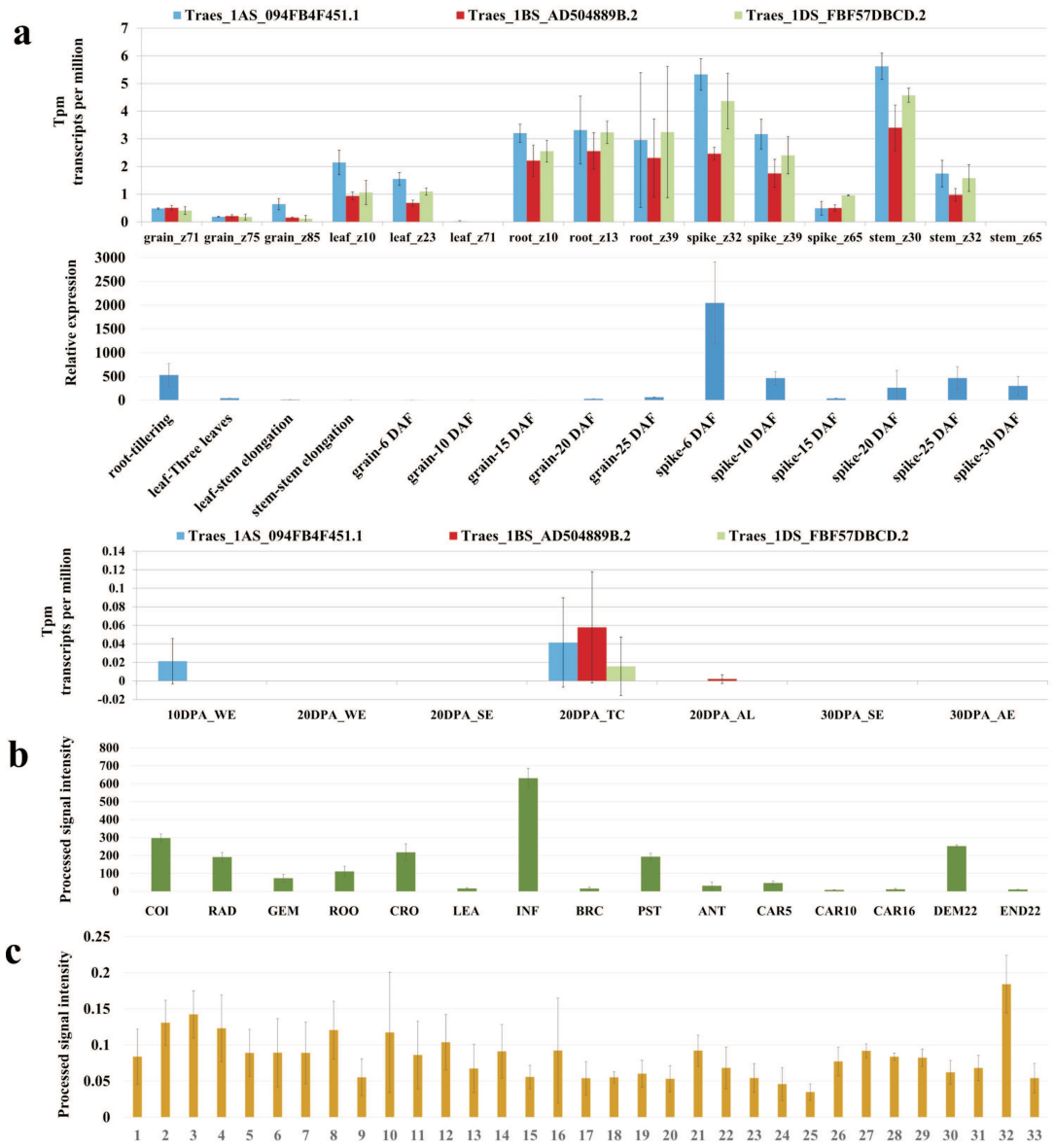


Fig. 4. Expression patterns of *GW5* in wheat (a), barley (b), and rice (c). The y-axis represents the expression value. For a *upper panel*, Zadoks scale for wheat growth stage was adopted and detailed information for each stage was provided in Table S2. For a *middle panel*, expression levels in different stages and tissues by RT-qPCR. For a *lower panel*, 10DPA_WE, 10DPA_Whole_Endosperm; 20DPA_WE, 20DPA_Whole_Endosperm; 20DPA_SE, 20DPA_Starchy_Endosperm, 20DPA_TC, 20DPA_Transfer_Cells; 20DPA_AL, 20DPA_ Aleurone_Layer; 30DPA_SE, 30DPA_Starchy_Endosperm; 30DPA_AE, 30DPA_Aleurone_plus_ Endosperm. For b, symbols are COL for coleoptiles, RAD for radicel (seminal root), GEM for embryo from germinating seed, ROO for 10 cm seedling root, CRO for 10 cm seedling crown, LEA for 10 cm seedling leaf, INF for immature inûorescence, BRC for bracts, PST for pistil, ANT for anther, CAR5 for 5 DAP (days after pollination) caryopsis, CAR10 for 10 DAP caryopsis, CAR16 for 16 DAP caryopsis, DEM22 for 22 DAP embryo, END22 for 22 DAP endosperm. For c, 1, anther at an1; 2, anther at mei1; 3, anther at m1; 4, anther at m2; 5, anther at m3; 6, anther at p1; 7, anther at p2; 8, anther at p3; 9, stigma control; 10, ovary control; 11, pollinating stigma; 12, pollinating ovary; 13, pollen tube growth; 14, fertilization; 15, zygote formation; 16, embryosaccontrol_bottom; 17, 1 DAP embryosac_bottom; 18, 2 DAP embryosac_bottom; 19, 3 DAP embryosac_bottom; 20, 4 DAP embryosac_bottom; 21, embryosaccontrol_top; 22, 1 DAP embryosac_top; 23, 2 DAP embryosac_top; 24, 3 DAP embryosac_top; 25, 4 DAP embryosac_top; 26, growing callus; 27, regenerating callus 2 days; 28, regenerating callus 4 days; 29, regenerating callus 6 days; 30, regenerating callus 8 days; 31, root; 32, shoot; 33, young leaf

detectable node appears (spike_z32) and in stems when the spike was 1 cm in length (stem_z30), and the least expression level of three *GW5* homeologs in grain at various stages (graint_z71, z75, and z85, Fig. 4a *upper panel*). The expression of each of the three *GW5* homeologs in wheat could only be detected in 20DPA_Transfer_Cells (Fig. 4a *lower panel*). Our RT-qPCR analysis indicated similar expression patterns: the *GW5* homeologs were mostly expressed in spikes and roots (Fig. 4a, *middle panel*). Moreover, this expression pattern was verified by analyzing another processed expression database, WheatExp, reported by Pearce et al. (2015) (Suppl. Fig. S1a).

In barley, *HvGW5* was mainly expressed in immature inflorescence (INF) followed by coleoptiles (COL), and the least expression was detected in endosperm (END22) followed by caryopsis at 10 days after pollination (DAP; CAR10; Fig. 4b) by microarray-analysis. In addition, transcriptome expression data analysis showed that *HvGW5* (*HORVU1Hr1G028250.6*) was mainly expressed in rachis (RAC) followed by root tissue (ROO2), substantial expression was also detected in whole developing inflorescence tissue (INF1 and INF2) and embryonic tissue (EMB), the least expression was detected in senescing leaf (SEN) and epidermal strips (EPI) followed by etiolated leaf (ETI; Fig. S1b). In rice, microarray-analysis showed that OsGW5 has the highest expression in anther at mei1 and at m1 and shoot (Fig. 4c). The results in Rice Expression Database (RED) (Lin et al. 2017) showed that *OsGW5* was highly expressed in the pistil (Fig. S1c).

A correlation analysis of the expression patterns of five predicted transcription factors (Suppl. Table S8) and *GW5* in rice, barley, and wheat was conducted as listed in Table S9. Significant correlation between *ABF2* and *OsGW5* only in rice was detected based on microarray data (Suppl. Table S9). In wheat, both drought and heat treatments had no significant effects on the expression of *TaGW5* (Suppl. Fig. S2). No significant changes for the expression of *TaGW5* under the investigated biotic stresses were detected as well (Suppl. Fig. S2). In barley and rice, no significant differences for *GW5* were detected under abiotic and biotic treatments as well (Suppl. Fig. S2).

