The Conserved and Unique Genetic Architecture of Kernel Size and Weight in Maize and Rice[1][OPEN]

Jie Liu,[a] Juan Huang,[a] Huan Guo,[a] Liu Lan,[a] Hongze Wang,[a] Yuancheng Xu,[a] Xiaohong Yang,[b] Wenqiang Li,[a] Hao Tong,[a] Yingjie Xiao,[a] Qingchun Pan,[a] Feng Qiao,[a] Mohammad Sharif Raihan,[a] Haijun Liu,[a] Xuehai Zhang,[a] Ning Yang,[a] Xiaqing Wang,[a] Min Deng,[a] Minliang Jin,[a] Lijun Zhao,[a] Xin Luo,[a] Yang Zhou,[a] Xiang Li,[a] Wei Zhan,[a] Nannan Liu,[a] Hong Wang,[a] Gengshen Chen,[a] Qing Li,[a,2] and Jianbing Yan[a,2]

[a]National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan 430070, China

[b]National Maize Improvement Center of China, MOA Key Laboratory of Maize Biology, Beijing Key Laboratory of Crop Genetic Improvement, China Agricultural University, Beijing 100193, China

ORCID IDs: 0000-0002-1129-9584 (J.L.); 0000-0001-9190-2605 (J.H.); 0000-0001-6726-7576 (H.G.); 0000-0001-7734-706X (X.Z.); 0000-0001-8630-3464 (Y.Z.); 0000-0002-7015-6006 (X.L.); 0000-0003-3232-9479 (Q.L.); 0000-0001-8650-7811 (J.Y.).

Maize (*Zea mays*) is a major staple crop. Maize kernel size and weight are important contributors to its yield. Here, we measured kernel length, kernel width, kernel thickness, hundred kernel weight, and kernel test weight in 10 recombinant inbred line populations and dissected their genetic architecture using three statistical models. In total, 729 quantitative trait loci (QTLs) were identified, many of which were identified in all three models, including 22 major QTLs that each can explain more than 10% of phenotypic variation. To provide candidate genes for these QTLs, we identified 30 maize genes that are orthologs of 18 rice (*Oryza sativa*) genes reported to affect rice seed size or weight. Interestingly, 24 of these 30 genes are located in the identified QTLs or within 1 Mb of the significant single-nucleotide polymorphisms. We further confirmed the effects of five genes on maize kernel size/weight in an independent association mapping panel with 540 lines by candidate gene association analysis. Lastly, the function of *ZmINCW1*, a homolog of rice *GRAIN INCOMPLETE FILLING1* that affects seed size and weight, was characterized in detail. *ZmINCW1* is close to QTL peaks for kernel size/weight (less than 1 Mb) and contains significant single-nucleotide polymorphisms affecting kernel size/weight in the association panel. Overexpression of this gene can rescue the reduced weight of the Arabidopsis (*Arabidopsis thaliana*) homozygous mutant line in the *AtcwINV2* gene (Arabidopsis ortholog of *ZmINCW1*). These results indicate that the molecular mechanisms affecting seed development are conserved in maize, rice, and possibly Arabidopsis.

Maize (*Zea mays*) is one of the most important crops and is cultivated worldwide as a source of staple food, animal feed, and industrial materials. According to the Food and Agriculture Organization, the production of maize was 1,016.7 million tons in 2013, which was far more than rice (*Oryza sativa*) and wheat (*Triticum aestivum*; 745.7 and 713.1 million tons, respectively). Yield improvement is a central goal of maize breeding. Kernel size and weight are two significant components of maize yield, and many attempts have been made to elucidate the genetic basis of kernel size and weight.

Many studies have mapped quantitative trait loci (QTLs) for natural variations in kernel size and weight. For example, Liu et al. (2014) identified 55 QTLs for kernel size and weight in an F2 population. Raihan et al. (2016) mapped 16 major QTLs for kernel traits in a recombinant inbred line (RIL) population. Jiang et al. (2015) mapped 28 QTLs in a test cross population. Most of these studies used two diverse inbred lines to develop the segregating population and used a limited number of genetic markers to construct the linkage map, which greatly limited the resolution and power to detect rare and/or small-effect QTLs. Large-scale QTL mapping studies including more diverse genetic backgrounds and dense genetic markers would provide more insight into the number and effect of QTLs controlling the natural variations of kernel size and weight in maize.

[1] This work was supported by grants from the National Science Foundation of China (91635303), by the National Key Research and Development Program of China (2016YFD0101003), by Genetically Modified Organisms Breeding Major Projects (2014ZX08009-044B), and by the Huazhong Agricultural University Scientific and Technological Self-Innovation Foundation.

[2] Address correspondence to qingli@mail.hzau.edu.cn or yjianbing@mail.hzau.edu.cn.

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantphysiol.org) is: Jianbing Yan (yjianbing@mail.hzau.edu.cn).

J.Y. and Q.L. designed and supervised this study; J.L., H.G., and J.H. analyzed the data; J.L., J.H., H.G., L.L., H.W., Y.X., X.Y., W.L., H.T., Y.X., Q.P., F.Q., M.S.R., H.L., X.Z., N.Y., X.W., M.D., M.J., L.Z., X.L., Y.Z., X.L., W.Z., N.L., H.W., and G.C. performed the experiments; J.L., Q.L., and J.Y. prepared the article; all authors read and approved the article.

[OPEN] Articles can be viewed without a subscription.

www.plantphysiol.org/cgi/doi/10.1104/pp.17.00708

Despite the large number of QTLs that have been identified for maize kernel size and weight, none of them has been delimited to the causative variation. On the other hand, many genes controlling maize kernel development have been cloned using kernel mutants identified from Robertson's *Mutator* stocks, including *emp2* (*empty pericarp2*), *emp4*, *emp5*, *emp16*, *dek1* (*defective kernel1*), *dek35*, *maize pentatricopeptide repeat6*, *small kernel1*, *embryo defective14*, *U6 biogenesis-like1*, and many others (Fu et al., 2002; Lid et al., 2002; Gutiérrez-Marcos et al., 2007; Manavski et al., 2012; Liu et al., 2013; Li et al., 2014, 2015, 2017; Chen et al., 2016; Xiu et al., 2016). Mutations in these genes usually have severe phenotypes in kernels, such as empty pericarp, where both embryo and endosperm cannot develop properly. It is unclear whether weak mutations (or genetic variations) of these genes exist in nature and whether such genetic variants can contribute to phenotypic diversity in maize kernel. A comparative analysis between these mutant genes and kernel size/weight QTLs will not only provide candidate genes for QTLs but also shed light on the extent to which genes identified through mutant studies can contribute to natural variations in maize kernel phenotypes.

Maize shares common ancestors with rice (Murat et al., 2017). Comparative QTL studies between species showed that similar traits were usually controlled by QTLs that are located within syntenic regions among the species (Paterson et al., 1995). This idea has been illustrated further by comparative functional studies at the single gene level. Many genes that could affect the seed shape and weight have been fine-mapped and cloned in rice, such as *GS3* (Fan et al., 2006; Mao et al., 2010), *GW2* (Song et al., 2007), and *GS5* (Li et al., 2011). Li et al. (2010a, 2010b) isolated the maize orthologs of rice *GS3* and *GW2* and showed that the maize genes also control similar traits, although with different genetic variations. Similarly, Liu et al. (2015) showed that *GS5* contributes to kernel size variation in maize as well as in rice. Notably, the wheat orthologs of rice *GW2* and *GS5* also were associated significantly with wheat kernel size and weight (Su et al., 2011; Hong et al., 2014; Qin et al., 2014; Jaiswal et al., 2015; Wang et al., 2015, 2016; Ma et al., 2016; Simmonds et al., 2016). In addition to *GS3*, *GW2*, and *GS5*, many other genes controlling rice kernel size/weight have been cloned, such as genes involved in G-protein signaling (*DEP1* and *D1*) and genes from phytohormone pathways (*DST* and *Gn1a* for cytokinin; *D11*, *SRS5*, *D61*, *qGL3*, and *SMG1* for brassinosteroid; and *TGW6* for auxin). It remains an open question whether their maize homologous counterparts have a similar function in the phenotypic diversity of kernel size/weight. A systematic investigation of the function of these genes in maize would provide more insights into the genetic mechanisms controlling kernel development in the two closely related important crops.

In this study, we used 10 RIL populations to dissect the genetic basis of maize kernel size and weight with three models: separate linkage mapping (SLM), joint linkage mapping (JLM), and genome-wide association mapping (GWAS). Many QTLs with major and minor effects were identified. A comparison between these QTLs and the genes from maize mutant studies and maize homologs of well-known rice seed size/weight genes suggested that many of these genes have roles in controlling natural variations of kernel size and weight. There are also many QTLs that are not coincident with any known candidate genes or that contain candidate genes that do not affect natural variations in kernel size and weight. These results suggest both a conserved and species-specific genetic architecture of kernel traits between rice and maize. Furthermore, we found that *ZmINCW1*, an ortholog of the rice seed weight gene *GRAIN INCOMPLETE FILLING1* (*GIF1*), had a conserved function that affects kernel/seed development in maize and Arabidopsis (*Arabidopsis thaliana*). Our results help to elucidate the genetic basis of maize kernel size and weight.

