JIPB Journal of Integrative Plant Biology

Maize orthologs of rice GS5 and their trans-regulator are associated with kernel development

Jie Liu, Min Deng, Huan Guo, Sharif Raihan, Jingyun Luo, Yuancheng Xu, Xiaofei Dong and Jianbing Yan*

National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan 430070, China. *Correspondence: yjianbing@mail.hzau.edu.cn

Abstract Genome information from model species such as rice can assist in the cloning of genes in a complex genome, such as maize. Here, we identified a maize ortholog of rice *GS5* that contributes to kernel development in maize. The genome-wide association analysis of the expression levels of *ZmGS5*, and 15 of its 26 paralogs, identified a *trans*-regulator on chromosome 7, which was a *BAK1*-like gene. This gene that we named as *ZmBAK1-7* could regulate the expression of *ZmGS5* and three of the paralogs. Candidate-gene association analyses revealed that these five genes were associated with maize kernel development-related traits. Linkage analyses also detected that *ZmGS5* and *ZmBAK1-7* co-localized with mapped QTLs. A transgenic analysis of *ZmGS5* in *Arabidopsis thaliana* L. showed a significant increase in seed weight and cell number, suggesting that *ZmGS5* may have a conserved function among different plant species that affects seed development.

Keywords: Association analysis; kernel development; maize; *ZmBAK1-7*; *ZmGS5*
Citation: Liu J, Deng M, Guo H, Raihan S, Luo J, Xu Y, Dong X, Yan J (2015) Maize orthologs of rice GS5 and their trans-regulator are associated with kernel development. **J Integr Plant Biol** 57: 943–953 doi: 10.1111/jipb.12421
Edited by: Dabing Zhang, Shanghai Jiao Tong University, China
Received Jun. 20, 2015; **Accepted** Aug. 15, 2015
Available online on Aug. 17, 2015 at www.wileyonlinelibrary.com/journal/jipb

Research Article

Free Access

INTRODUCTION

Maize, a widely grown staple food, plays a critical role in supporting the growing world population. Grain size and weight are two important components of grain yield, but the genetic bases of these traits in maize are insufficiently understood.

Linkage and association mapping are two powerful and complementary methods to dissect the genetic bases of complex traits. For rice, using high-density genetic linkage maps and different types of mapping populations, researchers have resolved hundreds of QTLs and genes associated with yield traits (Xing and Zhang 2010). Map-based cloning of QTLs has identified tens of genes controlling grain size and grain yield in rice (Ashikari et al. 2005; Fan et al. 2006; Song et al. 2007; Shomura et al. 2008; Wang et al. 2008; Xue et al. 2008; Huang et al. 2009; Jiao et al. 2010; Li et al. 2010; Mao et al. 2010; Miura et al. 2010; Zhang et al. 2012), as well as many QTLs associated with yield and yield-related traits in maize. For example, the famous nested association mapping (NAM) population was used to map QTLs and elucidate the genetic architecture of complex agronomic and yield traits (Buckler et al. 2009; Kump et al. 2011; Tian et al. 2011). However, few genes have been cloned by map-based methods in maize. The genome-wide association study (GWAS) is a powerful method for gene identification in maize (Yan et al. 2011). More than one million SNPs were developed in a diverse panel using RNA sequencing strategy in maize, and 74 genes associated with oil concentration and fatty acid components were identified by GWAS (Li et al. 2013). However, using a similar strategy and the same panel, only a few loci were identified for yield and agronomic traits (Yang et al. 2014), which implies that grain yield is a more complex trait compared with fatty acids and that many genetic factors are probably involved.

Comparative genomics has proven to be a useful alternative approach to identifying the genes underlying QTLs for complex traits. Using high-resolution linkage mapping and expression analysis, Miller et al. (2007) found that *cis*-regulatory changes in *Kit ligand* (*Kitlg*) affect the gill and skin tissue colors of sticklebacks and, through admixture mapping, that the human *KITLG* genomic region has a significant effect on human skin color. The maize *TEOSINTE BRANCHED 1* (*TB1*) gene is involved in regulating the growth of axillary buds (Clark et al. 2006; Studer et al. 2011). The rice *OsTB1* gene, which was identified based on its sequence similarity with maize TB1, also negatively regulates lateral branching (Takeda et al. 2003). Orthologous genes of rice kernel size-related *GS3* (Fan et al. 2006) and *GW2* (Song et al. 2007) were also identified in maize (Li et al. 2011a, 2011b) and wheat (Bednarek et al. 2012; Zhang et al. 2014), and they have similar roles, providing us the opportunity to functionally identify and clone orthologous genes using the well-annotated rice genome.

GS5 controls grain size and weight through the regulation of grain width and filling in rice (Li et al. 2011). Enhanced expression of *GS5* competitively inhibits the interaction between *OsBAK1-7* and *OsMSBP1* by occupying the extracellular leucine-rich repeat (LRR) domain of *OsBAK1-7* (Xu et al. 2015). This inhibition could prevent *OsBAK1-7* from endocytosis caused by interacting with *OsMSBP1* which may explain how *GS5* could affect grain size. In the present study, the orthologous gene of rice *GS5*, *ZmGS5*, and 26 paralogs of *ZmGS5* were identified in maize. An eQTL analysis of *ZmGS5* and its paralogs found that *ZmGS5*and other three *GS5* homologous genes were regulated by a *trans*-regulator, *ZmBAK1-7* which was a *BAK1*-like gene, on chromosome 7. Candidate-gene association analysis showed that *ZmGS5* and

ZmBAK1-7 were associated with kernel-related traits. These two genes were also co-localized with mapped QTLs in present and previous studies. Transgenic analysis in *Arabidopsis* indicated that *ZmGS5* could regulate kernel development, thus enhancing seed weight.

RESULTS

Identification and characterization of *ZmGS5*

BLAST searches using the rice *GS5* protein sequence (GenBank: AEO37083) as the query against the maize B73-filtered gene set version 5b.60 RefGen_V2 database identified the protein (GRMZM2G123815) with the highest similarity to GS5 (E-Value = 0, Identity = 73.63%). This gene, which we named *ZmGS5*, contains 10 exons, as does *GS5* in rice.

The comparative genomic analysis, using Symap software (Soderlund et al. 2011), between rice chromosome 5 and maize chromosome 3 identified syntenic fragments corresponding to rice *GS5* and maize *ZmGS5* regions. Both genes belonged to the peptidase family S10, which has a serine-type carboxypeptidase activity.

In addition to *ZmGS5*, we identified 26 other genes that were paralogs of *ZmGS5* (Table S1). Most of these genes had no description, but some were predicted to have serine carboxypeptidase activity.

ZmGS5 and the other three homologous genes are regulated by a same *trans*-regulator

More than one million high-quality SNPs and the expression levels of 28,769 genes at 15 d after pollination were publicly available in 368 diverse maize lines (Li et al. 2013; Fu et al. 2013). SNPs and expression levels for *ZmGS5* and 12 paralogs were identified in the 368 maize lines. Using these datasets, GWAS on the expression levels of these genes were performed, and four genes, including *ZmGS5*, were found to be regulated by the same *trans*-eQTL located on chromosome 7, which we named as*ZmBAK1-7*(GRMZM2G149051, Figure 1A) because of its BAK1-like domain. The alignments of the BAK1-like domain between *ZmBAK1-7* with *OsBAK1* and *OsBAK1-7*, which could interact with rice GS5 showed high similarity (E-value = 8e-49 and 5e-54, Identity = 34 and 36%, respectively).

Sequence analysis of the protein encoded by *ZmBAK1-7* was performed through searching InterProScan and it indicated that this protein had putative leucine-rich repeat receptor-like serine/threonine-protein kinase (LRR-RLKs) activity. The HMM-HMM-based lightning-fast iterative sequence search method (Remmert et al. 2012) was used to search the Protein Data Bank, which revealed that this protein has a brassinosteroid insensitive 1 structure and a brassinosteroid insensitive 1-associated receptor structure. The secondary structure prediction using SOSUI (Hirokawa et al. 1998) showed that this protein had transmembrane helices.

The eQTL analysis demonstrated that the *cis*-element near this gene strongly regulates its expression level (Figure 1B). It was revealed that *ZmBAK1-7* could regulate the expression of *ZmGS5*, as well as the other three homologous genes of *GS5* (Figure 1 C–F) and that these five genes were highly co-expressed ($r > 0.5$, Figure 1A). These data implied that *ZmGS5* and another three homologous genes might be involved in the same pathway or participate in coordinated developmental processes.

