

# Gene duplication confers enhanced expression of 27-kDa $\gamma$ -zein for endosperm modification in quality protein maize

Hongjun Liu<sup>a,1</sup>, Junpeng Shi<sup>b,1</sup>, Chuanlong Sun<sup>a</sup>, Hao Gong<sup>c</sup>, Xingming Fan<sup>d</sup>, Fazhan Qiu<sup>e</sup>, Xuehui Huang<sup>c</sup>, Qi Feng<sup>c</sup>, Xixi Zheng<sup>a</sup>, Ningning Yuan<sup>a</sup>, Changsheng Li<sup>a</sup>, Zhiyong Zhang<sup>a</sup>, Yiting Deng<sup>a</sup>, Jiechen Wang<sup>a</sup>, Guangtang Pan<sup>f</sup>, Bin Han<sup>c</sup>, Jinsheng Lai<sup>b</sup>, and Yongrui Wu<sup>a,2</sup>

<sup>a</sup>National Key Laboratory of Plant Molecular Genetics, Chinese Academy of Sciences Center for Excellence in Molecular Plant Sciences, Institute of Plant Physiology & Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200032, China; <sup>b</sup>State Key Laboratory of Agrobiotechnology and National Maize Improvement Center, Department of Plant Genetics and Breeding, China Agricultural University, Beijing 100193, China; <sup>c</sup>National Center for Gene Research, Institute of Plant Physiology & Ecology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200233, China; <sup>d</sup>Institute of Food Crops, Yunnan Academy of Agricultural Sciences, Kunming 650205, China; <sup>e</sup>National Key Laboratory of Crop Genetic Improvement, Huazhong Agricultural University, Wuhan 430070, China; and <sup>f</sup>Key Laboratory of Biology and Genetic Improvement of Maize in Southwest Region, Ministry of Agriculture; Maize Research Institute, Sichuan Agricultural University, Chengdu 611130, China

Edited by Ronald L. Phillips, University of Minnesota, St. Paul, MN, and approved March 25, 2016 (received for review January 25, 2016)

The maize *opaque2* (*o2*) mutant has a high nutritional value but it develops a chalky endosperm that limits its practical use. Genetic selection for *o2* modifiers can convert the normally chalky endosperm of the mutant into a hard, vitreous phenotype, yielding what is known as quality protein maize (QPM). Previous studies have shown that enhanced expression of 27-kDa  $\gamma$ -zein in QPM is essential for endosperm modification. Taking advantage of genome-wide association study analysis of a natural population, linkage mapping analysis of a recombinant inbred line population, and map-based cloning, we identified a quantitative trait locus (*q27*) affecting expression of 27-kDa  $\gamma$ -zein. *q27* was mapped to the same region as the major *o2* modifier (*o2 modifier1*) on chromosome 7 near the 27-kDa  $\gamma$ -zein locus. *q27* resulted from a 15.26-kb duplication at the 27-kDa  $\gamma$ -zein locus, which increases the level of gene expression. This duplication occurred before maize domestication; however, the gene structure of *q27* appears to be unstable and the DNA rearrangement frequently occurs at this locus. Because enhanced expression of 27-kDa  $\gamma$ -zein is critical for endosperm modification in QPM, *q27* is expected to be under artificial selection. This discovery provides a useful molecular marker that can be used to accelerate QPM breeding.

QPM | *o2* modifiers | endosperm | artificial selection | gene duplication

By 2030, the world population is predicted to reach 8.5 billion people. As a consequence, food production will need to be increased by more than 50% (1). However, the rate of food production has not kept pace with explosive population growth. Enhancing the nutritional quality of staple crops is one strategy for addressing the emerging food crisis (2, 3).

Maize (*Zea mays*) is the highest-yielding crop in the world, but it cannot be used as the sole protein source for humans and monogastric livestock because its main storage proteins, zeins, are deficient in the essential amino acids, lysine and tryptophan (4). The poor protein quality of maize can be improved by the *opaque2* (*o2*) mutation, which increases the lysine and tryptophan levels by decreasing the synthesis of zeins and compensatorily increasing other (nonzein) seed proteins. However, unfortunately, the chalky and soft texture of *o2* kernels limits utilization of this mutant (5). The creation of quality protein maize (QPM) was based on modification of *o2* by accumulating quantitative trait loci (QTLs), called *o2* modifiers, that lead to a hard, vitreous endosperm (5, 6). The development of QPM has greatly improved the lives of people who suffer from malnutrition in the developing countries (7).

Although QPM breeding has gone on more than 50 y, neither the mechanism nor the genetic components controlling endosperm modification are well understood. Seven *o2* modifiers have been located on six chromosomes (8), including one, designated *o2 modifier1* in bin 7.02 near the 27-kDa  $\gamma$ -zein locus, which has a

major effect on endosperm modification; the other six loci contribute smaller effects (8). Coincidentally, the RNA transcript and protein levels of 27-kDa  $\gamma$ -zein accumulate two- to threefold higher in QPM than in an isogenic wild-type and *o2* mutant (9, 10). The extent of *o2* modification is positively correlated to the level of 27-kDa  $\gamma$ -zein in an *F2* population and recombinant inbred lines created by a cross between QPM and an *o2* mutant (10–12). Silencing of  $\gamma$ -zeins or a deletion that eliminates 27-kDa  $\gamma$ -zein expression abolished the formation of vitreous endosperm in QPM (13, 14).

In maize endosperm, zein proteins are stored in rough endoplasmic reticulum-retained protein bodies (PBs) (15–17). The 27-kDa  $\gamma$ -zein, along with the paralogous 16-kDa  $\gamma$ -zein and 15-kDa  $\beta$ -zein, is thought to play an important role in initiation and stabilization of PB formation (18). Knockdown or knockout of  $\gamma$ -zeins in QPM results in irregular, clumped PBs in reduced numbers and an opaque phenotype, indicating that  $\gamma$ -zeins are necessary, if not sufficient, for endosperm modification (13, 14). In this study, we identified a major QTL (*q27*) that controls enhanced expression of

## Significance

More than half a century ago, Oliver Nelson and Edwin Mertz at Purdue University found the maize *opaque2* (*o2*) mutation produces doubling of the endosperm lysine content, creating the foundation for quality protein maize (QPM) breeding. QPM has the potential to benefit millions of people in developing countries who consume maize as their sole protein source. However, breeding new QPM hybrids takes longer than regular hybrids, primarily because of the complex and unknown components of *o2* endosperm modification; this has limited expansion of QPM worldwide. We identified a quantitative trait locus, a gene duplication at the 27-kDa  $\gamma$ -zein locus, which confers enhanced expression of this protein and leads to endosperm modification. This knowledge can effectively be applied in QPM breeding.