RESULTS

Phenotypic Variation and Heritability of Kernel Size and Weight

We used 10 RIL populations derived from 14 diverse maize inbred lines (Pan et al., 2016) to dissect the genetic architecture of kernel size and weight in maize. These lines were grown under multiple environments. Seven of them (B73 × BY804, KUI3 × B77, K22 × CI7, DAN340 × K22, ZHENG58 × SK, YU87-1 × BK, and ZONG3 × YU87-1) were planted in eight environments, while the other three (DE3 × BY815, K22 × BY815, and BY815 × KUI3) were planted in four environments. Five maize kernel traits were measured for each line in these 10 RIL populations, including hundred kernel weight (HKW; weight of 100 kernels), kernel test weight (KTW; weight of 250 mL of kernels), kernel length (KL), kernel thickness (KT) and kernel width (KW; Fig. 1A). Best linear unbiased prediction (BLUP) values of each line were used to represent the phenotypic value. Both the 14 parental lines and the 10 RIL populations showed significant variations in these five kernel traits (Fig. 1, B and C; Supplemental Fig. S1). Broad-sense heritability ranged from 0.53 (KT in DE3 × BY815) to 0.94 (HKW in ZHENG58 × SK), with most higher than 0.8 (Supplemental Table S1). This suggests that phenotypic variation is controlled largely by genetic factors and can be genetically mapped. We calculated the correlation coefficient between HKW and the three kernel size traits, KL, KW, and KT (Supplemental Table S2), and found that KW and KT were significantly positively correlated with HKW in all 10 RIL populations, while KL was positively correlated with HKW in seven RIL populations. The correlation coefficients for KL were usually smaller compared with KW and KT, suggesting that KW and KT may play more important roles for kernel weight in maize.

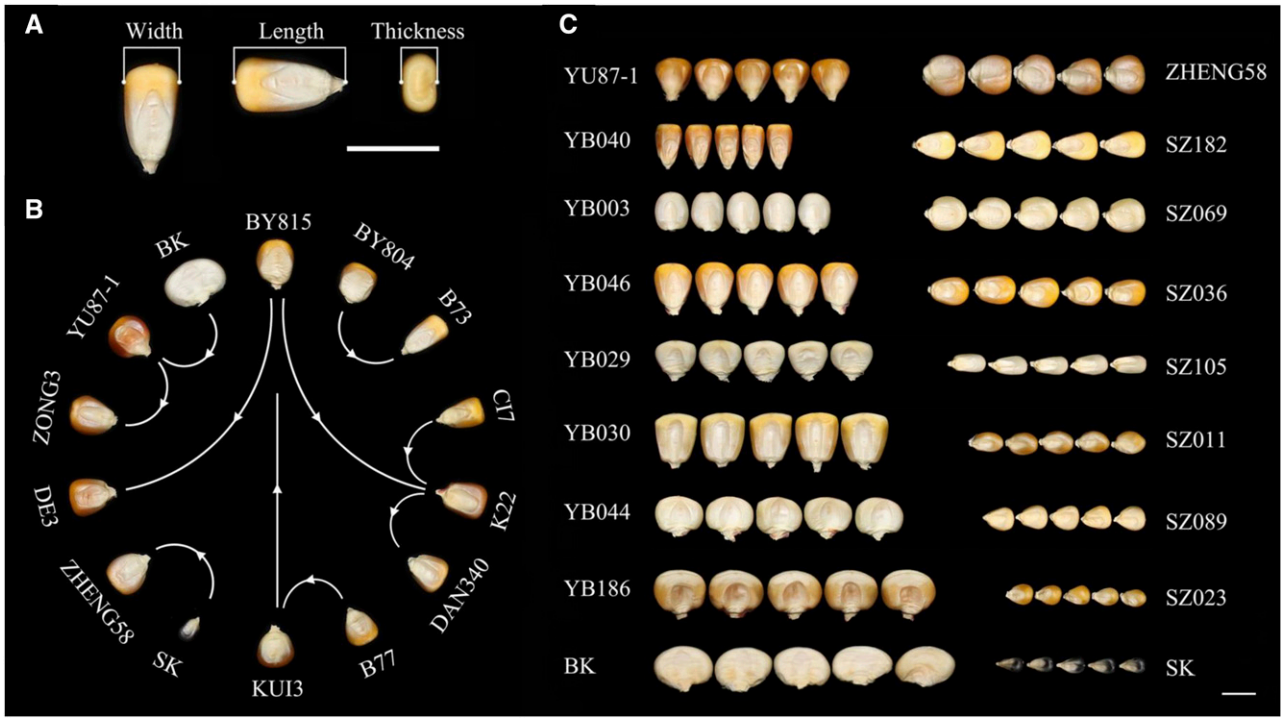

Figure 1. Measurements of kernel traits and variations of kernel size among 14 parental lines and representative lines in two RIL populations. A, Measurements of KL, KW, and KT illustrated with a B73 kernel. Bar = 1 cm. B, Fourteen parental inbred lines used in this study showed considerable variations of kernel size. The arrows point from paternal lines to maternal lines. C, Kernels of representative lines in YU87-1 × BK (left) and ZHENG58 × SK (right) RIL populations. Bar = 1 cm.

Dissection of the Genetic Architecture of Maize Kernel Size and Weight with Three Methods

First, we performed SLM in each RIL population with the composite interval mapping method (Zeng, 1994). In total, we identified 373 QTLs for kernel size and weight, including 90, 70, 61, 89, and 63 QTLs for HKW, KTW, KL, KW, and KT, respectively (Fig. 2A; Table I; Supplemental Table S3; Supplemental Figs. S2 and S3). The phenotypic variation explained by each QTL ranged from 2.91% to 19.43%, with an average of 7%. Out of these 373 QTLs, 267 (72%) were identified in only one population. Some QTLs could be identified in at least two populations for the same trait, including 27 QTLs for HKW, 24 for KTW, 15 for KL, 25 for KW, and 15 for KT. The presence of population-common and specific QTLs may reflect differences in allele frequency of the underlying causative sites and suggests that populations from diverse genetic backgrounds are needed to comprehensively understand the genetic architecture of kernel size and weight. Importantly, 10, 11, 10, 17, and nine major QTLs ($R^2 > 10\%$, i.e. QTLs that can explain more than 10% of the phenotypic variation) were identified for HKW, KTW, KL, KW, and KT, respectively. Among these major QTLs, 17 could be detected in more than one population for the same trait. We also detected 18 major QTLs that can affect more than one trait (Supplemental Tables S4 and S5). An example QTL for KW, KT, and HKW is shown in Figure 2B. Detailed information about these 373 QTLs is provided in Supplemental Table S6.

We also performed JLM and GWAS by analyzing the 10 RIL populations jointly (see "Materials and Methods"). In JLM, we identified 56, 59, 55, 68, and 62 QTLs for HKW, KTW, KL, KW, and KT, respectively (Fig. 2A; Table I; Supplemental Table S7; Supplemental Fig. S3). In GWAS, we detected between 123 and 198 significant single-nucleotide polymorphisms (SNPs) for each trait (Supplemental Table S8). To avoid redundancy of the significant SNPs caused by linkage disequilibrium, we performed a backward regression procedure (see "Materials and Methods"). After this analysis, 30, 22, 26, 32, and 25 independent SNPs for HKW, KTW, KL, KW, and KT, respectively, were obtained (Table I; Supplemental Table S9). Some SNPs can control two or more traits simultaneously. On average, each of the identified SNPs with GWAS could explain only a very small amount of phenotypic variation (between 1.06% and 1.35%) compared with QTLs identified with SLM, but they could jointly explain a large portion of phenotypic variation (55.85%, 59.17%, 59.91%, 75.72%, and 36.4% for HKW, KTW, KL, KW, and KT, respectively).

Notably, a considerable number of loci could be identified by more than one model (Fig. 2C). For example, 61.9% of QTLs identified with SLM also can be detected

