Extensive, clustered parental imprinting of protein-coding and noncoding RNAs in developing maize endosperm

Mei Zhang[1], Hainan Zhao[1], Shaojun Xie[1], Jian Chen, Yuanyuan Xu, Keke Wang, Haiming Zhao, Haiying Guan, Xiaojiao Hu, Yinping Jiao, Weibin Song, and Jinsheng Lai[2]

State Key Laboratory of Agrobiotechnology and National Maize Improvement Center, Department of Plant Genetics and Breeding, China Agricultural University, Beijing 100193, People's Republic of China

Edited* by Brian A. Larkins, University of Arizona, Tuscon, AZ, and approved October 27, 2011 (received for review July 27, 2011)

Although genetic imprinting was discovered in maize 40 years ago, its exact extent in the triploid endosperm remains unknown. Here, we have analyzed global patterns of allelic gene expression in developing maize endosperms from reciprocal crosses between inbreds B73 and Mo17. We have defined an imprinted gene as one in which the relative expression of the maternal and paternal alleles differ at least fivefold in both hybrids of the reciprocal crosses. We found that at least 179 genes (1.6% of protein-coding genes) expressed in the endosperm are imprinted, with 68 of them showing maternal preferential expression and 111 paternal preferential expression. Additionally, 38 long noncoding RNAs were imprinted. The latter are transcribed in either sense or antisense orientation from intronic regions of normal protein-coding genes or from intergenic regions. Imprinted genes show a clear pattern of clustering around the genome, with a number of imprinted genes being adjacent to each other. Analysis of allele-specific methylation patterns of imprinted loci in the hybrid endosperm identified 21 differentially methylated regions (DMRs) of several hundred base pairs in length, corresponding to both imprinted genes and noncoding transcripts. All DMRs identified are uniformly hypomethylated in maternal alleles and hypermethylated in paternal alleles, regardless of the imprinting direction of their corresponding loci. Our study indicates highly extensive and complex regulation of genetic imprinting in maize endosperm, a mechanism that can potentially function in the balancing of the gene dosage of this triploid tissue.

Genetic imprinting is an epigenetic phenomenon known in higher plants and mammals by which a subset of genes is expressed in a parent-of-origin–dependent manner. In plants, the triploid endosperm is the primary tissue where gene imprinting occurs (1). The importance of imprinted genes on endosperm development has been well illustrated from the pioneering studies of imprinted genes such as *MEA* and *FIS2* (2–7) (see recent reviews; refs. 8 and 9) in *Arabidopsis*.

Hundreds of imprinted genes were identified in mammals, with many of them clustered together and reported to associate with a variety of complex disorders or to play important roles in embryonic and brain development (10, 11). Meanwhile, a number of imprinted noncoding RNAs, such as *kcnq1ot1* antisense noncoding RNA and *Air* noncoding RNA (12, 13), were reported to participate in the developmental regulation of imprinted expression of protein-coding genes and to act on transcriptional silencing. Recent studies in *Arabidopsis* and rice suggest that there exist a large number of imprinted genes in plants as well (14–17).

Although genetic imprinting was first discovered in maize (18), so far there are only seven gene-specific imprinted genes reported in maize (including *Fie1, Fie2*, *Peg1*, *Nrp1, Mez1*, *Meg1*, and *Mee1*) (9), most of which are preferentially expressed in the endosperm. All except *Peg1* show maternal-specific expression (19). However, the exact extent of genetic imprinting in maize endosperm has not been explored.

DNA methylation is known to be involved in the regulation of some imprinted genes. Differential methylation between the paternal and maternal alleles in the endosperm is shown to correlate with allele-specific gene expression of several imprinted genes in both *Arabidopsis* (*HDG3* and *HDG9*) and maize (*Fie1* and *Mez1*) (19–24), although DNA methylation does not always correlate with the allelic expression of imprinted genes (20, 21). Screening of differentially methylated regions (DMRs) between embryo and endosperm resulted in the identification of five imprinted genes in *Arabidopsis* (22). Mutation of genes responsible for methylation maintenance (*MET1*) or demethylation (*DME*) had effects on a subset of imprinted genes in *Arabidopsis* (14, 25).

Upon analysis of large-scale sequencing data of maize endosperm transcriptomes, we found the extent and complexity of gene imprinting in maize endosperm is much more than previously anticipated. Our study provides a comprehensive analysis of imprinted noncoding RNAs and the discovery that imprinted loci (protein-coding genes and noncoding transcripts) can often cluster together in the maize genome. Allele-specific methylation analysis on hybrid maize endosperm has also resulted in the identification of a number of DMRs corresponding to both imprinted genes and noncoding sequences.

Results

Genome-Wide Scanning of Gene Imprinting. To accurately assess allelic expression patterns of maize genes in endosperm, we conducted mRNA-seq of 10 d after pollination (DAP) endosperm for the hybrids of reciprocal crosses (B73 × Mo17 and Mo17 × B73). A total of 149 million 100-bp paired-end reads (14.9 Gb) were obtained. Reads were mapped to the reference B73 genome (26) together with mRNA-seq data of inbreds B73 and Mo17 (see *SI Appendix*, Table S1) to call the SNP between them (*SI Appendix*, *SI Methods*). A total of 51,416 high-quality SNP sites (falling in 11,370 genes), which were covered by at least 10 sequencing reads from both reciprocal crosses, were identified and used to investigate allelic gene expression.

We first calculated the expression ratio between the maternal and paternal alleles at each SNP site in dissected endosperm tissue (*SI Appendix*, *SI Methods*). As expected, the majority of the 11,370 genes in 10 DAP endosperm exhibited a maternal to paternal ratio of 2:1 (*SI Appendix*, Fig. S1), consistent with the

Author contributions: M.Z. and J.L. designed research; M.Z., Hainan Zhao, S.X., J.C., Y.X., K.W., Haiming Zhao, H.G., X.H., Y.J., W.S., and J.L. performed research; M.Z., Hainan Zhao, and S.X. analyzed data; and M.Z., Hainan Zhao, and J.L. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

Freely available online through the PNAS open access option.

