

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

---

Faculty Publications from the Center for Plant  
Science Innovation

Plant Science Innovation, Center for

---

9-2013

## Dynamic Expression of Imprinted Genes Associates with Maternally Controlled Nutrient Allocation during Maize Endosperm Development

Mingming Xin  
*University of Arizona*


Ruolin Yang  
*University of Arizona*

Guosheng Li  
*University of Arizona*

Hao Chen  
*Boston University*

John Laurie  
*University of Nebraska - Lincoln*

Follow this and additional works at: <https://digitalcommons.unl.edu/plantscifacpub>  
See next page for additional authors

 Part of the [Plant Biology Commons](#), [Plant Breeding and Genetics Commons](#), and the [Plant Pathology Commons](#)

---

Xin, Mingming; Yang, Ruolin; Li, Guosheng; Chen, Hao; Laurie, John; Ma, Chuang; Wang, Dongfang; Yao, Yingyin; Larkins, Brian A.; Sun, Qixin; Yadegari, Ramin; Wang, Xiangfeng; and Ni, Zhongfu, "Dynamic Expression of Imprinted Genes Associates with Maternally Controlled Nutrient Allocation during Maize Endosperm Development" (2013). *Faculty Publications from the Center for Plant Science Innovation*. 110. <https://digitalcommons.unl.edu/plantscifacpub/110>

This Article is brought to you for free and open access by the Plant Science Innovation, Center for at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Faculty Publications from the Center for Plant Science Innovation by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

---

**Authors**

Mingming Xin, Ruolin Yang, Guosheng Li, Hao Chen, John Laurie, Chuang Ma, Dongfang Wang, Yingyin Yao, Brian A. Larkins, Qixin Sun, Ramin Yadegari, Xiangfeng Wang, and Zhongfu Ni

LARGE-SCALE BIOLOGY ARTICLE

# Dynamic Expression of Imprinted Genes Associates with Maternally Controlled Nutrient Allocation during Maize Endosperm Development<sup>WJOPEN</sup>

Mingming Xin,<sup>a,1</sup> Ruolin Yang,<sup>a,1</sup> Guosheng Li,<sup>a</sup> Hao Chen,<sup>a,2</sup> John Laurie,<sup>a,3</sup> Chuang Ma,<sup>a</sup> Dongfang Wang,<sup>a,4</sup> Yingyin Yao,<sup>b</sup> Brian A. Larkins,<sup>a,5</sup> Qixin Sun,<sup>b</sup> Ramin Yadegari,<sup>a</sup> Xiangfeng Wang,<sup>a,6</sup> and Zhongfu Ni<sup>b</sup>

<sup>a</sup>School of Plant Sciences, University of Arizona, Tucson, Arizona 85721-0036

<sup>b</sup>State Key Laboratory for Agrobiotechnology, Key Laboratory of Crop Heterosis Utilization (Ministry of Education), Beijing Key Laboratory of Crop Genetic Improvement, China Agricultural University, Beijing 100193, China

ORCID IDs: 0000-0002-3306-5594 (M.X.); 0000-0001-6241-6317 (R.Y.); 0000-0002-6406-5597 (X.W.).

In angiosperms, the endosperm provides nutrients for embryogenesis and seed germination and is the primary tissue where gene imprinting occurs. To identify the imprintome of early developing maize (*Zea mays*) endosperm, we performed high-throughput transcriptome sequencing of whole kernels at 0, 3, and 5 d after pollination (DAP) and endosperms at 7, 10, and 15 DAP, using B73 by Mo17 reciprocal crosses. We observed gradually increased expression of paternal transcripts in 3- and 5-DAP kernels. In 7-DAP endosperm, the majority of the genes tested reached a 2:1 maternal versus paternal ratio, suggesting that paternal genes are nearly fully activated by 7 DAP. A total of 116, 234, and 63 genes exhibiting parent-specific expression were identified at 7, 10, and 15 DAP, respectively. The largest proportion of paternally expressed genes was at 7 DAP, mainly due to the significantly deviated parental allele expression ratio of these genes at this stage, while nearly 80% of the maternally expressed genes (MEGs) were specific to 10 DAP and were primarily attributed to sharply increased expression levels compared with the other stages. Gene ontology enrichment analysis of the imprinted genes suggested that 10-DAP endosperm-specific MEGs are involved in nutrient uptake and allocation and the auxin signaling pathway, coincident with the onset of starch and storage protein accumulation.

## INTRODUCTION

Angiosperm seed development depends on the coordinated regulation of gene activities in the filial embryo and endosperm as well as the maternal tissues surrounding these structures (Lopes and Larkins, 1993; Ray, 1997; Berger et al., 2006; Ingram, 2010; Li and Berger, 2012). The diploid embryo (1 maternal:1 paternal, 1m:1p hereafter) and the triploid endosperm (2m:1p) arise as a result of double fertilization of the central cell (2n) and the egg cell (1n), respectively, by one of two sperm cells (1n) delivered by the pollen tube (Yadegari and Drews, 2004). Although a terminal

tissue, the endosperm not only plays a crucial role by providing nutrients to support embryogenesis and seed germination, it may play a regulatory role in coordinating proper seed development (Berger et al., 2006). In most angiosperms, the endosperm undergoes an initial phase of syncytial development immediately after fertilization, where nuclear divisions occur in the absence of cytokinesis. This is followed by a period of cellularization, during which existing endosperm nuclei are bound by cell walls (Berger, 2003; Olsen, 2004; Sabelli and Larkins, 2009). In most dicots, including *Arabidopsis thaliana*, the cellularized endosperm does not undergo significant proliferative growth and is absorbed by the developing embryo during seed development (Brown et al., 1999). However, in grasses, immediately following cellularization, the endosperm undergoes an intense period of cell division, followed by differentiation and specialization to form a nutritive tissue that supports embryo and seedling development (Olsen, 2001; Sabelli and Larkins, 2009). In maize (*Zea mays*), this mitotic phase begins in the central region at ~4 to 5 d after pollination (DAP) and lasts until 8 to 12 DAP, but it may continue until ~20 to 25 DAP in the peripheral regions of the endosperm. Differentiation of endosperm cells follows these proliferative processes and results in four primary regions or compartments: aleurone, starchy endosperm, transfer layer, and the embryo-surrounding region (Sabelli and Larkins, 2009).

How the expression of the filial genomes is regulated to control development of the endosperm and indirectly the embryo and seedling through the proper control of the absorptive functions of

<sup>1</sup> These authors contributed equally to this work.

<sup>2</sup> Current address: Bioinformatics Graduate Program, Boston University, Boston, MA 02215.

<sup>3</sup> Current address: Center for Plant Science Innovation, University of Nebraska, Lincoln, NE 68588.

<sup>4</sup> Current address: Department of Biology, Spelman College, Atlanta, GA 30314.

<sup>5</sup> Current address: 230 Whittier Research Center, University of Nebraska, Lincoln, NE 68583-0857.

<sup>6</sup> Address correspondence to xwang1@cals.arizona.edu.

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.plantcell.org) is: Xiangfeng Wang (xwang1@cals.arizona.edu).

<sup>WJOPEN</sup> Online version contains Web-only data.

<sup>OPEN</sup> Articles can be viewed online without a subscription.

www.plantcell.org/cgi/doi/10.1105/tpc.113.115592

the endosperm is not understood. Specifically, how unequal contribution of the paternal and maternal genomes and individual genes control these processes has not been fully explored. Some recent studies have indicated that transcriptional activation of the two parental genomes might not occur simultaneously in the zygote during the first 1 to 3 DAP, resulting in their unbalanced expression during early seed development (Vielle-Calzada et al., 2000; Grimanelli et al., 2005; Autran et al., 2011). However, other studies showed the paternal and maternal genomes might contribute equally in early stages of embryonic development (Weijers et al., 2001; Meyer and Scholten, 2007; Nodine and Bartel, 2012). Analyses of early gene expression in *Arabidopsis* and maize endosperm suggested that the activation of paternal alleles lags behind that of maternal alleles (Vielle-Calzada et al., 2000; Köhler and Grossniklaus, 2005; Autran et al., 2011). Thus, our understanding of the dynamics of paternal genome activation, that is, the timing and the proportion of the paternal to maternal genome activity during early endosperm development, is ambiguous.

Genomic imprinting is a form of allele-biased expression involving the uniparental transcription of select genes in a parent-of-origin manner. In angiosperms, imprinted genes are primarily found in the endosperm, with only one exception (*maternally expressed in embryo 1*) in the maize embryo (Jahnke and Scholten, 2009). However, due to its small size, early-developing endosperm tissue has proven relatively intractable for the standard molecular studies necessary to understand genomic imprinting mechanisms. Genome-wide studies in hybrid endosperms of *Arabidopsis*, rice (*Oryza sativa*), and maize identified several hundred imprinted genes via high-throughput transcriptome sequencing (RNA-Seq) technology; however, the actual number of imprinted genes may have been largely underestimated because most of these experiments involved only one developmental stage (Gehring et al., 2011; Hsieh et al., 2011; Luo et al., 2011; Waters et al., 2011; Wolff et al., 2011; Zhang et al., 2011). Furthermore, only a small portion of the imprinted genes were commonly identified in independent studies of *Arabidopsis* and maize. This lack of corroboration is likely for multiple reasons, including the use of different ecotypes, different developmental stages, different sets of polymorphic nucleotide information, and different bioinformatic criteria to define imprinted gene sets (Gehring et al., 2011; Hsieh et al., 2011; Waters et al., 2011; Wolff et al., 2011; Zhang et al., 2011; Ikeda, 2012).

In plants, the molecular mechanisms underlying gene imprinting are best understood from studies of several *Arabidopsis* imprinted genes encoding components of the fertilization-independent seed Polycomb repressive complex 2 (PRC2). These genes include *FERTILIZATION-INDEPENDENT ENDOSPERM (FIE)*, *MULTICOPY SUPPRESSOR OF IRA1*, *MEDEA (MEA)*, and *FERTILIZATION-INDEPENDENT SEED2 (FIS2)*. Mutations in the PRC2 genes cause autonomous formation of endosperms without fertilization and abnormal seed development if fertilized (Gehring et al., 2004; Guitton and Berger, 2005). The two homologs of the *Arabidopsis FIE* gene in maize are *fie1* and *fie2*, a pair of duplicated genes exhibiting different expression patterns and modes of imprinting during endosperm development (Danilevskaya et al., 2003). *fie1* is specifically expressed in the endosperm, and only maternal transcripts can be detected at 6

DAP; they peak at 9 DAP and then gradually decrease and become undetectable by 12 DAP (Danilevskaya et al., 2003). By contrast, *fie2* is nearly universally expressed in maize kernel and vegetative tissues (Springer et al., 2002). In the endosperm, maternal transcripts of *fie2* are detected as early as 2 DAP, and the paternal allele of *fie2* is activated after 5 DAP, followed by activation of biallelic expression after 9 DAP (Danilevskaya et al., 2003). Therefore, the parent-of-origin expression of imprinted genes in these plants appears to be transient and is likely correlated with specific cellular events that occur during particular stages of endosperm development.

The biological implications of genomic imprinting in plants have been inferred primarily from phenotypic studies of reciprocal interploidy crosses (Lin, 1984; Scott et al., 1998). In *Arabidopsis*, whereas paternal-excess crosses promoted endosperm (2m:2p) and embryo (2m:4p) development, maternal-excess crosses inhibited growth of the endosperm (4m:1p) and embryo (2m:1p). On average, seeds produced from the paternal-excess crosses were almost 3 times heavier than those from the maternal-excess crosses (Scott et al., 1998). These results and similar observations in maize interploidy crosses support the kinship theory of imprinting, which hypothesizes a conflicting relationship between the two parental genomes: Maternally expressed alleles tend to restrict the growth of offspring by limiting nutrient allocation to the offspring, whereas the paternally expressed alleles would have the opposite tendency (Haig and Westoby, 1989; Wilkins and Haig, 2003). Although this model does not provide a mechanistic explanation of genomic imprinting at the molecular level, it has been supported by functional studies of a number of individual genes. For instance, the loss of function of two maternally expressed genes (MEGs; *MEA* and *FIS2*) in *Arabidopsis* allows fertilization-independent endosperm development, suggesting these genes might act as repressive regulators of endosperm development (Ohad et al., 1996; Chaudhury et al., 1997; Grossniklaus et al., 1998; Kiyosue et al., 1999; Luo et al., 1999). However, demonstration of the function of maize *meg1*, which is expressed in the basal endosperm transfer layer (BETL) cells between 4 and 25 DAP, as an endosperm growth enhancer and promoter of nutrient uptake from maternal tissues into the endosperm, has been used to suggest an alternative adaptive model for evolution of gene imprinting (Wolf and Hager, 2006; Costa et al., 2012).

To identify additional seed imprinted genes and understand the developmental basis of gene imprinting in maize endosperm, we performed a genome-wide search for imprinted genes expressed in developing kernels at 0, 3, and 5 DAP and endosperm at 7, 10, and 15 DAP from reciprocal crosses of maize B73 and Mo17 inbred lines. These six stages are inclusive of the key developmental events central to endosperm function as an absorptive structure, including early proliferation and differentiation and the initial stages of starch and storage protein accumulation (Sabelli and Larkins, 2009). We observed gradual transcriptional activation of the paternal genome in 3- and 5-DAP kernels, followed by complete paternal genome activation in 7-DAP endosperm. We identified 290 imprinted genes in hybrid maize endosperm using RNA-Seq analyses and experimentally validated a number of these genes by RT-PCR and cleaved amplified polymorphic sequence (CAPS) assays. Our analysis indicates the expression of

imprinted genes changes dynamically during endosperm development, with the majority of MEGs uniquely expressed at 10 DAP and the majority of the paternally expressed genes (PEGs) expressed predominantly at 7 DAP, compared with other stages of development. Additionally, gene ontology (GO) analyses revealed the MEGs and PEGs are enriched in distinct functional pathways.

## RESULTS

### Mo17 Genome Assembly, Single Nucleotide Polymorphism Identification, and RNA-Seq Data Processing

We performed reference-guided assembly of the Mo17 genome based on the B73 genome (release 5b.60, <http://maizesequence.org>) using 650-Gb Illumina short reads obtained from the Joint Genome Institute. The resulting Mo17 genome assembly comprised 2,058,527,894 bases containing a total of 117,847,390-bp short gaps. Thus, the coverage of the Mo17 genome was ~94.3%, producing a similar genome size to the 2,066,432,971 bases in the B73 reference. The identification of single nucleotide polymorphisms (SNPs) and insertions and deletions (INDELs) was performed using the SHORE pipeline (Schneeberger et al., 2011), which identified 6,557,611 SNPs, 157,994 insertions, and 191,549 deletions between the B73 and Mo17 genomes. In the working gene set (WGS), 38,557 protein-coding genes contained at least one SNP (see Supplemental Figure 1 online). The 90-nucleotide pair-end RNA-Seq reads from the 12 libraries were mapped to the B73 and Mo17 genomes using Tophat2 (Kim et al., 2013). The B73-unique, Mo17-unique, and B73/Mo17 common reads were distinguished based on SNP information. After excluding genes with few reads in the 12 samples, 24,984 SNP-containing genes expressed (at least six reads from combined 12 samples) were retained to measure allelic expression.

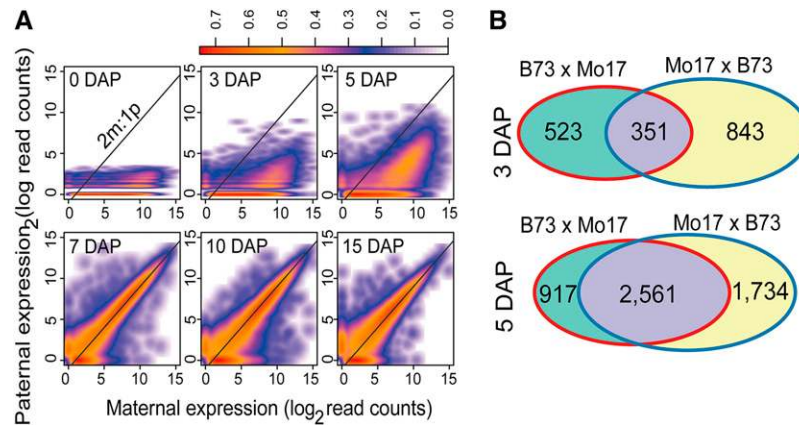
Because of short gaps in the Mo17 genome assembly that resulted from insufficient read coverage or high levels of variation between Mo17 and B73, ~10 to 15% of the RNA-Seq reads from the 12 samples of the hybrid kernel and endosperm could not be mapped to the Mo17 genome, but they could be mapped to the B73 genome. This issue potentially led to biased reporting of higher expression for many B73 alleles, which might influence the identification of the imprinted genes. Thus, we normalized the expression levels of the Mo17 and B73 alleles according to the gap size (see Supplemental Figure 2A and Supplemental Methods1 online). Additionally, genes containing gaps in their coding regions longer than 100 bp were not included in further analyses because such long gaps might influence the accuracy of the allelic expression measurements (see Supplemental Figure 2B online). The expression information for each gene per sample included the count of reads common to the B73 and Mo17 alleles and of reads unique to either B73 or Mo17 alleles, as distinguished by SNP information. Whereas the B73/Mo17 common reads exhibited a 1:1 ratio, the B73- and Mo17-unique reads with different ratios were used to measure allelic expression (see Supplemental Figure 3 online). Specifically, the fraction of maternal reads, including the B73-unique reads in the B73×Mo17 cross and the Mo17-unique reads in the Mo17×B73 cross, represents the expression of

maternal alleles and vice versa for paternal expression. The allelic expression inferred from allele-specific reads was proportional to total gene expression, indicating the ratio of allele-specific reads properly reflected the ratio of the maternal and paternal transcripts in the RNA samples (see Supplemental Figure 4 online).

### Postfertilization Activation of the Paternal Transcriptome

Upon double fertilization, the paternal genome inherited from the sperm cell is activated in the embryo and the endosperm. With support of informative SNPs and high-throughput sequencing, we examined the process of paternal genome activation in 3- and 5-DAP maize hybrid kernels. To minimize artifacts from ambiguous mapping, genes identified with more than five paternally derived reads at 0 DAP were excluded from further analysis. By plotting the paternal versus maternal expression for all retained genes, we observed a gradual activation of the paternal genome in 3- and 5-DAP kernels in terms of both the number of activated genes and their expression levels (Figure 1A). The number of genes with activated paternal alleles (>10 paternal reads from reciprocal crosses) increased from 0 at 0 DAP to 941 at 3 DAP and then to 4063 at 5 DAP. The corresponding fraction of paternal reads to total reads for the activated paternal alleles rose from 0 at 0 DAP to 0.0012 at 3 DAP and then to 0.0081 at 5 DAP. In 7-DAP endosperm, 11,027 genes containing SNPs were expressed; of these, 8128 (73.71%) exhibited the expected 2m:1p ratio ( $\chi^2$  test,  $P$  value > 0.001), whereas 2899 (26.29%) genes exhibited allele-biased expression ( $\chi^2$  test,  $P$  value  $\leq$  0.001). Similar proportions were observed in 10-DAP endosperm, with 72.75% (7691) of the genes exhibiting the expected ratio and 27.25% (2882) exhibiting an allelic bias. These results suggest that complete transcriptional activation of the paternal genome was achieved by 7 DAP in the endosperm. This observation is consistent with a previous report (Grimanelli et al., 2005). However, identification of paternal transcripts in the complex structure of the maize kernel indicated that at least a small fraction (<5%) of the paternal alleles might already be activated by 3 DAP. Although these genes had substantially fewer paternal transcripts than maternal transcripts at 3 and 5 DAP, we still cannot conclude an earlier activation of maternal compared with paternal alleles during maize seed development because ~97.5% of the SNP-containing genes detected with at least 10 reads in the endosperm stages could also be detected in the 0-DAP kernel stage.