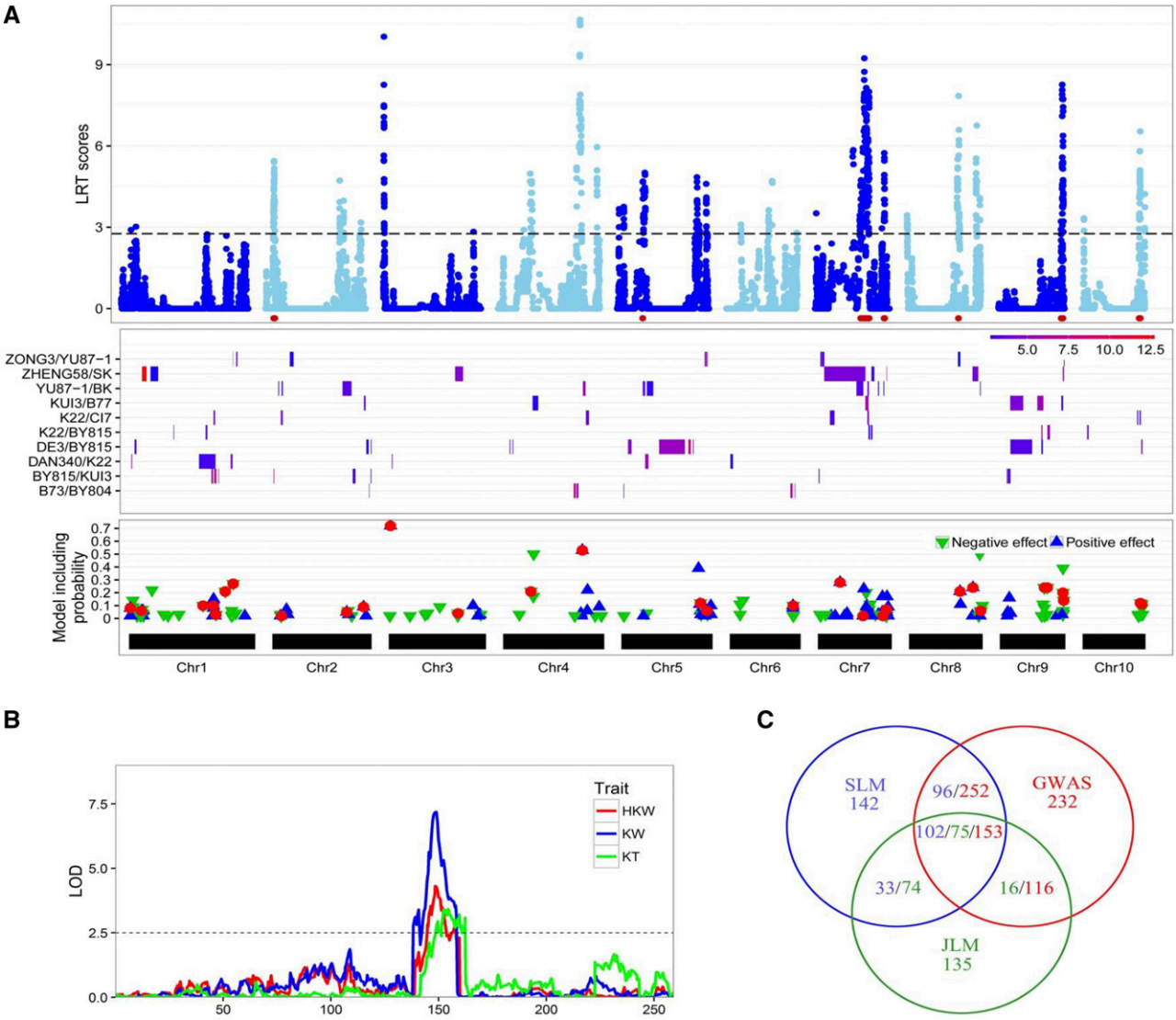

Figure 2. Overview of the QTLs and significant SNPs for HKW identified with three models. A, The top graph (Manhattan plot) shows likelihood ratio test (LRT) scores from JLM. The red points under the *x* axis indicate the significant SNPs identified in all three models. The middle graph shows the results of SLM in each of the 10 RIL populations. The colored rectangles indicate the QTL regions in each RIL population, and the color density is proportional to the LOD values. The bottom graph shows the results of GWAS. The blue upward triangles indicate that the minor allele increases HKW relative to the major allele, the green downward triangles indicate the opposite effect, and red dots indicate the candidate SNPs identified by the backward regression model. B, A pleiotropic QTL that was identified for KW, KT, and HKW (R^2 = 12.4%, 5.13%, and 6.92%, respectively) in the K22 × CI7 population. C, Number of overlapped QTLs or SNPs identified with three methods. Blue numbers are for SLM, red for GWAS, and green for JLM. For example, 102/75/116 means that 102 QTLs identified with the SLM model overlapped with 75 QTLs identified with the JLM model and 116 SNPs from the GWAS model.

using JLM and/or GWAS. Similarly, 55% of JLM QTLs and 69.2% of GWAS SNPs can be identified by the other two models. More importantly, 22 major QTLs can be identified in all three models. These results confirm the reliability of the identified QTLs and suggest that the three statistical models are complementary to each other. The integrated use of these models can provide more insight into the genetic architecture of phenotypic variation.

Natural Variations of Some Maize Mutant Genes Were Significantly Associated with Kernel Development

To identify candidate genes for the QTLs, we collected 36 maize genes that had been cloned using maize kernel mutants and were reported to be involved in maize kernel development (Supplemental Table S10). Of these 36 genes, 21 were located in the QTLs identified by SLM, seven were located in the QTLs identified by JLM, and 15 were located within a 1-Mb region of the

Table I. *QTLs or significant SNPs for kernel size and weight detected with three methods*

Mapping Method	HKW	KTW	KL	KW	KT
SLM[a]	90, 27, 10, 2	70, 24, 11, 4	61, 15, 10, 2	89, 25, 17, 7	63, 15, 9, 2
JLM	56	59	55	68	62
GWAS[b]	194, 30	136, 22	138, 26	163, 32	123, 25

[a]The number of QTLs identified in one population, more than one population, major QTLs in one population ($R^2 > 10\%$), and major QTLs in at least two populations, respectively. [b]The number of significant SNPs identified by GWAS and backward regression, respectively.

significant SNPs identified by GWAS (Supplemental Table S10). To further confirm the function of these genes, we used an independent association panel consisting of 540 lines. This panel has been genotyped with 1.25 million SNPs (Liu et al., 2017). Between one and 209 SNPs were identified in these 36 genes and were used to identify loci that are significantly associated with variations in kernel size and weight. We found that seven of these 36 genes affect at least one kernel trait (Supplemental Table S10; Supplemental Fig. S5). For example, *Dek36* (GRMZM5G892151) was significantly associated with both HKW ($P = 5.13 \times 10^{-4}$) and KTW ($P = 6.30 \times 10^{-4}$; Supplemental Fig. S5B). Many of the most significant SNPs were either located in the untranslated region or represented synonymous substitutions. This is reasonable considering that loss-of-function alleles of most genes usually lead to defective kernels with limited/no viability; thus, genetic variation greatly affecting gene function would be unfavorable under natural conditions. It is also possible that the synonymous variants are in linkage disequilibrium with the causal variants that were not assayed. Nevertheless, these findings provide evidence that many of the genes identified from mutant studies contain natural genetic variations and that many of them contribute to phenotypic diversity in maize kernel size and weight.

Many Rice Seed Size/Weight Genes Have Conserved Functions in Maize

To provide insight into whether rice seed size/weight genes have similar functions in maize, we investigated the functions of the maize orthologs of 18 rice genes that have been shown to affect seed size or weight (Supplemental Table S11). These 18 genes are involved in proteasomal degradation, phytohormones (auxin, cytokonin, and brassinosteroid), G-protein signaling, and other processes.

Based on the MSU Rice Genome Annotation Project (Kawahara et al., 2013; http://rice.plantbiology.msu.edu/cgi-bin/gbrowse/rice/), we identified 30 maize orthologs of these 18 rice genes (Supplemental Table S11), with nine genes having one ortholog, six genes having two orthologs, and three genes having three orthologs. Colocalization of these 30 genes and the identified QTLs highlighted several interesting findings. (1) Twenty-four of the 30 genes were located in the QTL confidence intervals or within 1-Mb flanking regions of the significant SNPs (Fig. 3; Supplemental Table S11). (2) Three genes were located within 500 kb of the peak, and two genes were very close to the peak (3.4 kb for *ZmGW7-2* and 59.9 kb for *ZmBG2*). (3) Six genes were located in major QTLs; for example, *ZmSLG* was located in a major QTL ($R^2 = 12.40\%$) for KW in the K22 × CI7 population and a major QTL ($R^2 = 12.89\%$) for HKW in the BY815 × KUI3 population (Supplemental Table S11). (4) Five genes were located within QTLs identified in all three models. These results suggest that the rice orthologous genes provide good candidates for the maize QTLs.

We investigated the functions of these 30 genes using the association panel consisting of 540 inbred lines. Between nine and 174 SNPs (58.3 on average) were identified within each gene and were used for candidate gene association analysis. Out of these 30 genes, five were significantly associated with at least one kernel size or weight trait by candidate gene association analysis (Fig. 3; Supplemental Fig. S4; Supplemental Table S11), and all of these five genes were located in the QTLs identified in this study. *ZmSLG* (GRMZM2G179703) was significantly associated with KW ($P = 1.06 \times 10^{-3}$; Supplemental Fig. S4A); *ZmGLW7-1* (GRMZM2G113779) was significantly associated with KT ($P = 9.44 \times 10^{-5}$) and HKW ($P = 1.04 \times 10^{-3}$; Supplemental Fig. S4B); *ZmGL2* (GRMZM2G034876) was significantly associated with KW ($P = 2.65 \times 10^{-5}$) and HKW ($P = 2.97 \times 10^{-5}$; Supplemental Fig. S4C); *ZmGW7-2* (GRMZM2G370081) was significantly associated with KW ($P = 4.57 \times 10^{-4}$; Supplemental Fig. S4D); and *ZmSRS1-2* (GRMZM2G414043) was significantly associated with KL ($P = 4.56 \times 10^{-4}$; Supplemental Fig. S4E).