[1]M.Z., Hainan Zhao, and S.X. contributed equally to this work.

[2]To whom correspondence should be addressed. E-mail: jlai@cau.edu.cn.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1112186108/-/DCSupplemental.

expected parental genomic contribution in the triploid endosperm. However, a significant number of SNP sites (1,686, 3% of total SNPs, in 699 genes) deviated from the expected ratio in hybrids of both B73 × Mo17 and Mo17 × B73 (Pearson χ^2 test, 0.05). Among 699 genes, 127 showed maternally preferred expression, whereas 572 were paternally preferred. Such a deviated expression pattern cannot be explained by inbred-specific allelic differences, because the same patterns were shown in both directions of reciprocal crosses. Although other maternal effects, e.g., those mediated by the deposition of long-lived transcripts produced in the embryo sac, cannot be excluded, we believe this possibility to be unlikely as we used 10 DAP endosperm. The most likely explanation is that the expression of genes with a deviating allelic expression ratio is affected by the parental origin of the alleles.

To obtain a set of high-confidence imprinted genes, we took an arbitrary standard such that the level of expression for the actively expressed allele is at least five times more than that of the repressed allele in both hybrids of reciprocal crosses (*SI Appendix*, *SI Methods*). Three previously reported imprinted genes (*Fie1, Mez1,* and *Nrp1*) that have SNP polymorphisms in our sequencing data qualified as imprinted genes by using this criterion, and they all showed unambiguous imprinting effects in all SNP sites identified (*SI Appendix*, Fig. S2), confirming the reliability of our screening approach. Using this criterion, we identified 179 candidate imprinted protein-coding genes, including the three known imprinted genes (*SI Appendix*, Table S2). Among them, 111 were paternally expressed gene (PEGs) and 68 were maternally expressed genes (MEGs). Gene ontology analysis showed that PEGs are slightly enriched in ion or nucleic acid binding according to their molecular function (*SI Appendix*, Fig. S3), similar to a recent report in *Arabidopsis* (22). MEGs show enrichment in establishment of localization, localization, and transport according to their biological processes (Pearson χ^2 test; 0.05) (*SI Appendix*, Fig. S3).

As further experimental verification (*SI Appendix*, *SI Methods*), we randomly tested eight candidate imprinted genes (four MEGs and four PEGs) by reverse transcription PCR (RT-PCR) by using RNA samples from 10 DAP endosperm. The analyses of RT-PCR products digested with allele-specific restriction enzymes showed that all eight genes have the same parent-of-origin–dependent expression pattern, as suggested by mRNA-seq data (Fig. 1). We therefore concluded that the majority of the 68 MEGs and 111 PEGs are subject to genomic imprinting.

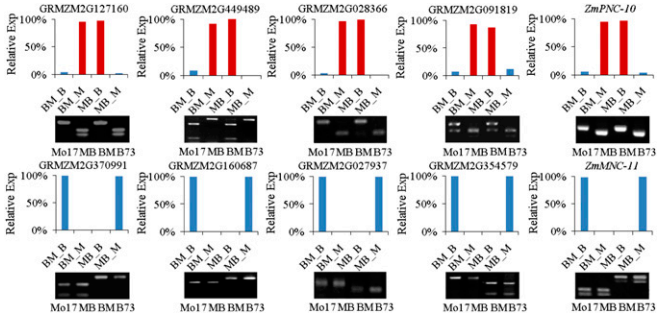

Fig. 1. Confirmation of candidate imprinted protein-coding genes and long noncoding RNAs. *GRMZM2G127160*, *GRMZM2G449489*, *GRMZM2G028366*, and *GRMZM2G091819* are four PEGs, and *GRMZM2G370991*, *GRMZM2G160687*, *GRMZM2G027937*, and *GRMZM2G354579* are four MEGs. *ZmMNC-11* is a candidate MNC, whereas *ZmPNC-10* is a PNC. BM_B, BM_M, MB_B, and MB_M indicate the maternal or paternal allele (B for B73, M for Mo17) in B73 × Mo17 (BM) or Mo17 × B73 (MB) crosses, and the *y* axis shows the relative expression levels as percentages of expression for "B" or "M" allele in samples of BM or MB. Red columns represent the paternal alleles, whereas blue is for maternal alleles. Gel photos are RT-PCR products of 10 DAP endosperm digested by allele-specific restriction enzymes.

Because some of the previously reported imprinted genes in maize (*Dzr-1* and *R*) showed allele-specific imprinting, we therefore examined the imprinting pattern for some of the newly identified imprinted genes in different genetic backgrounds. Taking advantage of recent resequencing results (27), we tested five genes from reciprocal crosses between inbreds Zheng58 and Chang7-2. Our results showed that all of the tested genes had the same imprinting effect in Zheng58 and Chang7-2 as in B73 and Mo17 (*SI Appendix*, Fig. S4), suggesting that most of the imprinted genes we identified are likely to be gene specific, rather than allele specific.

To investigate the imprinting pattern of the newly identified genes at different stages of endosperm development, we analyzed eight imprinted genes from 2 DAP to 20 DAP. The results showed that all eight genes were expressed from 2 DAP to 20 DAP (*SI Appendix*, Fig. S5), and most of them displayed the highest expression at ≈10–12 DAP (*SI Appendix*, Fig. S6). However, only four PEGs and one MEG (*GRMZM2G370991*) maintained their imprinting pattern until 20 DAP. The imprinting pattern of the other three genes (*GRMZM2G027973*, *GRMZM2G354579*, and *GRMZM2G160687*) can be seen from 2 DAP to 16 DAP (*SI Appendix*, Fig. S5). These results showed that the imprinted expression of certain endosperm genes depends on developmental stages, with some of them being imprinted only in early stages, consistent with previous reports for *Fie1, Fie2*, and *Meg1* genes (20, 28).

To assess the tissue specificity of the identified imprinted genes, we conducted RT-PCR by using samples from various tissues. Results of eight selected genes showed that most of them were expressed at a relatively low level in tissues other than endosperm (*SI Appendix*, Fig. S7).