We observed higher numbers and transcript abundances of genes with activated paternal alleles in the Mo17×B73 cross than in the B73×Mo17 cross at 3 and 5 DAP (Figure 2B; see Supplemental Figure 5 online). Namely, at 3 DAP, 1194 (8.3%) genes with activated paternal alleles (more than five paternal reads) were identified in the Mo17×B73 compared with 874 (6.4%) in B73×Mo17 cross, respectively, and only 351 (20.4%) of these were commonly activated in the two reciprocal crosses (Figure 1B); at 5 DAP, 4295 (25.3%) genes were paternally activated in the Mo17×B73 compared with 3478 (20.3%) in B73×Mo17 crosses, and, surprisingly, only 2561 (49.1%) of these genes were commonly activated (Figure 1B), while in 7-, 10-, and 15-DAP endosperm, the proportion of commonly activated genes accounted for 74.3, 71.3, and 70.9%, respectively. To determine whether the activated paternal alleles in the two reciprocal



**Figure 1.** Paternal Genome Activation in the Maize Kernels and Endosperms.

**(A)** Full activation of the paternal genome is likely achieved by 7 DAP. The expression level of paternal alleles (y axis) and maternal alleles (x axis) are represented by the log<sub>2</sub>-transformed sum of the paternally and maternally derived reads in the reciprocal crosses, respectively. The color scale in blue (low), yellow (medium), and red (high) represents the relative density of the genes. The solid diagonal line represents the expected 2m:1p ratio.

**(B)** Numbers of genes with transcriptionally activated paternal alleles in the reciprocal crosses at 3 and 5 DAP.

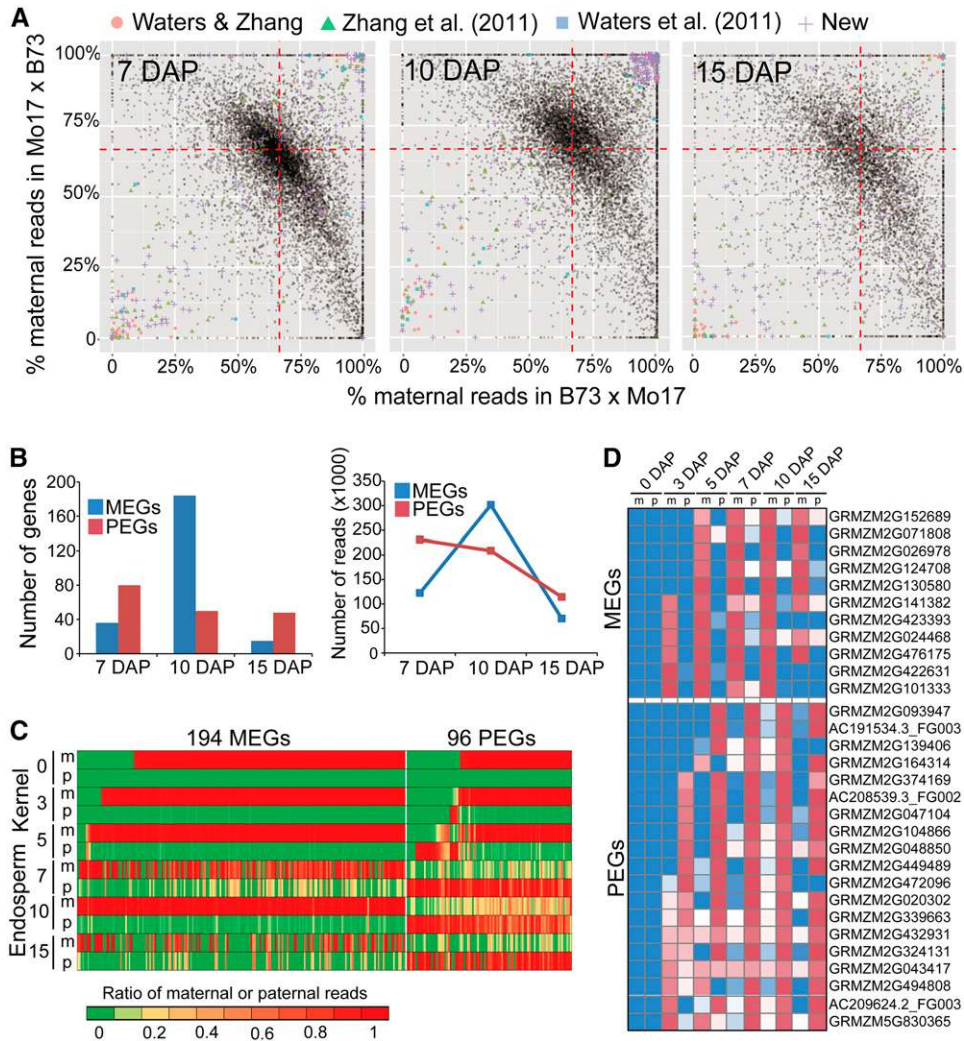
crosses were functionally different, we compared the GO enrichments of the paternally activated genes in the 3- and 5-DAP kernels of the Mo17×B73 and B73×Mo17 crosses. Overall, their enriched GO terms were very similar, without any GO class demonstrating significant differences (see Supplemental Figure 6 online). This similarity suggested the early-activated genes in both crosses might belong to the same functional pathways, even though their paternal alleles were not simultaneously activated at the same stage.

### Computational Identification of Imprinted Genes in the Maize Hybrid Endosperm

Because most of the genes expressed in endosperm samples were also expressed in the whole kernel, we decided to identify imprinted genes among the 11,027, 10,573, and 7777 SNP-containing genes (more than five allele-specific reads in both reciprocal crosses) identified in the 7-, 10-, and 15-DAP endosperm samples, respectively (see Supplemental Data Set 1 online). We first performed a  $\chi^2$  test to determine whether the ratio of maternally to paternally derived reads significantly deviated from the normal ratio of 2m:1p in each cross. The genes exhibiting parent-specific patterns in both reciprocal crosses were selected as candidate imprinted genes. At a significance level of  $P = 0.001$ , 284, 606, and 190 genes were determined to have allele-biased, parent-of-origin-specific expression patterns at 7, 10, and 15 DAP, respectively (see Supplemental Figure 7A online). In total, 300 PEGs and 499 MEGs were identified from the three developing endosperm stages by a  $\chi^2$  test (see Supplemental Figure 7B online). Among the 499 MEGs, 418 (83.8%) were identified at 10 DAP only, and only 15 (3.0%) were identified at all three time points (see Supplemental Figure 7C online). Among the 300 PEGs, 213, 130, and 163 were observed at 7, 10, and 15 DAP, respectively. Of these, 75 (25%) were present at all three time points (see Supplemental Figure 7C online).

We further narrowed the list of candidate imprinted genes using more stringent criteria to precisely reflect the parent-of-origin expression status of the allele-specific genes; namely, 90% of the maternal reads and 70% of the paternal reads of the total SNP-associated reads (with a minimum 40 reads) that mapped to a gene in both reciprocal crosses were used to define the MEGs and PEGs, respectively (Figure 2A). The ratio-based cutoff shortened the list to 290 imprinted genes identified at three developmental stages. These included 36, 184, and 15 MEGs and 80, 50, and 48 PEGs at 7, 10, and 15 DAP, respectively (Figure 2B). Among these imprinted genes, 38 were considered to be imprinted at all three stages, 51 were considered to be imprinted at two stages, and the remaining 206 were considered to be imprinted at only one stage, with the majority of this last category imprinted at 10 DAP. In addition, whereas the average expression level of the PEGs gradually decreased from 7 to 15 DAP, the MEGs reached their highest expression level at 10 DAP (Figure 2B). Subsequently, we traced the 290 imprinted genes identified in endosperm back to the early 3- and 5-DAP kernel stages to examine their early imprinting status (Figure 2C). Although we cannot completely exclude the influence of maternal tissues, especially for the MEGs, the genes that were not expressed at 0 DAP are likely to exhibit postfertilization imprinting behavior within the endosperm. We only found 11 MEGs and 19 PEGs that were not expressed in 0-DAP kernels (Figure 2D). Therefore, these genes were considered to be newly activated after fertilization. The imprinted expression patterns of these genes were evidently established at either 3 or 5 DAP, and all were maintained as such at 7 and 10 DAP. Interestingly, some of the same genes exhibited a loss of their parent-specific, imprinted expression pattern subsequently at 15 DAP (Figure 2D).

A number of MEGs and PEGs we identified were found to have functional counterparts in other plants. Among the 11 MEGs, an APETALA2/B3-like transcription factor gene (GRMZM2G423393) (Figure 2D) is highly similar to *NGATHA1*



**Figure 2.** Computational Identification of Imprinted Genes in the Maize Endosperm.

**(A)** Ratio-based cutoff to identify MEGs and PEGs. Spots clustered at the top right corners have >90% maternal reads (MEGs), whereas spots clustered in the bottom left corners have <30% of the maternal reads (PEGs). Imprinted genes in Waters et al. (2011), Zhang et al. (2011), and identified in this study are represented in a circle, triangle, square, and cross, respectively.

**(B)** Numbers and average expression levels of imprinted genes in 7-, 10-, and 15-DAP endosperm.

**(C)** Hierarchical clustering of the 290 imprinted genes. Parent-of-origin expression patterns of imprinted genes were exhibited in the 0-, 3-, and 5-DAP kernels and 7-, 10-, and 15-DAP endosperm. The color scale from green (low) to red (high) represents the ratio of maternal or paternal reads, m represents maternal allele, and p represents paternal allele.

**(D)** Heat map of 30 imprinted genes expressed after fertilization. Their imprinting status was traced back to as early as 3 DAP, and 25 imprinted genes are persistent to 15 DAP. The color scale in blue (low), white (medium) and red (high) represents the relative expression level of maternal or paternal alleles, m represents maternal allele, and p represents paternal allele.

(*NGA1*) in *Arabidopsis*, which redundantly together with three other *NGA* genes (*NGA2* to *NGA4*) direct style development in the gynoecium through the *YUCCA* (*YUC*)–mediated auxin synthesis pathway (Alvarez et al., 2009). Additionally, overexpression of the *Brassica rapa* gene *Br-NGA1* in *Arabidopsis* reduces lateral organ growth by inhibiting cell proliferation (Kwon et al., 2011). Of the 19 PEGs, 14 were paternally expressed as early as 3 DAP, and most (13 of 14) of their imprinted expression persisted until 15 DAP (Figure 2D). Among these genes, GRMZM2G472096 encodes

a MADS box transcription factor gene that is highly similar to *AGAMOUS-LIKE61*, which was shown to be required for proper development of the central cell and seed in *Arabidopsis* (Steffen et al., 2008). Two additional PEGs, a zinc finger gene (AC191534.3\_FG003) and a homeobox gene (GRMZM2G047104) are similar to *Arabidopsis* *VARIANT IN METHYLATION1* (*VIM1*) and *SAWADEE HOMEODOMAIN HOMOLOG1* (*SHH1*), respectively; both were reported to be associated with epigenetic regulation. In *Arabidopsis*, *vim1* mutations produced DNA hypomethylation of the

centromere and subsequent decondensation of heterochromatin (Woo et al., 2008), whereas *SHH1* was shown to be required for de novo and maintenance methylation in the RNA-directed DNA methylation pathway (Law et al., 2011). Although the functional roles of these early expressed genes are largely unknown in maize, identification of these MEGs and PEGs suggests an important role for transcription and epigenetic reprogramming in control of early endosperm development.

### Experimental Validation of Imprinted Genes

Two previous RNA-Seq studies identified 179 and 100 imprinted genes (223 genes in total) in 10- and 14-DAP endosperm samples (independently staged material), respectively, from reciprocal B73 and Mo17 crosses. However, only 56 of the imprinted genes were common (Waters et al., 2011; Zhang et al., 2011). We compared the 290 imprinted genes identified in our study with the 223 imprinted genes reported by Waters et al. (2011) and Zhang et al. (2011). Interestingly, only 81 of the imprinted genes identified in our study were also identified in one of the two previous studies (Figure 3A).

We selected two groups of genes for experimental validation using RT-PCR experiments followed by CAPS assays. The first group included seven genes previously identified by either Waters et al. (2011) or Zhang et al. (2011) that existed in our data. Based on this analysis, five of the seven genes, including three MEGs (GRMZM2G150134, GRMZM2G160687, and GRMZM2G370991) and two PEGs (GRMZM2G127160 and GRMZM2G028366), exhibited parent-specific imprinted expression patterns consistent with the previous reports (Figure 3B). The other two genes (GRMZM2G449489 and AC208539.3\_FG002) did not exhibit any polymorphisms between the parental alleles when using *Bgl*II and *Nla*III restriction enzymes, as reported by Zhang et al. (2011). Instead, when using the enzyme *Hind*III for GRMZM2G449489 in the CAPS assay, we observed a clear paternally biased expression pattern (Figure 3B). The second group included eight imprinted genes that were only identified in this study. All eight genes exhibited imprinting patterns by CAPS, consistent with the results from the RNA-Seq data (Figure 3B).

### A Large Number of Both Nonimprinted Allele-Specific Genes and Inbred Line-Dependent Imprinted Genes Are Expressed in Endosperm

In addition to the imprinted genes that have allele-specific expression dependent on parent of origin, two more types of allele-specifically expressed genes were also investigated: nonimprinted, allele-specifically expressed genes and inbred line-dependent imprinted genes. Nonimprinted, allele-specific expression is widespread and believed to be responsible for quantitative variations in the phenotypic traits of hybrid offspring (Guo et al., 2008). In the three staged samples of endosperm, 1688 B73 allele-specific and 1130 Mo17 allele-specific genes were detected using a  $\chi^2$  test that exhibited expression dosages deviating from the 2m:1p ratio ( $P \leq 0.001$ ). Among these, 303 B73 genes and 128 Mo17 genes exhibited a non-parent-specific, monoallelic expression pattern under our stringent ratio-based criterion ( $\geq 90\%$  reads in both reciprocal crosses; Figure 4A). Three genes

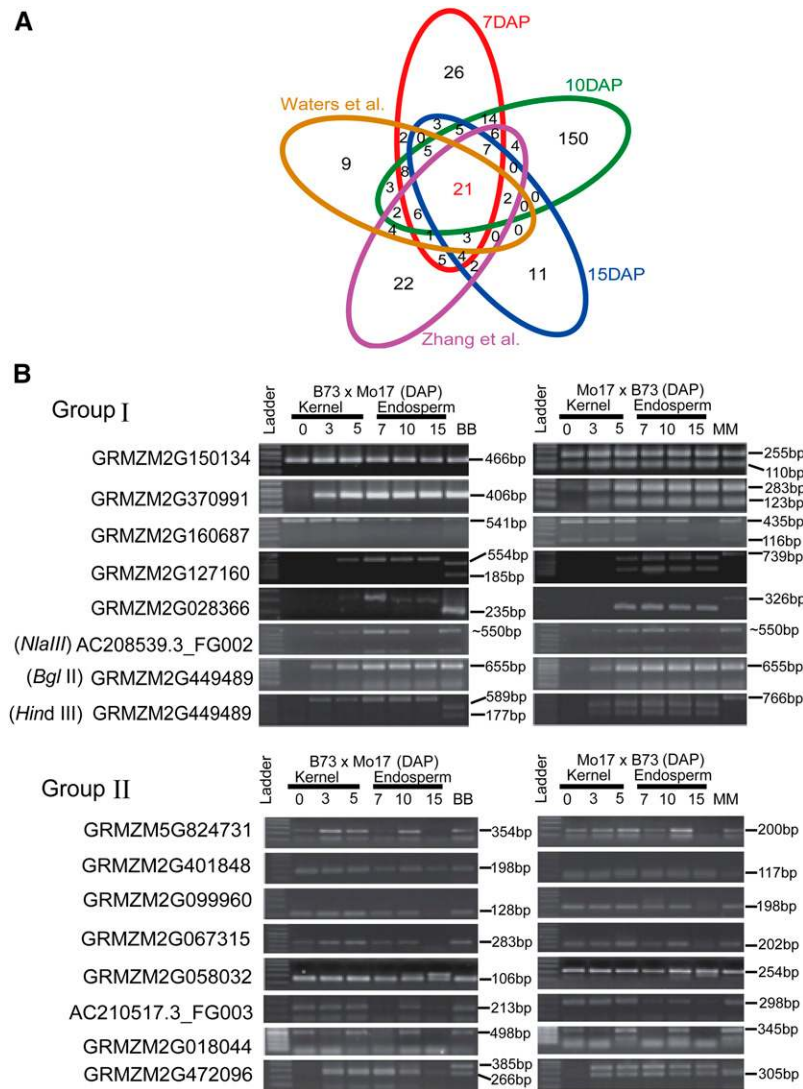
(GRMZM2G446999, GRMZM2G000361, and AC199068.2\_FG017) of this type were experimentally validated by CAPS experiments (Figure 4B).

The second type of gene, previously reported in *Arabidopsis*, rice, and maize, is an inbred line-dependent imprinted gene that exhibits allele-specific expression in one direction of each reciprocal pair of crosses but biallelic expression in the other direction (Gehring et al., 2011; Hsieh et al., 2011; Waters et al., 2011; Wolff et al., 2011). Using the same threshold for identifying imprinted genes but only considering a single direction of allele-specific expression, we identified 377 inbred line-dependent MEGs (ILMEGs) and 129 inbred line-dependent PEGs (ILPEGs) in the B73×Mo17 cross and 292 ILMEGs and 235 ILPEGs in the Mo17×B73 cross (Figure 4A). We used CAPS to experimentally validate the inbred line-dependent imprinting pattern for GRMZM2G147226 (encoding a Epsin NH2-Terminal Homology /Vps27p/Hrs/Stam family protein) (Figure 4B). Overall, these data indicate that our RNA-Seq approach was sufficiently robust to identify multiple allele-specific expression patterns in maize endosperm.

### Endosperm-Expressed MEGs and PEGs Encode Distinct Functional Groups

We performed Blast2GO analysis (Conesa et al., 2005; Conesa and Götting, 2008; Götting et al., 2008; Götting et al., 2011) to examine the functional distribution of the 290 imprinted genes identified in our study. Only four GO categories exhibited significantly higher functional enrichments compared with the background gene set ( $P \leq 0.01$ ). These groups included “regulation of gene expression by genetic imprinting,” “endosperm development,” “response to hormone stimulus (auxin response),” and “DNA binding (sequence-specific DNA binding)” (see Supplemental Figure 8A online). These functional categories indicated involvement of the imprinted genes in hormone signaling pathways and transcriptional regulation of endosperm development. When the 194 MEGs and 96 PEGs were examined separately, we found only two categories and one category, respectively, exhibiting significant enrichments for MEGs and PEGs ( $P \leq 0.01$ ) (see Supplemental Table 1 and Supplemental Figures 8B and 8C online). The first category of MEGs was the “response to hormone stimulus” group, which included 13 genes, six of which were involved in the auxin-mediated transcriptional response. The second was the “cytoplasmic membrane-bounded vesicle” category, which included 26 genes encoding a variety of enzymes, transporter proteins, and cell wall formation proteins located in the cytoplasmic membrane, suggesting these MEGs might be involved in intercellular nutrient transport and signal transduction (see Supplemental Table 1 and Supplemental Figure 8B online). The only one enriched among the 29 PEGs was the “binding” category. Close inspection of the 29 PEGs revealed this category represented various forms of molecular interactions, such as DNA binding, RNA binding, protein binding, and ATP binding (see Supplemental Table 1 and Supplemental Figure 8C online). Although the exact functions of the MEGs and PEGs remain largely unknown, the two groups of imprinted genes enriched in distinct functional pathways might play specific roles in regulating endosperm development: Whereas the MEGs may be involved in





**Figure 3.** Experimental Validation of Previously Reported and Newly Identified Imprinted Genes by CAPS Assay.

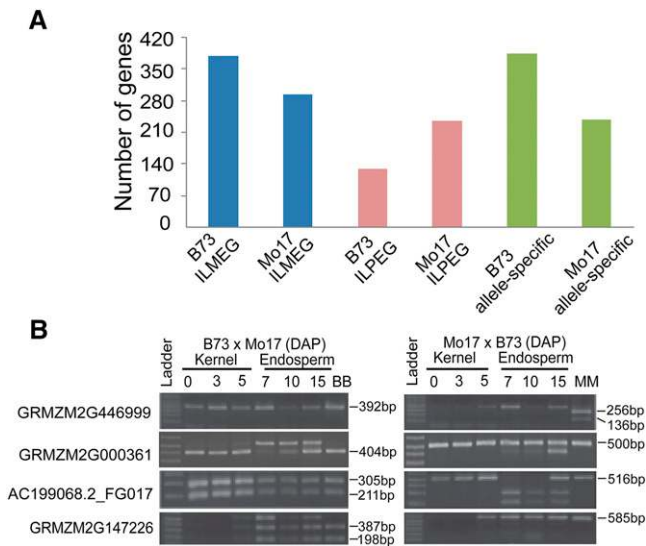
**(A)** Venn diagram analysis of imprinted genes. The imprinted genes identified at 7, 10, and 15 DAP in this study and the imprinted genes previously reported by Waters et al. (2011) and Zhang et al. (2011) are represented in red, blue, green, purple, and orange.