The most significant SNPs in *ZmGL2* and *ZmGW7-2* are missense variants that lead to amino acid changes (Gly/Ser-211 and His/Gln-271), while the most significant SNPs in *ZmSLG*, *ZmGLW7-1*, and *ZmSRS1-2* are located in the 5′ untranslated region (UTR) or the 3′ UTR. GWAS for expression levels of these five genes showed that the expression level of *ZmSLG* was significantly associated with SNPs located in this gene ($P = 3.86 \times 10^{-11}$; Supplemental Fig. S4F). We also found a significant correlation between the expression level of *ZmSLG* and KW ($P < 0.01$, $r = -0.18$; Supplemental Fig. S4G). This indicates that the cis-element near *ZmSLG* might regulate its expression to affect kernel development. Notably, *ZmSLG* also is located within a region that has been shown to be under artificial selection during the generation of small seed and big seed

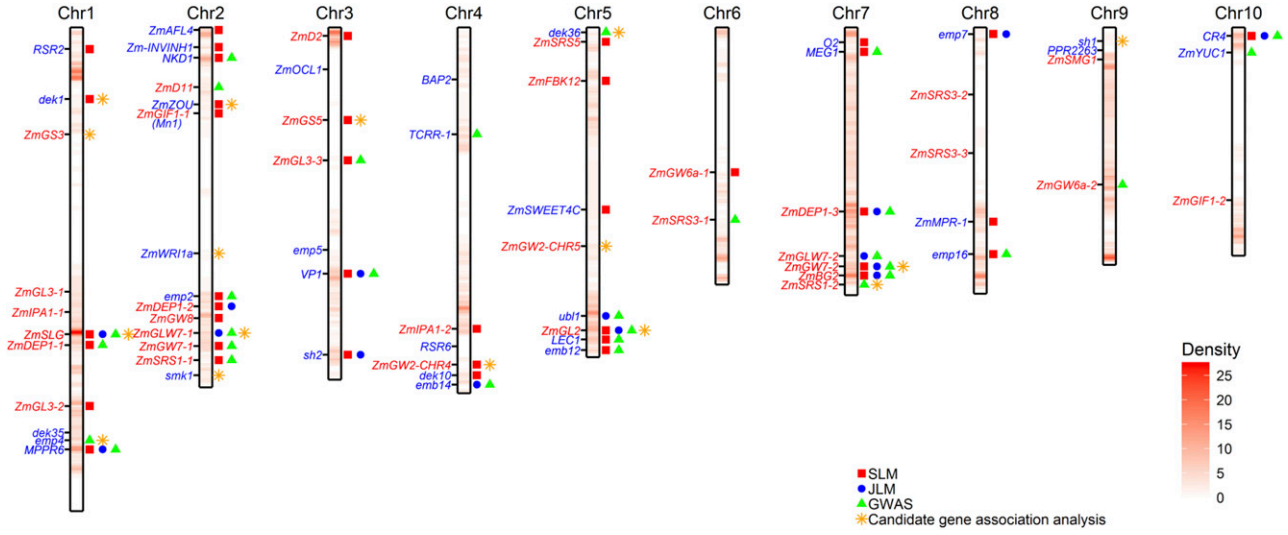

Figure 3. Comparative analysis of QTLs and genes identified from maize mutant studies or based on rice seed size or weight genes. A total of 21 rice genes (18 from this study and *GS3*, *GW2*, and *GS5* from previous studies; shown in red) and 36 maize genes (shown in blue) reported to be involved in maize kernel development in mutant studies are shown. Points with different color and shape indicate that genes were significantly associated with maize kernel size or weight by different methods. The heat map in the chromosome region indicates the density of QTLs for kernel traits (see scale at bottom right). The window size is 1 Mb.

populations by Hirsch et al. (2014a). These results provide evidence that some maize orthologous genes might have similar and conserved functions as their rice counterparts.

ZmINCW1, Whose Protein Sequence Has High Similarity to Rice GIF1, Affects Kernel Development in Maize

The maize gene *Mn1* (GRMZM2G119689; *Miniature-1*, also named *incw2*) is critical in maize kernel development, and the *mn1* seed weighs only 20% of the normal seed (Lowe and Nelson, 1946). *Mn1* encodes a cell wall invertase, and extremely low invertase activity is the causal basis of the mutant phenotype (Miller and Chourey, 1992). Because of the tetraploid origin of maize, *Mn1* has a paralog in the maize genome, GRMZM2G095725. Both *Mn1* and GRMZM2G095725 are orthologs of rice *GIF1* (Supplemental Fig. S4), which affects rice grain filling and weight (Wang et al., 2008). Interestingly, both *Mn1* and GRMZM2G095725 are located in the confidence intervals of QTLs for maize kernel traits in this study.

In addition to the two genes, we identified a third gene, GRMZM2G139300, that shares high protein sequence similarity to rice GIF1 (identity = 71.28%, E-value = 0; Supplemental Fig. S6). GRMZM2G139300 had previously been designated as the *incw1* locus on maize chromosome 5 and encodes a cell wall invertase (Shanker et al., 1995). Hereafter, it is named *ZmINCW1*. The best homolog of *ZmINCW1* in maize is *Mn1*, suggesting that *ZmINCW1* may affect kernel size and weight in maize, which also is supported by our QTL mapping and candidate gene association analysis results. QTL mapping shows that *ZmINCW1* is located in the QTLs for HKW and KW in the B73 × BY804 population (Fig. 4A), and the phenotypic variations explained by these QTLs are 14% (HKW-12YN), 8% (KW-11DHN), and 3.7% (KW-BLUP). The distances between *ZmINCW1* and the peaks of these three QTLs are less than 1 Mb (550, 417, and 786 kb, respectively), providing a good candidate for these QTLs. We also detected QTLs covering *ZmINCW1* for KL in the DAN340 × K22 population (Supplemental Fig. S7A) and for HKW in the DE3 × BY815 population (Supplemental Fig. S7B; phenotypic variations explained by these two QTLs are 8.5% and 19%, respectively). *ZmINCW1* also was significantly associated with KL ($P = 7.10 \times 10^{-6}$), KW ($P = 4.27 \times 10^{-4}$), KT ($P = 5.71 \times 10^{-4}$), and HKW ($P = 8.73 \times 10^{-3}$; Fig. 4B) by candidate gene association analysis in an independent association panel with 540 diverse inbred lines. Four SNPs, chr5.S_169457546, chr5.S_169458449, chr5.S_169458176, and chr5.S_169458817, were the most significant SNPs for KL, KW, KT, and HKW, respectively. chr5.S_169457546 is located in the second intron of *ZmINCW1*; chr5.S_169458449 and chr5.S_169458176 are synonymous variants; and chr5.S_169458817 is located in the 3′ UTR of *ZmINCW1*. It is likely that these SNPs either affect phenotypes through regulatory roles or are in linkage disequilibrium with the causal polymorphisms.

ZmINCW1 showed abundant variations in gene expression in the association panel (Fig. 4D), implying that expression changes in this gene may contribute to phenotypic diversity. To further investigate the functional mechanisms of *ZmINCW1*, we performed GWAS of the expression of *ZmINCW1* in an association mapping population consisting of 368 diverse maize inbred lines that are a subset of the 540 inbred lines. We identified four SNPs that showed significant correlations with the expression of *ZmINCW1* ($P < 7.97 \times 10^{-7}$). Notably, the

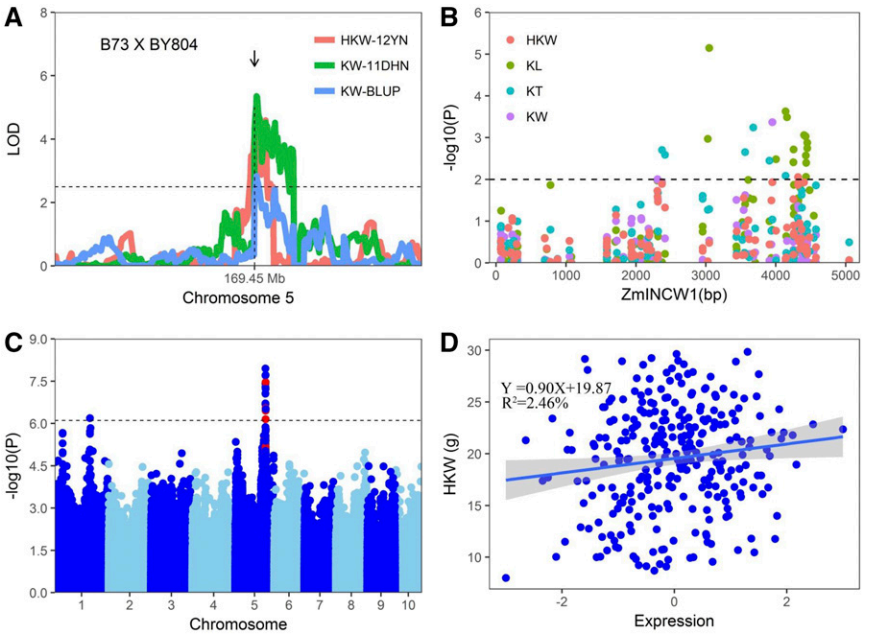

Figure 4. *ZmINCW1* was significantly associated with maize kernel development. A, *ZmINCW1* is located in the QTLs identified in the B73 × BY804 population for kernel size and weight. 12YN, Yunnan province in 2012; 11DHN, Hainan province in 2011. The arrow indicates the position of *ZmINCW1*. B, SNPs in *ZmINCW1* were significantly associated with kernel size and weight in an association panel. C, GWAS of the expression level of *ZmINCW1*. The red points indicate the SNPs located in *ZmINCW1*. D, The expression level of *ZmINCW1* was significantly positively correlated with HKW in 2011 Yunnan ($n = 292$, $r = 0.16$, $P = 7.30 \times 10^{-3}$).

most significant SNP (chr5.S_169456915) also showed significant correlation with KT ($P = 2.57 \times 10^{-3}$). Besides the significant SNPs located within *ZmINCW1*, there were some significant SNPs located ~220 kb upstream of the gene (three SNPs in the intergenic region and four SNPs in three other genes; Fig. 4C). We then performed GWAS conditioning on the most significant SNPs in *ZmINCW1* and found no other significant SNPs for the expression of *ZmINCW1*. This result indicates that the expression of *ZmINCW1* is regulated mainly by nearby variations. Interestingly, the expression levels of *ZmINCW1* are associated significantly with kernel traits in two of 12 environments (KL in 2011 Hainan, $r = 0.17$, $P = 1.87 \times 10^{-3}$; HKW in 2011 Yunnan, $r = 0.16$, $P = 7.3 \times 10^{-3}$; Fig. 4D). The failure to detect significant associations in the other 10 environments indicates an environmental effect of this gene. Together, these results support the notion that cis-variations around *ZmINCW1* affect its expression, which, in turn, can control kernel size and weight in maize.