Author contributions: B.H., J.L., and Y.W. designed research; H.L., J.S., C.S., H.G., Q.F., X.Z., N.Y., C.L., Z.Z., Y.D., J.W., and Y.W. performed research; X.F., F.Q., G.P., B.H., and J.L. contributed new reagents/analytic tools; H.L., J.S., X.H., J.W., and Y.W. analyzed data; and H.L. and Y.W. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

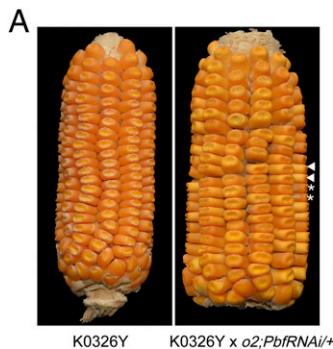
Freely available online through the PNAS open access option.

Data deposition: The sequences reported in this paper have been deposited in the GenBank/EMBL database (accession nos. KU593569 and KU593570).

<sup>1</sup>H.L. and J.S. contributed equally to this work.

<sup>2</sup>To whom correspondence should be addressed. Email: yrwu@sibs.ac.cn.

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1601352113/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1601352113/-DCSupplemental).



**Fig. 1.** Silencing of *Pbf* in QPM. (A) Ear phenotypes of K0326Y and K0326Y × *o2;PbfRNAi/+*. In K0326Y × *o2;PbfRNAi/+*, half of the progeny inherited the *PbfRNAi* transgene and expressed a reduced level of 27-kDa γ-zein and exhibited an opaque phenotype. Two representative vitreous and opaque seeds are indicated by asterisks and arrowheads. (B) Truncated seed phenotypes. The left two seeds not inheriting the *PbfRNAi* are vitreous, whereas the right two positive seeds are opaque. (C) The seeds inheriting the *PbfRNAi* expressed a decreased level of 27-kDa γ-zein. The 27-kDa γ-zein is indicated by the arrow.

the 27-kDa γ-zein gene in QPM. Its DNA structure, derivative alleles from DNA rearrangement, and potential application in QPM breeding are discussed.

## Results

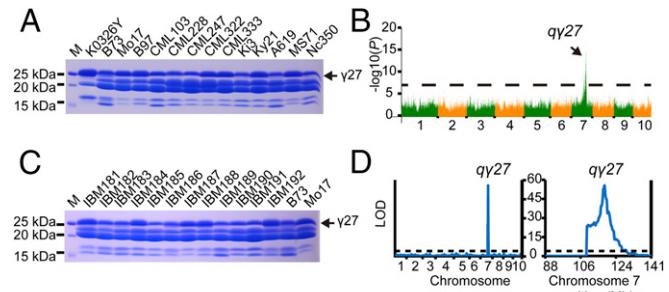
**Genetic Regulation of 27-kDa γ-Zein Expression in QPM.** An *o2* mutant with a starchy endosperm phenotype can be modified to create a vitreous QPM by *o2* modifiers, which act in a semidominant mode (8). Previous research showed there is a major QTL associated with the 27-kDa γ-zein locus that affects *o2* modification (8–12). Additionally, evidence exists that an enhanced level of this protein is essential for creating vitreous endosperm in QPM (13, 14). These observations suggest that *o2* modifier-mediated endosperm modification in QPM is associated with enhanced expression of 27-kDa γ-zein and the QTL (designated here as *qγ27*) directly regulates the elevation of 27-kDa γ-zein, which acts as the *o2* modifier. To investigate the genetic action of *qγ27* on 27-kDa γ-zein expression, K0326Y QPM (with *qγ27*) and B73 (without *qγ27*) were reciprocally crossed. Because the endosperm is triploid (two from the female and one from the male), the four genetic materials—(i) K0326Y QPM, (ii) K0326Y QPM × B73, (iii) B73 × K0326Y QPM, and (iv) B73—have three, two, one, and zero dosage of *qγ27*, respectively. Phenotypic analysis of the resulting progeny revealed the protein level of 27-kDa γ-zein is positively correlated with the genetic dosage of *qγ27* in QPM (Fig. S1).

Unlike the 22-kDa α-zein genes, the expression of the 27-kDa γ-zein gene is not significantly affected in *o2* (19). The 27-kDa γ-zein gene is transcriptionally regulated by the proline-rich binding factor (PBF) and O2 heterodimerizing proteins (OHP1 and OHP2), the latter being homologs of O2 at other loci in the maize genome (19, 20). We previously created *PbfRNAi* transgenic lines with suppressed expression of *Pbf*, which lead to a decreased level of 27-kDa γ-zein (20). Because QPM is homozygous for *o2*, *PbfRNAi* was introgressed into W64Ao2 by several rounds of back-crossing, yielding the genotype *PbfRNAi/+;o2* (homozygous for *o2* and heterozygous for *PbfRNAi*) (19). To investigate the genetic relationship between *qγ27* and the transcriptional regulation of 27-kDa γ-zein, K0326Y QPM and *PbfRNAi/+;o2* were crossed. Half of the progeny (120 of 250) inherited the *PbfRNAi* transgene. They expressed a reduced level of 27-kDa γ-zein and exhibited an opaque phenotype, indicating that *PbfRNAi* is dominant over *qγ27* on 27-kDa γ-zein expression. In other words, *qγ27* acts downstream of the transcriptional regulation of 27-kDa γ-zein by PBF (Fig. 1). We also examined the *cis*- and *trans*-action of *qγ27* on elevated expression of 27-kDa γ-zein. To do this, the two 27-kDa γ-zein alleles in the cross of a QPM and a wild-type without *qγ27* must be