**(B)** Validation of the status of 15 imprinted genes. Group I, seven genes previously identified by either Waters et al. (2011) or Zhang et al. (2011) also found in our data. Two genes (AC208539.3\_FG002 and GRMZM2G449489) did not show any polymorphisms between parental alleles using *Nla*III and *Bgl*II restriction enzymes, respectively, as reported by Zhang et al. (2011), but when using *Hind*III instead, GRMZM2G449489 exhibited parent-specific expression patterns; Group II, eight imprinted genes only identified in our data. BB, B73 × B73; MM, Mo17 × Mo17.

nutrient transport and hormone signaling, the PEGs are likely to be involved in the regulation of gene transcription.

To further determine whether the functions of inbred line-dependent imprinted genes were different from those of imprinted genes, we applied Blast2GO analysis separately on the 669 ILMEGs and 364 ILPEGs. We observed five GO categories of ILMEGs exhibiting significantly higher enrichment compared with the background ( $P \leq 0.01$ ) (see Supplemental Table 2 and Supplemental Figures 8D and 8E online). The “cellular lipid metabolic process” category included 39 ILMEGs that encode for numerous enzymes bound to the cell membrane. The second

category was “inorganic anion trans-membrane transporter activity,” which included eight genes belonging to transporter protein families: an ATP-Binding Cassette transporter, a sulfate transporter, and several phosphate transporter proteins. The third category was the “lactate metabolic process” category, and the other two categories were the “cellular response to extracellular stimulus” and the “regulation of response to stress” categories. These findings are in contrast with the MEG categories, which suggested involvement in “response to organic substance” and “response to endogenous stimulus” (see Supplemental Table 2 and Supplemental Figure 8D online). Thus, the allele-specific



**Figure 4.** Nonimprinted, Allele Specifically Expressed Genes and Inbred Line-Dependent Imprinted Genes.

**(A)** Numbers of inbred line-dependent imprinted genes and nonimprinted allele specifically expressed genes.

**(B)** CAPS validation of four genes showing allele-specific expression or inbred line-dependent imprinting patterns. BB, B73 × B73; MM, Mo17 × Mo17.

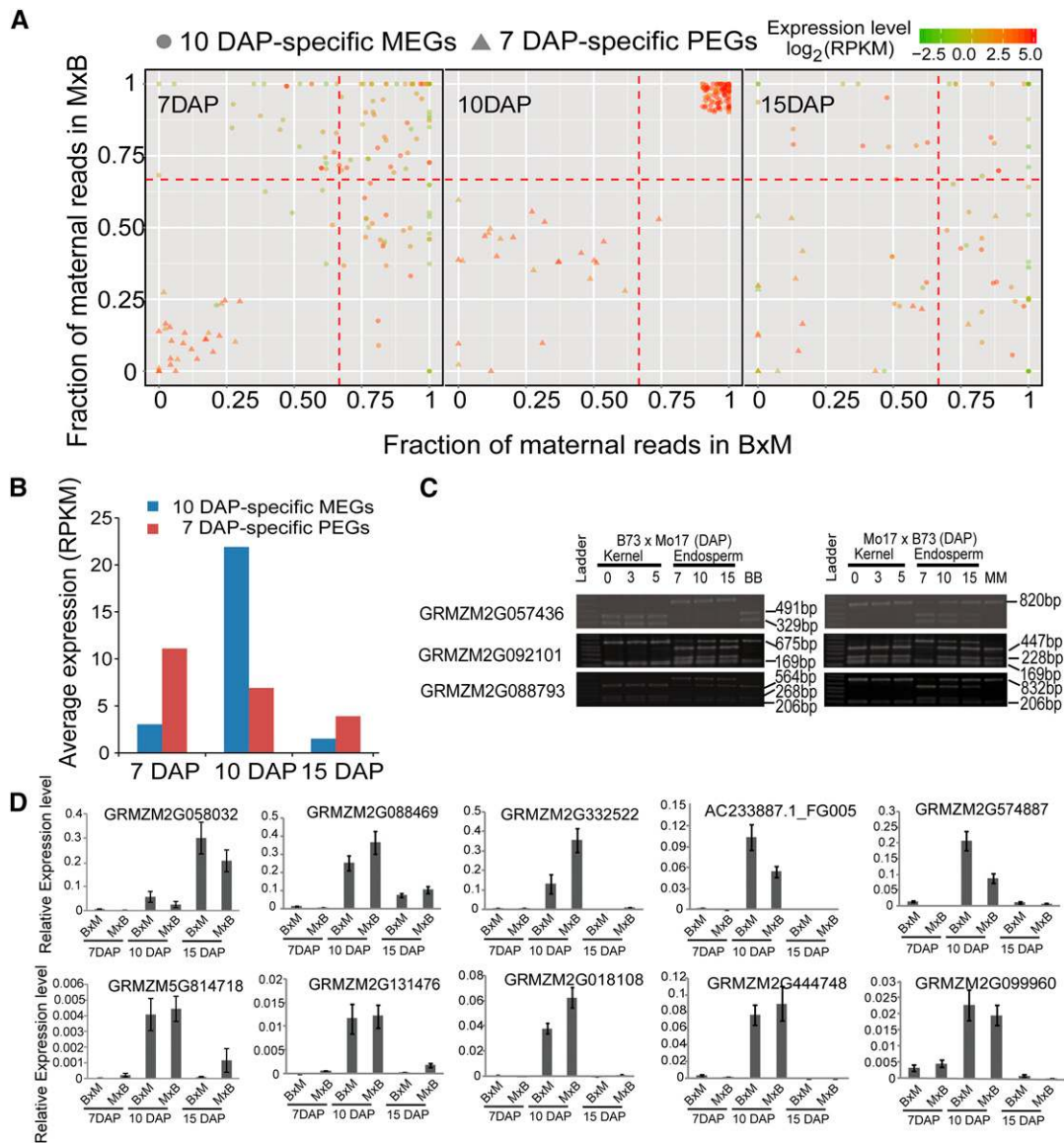
expression of ILMEGs may be mainly involved in the response to environmental cues during endosperm development, similar to the hypothesis put forth by Guo et al. (2004). By contrast, we found that only nine of the 364 ILPEGs had significant enrichment in two categories: Four genes were enriched in the “histone acetyltransferase activity” category, and five genes were enriched in the “membrane lipid biosynthesis” category (see Supplemental Table 2 and Supplemental Figure 8E online). The four genes with histone acetylation activity included three members of the histone acetyltransferase family that are similar to *Arabidopsis* *HISTONE ACETYLTRANSFERASE1* (*HAC1*) and *HAC12*. Both must be recruited to specific promoters to activate the transcription initiation of target genes by acetylating the N terminus of histone H3 or H4. The fourth gene is a homolog of *RELATIVE OF EARLY FLOWERING6* (*REF6*), also known as *JUMONJI DOMAIN-CONTAINING PROTEIN12* (*JMJ12*). The *Arabidopsis* *REF6* gene has been shown to function as a histone H3 Lys 27 demethylase that facilitates removal of Histone H3 lysine 27 trimethylation (H3K27me<sub>3</sub>) during the gene transcription initiation process (Lu et al., 2011). Therefore, our data suggest that allelic expression may control specific cellular or biochemical processes that are essential for endosperm or kernel development.

### Distinct Patterns of MEG and PEG Imprinting and Related Gene Functions Are Associated with Specific Stages of Endosperm Development

We found that MEGs and PEGs exhibited distinct patterns of gene imprinting during endosperm development, with PEGs predominantly detected at 7 DAP and MEGs predominantly detected at 10 DAP (Figure 2). To determine dynamic patterns of gene

imprinting during development, we identified and tracked the patterns of imprinting and levels of expression for individual MEGs and PEGs. Remarkably, of the 194 MEGs identified in the three stages of endosperm development (Figures 2B and 2C), 150 (78.5%) were shown to be expressed maternally in both crosses uniquely at 10 DAP (Figure 5A; hereafter referred to as 10-DAP-specific). By contrast, the 7-DAP endosperm contained the highest number of PEGs (80 out of 96; Figures 2B and 2C), 26 of which were shown to be paternally expressed only at this developmental stage in both crosses (Figure 5A; 7-DAP-specific). Furthermore, we found the occurrence of 7-DAP-specific PEGs were mainly (22 of 26) due to the significantly deviated parental allele ratio, whereas the 10-DAP-specific MEGs could be primarily attributed to the sharply increased expression levels at the respective stages (Figures 5A and 5B). For example, among the latter group, 112 (74.6%) were expressed at least threefold higher at 10 DAP than at 7 and 15 DAP (Figure 5B).

To validate the transient imprinting patterns of the 7-DAP-specific PEGs and 10-DAP-specific MEGs, we performed CAPS and quantitative RT-PCR experiments on three 7-DAP-specific PEGs and nine 10-DAP-specific MEGs (including one contrast gene, GRMZM2G058032, with an expression level that increased from 7 DAP to 10 and 15 DAP), respectively. All the three PEGs (GRMZM2G057436, GRMZM2G092101, and GRMZM2G088793) exhibited paternally biased expression patterns at 7 DAP in both reciprocal crosses, although their maternal alleles were weakly detected, but at later developmental stages, both parental alleles were strongly expressed in Mo17 × B73 cross for GRMZM2G057436 and in reciprocal crosses for GRMZM2G092101 and GRMZM2G088793 (Figure 5C). All nine MEGs tested showed a predominant pattern of expression at 10 DAP with very low to nondetectable levels of expression at 7 or 15 DAP in both reciprocal crosses (Figure 5D). GO analysis of the 150 MEGs confirmed the observed enrichment in MEGs for the “response to organic substance,” “cytoplasmic membrane-bounded vesicles,” and “response to endogenous stimulus” categories were mostly due to the contribution of the 10-DAP-specific MEGs (see Supplemental Figure 9 online). The stage of maize endosperm development beginning around 10 DAP is likely critical for accumulation of nutrients prior to starch and zein storage protein synthesis, and it is likely coincident with the abrupt increase of auxin levels, as described for comparable stages of kernel development (Lur and Setter, 1993; Sabelli and Larkins, 2009). An iodine potassium iodide staining assay indicated the accumulation of starch granules began in 10-DAP endosperm under our specific growth conditions in Arizona (Figure 6A). Previous studies suggested the 10-DAP-specific MEGs identified in our study were likely associated with nutrient synthesis and transport mediated by auxin signaling pathways. For instance, *pin-formed1* (*pin1*; GRMZM2G098643), which encodes an auxin efflux transporter highly expressed in the maize BETL and embryo-surrounding region (ESR) (Forestan et al., 2010), was identified as a 10-DAP-specific MEG (see Supplemental Data Sets 2 and 3 online). Expression of *pin1* is coincident with a rapid auxin increase, which precedes starch and zein protein accumulation (Lur and Setter, 1993; Sabelli and Larkins, 2009). Additionally, among the 10-DAP-specific MEGs, two were identified to encode auxin response factors belonging to the auxin/indole-3-acetic acid family (GRMZM2G317900 and GRMZM2G066219).



**Figure 5.** Development-Dependent, Dynamic Expression of Imprinted Genes.

**(A)** Imprinting status of 7-DAP-specific PEGs and 10-DAP-specific MEGs. The 26 PEGs were detected at 7 DAP but not at 10 and 15 DAP mainly because the parental allele expression ratios of most PEGs (22/26) significantly deviated from 2m:1p only at 7 DAP but not at 10 and 15 DAP, whereas 150 MEGs were only detected in 10-DAP endosperm due to their much higher expression levels compared with 7- and 15-DAP endosperms. RPKM: Reads per kilobase per million.

**(B)** Average expression levels of 7-DAP-specific PEGs and 10-DAP-specific MEGs at the three endosperm developmental stages according to RNA-Seq data.

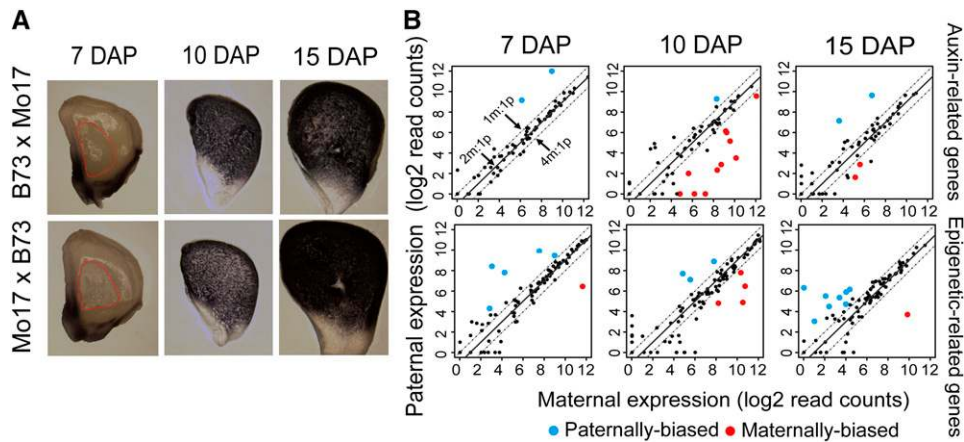
**(C)** Validation of three 7-DAP-specific PEGs in developmental kernels and endosperms. BB, B73  $\times$  B73; MM, Mo17  $\times$  Mo17.

**(D)** Quantitative RT-PCR analysis of nine 10-DAP-specific MEGs. GRMZM2G058032 was used as a control, with expression level increased from 7 to 15 DAP in reciprocal crosses, and the maize thioredoxin gene (gene ID GRMZM2G066612) was used as an internal reference gene. The results indicated that the 10-DAP-specific MEGs were more abundantly expressed at 10 DAP compared with 7 and 15 DAP.

Taken together, these data suggest auxin-mediated nutrient uptake from maternal tissues to the endosperm is under control of the maternal genome.

To address this hypothesis using available data and further understand the nature of imprinted gene functions, we selected

82 SNP-containing genes that are potentially involved in auxin biosynthesis, transport, response, and ubiquitination pathways (auxin-related) based on Maize Genetics and Genomics Database annotations. We used a less stringent set of criteria, namely  $\chi^2$  ( $P \leq 0.05$ ) with 4m:1p and 1m:1p ratio cutoffs, to examine whether



**Figure 6.** Maternally Biased Expression of Auxin-Related Genes and Paternally Biased Expression of Epigenetic-Related Genes.

**(A)** Iodine potassium iodide staining analysis of the 7-, 10-, and 15-DAP endosperms from reciprocal crosses. The results indicated that starch accumulation began at 10 DAP. Micropylar region of the kernel is located in the lower left-hand side of each panel.

**(B)** Maternally biased expression of auxin-related genes and paternally biased expression of epigenetic-related genes. Top panel: 11 of 82 auxin-related genes exhibit maternally biased expression at 10 DAP with a  $\chi^2$  test  $\leq 0.05$  and a parental expression ratio of over 4m:1p in reciprocal crosses. Bottom panel: Eight of 111 epigenetic-related genes exhibit paternally biased expression at 15 DAP with a  $\chi^2$  test  $\leq 0.05$  and a parental expression ratio of less than 1m:1p in reciprocal crosses.

the expression dosages of their parental alleles deviated from the expected 2m:1p ratio. We then plotted the maternal and paternal expression ratios separately for genes expressed at 7, 10, and 15 DAP (Figure 6B). At 7 DAP, 80 genes were centered on the diagonal line of 2m:1p ratio, except for two genes exhibiting paternal bias. At 10 DAP, 11 genes exhibited maternally biased expression, including four that had been classified as 10-DAP-specific MEGs; only one 10-DAP gene exhibited paternal bias (Figure 6B). At 15 DAP, the number of maternally biased auxin-related genes decreased to two genes (Figure 6B). This indicates a preponderance of auxin-related genes among maternally biased genes expressed at 10 DAP.

Also, because we detected certain PEG functions to be associated with epigenetic regulation during endosperm development, we performed a similar analysis of the 111 SNP-containing epigenetic-related genes collected from ChromDB (Gendler et al., 2008). We found that five, three, and eight genes exhibited paternally biased expression at 7, 10, and 15 DAP, respectively, whereas one, four, and one genes exhibited maternally biased expression at the same three stages (Figure 6B). Interestingly, in contrast with the 10-DAP-prevalent pattern of maternally biased auxin-related genes, the paternally biased epigenetic-related genes were instead more prevalent in the 7- and 15-DAP endosperm. Together, these results suggest that the maternally biased, auxin-related genes and the paternally biased, epigenetic-related genes have coordinated and yet complementary expression patterns during maize endosperm development.

## DISCUSSION

### The Number of Imprinted Genes in Maize Endosperm Remains Largely Underestimated

In this study, we performed a genome-wide survey of imprinted genes in hybrid maize endosperm during the early developmental

stages of 7, 10, and 15 DAP. Our analyses revealed that 799 genes exhibited allele-biased, parent-specific expression patterns based on a  $\chi^2$  test ( $P \leq 0.001$ ) (see Supplemental Figure 7 online). Of these, 290, including 194 MEGs and 96 PEGs (Figures 2B, 2C, and 3A), were defined as imprinted genes based on the more stringent ratio-based cutoff. Similar work by Zhang et al. (2011) and Waters et al. (2011) with 10- and 14-DAP maize endosperm (defined under local growth conditions) identified 179 and 100 imprinted genes (223 genes in total), respectively. Of these, only 56 (25.1%) genes were common to both studies (Waters et al., 2011; Zhang et al., 2011) (Figure 3A). We identified 81 of the 223 imprinted genes reported by Waters et al. (2011) and Zhang et al. (2011) (Figure 3A). The remaining 142 genes not documented in our study were excluded by our preset conditions: 36 genes contained gaps longer than 100 bp in the coding regions, 47 contained more than five paternally derived reads at 0 DAP, five had no SNPs, 19 contained  $<40$  SNP-associated reads between the B73 and Mo17 alleles, and the remaining 35 genes did not satisfy the imprinted expression criteria used in our analysis. Additionally, the inclusive SNP information based on the Mo17 whole-genome assembly and the three developmental stages of endosperm allowed us to identify 209 more imprinted genes in maize endosperm compared with the two previous studies (Waters et al., 2011; Zhang et al., 2011). Among our 290 imprinted genes, only 11 MEGs and 27 PEGs exhibited a persistent imprinted expression pattern in the 7-, 10-, and 15-DAP endosperm stages (Figure 3A; see Supplemental Data Set 2 online). The persistent MEGs included *mez1* and *fie1*, which encode components of PRC2 (Springer et al., 2002; Danilevskaya et al., 2003). The persistent PEGs included *yuc1*, which encodes a flavin monooxygenase that catalyzes auxin biosynthesis; its mutants exhibit a small-sized kernel, retarded endosperm endoreduplication, and reduced dry matter in the endosperm (Bernardi et al., 2012). Of the 290 imprinted genes,

77% (150 of 194) are MEGs and 35% (26 of 96) PEGs specific to 10- and 7-DAP endosperm, respectively, indicating the majority of the imprinted genes are expressed uniquely from a given allele only during specific developmental stages (Figures 2 and 5). The dynamic expression patterns of the maize imprinted genes suggest genomic imprinting is likely associated with specific processes requiring distinct contributions from the parental genomes during early endosperm development (see below).

Our data suggest that assessing a limited number of maize endosperm developmental stages can lead to overlooking imprinted genes because of stage-specific timing of transcriptional activation as well as potential posttranscriptional processes that impinge on allele mRNA stability. Additionally, the identification of imprinted genes in B73 and Mo17 endosperm based on RNA-Seq analysis could miss alleles that lack polymorphic nucleotides, which is true for nearly 40% of them (see Supplemental Figure 1 online). Some previously known maize imprinted genes were not observed among the 290 genes reported here; for example, *meg1* was not identified in our study due to the absence of SNPs between the B73 and Mo17 alleles (Gutiérrez-Marcos et al., 2004). Consequently, it is reasonable to hypothesize that the number of imprinted genes in the endosperm of plants is significantly underestimated due to the technical issues associated with detection and the inherently dynamic nature of gene imprinting in angiosperms.