To assess the conservation of imprinted genes, we compared all 179 imprinted protein-coding genes with the recently reported imprinted genes in *Arabidopsis* and rice (14–17) (*SI Appendix*, *SI Methods*). Results show that there are 15 imprinted genes conserved between rice and maize and 6 between maize and *Arabidopsis*. There are two genes (*VIM5*, *YUC10*) showing paternal preferential expression in all three species (*SI Appendix*, Table S4). Overall, our results showed that the conservation of imprinting in plants is very limited.

Identification of Imprinted Noncoding RNAs. Although a number of imprinted noncoding RNAs have been reported in mammals (29), so far there has been no systematic characterization of imprinted noncoding RNAs in plants. Because the mRNA-seq data covered noncoding regions as well, we therefore searched for imprinted endosperm noncoding RNAs. We annotated a transcript as noncoding RNA by a set of strict screening protocols (*SI Appendix*, *SI Methods*). Using the same high stringency criterion as for identifying the imprinted protein-coding genes, we identified a total of 38 noncoding imprinted RNAs (*SI Appendix*, Table S5 and Table S6). Among them, 25 are maternally expressed transcripts (MNC), whereas 13 are paternally expressed transcripts (PNC). These noncoding RNAs are transcribed from two different genomic regions. Nine (six MNC and three PNC) are from intergenic regions; 29 (19 MMC, 10 PNC) are from intronic regions of annotated protein-coding genes, with some having transcripts extending to adjacent exons (*SI Appendix*, Table S5). The imprinted expression of these noncoding transcripts was further validated by RT-PCR for four selected transcripts (*ZmMNC-11*, *ZmPNC-10* in Fig. 1, *ZmMNC-18* in Fig. 2, and *ZmMNC-24* in Fig. 3).

These imprinted noncoding transcripts have an average length of 468 bp, ranging from 79 bp to 1,650 bp (*SI Appendix*, Table S5), as estimated from regions covered by the sequencing reads. The imprinted noncoding transcripts appear to be a class of long noncoding RNAs, because they lack the conserved structure of snoRNAs and the typical hairpin structure of microRNAs, together with relatively greater lengths.

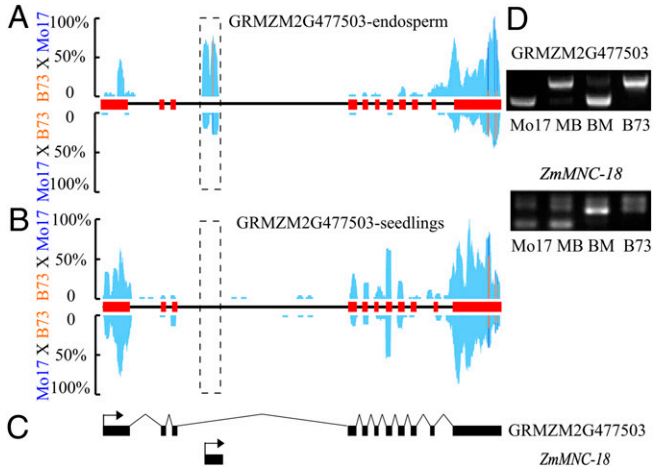

Fig. 2. Imprinted intronic noncoding RNA *ZmMNC-18* transcribed sense to a paternally expressed protein-coding gene *GRMZM2G477503*. (*A* and *B*) Expression and imprinting profile of the genomic region of *GRMZM2G477503* (PEG) in 10 DAP endosperm (*A*) and seedling transcriptomes (*B*). The dashed box represents the genomic region of imprinted noncoding RNA *ZmMNC-18*. Overall expression level of transcribed regions is shown in light blue for both B73 × Mo17 and Mo17 × B73. The relative expressional levels for specific SNP sites are shown for both B73 and Mo17 alleles, orange lines for B73 and blue lines for Mo17. Red rectangle, exon; black line, intron. (*C*) Gene structures of two confirmed transcripts from the *GRMZM2G477503* genomic region in 10 DAP endosperm. The gene structure of *GRMZM2G477503* was confirmed by Sanger sequencing of RT-PCR products, whereas the transcript of *ZmMNC-18* was defined by 5′ and 3′ RACE experiments. Exons (black rectangle) of *GRMZM2G477503* are linked by gray lines; arrows indicate the transcription direction. (*D*) Confirmation of *GRMZM2G477503* and *ZmMNC-18* being paternally and maternally expressed imprinted genes in 10 DAP endosperm. B73, B73 inbred; BM, B73 × Mo17; MB, Mo17 × B73; Mo17, Mo17 inbred.

Interestingly, four maternal specifically expressed intronic noncoding transcripts (*ZmMNC-2*, *ZmMNC-9*, *ZmMNC-18*, and *ZmMNC-24*) are transcribed from within four PEGs (*GRMZM2G128663*, *GRMZM5G854045*, *GRMZM2G477503*, and *GRMZM2G406553*, respectively) (*SI Appendix*, Table S5). Their corresponding intronic regions show an average length of 1,537 bp with a minimum of 1,059 bp. For example, the *GRMZM2G477503* gene region has one maternal allelic expression SNP in its third intron, whereas the other four SNPs in its predicted 5′ UTR region express paternally in 10 DAP endosperm (Fig. 2*A*). *GRMZM2G477503* is annotated as a glycosyl transferase, a protein that can potentially be involved in polycomb repression (30). Results of 5′ and 3′ RACE showed that *ZmMNC-18* is transcribed in the same direction as *GRMZM2G477503* and has a poly(A) tail (Fig. 2*C*). In our endosperm transcriptome data, *GRMZM2G477503* showed significant paternal preferential expression, whereas *ZmMNC-18* exhibited perfect maternal expression (Fig. 2*A*). However, in seedling transcriptome data, *GRMZM2G477503* exhibited normal biallelic expression, whereas *ZmMNC-18* showed no expression at all (Fig. 2 *A* and *B*). Imprinting of both *GRMZM2G477503* and *ZmMNC-18* in endosperm was further validated by RT-PCR (Fig. 2*D*).