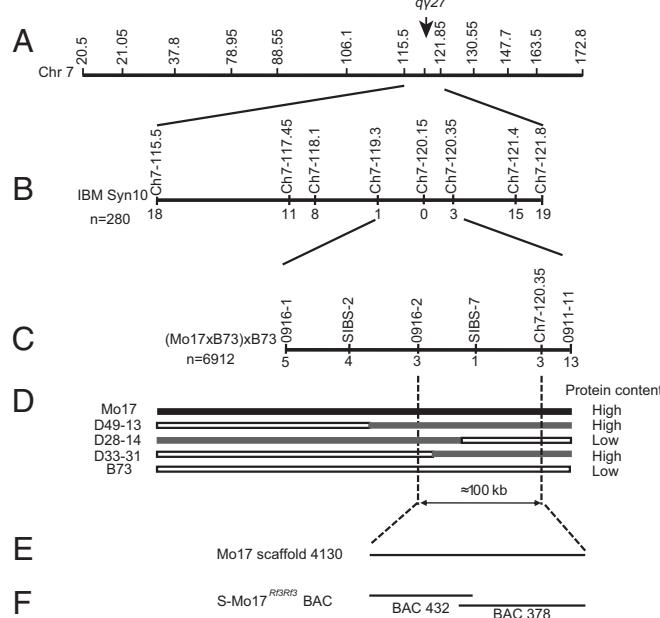
distinguishable by a DNA polymorphism. XF134 is an inbred line that has a 27-kDa γ-zein gene 18-bp shorter than the B73-*γ27* and K0326Y-*γ27* alleles (Fig. S2A and B). If *qγ27* is a *trans*-factor that equally affects the expression of K0326Y-*γ27* and XF134-*γ27* alleles, one would expect that the cDNA ratio of the two alleles should be 2:1 in the cross of K0326Y QPM × XF134 as the same in the cross of B73 × XF134 because of the triploid endosperm. However, the observed ratio of K0326Y-*γ27* and XF134-*γ27* alleles (4.25:1) was significantly higher than the expected value ( $\chi^2 = 48.0$ ) (Fig. S2C), suggesting *qγ27* acts in a *cis* manner, resulting in the 27-kDa γ-zein allele from K0326Y QPM being preferentially expressed compared with its counterpart allele from XF134.

**Cloning of *qγ27*.** QPM accumulates a higher level of 27-kDa γ-zein protein than some wild-type inbred lines, like B73 and CM105, although the level of 27-kDa γ-zein was found to vary in different inbred lines (21) (Fig. 2A), suggesting genetic variation for this trait. This finding also indicates that *qγ27*, controlling the enhanced expression of 27-kDa γ-zein in QPM, might be selected from QTLs in the natural population. To evaluate the number of QTLs, we collected 492 inbred lines and their levels of 27-kDa γ-zein were phenotyped by estimating the abundance of this protein based on SDS/PAGE (Fig. 2A and Dataset S1). A genome-wide association study (GWAS) analysis identified a major peak ( $P = 8.568517806e^{-17}$ ) of 27-kDa γ-zein accumulation associated with the physical coordinate 120 Mb near the 27-kDa γ-zein locus on chromosome 7 (Figs. 2B and 3A), which falls into the same region as *o2 modifier1*, the major *o2* modifier for endosperm modification (8, 22), suggesting there is only one major QTL in the natural population and it has been selected for enhanced expression of 27-kDa γ-zein in QPM.

Because Mo17 expresses an overtly higher level of 27-kDa γ-zein than B73 (Fig. 2A), we were able to take advantage of the intermated B73 × Mo17 synthetic 10 double-haploid population (IBM Synthetic 10 DH) to further map the QTL (23). This analysis confirmed the location of the QTL in the same region as *o2 modifier1* (Figs. 2C and D and 3B). Furthermore, the ratio of the Mo17-*γ27* allele expressed in the endosperms of Mo17 × XF134 and XF134 × Mo17 was significantly higher than expected (Fig. S2C), similar to the cross of K0326Y QPM and XF134, indicating that the two QTLs are identical. This finding



**Fig. 2.** GWAS and linkage mapping of *qγ27*. (A) SDS/PAGE analysis of zein proteins from different inbred lines. K0326Y QPM, Mo17, B97, CML322, CML333, Ky21, and MS71 expressed higher levels of 27-kDa γ-zein than B73, CML103, CML228, CML247, Ki3, A619, and Nc350. The 27-kDa γ-zein is indicated by the arrow. Total zein from 200 μg of corn flour was loaded in each lane. M, protein markers from top to bottom correspond to 25, 20, and 15 kDa. The zein protein analysis was the same in all of the following investigations. (B) Manhattan plot of the 27-kDa γ-zein levels. The numbers on the horizontal axis indicate the maize chromosomes. (C) SDS/PAGE analysis of zein proteins in different IBM lines. The 27-kDa γ-zein is expressed at a higher level in lines 181, 184, 187, 188, and 192 than in 182, 183, 185, 186, 189, 190, and 191, which are similar to levels of their parent lines Mo17 and B73, respectively. (D) Linkage analysis of the 27-kDa γ-zein expression. (Left) The logarithm of the odds (LOD) score profile with the permutation threshold indicated by the horizontal line; right, the LOD score profile for the enlarged chromosome 7 region.



**Fig. 3.** Map-based cloning of *qγ27*. (A) Location of *qγ27* on the physical map by GWAS. (B) Location of *qγ27* on the physical map by the linkage analysis of 280 IBM individuals. (C) Fine-mapping of *qγ27* using the population of (*Mo17* × *B73*) × *B73* with 6912 seeds. The numbers under each bar indicate the number of recombinants between *qγ27* and the molecular marker. (D) Genotypes and phenotypes of the recombinants. (E) *Mo17* scaffold 4130. (F) BACs 432 and 378 are overlapped and yield a continuous 258-kb sequence.

enabled us to use *Mo17*, the genome of which is being sequenced, to fine-map *qγ27*.

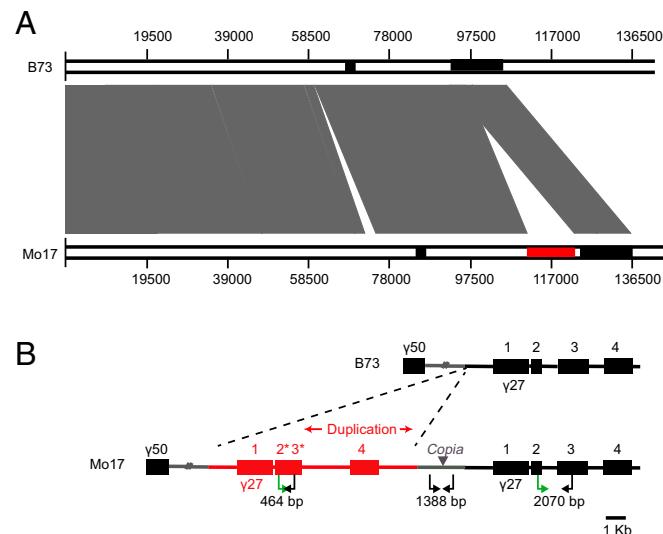
The effect of *qγ27* on 27-kDa γ-zein expression is dosage-dependent (Fig. S1). To clearly score the protein phenotype, we created an *F<sub>1</sub>BC<sub>1</sub>* population of (*Mo17* × *B73*) × *B73* to fine-map this QTL. Each individual seed in this population was cut in half: one for zein protein extraction and the other for genomic DNA isolation. The zein protein was analyzed by SDS/PAGE for phenotyping the abundance of 27-kDa γ-zein (Fig. S3). The segregation ratio of high and low expression of 27-kDa γ-zein was 1:1, confirming that a single QTL controls this trait (Fig. S3). Using 6,912 individuals, the QTL was narrowed down to a region between the markers 0916-2 and bin 120.35 (Fig. 3C). Based on the *B73* genome, these two markers span ~100 kb, where only five genes were annotated: namely, the 50-kDa γ-zein gene, the 27-kDa γ-zein gene, GRMZM2G565441, GRMZM2G138976, and GRMZM5G873335. GRMZM2G565441 is an unknown gene and appears not to be expressed in all examined tissues based on RNA-seq analysis (24). The latter two genes encode an ARID (AT-rich interaction domain) transcription factor and a protein with unknown function, respectively.