#### Paternal Genome Activation Is Near Fully Achieved in the Endosperm by 7 DAP

Our knowledge of the timing and extent of postfertilization activation of the maternal and paternal genomes in endosperm is largely incomplete. Studies of *Arabidopsis* and maize seed suggested the activation of paternal alleles could be relatively delayed in the embryo, although contrary evidence indicates the two sets of parental alleles are activated simultaneously (Vielle-Calzada et al., 2000; Weijers et al., 2001; Autran et al., 2011; Nodine and Bartel, 2012), and the available evidence suggests that activation of paternal alleles is delayed in the endosperm (Vielle-Calzada et al., 2000; Grimanelli et al., 2005). Our RNA-Seq data allowed us to investigate the extent of paternal genome activation even with the compound structure of the kernel. At 3 DAP, ~941 genes were shown to produce paternal transcripts, albeit at lower abundances than the maternal transcripts. Of the 941 genes, 923 ( $\geq 98\%$ ) were also expressed at 0 DAP, suggesting the apparent expression bias may not necessarily be a reflection of a delay in paternal genome activation but rather due to expression or presence of the transcripts in the maternal kernel tissues (e.g., nucellus and pericarp). Nevertheless, it is reasonable to conclude that paternal genome activation was nearly fully achieved by 7 DAP or earlier because the majority of the 11,027 genes with SNPs were shown to be biallelically expressed (Figure 1A) and the proportion of the expressed genes with the expected 2m:1p ratio were similar in the 7- and 10-DAP endosperm samples.

#### Potential Antagonistic Functions of MEGs and PEGs in Regulating Gene Expression

Among the paternally activated genes expressed at 3 DAP, 14 genes were shown to exhibit uniparental expression (Figure 2D).

Interestingly, one of the PEGs (GRMZM2G374169) encoding a RNA binding protein similar to *Arabidopsis* *LHP1-INTERACTING FACTOR2* (*LIF2*) was found to be persistently imprinted from 3 to 15 DAP. *LIF2* is an RNA binding partner of the LHP1 protein in the PCR1-like complex (Latrasse et al., 2011). In the *lif2* mutant, the level of the repressive H3K27me3 chromatin mark at the *FLOWERING LOCUS C* locus was increased compared with the wild type, indicating that *LIF2* may negatively control the deposition of H3K27me3 (Latrasse et al., 2011). This finding is reminiscent of the two known MEGs, *fie1* and *mez1*, which are specifically expressed in maize endosperm and are similar to the *Arabidopsis* *FIE* and *MEA* genes. *FIE* and *MEA* are components of the PRC2 complex that has been shown to silence the paternal *MEA* allele and other genes by depositing H3K27me3 marks (Köhler et al., 2003; Gehring et al., 2006; Jullien et al., 2006; Huh et al., 2008). Therefore, the paternally expressed maize *lif2* and the maternally expressed *fie1* and *mez1*, which act by depositing or removing H3K27 mark, may exemplify the opposing molecular functions of PEGs and MEGs during seed development.

#### Interplay between PEGs and MEGs in Auxin-Mediated Regulation of Nutrient Allocation in the Endosperm

Our analyses identified imprinted genes that are involved in a wide range of cellular functions, including transcriptional activation, chromatin remodeling, cross-membrane molecule transport, signal transduction, hormone responses, protein posttranslational modification, cell wall formation, and various enzymatic activities in starch metabolic pathways. GO analyses also indicated that MEGs and PEGs are enriched in distinct functional categories, although a clear antagonistic relationship was not observed between MEG and PEG functions (see Supplemental Table 1 online). In mammals, some MEGs (including *H19*, *Igf2r*, *Phlda2*, *Grb10*, and *Cdkn1c*) appear to reduce nutrient supply to the embryo by decreasing efficiency of allocation (Wang et al., 1994; Leighton et al., 1995; Takahashi et al., 2000; Charalambous et al., 2003; Salas et al., 2004), whereas several PEGs (including *Igf2*, *Igf2P0*, and *peg1*) act to increase nutrient transport to the fetus by maximizing placental efficiency (Baker et al., 1993; Lefebvre et al., 1998; Constância et al., 2002), which seems to support the parental conflict theory. In plants, the only imprinted gene known to function in nutrient supply is maize *meg1*, which is preferentially expressed in the BETL of 10-DAP endosperm; it is imprinted early (before 12 DAP) but is biallelically expressed later (at and after 12 DAP) (Gutiérrez-Marcos et al., 2004; Costa et al., 2012). As a MEG, *meg1* is expected to restrict maternal allocation to the seed; however, in a recent study (by knocking down activity and varying the *meg1* gene dosage with imprinted and nonimprinted transgenes), *meg1* in fact was shown to positively regulate transfer tissue development and function, increasing nutrient uptake of the seed and producing a larger endosperm and seed (Costa et al., 2012). This may imply the need for reconsidering the parental conflict model for the adaptive evolution of gene imprinting (Jullien and Berger, 2010; Costa et al., 2012; Jiang and Köhler, 2012; Li and Berger, 2012). Alternatively, *meg1* and similar MEGs (and PEGs) may perform different roles at distinct stages of endosperm development related either to processes required for establishment of structures or cellular and biochemical processes that control storage or assimilation.

Our GO analysis of gene functions for imprinted genes specific to 10 DAP identified five genes (GRMZM2G030465, GRMZM2G121468, GRMZM2G317900, GRMZM2G098643, and GRMZM2G043191) involved in the auxin signaling pathway (see Supplemental Figure 9 and Supplemental Table 1 online). This is the stage at which the majority of MEGs were detected (150 out of 194) (Figure 3A) and is coincident with the onset of nutrient accumulation for starch and zein storage protein synthesis (Lur and Setter, 1993; Sabelli and Larkins, 2009) and a rapid increase in auxin accumulation (Lur and Setter, 1993). A previous study showed auxin accumulation occurs mainly in the BETL, ESR, and aleurone regions of maize endosperm (Forestan et al., 2010), which are involved in mediating interactions between the endosperm and the embryo, and the endosperm and maternal tissues. This relationship is reflected by two significantly enriched functional categories of 10-DAP-specific MEGs: the “cytoplasmic membrane-bound vesicle” and the “auxin-mediated hormone response” (see Supplemental Figure 9 online). The former GO category includes numerous genes encoding transmembrane transport proteins involved in the inter- and intracellular distribution of large substances such as proteins, and certain amino acids and polysaccharides. The latter group includes three auxin-responsive factors (GRMZM2G066219, GRMZM2G317900, and GRMZM2G030465) with uncharacterized functions in maize endosperm. In addition, the MEG-encoded auxin efflux carrier protein *pin1* (GRMZM2G098643), found to be expressed predominantly at 10 DAP in our analysis, was shown to be highly expressed in the BETL and ESR compartments of the early developing maize endosperm (Forestan et al., 2010). Expression of *pin1* could not be detected in the maize *defective endosperm-B18* (*de\*-B18*) mutant, which has an endosperm phenotype characterized by decreased starch and protein accumulation (Bernardi et al., 2012). The *de\*-B18* mutant was recently characterized as a loss-of-function mutation in *Zm Yuc1*, which encodes a flavin monooxygenase essentially involved in a critical step to catalyze the conversion of indole-3-pyruvic to indole-3-acetic acid (Bernardi et al., 2012). Our data indicate *yuc1* is a PEG expressed as early as 5 DAP, reaching its peak abundance at 7 and 10 DAP and decreasing at 15 DAP (see Supplemental Data Set 2 online). Moreover, the homologs of *yuc1* in rice (*Os12g08780*) and *Arabidopsis* (AT1G48910) were found to have paternal expression only (Hsieh et al., 2011; Luo et al., 2011), suggesting that paternal expression of *yuc1* in plants could play a critical role in endosperm development or function. Another PEG expressed at 7, 10, and 15 DAP and involved in the auxin signaling pathway is GRMZM2G037368 (see Supplemental Data Set 2 online), which encodes an auxin/indole-3-acetic acid transcription factor highly similar to tomato (*Solanum lycopersicum*) *IAA27*. Although no endosperm or seed phenotype is described, downregulation of this gene in tomato can increase sensitivity to auxin and produce reduced fertility and small-sized fruit (Bassa et al., 2012). A third PEG that can interact with the auxin signaling pathway is GRMZM2G084462, which is expressed primarily at 7 and 10 DAP (see Supplemental Data Set 2 online). This PEG is similar to the *Arabidopsis* gene, *ISOPENYNYL TRANSFERASE3* (AT3G63110), which encodes a cytokinin synthase. High concentrations of cytokinin can promote the bisymmetric distribution of *PIN1* proteins and thereby activate polar transport of auxin to central regions in the root vasculature (Bishopp et al., 2011). Although the specific role of auxin

and auxin-signaling pathways during early endosperm development is unknown, the results of our analysis of endosperm imprinted genes and their associated GO functions highlight a potential role for auxin-mediated control of nutrient uptake from maternal tissues through a complementary set of functions performed by distinct MEGs and PEGs. As such, this study supports the recent observations of the effects of parental genome excess on the timing of gene expression during the same developmental period in inter-ploidy maize crosses (Li and Dickinson, 2010). Our data indicate potential interplay between the relatively rapid upregulation of MEGs involved in nutrient accumulation at 10 DAP and the concomitant downregulation of an earlier expressed set of PEGs with potential cell proliferation regulatory functions. They suggest that a coordinated set of regulatory programs control imprinting of genes that are required for proper control of endosperm development to prepare it as a functional absorptive tissue.

## METHODS

### Plant Materials

The maize (*Zea mays*) inbred lines B73 and Mo17 were grown under greenhouse conditions (16 h day/8 h night, 30°C/25°C) at the University of Arizona during March to May 2011. Reciprocal crosses and self-pollination were performed as follows: Ears were bagged before silking; when the silk grew to ~2- to 3-cm long, the ears were cut 2 cm from the top and tassels were bagged. Pollination was conducted the next day using appropriate pollen. Unpollinated kernels (0 DAP), kernels at 3 and 5 DAP, and endosperms from 7, 10, and 15 DAP were obtained by hand dissection. Endosperm and kernels were collected from at least three different ears to create three biological replicates and were immediately frozen in liquid nitrogen.

### RNA Extraction

Total RNA was isolated from 36 (12 samples × three replicates) groups of plant materials using an SDS-phenol method (Shirzadegan et al., 1991) with the following modification: ~1 g of tissue was ground to a fine powder in liquid nitrogen and homogenized with 6 mL of buffer containing 1% SDS, 50 mM Tris-HCl, pH 8.0, 150 mM lithium chloride (LiCl), 5 mM EDTA, and 10 mM DTT. The homogenate was mixed with 6 mL phenol:chloroform (5:1, pH 4.5; Ambion AM9720) and incubated on ice for 5 min. The mixture was transferred to a 50-mL Phase-Lock-Gel tube (5 Prime) and centrifuged at 5000 rpm (SS-34 rotor) for 10 min at 4°C. The aqueous phase was transferred to a new 50-mL Phase-Lock-Gel tube. The previous step was repeated with phenol:chloroform (1:1) and chloroform alone. The RNA was precipitated with 2.5 M LiCl on ice overnight, washed with ice-cold 2 M LiCl, dissolved in Trypsin/EDTA, and mixed with a one-ninth volume of 3 M NaAC, pH 5.2, and 2.5 volumes of ethanol. The mixture was held in a -80°C freezer for at least 4 h, after which the RNA was pelleted by centrifugation at 14,000 rpm for 15 min at 4°C. After rinsing with 75% ethanol and air-drying, the RNA was dissolved in diethylpyrocarbonate-treated water. DNA was removed by TURBO DNase I (Ambion) and purified using an RNeasy column (Qiagen).

### RNA-Seq Library Preparation

The standard protocols provided by the Beijing Genomics Institute (BGI) for Illumina sequencing were used to prepare the RNA-Seq libraries. First, the poly(A)-containing mRNA molecules were purified from the total RNAs using ~10 µg total RNA using poly-T oligo-attached magnetic beads. Second, the extracted mRNAs were fragmented into ~200-bp-long pieces using divalent cations under elevated temperature. Third, the cleaved

mRNA fragments were copied into first-strand cDNA using reverse transcriptase and random primers, followed by second-strand cDNA synthesis using DNA Polymerase I and RNase H. Then, the cDNA fragments went through an end repair process, the addition of a single “A” base, and ligation of the adapter sequences. Finally, the cDNA products were purified and enriched with PCR assay to construct libraries for non-strand-specific RNA-Seq.

### Illumina Sequencing

RNA samples were sent to the BGI for mRNA library construction and high-throughput sequencing using the Illumina HiSeq2000 platform. Before library construction, quality of the 36 RNA samples was examined using an Agilent 2100 bioanalyzer. The three biological replicates for each sample were of high quality and were thus combined to construct one mRNA library for RNA-Seq based on BGI’s standard protocol. Between 55 and 65 million 90-nucleotide RNA-Seq reads were generated for each sample to ensure 5.5- to 6.5-Gb coverage of the transcriptomes for each sample. FastQC software was used to examine the sequencing quality of the reads in each sample (<http://www.bioinformatics.babraham.ac.uk>). The RNA-Seq reads used for this study are deposited at the National Center for Biotechnology Information Short Read Archive under accession number GSE48425.

### Mo17 Genome Assembly and SNP Identification

We performed a reference-guided assembly of the Mo17 genome based on the B73 genome (release 5b.60; <http://maizesequence.org>), using 650-Gb Illumina short reads obtained from the Joint Genome Institute, Department of Energy (SRA accession no. SRP003567). The sequence quality of the short reads was first examined by the FastQC tool (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Then, the short reads were aligned to the B73 reference genome sequence using Bowtie2 with the following parameters: `-score-min L,-15,-0.4 -3 26 -very-sensitive`. The alignment results were imported into the SHORE pipeline (Schneeberger et al., 2009) to identify the SNPs and INDELS between the B73 and Mo17 genomes. Finally, we used a customized Perl script to generate the consensus sequences (contigs) based on the alignment of Mo17 reads to the B73 genome and connect the Mo17 contigs to form pseudo-chromosome sequences based on their positions in the B73 genome. The chromosomal coordinates of the nucleotides in the Mo17 assembly were adjusted based on the sizes of deletions or insertions occurred between the two genomes. Moreover, if a relatively long genomic portion in the B73 genome was not covered by Mo17 reads, the same number of “N” was filled in the Mo17 assembly to indicate the existence of a gap. The corresponding chromosomal coordinates of the SNPs and INDELS in the two genomes were correlated for the differentiation of allele-specific reads. The Mo17 assembly and the SNP information between the B73 and Mo17 genomes are available via MaizeGDB (Sen et al., 2009; <http://www.maizegdb.org>) through front page links to “Data Downloads” with direct linkage available at [http://ftp.maizegdb.org/MaizeGDB/FTP/Mo17/2013\\_XinEtAl\\_PlantCell](http://ftp.maizegdb.org/MaizeGDB/FTP/Mo17/2013_XinEtAl_PlantCell). Alignments of the Mo17 assembly to the B73 reference genome (including representations of associated SNPs) within the context of the MaizeGDB genome browser are planned for release in November of 2013.

### Alignment of RNA-Seq Reads to B73 and Mo17 Genomes

The Illumina sequencing experiment generated on average 63 million paired-end RNA-Seq reads for each sample. The RNA-Seq reads were separately mapped to the B73 and Mo17 genomes using the Tophat2 with parameters `-bowtie1 -r 200 -l 5000 -mate-std-dev 100 -g 5 -max-segment-intron 5000 -n 3 -genome-read-mismatch 4 -read-mismatches 4` (Kim et al., 2013). Reads with no more than five mapped positions were

retained for further analysis, and ~95% of the mapped reads had unique positions in the B73 and Mo17 genomes. The read mapping statistics of the 12 samples on the B73 and Mo17 genomes is shown in Supplemental Table 3 online. To count the number of reads mapped to a gene-encoding region based on the maize genome annotation (release 5b.60), we used the `intersectBed` command in the BEDTools package (<https://code.google.com/p/bedtools/>). The expression levels of the ~110,000 annotated genes in the maize WGS across all of the 12 samples were normalized to a baseline read count of one million reads using Trimmed Mean of M-values normalization algorithm in the edgeR package (Robinson et al., 2010).

### Quantitative RT-PCR Validation

DNase I–treated total RNAs were reverse transcribed with oligo(dT) primers using a RETROscript kit (Ambion), following the manufacturer’s instructions. Quantitative RT-PCR was performed using a LightCycler 1.5 instrument in a 32-capillary format (Roche) with the same conditions as previously reported (Wang et al., 2010). Cycle threshold values were calculated using the standard approach provided in the LightCycler software 4.0 (Roche). Maize thioredoxin (Zm thioredoxin; gene ID GRMZM2G066612) was used as an internal reference gene. Each quantitative RT-PCR reaction was run in a 10- $\mu$ L reaction volume using an aliquot of master mix from the LightCycler FastStart DNA MasterPLUS SYBR Green I kit according to manufacturer’s instructions (Roche).

### CAPS

Experimental validation of imprinted expression patterns of selected genes was performed using CAPS assays as previous described (Konieczny and Ausubel, 1993). RT-PCR was performed using the primers listed in Supplemental Table 4 online. Amplification products were digested with the restriction enzymes listed in Supplemental Table 3 online to differentiate the transcripts derived from the B73 or Mo17 alleles.

### Accession Numbers

The raw RNA-Seq data and normalized expression data of the ~110,000 WGS genes were deposited to the National Center for Biotechnology Information Gene Expression Omnibus under accession number GSE48425. The Mo17 assembly and the SNP information identified between the B73 and Mo17 genome are available at [http://ftp.maizegdb.org/MaizeGDB/FTP/Mo17/2013\\_XinEtAl\\_PlantCell/](http://ftp.maizegdb.org/MaizeGDB/FTP/Mo17/2013_XinEtAl_PlantCell/).

### Supplemental Data

The following materials are available in the online version of this article.

**Supplemental Figure 1.** SNP Distribution per Gene.

**Supplemental Figure 2.** Adjustment of Parental Gene Expression Ratios Based on the Gap Length between B73 and Mo17 Genomes.

**Supplemental Figure 3.** Parental Expression Ratios Plotted Separately for Each B73 and Mo17 Cross.

**Supplemental Figure 4.** Allele-Specific Expression Inferred from the SNP-Associated Reads Is Proportional to the Total Expression Level of an Imprinted Gene.

**Supplemental Figure 5.** Paternal Genome Activation in the Mo17  $\times$  B73 (MB) Kernel Is Faster Than in the B73  $\times$  Mo17 (BM) Kernel.

**Supplemental Figure 6.** GO Comparison of Paternally Activated Genes at 3 and 5 DAP in the Reciprocal Crosses.

**Supplemental Figure 7.** Identification of Genes Showing Maternally and Paternally Biased Expression Using the  $\chi^2$  Test.

**Supplemental Figure 8.** GO Enrichment Analysis of Imprinted Genes.

**Supplemental Figure 9.** GO Analysis of 10-DAP-Specific MEGs.

**Supplemental Table 1.** GO Categories Enriched with Imprinted Genes.

**Supplemental Table 2.** GO Categories Enriched with Inbred Line-Dependent Imprinted Genes.

**Supplemental Table 3.** Summary of the Alignment of RNA-Seq Reads to the B73 and Mo17 Genomes.

**Supplemental Table 4.** Primers and Enzymes Used for qRT-PCR and CAPS Assays.

**Supplemental Methods 1.** Detailed Information of Gap Size Adjustment and Comparison of Three Imprinted Gene Sets.

**Supplemental Data Set 1.** Allele-Specific Expression of the 11,027, 10,573, and 7777 SNP-Containing Genes in 7-, 10-, and 15-DAP Endosperm.

**Supplemental Data Set 2.** Allele-Specific Expression of the 290 Imprinted Genes.

**Supplemental Data Set 3.** Functional Annotation of the 290 Imprinted Genes.

## ACKNOWLEDGMENTS

This work was supported by National Key Scientific Program (2012CB910900), the National Natural Science Foundation of China (30925023 and 31230054), and U.S. National Science Foundation Grant IOS-0923880 (to B.A.L. and R.Y.). MaizeGDB's representation of the Mo17 assembly data is supported by the USDA-Agricultural Research Service. Maize Mo17 shotgun sequence data were produced by the U.S. Department of Energy Joint Genome Institute under contract from the U.S. Department of Energy and the USDA.