Overexpression of Maize *ZmINCW1* in Arabidopsis Can Increase Seed Weight

To further verify the function of *ZmINCW1*, we identified a T-DNA mutant of the *AtcwINV2* gene (stock, SALK_068113C; http://www.arabidopsis.org/), which is the ortholog of *ZmINCW1* in Arabidopsis (Supplemental Fig. S5). This line has a T-DNA insertion in the fourth exon of *AtcwINV2* (Fig. 5, A and B), and this insertion disturbs the glycosyl hydrolase C-terminal domain. Compared with wild-type (Columbia-0) controls, the homozygous mutants showed normal growth and could produce normal seed (Fig. 5, C and D). However, the thousand seed weight was reduced significantly in the mutant plant (16.85 mg versus 16.14 mg, one-way ANOVA, $n = 20/32$, $P = 7.34 \times 10^{-6}$; Fig. 5E), suggesting that *AtcwINV2* could affect Arabidopsis seed development.

Next, we overexpressed *ZmINCW1* in this mutant, and the expression level of *ZmINCW1* was confirmed using RT-PCR and western blot (Fig. 5, F and G). We screened two positive transgenic lines (T1) that expressed *ZmINCW1* and eight negative transgenic lines that had no detectable expression of *ZmINCW1* or that did not have the transgene. Both positive and negative T1 transgenic lines bear normal seeds (Fig. 5, H and I); however, the positive T1 transgenic lines had increased thousand seed weight compared with the negative T1 transgenic lines by 28.35% (21.52 mg versus 16.77 mg, one-way ANOVA, $n = 6/26$, $P = 1.51 \times 10^{-12}$; Fig. 5J). We also found significant differences between wild-type and negative T2 transgenic lines ($n = 7/5$, $P = 9.91 \times 10^{-3}$; Fig. 5K) and between positive transgenic lines and negative transgenic lines in the T2 generation ($n = 6/5$, $P = 0.03$; Fig. 5K). These findings suggest that *ZmINCW1* has conserved function for seed development in maize and Arabidopsis.

DISCUSSION

Maize has tremendous phenotypic and genotypic diversity. It has been estimated that there is a polymorphic site in every 44 bp on average, that the B73 reference genome sequence may capture only ~70% of the low-copy genome fraction represented by 27 diverse maize inbred lines (Gore et al., 2009), and that ~50% of the representative transcript assemblies identified from 503 maize inbred lines are not present in B73

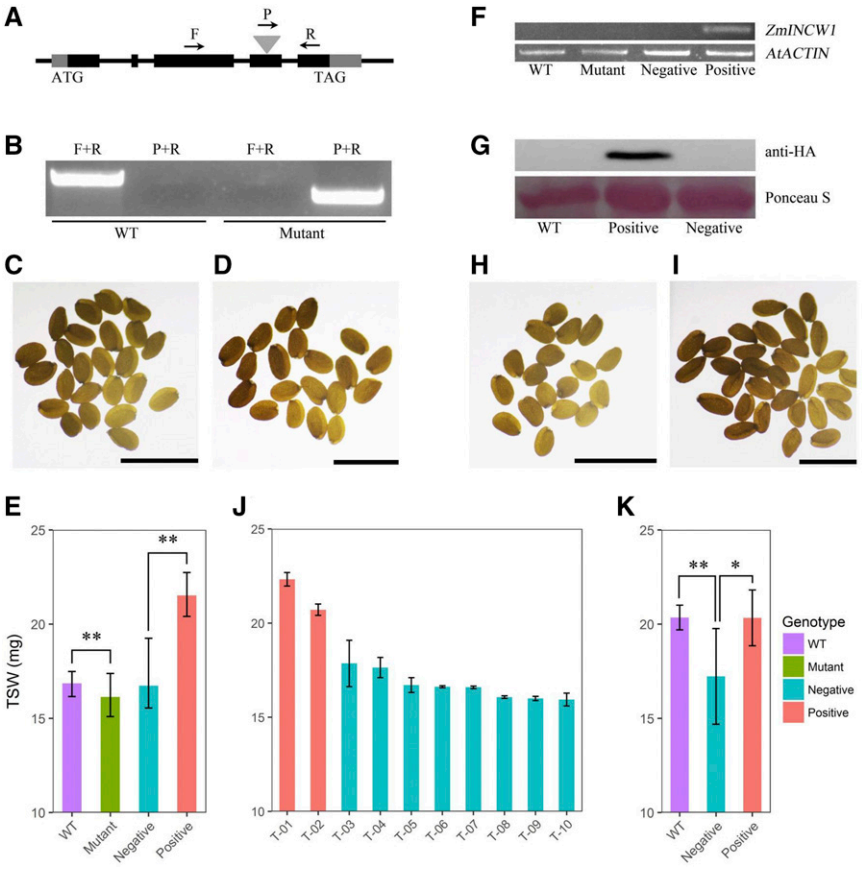

Figure 5. Overexpression of *ZmINCW1* rescues the reduced thousand seed weight in the Arabidopsis *AtcwINV2* T-DNA mutant. A and B, The T-DNA was inserted into the fourth exon of *AtcwINV2*, and three primers were used to confirm the insertion (F + R and P + R). C and D, Both the wild type (C) and the T-DNA insertion mutant (D) had normal seeds. Bars = 1 mm. E, The T-DNA insertion mutant had decreased thousand seed weight (TSW) compared with the wild type (WT; $n = 20/32$, $P = 7.34 \times 10^{-6}$). F and G, The expression levels of *ZmINCW1* in Arabidopsis were confirmed by reverse transcription-PCR (F) and western blot (G). H and I, Both negative (H) and positive (I) transgenic lines had normal seeds. Bars = 1 mm. J, T1 positive transgenic lines had increased thousand seed weight compared with negative transgenic lines ($n = 6/26$, $P = 1.51 \times 10^{-12}$). The combined result is shown in E. K, There was a significant difference in thousand seed weight between positive lines and negative lines in the T2 generation ($n = 6/5$, $P = 0.03$). **, $P < 0.01$ and *, $P < 0.05$.

(Hirsch et al., 2014b). Jin et al. (2016) identified 13,382 genes with expression presence/absence in 368 maize inbred lines and found that 788 novel genes were associated with 487 metabolic traits and a novel gene was associated with kernel width. These pan-genome and pan-transcriptome analyses showed great diversity and the importance of presence/absence or expression presence/absence for agronomic traits (Lai et al., 2010; Jin et al., 2016).

In this study, we mapped 373 QTLs with the SLM model for natural variation of maize kernel size and weight with 10 RIL populations. This represented 309 independent loci. The total numbers of QTLs identified by other studies, which used only one population, ranged from 12 to 55 (Liu et al., 2014; Zhang et al., 2014; Jiang et al., 2015; Chen et al., 2016; Raihan et al., 2016). Compared with the results published previously, we identified many more QTLs for kernel size and weight. There might be two main reasons for the greater number of QTLs identified in this study. First, RIL populations had more recombinant events compared with F2 populations used by other groups. Second, more diversity was introduced with more parental inbred lines (14 in this study) compared with two or four lines used by others. Neuffer and Sheridan (1980) estimated that maize kernel mutants map to 285 loci. This number is very close to the number of loci (309) identified for natural variation of kernel size and weight by this study. This estimate and our results show the complexity of the genetic basis of maize kernel size and weight.