Because *B73* lacks *qγ27*, we were unlikely to clone the gene from its genome. However, the *Mo17* genome is being sequenced and assembled at China Agricultural University, yielding a scaffold (designated 4130) that was identified to cover this region. Although there are a few gaps in this scaffold, duplication of a fragment containing two copies of the 27-kDa γ-zein gene, GRMZM2G565441, GRMZM2G138976, and GRMZM5G873335 was revealed (Fig. 4). This duplication, which was confirmed by Southern blotting (Fig. S4A), could potentially explain the high-level expression of the *Mo17*-*γ27* allele in the crosses of *Mo17* and XF134 (Fig. S2C). The duplication is spaced by a 1,388-bp filler DNA in which a *Copia* retrotransposon is inserted (Fig. 4B). Sequence alignment of the two 27-kDa γ-zein gene copies, including the promoter regions (~1 kb upstream of the start codon), CDSs, and 3'UTRs showed they are identical (Fig. S5). Real-time PCR revealed that not only the 27-kDa γ-zein gene itself, but also GRMZM2G138976 and

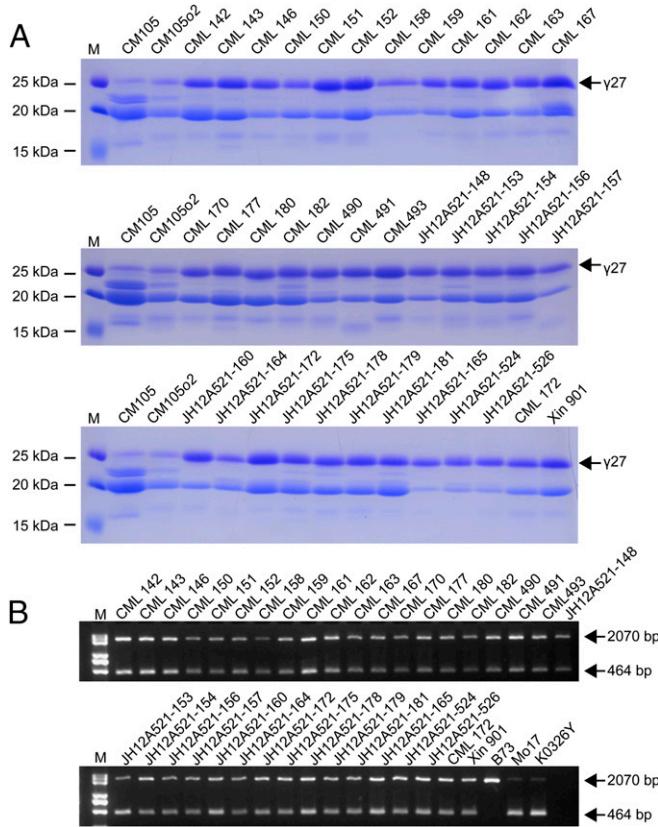
GRMZM5G873335, are expressed nearly twofold higher in *Mo17* than *B73* (Fig. S4B), consistent with the effect of the copy-number difference (Fig. 4 and Fig. S4A). We also screened a *Mo17* BAC library, which was constructed with a special *Mo17* line containing the male sterile restoring gene (*S-Mo17<sup>Rf3Rf3</sup>*). Sequence determination revealed the same duplication in the continuous 258-kb sequence, although many variations were discovered between *S-Mo17<sup>Rf3Rf3</sup>* and the standard *Mo17* genome that was used for mapping and genomic sequencing (Fig. 3F).

**Elevated Expression of 27-kDa γ-Zein in QPM Is Associated with a Gene Duplication.** To investigate if K0326Y QPM also has the 27-kDa γ-zein gene duplication, genomic DNA was blotted with a 27-kDa γ-zein probe, which revealed two copies, as observed in *Mo17* (Fig. S4A). CM105, CM105<sub>a</sub>2, and CM105<sub>Mo2</sub>, a QPM line, are nearly isogenic lines. The duplication giving rise to two copies of a 27-kDa γ-zein gene would be expected to be associated with elevated gene expression in QPM lines, so CM105<sub>Mo2</sub> would be predicted to have the duplication, but not the other two nearly isogenic lines. To test this, we designed a primer pair (0707-1) (Table S1 and Dataset S2) that flanks a 1,606-bp deletion in the 3' region of the duplicated GRMZM2G138976 gene that would give rise to one or two DNA bands, depending on the absence or presence of the duplication (Fig. 4B). Two bands characteristic of the duplication were found in CM105<sub>Mo2</sub>, but not CM105<sub>a</sub>2 or CM105 wild-type (Fig. S6). We then checked 36 other QPMs and found they all contain the duplicated locus (Fig. 5), confirming that *qγ27* is the 27-kDa γ-zein gene duplication. Among the 492 inbred lines used in the GWAS, the 97 lines accumulating the highest level of 27-kDa γ-zein were all found to contain the 27-kDa γ-zein duplication (Dataset S2).

**A DNA Rearrangement Is Associated with Reduced 27-kDa γ-Zein Gene Expression.** Some inbred lines, like W22 and A188, were previously found to carry the so-called standard allele (*S* allele) of 27-kDa γ-zein gene, which was defined as consisting of two (A and



**Fig. 4.** Sequence alignment of the two haplotypes of *B73* and *Mo17*. (A) The contiguous sequence of the *Mo17*-*γ27* locus determined in this study was aligned with *B73* orthologous region. The small and long black bars indicate the 50-kDa and 27-kDa γ-zein locus, respectively. The red bar in *Mo17* represents the duplicated fragment of the 27-kDa γ-zein locus. Sequence homology between the two inbreds is connected by vertical gray areas. Scale markers for both contiguous sequences are indicated. (B) Gene duplication of the 27-kDa γ-zein locus in *Mo17*. The solid boxes represent genes. Compared with *B73*, the duplicated region is illustrated in red. Genes 1, 2, 3, and 4 represent the 27-kDa γ-zein gene, GRMZM2G565441, GRMZM2G138976, and GRMZM5G873335, respectively.