## AUTHOR CONTRIBUTIONS

Z.N., X.W., R.Y., and Q.S. conceived the project. M.X., G.L., J.L., and D.W. collected the plant materials. M.X., R.Y., Y.Y., H.C., and C.M. analyzed data. M.X., X.W., R.Y., B.A.L., and Z.N. wrote the article.

Received June 28, 2013; revised August 23, 2013; accepted August 30, 2013; published September 20, 2013.

## REFERENCES

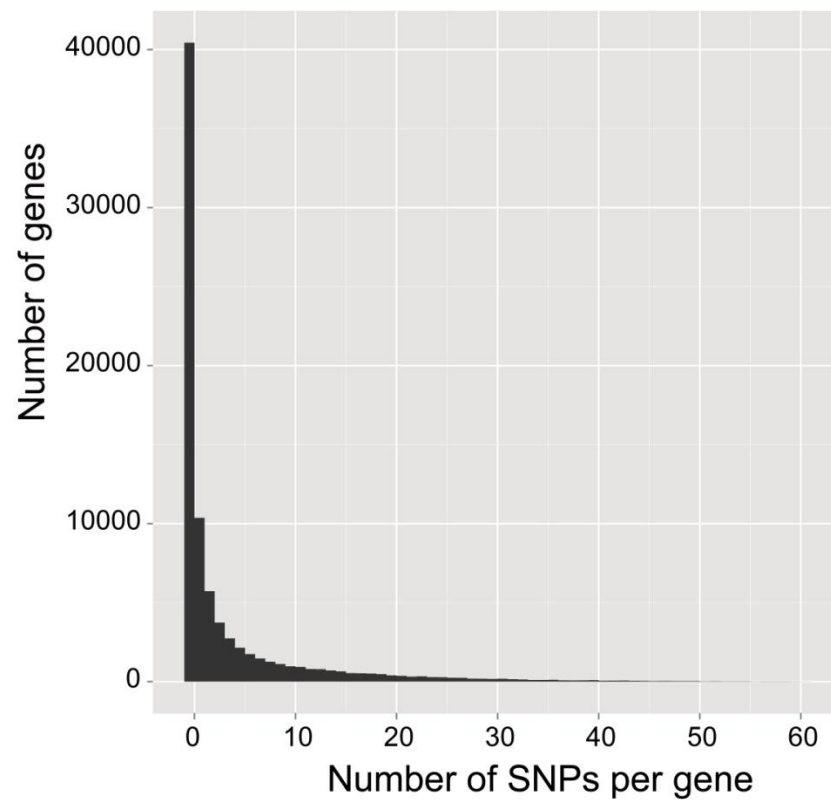
- Alvarez, J.P., Goldshmidt, A., Efroni, I., Bowman, J.L., and Eshed, Y.** (2009). The NGATHA distal organ development genes are essential for style specification in *Arabidopsis*. *Plant Cell* **21**: 1373–1393.
- Autran, D., et al.** (2011). Maternal epigenetic pathways control parental contributions to *Arabidopsis* early embryogenesis. *Cell* **145**: 707–719.
- Baker, J., Liu, J.P., Robertson, E.J., and Efstratiadis, A.** (1993). Role of insulin-like growth factors in embryonic and postnatal growth. *Cell* **75**: 73–82.
- Bassa, C., Mila, I., Bouzayen, M., and Audran-Delalande, C.** (2012). Phenotypes associated with down-regulation of SI-IAA27 support functional diversity among Aux/IAA family members in tomato. *Plant Cell Physiol.* **53**: 1583–1595.
- Berger, F.** (2003). Endosperm: The crossroad of seed development. *Curr. Opin. Plant Biol.* **6**: 42–50.
- Berger, F., Grini, P.E., and Schnittger, A.** (2006). Endosperm: An integrator of seed growth and development. *Curr. Opin. Plant Biol.* **9**: 664–670.
- Bernardi, J., Lanubile, A., Li, Q.B., Kumar, D., Kladnik, A., Cook, S. D., Ross, J.J., Marocco, A., and Chourey, P.S.** (2012). Impaired auxin biosynthesis in the defective endosperm18 mutant is due to mutational loss of expression in the ZmYuc1 gene encoding endosperm-specific YUCCA1 protein in maize. *Plant Physiol.* **160**: 1318–1328.
- Bishopp, A., Help, H., El-Showk, S., Weijers, D., Scheres, B., Friml, J., Benková, E., Mähönen, A.P., and Helariutta, Y.** (2011). A mutually inhibitory interaction between auxin and cytokinin specifies vascular pattern in roots. *Curr. Biol.* **21**: 917–926.
- Brown, R.C., Lemmon, B.E., Nguyen, H., and Olsen, O.A.** (1999). Development of endosperm in *Arabidopsis thaliana*. *Sex. Plant Reprod.* **12**: 32–42.
- Charalambous, M., Smith, F.M., Bennett, W.R., Crew, T.E., Mackenzie, F., and Ward, A.** (2003). Disruption of the imprinted Grb10 gene leads to disproportionate overgrowth by an Igf2-independent mechanism. *Proc. Natl. Acad. Sci. USA* **100**: 8292–8297.
- Chaudhury, A.M., Ming, L., Miller, C., Craig, S., Dennis, E.S., and Peacock, W.J.** (1997). Fertilization-independent seed development in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA* **94**: 4223–4228.
- Conesa, A., and Götz, S.** (2008). Blast2GO: A comprehensive suite for functional analysis in plant genomics. *Int. J. Plant Genomics* **2008**: 619832.
- Conesa, A., Götz, S., García-Gómez, J.M., Terol, J., Talón, M., and Robles, M.** (2005). Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**: 3674–3676.
- Constância, M., Hemberger, M., Hughes, J., Dean, W., Ferguson-Smith, A., Fundele, R., Stewart, F., Kelsey, G., Fowden, A., Sibley, C., and Reik, W.** (2002). Placental-specific IGF-II is a major modulator of placental and fetal growth. *Nature* **417**: 945–948.
- Costa, L.M., Yuan, J., Rouster, J., Paul, W., Dickinson, H., and Gutierrez-Marcos, J.F.** (2012). Maternal control of nutrient allocation in plant seeds by genomic imprinting. *Curr. Biol.* **22**: 160–165.
- Danilevskaya, O.N., Hermon, P., Hantke, S., Muszynski, M.G., Kollipara, K., and Ananiev, E.V.** (2003). Duplicated fie genes in maize: Expression pattern and imprinting suggest distinct functions. *Plant Cell* **15**: 425–438.
- Forestan, C., Meda, S., and Varotto, S.** (2010). ZmPIN1-mediated auxin transport is related to cellular differentiation during maize embryogenesis and endosperm development. *Plant Physiol.* **152**: 1373–1390.
- Gehring, M., Choi, Y., and Fischer, R.L.** (2004). Imprinting and seed development. *Plant Cell* **16** (suppl.): S203–S213.
- Gehring, M., Huh, J.H., Hsieh, T.F., Penterman, J., Choi, Y., Harada, J.J., Goldberg, R.B., and Fischer, R.L.** (2006). DEMETER DNA glycosylase establishes MEDEA polycomb gene self-imprinting by allele-specific demethylation. *Cell* **124**: 495–506.
- Gehring, M., Missirian, V., and Henikoff, S.** (2011). Genomic analysis of parent-of-origin allelic expression in *Arabidopsis thaliana* seeds. *PLoS ONE* **6**: e23687.
- Gendler, K., Paulsen, T., and Napoli, C.** (2008). ChromDB: The chromatin database. *Nucleic Acids Res.* **36** (Database issue): D298–D302.
- Götz, S., Arnold, R., Sebastián-León, P., Martín-Rodríguez, S., Tischler, P., Jehl, M.A., Dopazo, J., Rattei, T., and Conesa, A.**



- (2011). B2G-FAR, a species-centered GO annotation repository. *Bioinformatics* **27**: 919–924.
- Götz, S., García-Gómez, J.M., Terol, J., Williams, T.D., Nagaraj, S.H., Nueda, M.J., Robles, M., Talón, M., Dopazo, J., and Conesa, A.** (2008). High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res.* **36**: 3420–3435.
- Grimanelli, D., Perotti, E., Ramirez, J., and Leblanc, O.** (2005). Timing of the maternal-to-zygotic transition during early seed development in maize. *Plant Cell* **17**: 1061–1072.
- Grossniklaus, U., Vielle-Calzada, J.P., Hoepfner, M.A., and Gagliano, W.B.** (1998). Maternal control of embryogenesis by MEDEA, a polycomb group gene in *Arabidopsis*. *Science* **280**: 446–450.
- Guitton, A.E., and Berger, F.** (2005). Control of reproduction by Polycomb Group complexes in animals and plants. *Int. J. Dev. Biol.* **49**: 707–716.
- Guo, M., Rupe, M.A., Zinselmeier, C., Habben, J., Bowen, B.A., and Smith, O.S.** (2004). Allelic variation of gene expression in maize hybrids. *Plant Cell* **16**: 1707–1716.
- Guo, M., Yang, S., Rupe, M., Hu, B., Bickel, D.R., Arthur, L., and Smith, O.** (2008). Genome-wide allele-specific expression analysis using Massively Parallel Signature Sequencing (MPSS) reveals cis- and trans-effects on gene expression in maize hybrid meristem tissue. *Plant Mol. Biol.* **66**: 551–563.
- Gutiérrez-Marcos, J.F., Costa, L.M., Biderre-Petit, C., Khbaya, B., O’Sullivan, D.M., Wormald, M., Perez, P., and Dickinson, H.G.** (2004). Maternally expressed gene1 is a novel maize endosperm transfer cell-specific gene with a maternal parent-of-origin pattern of expression. *Plant Cell* **16**: 1288–1301.
- Haig, D., and Westoby, M.** (1989). Parent-specific gene expression and the triploid endosperm. *Am. Nat.* **134**: 147–155.
- Hsieh, T.F., Shin, J., Uzawa, R., Silva, P., Cohen, S., Bauer, M.J., Hashimoto, M., Kirkbride, R.C., Harada, J.J., Zilberman, D., and Fischer, R.L.** (2011). Regulation of imprinted gene expression in *Arabidopsis* endosperm. *Proc. Natl. Acad. Sci. USA* **108**: 1755–1762.
- Huh, J.H., Bauer, M.J., Hsieh, T.F., and Fischer, R.L.** (2008). Cellular programming of plant gene imprinting. *Cell* **132**: 735–744.
- Ikeda, Y.** (2012). Plant imprinted genes identified by genome-wide approaches and their regulatory mechanisms. *Plant Cell Physiol.* **53**: 809–816.
- Ingram, G.C.** (2010). Family life at close quarters: Communication and constraint in angiosperm seed development. *Protoplasma* **247**: 195–214.
- Jahnke, S., and Scholten, S.** (2009). Epigenetic resetting of a gene imprinted in plant embryos. *Curr. Biol.* **19**: 1677–1681.
- Jiang, H., and Köhler, C.** (2012). Evolution, function, and regulation of genomic imprinting in plant seed development. *J. Exp. Bot.* **63**: 4713–4722.
- Jullien, P.E., and Berger, F.** (2010). Parental genome dosage imbalance deregulates imprinting in *Arabidopsis*. *PLoS Genet.* **6**: e1000885.
- Jullien, P.E., Katz, A., Oliva, M., Ohad, N., and Berger, F.** (2006). Polycomb group complexes self-regulate imprinting of the Polycomb group gene MEDEA in *Arabidopsis*. *Curr. Biol.* **16**: 486–492.
- Kim, D., Perte, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L.** (2013). TopHat2: Accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**: R36.
- Kiyosue, T., Ohad, N., Yadegari, R., Hannon, M., Dinneny, J., Wells, D., Katz, A., Margossian, L., Harada, J.J., Goldberg, R.B., and Fischer, R.L.** (1999). Control of fertilization-independent endosperm development by the MEDEA polycomb gene in *Arabidopsis*. *Proc. Natl. Acad. Sci. USA* **96**: 4186–4191.
- Köhler, C., and Grossniklaus, U.** (2005). Seed development and genomic imprinting in plants. *Prog. Mol. Subcell. Biol.* **38**: 237–262.
- Köhler, C., Hennig, L., Spillane, C., Pien, S., Griessem, W., and Grossniklaus, U.** (2003). The Polycomb-group protein MEDEA regulates seed development by controlling expression of the MADS-box gene PHERES1. *Genes Dev.* **17**: 1540–1553.
- Konieczny, A., and Ausubel, F.M.** (1993). A procedure for mapping *Arabidopsis* mutations using co-dominant ecotype-specific PCR-based markers. *Plant J.* **4**: 403–410.
- Kwon, S.H., Chang, S.C., Ko, J.H., Song, J.T., and Kim, J.H.** (2011). Overexpression of *Brassica rapa* NGATHA1 gene confers de-etiolation phenotype and cytokinin resistance on *Arabidopsis thaliana*. *J. Plant Biol.* **54**: 119–125.
- Latrasse, D., et al.** (2011). Control of flowering and cell fate by LIF2, an RNA binding partner of the polycomb complex component LHP1. *PLoS ONE* **6**: e16592.
- Law, J.A., Vashisht, A.A., Wohlschlegel, J.A., and Jacobsen, S.E.** (2011). SHH1, a homeodomain protein required for DNA methylation, as well as RDR2, RDM4, and chromatin remodeling factors, associate with RNA polymerase IV. *PLoS Genet.* **7**: e1002195.
- Lefebvre, L., Viville, S., Barton, S.C., Ishino, F., Keverne, E.B., and Surani, M.A.** (1998). Abnormal maternal behaviour and growth retardation associated with loss of the imprinted gene Mest. *Nat. Genet.* **20**: 163–169.
- Leighton, P.A., Ingram, R.S., Eggenschwiler, J., Efstratiadis, A., and Tilghman, S.M.** (1995). Disruption of imprinting caused by deletion of the H19 gene region in mice. *Nature* **375**: 34–39.
- Li, J., and Berger, F.** (2012). Endosperm: Food for humankind and fodder for scientific discoveries. *New Phytol.* **195**: 290–305.
- Li, N., and Dickinson, H.G.** (2010). Balance between maternal and paternal alleles sets the timing of resource accumulation in the maize endosperm. *Proc. Biol. Sci.* **277**: 3–10.
- Lin, B.Y.** (1984). Ploidy barrier to endosperm development in maize. *Genetics* **107**: 103–115.
- Lopes, M.A., and Larkins, B.A.** (1993). Endosperm origin, development, and function. *Plant Cell* **5**: 1383–1399.
- Lu, F.L., Cui, X., Zhang, S.B., Jenuwein, T., and Cao, X.F.** (2011). *Arabidopsis* REF6 is a histone H3 lysine 27 demethylase. *Nat. Genet.* **43**: 715–719.
- Luo, M., Bilodeau, P., Koltunow, A., Dennis, E.S., Peacock, W.J., and Chaudhury, A.M.** (1999). Genes controlling fertilization-independent seed development in *Arabidopsis thaliana*. *Proc. Natl. Acad. Sci. USA* **96**: 296–301.
- Luo, M., Taylor, J.M., Spriggs, A., Zhang, H.Y., Wu, X.J., Russell, S., Singh, M., and Koltunow, A.** (2011). A genome-wide survey of imprinted genes in rice seeds reveals imprinting primarily occurs in the endosperm. *PLoS Genet.* **7**: e1002125.
- Lur, H.S., and Setter, T.L.** (1993). Role of auxin in maize endosperm development (timing of nuclear DNA endoreduplication, zein expression, and cytokinin). *Plant Physiol.* **103**: 273–280.
- Meyer, S., and Scholten, S.** (2007). Equivalent parental contribution to early plant zygotic development. *Curr. Biol.* **17**: 1686–1691.
- Nodine, M.D., and Bartel, D.P.** (2012). Maternal and paternal genomes contribute equally to the transcriptome of early plant embryos. *Nature* **482**: 94–97.
- Ohad, N., Margossian, L., Hsu, Y.C., Williams, C., Repetti, P., and Fischer, R.L.** (1996). A mutation that allows endosperm development without fertilization. *Proc. Natl. Acad. Sci. USA* **93**: 5319–5324.
- Olsen, O.A.** (2001). Endosperm development: Cellularization and cell fate specification. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **52**: 233–267.
- Olsen, O.A.** (2004). Nuclear endosperm development in cereals and *Arabidopsis thaliana*. *Plant Cell* **16** (suppl.): S214–S227.

- Ray, A.** (1997). Three's company: Regulatory cross-talk during seed development. *Plant Cell* **9**: 665–667.
- Robinson, M.D., McCarthy, D.J., and Smyth, G.K.** (2010). edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**: 139–140.
- Sabelli, P.A., and Larkins, B.A.** (2009). The development of endosperm in grasses. *Plant Physiol.* **149**: 14–26.
- Salas, M., John, R., Saxena, A., Barton, S., Frank, D., Fitzpatrick, G., Higgins, M.J., and Tycko, B.** (2004). Placental growth retardation due to loss of imprinting of *Phlda2*. *Mech. Dev.* **121**: 1199–1210.
- Schneeberger, K., Ossowski, S., Lanz, C., Juul, T., Petersen, A.H., Nielsen, K.L., Jørgensen, J.E., Weigel, D., and Andersen, S.U.** (2009). SHOREmap: Simultaneous mapping and mutation identification by deep sequencing. *Nat. Methods* **6**: 550–551.
- Schneeberger, K., et al.** (2011). Reference-guided assembly of four diverse *Arabidopsis thaliana* genomes. *Proc. Natl. Acad. Sci. USA* **108**: 10249–10254.
- Scott, R.J., Spielman, M., Bailey, J., and Dickinson, H.G.** (1998). Parent-of-origin effects on seed development in *Arabidopsis thaliana*. *Development* **125**: 3329–3341.
- Sen, T.Z., Andorf, C.M., Schaeffer, M.L., Harper, L.C., Sparks, M.E., Duvick, J., Brendel, V.P., Cannon, E., Campbell, D.A., and Lawrence, C.J.** (2009). MaizeGDB becomes 'sequence-centric'. *Database (Oxford)* **2009**: bap020.
- Shirzadegan, M., Christie, P., and Seemann, J.R.** (1991). An efficient method for isolation of RNA from tissue cultured plant cells. *Nucleic Acids Res.* **19**: 6055.
- Springer, N.M., Danilevskaya, O.N., Hermon, P., Helentjaris, T.G., Phillips, R.L., Kaepler, H.F., and Kaepler, S.M.** (2002). Sequence relationships, conserved domains, and expression patterns for maize homologs of the polycomb group genes *E(z)*, *esc*, and *E(Pc)*. *Plant Physiol.* **128**: 1332–1345.
- Steffen, J.G., Kang, I.H., Portereiko, M.F., Lloyd, A., and Drews, G.N.** (2008). AGL61 interacts with AGL80 and is required for central cell development in *Arabidopsis*. *Plant Physiol.* **148**: 259–268.
- Takahashi, K., Kobayashi, T., and Kanayama, N.** (2000). p57(Kip2) regulates the proper development of labyrinthine and spongiotrophoblasts. *Mol. Hum. Reprod.* **6**: 1019–1025.
- Vielle-Calzada, J.P., Baskar, R., and Grossniklaus, U.** (2000). Delayed activation of the paternal genome during seed development. *Nature* **404**: 91–94.
- Wang, D., Zhang, C., Hearn, D.J., Kang, I.H., Punwani, J.A., Skaggs, M.I., Drews, G.N., Schumaker, K.S., and Yadegari, R.** (2010). Identification of transcription-factor genes expressed in the *Arabidopsis* female gametophyte. *BMC Plant Biol.* **10**: 110.
- Wang, Z.Q., Fung, M.R., Barlow, D.P., and Wagner, E.F.** (1994). Regulation of embryonic growth and lysosomal targeting by the imprinted *Igf2/Mpr* gene. *Nature* **372**: 464–467.
- Waters, A.J., Makarevitch, I., Eichten, S.R., Swanson-Wagner, R.A., Yeh, C.T., Xu, W., Schnable, P.S., Vaughn, M.W., Gehring, M., and Springer, N.M.** (2011). Parent-of-origin effects on gene expression and DNA methylation in the maize endosperm. *Plant Cell* **23**: 4221–4233.
- Weijers, D., Geldner, N., Offringa, R., and Jürgens, G.** (2001). Seed development: Early paternal gene activity in *Arabidopsis*. *Nature* **414**: 709–710.
- Wilkins, J.F., and Haig, D.** (2003). What good is genomic imprinting: The function of parent-specific gene expression. *Nat. Rev. Genet.* **4**: 359–368.
- Wolf, J.B., and Hager, R.** (2006). A maternal-offspring coadaptation theory for the evolution of genomic imprinting. *PLoS Biol.* **4**: e380.
- Wolff, P., Weinhofer, I., Seguin, J., Roszak, P., Beisel, C., Donoghue, M.T., Spillane, C., Nordborg, M., Rehmsmeier, M., and Köhler, C.** (2011). High-resolution analysis of parent-of-origin allelic expression in the *Arabidopsis* endosperm. *PLoS Genet.* **7**: e1002126.
- Woo, H.R., Dittmer, T.A., and Richards, E.J.** (2008). Three SRA-domain methylcytosine-binding proteins cooperate to maintain global CpG methylation and epigenetic silencing in *Arabidopsis*. *PLoS Genet.* **4**: e1000156.
- Yadegari, R., and Drews, G.N.** (2004). Female gametophyte development. *Plant Cell* **16** (suppl.): S133–S141.
- Zhang, M., Zhao, H.N., Xie, S.J., Chen, J., Xu, Y.Y., Wang, K.K., Zhao, H.M., Guan, H.Y., Hu, X.J., Jiao, Y.P., Song, W.B., and Lai, J.S.** (2011). Extensive, clustered parental imprinting of protein-coding and noncoding RNAs in developing maize endosperm. *Proc. Natl. Acad. Sci. USA* **108**: 20042–20047.