Natural variations in *ZmGS5* and *ZmBAK1-7* affect maize kernel-related traits

Kernel-related traits, including kernel length (KL), kernel width (KW), kernel thickness (KT) and 100-kernel weight (HKW), of 368 diverse maize inbred lines were measured under four or more environments (location and year), including Sichuan, Yunnan and Hainan Provinces in 2009, 2010, and 2011, respectively. Significant phenotypic variations were identified in the four measured traits. The smallest variation was for KL, ranging from 7.04 to 11.26 mm, and the greatest variation was for HKW, ranging from 11.11 to 29.79 g (Table 1). The broad-sense heritability for these four traits was >0.85 (Table 1).

For a candidate gene association analysis, we combined two sets of genotypic data from *ZmGS5* and *ZmBAK1-7*. One set was the SNPs generated through RNA-Seq and the other was the polymorphisms obtained by re-sequencing the 5′ upstream regions of these two genes. We re-sequenced *ZmGS5* in 155 diverse maize inbred lines (Yang et al. 2010) and found another 18 polymorphic sites, including four insertions and/or deletions (InDels). Re-sequencing of *ZmBAK1-7* was performed in 508 diverse maize inbred lines and generated another 43 polymorphisms, including 13 InDels. Overall, we identified 43 and 67 polymorphisms with minor allele frequencies (MAFs) ≥0.05 in *ZmGS5* and *ZmBAK1-7*, respectively. Genotypic data for the other three *GS5* homologous genes, GRMZM2G052507, GRMZM2G123940 and GRMZM5G879749, were derived from RNA-Seq results and contained 35, 7 and 20 SNPs, respectively, with MAFs ≥0.05.

The candidate gene association analysis was performed using TASSEL software (Bradbury et al. 2007) with a mixed linear model (Yu et al. 2006) that takes population structure and kinship into consideration. *ZmBAK1-7* was associated with all of the investigated kernel traits in at least two environments, and *ZmGS5* was associated with KL, KW and HKW in one environment (Tables 2, S2). GRMZM2G052507, GRMZM5G879749 and GRMZM2G123940 were associated with at least one kernel-related trait (Tables 2, S2). Most SNPs in *ZmBAK1-7* and *ZmGS5* were significantly associated with kernel traits in two environments. However, SNP M7c130937610 showed significant associations with KL in four environments (Table S2).

SNP M7c130937690 from *ZmBAK1-7* was the most significant polymorphism in this gene and was associated with KT in Hainan Province, 2011 ($n = 320$, $P = 3.39 \times 10^{-4}$) (Figure 2A). It could lead to an amino acid change (Thr/Ser^{176}) which was predicted to locate in the leucine-rich repeat receptor-like protein kinase domain and outside the cell membrane (Figure 2B). This SNP was in strong linkage disequilibrium (LD) with three other significant SNPs ($r^2 > 0.2$, Figure 2C). Among these three SNPs, two were synonymous and the remaining one could lead to amino acid change (Ala/Ser), but it was not located in reported domains. SNP M3c61242466 from *ZmGS5* was the most significant variant and associated with KL in Hainan Province, 2011 ($n = 326$, $P = 3.90 \times 10^{-4}$) (Figure 3A). This SNP and other significant SNPs were all synonymous and could not affect the amino acid sequence (Figure 3B). Notably, these SNPs were in strong LD ($r^2 > 0.5$) and located near the 3′ UTR region (Figure 3C). This suggested that the potential functional variation might be a regulatory element in the 3′ end of the gene. These two SNPs contributed 6.7% and 6.3% to the phenotypic variation of KT and KL, respectively. Comparisons of

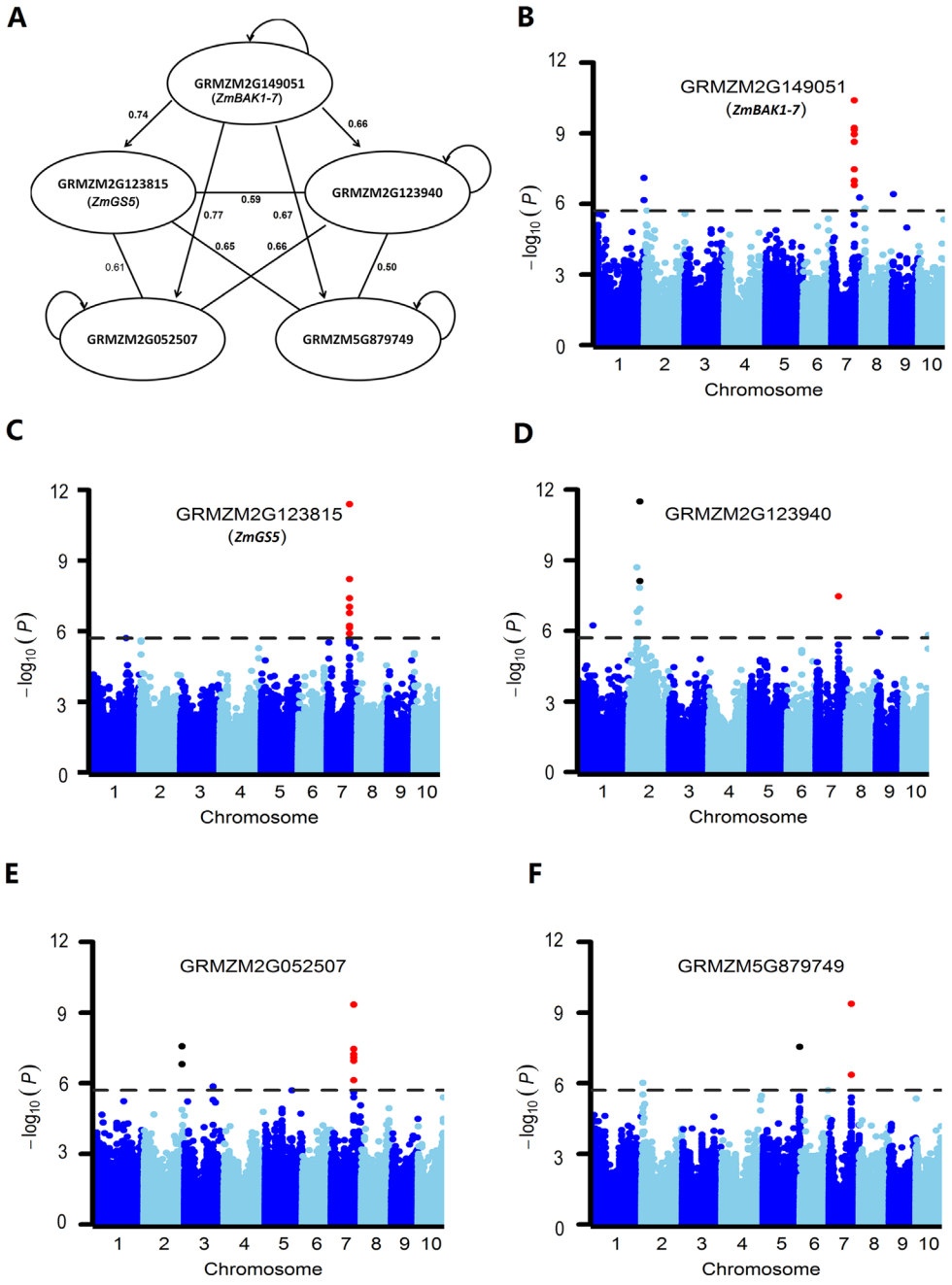

Figure 1. Co-expression network and Manhattan plots of five genes, which were associated with kernel-related traits
(**A**) The co-expression network of five genes. The numbers underling the edge lines meant the correlation coefficients of expression levels and the arrows meant that *ZmBAK1-7* regulated the other four genes. (**B–F**) Manhattan plots for GWAS results of expression levels of *ZmBAK1-7*, *ZmGS5*, GRMZM2G123940, GRMZM2G052507 and GRMZM5G879749, respectively. The red dots in *ZmBAK1-7*indicated that expression of this gene is strongly controlled by *cis*-element and that in the other four genes indicated that *ZmBAK1-7* was trans-regulator for these four genes. The black dots in GRMZM2G123940, GRMZM2G052507 and GRMZM5G879749 showed the significant SNPs within these three genes, respectively.

these two significant *P*-values with the *P*-value distribution of 500 randomly chosen SNPs from 500 different genes suggested that these two significant associations were not due to false positives (Figures 2D, 3D). Considering these results, we concluded that *ZmGS5* and *ZmBAK1-7* were associated with kernel development.