**Fig. 5.** Association of the elevated expression of 27-kDa  $\gamma$ -zein and the gene duplication. (A) SDS/PAGE analysis of zein proteins in different QPM lines. All QPMs accumulate significantly higher levels of 27-kDa  $\gamma$ -zein than the wild-type inbred CM105 and its  $o2$  mutant CM105o2. The 27-kDa  $\gamma$ -zein is indicated by the arrow. (B) There is gene duplication at the 27-kDa  $\gamma$ -zein locus in all QPMs. The size of each band is indicated beside. M, DNA markers.

B) 27-kDa  $\gamma$ -zein genes. The A and B genes can be distinguished by several conserved SNPs in the CDS (25, 26) (Fig. S7). Using the 0707-1 primer as performed in Fig. 5 and Fig. S6, genes were found to result from the same gene duplication, indicating that the two copies of the 27-kDa  $\gamma$ -zein gene in some lines diverged after duplication. We found that 29 of 97 inbreds expressing high levels of 27-kDa  $\gamma$ -zein have the A and B genes, designated the *Sab* allele; the remaining 68 carry the *Saa* allele, which consists of two identical A genes (Dataset S2). Of 38 QPM lines, 22 carry the *Sab* allele and 16 have the *Saa* allele (Dataset S2).

The 27-kDa  $\gamma$ -zein gene duplicated locus appears to be unstable. In a special stock of W22 and in A188, a single copy of the 27-kDa  $\gamma$ -zein gene was identified and appears to be generated by a DNA rearrangement (25, 27). The W22-Rb stock in our laboratory contains only the B gene, and it accumulates a much lower level of 27-kDa  $\gamma$ -zein than the wild-type W22, which contains the A and B genes (Fig. 6 and Fig. S7). Thus, this result is also consistent with the hypothesis that the 27-kDa  $\gamma$ -zein duplication is associated with its elevated protein synthesis and accumulation.

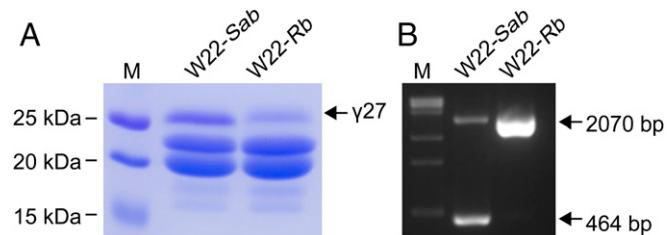
## Discussion

***qy27* Is a Duplication of 27-kDa  $\gamma$ -Zein Gene.** A vitreous kernel with a hard endosperm texture is an important agronomic trait selected by maize breeders because it provides mechanical strength during seed harvest, transportation, and storage. In a wild-type kernel, the outer region of the endosperm is vitreous and the inner area is starchy (28, 29). It appears that vitreous endosperm formation involves the interaction of PBs, starch granules, and cytoplasm (30). During endosperm development, the starch granules are surrounded by PBs in the cytoplasm. However, starch granules in

the periphery of the endosperm are fewer and smaller than those in the center; in contrast, number and size of PBs are the inverse in these two regions (18, 29–31). At seed maturity, the endomembrane system degrades as a result of programmed cell death and cell desiccation. Consequently, starch granules in the outer region of the endosperm become embedded in a proteinaceous matrix composed of soluble proteins and zein PBs, which have cysteine-rich proteins ( $\gamma$ - and  $\beta$ -zeins) on their surface. Condensation of these components leads to the formation of vitreous endosperm, whereas the inner area forms soft, starchy endosperm (29–31). In the *o2* mutant, the number and size of PBs is dramatically reduced, resulting in naked starch granules throughout the endosperm; as a consequence, the phenotype of the kernel becomes starchy and opaque (15). However, this phenotype can be overcome in QPM by *o2* modifiers, which have the ability to partially restore the proteinaceous matrix around the starch granules (14, 29).

*o2* endosperm modification is the result of a complex genetic action that involves multiple *o2* modifier loci. Seven *o2* modifiers have been mapped to six chromosomes (8), among which one, designated *o2 modifier1* near the 27-kDa  $\gamma$ -zein locus, is tightly linked to elevated expression of 27-kDa  $\gamma$ -zein and functions as a major QTL for *o2* endosperm modification (8–10, 12). Down-regulation, silencing, or elimination of the 27-kDa  $\gamma$ -zein gene expression have similar effects in disrupting endosperm modification (Fig. 1) (13, 14), indicating that *o2* modifier-mediated endosperm modification in QPM depends on enhanced synthesis of 27-kDa  $\gamma$ -zein. Furthermore, the correlation between the degree of modification and the level of 27-kDa  $\gamma$ -zein in different populations created by a cross between QPM and a starchy *o2* mutant suggests the QTL (*qy27*) controlling enhanced expression of 27-kDa  $\gamma$ -zein is potentially an *o2* modifier. Based on this finding, we designed a series of experiments to map *qy27* by phenotyping the amount of 27-kDa  $\gamma$ -zein protein. We found that *qy27* increases 27-kDa  $\gamma$ -zein expression in a *cis* manner (Fig. S2C), excluding the possibility of it being a *trans*-factor that transcriptionally or posttranscriptionally regulates 27-kDa  $\gamma$ -zein gene expression. It is also unlikely that *qy27* is an enhancer affecting transcription of 27-kDa  $\gamma$ -zein gene in *cis*. With a F<sub>1</sub>BC<sub>1</sub> population of 6,912 individuals, *qy27* was mapped into a region ~100 kb, in which we didn't identify possible enhancer-like elements by comparison of Mo17 and B73 sequences.