Discussion

The conserved gene sequence characteristic and protein structure were detected with respect to *GW5* gene among wheat, barley, rice, *B. distachyon* and *T. turgidum*. Similarity of exons and introns by multiple-alignments was detected, which ranged from 65.29 to 88.70 per cent. Inaddition, a highly conserved domain of unknown function (DUF 4005) and two highly conserved IQ domains at GW5 protein were also detected. Highly conserved domains usually play important role in performing the function of the corresponding protein. Thus, in conserved domains, mutations often change the function of a given protein thereby altering the plant phenotype. For instance, a SNP (T/G) was found in the conserved exons of wheat *TaGS5-3A* whose rice ortholog *OsGS5* regulates the thousand grain weight and grain size in rice. Based on the correlation analysis by this SNP marker, it was indicated that the mutant *TaGS5-3A-T* allele significantly increased thousand grain weight and grain size (Ma et al. 2016). These identified conserved domains of *GW5* provide a reference to the dissection of the function of *GW5*. In rice, the *GW5* gene encodes a protein with two IQ domains and an unknown function domain DUF4005. IQ domain-containing proteins usually regulate calmodulin-binding proteins to control plant development and plant stress responses. It was reported that *GW5* may be involved in calcium signaling to regulate grain size in rice (Duan et al. 2017). The protein analysis showed that the studied species have two same IQ domains and an unknown DUF4005 as rice. Thus, it is likely that *GW5* in *H. vulgare*, *A. tauschii*, *T. turgidum*, *T. aestivum* and *B. distachyon* may have similar function similar to *O. sativa*. So far, the clear function of domain DUF4005 is unknown and the effect of its conservative structure on controlling grain size still needs further investigation.

In this study, the detected *OsGW5* was expressed mainly in reproductive organs including anther and pistil, being in accordance with the previous study by RT–qPCR and *pGW5::GUS* reporter gene analyses, wherein it was reported that the highest levels of *GW5* expression was detected in young panicles (Liu et al. 2017). It was verified that *OsGW5* regulates grain size predominantly by influencing cell proliferation in spikelet hulls in rice (Duan et al. 2017; Liu et al. 2017). The study with qPCR and public database based on transcriptome analysis revealed that *GW5* in wheat and barley were detected to be highly expressed in spikes or immature inflorescence indicating their similar expression pattern with its ortholog in rice. Thus, it is presumed that *GW5* in wheat and barley may also have a role in controlling grain size. However, there was no significant effects of abiotic (drought and heat) and biotic treatments on

the expression of *GW5* was in wheat, barley and rice, respectively (Fig. S2). According to the present results, it could not be deduced that *GW5* is expressed selectively in spike and root and nearly no expression was detected in other tissues. As the treatments were conducted against non-spike and non-root organs, it is understandable that no significant differences were detected under the treatments. Further, stress treatments analyses could be carried out against spike and root to check their possible effects on *GW5*.

Transcription factors play a key role in the regulation of gene expression (Sakuma et al. 2006; Ito et al. 2006). In this study, we were aiming at identifying potential transcription factors significantly associated with the *GW5* in wheat, barley and rice by analyzing their expression patterns. Correlation analysis showed that only *ABF2* was detected to be significantly correlated to *GW5* in rice suggesting its possible relationship with *GW5*. It was widely accepted that the prediction and detection of regulatory regions in higher eukaryotes is more complicated than in model organisms with smaller genomes due to the lack of general principles processing the locations of DNA regulatory elements in larger genomes (Bulyk 2003). Here, we used the known data only in *Arabidopsis*, a model plant with smaller genomes to infer the transcription factor binding sites. Therefore, compared to rice with simple genomes, it is more challenging to identify the true regulatory motifs in wheat and barley with complex genomes.

Authors' contribution

Conceptualization of research (JM, JXL); Designing of the experiments (JM, JXL); Contribution of experimental materials (YM, HTP); Execution of field/ lab experiments and data collection (PYD, YJK, WL); Analysis of data and interpretation (YYZ, TL); Preparation of manuscript (JM, PYD).

Declaration

The authors declare no conflict of interest.

Acknowledgements

This publication is based upon work supported by the projects from the National Key Research and Development Program of China (2016YFD0102000), the National Natural Science Foundation of China (31570335), and the Applied Basic Research Programs of Science and Technology Department of Sichuan Province (2016JY0010). We thank Dr. Cristobal Uauy and Dr. Philippa Borrill for providing the wheat transcriptome data. We appreciate the anonymous referees for critical reading of the manuscript.

Reference

Borrill P., Ramirez-Gonzalez R. and Uauy C. 2016. expVIP: A customisable RNA-seq data analysis and visualisation platform opens up gene expression analysis. Plant Physiol., **170**(4): 2172.

Brocklehurst P. A. 1977. Factors controlling grain weight in wheat. Nature, **24** (5600): 348-349.

Brown M. E. and Funk C. C. 2008. Food Security Under Climate Change. Science **319**(5863): 580-581.

Bulyk M. L. 2003. Computational prediction of transcription-factor binding site locations. Genome Biol., **5**(1): 201.

Che R., Tong H., Shi B., Liu Y., Fang S., Liu D., Xiao Y., Hu B., Liu L., Wang H. 2015. Erratum: Control of grain size and rice yield by GL2-mediated brassinosteroid responses. Nat. Plants, **2**(1): 15195.

Choi H. K., Mun J. H., Kim D. J., Zhu H., Baek J. M., Mudge J., Roe B., Ellis N., Doyle J. and Kiss G. B. 2004. Estimating genome conservation between crop and model legume species. Proc. Natl. Acad. Sci. USA, **101**(43): 15289.

Cui F., Zhao C., Ding A., Li J., Wang L., Li X., Bao Y., Li J. and Wang H. 2014. Construction of an integrative linkage map and QTL mapping of grain yield-related traits using three related wheat RIL populations. Theor. Appl. Genet., **127**(3): 659-675.

Dash S., Hemert J. V., Hong L., Wise R. P. and Dickerson J. A. 2012. PLEXdb: Gene expression resources for plants and plant pathogens. Nucleic Acids. Res., **40**: D1194-D1201.

Duan P., Xu J., Zeng D., Zhang B., Geng M., Zhang G., Huang K., Huang L., Xu R., Ge S. 2017. Natural Variation in the Promoter of *GSE5* Contributes to Grain Size Diversity in Rice. Mol. Plant, **10**(5): 685-694.

Gegas V. C., Nazari A., Griffiths S., Simmonds J., Fish L., Orford S., Sayers L., Doonan J. H. and Snape J. W. 2010. A genetic framework for grain size and shape variation in wheat. Plant Cell, **22**(4): 1046-1056.

Hackett R. M., Cadwallader G. and Franklin F. C. H. 1996. Functional Analysis of a Brassica oleracea *SlR7* Gene Promoter. Plant Physiol., **112**(4): 1601-1607.

Hu Z., Lu S. J., Wang M. J., He H., Sun L., Wang H., Liu X. H., Jiang L., Sun J. L. and Xin X. 2018. A Novel QTL q*TGW3* Encodes the GSK3/SHAGGY-Like Kinase OsGSK5/OsSK41 that Interacts with OsARF4 to Negatively Regulate Grain Size and Weight in Rice. Mol Plant Doi:10.1016/j.molp.2018.03.005.

Ishimaru K., Hirotsu N., Madoka Y., Murakami N., Hara N., Onodera H., Kashiwagi T., Ujiie K., Shimizu B. and Onishi A. 2013. Loss of function of the IAA-glucose

hydrolase gene *TGW6* enhances rice grain weight and increases yield. Nat. Genet., **45**(6): 707.

Ito Y., Katsura K., Maruyama K., Taji T., Kobayashi M., Seki M., Shinozaki K. and Yamaguchishinozaki K. 2006. Functional Analysis of Rice DREB1/CBF-type Transcription Factors Involved in Cold-responsive Gene Expression in Transgenic Rice. Plant Cell Physiol., **47**(1): 141-153.

The International Wheat Genome Sequencing Consortium (IWGSC). 2014. A chromsome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. Science, **345**(6194): 1251788.

Jiang Q. T., Liu T., Ma J., Wei Y. M., Lu Z. X., Lan X. J., Dai S. F. and Zheng Y. L. 2011. Characterization of barley *Prp1* gene and its expression during seed development and under abiotic stress. Genetica, **139**(10): 1283-1292.

Kumar S., Stecher G. and Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol. Biol. Evol., **33**(7): 1870.

Lescot M., Déhais P., Thijs G., Marchal K., Moreau Y., Van dP. Y., Rouzé P. and Rombauts S. 2002. PlantCARE, a database of plant cis-acting regulatory elements and a portal to tools for in silico analysis of promoter sequences. Nucleic Acids Res., **30**(1): 325-327.

Li M., Wang Z., Liang Z., Shen W., Sun F., Xi Y. and Liu S. 2015. Quantitative trait loci analysis for kernel-related characteristics in common wheat (*Triticum aestivum* L.). Crop Sci., **326** (5958): 1357-1358.

Lin X., Dong Z., Jian S., Xu X., Yin H., Li M., Wu S., Hu S., Hao L. and Zhang Z. 2017. Rice Expression Database (RED): An integrated RNA-Seq-derived gene expression database for rice. J. Genet. Genomics, **44**(5): 235-241.

Liu J., Chen J., Zheng X., Wu F., Lin Q., Heng Y., Tian P., Cheng Z., Yu X. and Zhou K. 2017. *GW5* acts in the brassinosteroid signalling pathway to regulate grain width and weight in rice. Nat. Plants, **3**: 17043.