ZmGS5 and *ZmBAK1-7* are located in mapped kernel-related QTL intervals

One major QTL affecting KW and HKW was identified on chromosome 7 in a recombinant inbred lines (RIL) population derived from an inbred line BK selected from a tropical landrace with a big kernel size and the elite Chinese breeding

Table 1. Phenotypic analysis and repeatability of four kernel related traits

Category	Source of variation	DF	HKW (g)	KL (mm)	KW (mm)	KT (mm)
ANOVA	Environments	6	998.87**	121.35**	23.65**	6.62**
	Genotypes	367	62.59**	4.13**	2.34**	0.82**
	Error	1,619	11.66	0.40	0.18	0.10
	Heritability		0.85	0.93	0.94	0.90
	Range		11.11–29.79	7.04–11.26	5.03–9.90	3.78–5.64
	Mean ± *SD*		22.53 ± 2.76	9.05 ± 0.77	8.14 ± 0.58	4.68 ± 0.34

HKW, one-hundred kernel weight; KL, kernel length; KT, kernel thickness; KW, kernel width. **, $P < 0.01$.

line Yu8701. *ZmBAK1-7* was mapped near the peak region of the QTL (Figure 4A). Another QTL affecting the kernel test weight was also identified on chromosome 3 in a BC_2F_6 population derived from a teosinte accession and the elite breeding line M017. *ZmGS5* was mapped near the peak region of the QTL (Figure 4B). Previous studies (Beavis et al. 1994; CIMMYT 1994; Doebley et al. 1994; Veldboom and Lee 1994, 1996; Ajmone-Marsan et al. 1995; Austin and Lee 1996; Melchinger et al. 1998) also identified QTLs for KW and yield in similar regions (bin 3.04 and 7.03) to where *ZmGS5* and *ZmBAK1-7* were located, respectively (Table S3). The QTL where *ZmGS5* was located explained 8.5% of the kernel test weight variation in the Teosinte/M017 BC_2F_6 population, whereas the QTL where *ZmBAK1-7* was located explained 10.6 and 9.8% of the KW and HKW variation, respectively, in the BK/Yu8701 RIL population. The two candidate genes were of great interest and might be the underlying genes of the two QTLs although more studies were required.

ZmGS5 affects seed size in transgenic *Arabidopsis*

The full-length cDNA sequence of *ZmGS5* from maize inbred line B73 was cloned and transformed into *Arabidopsis* with expression vector pBinGlyRed3 consisting of CaMV35S promoter (Zhang et al. 2013). In total, 14 independent *ZmGS5* transgenic lines were obtained (Figure S1), and the 1,000-seed weights of the T2 generations from 10 of 14 transgenic lines were higher than that of the wild type. In particular, the 1,000-seed weights of *ZmGS5-1*, *ZmGS5-13* and *ZmGS5-20* were significantly higher than the wild-type. To confirm the results, these positive transgenic lines, *ZmGS5-1*, *ZmGS5-13* and *ZmGS5-20*, were selected for further analysis. Subsequently, the seeds of the third generation of transgenic individuals were used to measure 1,000-seed weight. The 1,000-seed weight from transgenic lines *ZmGS5-1*, *ZmGS5-13* and *ZmGS5-20* increased 6.8, 6.6 and 17.0%, respectively, over wild type (Figure 5A). The data from the T3 generation showed similar trends to those of the T2 generation. Semi-quantitative PCR was performed as an expression analysis of *ZmGS5* in the measured materials, and a positive correlation was observed between seed size and gene expression (Figure 5B).

We also investigated the size of embryos, including cotyledons and radicals, from mature seeds (Figure 5C). The cotyledon sizes of the transgenic lines *ZmGS5-1*, *ZmGS5-13* and *ZmGS5-20* were 19.6, 24.1 and 9.9% greater, respectively, than that of wild type, and the radical sizes of the transgenic lines *ZmGS5-1*, *ZmGS5-13* and *ZmGS5-20* were 20.5, 31.0 and 9.9% greater, respectively, than that of wild type (Figure 5E). These results were in accordance with the seed weights. To understand the basis for the increased cotyledon size in transgenic plants, we analyzed the cell sizes in the central region of the cotyledons (Figure 5D). The average cell areas were significantly increased in the three transgenic lines by 11.0, 14.2 and 7.1%, respectively, over that of wild type (Figure 5F). However, the cell number in cotyledons, obtained by dividing the cotyledon size by the embryo cell area, showed no significant difference between the transgenic lines *ZmGS5-1* and *ZmGS5-13* and the wild type; however, *ZmGS5-20* showed a non-significant decrease of 6.4% (Figure 5F). This suggested that *ZmGS5* affected both seed weight and cotyledon size.

Table 2. Number of environments in which the investigated genes was associated with kernel-related traits

Gene	KL	KW	KT	HKW
ZmBAK1-7	4	2	3	2
ZmGS5	1	1	0	1
GRMZM2G052507	5	1	0	0
GRMZM5G879749	0	0	1	0
GRMZM2G123940	0	4	1	0

DISCUSSION

Candidate gene mining and association analyses based on a comparative genomic strategy are useful tools for identifying key genes affecting complex agronomic traits, especially in crops with large genomes. With the rapid development of next-generation sequencing technology, it has become cheaper and quicker to determine the reference genome sequences of many crop species. Unlike conserved marker-based comparative genomic analyses (Moore et al. 1995), the new technology allows us to compare the whole genome sequences of many crop species and identify the conserved sequences and functions of agronomically important genes (Van Bel et al. 2012). This will lead to a better understanding of the evolutionary history, domestication and improvement of crops for agronomically and economically important traits, thus enhancing plant breeding.

The method of comparative genomics has been used to identify maize orthologs of rice genes controlling grain size and weight in previous studies (Li et al. 2010a, 2010b). In this study, we also found significant associations between *ZmGS5* and kernel-related traits. Three haplotypes in the promoter region of rice *GS5* seem to be associated with KW (Li et al.