The genomic region of *GRMZM2G406553* is even more complex (Fig. 3*C*). Transcriptome sequencing of 10 DAP endosperm showed that the *GRMZM2G406553* gene region has one maternal expression SNP in the sixth intron, whereas all 11 SNPs in the last exon showed paternal-specific expression (Fig. 3*A*). By designing targeted primers and Sanger sequencing, we confirmed that the two differentially expressed regions actually represent two different transcripts in 10 DAP endosperm (one for a protein-coding gene *GRMZM2G406553* and another noncoding *ZmMNC-24*) (Fig. 3*A*), whereas only the transcript of *GRMZM2G406553* was detected in seedlings (Fig. 3*B*). Paternal-specific expression of *GRMZM2G406553* and maternal-specific expression of *ZmMNC-24* were further verified (Fig. 3*D*). Results of 5′ and 3′ RACE for *ZmMNC-24* showed it is transcribed in the reverse direction, compared with the protein-coding gene *GRMZM2G406553* (Fig. 3*C*), and *ZmMNC-24* has many isoforms, including the *ZmMNC-24a*, *ZmMNC-24b*, *ZmMNC-24c*, *ZmMNC-24d*, and *ZmMNC-24e* transcripts (Fig. 3*C*). *ZmMNC-24c*, *ZmMNC-24d*, and *ZmMNC-24e* were validated as maternally expressed by using sequence-specific primers (Fig. 3 *C* and *D*). Moreover, the transcript region shared by *ZmMNC-24a*, *ZmMNC-24b*, and *ZmMNC-24c* was confirmed to be from maternal expression by allele-specific RT-PCR (Fig. 3 *C* and *D*).

Clustering of Imprinted Genes. A hallmark of imprinted genes reported for human and mouse is that they tend to cluster in the genome (31). The relatively large number of the imprinted transcripts identified in this study allowed us to test whether maize imprinted genes are similarly clustered. We scanned the genome for candidate clusters containing at least two imprinted genes within a region of 1 Mb, a standard that is similar to that used in mammals (10). Seventy-four imprinted genes and noncoding RNAs were found to fall into 33 clusters (*SI Appendix*, Table S7), which is significantly higher than expected by chance ($P = 0.01$) (*SI Appendix*, *SI Methods*). Alternatively, the number of clusters is significantly higher ($P = 7.41e^{-7}$) proportionally than the number of clusters for all 31,193 genes expressed in maize endosperm using the same criterion. Although some clusters are located to gene-rich regions, there are clusters in gene-poor or average gene density regions (*SI Appendix*, Fig. S8). Thirteen of 38 imprinted noncoding genes are spread in 13 clusters (*SI Appendix*, Table S7). Some imprinted genes are found to be adjacent to each other. For example, *GRMZM2G147226* (validated MEG) annotated as a "regulation of nuclear pre-mRNA domain-containing (RPRD) protein" and *GRMZM2G447406* (validated PEG) encoding a hypothetical protein are ≈5 kb apart (Fig. 4), with a fragment of a Gypsy retrotransposon and two RTE1 like non-LTR retroelements located in between. A previously reported MEG, *Mez1* (24), and a validated imprinted noncoding gene, *ZmMNC-11* (Fig. 1), are also adjacent (Fig. 4*A* and *SI Appendix*, Table S7), separated by ≈9 kb containing two degenerate transposons.

Characterization of Allele-Specific DMRs for Imprinted Genes. Because DNA methylation is generally believed to be involved in epigenetic regulation, we conducted bisulfite sequencing (*SI Appendix*, *SI Methods*) to assess patterns of DNA methylation of both maternal and paternal alleles of imprinted loci identified in this study. Using DNA from 12 DAP endosperm of the B73 × Mo17 hybrid, 866 million 100-bp paired-end reads were generated, accounting for ≈16-fold genome coverage.

To be analyzed for potential allele-specific methylation, a gene required sufficient read coverage for both maternal and paternal genomic alleles in regions with at least one SNP between B73 and Mo17. A total of 98 genes (39 MEGs, 59 PEGs) met these criteria. Of these genes, 9 MEGs and 8 PEGs (17.3% of analyzed genes) showed differential CpG methylation between the two parental alleles (*SI Appendix*, Table S8). Similarly, we found 4 DMRs for 4 noncoding RNAs among 13 analyzable transcripts (including 11 MNCs and 2 PNCs). All of these four noncoding RNAs belong to MNCs, with three of them being intronic and one intergenic (*SI Appendix*, Table S8). For example, the CpG methylation level of the maternal alleles of a DMR region of *GRMZM2G406553* is 5%, whereas that of the paternal allele is 97% (Fig. 3 *E* and *F*). Interestingly, this DMR of at least 540 bp overlaps with the maternally expressed noncoding RNA, *ZmMNC-24*.

Surprisingly, all of the identified DMRs exhibit hypomethylation in maternal alleles and hypermethylation in paternal alleles.