Long X.-Y., Wang J.-R., Ouellet T., Rocheleau H., Wei Y.-M., Pu Z.-E., Jiang Q.-T., Lan X.-J. and Zheng Y.-L. 2010. Genome-wide identification and evaluation of novel internal control genes for Q-PCR based transcript normalization in wheat. Plant Mol. Biol., **74**(3): 307-311.

Ma J., Ding P., Qin P., Liu Y. X., Xie Q., Chen G., Li W., Jiang Q., Chen G. and Lan X. J. 2017. Structure and expression of the *TaGW7* in bread wheat (*Triticum aestivum* L.). Plant Growth Regul., **6**: 1-11.

Ma J., Qin N., Cai B., Chen G., Ding P., Zhang H., Yang C., Huang L., Mu Y., Tang H., Liu Y., Wang J., Qi P., Jiang Q., Zheng Y., Liu C., Lan X. and Wei Y. 2019. Identification and validation of a novel major QTL for all-stage stripe rust resistance on 1BL in the winter wheat line 20828. Theor. Appl. Genet., **132**: 1363-137.

Ma J., Stiller J., Berkman P. J., Wei Y., Rogers J., Feuillet C., Dolezel J., Mayer K. F., Eversole K. and Zheng Y. L. 2013. Sequence-based analysis of translocations and inversions in bread wheat (*Triticum aestivum* L.). Plos One, **8**(11): e79329.

Ma L., Li T., Hao C., Wang Y., Chen X. and Zhang X. 2016. *TaGS5-3A*, a grain size gene selected during wheat improvement for larger kernel and yield. Plant Biotech. J., **14**(5): 1269-1280.

Murray M. G. and Thompson W. F. 1980. Rapid isolation of high molecular weight plant DNA. Nucleic Acids Res., **8**(19): 4321-4325.

Pearce S., Vazquezgross H., Herin S. Y., Hane D., Wang Y., Gu Y. Q. and Dubcovsky J. 2015. WheatExp: an RNA-seq expression database for polyploid wheat. BMC Plant Biol., **15**(1): 299.

Sakuma Y., Maruyama K., Osakabe Y., Qin F., Seki M., Shinozaki K. and Yamaguchishinozaki K. 2006. Functional Analysis of an Arabidopsis Transcription Factor, *DREB2A*, Involved in Drought-Responsive Gene Expression. Plant Cell, **18**(5): 1292-1309.

Simmonds J., Scott P., Brinton J., Mestre T. C., Bush M., Blanco A. D., Dubcovsky J. and Uauy C. 2016. A splice acceptor site mutation in *TaGW2-A1* increases thousand grain weight in tetraploid and hexaploid wheat through wider and longer grains. Theor. Appl. Genet., **129**(6): 1099-1112.

Su Z., Jin S., Lu Y., Zhang G., Chao S. and Bai G. 2016. Single nucleotide polymorphism tightly linked to a major QTL on chromosome 7A for both kernel length and kernel weight in wheat. Mol. Breeding., **36**(2): 1-11.

Sun L., Li X., Fu Y., Zhu Z., Tan L., Liu F., Sun X., Sun X. and Sun C. 2013. *GS6*, A Member of the GRAS Gene Family, Negatively Regulates Grain Size in Rice. Mol. Plant, **55**(10): 938-949.

Sun X. Y., Ke W., Yan Z., Kong F. M., Han G. Z., Jiang H. M., Huang X. J., Li R. J., Wang H. G. and Li S. S. 2009. QTL analysis of kernel shape and weight using recombinant inbred lines in wheat. Euphytica, **165**(3): 615.

Takeda S. and Matsuoka M. 2008. Genetic approaches to crop improvement: responding to environmental and population changes. Nat. Rev. Genet., **9**(6): 444-457.

Thompson J. D., Higgins D. G. and Gibson T. J. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res., **22**(22): 4673-4680.

Vogel J. P., Garvin D. F., Mockler T. C., Schmutz J., Dan R., Bevan M. W., Barry K., Lucas S., Harmonsmith M. and Lail K. 2010. Genome sequencing and analysis of the model grass *Brachypodium distachyon*. Hereditas, **463**(7282): 763.

Wang S., Zhang X., Chen F. and Cui D. 2015. A Single-Nucleotide Polymorphism of *TaGS5* Gene Revealed its Association with Kernel Weight in Chinese Bread Wheat. Front Plant Sci., **6.**

Weng J., Gu S., Wan X., Gao H., Guo T., Su N., Lei C., Zhang X., Cheng Z. and Guo X. 2008. Isolation and initial characterization of *GW5*, a major QTL associated with rice grain width and weight. Cell Res., **18**(12): 1199.

Wu Q. H., Chen Y. X., Zhou S. H., Fu L., Chen J. J., Xiao Y., Zhang D., Ouyang S. H., Zhao X. J. and Cui Y. 2015. High-density genetic linkage map construction and QTL mapping of grain shape and size in the wheat population Yanda1817 × Beinong6. Plos One, **10**(2): e0118144.

Xia D., Zhou H., Liu R., Dan W., Li P., Wu B., Chen J., Wang L., Gao G. and Zhang Q. 2018. *GL3.3*, a Novel QTL Encoding a GSK3/SHAGGY-like Kinase, Epistatically Interacts with *GS3* to Form Extra-long Grains in Rice. Mol. Plant, Doi: 10.1016/j.molp. 2018.03.006.

Xia L., Zou D., Sang J., Xu X., Yin H., Li M., Wu S., Hu S., Hao L. and Zhang Z. 2017. Rice Expression Database (RED): An integrated RNA-Seq-derived gene expression database for rice. J. Genet. Genomics, **44**(5): 235-241.

Xu C., Yu L., Li Y., Xu X., Xu C., Li X., Xiao J. and Zhang Q. 2015. Differential expression of *GS5* regulates grain size in rice. J. Exp. Bot., **66**(9): 2611.

Yang Z., Bai Z., Li X., Wang P., Wu Q., Yang L., Li L. and Li X. 2012. SNP identification and allelic-specific PCR markers development for *TaGW2*, a gene linked to wheat kernel weight. Theor. Appl. Genet., **125**(5): 1057-1068.

Ying J. Z., Ma M., Bai C., Huang X., Liu J. L., Fan Y. Y. and Song X. J. 2018. *TGWD*, a Major QTL that Negatively Modulates Grain Length and Weight in Rice. Mol. Plant. Doi: 10.1016/j.molp.2018.03.007.

Yu M., Mao S. L., Hou O. B., Chen G. Y., Pu Z. E., Li W., Lan X. J., Jiang Q. T., Liu Y. X., Deng M., Wei Y. M. 2018. Analysis of contributors to grain yield in wheat at the individual quantitative trait locus level. Plant Breed., **137**: 35-49.

Zadoks J. C., Chang T. T. and Konzak C. F. 2010. A decimal code for the growth stages of cereals. Weed Res., **14**(6): 415-421.

Zhang H., Ma J., Liu J., Mu Y., Tang H., Liu Y., Chen G., Jiang Q., Chen G., Wei Y., Zheng Y. and Lan X. 2018. Molecular characterization of the *TaWTG1* in bread wheat (*Triticum aestivum* L.). Gene, **678**: 23-32.

Zhao D. S., Li Q. F., Zhang C. Q., Zhang C., Yang Q. Q., Pan L. X., Ren X. Y., Lu J., Gu M. H. and Liu Q. Q. 2018. *GS9* acts as a transcriptional activator to regulate rice grain shape and appearance quality. Nat. Commun., **9**(1): 1240.