Cell wall invertase, which hydrolyzes Suc into Glc and Fru, plays an important role in plant growth and development. Transgenic carrot (*Daucus carota*) plants with reduced cell wall and vacuolar invertase activity had altered phenotypes at the very early stages of development and reduced tap root development leading to smaller organ size (Tang et al., 1999). Rice *GIF1*, a cell wall invertase, was reported to affect grain filling and weight (the *gif1* mutant had ~24% lighter seeds compared with the wild type) and was a domestication gene (Wang et al., 2008). In maize, *Mn1* (*incw2*), which also encodes a cell wall invertase, was confirmed to be important for kernel development. Kernel weight of the *Mn1* mutant was about 20% of wild-type weight due to the low cell wall invertase activity (Lowe and Nelson, 1946; Miller and Chourey, 1992). *ZmINCW1* also encoded a cell wall invertase and was located in QTLs mapped in three RIL populations and was significantly associated with kernel size and weight. T-DNA insertion lines of the Arabidopsis ortholog (*AtcwINV2*) of *ZmINCW1* had reduced seed weight. Transformation of *ZmINCW1* into this mutant increased seed weight, which indicates that *ZmINCW1* had conserved function for kernel/seed development in maize and Arabidopsis. Similar conserved function also was reported

for *Mn1*, *OsGIF1*, and *AtcwINV1*. Constitutive expression of *Mn1*, *AtcwINV1*, and *OsGIF1* via a transgenic method in an elite maize inbred line (Ye478) produced larger cobs and kernels, leading to up to 145.3% improvement in grain yield (Li et al., 2013). These results suggest that genes encoding cell wall invertase might be a good choice for yield improvement through marker-assisted selection or genetic engineering.

Comparative genetic analysis is a powerful method for identifying genes that have conserved functions across species, such as flowering time (*Ghd7* in rice and *ZmCCT* in maize; Xue et al., 2008; Hung et al., 2012; Yang et al., 2013) and branching regulation (*tb1* in maize and *OsTB1* in rice; Takeda et al., 2003; Clark et al., 2006). Seed size and weight are two of the most important agronomic traits for yield and undergo selection during domestication. In rice, many genes affecting kernel development have been cloned (Supplemental Table S7), such as *GS3*, *GW2*, and *GS5*. Their orthologs in maize, *ZmGS3* (Li et al., 2010a), *ZmGW2-CHR4* and *ZmGW2-CHR5* (Li et al., 2010b), and *ZmGS5* (Liu et al., 2015), also were found to be involved in kernel development, but with different mechanisms. Here, we found that *ZmINCW1* has conserved function for kernel/seed weight in maize and Arabidopsis and that expression regulation by cis-elements might be the cause of the phenotypic change. This is very different from rice *GIF1*, where a one-nucleotide deletion caused the premature termination of its open reading frame. These findings suggest that even though these genes have conserved functions, the types of genetic variation important for the phenotype may be different between species.

We used comparative genetic analysis and identified 30 genes that are orthologs of 18 cloned rice genes for seed size or weight. Among these 30 genes, 26 are located in the candidate region mapped by at least one method (SLM, JLM, and GWAS) in the RIL populations and five were found to be significantly associated with kernel traits by candidate gene association mapping in a large association panel. Given the conserved functions of many of the known genes for kernel development in maize, rice, and wheat (Su et al., 2011; Hong et al., 2014; Qin et al., 2014; Jaiswal et al., 2015; Wang et al., 2015, 2016; Ma et al., 2016; Simmonds et al., 2016), these genes represent additional candidates for kernel development across various species.

In summary, our findings shed light on the genetic basis of kernel size and weight in maize. We provided candidate genes for many of the loci that contribute to natural variation in maize kernel size and weight. We also provided evidence for a conserved and unique genetic architecture of kernel traits in maize compared with rice.

MATERIALS AND METHODS

Plant Materials and Phenotype Measurements

Seven RIL populations (B73 × BY804, KUI3 × B77, K22 × CI7, DAN340 × K22, ZHENG58 × SK, YU87-1 × BK, and ZONG3 × YU87-1) were planted in eight trials in Hubei, Chongqing, Henan, Yunnan, and Hainan province in China during 2011 and 2012, while the other three RIL populations (DE3 × BY815, K22 × BY815, and BY815 × KUI3) were planted in four trials (Chongqing, Hubei, Henan, and Yunnan province in China during 2012) because of insufficient seeds for field trials in 2011. An association mapping population consisting of 540 inbred lines (AM540) also was planted in these eight environments during 2011 and 2012. All populations were planted with one random block replication per location. For each line, we planted 11 plants per row and selected five well-pollinated ears in the middle of the row to measure five kernel size and weight traits (i.e. KL, KW, KT, HKW, and KTW). Before measuring traits, we first mixed kernels of these five ears and used a digital ruler to measure KL, KW, and KT of 30 single kernels (illustrated in Fig. 1A). The average of these 30 kernels was used to represent the trait measurement. We measured HKW three times for a single line and used the mean value to represent HKW for that line. We used 250 mL of kernels to measure the KTW for each line, and if there were not enough kernels to measure 250 mL, we used at least 50 mL of kernels to measure the weight and then converted it to the 250-mL weight.

We used the BLUP value of each line to perform data analysis, including phenotype statistics, correlation analysis, and QTL mapping. BLUP values were computed by PROC MIXED in the Statistical Analysis System, and Pearson correlation coefficients were calculated with Excel. The heritability for each trait was calculated as follows: $H^2 = \delta^2_g/(\delta^2_g + \delta^2_e/n)$, where δ^2_g is the genetic variance, δ^2_e is the residual variance, and n is the number of environments.

Genotype

Ten RIL populations used in this study have been genotyped with the Illumina MaizeSNP50 BeadChip, with each population having 11,360 to 15,285 polymorphic markers, and these polymorphic markers were used to construct a high-density linkage map (Pan et al., 2016). These 10 populations contain 1,979 to 3,071 genetic blocks in which no recombinant events occur. The 14 inbred lines used to construct the 10 RIL populations also were contained in the 368 lines that were genotyped by RNA sequencing in a previous study (Fu et al., 2013). Thus, we projected the 1.03 million SNP genotypes of the 14 parental lines onto their 1,887 offspring RILs using a two-step imputation strategy. We first used a method similar to the aforementioned imputation to separately project high-density SNPs from two parents onto offspring RILs based on the linkage map for each population, and then we mapped the projected genotypes of RILs to base pairs according to the parental genotypes. In total, there were 14,612 genetic blocks for JLM and 185,212 blocks for GWAS.

SLM, JLM, and GWAS in the RIL Populations

SLM was performed by composite interval mapping (CIM) using the Windows QTL Cartographer software version 2.5 (Wang et al., 2012) in each RIL population. The program settings were as follows: CIM model = model 6: Standard model; control markers numbers = 5; window size = 10 centimorgan; regression method = Backward Regression Method; walk speed = 0.5 centimorgan. We used LOD = 2.5 as the threshold, and the 2-LOD interval was considered as the QTL candidate region.

We combined 10 RIL populations to perform JLM and GWAS (Xiao et al., 2016). For JLM, a linear mixed model was used to detect significant recombination blocks. The model is as follows: $y = X\beta + Z\gamma + \xi + \varepsilon$, where $X\beta$ represents fixed effects, Z is an N × P matrix for the genotype (N is the total number of SNPs and P is the number of lines used to construct RIL populations), γ is a vector of genetic effects for markers, ξ is a vector of polygenic effects, and ε is a vector of the residual errors. The restricted maximum likelihood was used to estimate the parameters, and a permutation test of 500 permutated samples was used to determine the threshold of likelihood ratio test scores. At the type I error rate of 0.05, the threshold of likelihood ratio test scores was 2.76.

For GWAS, we used a stepwise regression method (Tian et al., 2011) with minor modification. To control the polygenic background effect, the GWAS was performed one chromosome at a time. For each chromosome, we forced population effects and the effects of QTLs detected by SLM and JLM from other chromosomes to be included in a general linear model. The residual of this model was then used as the dependent variable to test all SNPs on the current chromosome. We used both forward and backward regressions to select variables, and the cutoff P value for SNPs entering or leaving the model was determined by 500 permutations. The SNPs in the final model were regarded as significant SNPs, and the P value was calculated from the marginal F values of the SNPs. To reduce SNP redundancy, we performed a final backward

regression for the significant SNPs. For SNPs falling within the QTL regions, a backward regression was conducted one QTL at a time, where population effects and all other QTLs were fitted to the model. For the remaining SNPs falling outside the QTL regions, a backward regression was conducted by forcing population effects and all QTLs in the model. The median cutoff P value of 10 chromosomes was used as the threshold of the marker resulting in the final backward model.

Overlapping Analysis

To analyze the overlap between QTLs identified by SLM in each of the 10 RIL populations, we compared the confidence intervals of the mapped QTLs. When two QTLs overlapped, they were considered to represent a single unique QTL. For the analysis of overlapping between genes and QTLs, we compared the positions of the genes and the confidence interval of QTLs identified with three methods. If the candidate region of a QTL identified by SLM was larger than 5 Mb, we limited the candidate region to 2.5 Mb on each side of the peak position. Genes that fell into the candidate regions of QTLs identified by SLM and JLM or the 1-Mb flanking region of the significant SNPs identified by GWAS were considered to be located in the mapped QTLs. A significant level was obtained by comparing the number of genes falling within QTLs in our observation with the numbers resulting from 10,000 permutations. For each permutation, we randomly selected 30 genes and counted the number of genes falling within QTL regions. To evaluate the overlaps between QTLs identified by the three different models, a candidate region was used for each QTL identified by SLM and JLM, while the 1-Mb flanking region of the significant SNPs was used for GWAS. When there was an overlap between two QTLs, we considered them as one unique QTL.