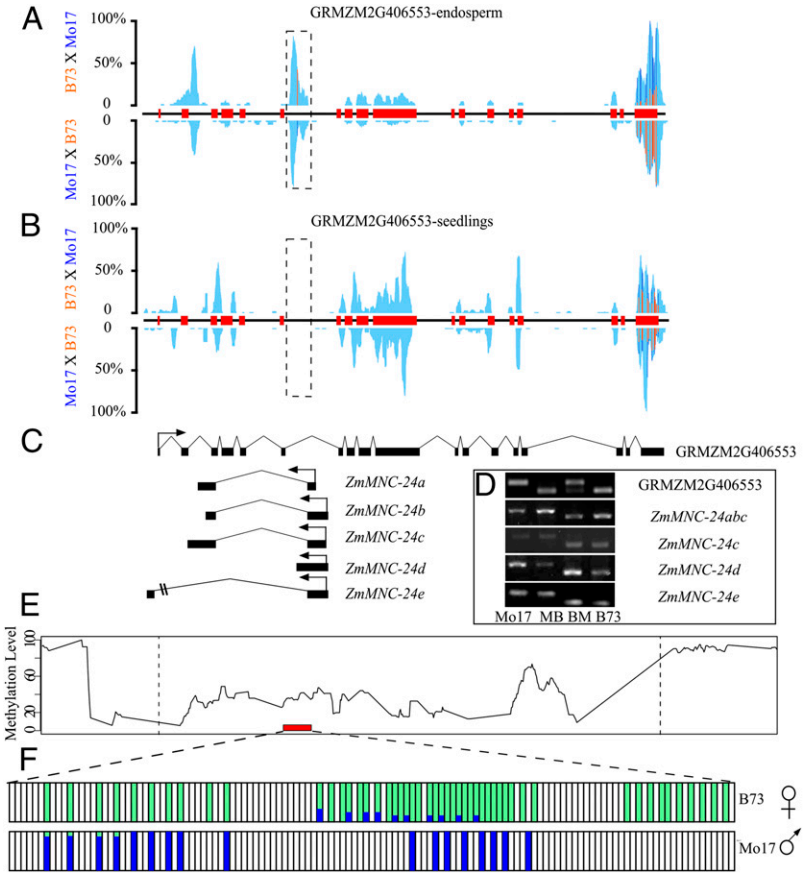

Fig. 3. Imprinting and differential methylation of *GRMZM2G406553* and *ZmMNC-24*. (*A* and *B*) Expression and imprinting profile of the genomic region of *GRMZM2G406553* (PEG) in 10 DAP endosperm (*A*) and seedling transcriptomes (*B*). The dashed box represents the genomic region of imprinted noncoding RNA *ZmMNC-24*. Overall expression level of transcribed regions is shown in light blue for both B73 × Mo17 and Mo17 × B73. The relative expressional levels for specific SNP sites are shown for both B73 and Mo17 alleles, orange lines for B73 and blue lines for Mo17. Red rectangle, exon; black line, intron. (*C*) Gene structures of *GRMZM2G406553* and its related five isoforms. The gene structure of *GRMZM2G406553* was confirmed by Sanger sequencing of RT-PCR products, whereas the transcripts of *ZmMNC-24* were defined by 5′ and 3′ RACE experiments. Exons (black rectangle) of *GRMZM2G406553* and *ZmMNC-24* are linked by gray lines; arrows indicate the transcription direction. (*D*) Confirmation of *GRMZM2G406553* and *ZmMNC-24* as paternally and maternally expressed genes in 10 DAP endosperm. The gel photo shows the results of RT-PCR products using transcript-specific primers digested by allele-specific enzymes. B73, B73 inbred; BM, B73 × Mo17 hybrid; MB, Mo17 × B73 hybrid; Mo17, Mo17 inbred. (*E*) Methylation profile of *GRMZM2G406553* genomic region including 2 kb up- and downstream in B73 × Mo17 endosperm. Red rectangle, DMR; dotted line, the boundary of *GRMZM2G406553* genomic sequence. (*F*) Methylation status for both the maternal (B73) and paternal (Mo17) alleles of the DMR in the B73 × Mo17 hybrid endosperm. The upper track represents the methylation status of maternal allele (B73), whereas the lower track is for the paternal allele (Mo17). Light green rectangular blocks represent all analyzable cytosines, with the height of the blue column indicating the methylation level. White blocks represent the cytosines that are not covered by enough reads (<3 reads) and not analyzed in this study.

The identified DMRs, ranging from 220 bp to 540 bp, have an average of 14% CpG methylation in maternal alleles and 94% in paternal alleles. DMRs for MEGs locate mostly in the upstream or downstream sequences of genes, whereas those of PEGs can be sometimes in the middle of the genes (*SI Appendix*, Table S8).

Discussion

Extent of Genetic Imprinting in Maize. The identification of imprinted genes has long been of interest for maize genetics research (19, 20, 32). We report here that at least 699 genes are potentially imprinted in maize endosperm. Using highly stringent criteria, we identified 111 PEGs, 68 MEGs, and 38 long noncoding RNAs. Our result has significantly expanded the number of imprinted genes known in maize. However, we believe a number of imprinted genes, particularly MEGs specific for early or late stages of maize endosperm development, were likely to be missed from this study. This assumption is exemplified by *Fie2*, which was reported to be an early stage maternally expressed imprinted gene (28) but was not detected in our screen because we used 10 DAP endosperm.

Semiquantitative RT-PCR conducted for eight selected protein-coding genes (*SI Appendix*, Fig. S7) indicated they all have endosperm-preferential expression. This result is consistent with previous reports of imprinted genes in maize and *Arabidopsis* (8, 33). However, analysis of mRNA-seq data generated from 14-d seedlings (*SI Appendix*, Table S1) did not identify any imprinted genes, suggesting that imprinting in maize could be limited to seeds. There is one report identifying an imprinted gene, *Mee1* (21) in maize embryos; however, systematic screening in *Arabidopsis* did not find any embryo imprinted genes (14), whereas three imprinted genes were identified in the rice embryo (17). Whether imprinting is prevalent in the embryo of maize remains to be determined.

Imprinted Noncoding RNAs in Plants. In mammals, there are hundreds of imprinted long noncoding RNAs, such as *Air*, *H19*, and *Kcnq1ot1* (13, 34, 35). Most of these RNAs play important regulatory roles within their imprinted clusters (29). We report here the identification and detailed characterization of 38 long noncoding imprinted RNAs in maize. It is possible these newly identified imprinted long noncoding RNAs also have regulatory functions. The expression patterns of four maternal specifically expressed intronic noncoding RNAs, complementing their associated paternal specifically expressed protein-coding genes, strongly suggest they could be involved in the regulation of imprinting, which resembles the developmental regulation of *Airn* ncRNA to *Igf2r* imprinted expression (36). For example, *MNC-24* is not expressed in seedlings, but is maternally expressed in endosperm, whereas its associated gene, *GRMZM2G406553*, is expressed biallelically in seedlings, but shows paternal-preferred expression in the endosperm (Fig. 3). This result is highly suggestive that the *MNC-24* may be involved in the imprinted expression of *GRMZM2G406553*. *MNC-18* in the intronic region of *GRMZM2G477503* is exactly the same (Fig. 2). We speculate that the maternal specifically expressed noncoding RNAs function in recruitment of a complex to repress maternal alleles, leading to the paternal-specific expression of PEGs. In fact, a sense, long noncoding RNA, *COLDAIR*, has been recently reported to mediate the repression of the *FLC* locus in *Arabidopsis* through the PRC2 complex (37). However, the corresponding genes for eight intronic noncoding RNAs identified in our study express biallelically, suggesting that functions of different noncoding RNAs can be variable.