Supplementary Table S1. Information of reference sequences for *GW5* retrieved from public database

Species	Genomic DNA						
	Data source or database	Chromosome or contig No.	Promoter locus**	ATG locus	TAG locus	Data source or database	Accession NO.
Tritcium aestivum	http://wheat-urgi.versailles.inra.fr/	chr1A	*	140961020	140962582	http://wheat-urgi.versailles.inra.fr/	TraesCS1A01G122900.1
		chr1B	*	190630415	190632007		TraesCS1B01G142100.1
		chr1D	*	126472959	126474550		TraesCS1D01G123800.1
Triticum turgidum	https://www.dropbox.com/sh/3dm05grokhl0nbv/AAC3wvlYmAher8fY0srX3gX9a?dl=0%22	chr1A	-	142379764	142381359	https://www.dropbox.com/sh/3dm05grokhl0nbv/AAC3wvlYmAher8fY0srX3gX9a?dl=0%22	TRIDC1AG017640.1 IQ-domain 26
		chr1B		185320200	185321816		TRIDC1BG021520.1 IQ-domain 26
Aegilops tauschii	https://www.ncbi.nlm.nih.gov/assembly/GCA_002575655.1/#/def_asm_Primary_Assembly	chr1D	-	129213269	129214860	https://www.ncbi.nlm.nih.gov/assembly/GCA_002575655.1/#/def_asm_Primary_Assembly	AET1Gv20304400.4
Hordeum vulgare	ftp://ftp.ensemblgenomes.org/pub/plants/release-32/fasta/	chr1H	*	151342637	151340811	http://webblast.ipk-gatersleben.de/barley_ibsc/downloads/	HORVU1Hr1G028250.6
Brachypodium distachyon	http://www.phytozome.org/	Bd2	-	33465745	33464104	http://www.phytozome.org/	Bradi2g33370.1
Oryza sativa	http://www.phytozome.org/	Chr5	*	5365122	5366701	http://www.phytozome.org/	Os05g09520.1

**: promoters of *Triticum aestivum*, *Hordeum vulgare*, and *Oryza sativa* only were used for promoter analysis.

Supplementary Table S2. Zadoks scale table*

Stage	Wheat growth stage	Zadoks scale	Leaves	Root	Stem	Spike	Grain
Seedling	First leaf through coleoptile	10	x	x			
Three leaves	3 leaves unfolded	13		x			
Three tillers	Main shoot and 3 tillers	23	x				
Spike at 1 cm	Pseudostem erection	30			x		
Two nodes	2nd detectable node	32			x	x	
Meiosis	Flag leaf ligule and collar visible	39		x		x	
Anthesis	1/2 of flowering complete	65			x	x	
2 DAA (50 ℃.days)	Kernel (caryopsis) watery ripe	71	x				x
14 DAA (350 ℃.days)	Medium Milk	75					x
30 DAA (700 ℃.days)	Soft dough	85					x

* The table was cited from "Zadoks JC, Chang TT, Konzak CF. (1974) A decimal code for the growth stages of cereals. Weed Research 14: 415-21." "x" represents the tissue used for expression analysis in this study.

Supplementary Table S3. Comparison of similarity and size (bp) for *GW5* exons and introns among different species

Exon or intron	wheat (1AS)	wheat (1BS)	wheat (1DS)	*T. turgidum* (1A)	*T. turgidum* (1B)	*A. tauschii*	barley	*B. distachyon*	rice	Exon similarity (%)	Intron similarity (%)
exon1	434	434	434	467	458	434	647	446	443	65.29	
intron1	77	107	107	77	107	107	129	140	86		64.64
exon2	207	207	207	232	207	207	207	201	1051	86.42	
intron2	121	121	120	100	121	120	114	83	-		78.34
exon3	724	724	724	720	724	724	730	772	-	88.70	
exon total	1365	1365	1365	1419	1389	1365	1584	1419	1494	79.47	
intron total	198	228	227	177	228	227	243	223	86		75.44
Total	1563	1593	1592	1596	1617	1592	1827	1642	1580		

'-' indicates that the corresponding sequences are absent.

Supplementary Table S4. Identified domains of GW5 proteins in different species*

Species	Domain Name	Start	End	E-value	Sequence
Brachypodium distachyon	low complexity	49	74	N/A	AEAAAAAAATAQQGGNAAIARAAEAA
Brachypodium distachyon	low complexity	91	116	N/A	AIAVAAATAAAADAAVAAAHAAVAVV
Brachypodium distachyon	IQ	133	155	1.49	GPAAAAVRIQTAFRGFLAKKALR
Brachypodium distachyon	IQ	156	177	6.59	ALKALVKLQALVRGYLVRKQAA
Brachypodium distachyon	low complexity	188	201	N/A	RAQAAMRAHRAGAA
Brachypodium distachyon	Pfam:DUF4005	263	447	0.00000012	PKSRSSSRRASSPLQLDPCDEYWCANNSSNNPMSSPLLPPARIAVAA
					PTPRHGHFPEYDWCAMEKARPATAQSTPRYMSINFNANNNAPATP
					TKSVCGAGGYLYSSLNCPGYMSSTQSFEAKTRSHSAPKQRPEPPPAN
					GRRQRVPLSEVVVVESSRASLSGAVGMQRSCNRASTTQQEAFNFKT
Hordeum vulgare	low complexity	24	36	N/A	PTPSTRPLPLPSL
Hordeum vulgare	low complexity	117	142	N/A	AEAAAVAAATPAHGGNAAMARAAEAA
Hordeum vulgare	low complexity	159	184	N/A	AIAVAAATQAAADAAVAAAHAAVAVV
Hordeum vulgare	IQ	200	222	1.49	GPAAAAVRIQTAFRGFLAKKALR
Hordeum vulgare	IQ	223	244	6.59	ALKALVKLQALVRGYLVRKQAA
Hordeum vulgare	low complexity	303	314	N/A	SRRLSASIESSS
Hordeum vulgare	Pfam:DUF4005	320	500	4.20E-09	SPKIVEMDTGRPKSRSSSRRASSPLLDPCEEWCAAANPMASPLLPCH
					MPGGAPPRIAVPTPGHLPEYDWCAMEKARPATAQCTPRYMNTPAT
					PTKSVCGGGGYSASSLLNCPSYMSSTQSFEAKVRSHSAPKQRPEPPA
					AASTNRKRVPLSEVVVVESRASLSGVGMQRSCNRVEEAFNFK
Oryza sativa	low complexity	41	68	N/A	AEAAAAAAAAAAAEASGGNAAIARAAEAA
Oryza sativa	low complexity	85	110	N/A	AIAVAAATAAAADAAVAAAQAAVAVV
Oryza sativa	IQ	132	154	0.0707	SLAAAAVRIQTAFRGFLAKKALR
Oryza sativa	IQ	155	176	8	ALKALVKLQALVRGYLVRRQAA
Oryza sativa	low complexity	198	212	N/A	GAGAAANLPHLHHAP
Oryza sativa	low complexity	265	276	N/A	SRRLSASIESSS
Oryza sativa	low complexity	295	305	N/A	SRSSSSRRASS

Oryza sativa	Pfam:DUF4005	355	471	2.90E-15	DWCALEKARPATAQSTPRYAHAPPTPTKSVCGGGGGGGIHSS PLNCPNYMSNTQSFEAKVRSQSAPKQRPETGGAGAGGGRKRV PLSEVVVVESRASLSGVGMQRSCNRVQEAFNFK
Triticum turgidum-1A	low complexity	56	68	N/A	SAEAAATAAATSA
Triticum turgidum-1A	low complexity	74	82	N/A	AIARAAEAA
Triticum turgidum-1A	low complexity	99	124	N/A	AIAVAAATQAAADAAVAAAHAAVAVV
Triticum turgidum-1A	IQ	140	162	1.49	GPAAAAVRIQTAFRGFLAKKALR
Triticum turgidum-1A	IQ	163	184	6.59	ALKALVKLQALVRGYLVRKQAA
Triticum turgidum-1A	low complexity	250	261	N/A	SRRLSASIESSS
Triticum turgidum-1B	low complexity	52	79	N/A	SSRDSAEAAAATSARGGNAAIARAAEAA
Triticum turgidum-1B	low complexity	96	121	N/A	AIAVAAATQAAADAAVAAAHAAVAVV
Triticum turgidum-1A	Pfam:DUF4005	277	445	0.0000013	RPKSRSSSRRASSPLLDPCEEWCAAANPMSSPLLLPCHMPGGAPPRI AVPTPRHLPEYDWCAMEKARPATAQCTPRYMNANAPATPTKSVC GGGYSSSSLLNCPSYMSSTQSFEAKVRSHSAPKQRPEPPTNRKRVPL SEVVVVESRASLSGVGMQRSCNRVEEAFNFK
Triticum turgidum-1B	IQ	137	159	2.99	GPAAAAVRIQTTFRGFLAKKALR
Triticum turgidum-1B	IQ	160	181	6.59	ALKALVKLQALVRGYLVRKQAA
Triticum turgidum-1B	low complexity	240	251	N/A	SRRLSASIESSS
Triticum turgidum-1B	Pfam:DUF4005	269	435	9.40E-07	KSRSSSRRASSPLLDPCEEWCAATNPMSSPLLLPCHMPGGAPPRIAV PTPRHLPEYDWCAMEKARPATAQCTPRYMNANAPATPTKSVCGSG YSSSSLLNCPSYMSSTQSFEAKVRSHSAPKQRPEPPTNRKRVPLSEVV VVESRASLSGVGMQRSCNRVEEAFNFK
Tritcium aestivum-1A	low complexity	45	57	N/A	SAEAAATAAATSA
Tritcium aestivum-1A	low complexity	63	71	N/A	AIARAAEAA
Tritcium aestivum-1A	low complexity	88	113	N/A	AIAVAAATQAAADAAVAAAHAAVAVV
Tritcium aestivum-1A	IQ	129	151	1.49	GPAAAAVRIQTAFRGFLAKKALR
Tritcium aestivum-1A	IQ	152	173	6.59	ALKALVKLQALVRGYLVRKQAA
Tritcium aestivum-1A	low complexity	232	243	N/A	SRRLSASIESSS