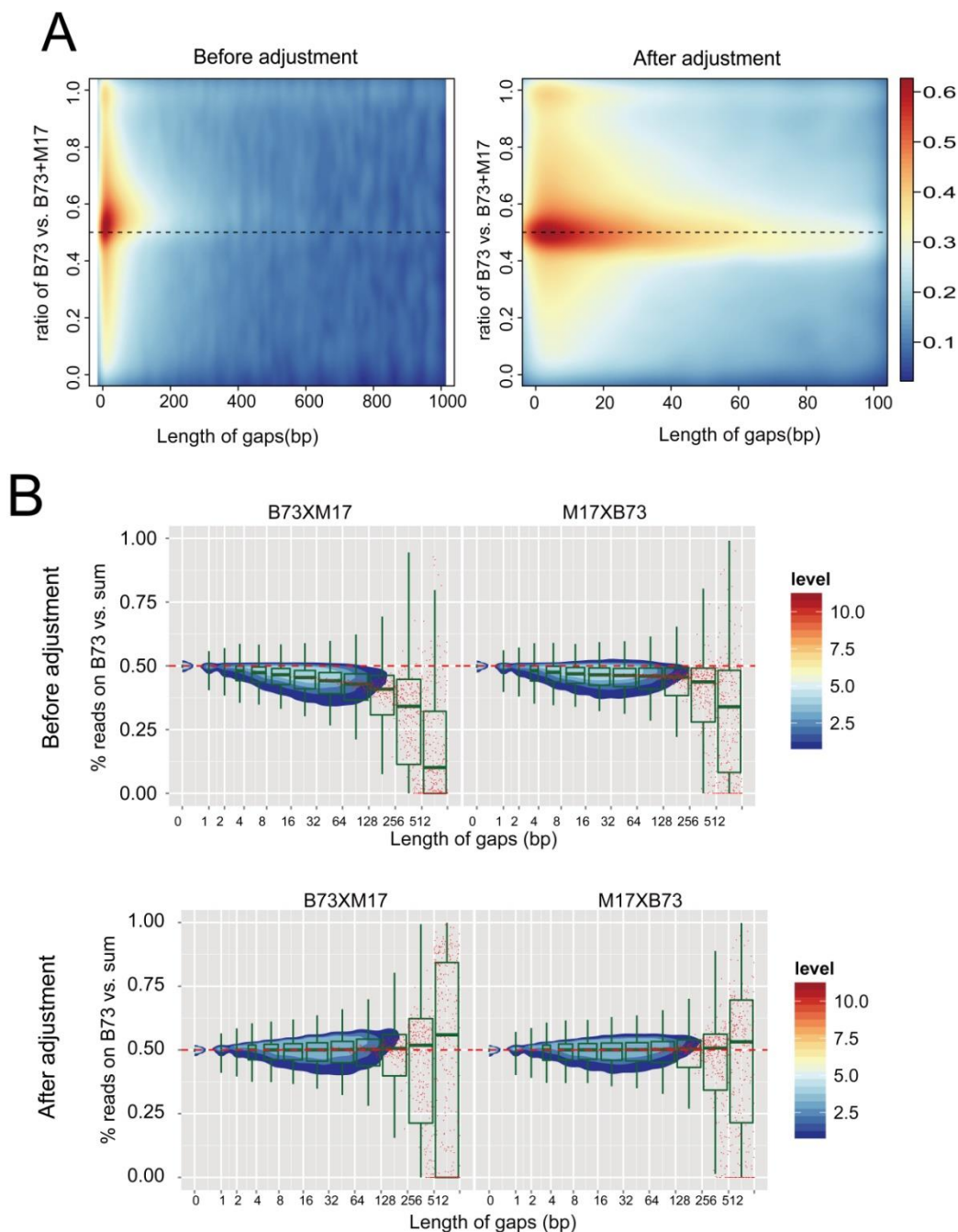
### Supplemental Figure 1



#### Supplemental Figure 1.SNP distribution per gene

More than 40,000 genes have no SNPs between the B73 and Mo17 inbred lines, and gene number decreased with the increase number of SNPs per gene increased.

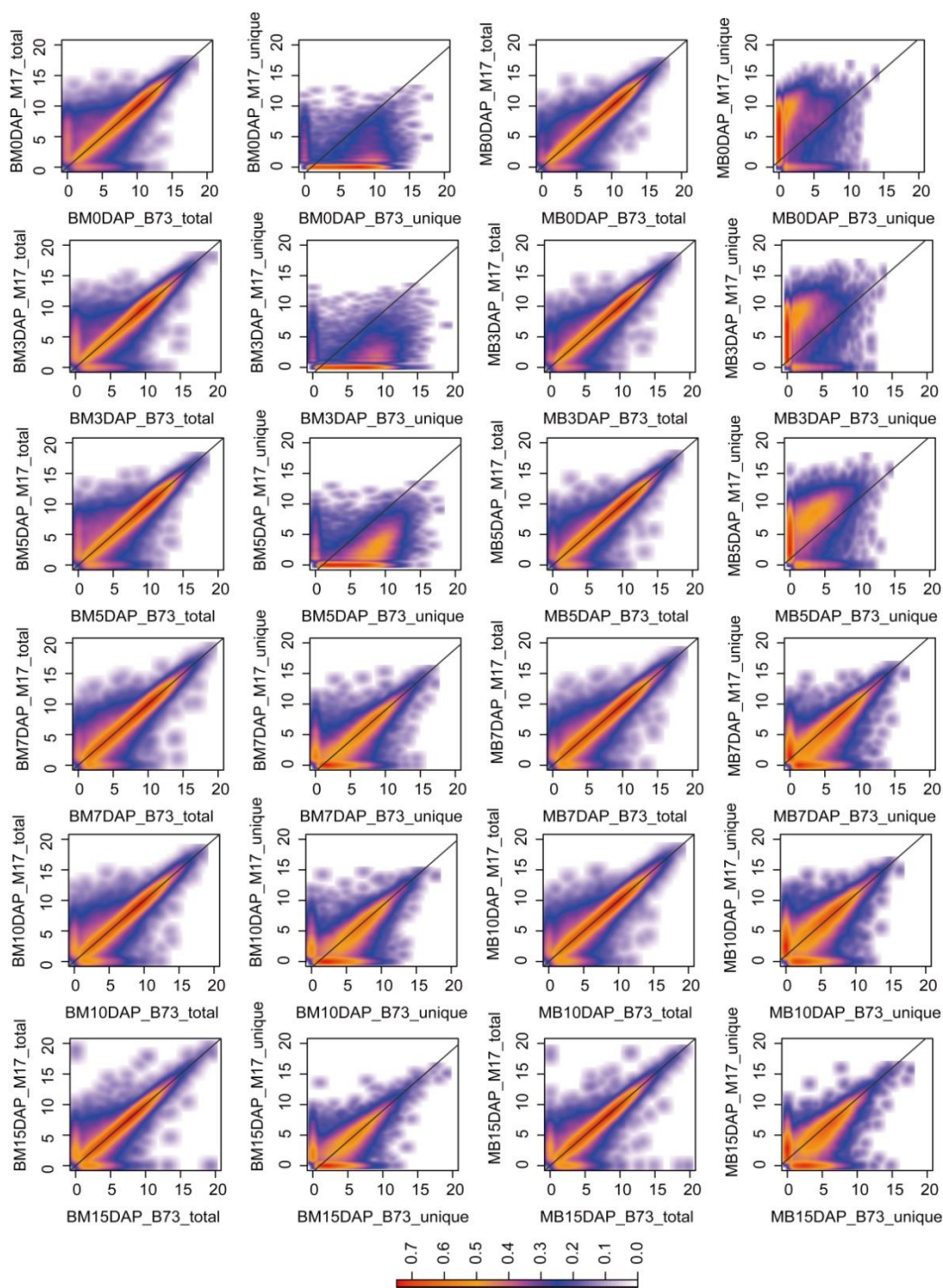
## Supplemental Figure 2



### Supplemental Figure 2. Adjustment of parental gene expression ratios based on the gap length between B73 and Mo17 genomes

- (A) The deviation of reads ratio of the B73 versus Mo17 before and after adjustment.  
 (B) The influences of gap lengths to the allele expression before and after adjustment.

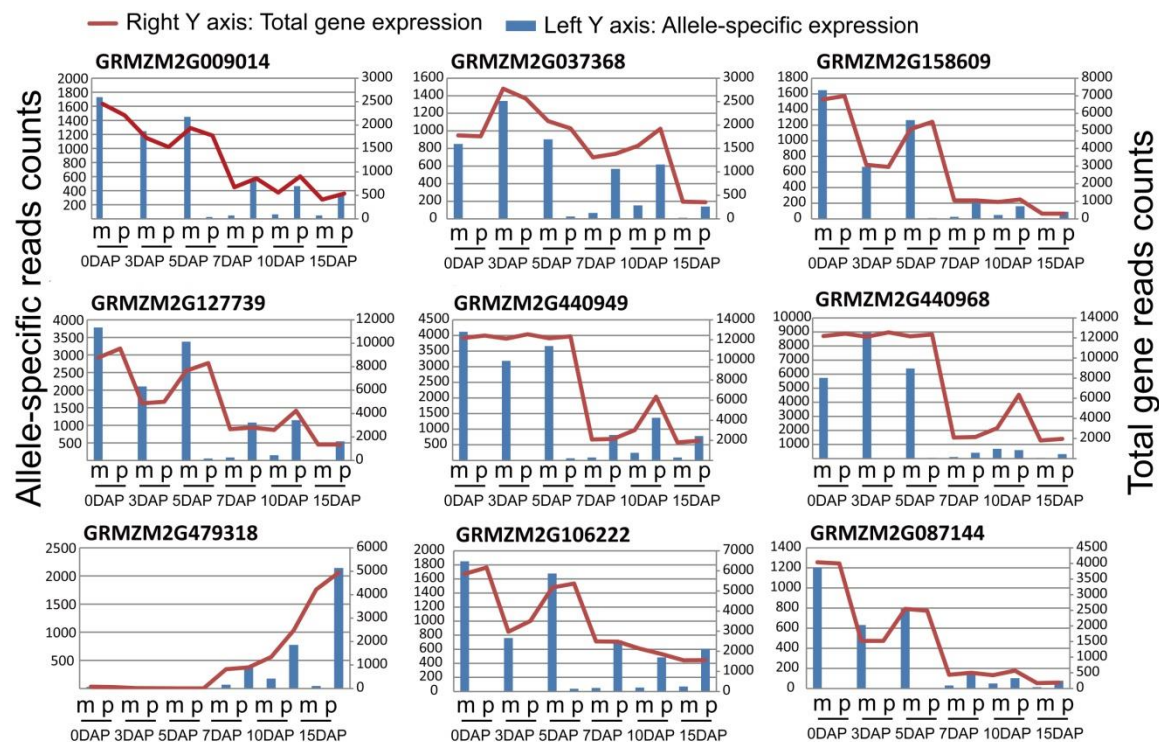
### Supplemental Figure 3



**Supplemental Figure 3. Parental expression ratios plotted separately for each B73 and Mo17 cross**

BM represents B73×Mo17, MB represents Mo17×B73.

### Supplemental Figure 4

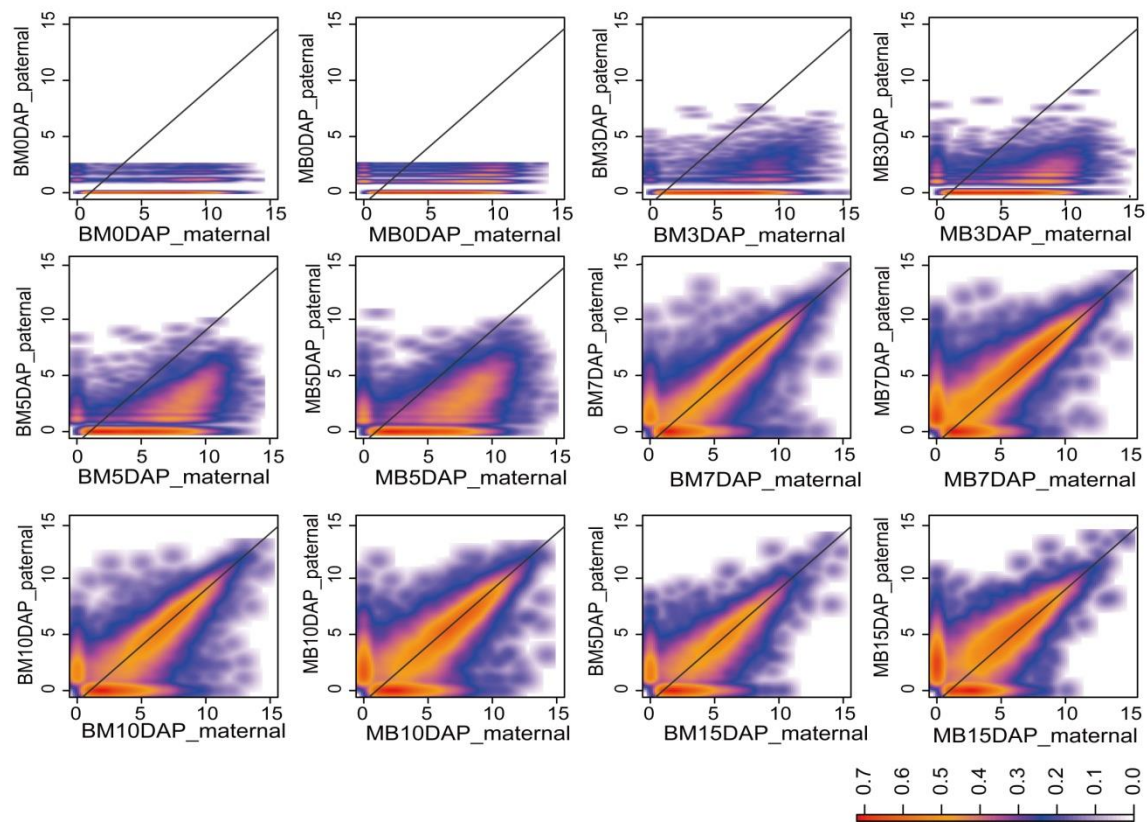


GRMZM2G009014: expressed protein  
 GRMZM2G037368: (IAA27,PAP2, Aux/IAA): phytochrome-associated protein 2  
 GRMZM2G158609: expressed protein  
 GRMZM2G127739: (XLG1): extra-large G-protein 1  
 GRMZM2G440949: (NF-YC11): nuclear factor Y, subunit C11  
 GRMZM2G440968: (CYSB): cystatin B, cysteine proteinase inhibitor precursor protein  
 GRMZM2G479318: expressed protein  
 GRMZM2G087144: Ypt/Rab-GAP domain of gyp1p superfamily protein  
 GRMZM2G106222: (UBP8): ubiquitin-specific protease 8

### Supplemental Figure 4. Allele-specific expression inferred from the SNP-associated reads was proportional to the total expression level of an imprinted gene

The allele-specific gene counts and total reads counts are normalized according to the sequencing library size. For total gene reads counts, each stage has two numerical values, the first one is reads counts mapped to B73 genome and the second one is reads counts mapped to assembled Mo17 genome.