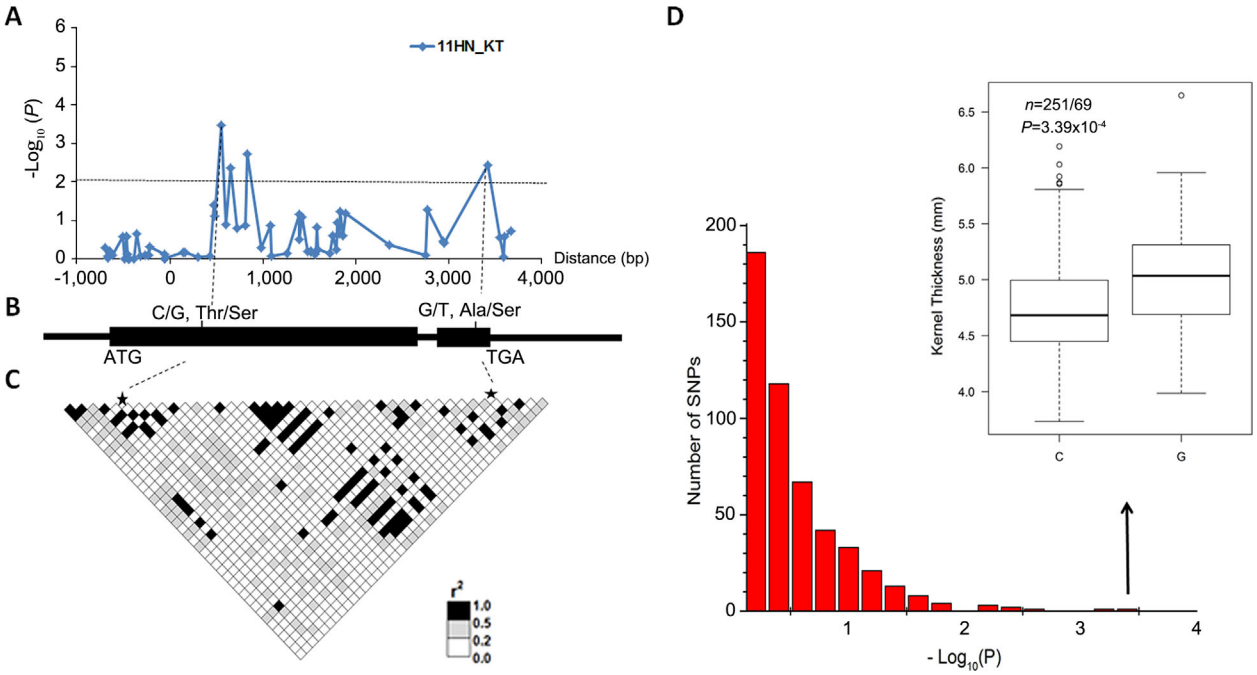

Figure 2. Significant associations between natural variations of *ZmBAK1-7* and kernel thickness
(**A**) Candidate gene association analysis between single nucleotide polymorphisms (SNPs) of *ZmBAK1-7* and kernel thickness. (**B**) Gene structure of *ZmBAK1-7*. Two significant SNPs were non-synonymous. (**C**) LD matrix across *ZmBAK1-7*. (**D**) Permutation test for the identified strongest associated SNP. Comparison of the strongest significant SNP ($n = 251/69$, $P = 2.96 \times 10^{-6}$) identified in present study with the association results of 500 randomly selected SNPs for the same trait. The results showed that the identified association is more significant than the 500 times random tests which demonstrated the identified association was not false positive but real association.

2011). However, according to our results, *ZmGS5* might not affect grain size in the same way as the rice gene because its relevant sites were not in the promoter region. This result was also consistent with the results of *ZmGS3*, *ZmGW2-CHR4* and *ZmGW2-CHR5* (Li et al. 2010a, 2010b). They suggest that maize and rice might have undergone different evolutionary pressures after divergence.

Seed weight and cotyledon cell number increased significantly compared with wild type in *A. thaliana* over-expressing *ZmGS5* (Figure 5). Thus *GS5* seems to be functionally conserved across different plant species. The increased seed weight, cell number and faster root growth of transgenic lines compared with wild type were consistent with the function of rice *GS5* as a positive modulator upstream of cell cycle genes (Li et al. 2011). It would be interesting to determine whether *GS5* is also functional in other crops, such as wheat and sorghum.

Although many genes underlying QTLs for grain size and weight have been positionally cloned in other species (Ashikari et al. 2005; Fan et al. 2006; Song et al. 2007; Shomura et al. 2008; Wang et al. 2008; Xue et al. 2008; Huang et al. 2009; Jiao et al. 2010; Mao et al. 2010; Li et al. 2010; Miura et al. 2010; Zhang et al. 2012), little is known about the pathways they involve in. Here, in addition to the association between *ZmGS5* and kernel-related traits, we have also discovered using eQTL analysis that *ZmBAK1-7* could regulate the expressions of *ZmGS5* and other three homologous genes and all these five genes were highly co-expressed. *ZmBAK1-7* was a *BAK1*-like gene and contained LRR-RLKs domain. Mutations in various LRR-RLKs can affect diverse developmental processes in plants, such as the perception of the hormones brassinosteroid (Li and Chory 1997) and system in (Montoya et al. 2002; Scheer and Ryan 2002), meristem differentiation (Clark et al. 1993), endosperm and pollen development (Canales et al. 2002; Zhao et al. 2002), ovule development and early embryogenesis (Hecht et al. 2001).

Recently, it was found that rice *GS5* could interact with *OsBAK1-7*and its expression could be suppressed by BR. This demonstrated that *GS5* involved in brassinosteroid signaling. In our study, we also found that *ZmGS5* could interact with a BAK1-like gene (*ZmBAK1-7*). This consistence indicated that *ZmGS5* might also involve in brassinosteroid signaling. The similar domain between OsBAK1-7 and ZmBAK1-7 implied that *GS5* might have conserved function among different plant species. Meanwhile, we also found that *ZmBAK1-7* was one of the 12 regulatory hotspots which could regulate the expressions of more than 100 genes (data not shown). When a very stringent threshold for GWAS was set ($P = 1.0 \times 10^{-10}$), *ZmBAK1-7* was found to regulate expressions of 67 genes (Table S4). If the threshold was set to a less stringent level ($P = 1.8 \times 10^{-6}$), then the number of regulated genes reached ≈900. Gene ontology (GO) analysis of these 67 genes showed that these genes could play role as binding factor, catalytic factor, transcription regulator and transporter and they mainly affected the process of

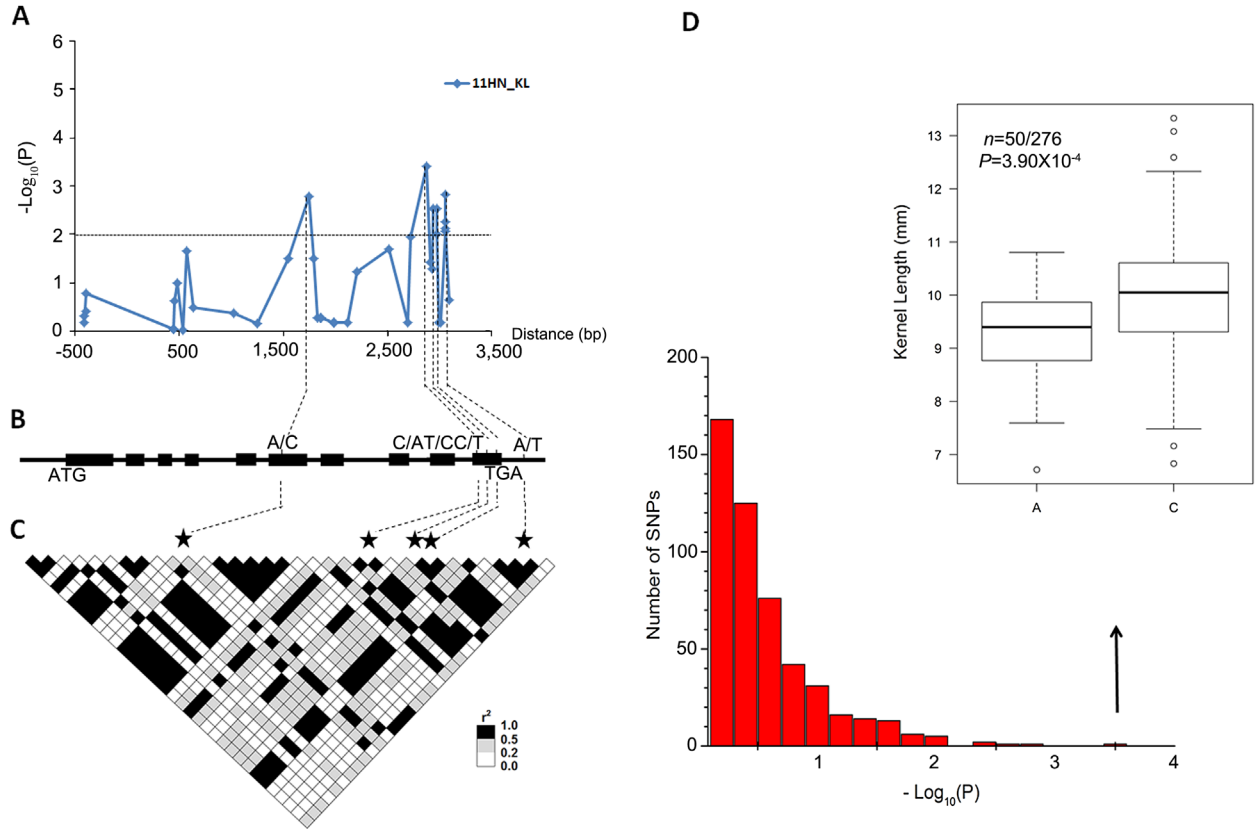

Figure 3. Significant associations between natural variations of *ZmGS5* and kernel length
(**A**) Candidate gene association analysis between single nucleotide polymorphisms (SNPs) of *ZmGS5* and kernel length. (**B**) Gene structure of *ZmGS5*. Significant SNPs were synonymous. (**C**) LD matrix across *ZmGS5*. (**D**) Permutation test for the identified strongest associated SNP. Comparison of the strongest significant SNP ($n = 50/276$, $P = 4.82 \times 10^{-6}$) identified in the present study with the association results of 500 randomly selected SNPs for the same trait. The results showed that the identified association is more significant than the 500 times random tests which demonstrated the identified association was not false positive but real association.

biological regulation, metabolic reaction, response to stimulus (Figure S2). Thus, *ZmBAK1-7* seems to be a key regulator that interacts with *ZmGS5* to affect maize kernel development. Further studies are required to reveal the function of *ZmBAK1-7* that will be very helpful in understanding the maize kernel development as well as facilitating the genetic improvement of grain yield.