Tritcium aestivum-1A	Pfam:DUF4005	249	427	2.90E-07	SPKIVEMDTGRPKSRSSSRRASSPLLDPCEEWCAAANPMSSPLLLPCH MPGGAPPRIAVPTPRHLPEYDWCAMEKARPATAQCTPRYMNANAPA TPTKSLCGGGYSSSSLLNCPSYMSSTQSFEAKVRSHSAPKQRPEPPTNR KRVPLSEVVVVESRASLSGVGMQRSCNRVEEAFNFK
Tritcium aestivum-1B	low complexity	41	71	N/A	SSRDSAEAAATAAATSARGGNAAIARAAEAA
Tritcium aestivum-1B	low complexity	88	113	N/A	AIAVAAATQAAADAAVAAAHAAVAVV
Tritcium aestivum-1B	IQ	129	151	1.49	GPAAAAVRIQTAFRGFLAKKALR
Tritcium aestivum-1B	IQ	152	173	6.59	ALKALVKLQALVRGYLVRKQAA
Tritcium aestivum-1B	low complexity	232	243	N/A	SRRLSASIESSS
Tritcium aestivum-1B	Pfam:DUF4005	249	427	6.00E-07	SPKIVEMDTGRPKSRSSSRRASSPLLDPCEEWCAATNPMSSPLLLPCHM PGGAPPRIAVPTPRHLPEYDWCAMEKARPATAQCTPRYMNANAPATP TKSVCGSGYSSSSLLNCPSYMSSTQSFEAKVRSHSAPKQRPEPPTNRKR VPLSEVVVVESRASLSGVGMQRSCNRVEEAFNFK
Tritcium aestivum-1D	low complexity	46	71	N/A	AEAAAAAAATSARGGNAAIARAAEAA
Tritcium aestivum-1D	low complexity	88	113	N/A	AIAVAAATQAAADAAVAAAHAAVAVV
Tritcium aestivum-1D	IQ	129	151	1.49	GPAAAAVRIQTAFRGFLAKKALR
Tritcium aestivum-1D	IQ	152	173	6.59	ALKALVKLQALVRGYLVRKQAA
Tritcium aestivum-1D	low complexity	232	243	N/A	SRRLSASIESSS
Tritcium aestivum-1D	Pfam:DUF4005	249	427	2.90E-07	SPKIVEMDTGRPKSRSSSRRASSPLLDPCEEWCAAANPMSSPLLLPC HMPGGAPPRIAVPTPRHLPEYDWCAMEKARPATAQCTPRYMNAN APATPTKSVCGGGYSSSSLLNCPSYMSSTQSFEAKVRSHSAPKQRPE PPTNRKRVPLSEVVVVESRASLSGVGMQRSCNRVEEAFNFK

*, The identification of domains was conducted by SMART on http://smart.embl-heidelberg.de/smart/set_mode.cgi?NORMAL=1#

Note: the amino acids in blue font are the highly conserved domain of DUF4005, those in green font are the highly conserved IQ domains and in orange font are the low complexity regions shown in Fig 3.

Supplementary Table S5. Putative motifs of *GW5* promoters in barley, rice, and wheat

Species	Motifs name	Function
barley	5UTR Py-rich stretch	cis-acting element conferring high transcription levels
	A-box	cis-acting regulatory element sequence conserved in alpha-amylase promoters
	AAGAA-motif	unknown
	ABRE	cis-acting element involved in the abscisic acid responsiveness
	ARE	cis-acting regulatory element essential for the anaerobic induction
	ATCT-motif	part of a conserved DNA module involved in light responsiveness
	Box II	part of a light responsive element
	CAAT-box	common cis-acting element in promoter and enhancer regions
	CATT-motif	part of a light responsive element
	CCAAT-box	MYBHv1 binding site
	CCGTCC-box	cis-acting regulatory element related to meristem specific activation
	CGTCA-motif	cis-acting regulatory element involved in the MeJA-responsiveness
	G-Box	cis-acting regulatory element involved in light responsiveness
	GAG-motif	part of a light responsive element
	GC-motif	enhancer-like element involved in anoxic specific inducibility
	MNF1	light responsive element
	O2-site	cis-acting regulatory element involved in zein metabolism regulation
	P-box	gibberellin-responsive element
	RY-element	cis-acting regulatory element involved in seed-specific regulation
	SARE	cis-acting element involved in salicylic acid responsiveness
	Skn-1_motif	cis-acting regulatory element required for endosperm expression
	Sp1	light responsive element
	TATA-box	core promoter element around -30 of transcription start
	TCCC-motif	part of a light responsive element
	TGACG-motif	cis-acting regulatory element involved in the MeJA-responsiveness
	TGG-motif	part of a light responsive element
	box S	none
	chs-CMA1a	part of a light responsive element
	circadian	cis-acting regulatory element involved in circadian control
	3-AF1 binding site	light responsive element
rice	5UTR Py-rich stretch	cis-acting element conferring high transcription levels
	A-box	cis-acting regulatory element
	AAGAA-motif	unknown
	ABRE	cis-acting element involved in the abscisic acid responsiveness

	ARE	cis-acting regulatory element essential for the anaerobic induction
	AT-rich element	binding site of AT-rich DNA binding protein (ATBP-1)
	Box 4	part of a conserved DNA module involved in light responsiveness
	CAAT-box	common cis-acting element in promoter and enhancer regions
	CATT-motif	part of a light responsive element
	CCGTCC-box	cis-acting regulatory element related to meristem specific activation
	CGTCA-motif	cis-acting regulatory element involved in the MeJA-responsiveness
	G-Box	cis-acting regulatory element involved in light responsiveness
	GAG-motif	part of a light responsive element
	GARE-motif	gibberellin-responsive element
	GC-motif	unknown
	HD-Zip 3	protein binding site
	HSE	cis-acting element involved in heat stress responsiveness
	LTR	cis-acting element involved in low-temperature responsiveness
	MBS	MYB binding site involved in drought-inducibility
	Skn-1_motif	cis-acting regulatory element required for endosperm expression
	Sp1	light responsive element
	TATA-box	core promoter element around -30 of transcription start
	TCA-element	cis-acting element involved in salicylic acid responsiveness
	TGA-element	auxin-responsive element
	TGACG-motif	cis-acting regulatory element involved in the MeJA-responsiveness
	3-AF1 binding site	light responsive element
	A-box	cis-acting regulatory element
wheat-1A	AAGAA-motif	unknown
	ABRE	cis-acting element involved in the abscisic acid responsiveness
	ATC-motif	part of a conserved DNA module involved in light responsiveness
	Box 4	part of a conserved DNA module involved in light responsiveness
	Box II	part of a light responsive element
	CAAT-box	common cis-acting element in promoter and enhancer regions
	CCAAT-box	MYBHv1 binding site
	CCGTCC-box	cis-acting regulatory element related to meristem specific activation
	CGTCA-motif	cis-acting regulatory element involved in the MeJA-responsiveness
	G-Box	cis-acting regulatory element involved in light responsiveness
	GAG-motif	part of a light responsive element
	GATA-motif	part of a light responsive element
	I-box	part of a light responsive element
	P-box	gibberellin-responsive element
	RY-element	cis-acting regulatory element involved in seed-specific regulation

	Skn-1_motif	cis-acting regulatory element required for endosperm expression
	Sp1	light responsive element
	TATA-box	core promoter element around -30 of transcription start
	TCA-element	cis-acting element involved in salicylic acid responsiveness
	TCCC-motif	part of a light responsive element
	TGACG-motif	cis-acting regulatory element involved in the MeJA-responsiveness
	TGG-motif	part of a light responsive element
	circadian	cis-acting regulatory element involved in circadian control
wheat-1B	3-AF1 binding site	light responsive element
	A-box	cis-acting regulatory element
	AAGAA-motif	unknown
	ABRE	cis-acting element involved in the abscisic acid responsiveness
	ACE	cis-acting element involved in light responsiveness
	ARE	cis-acting regulatory element essential for the anaerobic induction
	Box 4	part of a conserved DNA module involved in light responsiveness
	Box II	part of a light responsive element
	CAAT-box	common cis-acting element in promoter and enhancer regions
	CCAAT-box	MYBHv1 binding site
	CCGTCC-box	cis-acting regulatory element related to meristem specific activation
	CGTCA-motif	cis-acting regulatory element involved in the MeJA-responsiveness
	G-Box	cis-acting regulatory element involved in light responsiveness
	GA-motif	part of a light responsive element
	GAG-motif	part of a light responsive element
	GATA-motif	part of a light responsive element
	GC-motif	enhancer-like element involved in anoxic specific inducibility
	I-box	part of a light responsive element
	O2-site	cis-acting regulatory element involved in zein metabolism regulation
	P-box	gibberellin-responsive element
	RY-element	cis-acting regulatory element involved in seed-specific regulation
	Skn-1_motif	cis-acting regulatory element required for endosperm expression
	Sp1	light responsive element
	TATA-box	core promoter element around -30 of transcription start
	TCA-element	cis-acting element involved in salicylic acid responsiveness
	TCCC-motif	part of a light responsive element
	TGACG-motif	cis-acting regulatory element involved in the MeJA-responsiveness
	TGG-motif	part of a light responsive element
	circadian	cis-acting regulatory element involved in circadian control