In this study, we identified *qy27* as a gene duplication at the 27-kDa  $\gamma$ -zein locus (Fig. 4). This duplication affects four genes, of which two copies of the 27-kDa  $\gamma$ -zein gene are clearly associated with elevated transcription (Fig. S4B) and synthesis of 27-kDa  $\gamma$ -zein protein in the wild-type population and in QPM lines (Figs. 2 and 5 and Dataset S2). We investigated 38 different QPM lines and found they all bear the gene duplication (Fig. S5), indicating it has been selected during QPM breeding and is a likely candidate for the QTL. Loss of one gene copy in W22-Rb is associated with reduced accumulation of 27-kDa  $\gamma$ -zein compared with that in W22-Sab (Fig. 6). The identity of *qy27* as a duplication of the 27-kDa  $\gamma$ -zein gene is consistent with the overrepresented expression of the K0326y-*qy27* and Mo17-*qy27* alleles in the crosses with XF134 and our genetic analyses showing that *qy27* has a dosage effect on 27-kDa  $\gamma$ -zein expression and appears to act downstream of its regulatory



**Fig. 6.** Comparison of 27-kDa  $\gamma$ -zein expression in the wild-type W22 and W22-Rb stock. (A) Comparatively lower accumulation of 27-kDa  $\gamma$ -zein in W22-Rb stock. (B) Absence of the duplication in the W22-Rb stock.

transcription factor, PBF (Fig. 1 and Figs. S1 and S2). Coincidentally, *qy27* was mapped to the same region as the *o2 modifier1* locus, which has been shown to confer a major contribution to the vitreous phenotype in QPM (8). Although the *o2 modifier1* locus may contain a different QTL that is tightly linked to *qy27*, the elevated amount of 27-kDa  $\gamma$ -zein from the duplication is consistent with phenotype and clearly contributes to endosperm modification.

Copy-number variation resulting in altered expression of a target gene, and in turn plant development and manifestation of agronomic traits, has been demonstrated in many reports (32–35). Evidence from this and other studies (8, 10, 11, 13, 14) support the hypothesis that enhanced expression of 27-kDa  $\gamma$ -zein is a function of gene duplication. Previous studies indicated that zein proteins play different roles in PB formation (36, 37). The  $\gamma$ - and  $\beta$ -zeins are expressed a little earlier than  $\alpha$ - and  $\delta$ -zeins and are thought to initiate PB formation, whereas  $\alpha$ - and  $\delta$ -zeins are subsequently deposited into the inner region of PBs and enlarge their size (18). Because the expression of the  $\beta$ -zein gene is dramatically reduced in QPM because of the *o2* mutation (38), an elevated level of 27-kDa  $\gamma$ -zein becomes crucial to facilitate the formation of numerous small protein bodies that surround the starch granules and contribute to rebuilding the proteinaceous matrix. Biochemical analysis of the starch in QPM indicates that endosperm modification might also be associated with altered starch structure, probably a result of altered expression of granule-bound starch synthase I, pullulanase, and starch synthase III (29, 39). This feature, along with the altered composition of starch granules and a more extensive cytoplasmic proteinaceous matrix, is critical for restoration of kernel vitreousness (14, 29).

It should be noted that the level of 27-kDa  $\gamma$ -zein is higher in K0326Y QPM than in Mo17 wild-type, and it also varies among QPM lines (Figs. 2A and 5A). It is likely that K0326Y QPM has additional *o2* modifiers that don't exist in Mo17; some of these could further increase synthesis of 27-kDa  $\gamma$ -zein. For example, a locus linked to *o15*, a mutation on the distal arm of chromosome 7, affects 27-kDa  $\gamma$ -zein synthesis (8, 40). It has been suggested that some *o2* modifiers regulate 27-kDa  $\gamma$ -zein gene expression through a posttranscriptional mechanism (41). However, the nature and mechanism of the involved *o2* modifiers remains to be identified.

**Instability of *qy27* and Artificial Selection for QPM Breeding.** Because *qy27* is a duplication of the 27-kDa  $\gamma$ -zein locus, we were curious to know where it originates from and how it was selected. Therefore, we investigated a number of *Z. mays* ssp. *Parviglumis* accessions (maize wild progenitor) and found most contain this gene duplication (Fig. S8). This finding indicates that the duplication occurred before maize domestication and that the single-copy 27-kDa  $\gamma$ -zein alleles in modern maize might result from DNA rearrangement. Indeed, the duplication appears to be unstable and the DNA rearrangement was determined to occur frequently in some inbred lines (25, 27), as evidenced by W22-R<sub>b</sub>, a W22 stock with the only B gene, resulting from the rearranged W22-Sab allele (Fig. 6). However, DNA rearrangement causes a reduction in 27-kDa  $\gamma$ -zein expression (Fig. 6). We examined 159 inbred lines that express low levels of 27-kDa  $\gamma$ -zein and found they contain only one gene copy, of which 152 and 7 have A and B genes, respectively (Dataset S2). Although loss of one 27-kDa  $\gamma$ -zein gene causes no obvious change of phenotype in wild-type maize, the elevated expression of 27-kDa  $\gamma$ -zein is critical for the vitreous phenotype in QPM. We propose that the gene structure of *qy27* is under artificial selection and thus stabilized during QPM breeding (Fig. 5).

## Materials and Methods

**Plant Materials.** The 492 temperate maize inbred lines used for GWAS mapping were collected and genotyped by J.L.'s laboratory at China Agricultural University. The 280 lines of the IBM Synthetic 10 DH population for the linkage mapping were originally from Michael Lee's laboratory at Iowa

State University and their genotypes were provided by G.P.'s laboratory at Sichuan Agricultural University (23). Of 38 QPM inbred lines, 36 were obtained from X.F.'s laboratory at the Institute of Food Crops, Yunnan Academy of Agricultural Sciences. The isogenic lines of CM105, CM105o2, and CM105Mo2 (a QPM line developed from CIMMYT Pool 33) and K0326Y (a QPM line originally from South Africa) were provided by David Holding's laboratory at University of Nebraska, Lincoln, NE. The teosinte lines of *Z. mays* ssp. *Parviglumis* were provided by Zhongwei Lin at China Agricultural University, Beijing, China. All of the other materials were maintained from our own laboratory.

**Total Zein Extraction and Phenotypic Evaluation.** At least 20 kernels from each inbred or IBM line were finely ground with a coffee grinder (Electric Grinder) for 2 min, and 100 mg of flour were used for zein protein extraction. In brief, the flour was mixed and vortexed with 1 mL zein extraction buffer [70% (vol/vol) ethanol, 2% (vol/vol) 2-mercaptoethanol, 3.75 mM sodium borate (PH 10), 0.3% SDS] in a 2-mL tube (Axygen), incubated on the bench at room temperature for 2 h. Then the tube was centrifuged at 15,700  $\times g$  (Eppendorf) for 20 min, and 100  $\mu$ L supernatant liquid was transferred into a new tube with an additional 10  $\mu$ L 10% (wt/vol) SDS, the mixture was evacuated for 70 min at the condition of 45 °C, VAL model (Eppendorf). Next, 100  $\mu$ L of distilled water was then added to dissolve the protein. Finally, a mixture of 2- $\mu$ L protein solution and 8- $\mu$ L protein loading buffer was analyzed by SDS/PAGE gel [15% (wt/vol)] for a 27-kDa  $\gamma$ -zein accumulation pattern (Bio-Rad). B73 and Mo17 were selected as standards, and levels in B73 and Mo17 were recorded as 1 and 2, respectively, from SDS/PAGE gel for phenotype collection.