### Supplemental Figure 5

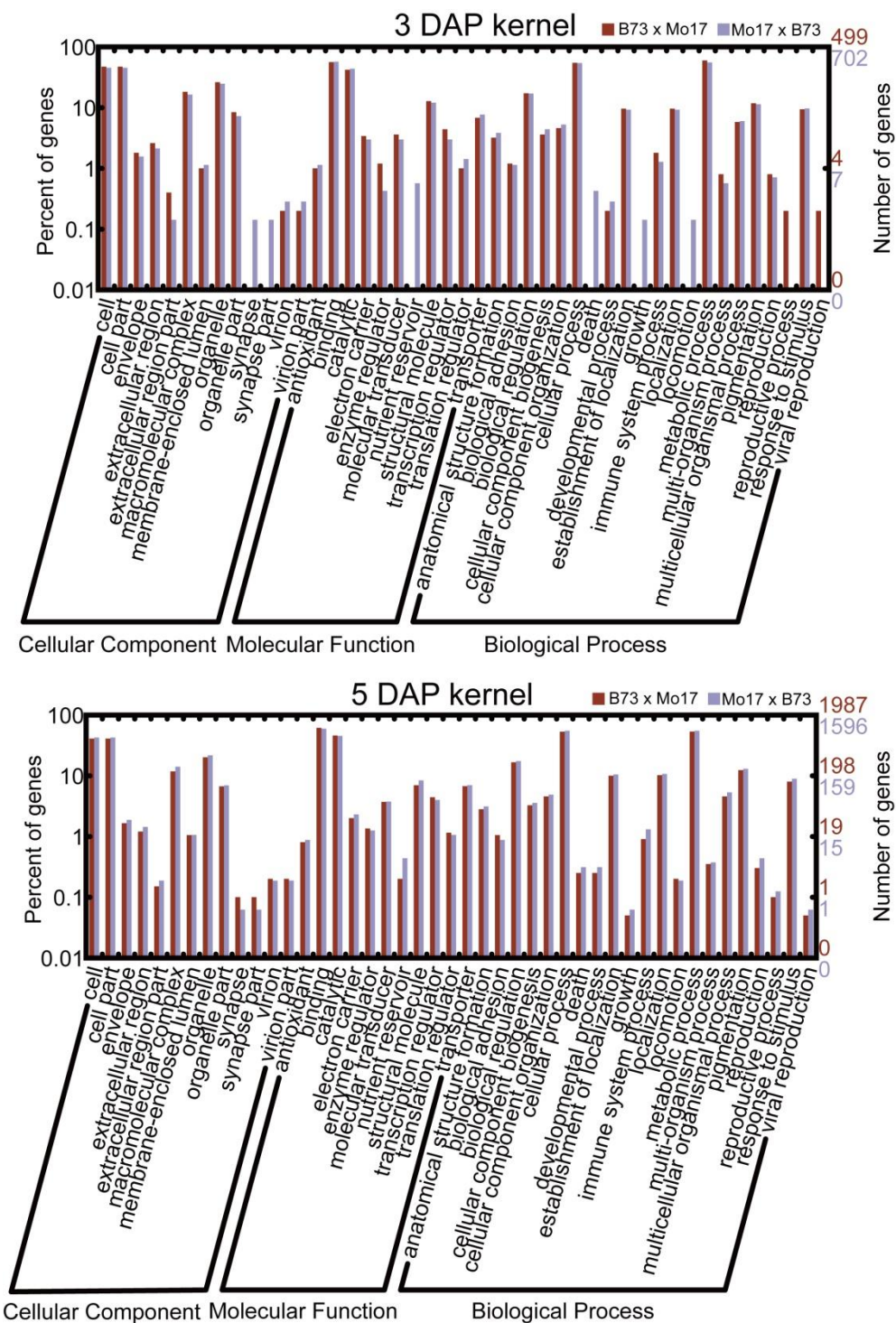


### Supplemental Figure 5. Paternal genome activation in the Mo17 x B73 (MB)

kernel was faster than in the B73 x Mo17 (BM) kernel

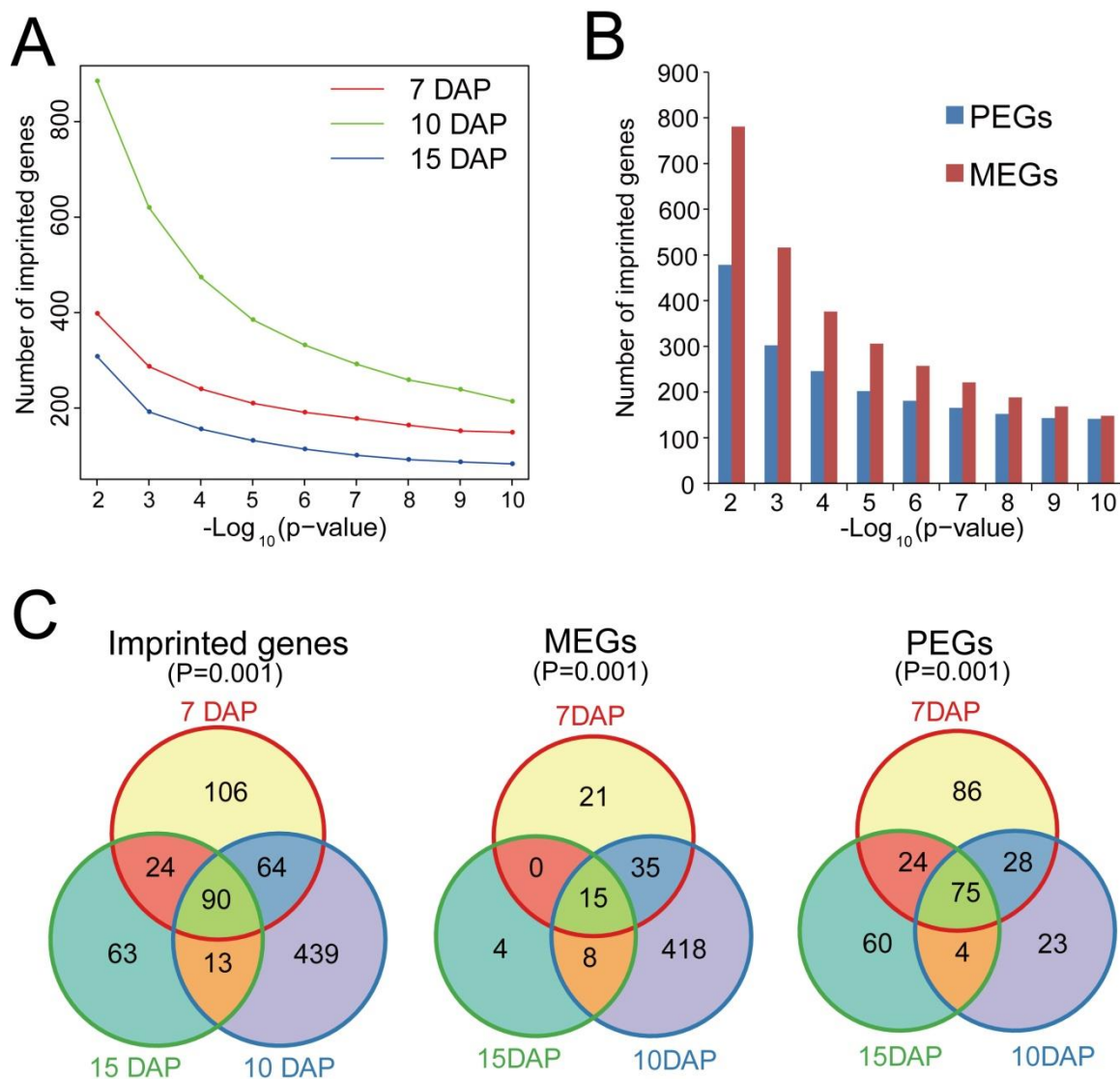
Y axis:  $\log_2$ -transformed paternally-derived reads. X axis:  $\log_2$ -transformed maternally-derived reads

**Supplemental Figure 6**



**Supplemental Figure 6. GO comparison of paternally activated genes at 3 and 5 DAP in the reciprocal crosses**



**Supplemental Figure 7**

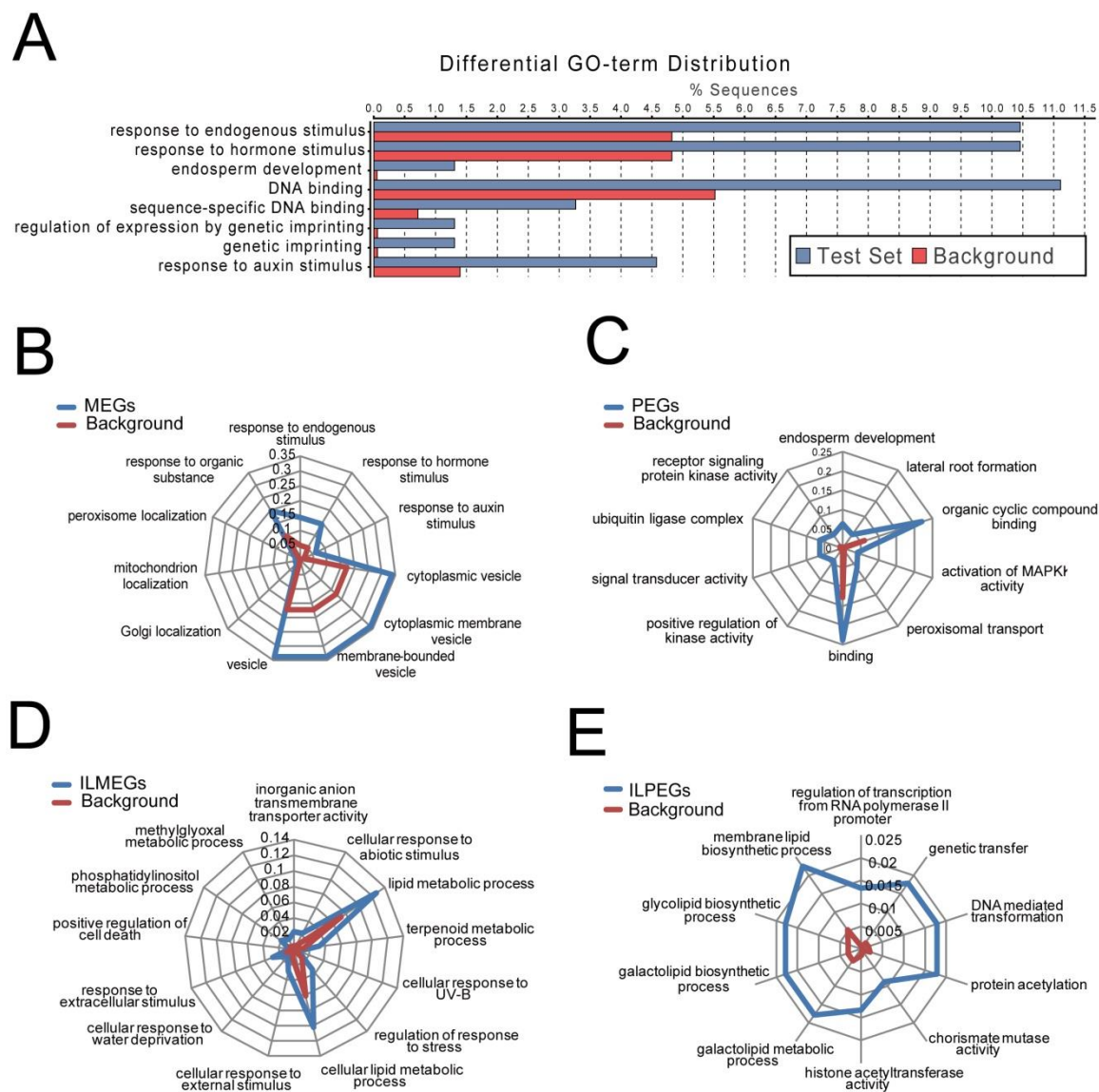
**Supplemental Figure 7. Identification of genes showing maternally and paternally biased expression using the  $\chi^2$  test**

(A) Numbers of parent-specifically expressed genes at different  $\chi^2$  test p values at 7, 10, and 15 DAP.

(B) Numbers of maternally and paternally biased expressed genes at different  $\chi^2$  test p values at the three stages.

(C) Venn diagram of the prescreen of imprinted genes at the three stages by the  $\chi^2$  test.

## Supplemental Figure 8



### Supplemental Figure 8. GO enrichment analysis of imprinted genes.

(A) GO enrichment analysis of 290 imprinted genes.

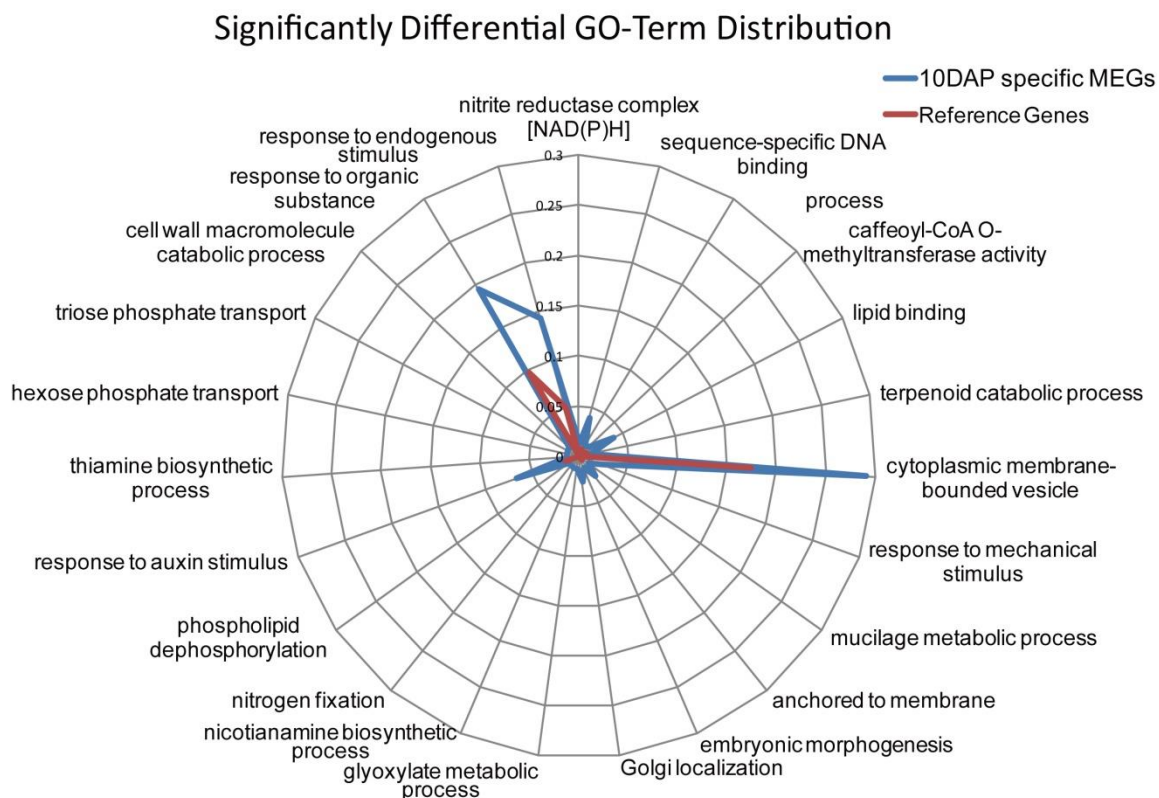
(B) GO enrichment analysis of 194 MEGs.

(C) GO enrichment analysis of 96 PEGs.

(D) GO enrichment analysis of 660 ILMEGs.

(E) GO enrichment analysis of 364 ILPEGs.

**Supplemental Figure 9**



**Supplemental Figure 9. GO analysis of 10DAP-specific MEGs**

**Supplemental Table 1. Gene ontology (GO) categories enriched with imprinted genes**

Maize ID	Putative function	Arabidopsis homolog
<b><i>13 MEGs enriched in "Response to hormone stimulus (GO:0009725)" P = 0.00151</i></b>		
GRMZM2G024468	MYB family transcription factor	AT3G47600 (MYB94)
AC207342.3_FG008	DCD (Development and Cell Death) domain protein	AT5G42050 (N/A)
GRMZM2G122543	Poly (ADP-ribose) polymerase catalytic domain protein (ATP8)	AT1G32230 (ATP8)
GRMZM2G112925	Protein phosphatase 2C	AT1G07630 (PLL5)
GRMZM2G131476	Homeobox-leucine zipper transcription factor family	AT4G37790 (HAT22)
GRMZM2G141382	Polyketide cyclase/dehydrase and lipid transport superfamily protein	AT5G05440 (PYL5)
GRMZM2G152689	Phosphatidylethanolamine-binding protein 10 (zcn10)	AT1G18100 (MFT)
GRMZM2G030465	AUX/IAA transcriptional regulator family protein,	AT5G43700 (IAA4)
GRMZM2G014119	Ubiquitin family protein	AT1G31340 (RUB1)
GRMZM2G121468	Co-factor for nitrate, reductase and xanthine dehydrogenase 7 (vp15)	AT4G10100 (CNX7)
GRMZM2G317900	Auxin-responsive B3 family transcription factor	AT5G20730 (ARF7)
GRMZM2G098643	Auxin efflux carrier family protein	AT1G73590 (PIN1)
GRMZM2G043191	Inositol polyphosphate 5-phosphatase 11	AT1G47510
<b><i>26 MEGs enriched in "Cytoplasmic membrane-bounded vesicle (GO:0016023)" P = 0.003687</i></b>		
GRMZM2G019317	BRASSINOSTEROID INSENSITIVE 1-associated receptor kinase	AT5G16000 (NIK1)
GRMZM2G158682	Citrate transmembrane transporter	AT1G02260
GRMZM2G099960	Expressed protein with unknown function	AT5G63500
GRMZM2G112792	FAD-binding and arabino-lactone oxidase domain protein	AT2G46750
GRMZM2G103937	GDSL-like Lipase/Acylhydrolase superfamily protein	AT2G42990
GRMZM2G087150	GDSL-like Lipase/Acylhydrolase superfamily protein	AT1G28590
GRMZM2G164090	Gibberellin-regulated GASA/GAST/Snakin family protein	AT1G75750 (GASA1)
GRMZM2G059029	Glycine and cysteine rich family protein precursor	AT4G21620
GRMZM2G019373	hydroxyproline-rich glycoprotein family protein	AT5G09530
GRMZM2G041065	Inorganic H pyrophosphatase family protein	AT1G15690 (AVP3)
GRMZM2G053720	Methylenetetrahydrofolate reductase family protein	AT5G38710
GRMZM5G814718	Multicopper oxidase domain (laccase 5) protein	AT2G40370(LAC5)
GRMZM2G168214	Nodulin MtN21 /EamA-like transporter family protein	AT3G30340
GRMZM2G065230	Nodulin MtN21 /EamA-like transporter family protein	AT1G09380
GRMZM2G341934	Peroxidase superfamily protein	AT1G05260 (RCI3)
GRMZM2G053779	Plastocyanin-like domain cupredoxin superfamily protein	AT1G72230
GRMZM2G023847	Plastocyanin-like domain uclacyanin 1 protein	AT2G32300 (UCC1)
GRMZM2G118873	Pollen allergen and expansin B4 protein	AT2G45110 (EXPB4)
GRMZM2G176595	Pollen allergen and expansin B4 protein	AT1G65680 (EXPB2)
GRMZM2G009465	Pollen Ole e I allergen and extensin family protein	AT2G21140 (PRP2)
GRMZM2G114356	Pollen Ole e I allergen and extensin family protein	AT2G21140 (PRP2)
GRMZM2G003909	Pollen Ole e I allergen and extensin family protein	AT2G21140 (PRP2)
GRMZM2G475948	Receptor lectin kinase	AT2G37710 (RLK)
GRMZM5G803318	Sapoin type B domain-containing protein	AT3G51730
GRMZM2G121293	Subtilase family protein	AT5G67360 (ARA12)
GRMZM2G041175	Tetraspanin family protein	AT2G23810 (TET8)

<b>20 PEGs enriched in “Binding (GO:0005488)” P = 0.0036</b>		
GRMZM5G866423	ARID/BRIGHT DNA-binding domain-containing protein	AT3G43240
GRMZM2G141273	Arogenate dehydratase 2 protein	AT3G07630 (ADT2)
GRMZM2G037368	Auxin-responsive Aux/IAA gene family member	AT4G29080 (IAA27)
GRMZM2G161680	Core histone H2A/H2B/H3/H4, nuclear factor Y	AT1G56170 (HAP5B)
GRMZM2G027750	Cullin 3 protein	AT1G26830 (CUL3)
GRMZM2G028366	DEAD-box ATP-dependent RNA helicase	AT1G54270 (EIF4A-2)
GRMZM2G064905	DNAJ heat shock N-terminal domain-containing protein	AT5G53150
GRMZM2G104866	Ethylene response factor/AP2 domain transcription factor	AT1G28360 (ERF12)
GRMZM2G127739	Extra-large G-protein alpha subunit	AT2G23460 (XLG1)
GRMZM2G324131	GATA zinc finger domain transcription factor	AT5G25830 (GATA12)
GRMZM2G030529	Primary amine oxidase activity, LSD1-like 3 protein	AT4G16310 (LDL3)
GRMZM2G419806	Magnesium-chelatase subunit chlI, chloroplast precursor (oyl)	AT5G45930 (CHLI2)
GRMZM2G121570	MYB family transcription factor	AT5G58850 (MYB119)
GRMZM2G097207	O-Glycosyl hydrolases family 17 protein	AT5G58090
GRMZM2G093947	RNA recognition (RNA-binding) protein	AT4G10610 (RBP37)
GRMZM2G045503	RNA recognition (RNA-binding) protein	AT2G43970
GRMZM2G374169	RNA recognition (RNA-binding) protein	AT4G00830
GRMZM2G152526	RNA recognition (RNA-binding) protein	AT3G13224
GRMZM2G347717	UDP-glucuronic acid decarboxylase 1	AT3G53520 (UXS1)
AC191534.3_FG003	Zinc finger (C3HC4-type RING finger) family protein	AT1G57820 (VIM1)
<b>9 PEGs enriched in “Phosphotransferase activity, alcohol group as acceptor (GO:0016773)” P = 0.0024</b>		
GRMZM2G138245	CAMK includes calcium/calmodulin dependent protein kinases	AT2G46700 (CRK3)
GRMZM2G118471	Expressed protein with unknown function	N/A
GRMZM5G845175	Homolog of yeast sucrose nonfermenting 4	AT1G09020 (SNF4)
GRMZM2G149903	Inositol 1,3,4-trisphosphate 5/6-kinase family protein	AT4G08170
AC217300.3_FG004	Serine/threonine-protein kinase	AT5G18610
GRMZM2G026301	Serine/threonine-protein kinase	AT3G59110
GRMZM2G167999	Transducin/WD40 G-beta repeat superfamily protein	AT3G05090 (LRS1)
GRMZM2G028037	Tyrosine protein kinase domain protein	AT5G15080
GRMZM2G103164	Tyrosine protein kinase domain protein	AT1G60800 (NIK3)