wheat-1D	3-AF1 binding site	light responsive element
	A-box	cis-acting regulatory element
	AAGAA-motif	unknown
	ABRE	cis-acting element involved in the abscisic acid responsiveness
	ACE	cis-acting element involved in light responsiveness
	ARE	cis-acting regulatory element essential for the anaerobic induction
	ATC-motif	part of a conserved DNA module involved in light responsiveness
	Box II	part of a light responsive element
	CAAT-box	common cis-acting element in promoter and enhancer regions
	CCAAT-box	MYBHv1 binding site
	CCGTCC-box	cis-acting regulatory element related to meristem specific activation
	CGTCA-motif	cis-acting regulatory element involved in the MeJA-responsiveness
	G-Box	cis-acting regulatory element involved in light responsiveness
	GAG-motif	part of a light responsive element
	GC-motif	enhancer-like element involved in anoxic specific inducibility
	I-box	part of a light responsive element
	LTR	cis-acting element involved in low-temperature responsiveness
	MNF1	light responsive element
	O2-site	cis-acting regulatory element involved in zein metabolism regulation
	P-box	gibberellin-responsive element
	RY-element	cis-acting regulatory element involved in seed-specific regulation
	Skn-1_motif	cis-acting regulatory element required for endosperm expression
	Sp1	light responsive element
	TATA-box	core promoter element around -30 of transcription start
	TCA-element	cis-acting element involved in salicylic acid responsiveness
	TCCC-motif	part of a light responsive element
	TGA-element	auxin-responsive element
	TGACG-motif	cis-acting regulatory element involved in the MeJA-responsiveness
	TGG-motif	part of a light responsive element
	chs-CMA1a	part of a light responsive element
	circadian	cis-acting regulatory element involved in circadian control

Table S6 Putative transcription factors of *GW5* promoters in wheat, barley and rice

	Matrix ID	Name	Predicted sequence
wheat-1AS	MA0123.1	abi4	CGCTGCCCCC
	MA0931.1	ABI5	GTACACGTGC
	MA0941.1	ABF2	GGGTACACGTGCA
	MA0564.1	ABI3	TCGCATGCT
	MA0005.1	AG	CCTATTTAAGC
	MA0930.1	ABF3	GCACGTGT
	MA1244.1	ABR1	CCGGCGAGAGGCGGCGGTG
wheat-1BS	MA0123.1	abi4	CGCTGCCCCC
	MA0564.1	ABI3	TTGCATGCT
	MA0005.1	AG	CCTATTTAAGC
	MA0931.1	ABI5	ACACACGTAG
	MA0941.1	ABF2	TAACACACGTAGA
	MA1277.1	Adof1	AACGATAAAAAGAAAAGAGAA
	MA0005.2	AG	TTTTCCAAGTAAGGTTTT
	MA1244.1	ABR1	CCGGCGAGAGGCGGCGGTG
wheat-1DS	MA0123.1	abi4	CGCTGCCCCC
	MA0931.1	ABI5	GTACACGTGC
	MA0941.1	ABF2	GGGTACACGTGCA
	MA0564.1	ABI3	TCGCATGCT
	MA0005.1	AG	CCTATTTAAGC
	MA0930.1	ABF3	GCACGTGT
	MA0005.2	AG	TTTTCCAAGTAAGGTTTT
	MA1244.1	ABR1	CCGGCGAGAGGCGGCGGTG
barley	MA0123.1	abi4	CGCTGCCCCT
	MA0564.1	ABI3	CTGCATGCC
	MA0931.1	ABI5	GCACACGTGC
	MA0005.1	AG	CCTATTTAAGC
	MA0941.1	ABF2	GGGCACACGTGCA
	MA0930.1	ABF3	ACACGTGC
	MA0005.2	AG	ACTACCTATTTAAGCAAC
	MA1277.1	Adof1	AGAAATAAAAAGAAATGGATT
rice	MA1277.1	Adof1	AAAACGAAAAAGAAAAAGAAG
	MA0123.1	abi4	CGGCGCGCTC
	MA0564.1	ABI3	GTGCATGCG
	MA0941.1	ABF2	AGTACCACGTAAG
	MA0005.2	AG	ATTACCATTTCAGTTAGT
	MA0005.1	AG	CCTAATTAATT
	MA0931.1	ABI5	TACCACGTAA
	MA1244.1	ABR1	CCAGAAATCGGCGGCGGCC

Supplementary Table S7. Predicted transcription factors in given species

	Description[#]	Location[#]	Aliases[#]	Arabidopsis thaliana[&]	wheat*	barley*	rice*
ABF2	abscisic acid responsive elements-binding factor 2 [Arabidopsis thaliana (thale cress)]	Chromosome 1, NC_003070.9 (17165125. 17167924, complement)	AT1G45249, ABSCISIC ACID RESPONSIVE ELEMENTS-BINDING PROTEIN 2, AREB1, ATAREB1, AtABF2, abscisic acid responsive elements-binding factor 2	AT1G45249.1	TraesCS1D01G306000.1	HORVU3Hr1G084360.2	LOC_Os02g52780.1
ABI5	Basic-leucine zipper (bZIP) transcription factor family protein [Arabidopsis thaliana (thale cress)]	Chromosome 2, NC_003071.7 (15204659. 15207636, complement)	AT2G36270, ABA INSENSITIVE 5, AtABI5, F2H17.12, F2H17_12, GIA1, GROWTH-INSENSITIVITY TO ABA 1	AT2G36270.1	TraesCS3D01G364900.1	HORVU3Hr1G084360.2	LOC_Os01g64000.1
abi4	Integrase-type DNA-binding superfamily protein [Arabidopsis thaliana (thale cress)]	Chromosome 2, NC_003071.7 (16796247. 16797585, complement)	AT2G40220, ABA INSENSITIVE 4, ABSCISIC ACID-INSENSITIVE PROTEIN 4, ATABI4, GIN6, GLUCOSE INSENSITIVE 6, IMPAIRED SUCROSE INDUCTION 3, ISI3, SALOBRENO 5, SAN5, SIS5, SUCROSE UNCOUPLED 6, SUGAR-INSENSITIVE 5, SUN6, T7M7.16	AT2G40220.1	TraesCS1A01G223400.1 TraesCS1B01G236700.1 TraesCS1D01G225000.1	HORVU1Hr1G060060.1	LOC_Os05g28350.1

ABI3	AP2/B3-like transcriptional factor family protein [Arabidopsis thaliana (thale cress)]	Chromosome 3, NC_003074.8 (8997370. 9001185)	AT3G24650, ABA INSENSITIVE 3, ABSCISIC ACID INSENSITIVE 3, AtABI3, SIS10, SUGAR INSENSITIVE 10	AT1G49720.1	TraesCS3B01G452200.1	HORVU3Hr1G092690.2	LOC_Os01g68370.3
AG	K-box region and MADS-box transcription factor family protein [Arabidopsis thaliana (thale cress)]	Chromosome 4, NC_003075.7 (10382856. 10388539)	AT4G18960, AGAMOUS, AGAMOUS PROTEIN, F13C5.130, F13C5_130	AT4G18960.1	TraesCS3D01G140200.2 TraesCS3A01G314300.1 TraesCS3B01G157500.4 TraesCS1A01G125800.1 TraesCS1D01G127700.1 TraesCS1B01G144800.1	HORVU3Hr1G026650.1	LOC_Os01g10504.1

#: The information was obtained from NCBI.

&: The genes were used to retrieve the orthologs in wheat, barley and rice

*: orthologs of predicted transcription factors in given species.