Clustering of Imprinted Genes in the Maize Genome. The identification of 33 imprinting clusters in maize is very similar to what is found in mammals, suggesting that imprinting in plants and

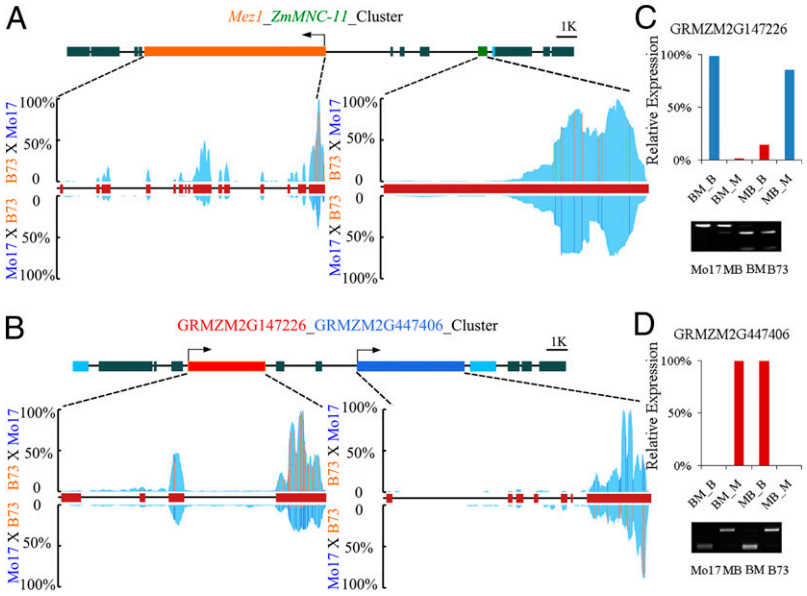

Fig. 4. Examples of neighboring imprinted genes. (*A* and *B*) Genomic region and imprinting profile of two candidate clusters: *Mez1* and *ZmMNC-11* cluster (*A*) and the cluster of two adjacent genes *GRMZM2G147226* and *GRMZM2G447406* (*B*). *Mez1*, shown as an orange rectangle, is a known MEG; *ZmMNC-11*, shown as a green rectangle, is a confirmed MNC (Fig. 1); the blue rectangle represents PEG gene *GRMZM2G447406*; and the red rectangle shows MEG gene *GRMZM2G147226*. Arrows indicate the transcriptional directions. The light blue rectangle represents neighboring genes that are not imprinted. Dark green represents transposons. Overall expression level of transcribed regions is shown in light blue for both B73 × Mo17 and Mo17 × B73. The relative expression levels for specific SNP sites are shown for both B73 and Mo17 alleles, orange lines for B73 and blue lines for Mo17. (*C* and *D*) Validation of imprinted expression for *GRMZM2G147226* (MEG) and *GRMZM2G447406* (PEG). BM_B, BM_M, MB_B, and MB_M indicate the maternal or paternal allele (B for B73, M for Mo17) in B73 × Mo17 (BM) or Mo17 × B73 (MB) crosses, and the *y* axis shows the relative expression level as a percentage of expression for the "B" or "M" allele in samples of BM or MB. Red columns represent the paternal alleles, whereas blue is for maternal alleles. Gel photos are RT-PCR products of 10 DAP endosperm digested by allele-specific restriction enzymes.

mammals could be regulated by a conserved system. In mammals, imprinting control regions (ICRs) within clusters have important regulatory functions (38). We identified a DMR region (overlapped with *ZmMMC-24*, around one PEG gene, *GRMZ2G406553*; Fig. 3) that is located in cluster 33 (*SI Appendix*, Table S7). This DMR region can potentially function as an ICR similar to that in mammals. Although only 13 of the identified imprinting clusters contain noncoding RNAs, we believe there could be more. Because our imprinting-screening pipeline relied on a relatively high expression level, imprinted noncoding RNAs could have been underestimated because of their relatively low level of expression (*SI Appendix*, Table S6). Also, some noncoding RNAs overlapping with exon and UTR regions of protein-coding genes could have been missed.

Using the similar clustering criteria in adjustment with their respective genome sizes (two imprinted genes within 54 kb in *Arabidopsis* and 182 kb in rice), 62 clusters in *Arabidopsis* and 55 clusters in rice can be identified by using all of the imprinted genes reported. Therefore, imprinting clustering can be a general feature in plants.

Potential Role of Gene Dosage Balancing for Imprinting in Maize Endosperm. Consistent with the parental conflict theory, there is enrichment of genes that function in certain biological processes, such as transport and localization, among the identified MEGs (*SI Appendix*, Fig. S3 and Table S3). However, many imprinted genes identified are apparently not directly involved in nutrient allocation, indicating that there must be other roles for imprinted genes in maize endosperm.

Previous studies suggested a potential role of transposable elements (TEs) in imprinting (15, 22, 25). In maize overall, there is no statistical difference in the appearance of TEs between the imprinted genes and nonimprinted genes, whereas a specific group of TEs (CACTA superfamily) is enriched around MEGs ($P = 3.46e^{-5}$) (*SI Appendix*, Table S9). However, the types of TEs shown to be related with imprinting in maize are not the same as in *Arabidopsis* (15). Additionally, several imprinted genes located in the syntenic chromosomal blocks in rice and maize are surrounded by totally different sets of TEs. Taken together, it seems that the contribution of TEs on imprinting in plants is very limited, even if they can affect specific subset of imprinted genes.