**Supplemental Table 2. GO categories enriched with inbred line-dependent imprinted genes**

Maize ID	Putative Function	Arabidopsis homolog
<b>Inbred line-dependent MEGs (ILMEGs)</b>		
<b>GO:0031668</b>	<b>cellular response to extracellular stimulus</b>	<b>P=0.0065</b>
GRMZM2G335618	glutathione S-transferase	AT1G17180 (GSTU25)
GRMZM2G092793	phosphate transporter	AT2G29650 (ANTR1)
GRMZM2G016602	recA DNA recombination family protein	AT1G79050
GRMZM2G164649	AFG1-like ATPase family protein	AT4G30490
GRMZM5G801969	N/A	N/A
GRMZM2G121303	ubiquitin-conjugating enzyme	AT3G17000 (UBC32)
GRMZM5G870932	Phosphoenolpyruvate carboxykinase	AT4G37870 (PCK1)
GRMZM2G158147	Pseudouridine synthase/archaeosine transglycosylase	AT1G19920 (APS2)
GRMZM2G065989	SPX domain-containing protein	AT2G45130 (SPX3)
GRMZM2G055575	CBL-interacting protein kinase	AT1G01140 (CIPK9)
<b>GO:0080134</b>	<b>regulation of response to stress</b>	<b>P=0.0059</b>
GRMZM5G895534	Transmembrane proteins	AT1G50740
GRMZM2G089944	Acetamidase/Formamidase family protein	AT4G37560
GRMZM2G027958	Leucine-rich repeat protein kinase family protein	AT2G31880 (EVR)
GRMZM2G092793	phosphate transporter	AT2G29650 (ANTR1)
GRMZM5G896173	N/A	N/A
GRMZM2G092174	phytochrome B (phyB)	AT2G18790 (HY3)
GRMZM2G356076	BRASSINOSTEROID INSENSITIVE 1 precursor	AT1G74360
GRMZM2G158734	DNA-binding protein phosphatase 1	AT2G25620 (DBP1)
GRMZM2G472643	CBL-interacting protein kinase	AT4G14580 (CIPK4)
GRMZM2G162413	Auxin-responsive GH3 family protein	AT2G46370 (FIN219)
GRMZM2G103247	protein phosphatase 2C	AT1G07630 (PLL5)
GRMZM2G065989	SPX domain-containing protein	AT2G45130 (SPX3)
<b>inorganic anion transmembrane transporter activity</b>		
<b>GO:0015103</b>	<b>P=0.00039</b>	
GRMZM2G042171	sulfate transporter	AT5G10180 (AST68)
GRMZM5G813886	ABC-2 type transporter family protein	AT5G06530
GRMZM2G092793	phosphate transporter	AT2G29650 (ANTR1)
GRMZM2G069198	natural resistance-associated macrophage protein	AT1G80830 (NRAMP1)
GRMZM5G801969	N/A	N/A
GRMZM2G304700	sulfate transmembrane transporters	AT1G80310
GRMZM2G076526	ABC-2 type transporter family protein	AT3G13220 (ABCG26)
GRMZM2G426922	sulfate transmembrane transporters	AT1G80310
<b>GO:0006089</b>	<b>lactate metabolic process</b>	<b>P=0.0015</b>
GRMZM2G032282	Coproporphyrinogen III oxidase	AT1G03475 (ATCPO-I)
GRMZM2G073916	Iron-sulphur cluster biosynthesis family protein	AT5G03905
GRMZM2G448001	WD40-like Beta Propeller Repeat family protein	AT1G21680
GRMZM5G824811	N/A	N/A
GRMZM2G164649	AFG1-like ATPase family protein	AT4G30490

<i>GO:0006629</i>	<i>lipid metabolic process P=0.0015</i>	
GRMZM2G091235	expressed protein with unknwon function	AT4G03600
GRMZM2G137151	Deoxyxylulose-5-phosphate synthase transketolase	AT4G15560 (CLA)
GRMZM2G105922	non-lysosomal glucosylceramidase	AT3G24180
GRMZM2G143917	NAD(P)-binding Rossmann-fold superfamily protein	AT4G35250
GRMZM2G070305	AAA-type ATPase family protein, peroxisome	AT5G08470 (PEX1)
GRMZM2G374827	non-lysosomal glucosylceramidase	AT5G49900
GRMZM2G078650	PLC-like phosphodiesterases superfamily protein	AT1G49740
GRMZM2G019681	expressed protein with unknwon function	AT5G22875
GRMZM2G529313	Putative Subtilisin homologue	AT5G19660 (S1P)
GRMZM5G850019	fatty acid hydroxylase, sterol	AT3G02580 (BUL1)
GRMZM2G092174	phytochrome	AT2G18790 (HY3)
GRMZM2G056975	1-deoxy-D-xylulose 5-phosphate reductoisomerase	AT5G62790 (DXR)
GRMZM5G824811	N/A	N/A
GRMZM2G120320	WRKY DNA-binding protein 40	AT1G80840 (WRKY40)
GRMZM2G163809	Solanesyl diphosphate synthase	AT1G78510 (SPS1)
GRMZM2G473533	Ergosterol biosynthesis ERG4/ERG24 family	AT3G52940 (ELL1)
GRMZM2G536644	Molecular chaperone Hsp40/DnaJ family protein	AT1G80030
GRMZM2G029065	AT-hook motif nuclear-localized protein 1	AT4G12080 (AHL1)
GRMZM2G104538	Phosphatidatecytidyl transferase, vitamin E pathway gene 5	AT5G04490 (VTE5)
GRMZM2G135277	NAD(P)-binding Rossmann-fold superfamily protein	AT3G59710
GRMZM2G149317	phytoene synthase, chloroplast precursor	AT5G17230 (PSY)
GRMZM2G136306	Expressed protein of unknown function	AT2G27290
GRMZM2G101457	enoyl-CoA hydratase/isomerase family protein	AT4G14440 (EC13)
GRMZM2G156365	Pectin acetylesterase family protein	AT4G19410
GRMZM2G347836	Phosphatidic acid phosphatase (PAP2) family protein	AT5G66450
GRMZM2G704005	Lactoylglutathione lyase /glyoxalase I family protein	AT2G32090
GRMZM2G090788	expressed protein with unknwon function	AT5G16520
GRMZM5G892035	uncharacterized Cys-rich domain containing protein	AT2G40935
GRMZM2G164649	AFG1-like ATPase family protein	AT4G30490
GRMZM2G158629	peroxisomal multifunctional enzyme type 2	AT1G76150 (ATECH2)
GRMZM2G046293	Phosphatidatecytidyl transferase family protein	AT5G58560
GRMZM2G095763	lecithine cholesterol acyltransferase	AT5G13640 (PDAT)
GRMZM2G121303	ubiquitin-conjugating enzyme	AT3G17000 (UBC32)
GRMZM2G144273	amine oxidase, flavin-containing, domain containing protein	AT1G57770
GRMZM2G559436	guanyl-nucleotide exchange factors	AT4G16340 (SPK1)
GRMZM2G057086	cytochrome P450	AT5G04660 (CYP77A4)
GRMZM2G110358	RNA ligase/cyclic nucleotide phosphodiesterase protein	AT5G40190
GRMZM2G324478	expressed protein with unknwon function	AT3G52070
GRMZM2G065989	SPX domain-containing protein	AT2G45130 (SPX3)
<b>Inbred line-dependent PEGs (ILPEGs)</b>		
<i>GO:0004402</i>	<i>histone acetyltransferase activity P=0.0018</i>	
GRMZM2G069886	histone acetyltransferase of the CBP family 1	AT1G79000 (HAC1)
GRMZM2G470556	histone acetyltransferase of the CBP family 12	AT1G16710 (HAC12)

GRMZM2G339379	C2H2 zinc finger protein, relative of early flowering 6	AT3G48430 (REF6)
GRMZM2G139977	histone acetyltransferase of the CBP family 1	AT1G79000 (HAC1)
<b>GO:0046467</b>	<b>membrane lipid biosynthetic process P=0.0088</b>	
GRMZM2G158008	phospholipase D. Active site motif family protein	AT3G16785 (PLD)
GRMZM2G481755	MBOAT (membrane bound O-acyl transferase) family protein	AT1G63050
GRMZM2G083655	SPX domain containing protein	AT5G20150 (SPX1)
GRMZM2G162670	phosphatidylinositolglycan synthase family protein	AT2G34980 (SETH1)
GRMZM2G053322	NAD dependent epimerase/dehydratase family protein	AT4G33030 (SQD1)



### Supplemental Table 3. Summary of alignment of RNA-seq reads to B73 and Mo17 genomes

RNA-Seq reads aligned to the B73 genome												
	0 DAP		3 DAP		5 DAP		7 DAP		10 DAP		15 DAP	
	B x M	M x B	B x M	M x B	B x M	M x B	B x M	M x B	B x M	M x B	B x M	M x B
Total	62,317,176	68,142,396	54,632,189	64,209,404	64,332,353	64,131,794	67,102,270	67,756,542	56,526,003	68,906,158	56,320,878	64,209,515
Mapped	44,812,939	42,972,229	41,440,426	40,898,931	45,554,576	40,748,973	44,164,752	43,293,809	38,223,974	44,127,031	39,930,456	42,966,187
Fraction	71.91%	63.06%	75.85%	63.70%	70.81%	63.54%	65.82%	63.90%	67.62%	64.04%	70.90%	66.92%
RNA-Seq reads aligned to the Mo17 genome												
	0 DAP		3 DAP		5 DAP		7 DAP		10 DAP		15 DAP	
	B x M	M x B	B x M	M x B	B x M	M x B	B x M	M x B	B x M	M x B	B x M	M x B
Total	62,317,176	68,142,396	54,632,189	64,209,404	64,332,353	64,131,794	67,102,270	67,756,542	56,526,003	68,906,158	56,320,878	64,209,515
Mapped	35,228,185	38,833,525	31,598,088	36,869,424	35,938,992	36,720,265	36,845,935	37,541,382	31,593,377	38,431,455	33,859,368	39,049,464
Fraction	56.53%	56.99%	57.84%	57.42%	55.86%	57.26%	54.91%	55.41%	55.89%	55.77%	60.12%	60.82%

### Supplemental Table 4. Primer sequences and enzymes used for qRT-PCR and CAPS validation analysis

Gene_ID	Primer_names	Primer_sequence	Purpose	Enzymes
GRMZM2G574887	RT-4887-F	ACGACGATGAAGCCATGGTG	qRT-PCR	
	RT-4887-R	CATCCGTCGTCTGAGTGAAGG	qRT-PCR	
GRMZM2G131476	RT-476-F	GCCACGCTCTATCATTTCATCG	qRT-PCR	
	RT-476-R	GTGTATGTGGTGTTCATGCTGTC	qRT-PCR	
GRMZM2G444748	RT-748-F	CTGCCGTTTCGTCTAGAGTTC	qRT-PCR	
	RT-748-R	GACGAAGGCAGAGACTGACAC	qRT-PCR	
AC233887.1_FG005	RT-887-L	TACGCTCGGAGGTTTCGAACAG	qRT-PCR	
	RT-887-R	CAAGGGTACCGACAGATTCCG	qRT-PCR	
GRMZM2G332522	RT-522-L	GCAAGGTCATCGCATTCGAC	qRT-PCR	
	RT-522-R	CCGGACGTAGTTAGGCTTGTC	qRT-PCR	
GRMZM5G814718	RT-718-F	CTACCACTTCTACGTGCTGGC	qRT-PCR	
	RT-718-R	GTAAGTCCACGTAGTGACGCAC	qRT-PCR	
GRMZM2G088469	RT-469-L	ACAGCAGCGTACAATCAGGC	qRT-PCR	
	RT-469-R	ATGGAGTCACACACCTCCTCC	qRT-PCR	
GRMZM2G018108	RT-108-F	AGCAATGTCTCCCAGACGTG	qRT-PCR	
	RT-108-R	CACGATCGCTATGATCGCATC	qRT-PCR	
GRMZM2G058032	58032-F	TACCGGATCGTCTGCGAGGA	qRT-PCR/CAPS	HaeIII
	58032-R	ACGGCGGTGTATACATGATGC	qRT-PCR/CAPS	
GRMZM2G099960	99960-F	CATACATGCTCAGTACGTACGC	qRT-PCR/CAPS	RsaI
	99960-R	TCGACGATGGATGGCCACAC	qRT-PCR/CAPS	
GRMZM5G824731	824731-F	ATGTATGTGTGCGCATGCATG	CAPS	BstUI
	824731-R	CAGACAGAAGTGGACTGCGAG	CAPS	
GRMZM2G018044	044-F	GTCTCACGGGAGATCCTAGAC	CAPS	KpnI
	044-R	AACAGACCGGTGTCGTCGTC	CAPS	
GRMZM2G067315	67315-F	GTGGTTACGGAGGACCAGGC	CAPS	BstUI
	67315-R	CAGAGCCGTATCCTCACCAC	CAPS	
AC210517.3_FG003	517-F	TGGAGGTCACCGTCCACAAG	CAPS	HindIII
	517-R	GCAGCGTAAGCTCTCTCCAGGC	CAPS	
GRMZM2G472096	2096-F	GCAGATATGCTTCACCAAGCG	CAPS	BstUI
	2096-R	CACCAAGCAGCTGGTACTGC	CAPS	
GRMZM2G401848	848-F	GCAGGACTATGAGGCATGTC	CAPS	RsaI
	848-R	TCATCTACCTAGTCTCAAACGC	CAPS	
GRMZM2G446999	UB_6999_F	AGGAGGCACAGACGTGCAATG	CAPS	BstUI
	UB_6999_R	CACAACATCCACACACACCATC	CAPS	
GRMZM2G000361	UB_0361-F	CATGGAGCCTAACCGTAAGCG	CAPS	StuI
	UB_0361-R	CATGTGCCTGACTGTCACAGG	CAPS	
AC199068.2_FG017	UB_9068-F	GATGCTGGTACAGCAACCTCAC	CAPS	PstI
	UB_9068-R	CGCAGAAGCATTACACAGCACG	CAPS	
GRMZM2G057436	436-F	CTTCATCAACGATGGATACACG	CAPS	HinfI
	436-R	GACTGTGAAAGGGCGACTAGAG	CAPS	

GRMZM2G092101	2101-F	CTGGGATCGTCGAGCAGATG	CAPS	HaeIII
	2101-R	GCAGAAGTTACCCATCAAGGAG	CAPS	
GRMZM2G088793	8793-F	ACCGATGACCATATCTCCATCC	CAPS	ScaFI
	8793-R	TGCTTCAAGCATCCTTGGGAC	CAPS	

## Supplemental Methods

### 1. Adjustment of Mo17 Allele-specific Expression based on the Gap Size in the Incomplete Assembly of Mo17 Genome

As shown in the above Table, the fractions of the mapped reads to the Mo17 genome of the 12 samples were always lower than the fractions of mapped reads to the B73 genome. This suggests a system bias caused by the gaps in the incomplete Mo17 genome: namely, the gaps in the assembled Mo17 genome may lead to systematically biased higher expression level for many B73 alleles compared to the corresponding Mo17 alleles. Therefore, we needed to perform further adjustment of Mo17 allele-specific read counts based on the sizes of the gap in the exon regions existing in the Mo17 assembly. Specifically, all of the genes were classified into 12 groups according to exonic gap size: 0, 1, 2, 3-4, 5-8, 9-16, 17-32, 33-64, 65-128, 129-256, 257-512, and 512-1000 nucleotides. For a gene in each group, we calculated the ratio of the total reads mapped to B73 against Mo17 genome, and then took the median ratio of all of the genes in one group as the scaling factor to adjust the counts of Mo17 allele-specific reads, whilst the counts of B73 allele-specific reads remained unchanged. This adjustment was performed in different size groups, because different sizes of gaps may have different degrees of bias on inferring allele-specific expression. In addition, because low reads counts may lead to a biased median ratio, genes with less than 30 mapped reads were not included to calculate the scaling factor, but adjustment was performed across all of the genes regardless of the minimum count of 30 reads. As indicated in the distribution before and after adjustment (**Supplemental Figure 2 online**), genes containing gaps longer than 100 bp in their exon regions were not included in further analyses, because long gaps might greatly affect the accuracy of measuring allelic expression. Evaluation of allele-specific gene expression and identification of imprinted genes were based on the gap size-adjusted counts of allele-specific reads without edgeR scaling normalization across the 12 samples.

### 2. Comparison of the Three Sets of Identified Imprinted Genes by Waters et al., Zhang et al. and Xin et al.

Comparison of the three sets of imprinted genes showed that 51 out of 279 genes commonly identified in Zhang et al and Waters et al's studies, and 81 out of 290 genes from our study found either in Zhang et al or Waters et al's study. This discrepancy was most likely due to the utilization of different SNP datasets and different developmental stages of endosperm. Based on the fully assembled Mo17 genome, we used 6,557,611 SNPs between B73 and Mo17 genomes to identify allele-specific expression, while Zhang et al. used 51,416 SNPs and Waters et al. used 1,550,000 SNPs in their studies. Additionally, our study included three stages (7, 10 and 15 DAP) of endosperm, while Zhang et al. and Waters et al. only included one stage, 10 and 14 DAP, respectively. Thus, we identified 209 novel imprinted genes that were not included in the two previous studies.

There are 142 imprinted genes identified by Zhang et al. or Waters et al., but not included in our final list of imprinted genes. 107 of them failed to pass the

prescreen criteria before the Chi-square analysis, including: 36 genes contained gaps longer than 100 bp in the exons of Mo17 alleles; 47 genes contained more than five paternally-derived reads at 0 DAP; 19 genes contained less than forty SNP-associated reads between the B73 and Mo17 alleles; and 5 genes contained no SNPs based on our comparison of B73 and Mo17 genome assemblies. The rest 35 genes did not pass the 90% (MEG) or 70% (PEG) criteria to define an imprinted gene in our data, among which 21 genes passed the  $P=0.001$  cutoff of Chi-square test in both reciprocal crosses, whilst 10 genes passed the cutoff in only one cross. The last 4 genes do not pass the Chi-square test.

Eight genes were confirmed by CAPS experiments but not identified in Water et al.'s and Zhang et al.'s studies, and we examined these eight genes in Zhang et al. and Waters et al.'s data. First of all, the eight genes were not included in Zhang et al.'s analysis, because the SNP information associated with these eight genes was not available. In Waters et al.'s data, we found four genes containing available SNP information, and were type differently from the original study based on the following evidence.

In our data, GRMZM2G058032 was identified as a MEG at 7 and 10 DAP, but became biallelically expressed at 15 DAP. Thus, Waters et al. failed to detect its imprinted expression at 14 DAP endosperm.

In Waters et al.'s data, GRMZM2G067315 exhibited a maternally-biased expression pattern, but failed to pass the ratio criteria (90%): the maternal expression ratio is 89.19% (Mo17xB73) and 77.11% (B73xMo17), respectively.

GRMZM2G099960 exhibited maternal expression in Waters' study, but it did not pass the Chi-square test. Thus, this gene was not included in the Waters et al.'s list of imprinted genes. Additionally, our CAPS analysis showed that GRMZM2G099960 was expressed at every low level at 15DAP (almost undetectable) (Figure 3B), and it was only identified as a MEG in 7 and 10DAP endosperm.

GRMZM2G401848 showed inbred line (B73) dependent allele-specific expression in Waters et al.'s data. But our validation showed that GRMZM2G401848 is a MEG in 7, 10 and 15 DAP endosperm in our data (Figure 3B).

In addition, 22 imprinted genes were confirmed (1 by CAPS) and (21 by re-sequencing) by Waters et al. and 16 of them were also identified in our data with the same imprinting status. We examined the other 6 genes not included in our list of imprinted genes. Four genes, including GRMZM2G127160 and GRMZM2G108285 showing paternally-biased expression, and GRMZM2G374088 and GRMZM2G158884 showing maternally-biased expression, were excluded by our prescreening criteria because paternally-derived reads were found at 0 DAP in either B73xMo17 or Mo17xB73 cross. GRMZM2G369203 also showed paternally-biased expression in our data, but the Mo17 allele contains 140 bp gap in the exon regions. At last, GRMZM2G016145 showed paternally-biased expression ratio passing the 70% criteria in the B73xMo17 cross, but failed passing the criteria in the Mo17xB73 cross.

Four genes confirmed by CAPS experiments in Zhang et al, but not included in our list of imprinted genes: GRMZM2G127160 had paternal reads at 0 DAP;

GRMZM2G160687 only passed the ratio criteria (90%) in Mo17xB73, but not in B73xMo17; at last, GRMZM2G027937 and CRMZM2G354579 were marked as “obsoleted” in the B73\_RefGen\_v2 genome annotation.

Thus, although these genes were not included in the final list of imprinted genes in our studies, their Chi-square testing results and allele-specific read counts were recorded in the Supplemental dataset 1 for future validation.