Supplementary Table S8. Expression patterns of the predicted transcription factors in given species

Data source or database	Species	transcription factors	gene ID	grain_z71	grain_z75	grain_z75	leaf_z23	leaf_z71	root_z10	root_z13	root_z39	spike_z32	spike_z39	stem_z30	stem_z32	stem_z65
expVIP	wheat	ABI3	TraesCS3B01G452200.1	3.1	1.5	4.2	2.6	3.5	1.2	2.1	2.9	5.3	5.6	1.4	2.4	3.5
		ABF2	TraesCS1A01G306300.1	2.7	1.3	5.2	1.2	1.1	2.7	1.9	2.1	6.6	4.5	2.6	2.4	1.3
			TraesCS1B01G317100.1	1.8	1.9	5.1	0.9	0.8	2.6	2.1	2.3	5.7	4.0	1.2	1.4	1.6
			TraesCS1D01G306000.1	2.1	1.1	5.5	0.9	0.6	1.7	1.5	1.1	4.8	3.0	1.9	1.1	1.4
		ABI5	TraesCS3D01G364900.1	-												
		abi4	TraesCS1A01G223400.1													
			TraesCS1B01G236700.1	-												
			TraesCS1D01G225000.1													
		AG	TraesCS3A01G314300.1	4.5	0.5	1.3	2.4	2.0	0.0	0.0	0.0	3.4	14.4	0.7	1.2	0.1
			TraesCS3B01G157500.4								-					
			TraesCS3D01G140200.2	3.1	0.2	0.0	0.2	0.4	0.0	0.1	0.0	2.1	9.5	0.1	0.7	0.1

Data source or database	Species	transcription factors	gene ID	grain_z71	grain_z75	grain_z85	leaf_z10	leaf_z23	leaf_z71	root_z10	root_z13	root_z39	spike_z29	spike_z32	spike_z65	stem_z30	stem_z32	stem_z65
wheatEXP	wheat	ABI3	TraesCS3B01G452200.1	0.7	10.5	60.6	0.4	0.1	0.1	0.0	0.2	0.2	0.1	0.2	0.1	0.1	0.0	0.1
		ABF2	TraesCS1A01G306300.1	2.7	1.9	5.4	1.3	1.3	1.2	3.0	2.1	2.5	6.7	4.4	3.4	2.8	2.5	1.6
			TraesCS1B01G317100.1	1.7	2.6	5.3	1.0	1.0	0.9	2.9	2.2	2.7	5.8	3.9	3.6	1.3	1.6	2.0
			TraesCS1D01G306000.1	2.4	1.7	6.3	1.5	1.3	0.7	2.0	1.8	1.5	5.2	3.1	2.3	2.4	1.7	1.9
		ABI5	TraesCS3D01G364900.1	0.2	3.3	2.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
		abi4	TraesCS1A01G223400.1															
			TraesCS1B01G236700.1	-														
			TraesCS1D01G225000.1															
		AG	TraesCS3A01G314300.1	3.5	0.5	1.1	0.4	2.4	1.9	0.0	0.0	0.0	2.9	10.9	25.6	0.5	1.0	0.1
			TraesCS3B01G157500.4	1.4	0.2	0.0	0.0	0.8	0.4	0.0	0.1	0.0	1.4	3.3	6.5	0.1	0.4	0.1
			TraesCS3D01G140200.2	3.7	0.3	0.0	0.0	0.5	0.7	0.0	0.1	0.0	2.4	9.0	24.0	0.1	1.3	0.2

Data source or database	Species	transcription factors	gene ID	COL	RAD	GEM	ROO	CRO	LEA	INF	BRC	PST	ANT	CAR5	CAR10	CAR16	DEM22	END22
Microarray	barley	ABI3	HORVU3Hr1G092690.2	773.9	253.8	484.7	27.7	16.4	10.6	30.3	10.3	16.7	12.9	80.5	601.9	999.0	3774.7	1115.1
		ABF2	HORVU3Hr1G084360.2	58.0	35.4	79.4	26.0	33.6	54.5	44.9	54.6	30.8	78.5	86.4	634.1	1127.2	375.5	1475.9
		ABI5	HORVU3Hr1G084360.2	58.0	35.4	79.4	26.0	33.6	54.5	44.9	54.6	30.8	78.5	86.4	634.1	1127.2	375.5	1475.9
		abi4	HORVU1Hr1G060060.1	66.8	48.1	162.5	39.8	31.3	39.4	41.5	36.7	37.8	75.2	35.7	52.9	58.7	531.0	33.9
		AG	HORVU3Hr1G026650.1	29.8	80.1	93.8	55.5	48.8	54.6	332.0	492.0	4152.2	10903.9	2640.5	1385.6	1205.6	36.1	1368.6

				LEA	ROO	ROO2	NOD	SEN	ETI	EPI	INF1	INF2	LOD	LEM	PAL	RAC	CAR5	CAR15	EMB
Transcriptome in this study	barley	ABI3	HORVU3Hr1G092690.2	0.0	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.3	0.0	0.0	0.0	0.0	107.8	204.3	65.8
		ABF2	HORVU3Hr1G084360.2	1.7	1.0	0.0	0.3	0.0	0.0	0.5	0.5	1.3	1.0	45.0	1.2	1.3	533.8	1413.5	8.0
		ABI5	HORVU3Hr1G084360.2	1.7	1.0	0.0	0.3	0.0	0.0	0.5	0.5	1.3	1.0	45.0	1.2	1.3	533.8	1413.5	8.0
		abi4	HORVU1Hr1G060060.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	7.7	5.2
		AG	HORVU3Hr1G026650.1	0.0	0.2	0.0	1.2	2.7	3.2	5.2	18.5	147.3	223.3	146.3	325.0	165.0	1117.2	163.3	0.3

				Anther at An1	Anther at Mei1	Anther at M1	Anther at M2	Anther at M3	Anther at P1	Anther at P2	Anther at P3	Stigma control	Ovary control	Pollinating stigma	Pollinating ovary	Pollen tube growth	Fertilization	Zygote formation	Embryosac control_bottom
Microarray	rice	ABI3	LOC_Os01g68370.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
		ABF2	LOC_Os02g52780.1	447.3	287.7	274.1	334.7	62.6	28.2	6.0	82.1	71.7	8.6	337.0	31.2	66.9	28.2	1.0	70.7
		ABI5	LOC_Os01g64000.1	0.0	0.1	0.1	0.1	0.1	0.1	6.7	6.8	0.2	0.1	0.2	0.1	0.1	0.1	0.1	0.5
		abi4	LOC_Os05g28350.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
		AG	LOC_Os01g10504.1	0.1	0.0	0.1	0.1	0.0	0.0	0.0	0.2	2.8	324.8	7.2	210.4	173.9	134.9	154.9	241.4

				1 DAP embryosac_bottom	2 DAP embryosac_bottom	3 DAP embryosac_bottom	4 DAP embryosac_bottom	Embryosac control_top	1 DAP embryosac_top	2 DAP embryosac_top	3 DAP embryosac_top	4 DAP embryosac_top	Growing callus	Regenerating callus 2 days	Regenerating callus 4 days	Regenerating callus 6 days	Regenerating callus 8 days	Root	Shoot
Microarray	rice	ABI3	LOC_Os01g68370.1	0.1	0.0	0.2	0.4	0.0	0.0	0.0	0.1	0.3	297.8	637.8	694.6	450.5	275.4	0.1	0.0
		ABF2	LOC_Os02g52780.1	23.1	20.7	12.1	19.2	6.1	6.2	20.8	14.1	20.1	7.7	6.9	11.0	17.2	38.2	270.7	168.4
		ABI5	LOC_Os01g64000.1	0.1	0.1	0.2	0.1	0.0	0.0	0.0	0.2	0.3	12.5	3.4	4.2	3.7	6.5	0.1	0.0
		abi4	LOC_Os05g28350.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
		AG	LOC_Os01g10504.1	143.1	167.2	203.3	178.9	307.2	184.0	163.9	127.3	165.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0

				Aleuro	Anther	Callus	Leaf	Panicl	Pistil	Root	Seed	Shoot
Xia et al, 2017	rice	ABI3	LOC_Os01g68370.1	74.7	0.0	30.9	0.0	0.0	0.1	0.0	0.0	0.0
		ABF2	LOC_Os02g52780.1	12.2	13.0	25.4	18.9	12.6	12.0	19.3	12.8	9.2
		ABI5	LOC_Os01g64000.1	23.6	1.1	45.3	0.5	4.4	0.3	0.4	0.1	0.0
		abi4	LOC_Os05g28350.1	0.0	0.0	2.5	0.0	1.3	0.0	0.0	0.0	0.0
		AG	LOC_Os01g10504.1	0.0	44.5	0.0	0.2	83.7	72.7	0.1	0.4	0.4

-: indicated no expression pattern data was found

Xia L, Zou D, Sang J, Xu X, Yin H, Li M, Wu S, Hu S, Hao L, Zhang Z (2017) Rice Expression Database (RED): An integrated RNA-Seq-derived gene expression database for rice. Journal of Genetics and Genomics 44 (5):235-241.