In our study, the total number of identified PEGs is larger than that of MEGs. Measuring parental allelic contributions based on high quality SNPs showed that there is a 40% overall paternal contribution in the 10 DAP endosperm transcriptome, a proportion that is significantly higher than the expected 33.3% (Pearson χ^2 test; 0.05) based on paternal genomic content in the triploid endosperm. We therefore propose that imprinting in higher plant endosperm may function in dosage compensation, a mechanism that has been discussed in *Arabidopsis* (39) and maize (40, 41). In other words, some genes are imprinted because of their sensitivity to gene dosage for proper endosperm function. Limited conservation of imprinting seen in rice, *Arabidopsis*, and maize may be due to different sets of dose-sensitive genes required for the endosperm development of these three species. Two conserved imprinted genes (*YUC10* and *VIM5*) were both reported to be regulated by the PRC2 complex in *Arabidopsis* (14), indicating that PRC2-mediated regulation of genetic imprinting is conserved between dicots and monocots.

DNA Methylation Is Involved in the Regulation of a Small Fraction of Imprinted Genes. DNA methylation has long been regarded as a key player in epigenetic regulation. Differential methylation has been shown for several known imprinted genes between the maternal and paternal alleles. Our results indicate that only 17.3% of the maize imprinted genes have differential methylation between the two parental alleles. The large number of imprinted genes showing no differential methylation between the parental alleles suggests that methylation is only involved in the

regulation of a small portion of imprinted genes. Surprisingly, both MEGs and PEGs having DMRs showed a pattern of maternal hypomethylation and paternal hypermethylation. In *Arabidopsis*, the *DME* gene functions in demethylation of maternal alleles for MEGs. The maternal hypomethylation of MEGs seen in maize suggests there could be a *DME*-like gene functioning in the maize central cell. Two genes in the maize genome (*GRMZM2G422464* and *GRMZM2G131756*) are highly homologous to DME in *Arabidopsis*, but their exact functions are still unknown. It is also possible that monocots use a different enzyme to mediate maternal demethylation (42). Because the previously identified imprinted genes are mostly MEGs, there is no previous report on parental differential methylation for PEGs. Silencing of the maternal allele of a PEG gene (*PHE1*) in *Arabidopsis* is reported to require demethylation of the maternal allele and depends on the PRC2 complex (43–45). All eight PEGs having DMRs are also maternally hypomethylated, suggesting that there could be an additional mechanism for the regulation of PEGs.

The large number of imprinted protein-coding genes, long noncoding RNAs, and their clustered distribution identified in this study clearly demonstrate complex parent-of-origin–dependent epigenetic regulation throughout maize endosperm development. Studies of the function and regulation of individual imprinted genes, including the noncoding RNAs within the imprinted clusters, will improve our understanding of the impact of genetic imprinting during maize endosperm development.

Materials and Methods

Tissue Collection and RNA Preparation. Seed and endosperm tissues from individual Mo17, B73, Mo17 × B73, and B73 × Mo17 ears were collected by manual dissection. Each sample was obtained from at least three ears of three different plants. Total RNA was extracted with TRIzol reagent (Invitrogen). Library construction and sequencing were performed according to Illumina instructions.

Identification of Imprinted Protein-Coding Genes and Noncoding RNAs in Maize 10 DAP Endosperm. To identify the imprinted SNPs, a two-tailed χ^2 test was conducted at each SNP site to test parental bias greater or less than 2:1 in both B73 × Mo17 and Mo17 × B73 ($\alpha = 0.05$). In a more stringent condition, a SNP was considered imprinted if the expression level of the actively expressed allele was at least five-times higher than the imprinted repressed allele. More detailed protocols for SNP discovery and imprinting genes identification are shown in *SI Appendix, SI Methods*.

Identification of Allele-Specific DMRs. Bisulfite sequencing was performed as described (46). Candidate allele-specific DMRs should fit the standard that the methylation level of one allele should be <30%, whereas the other allele is >70%. More detailed information is shown in *SI Appendix, SI Methods*.

ACKNOWLEDGMENTS. We thank Dr. Hugo Dooner and Dr. Philip W. Becraft for their critical reviews of the manuscript. This work is supported by National Basic Research Program (the "973" project) Grant 2009CB118400.