Candidate Gene Association Analysis in an Association Panel

We identified 30 maize (*Zea mays*) orthologs of 18 cloned rice (*Oryza sativa*) genes based on the MSU Rice Genome Annotation Project (Kawahara et al., 2013). This rice annotation project used 232,821 representative peptide sequences from rice (release 7), Arabidopsis (*Arabidopsis thaliana*; release 10), poplar (*Populus trichocarpa*; release 2.2), grapevine (*Vitis vinifera*; release 1_12x), sorghum (*Sorghum bicolor*; release 1.4), maize (release 5b filtered set), and *Brachypodium distachyon* (release 1.0) to identify orthologous groups with OrthoMCL software (Li et al., 2003).

Since 1.25 million SNPs had been mapped previously in AM540 (Liu et al., 2017), we used this genotype data set to identify the SNPs in these 30 genes and then performed candidate gene association analysis. We used the mixed linear model (Yu et al., 2006), which took population structure and kinship into consideration, to test the significance between the SNPs within candidate genes and kernel traits with TASSEL software (Bradbury et al., 2007). The threshold was determined by Bonferroni correction ($P < 0.05/N$, where N is the gene number) for each trait.

Phylogenetic Tree Construction

Amino acid sequences of cell wall invertase protein sequences in rice, maize, and Arabidopsis were aligned using the ClustalW program (Thompson et al., 1994). A phylogenetic tree was constructed with MEGA6 (Tamura et al., 2013). The statistical method was neighbor joining, and 1,000 bootstrap replications were used to test the phylogeny. The substitution model was p-distance, and the partial deletion option was selected to treat gaps/missing data. The genes used for the phylogenetic tree construction were as follows: GRMZM2G095725, *ZmINCW1* (GRMZM2G139300), *ZmINCW2* (GRMZM2G119689), *ZmINCW3* (GRMZM2G123633), and *ZmINCW4* (GRMZM2G119941) from maize; *OsCIN1* (LOC_Os02g33110), *OsCIN2* (*GIF1*; LOC_Os04g33740), *OsCIN3* (LOC_Os04g33720), *OsCIN4* (LOC_Os01g73580), *OsCIN5* (LOC_Os04g56930), *OsCIN6* (LOC_Os04g56920), *OsCIN7* (LOC_Os09g08072), and *OsCIN8* (LOC_Os09g08120) from rice; and *AtcwINV1* (At3G13790), *AtcwINV2* (At3G52600), *AtcwINV3* (At1G55120), *AtcwINV4* (At2G36190), *AtcwINV5* (At3G13784), and *AtcwINV6* (At5G11920) from Arabidopsis.

GWAS of the Expression Level of *ZmINCW1*

Expression QTLs for *ZmINCW1* were identified through GWAS in an association mapping population consisting of 368 maize inbred lines that were subsets of the 540 diverse lines used for candidate gene association. Gene expression was quantified in these 368 lines by RNA sequencing in a previous study (Fu et al., 2013). Since we had genotyped 368 lines with extra methods (Affymetrix Axiom Maize Genotyping 600K Array and genotyping by sequencing; Liu et al., 2017), we performed expression GWAS again with TASSEL software using a mixed linear (Q+K) model (Yu et al., 2006). The threshold was determined by Bonferroni correction, and it was $P < 7.97 \times 10^{-7}$ ($P < 1/n$, where n is the total number of SNPs).

Transgenic Analysis in Arabidopsis

The CTAB method was used to extract the Arabidopsis DNA, and primers AtcwINV2_T-DNA and p745 were used to confirm the T-DNA insertion. Primer ZmINCW1_CDS was used to amplify the maize open reading frame of *ZmINCW1*. The coding region of *ZmINCW1* fused with the hemagglutinin (HA) tag was then cloned behind the cauliflower mosaic virus 35S promoter into pCAMBIA99-1-3 vector and transformed into Arabidopsis by *Agrobacterium tumefaciens*-mediated transformation. Primer ZmINCW1_CDS was used to screen positive transgenic lines. Total RNA was extracted from fresh leaves using an RNA extraction kit (BioTeke), and cDNA was synthesized from the extracted RNA using the TransScript One-Step gDNA Removal and cDNA Synthesis SuperMix kit (TransGen). Quantitative PCR was performed for the gene expression using ZmINCW1_RT and AtACTIN primers. All primers used in this study are listed in Supplemental Table S12.

For protein expression analysis, proteins were extracted from aerial parts of three individual plants (six leaves) and separated by SDS-PAGE. Proteins were transferred to nitrocellulose membranes for western blot. HA detection was performed using a 1:5,000 dilution of anti-HA mouse monoclonal antibody, followed by hybridization with a 1:10,000 dilution of goat anti-mouse horseradish peroxidase secondary antibody. The horseradish peroxidase signal was detected by the ECL substrate kit.

When we measured the Arabidopsis seed weight, we first took digital photographs of the seeds and then measured the seed weight. With each photograph, we used ImageJ software to count the number of seeds. For each individual, we measured the weight of at least 200 seeds three to five times with replacement and then converted it to thousand seed weight.

Supplemental Data

The following supplemental materials are available.

Supplemental Figure S1. Box plots of kernel size and kernel weight in 10 RIL populations.

Supplemental Figure S2. QTLs detected in 10 RIL populations with separate linkage mapping.

Supplemental Figure S3. Overview of identified QTLs and significant SNPs for KTW, KL, KW, and KT.

Supplemental Figure S4. Significant associations between kernel traits and five maize genes that are orthologs of cloned rice genes for seed size or weight.

Supplemental Figure S5. Significant associations between seven maize mutant genes, which were reported to be involved in maize kernel development and kernel traits.

Supplemental Figure S6. Phylogenetic tree of cell wall invertase proteins in maize, rice, and Arabidopsis.

Supplemental Figure S7. *ZmINCW1* was located in candidate regions of QTLs identified in DAN340 × K22 for KL and in DE3 × BY815 for HKW.

Supplemental Table S1. Mean values, SD, and heritability of five kernel traits in 10 RIL populations.

Supplemental Table S2. Correlation coefficients between HKW and three other kernel size traits in 10 RIL populations.

Supplemental Table S3. QTL numbers for kernel size and weight in 10 RIL populations.

Supplemental Table S4. List of phenotypic variations explained by pleiotropic QTLs that could be detected in more than one population for the same traits.

Supplemental Table S5. List of phenotypic variations explained by pleiotropic QTLs that could be detected in more than one population for different traits.

Supplemental Table S6. Full list of identified QTLs with SLM for kernel size and weight in 10 RIL populations.

Supplemental Table S7. Full list of identified QTLs with JLM for kernel size and weight through combining 10 RIL populations.

Supplemental Table S8. Full list of identified significant SNPs with GWAS for kernel size and weight through combining 10 RIL populations.

Supplemental Table S9. Full list of candidate SNPs identified with GWAS for kernel size and weight through combining 10 RIL populations.

Supplemental Table S10. Significant associations between reported maize genes for seed development and maize kernel size and weight in this study.

Supplemental Table S11. Significant associations between maize orthologs of cloned rice genes and maize kernel size and weight.

Supplemental Table S12. Primers used in this study.

ACKNOWLEDGMENTS

We thank Xiongbing Yan for help in field work and Dr. Zhibing Lai and Dr. Hongze Wang for helping with the Arabidopsis transgenic experiments.

Received May 31, 2017; accepted August 12, 2017; published August 15, 2017.

LITERATURE CITED

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

Chen J, Zhang L, Liu S, Li Z, Huang R, Li Y, Cheng H, Li X, Zhou B, Wu S, et al (2016) The genetic basis of natural variation in kernel size and related traits using a four-way cross population in maize. PLoS ONE **11:** e0153428

Chen X, Feng F, Qi W, Xu L, Yao D, Wang Q, Song R (2017) Dek35 encodes a PPR protein that affects cis-splicing of mitochondrial nad4 intron 1 and seed development in maize. Mol Plant **10:** 427–441

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

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 Genet **112:** 1164–1171

Fu J, Cheng Y, Linghu J, Yang X, Kang L, Zhang Z, Zhang J, He C, Du X, Peng Z, et al (2013) RNA sequencing reveals the complex regulatory network in the maize kernel. Nat Commun **4:** 2832

Fu S, Meeley R, Scanlon MJ (2002) Empty pericarp2 encodes a negative regulator of the heat shock response and is required for maize embryogenesis. Plant Cell **14:** 3119–3132

Gore MA, Chia JM, Elshire RJ, Sun Q, Ersoz ES, Hurwitz BL, Peiffer JA, McMullen MD, Grills GS, Ross-Ibarra J, et al (2009) A first-generation haplotype map of maize. Science **326:** 1115–1117

Gutiérrez-Marcos JF, Dal Prà M, Giulini A, Costa LM, Gavazzi G, Cordelier S, Sellam O, Tatout C, Paul W, Perez P, et al (2007) empty pericarp4 encodes a mitochondrion-targeted pentatricopeptide repeat protein necessary for seed development and plant growth in maize. Plant Cell **19:** 196–210

Hirsch CN, Flint-Garcia SA, Beissinger TM, Eichten SR, Deshpande S, Barry K, McMullen MD, Holland JB, Buckler ES, Springer N, et al (2014a) Insights into the effects of long-term artificial selection on seed size in maize. Genetics **198:** 409–421