**RNA Extraction and Quantification.** Total RNA was extracted from the 18-DAP developing endosperms with TRIzol (Invitrogen) and purified with the RNeasy Mini Kit after digestion with DNase1 (Qiagen). cDNA was synthesized with the SuperScript III First Strand Kit (Invitrogen) using 2  $\mu$ g of RNA in a 20- $\mu$ L reaction system. Quantitative PCR was performed using SYBR Green (Takara) with the Bio-Rad CFX Connect. The maize *Actin* gene was used as a reference control and the comparative CT method ( $\Delta\Delta CT$  method) was used to calculate relative gene expression. All of the expression data generated by quantitative RT-PCR included at least five biological samples with three technical replications. The Student's *t* test was used to estimate statistical significance. All primers are listed in Table S1.

**Association Analysis.** The GWAS was performed by GAPIT (42) (gapit\_functions\_2.22) using the compressed MLM (43) approach with Q+K model. Kinship was obtained using emmax-kin (44) (version: beta-07Mar2010) with parameters: -v -h -s -d 10. Principal component analysis was calculated with gcta with the following two steps (45) (version: gcta64): (i) estimate genetic relationship matrix with parameters: -autosome-num 10-autosome-make-grm; (ii) input the genetic relationship matrix and output eigenvalues and eigenvectors with parameters: -grm -pca 20. The first three eigenvectors were used for GWAS analysis.

**Genotyping and QTL Analysis.** The IBM Syn10 DH population was used for QTL analysis by the GBS procedure. In general, each of 280 DH lines was sequenced with an average genome depth of 0.31x. The custom-made Mo17 pseudomolecules were generated with approximate 26.5x genome coverage for SNPs calling. Subsequently, the reads that only allow one mismatch perfectly mapped to the B73 reference were kept as high quality SNPs, recognized at this SNP site as genotype for of DH lines, and we discarded the reads that have multiple physical coordinate and low quality sequence (quality value  $\leq$  5). Bin markers were then generated with a sliding-window approach. Finally, by using two different software packages, *JoinMap* 4.1 and *MSTMap*, a genetic map was constructed with 6,618 recombination bins for QTL analysis (46, 47). QTL analysis was performed according to observing expanded map by using *QTL Cartographer* (48). Composite interval mapping pattern was used for population type (Zmap model 6).

**Mapping and Cloning of *qy27*.** With GWAS and linkage mapping, the major QTL controlling the high expression of 27-kDa  $\gamma$ -zein was located within a narrow interval of 1 Mb. Then, 6,912 seeds of (Mo17  $\times$  B73)  $\times$  B73 population were used for fine mapping. Each seed was cut into two halves, one half for zein protein extraction (phenotyping the high or low level of 27-kDa  $\gamma$ -zein), the other for DNA extraction. Taking advantage of the SSR and SNP markers, *qy27* was finally located into a 100-kb interval between the marker 0916-2 and Ch7-120.35.

**Mo17 BAC Sequencing.** The S-Mo17<sup>Rf3Rf3</sup> BACs were provided by F.Q. at Huazhong Agricultural University. BAC 378 and 432 were successfully screened by four pairs of primers (Table S1), which were designed based on the B73 genome in this region. Then, the two BACs were sequenced by PacBio RS II (Shanghai Hanyu Bio-tech). The reads were assembled using de novo assembly pipeline (49) of RS HGAP v2 (hierarchical genome-assembly process) from Pacific Biosciences SMRT Analysis software v2.3.0, yielding a continuous 258-kb sequence.

**Southern Blotting.** Approximately 50 µg of genomic DNA from Mo17, B73, and K0326Y seedlings was digested with HindIII and separated on a 1% agarose gel, denatured and blotted onto a Hybond-N<sup>+</sup> membrane (GE Healthcare). Hybridization with a specific 27-kDa  $\gamma$ -zein probe was performed according to the instructions of the North2South Chemiluminescent Hybridization and Detection Kit (Thermal). The luminescence signal was visualized on the Tanon-5200 Chemiluminescent Imaging System (Tanon Science and Technology).