MATERIALS AND METHODS

Genetic materials and phenotypic data collection

An association mapping population (AM508), consisting of 508 diverse maize inbred lines, was developed for the dissection of complex quantitative traits (Yang et al. 2011). One subset of 155 lines (CAM155) (Yang et al. 2010) was used for the re-sequencing of *ZmGS5*, and another subset composed of 368 lines was used for RNA-Seq (Li et al. 2013). The re-sequencing of *ZmBAK1-7* was done in AM508. Populations were planted during 2009, 2010 and 2011 in Sichuan, Yunnan and Hainan Provinces. Two linkage populations, BK/Yu8701 RIL and Teosinte/M017 BC_2F_7, were planted in Hainan in 2011. Kernel-related traits, such as KL, KW and KT, were measured 10 times for each line using a digital ruler, and HKW was measured three times for each line using an electronic scale. The values of each trait were then averaged to obtain the phenotypic value used in the analysis.

Identification of *ZmGS5* and 26 other genes with a high similarity to *GS5*

Based on the protein sequence of rice GS5, we searched reference protein databases and identified the protein with the highest similarity. We also conducted comparative genomic analyses between the *GS5* regions of rice and maize and the reference sequences and annotation files of the rice and maize genomes, respectively. A similar method was used to identify the paralogs of *ZmGS5* based on the protein sequence of *ZmGS5* and the maize B73-filtered gene set version 5b.60 RefGen_V2 database (Schnable et al. 2009).

Protein sequence analysis

First, we searched the InterProScan database to identify the domains present in these genes. This would identify, based on the amino acid sequences, families that these proteins belong

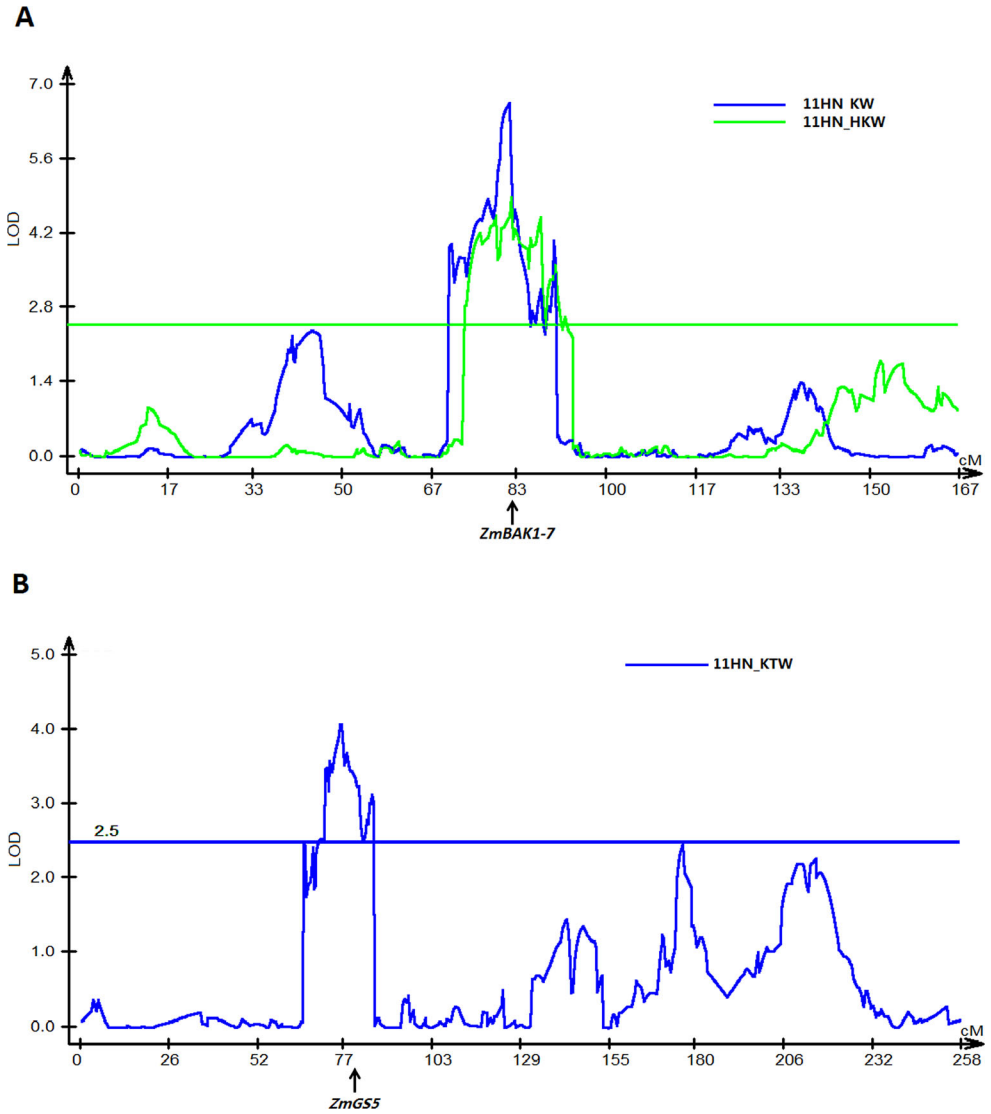

Figure 4. *ZmBAK1-7* and *ZmGS5* col-located with the mapped QTL for kernel width, 100-kernel weight and kernel test weight, respectively

(**A**) *ZmBAK1-7* fell into one of the mapped QTL intervals in BK/Yu8701 recombinant inbred line (RIL) population. The investigated traits were kernel width and 100-kernel weight in Hainan, 2011. The arrow indicated the position of *ZmBAK1-7*. (**B**) *ZmGS5* fell into one of the mapped QTL intervals in Teosinte/Mo17 BC_2F_6 population. The investigated trait was kernel test weight in Hainan, 2011. The arrow indicated the position of *ZmGS5*.

to. Then, a faster and more sensitive method than PSI-BLAST, HMM-HMM-based lightning-fast iterative sequence search (Remmert et al. 2012), was used to search the Protein Data Bank, confirm the structure and produce a more detailed functional annotation of the protein.

Genotypic data and gene expression analysis

For association tests, we used two sets of genotypic data. One consisted of the high-quality SNPs from RNA-Seq (Li et al. 2013), and the other was the re-sequencing data for *ZmGS5* and *ZmBAK1-7*. The primers (Table S5) used to re-sequence these two genes were designed based on the B73 reference genome sequence using Primer-BLAST in NCBI. The PCR products were sequenced, and multiple alignments of these sequences were performed using BioEdit software (Hall, 1999). Polymorphisms were generated from the alignment results using TASSEL software to extract InDels and remove SNPs with MAFs < 0.05.