Hirsch CN, Foerster JM, Johnson JM, Sekhon RS, Muttoni G, Vaillancourt B, Peñagaricano F, Lindquist E, Pedraza MA, Barry K, et al (2014b) Insights into the maize pan-genome and pan-transcriptome. Plant Cell **26:** 121–135

Hong Y, Chen L, Du LP, Su Z, Wang J, Ye X, Qi L, Zhang Z (2014) Transcript suppression of TaGW2 increased grain width and weight in bread wheat. Funct Integr Genomics **14:** 341–349

Hung HY, Shannon LM, Tian F, Bradbury PJ, Chen C, Flint-Garcia SA, McMullen MD, Ware D, Buckler ES, Doebley JF, et al (2012) ZmCCT and the genetic basis of day-length adaptation underlying the postdomestication spread of maize. Proc Natl Acad Sci USA **109:** E1913–E1921

Jaiswal V, Gahlaut V, Mathur S, Agarwal P, Khandelwal MK, Khurana JP, Tyagi AK, Balyan HS, Gupta PK (2015) Identification of novel SNP in promoter sequence of TaGW2-6A associated with grain weight and other agronomic traits in wheat (Triticum aestivum L.). PLoS ONE **10:** e0129400

Jiang L, Ge M, Zhao H, Zhang T (2015) Analysis of heterosis and quantitative trait loci for kernel shape related traits using triple testcross population in maize. PLoS ONE **10:** e0124779

Jin M, Liu H, He C, Fu J, Xiao Y, Wang Y, Xie W, Wang G, Yan J (2016) Maize pan-transcriptome provides novel insights into genome complexity and quantitative trait variation. Sci Rep **6:** 18936

Kawahara Y, de la Bastide M, Hamilton JP, Kanamori H, McCombie WR, Ouyang S, Schwartz DC, Tanaka T, Wu J, Zhou S, et al (2013) Improvement of the Oryza sativa Nipponbare reference genome using next generation sequence and optical map data. Rice (N Y) **6:** 4

Lai J, Li R, Xu X, Jin W, Xu M, Zhao H, Xiang Z, Song W, Ying K, Zhang M, et al (2010) Genome-wide patterns of genetic variation among elite maize inbred lines. Nat Genet **42:** 1027–1030

Li B, Liu H, Zhang Y, Kang T, Zhang L, Tong J, Xiao L, Zhang H (2013) Constitutive expression of cell wall invertase genes increases grain yield and starch content in maize. Plant Biotechnol J **11:** 1080–1091

Li C, Shen Y, Meeley R, McCarty DR, Tan BC (2015) Embryo defective 14 encodes a plastid-targeted cGTPase essential for embryogenesis in maize. Plant J **84:** 785–799

Li J, Fu J, Chen Y, Fan K, He C, Zhang Z, Li L, Liu Y, Zheng J, Ren D, et al (2017) The U6 biogenesis-like 1 plays an important role in maize kernel and seedling development by affecting the 3′ end processing of U6 snRNA. Mol Plant **10:** 470–482

Li L, Stoeckert CJ Jr, Roos DS (2003) OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res **13:** 2178–2189

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

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

Li XJ, Zhang YF, Hou M, Sun F, Shen Y, Xiu ZH, Wang X, Chen ZL, Sun SS, Small I, et al (2014) Small kernel 1 encodes a pentatricopeptide repeat protein required for mitochondrial nad7 transcript editing and seed development in maize (Zea mays) and rice (Oryza sativa). Plant J **79:** 797–809

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

Lid SE, Gruis D, Jung R, Lorentzen JA, Ananiev E, Chamberlin M, Niu X, Meeley R, Nichols S, Olsen OA (2002) The defective kernel 1 (dek1) gene required for aleurone cell development in the endosperm of maize grains encodes a membrane protein of the calpain gene superfamily. Proc Natl Acad Sci USA **99:** 5460–5465

Liu H, Luo X, Niu L, Xiao Y, Chen L, Liu J, Wang X, Jin M, Li W, Zhang Q, et al (2017) Distant eQTLs and non-coding sequences play critical roles in regulating gene expression and quantitative trait variation in maize. Mol Plant **10:** 414–426

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

Liu Y, Wang L, Sun C, Zhang Z, Zheng Y, Qiu F (2014) Genetic analysis and major QTL detection for maize kernel size and weight in multi-environments. Theor Appl Genet **127:** 1019–1037

Liu YJ, Xiu ZH, Meeley R, Tan BC (2013) Empty pericarp5 encodes a pentatricopeptide repeat protein that is required for mitochondrial RNA editing and seed development in maize. Plant Cell **25:** 868–883

Lowe J, Nelson OE (1946) Miniature seed: a study in the development of a defective caryopsis in maize. Genetics **31:** 525–533

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

Manavski N, Guyon V, Meurer J, Wienand U, Brettschneider R (2012) An essential pentatricopeptide repeat protein facilitates 5′ maturation and

translation initiation of rps3 mRNA in maize mitochondria. Plant Cell **24:** 3087–3105

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

Miller ME, Chourey PS (1992) The maize invertase-deficient miniature-1 seed mutation is associated with aberrant pedicel and endosperm development. Plant Cell **4:** 297–305

Murat F, Armero A, Pont C, Klopp C, Salse J (2017) Reconstructing the genome of the most recent common ancestor of flowering plants. Nat Genet **49:** 490–496

Neuffer MG, Sheridan WF (1980) Defective kernel mutants of maize. I. Genetic and lethality studies. Genetics **95:** 929–944

Pan Q, Li L, Yang X, Tong H, Xu S, Li Z, Li W, Muehlbauer GJ, Li J, Yan J (2016) Genome-wide recombination dynamics are associated with phenotypic variation in maize. New Phytol **210:** 1083–1094

Paterson AH, Lin YR, Li Z, Schertz KF, Doebley JF, Pinson SR, Liu SC, Stansel JW, Irvine JE (1995) Convergent domestication of cereal crops by independent mutations at corresponding genetic loci. Science **269:** 1714–1718

Qin L, Hao C, Hou J, Wang Y, Li T, Wang L, Ma Z, Zhang X (2014) Homologous haplotypes, expression, genetic effects and geographic distribution of the wheat yield gene TaGW2. BMC Plant Biol **14:** 107

Raihan MS, Liu J, Huang J, Guo H, Pan Q, Yan J (2016) Multi-environment QTL analysis of grain morphology traits and fine mapping of a kernel-width QTL in Zheng58 × SK maize population. Theor Appl Genet **129:** 1465–1477

Shanker S, Salazar RW, Taliercio EW, Chourey PS (1995) Cloning and characterization of full-length cDNA encoding cell-wall invertase from maize. Plant Physiol **108:** 873–874

Simmonds J, Scott P, Brinton J, Mestre TC, Bush M, Del Blanco A, Dubcovsky J, 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:** 1099–1112

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

Su Z, Hao C, Wang L, Dong Y, Zhang X (2011) Identification and development of a functional marker of TaGW2 associated with grain weight in bread wheat (Triticum aestivum L.). Theor Appl Genet **122:** 211–223

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

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol **30:** 2725–2729

Tang GQ, Lüscher M, Sturm A (1999) Antisense repression of vacuolar and cell wall invertase in transgenic carrot alters early plant development and sucrose partitioning. Plant Cell **11:** 177–189

Thompson JD, Higgins DG, Gibson TJ (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:** 4673–4680

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

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

Wang S, Basten CJ, Zeng ZB (2012) Windows QTL Cartographer 2.5. Department of Statistics, North Carolina State University, Raleigh, NC

Wang S, Yan X, Wang Y, Liu H, Cui D, Chen F (2016) Haplotypes of the TaGS5-A1 gene are associated with thousand-kernel weight in Chinese bread wheat. Front Plant Sci **7:** 783

Wang S, Zhang X, Chen F, Cui D (2015) A single-nucleotide polymorphism of TaGS5 gene revealed its association with kernel weight in Chinese bread wheat. Front Plant Sci **6:** 1166

Xiao Y, Tong H, Yang X, Xu S, Pan Q, Qiao F, Raihan MS, Luo Y, Liu H, Zhang X, et al (2016) Genome-wide dissection of the maize ear genetic architecture using multiple populations. New Phytol **210:** 1095–1106

Xiu Z, Sun F, Shen Y, Zhang X, Jiang R, Bonnard G, Zhang J, Tan BC (2016) EMPTY PERICARP16 is required for mitochondrial nad2 intron 4 cis-splicing, complex I assembly and seed development in maize. Plant J **85:** 507–519

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

Yang Q, Li Z, Li W, Ku L, Wang C, Ye J, Li K, Yang N, Li Y, Zhong T, et al (2013) CACTA-like transposable element in ZmCCT attenuated photoperiod sensitivity and accelerated the postdomestication spread of maize. Proc Natl Acad Sci USA **110:** 16969–16974

Yu J, Pressoir G, Briggs WH, Vroh Bi I, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, et al (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nat Genet **38:** 203–208

Zeng ZB (1994) Precision mapping of quantitative trait loci. Genetics **136:** 1457–1468

Zhang Z, Liu Z, Hu Y, Li W, Fu Z, Ding D, Li H, Qiao M, Tang J (2014) QTL analysis of kernel-related traits in maize using an immortalized F_2 population. PLoS ONE **9:** e89645