- Wheeler T, von Braun J (2013) Climate change impacts on global food security. *Science* 341(6145):508–513.
- Wu Y, Yuan L, Guo X, Holding DR, Messing J (2013) Mutation in the seed storage protein kafirin creates a high-value food trait in sorghum. *Nat Commun* 4:2217.
- Gilding EK, et al. (2013) Allelic variation at a single gene increases food value in a drought-tolerant staple cereal. *Nat Commun* 4:1483.
- Osborne TB, Mendel LB; With the cooperation of Edna LF and Alfred JW (1914) Nutritive properties of proteins of the maize kernel. *J Biol Chem* 18(1):1–16.
- Vasal SK, Villegas E, Bjarnason M, Gelaw B, Goertz P (1980) Genetic modifiers and breeding strategies in developing hard endosperm opaque-2 materials. *Improvement of Quality Traits of Maize for Grain and Silage Use*, eds Pollmer WG, Phillips RH (Martinus Nijhoff, London), pp 37–73.
- Paez AV, Helm JL, Zuber MS (1969) Lysine content of opaque-2 maize kernels having different phenotypes. *Crop Sci* 9(2):251–252.
- Nuss ET, Tanumihardjo SA (2011) Quality protein maize for Africa: Closing the protein inadequacy gap in vulnerable populations. *Adv Nutr* 2(3):217–224.
- Holding DR, et al. (2008) Genetic analysis of opaque2 modifier loci in quality protein maize. *Theor Appl Genet* 117(2):157–170.
- Geetha KB, Lending CR, Lopes MA, Wallace JC, Larkins BA (1991) opaque-2 modifiers increase gamma-zein synthesis and alter its spatial distribution in maize endosperm. *Plant Cell* 3(11):1207–1219.
- Holding DR, et al. (2011) Characterization of opaque2 modifier QTLs and candidate genes in recombinant inbred lines derived from the K0326Y quality protein maize inbred. *Theor Appl Genet* 122(4):783–794.
- Lopes MA, Larkins BA (1991) Gamma-zein content is related to endosperm modification in quality protein maize. *Crop Sci* 31(6):1655–1662.
- Lopes MA, Larkins BA (1995) Genetic analysis of opaque2 modifier gene activity in maize endosperm. *Theor Appl Genet* 91(2):274–281.
- Yuan L, Dou Y, Kianian SF, Zhang C, Holding DR (2014) Deletion mutagenesis identifies a haploinsufficient role for  $\gamma$ -zein in opaque2 endosperm modification. *Plant Physiol* 164(1):119–130.
- Wu Y, Holding DR, Messing J (2010) Gamma-zeins are essential for endosperm modification in quality protein maize. *Proc Natl Acad Sci USA* 107(29):12810–12815.
- Wolf MJ, Khoo U, Seckinger HL (1967) Subcellular structure of endosperm protein in high-lysine and normal corn. *Science* 157(3788):556–557.
- Larkins BA, Dalby A (1975) In vitro synthesis of zein-like protein by maize polyribosomes. *Biochem Biophys Res Commun* 66(3):1048–1054.
- Burr B, Burr FA (1976) Zein synthesis in maize endosperm by polyribosomes attached to protein bodies. *Proc Natl Acad Sci USA* 73(2):515–519.
- Lending CR, Larkins BA (1989) Changes in the zein composition of protein bodies during maize endosperm development. *Plant Cell* 1(10):1011–1023.
- Zhang Z, Yang J, Wu Y (2015) Transcriptional regulation of zein gene expression in maize through the additive and synergistic action of opaque2, prolamine-box binding factor, and O2 heterodimerizing proteins. *Plant Cell* 27(4):1162–1172.
- Wu Y, Messing J (2012) Rapid divergence of prolamain gene promoters of maize after gene amplification and dispersal. *Genetics* 192(2):507–519.
- Flint-Garcia SA, Bodnar AL, Scott MP (2009) Wide variability in kernel composition, seed characteristics, and zein profiles among diverse maize inbreds, landraces, and teosinte. *Theor Appl Genet* 119(6):1129–1142.
- Lopes MA, Takasaki K, Bostwick DE, Helentjaris T, Larkins BA (1995) Identification of two opaque2 modifier loci in quality protein maize. *Mol Gen Genet* 247(5):603–613.
- Liu H, et al. (2015) An ultra-high-density map as a community resource for discerning the genetic basis of quantitative traits in maize. *BMC Genomics* 16(1):1078.
- Chen J, et al. (2014) Dynamic transcriptome landscape of maize embryo and endosperm development. *Plant Physiol* 166(1):252–264.
- Das OP, Poliak E, Ward K, Messing J (1991) A new allele of the duplicated 27kD zein locus of maize generated by homologous recombination. *Nucleic Acids Res* 19(12):3325–3330.
- Das OP, Ward K, Ray S, Messing J (1991) Sequence variation between alleles reveals two types of copy correction at the 27-kDa zein locus of maize. *Genomics* 11(4):849–856.
- Das OP, Levi-Minzi S, Koury M, Benner M, Messing J (1990) A somatic gene rearrangement contributing to genetic diversity in maize. *Proc Natl Acad Sci USA* 87(20):7809–7813.
- Gibbon BC, Larkins BA (2005) Molecular genetic approaches to developing quality protein maize. *Trends Genet* 21(4):227–233.
- Gibbon BC, Wang X, Larkins BA (2003) Altered starch structure is associated with endosperm modification in quality protein maize. *Proc Natl Acad Sci USA* 100(26):15329–15334.
- Duvick DN (1961) Protein granules of maize endosperm cells. *Cereal Chem* 38:374–385.
- Wu Y, Messing J (2010) RNA interference-mediated change in protein body morphology and seed opacity through loss of different zein proteins. *Plant Physiol* 153(1):337–347.
- Wang Y, et al. (2015) Copy number variation at the GL7 locus contributes to grain size diversity in rice. *Nat Genet* 47(8):944–948.
- Tang YC, Amon A (2013) Gene copy-number alterations: A cost-benefit analysis. *Cell* 152(3):394–405.
- Cook DE, et al. (2012) Copy number variation of multiple genes at Rhg1 mediates nematode resistance in soybean. *Science* 338(6111):1206–1209.
- Stranger BE, et al. (2007) Relative impact of nucleotide and copy number variation on gene expression phenotypes. *Science* 315(5813):848–853.
- Coleman CE, Herman EM, Takasaki K, Larkins BA (1996) The maize gamma-zein sequesters alpha-zein and stabilizes its accumulation in protein bodies of transgenic tobacco endosperm. *Plant Cell* 8(12):2335–2345.
- Bagga S, Adams HP, Rodriguez FD, Kemp JD, Sengupta-Gopalan C (1997) Coexpression of the maize delta-zein and beta-zein genes results in stable accumulation of delta-zein in endoplasmic reticulum-derived protein bodies formed by beta-zein. *Plant Cell* 9(9):1683–1696.
- Cord Neto G, et al. (1995) The involvement of Opaque 2 on  $\beta$ -prolamain gene regulation in maize and Coix suggests a more general role for this transcriptional activator. *Plant Mol Biol* 27(5):1015–1029.
- Wu H, et al. (2015) Pullulanase and starch synthase III are associated with formation of vitreous endosperm in quality protein maize. *PLoS One* 10(6):e0130856.
- Dannenhoffer JM, Bostwick DE, Or E, Larkins BA (1995) opaque-15, a maize mutation with properties of a defective opaque2 modifier. *Proc Natl Acad Sci USA* 92(6):1931–1935.
- Or E, Boyer SK, Larkins BA (1993) opaque2 modifiers act post-transcriptionally and in a polar manner on gamma-zein gene expression in maize endosperm. *Plant Cell* 5(11):1599–1609.
- Lipka AE, et al. (2012) GAPIT: Genome association and prediction integrated tool. *Bioinformatics* 28(18):2397–2399.
- Zhang Z, et al. (2010) Mixed linear model approach adapted for genome-wide association studies. *Nat Genet* 42(4):355–360.
- Kang HM, et al. (2010) Variance component model to account for sample structure in genome-wide association studies. *Nat Genet* 42(4):348–354.
- Yang J, Lee SH, Goddard ME, Visscher PM (2011) GCTA: A tool for genome-wide complex trait analysis. *Am J Hum Genet* 88(1):76–82.
- VAN OIJEN JW, JV VANNO (2011) Multipoint maximum likelihood mapping in a full-sib family of an outbreeding species. *Genet Res* 93(5):343–349.
- Wu Y, Bhat PR, Close TJ, Lonardi S (2008) Efficient and accurate construction of genetic linkage maps from the minimum spanning tree of a graph. *PLoS Genet* 4(10):e1000212.
- Silva LdaC, Wang S, Zeng ZB (2012) Composite interval mapping and multiple interval mapping: Procedures and guidelines for using Windows QTL Cartographer. *Methods Mol Biol* 871:75–119.
- Chin CS, et al. (2013) Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nat Methods* 10(6):563–569.