Association and linkage analyses

The expression levels of *ZmGS5*, *ZmBAK1-7* and three other paralogs of *ZmGS5* were quantified by RNA-Seq (Li et al. 2013). GWAS for eQTLs and association tests between polymorphisms from candidate genes and kernel-related traits were performed using TASSEL software with a mixed linear model. The kinship matrix and population structure were estimated using 36,618 high-quality SNPs obtained from Illumina MaizeSNP50 BeadChip (Ganal et al. 2011; Li et al. 2012). For

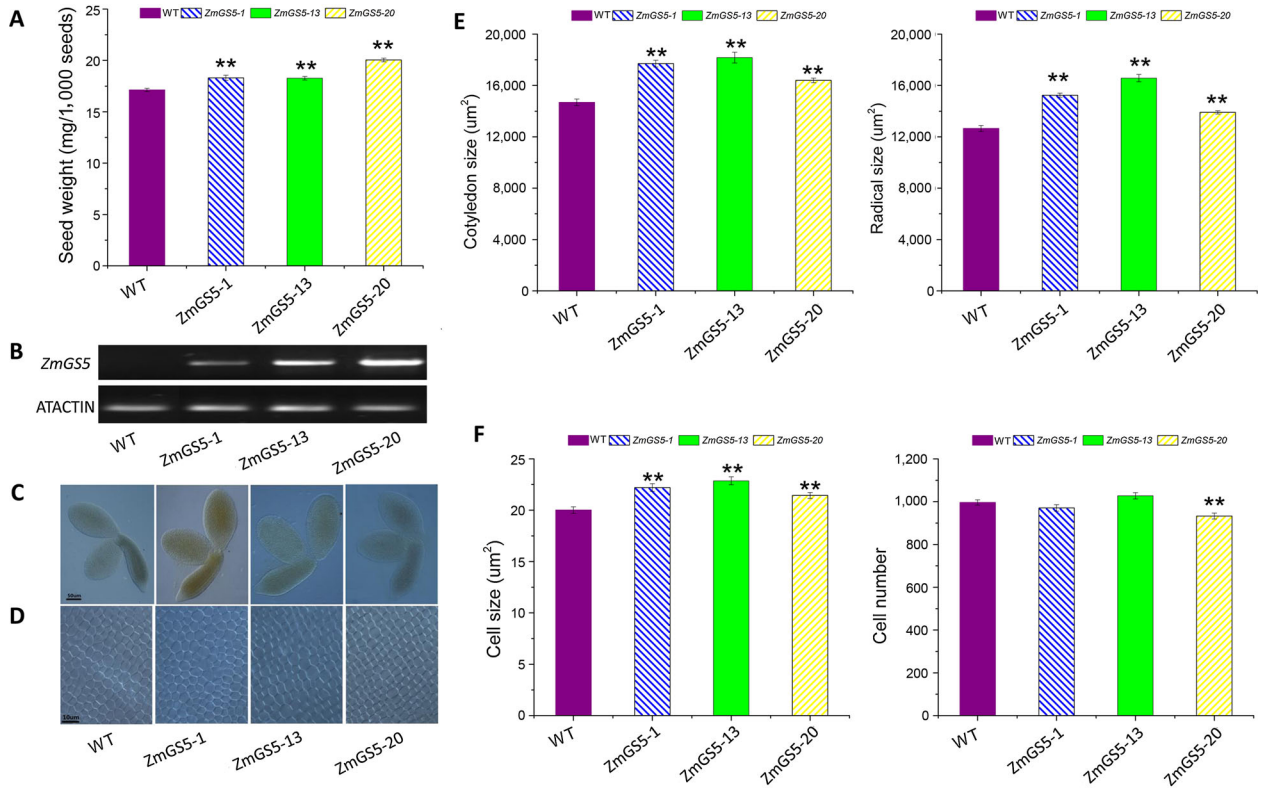

Figure 5. Transgenic validation of *ZmGS5* in *Arabidopsis*

(**A**) The seed weights from the third generation of *ZmGS5-1*/ *ZmGS5-13*/ *ZmGS5-20* transgenic lines and wild type. The bars represented the average seed weight based on five repeats and the error bars stood for the standard errors, the significant differences were based on *t*-test ($^{*}P < 0.05$; $^{**}P < 0.01$). (**B**) RNA expression of *ZmGS5* in different transgenic lines and WT, as determined by RT-PCR for different cycles. (**C**) Representative embryos from seeds of *ZmGS5-1*/ *ZmGS5-13*/ *ZmGS5-20* transgenic lines and WT. (**D**) Embryo cells in central region of cotyledons from seeds of *ZmGS5-1*/ *ZmGS5-13*/ *ZmGS5-20* transgenic lines and WT. (**E**) Cotyledon and radical size of mature embryos from the transgenic lines and WT. The bars represent mean ± standard errors and the significant differences were based on *t*-test, with $n > 50$ for cotyledon size analysis, $n > 20$ for radical size analysis ($^{*}P < 0.05$; $^{**}P < 0.01$). (**F**) Cell size and average cell number of mature embryos from the transgenic lines and WT. The bars represent mean ± standard errors and the significant differences were based on *t*-test, $n > 20$ ($^{*}P < 0.05$; $^{**}P < 0.01$).

GWAS, the threshold was set at $P = 1.8 \times 10^{-6}$ ($P = 1/n$, where n = the number of markers used). Linkage mapping was performed using Windows QTL Cartographer Version 2.5.

Transgenic analysis in *Arabidopsis*

Arabidopsis thaliana used in this study was the Col-0 ecotype (provided by Dr. Yongming Zhou, Huazhong Agricultural University, Wuhan, China). Wild-type and transgenic *Arabidopsis* seeds were surface sterilized and sown on 0.5 × MS medium containing 1% sucrose (w/v) and 0.8% agar (w/v). All plants were grown at 22 °C with a 16/8-h light/dark photoperiod.

The DNA for the genotyping analysis was extracted from the leaves of transgenic lines and wild type. For RNA extractions, 15-d-old seedlings grown on 0.5× MS medium were collected. Total RNA was prepared from fresh tissues at 15 d after sowing using an RNA extraction kit (BioTeke, China). For RT-PCR, the first-strand cDNA was synthesized from 4 μg total RNA using the TransScript One-Step gDNA Removal and cDNA Synthesis SuperMix kit (TransGen, China). Semi-quantitative PCR was performed for the gene expression analysis using gene-specific (15F-1 and 9R-8) and *A. thaliana* ACTIN (ATACTIN-F and ATACTIN-R) primers. The primer sequences are listed in Table S5.

The open reading frame of *ZmGS5* was amplified from the cDNA of maize inbred line B73 leaves by PCR using the gene-specific primers 15F-1M and 9R-8M (Table S5). The amplified fragment of *ZmGS5*, bearing the EcoRI and XbaI restriction sites, was inserted into the pEASY-Blunt vector (TransGen, China). Positive clones were sequenced using the primers M13F and M13R (Table S3). After sequencing, the cDNA was cloned into the expression vector pBinGlyRed3 (Zhang et al. 2013), which contained a DsRed marker for transgenic plant selection. The seed-specific glycinin promoter in pBinGlyRed3 was replaced by the CaMV35S promoter (provided by Dr. Chunyu Zhang, Huazhong Agricultural University, Wuhan, China). The final construct was introduced into *A. tumefaciens* strain GV3101 (provided by Dr. Yongming Zhou, Huazhong Agricultural University, Wuhan, China) for *Arabidopsis* transformations (Zhou et al. 2003).

For cytological observations, embryos and seeds of wild-type and transgenic lines were soaked in water for 4 h and then cleared for 4 h in Hoyer's solution containing 8:1:2 (w/v/v) hydrate/glycerol/water (Yin et al. 2012). All samples were observed and photographed using a Nikon-Eclipse80i differential interference contrast microscope equipped with a CCD camera. The sizes of cotyledons and radicals were measured with ImageJ (http://rsbweb.nih.gov/ij/index.html). Cell size was determined using the average cell size in areas of $>$2,000 μm^2 and the cleared cell number. The cell number was calculated by dividing the cotyledon size by the embryo cell size (Cheng et al. 2013).

ACKNOWLEDGEMENTS

This research was supported by the National Natural Science Foundation of China (31222041) and the National Hi-Tech Research and Development Program of China (2012AA10A307).