1. Huh JH, Bauer MJ, Hsieh T-F, Fischer RL (2008) Cellular programming of plant gene imprinting. *Cell* 132:735–744.
2. Luo M, Bilodeau P, Dennis ES, Peacock WJ, Chaudhury A (2000) Expression and parent-of-origin effects for FIS2, MEA, and FIE in the endosperm and embryo of developing Arabidopsis seeds. *Proc Natl Acad Sci USA* 97:10637–10642.
3. Grossniklaus U, Vielle-Calzada JP, Hoeppner MA, Gagliano WB (1998) Maternal control of embryogenesis by MEDEA, a polycomb group gene in Arabidopsis. *Science* 280:446–450.
4. Vielle-Calzada J-P, et al. (1999) Maintenance of genomic imprinting at the Arabidopsis medea locus requires zygotic DDM1 activity. *Genes Dev* 13:2971–2982.
5. Kinoshita T, Yadegari R, Harada JJ, Goldberg RB, Fischer RL (1999) Imprinting of the MEDEA polycomb gene in the Arabidopsis endosperm. *Plant Cell* 11:1945–1952.
6. Guitton A-E, et al. (2004) Identification of new members of Fertilisation Independent Seed Polycomb Group pathway involved in the control of seed development in Arabidopsis thaliana. *Development* 131:2971–2981.
7. Kiyosue T, et al. (1999) Control of fertilization-independent endosperm development by the MEDEA polycomb gene in Arabidopsis. *Proc Natl Acad Sci USA* 96:4186–4191.
8. Berger F, Chaudhury A (2009) Parental memories shape seeds. *Trends Plant Sci* 14: 550–556.
9. Raissig MT, Baroux C, Grossniklaus U (2011) Regulation and flexibility of genomic imprinting during seed development. *Plant Cell* 23:16–26.
10. Gregg C, et al. (2010) High-resolution analysis of parent-of-origin allelic expression in the mouse brain. *Science* 329:643–648.
11. Weaver JR, Susiarjo M, Bartolomei MS (2009) Imprinting and epigenetic changes in the early embryo. *Mamm Genome* 20:532–543.
12. Pandey RR, et al. (2008) Kcnq1ot1 antisense noncoding RNA mediates lineage-specific transcriptional silencing through chromatin-level regulation. *Mol Cell* 32:232–246.
13. Nagano T, et al. (2008) The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. *Science* 322:1717–1720.
14. Hsieh T-F, et al. (2011) Regulation of imprinted gene expression in Arabidopsis endosperm. *Proc Natl Acad Sci USA* 108:1755–1762.
15. Wolff P, et al. (2011) High-resolution analysis of parent-of-origin allelic expression in the Arabidopsis Endosperm. *PLoS Genet* 7:e1002126.
16. Gehring M, Missirian V, Henikoff S (2011) Genomic analysis of parent-of-origin allelic expression in Arabidopsis thaliana seeds. *PLoS ONE* 6:e23687.
17. Luo M, et al. (2011) A genome-wide survey of imprinted genes in rice seeds reveals imprinting primarily occurs in the endosperm. *PLoS Genet* 7:e1002125.
18. Kermicle JL (1970) Dependence of the R-mottled aleurone phenotype in maize on mode of sexual transmission. *Genetics* 66:69–85.
19. Gutierrez-Marcos JF, Pennington PD, Costa LM, Dickinson HG (2003) Imprinting in the endosperm: A possible role in preventing wide hybridization. *Philos Trans R Soc Lond B Biol Sci* 358:1105–1111.
20. Gutiérrez-Marcos JF, et al. (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.
21. Jahnke S, Scholten S (2009) Epigenetic resetting of a gene imprinted in plant embryos. *Curr Biol* 19:1677–1681.
22. Gehring M, Bubb KL, Henikoff S (2009) Extensive demethylation of repetitive elements during seed development underlies gene imprinting. *Science* 324:1447–1451.
23. Gutiérrez-Marcos JF, et al. (2006) Epigenetic asymmetry of imprinted genes in plant gametes. *Nat Genet* 38:876–878.
24. Haun WJ, et al. (2007) Genomic imprinting, methylation and molecular evolution of maize Enhancer of zeste (Mez) homologs. *Plant J* 49:325–337.
25. Hsieh T-F, et al. (2009) Genome-wide demethylation of Arabidopsis endosperm. *Science* 324:1451–1454.
26. Schnable PS, et al. (2009) The B73 maize genome: Complexity, diversity, and dynamics. *Science* 326:1112–1115.
27. Lai J, et al. (2010) Genome-wide patterns of genetic variation among elite maize inbred lines. *Nat Genet* 42:1027–1030.
28. Danilevskaya ON, et al. (2003) Duplicated fie genes in maize: Expression pattern and imprinting suggest distinct functions. *Plant Cell* 15:425–438.
29. Royo H, Cavaille J (2008) Non-coding RNAs in imprinted gene clusters. *Biol Cell* 100: 149–166.
30. Gambetta MC, Oktaba K, Müller J (2009) Essential role of the glycosyltransferase sxc/Ogt in polycomb repression. *Science* 325:93–96.
31. Brannan CI, Bartolomei MS (1999) Mechanisms of genomic imprinting. *Curr Opin Genet Dev* 9:164–170.
32. Guo M, Rupe MA, Danilevskaya ON, Yang X, Hu Z (2003) Genome-wide mRNA profiling reveals heterochronic allelic variation and a new imprinted gene in hybrid maize endosperm. *Plant J* 36:30–44.
33. Springer NM, et al. (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.
34. Bartolomei MS, Zemel S, Tilghman SM (1991) Parental imprinting of the mouse H19 gene. *Nature* 351:153–155.
35. Mitsuya K, et al. (1999) LIT1, an imprinted antisense RNA in the human KvLQT1 locus identified by screening for differentially expressed transcripts using monochromosomal hybrids. *Hum Mol Genet* 8:1209–1217.
36. Santoro F, Barlow DP (2011) Developmental control of imprinted expression by macro non-coding RNAs. *Semin Cell Dev Biol* 22:328–335.
37. Heo JB, Sung S (2011) Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA. *Science* 331:76–79.
38. Pauler FM, Koerner MV, Barlow DP (2007) Silencing by imprinted noncoding RNAs: Is transcription the answer? *Trends Genet* 23:284–292.
39. Tiwari S, et al. (2010) Transcriptional profiles underlying parent-of-origin effects in seeds of Arabidopsis thaliana. *BMC Plant Biol* 10:72.
40. Birchler JA, Bhadra U, Bhadra MP, Auger DL (2001) Dosage-dependent gene regulation in multicellular eukaryotes: Implications for dosage compensation, aneuploid syndromes, and quantitative traits. *Dev Biol* 234:275–288.
41. Birchler JA, Riddle NC, Auger DL, Veitia RA (2005) Dosage balance in gene regulation: Biological implications. *Trends Genet* 21:219–226.
42. Zemach A, et al. (2010) Local DNA hypomethylation activates genes in rice endosperm. *Proc Natl Acad Sci USA* 107:18729–18734.
43. Köhler C, Page DR, Gagliardini V, Grossniklaus U (2005) The Arabidopsis thaliana MEDEA Polycomb group protein controls expression of PHERES1 by parental imprinting. *Nat Genet* 37:28–30.
44. Köhler C, et al. (2003) The Polycomb-group protein MEDEA regulates seed development by controlling expression of the MADS-box gene PHERES1. *Genes Dev* 17:1540–1553.
45. Makarevich G, Villar CBR, Erilova A, Köhler C (2008) Mechanism of PHERES1 imprinting in Arabidopsis. *J Cell Sci* 121:906–912.
46. Lister R, et al. (2008) Highly integrated single-base resolution maps of the epigenome in Arabidopsis. *Cell* 133:523–536.