Supplementary Table S9. Correlations between predicted transcription factors and *GW5* for the expression patterns

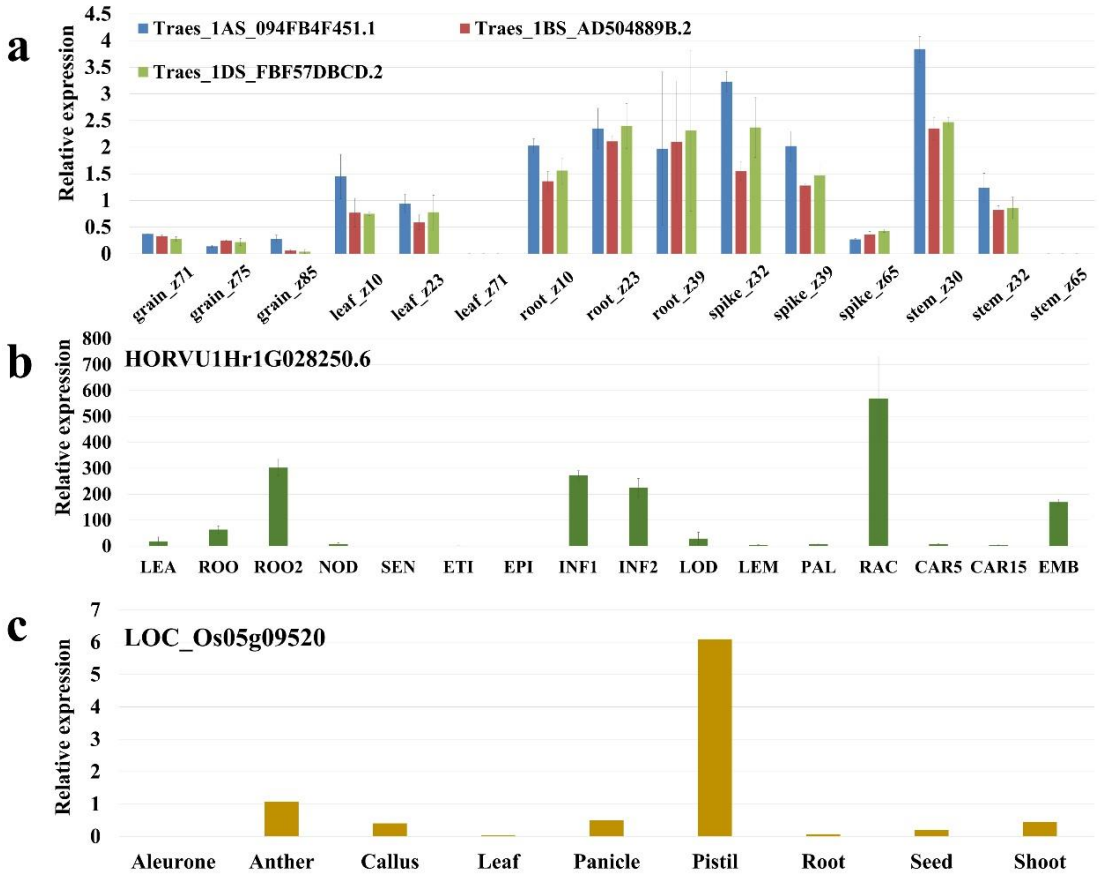
	ABI3		ABF2		ABI5		abi4		AG			Data source or database
	TraesCS3B01G452200.1#	TraesCS1A01G306300.1#	TraesCS1B01G317100.1#	TraesCS1D01G306000.1#	TraesCS3D01G364900.1#	TraesCS1A01G223400.1 #	TraesCS1B01G236700.1 #	TraesCS1D01G225000.1#	TraesCS3A01G314300.1#	TraesCS3B01G157500.4#	TraesCS3D01G140200.2#	
TaGW5-1A		0.482	0.335	0.274					0.117		0.150	
TaGW5-1B	-	0.302	0.187	0.103	-		-		0.019	-	0.078	expVIP
TaGW5-1D		0.395	0.268	0.175					0.047		0.096	
TaGW5-1A	-0.290	0.361	0.157	0.187	-0.391				-0.193	-0.153	-0.188	
TaGW5-1B	-0.336	0.188	0.059	-0.003	-0.386		-		-0.187	-0.162	-0.170	WheatExp
TaGW5-1D	-0.340	0.298	0.174	0.086	-0.400				-0.163	-0.123	-0.149	
HvGW5	0.076		-0.355		-0.355		0.150			-0.243		Microarray
HvGW5	-0.194		-0.222		-0.222		-0.082			-0.135		Transcriptome in this study
OsGW5	-0.040		0.406*		0.025		0.258			-0.245		Microarray
OsGW5	-0.225		-0.258		-0.207		-0.145			0.616		Xia et al, 2017

* Significant correlations at 0.05 level

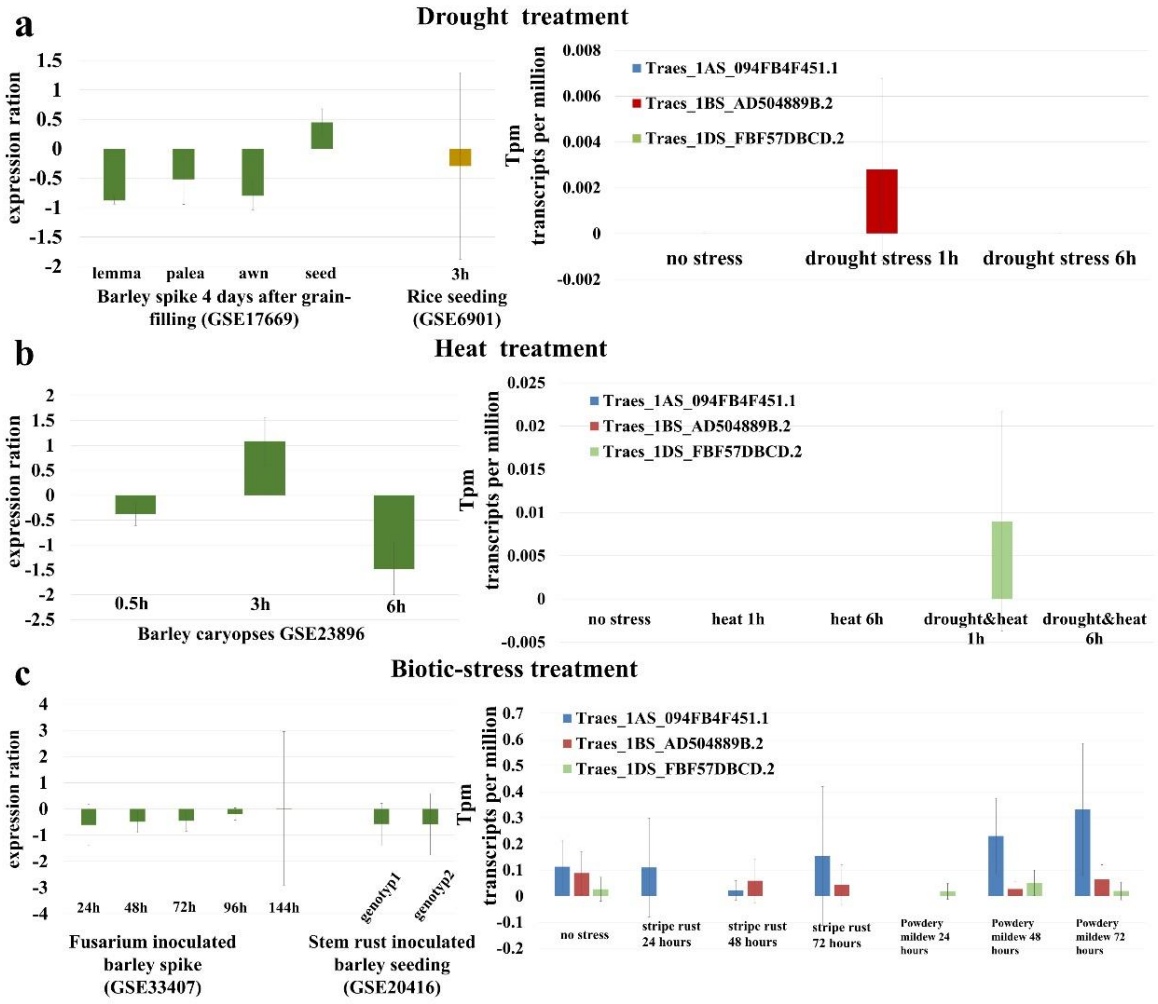
#: orthologs of predicted transcription factors in wheat

-: indicated no expression pattern data was found

Xia L, Zou D, Sang J, Xu X, Yin H, Li M, Wu S, Hu S, Hao L, Zhang Z (2017) Rice Expression Database (RED): An integrated RNA-Seq-derived gene expression database for rice. Journal of Genetics and Genomics 44 (5):235-241.



Supplementary Fig. S1. Validation of expression patterns of *GW5*. For a expression level of *GW5* in wheat by employment of another processed expression database WheatExp, Zadoks scale for wheat growth stage was adopted and detailed information for each stage was provided in Table S2; For b expression patterns of *GW5* in barley by transcriptome data calculated in this study and the detailed information for each stage can be found in Materials and Methods section.; For c expression patterns of *GW5* in rice by transcriptome data from Lin et al. (2017).



Supplementary Fig. S2. Expression profiling of *GW5* in barley and wheat under drought treatment (a); heat treatment (b); and biotic-stress treatment (Fusarium, stem rust and powdery mildew infections) (c). For barley and rice, value > 1 or value <-1 indicates the differences are significant (labelled by the asterisk above the error bar). For wheat, the asterisk labelled above the error bar indicates that *GW5* was differentially expressed (qval<0.05) (Borrill et al. 2016), and the fold changes were given above the asterisk.