REFERENCES

Ajmone-Marsan P, Monfredini G, Ludwig WF, Melchinger AE, Franceschini P, Pagnotto G, Motto M (1995) In an elite cross of maize a major quantitative trait locus controls one-fourth of the genetic variation for grain yield. **Theor Appl Genet** 90: 415–424

Ashikari M, Sakakibara H, Lin S, Yamamoto T, Takashi T, Nishimura A, Angeles ER, Qian Q, Kitano H, Matsuoka M (2005) Cytokinin oxidase regulates rice grain production. **Science** 309: 741–745

Austin DF, Lee M (1996) Comparative mapping in F2:3 and F6:7 generations of quantitative trait loci for grain yield and yield components in maize. **Theor Appl Genet** 92: 817–826

Beavis WD, Smith OS, Grant D, Fincher R (1994) Identification of quantitative trait loci using a small sample of topcrossed and F4 progeny from maize. **Crop Sci** 34: 882–896

Bednarek J, Boulaflous A, Girousse C, Ravel C, Tassy C, Barret P, Bouzidi MF, Mouzeyar S (2012) Down-regulation of the *TaGW2* gene by RNA interference results in decreased grain size and weight in wheat. **J Exp Bot** 63: 5945–5955

Bradbury PJ, Zhang Z, Kroon DE, Casstevens TM, Ramdoss Y, Buckler ES (2007) TASSEL: Software for association mapping of complex traits in diverse samples. **Bioinformatics** 23: 2633–2635

Buckler E, Holland JB, Bradbury PJ, Acharya CB, Brown PJ, Browne C, Ersoz E, Flint-Garcia S, Garcia A, Glaubitz JC, Goodman MM, Harjes C, Guill K, Kroon DE, Larsson S, Lepak NK, Li H, Mitchell SE, Pressoir G, Peiffer JA, Rosas MO, Rocheford TR, Romay MC, Romero S, Salvo S, Villeda HS, Silva HS, Sun Q, Tian F, Upadyayula N, Ware D, Yates H, Yu J, Zhang Z, Kresovich S, McMullen MD (2009) The genetic architecture of maize flowering time. **Science** 325: 714–718

Canales C, Bhatt AM, Scott R, Dickinson H (2002) EXS, a putative LRR receptor kinase, regulates male germline cell number and tapetal identity and promotes seed development in *Arabidopsis*. **Curr Biol** 12: 1718–1727

Cheng Y, Cao L, Wang S, Li Y, Shi X, Liu H, Li L, Zhang Z, Fowke LC, Wang H, Zhou Y (2013) Downregulation of multiple CDK inhibitor *ICK/KRP* genes upregulates the E2F pathway and increases cell proliferation, and organ and seed sizes in *Arabidopsis*. **Plant J** 75: 642–655

CIMMYT (1994) QTL data for populations Ki3 x CML139 and CML131 x CML67. http://www.agron.missouri.edu/locus.html

Clark RM, Wagler TN, Quijada P, Doebley J (2006) A distance upstream enhancer at the maize domestication gene *tb1* has pleiotropic effects on plant and inflorescent architecture. **Nat Genet** 38: 594–597

Clark SE, Running MP, Meyerowitz EM (1993) CLAVATA1, a regulator of meristem and flower development in *Arabidopsis*. **Development** 119: 397–418

Doebley J, Bacigalupo A, Stec A (1994) Inheritance of kernel weight in two maize-teosinte hybrid populations: Implications for crop evolution. **J Hered** 85: 191–195

Fan C, Xing Y, Mao H, Lu T, Han B, Xu C, Li X, Zhang Q (2006) *GS3*, a major QTL for grain length and weight and minor QTL for grain width and thickness in rice, encodes a putative transmembrane protein. **Theor Appl Genel** 112: 1164–1171

Fu J, Cheng Y, Linghu J, Yang X, Kang L, Zhang Z, Zhang J, He C, Du X, Peng Z, Wang B, Zhai L, Dai C, Xu J, Wang W, Li X, Zheng J, Chen L, Luo L, Liu J, Qiaan X, Yan J, Wang J, Wang G (2013) RNA sequencing reveals the complex regulatory network in the maize kernel. **Nat Commun** 4: 2832

Ganal MW, Durstewitz G, Polley A, Bérard A, Buckler ES, Charcosset A, Clarke JD, Graner EM, Hansen M, Joets J, Le Paslier MC, McMullen MD, Montalent P, Rose M, Schön CC, Sun Q, Walter H, Martin OC, Falque M (2011) A large maize (*Zea mays* L.) SNP genotyping array: Development and germplasm genotyping, and genetic mapping to compare with the B73 reference genome. **PLoS ONE** 6: e28334

Hall TA (1999) BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. **Nucl Acids Symp Ser** 41: 95–98

Hecht V, Vielle-Calzada JP, Hartog MV, Schmidt ED, Boutilier K, Grossniklaus U, de Vries SC (2001) The *Arabidopsis SOMATIC EMBRYOGENESIS RECEPTOR KINASE 1* gene is expressed in developing ovules and embryos and enhances embryogenic competence in culture. **Plant Physiol** 127: 803–816

Hirokawa T, Boon-Chieng S, Mitaku S (1998) SOSUI: Classification and secondary structure prediction system for membrane proteins. **Bioinformatics** 14: 378–379

Huang X, Qian Q, Liu Z, Sun H, He S, Luo D, Xia G, Chu C, Li J, Fu X (2009) Natural variation at the DEP1 locus enhances grain yield in rice. **Nat Genet** 41: 494–497

Jiao Y, Wang Y, Xue D, Wang J, Yan M, Liu G, Dong G, Zeng D, Lu Z, Zhu X, Qian Q, Li J (2010) Regulation of *OsSPL14* by OsmiR156 defines ideal plant architecture in rice. **Nat Genet** 42: 541–544

Kump KL, Bradbury PJ, Wisser RJ, Buckler ES, Belcher AR, Oropeza-Rosas MA, Zwonitzer JC, Kresovich S, McMullen MD, Ware D, Balint-Kurti PJ, Holland JB (2011) Genome-wide association study of quantitative resistance to southern leaf blight in the maize nested association mapping population. **Nat Genet** 43: 163–168

Li H, Peng Z, Yang X, Wang W, Fu J, Wang J, Han Y, Chai Y, Guo T, Yang N, Liu J, Warburton ML, Cheng Y, Hao X, Zhang P, Zhao J, Liu Y, Wang G, Li J, Yan J (2013) Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels. **Nat Genet** 45: 43–50

Li J, Chory J (1997) A putative leucine-rich repeat receptor kinase involved in brassinosteroid signal transduction. **Cell** 90: 929–938

Li Q, Yang X, Bai G, Warburton ML, Mahuku G, Gore M, Dai J, Li J, Yan J (2010) Cloning and characterization of a putative *GS3* ortholog involved in maize kernel development. **Theor Appl Genet** 120: 753–763

Li Q, Yang X, Warburton ML, Bai G, Dai J, Li J, Yan J (2010) Relationship, evolutionary fate and function of two maize co-orthogogs of rice *GW2* associated with kernel size and weight. **BMC Plant Biol** 10: 143

Li Q, Yang X, Xu S, Cai Y, Zhang D, Han Y, Li L, Zhang Z, Gao S, Li J, Yan J (2012) Genome-wide association studies identified three independent polymorphisms associated with α-tocopherol content in maize kernels. **PLoS ONE** 7: e36807

Li Y, Fan C, Xing Y, Jiang Y, Luo L, Sun L, Shao D, Xu C, Li X, Xiao J, He Y, Zhang Q (2011) Natural variation in *GS5* plays an important role in regulating grain size and yield in rice. **Nat Genet** 43: 1266–1269

Mao H, Sun S, Yao J, Wang C, Yu S, Xu C, Li X, Zhang Q (2010) Linking differential domain functions of the GS3 protein to natural variation of grain size in rice. **Proc Natl Acad Sci USA** 107: 19579–19584

Melchinger, AE, Utz HF, Schon CC (1998) Quantitative trait locus (QTL) mapping using different testers and independent population samples in maize reveals low power of QTL detection and large bias in estimates of QTL effects. **Genetics** 149: 383–403

Miller CT, Beleza S, Pollen AA, Schluter D, Kittles RA, Shriver M, Kinsley DM (2007) *Cis*-regulatory changes in *Kit Lig and* expression and parallel evolution of pigmentation in sticklebacks and humans. **Cell** 131: 1179–1189

Miura K, Ikeda M, Matsubara A, Song X, Ito M, Asano K, Matsuoka M, Kitano H, Ashikari M (2010) *OsSPL14* promotes panicle branching and higher grain productivity in rice. **Nat Genet** 42: 545–549

Montoya T, Nomura T, Farrar K, Kaneta T, Yokota T, Bishop GJ (2002) Cloning the tomato *curl3* gene highlights the putative dual role of the leucine-rich repeat receptor kinase tBRI1/SR160 in plant steroid hormone and peptide hormone signaling. **Plant Cell** 14: 3163–3176

Moore G, Debos KM, Wang Z, Gale M.D. (1995) Cereal genome evolution: Grasses line up and form a circle. **Curr Biol** 5: 737–739

Schnable SP, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C, Zhang J, Fulton L, Graves TA, Minx P, Reily AD, Courtney L, Kruchowski SS, Tomlinson C, Strong C, Delehaunty K, Fronick C, Courtney B, Rock SM, Belter E, Du F, Kim K, Abbott RM, Cotton M, Levy A, Marchetto P, Ochoa K, Jackson SM, Gillam B, Chen W, Yan L, Higginbotham J, Cardenas M, Waligorski J, Applebaum E, Phelps L, Falcone J, Kanchi K, Thane T, Scimone A, Thane N, Henke J, Wang T, Ruppert J, Shah N, Rotter K, Hodges J, Ingenthron E, Cordes M, Kohlberg S, Sgro J, Delgado B, Mead K, Chinwalla A, Leonard S, Crouse K, Collura K, Kudrna D, Currie J, He R, Angelova A, Rajasekar S, Mueller T, Lomeli R, Scara G, Ko A, Delaney K, Wissotski M, Lopez G, Campos D, Braidotti M, Ashley E, Golser W, Kim H, Lee S, Lin J, Dujmic Z, Kim W, Talag J, Zuccolo A, Fan C, Sebastian A, Kramer M, Spiegel L, Nascimento L, Zutavern T, Miller B, Ambroise C, Muller S, Spooner W, Narechania A, Ren L, Wei S, Kumari S, Faga B, Levy MJ, McMahan L, Van Buren P, Vaughn MW, Ying K, Yeh CT, Emrich SJ, Jia Y, Kalyanaraman A, Hsia AP, Barbazuk WB, Baucom RS, Brutnell TP, Carpita NC, Chaparro C, Chia JM, Deragon JM, Estill JC, Fu Y, Jeddeloh JA, Han Y, Lee H, Li P, Lisch DR, Liu S, Liu Z, Nagel DH, McCann MC, SanMiguel P, Myers AM, Nettleton D, Nguyen J, Penning BW, Ponnala L, Schneider KL, Schwartz DC, Sharma A, Soderlund C, Springer NM, Sun Q, Wang H, Waterman M, Westerman R, Wolfgruber TK, Yang L, Yu Y, Zhang L, Zhou S, Zhu Q, Bennetzen JL, Dawe RK, Jiang J, Jiang N, Presting GG, Wessler SR, Aluru S, Martienssen RA, Clifton SW, McCombie WR, Wing RA, Wilson RK. (2009) The B73 maize genome: Complexity, diversity, and dynamics. **Science** 326: 1112–1115

Remmert M, Biegert A, Hauser A, Söding J (2012) HHblits: Lightning-fast iterative protein sequence searching by HMM-HMM alignment. **Nat Methods** 9: 173–175

Scheer JM, Ryan CA JR (2002) The systemin receptor SR160 from *Lycopersicon peruvianum* is a member of the LRR receptor kinase family. **Proc Natl Acad Sci USA** 99: 9585–9590

Shomura A, Izawa T, Ebana K, Ebitani T, Kanegae H, Konishi S, Yano M (2008) Deletion in a gene associated with grain size increased yields during rice domestication. **Nat Genet** 40: 1023–1028

Soderlund C, Bomhoff M, Nelson W (2011) SyMAP v3.4: A turnkey synteny system with application to plant genomes. **Nucleic Acids Res** 39: e68

Song XJ, Huang W, Shi M, Zhu MZ, Lin HX (2007) A QTL for rice grain width and weight encodes a previously unknown RING-type E3 ubiquitin ligase. **Nat Genet** 39: 623–630

Studer A, Zhao Q, Ross-Ibarra J, Doebley J (2011) Identification of a functional transposon insertion in the maize domestication gene *tb1*. **Nat Genet** 43: 1160–1163

Takeda T, Suwa Y, Suzuki M, Kitano H, Ueguchi-Tanaka M, Ashikari M, Matsuoka M, Ueguchi C (2003) The *OsTB1* gene negatively regulates lateral branching in rice. **Plant J** 33: 513–520

Tian F, Bradbury PJ, Brown PJ, Hung H, Sun Q, Flint-Garcia S, Rocheford TR, McMullen MD, Holland JB, Buckler ES (2011) Genome-wide association study of leaf architecture in the maize nested association mapping population. **Nat Genet** 43: 159–162

Van Bel M, Proost S, Wischnitzki E, Movahedi S, Scheerlinck C, Van de Peer Y, Vandepoele K (2012) Dissecting plant genomes with the PLAZA comparative genomics platform. **Plant Physiol** 158: 590–600

Veldboom LR, Lee M (1994) Molecular-marker-facilitated studies of morphological traits in maize. II: Determination of QTLs for grain yield and yield components. **Theor Appl Genet** 89: 451–458

Veldboom LR, Lee M (1996) Genetic mapping of quantitative trait loci in maize in stress and nonstress environments: IGrain yield and yield components. **Crop Sci** 36: 1310–1319

Wang E, Wang J, Zhu X, Hao W, Wang L, Li Q, Zhang L, He W, Lu B, Lin H, Ma H, Zhang G, He Z (2008) Control of rice grain-filling and yield by a gene with a potential signature of domestication. **Nat Genet** 40: 1370–1374

Xing Y, Zhang Q (2010) Genetic and molecular bases of rice yield. **Annu Rev Plant Biol** 61: 421–442

Xu C, Liu Y, Li Y, Xu X, Xu C, Li X, Xiao J, Zhang Q (2015) Differential expression of GS5 regulates grain size in rice. **J Exp Bot** 66: 2611–2623

Xue W, Xing Y, Weng X, Zhao Y, Tang W, Wang L, Zhou H, Yu S, Xu C, Li X, Zhang Q (2008) Natural variation in *Ghd7* is an important regulator of heading date and yield potential in rice. **Nat Genet** 40: 761–767

Yan J, Warburton M, Crouch J (2011) Association mapping for enhancing maize (*Zea mays* L.) genetic improvement. **Crop Sci** 51: 433–449

Yang X, Gao S, Xu S, Zhang Z, Prasanna BM, Li L, Li J, Yan J (2011) Characterization of a global germplasm collection and its potential utilization for analysis of complex quantitative traits in maize. **Mol Breeding** 28: 511–526

Yang X, Yan J, Shah T, Warburton ML, Li Q, Li L, Gao Y, Chai Y, Fu Z, Zhou Y, Xu S, Bai G, Meng Y, Zheng Y, Li J (2010) Genetic analysis and characterization of a new maize association mapping panel for quantitative trait loci dissection. **Theor Appl Genet** 121: 417–431

Yin T, Pan G, Liu H, Wu J, Li Y, Zhao Z, Fu T, Zhou Y. (2012) The chloroplast ribosomal protein L21 gene is essential for plastid development and embryogenesis in *Arabidopsis*. **Planta** 235: 907–921

Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, Kresovich S,

Buckler ES (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. **Nat Genet** 38: 203–208

Zhang C, Cahoon RE, Hunter SC, Chen M, Han J, Cahoon EB (2013) Genetic and biochemical basis for alternative routes of tocotrienol biosynthesis for enhanced vitamin E antioxidant production. **Plant J** 73: 628–639

Zhang X, Wang J, Huang J, Lan H, Wang C, Yin C, Wu Y, Tang H, Qian Q, Li J, Zhang H (2012) Rare allele of *OsPPKL1* associated with grain length causes extra-large grain and a significant yield increase in rice. **Proc Natl Acad Sci USA** 10: 21534–21539

Zhang Y, Liu J, Xia X, He Z (2014) *TaGS-D1*, an ortholog of rice *OsGS3*, is associated with grain weight and grain length in common wheat. **Mol Breeding** 34: 1097–1107

Zhao DZ, Wang GF, Speal B, Ma H (2002) The excess microsporocytes1 gene encodes a putative leucine-rich repeat receptor protein kinase that controls somatic and reproductive cell fates in the *Arabidopsis* anther. **Genes Dev** 16: 2021–2031

Zhou Y, Li G, Brandizzi F, Fowke LC, Wang H (2003) The plant cyclin-dependent kinase inhibitor ICK1 has distinct functional domains for in vivo kinase inhibition, protein instability and nuclear localization. **Plant J** 35: 476–489

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article at the publisher's web-site.

Figure S1. The trends of 1,000-seed weight from the second generation of 14 transgenic lines and wild type

Figure S2. GO analysis of 67 genes whose expressions were regulated by *ZmBAK1-7*

Table S1. 26 paralogs of *ZmGS5*

Table S2. Significant associations between SNPs and kernel-related traits

Table S3. QTLs identified within bin 3.04 and 7.03 for kernel related traits in previous studies

Table S4. The genes whose expressions were regulated by *ZmBAK1-7*

Table S5. The primers used in this study