Title

Comprehensive analysis of imprinted genes in maize reveals allelic variation for imprinting and limited conservation with other species

Journal

Proceedings of the National Academy of Sciences of the United States of America, 110(48)

ISSN

0027-8424

Authors

Waters, Amanda J
Bilinski, Paul
Eichten, Steven R
et al.

Publication Date

2013-11-26

DOI

10.1073/pnas.1309182110

Peer reviewed

Comprehensive analysis of imprinted genes in maize reveals allelic variation for imprinting and limited conservation with other species

Amanda J. Waters[a], Paul Bilinski[b], Steven R. Eichten[a], Matthew W. Vaughn[c], Jeffrey Ross-Ibarra[b,d], Mary Gehring[e,f], and Nathan M. Springer[a,1]

[a]Microbial and Plant Genomics Institute and Department of Plant Biology, University of Minnesota, St. Paul, MN 55108; [b]Department of Plant Sciences and [d]The Genome Center and Center for Population Biology, University of California, Davis, CA 95616; [c]Texas Advanced Computing Center, University of Texas–Austin, Austin TX 78758; [e]Whitehead Institute for Biomedical Research, Cambridge, MA 02142; and [f]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139

Edited by Steven E. Jacobsen, University of California, Los Angeles, CA, and approved October 18, 2013 (received for review May 17, 2013)

In plants, a subset of genes exhibit imprinting in endosperm tissue such that expression is primarily from the maternal or paternal allele. Imprinting may arise as a consequence of mechanisms for silencing of transposons during reproduction, and in some cases imprinted expression of particular genes may provide a selective advantage such that it is conserved across species. Separate mechanisms for the origin of imprinted expression patterns and maintenance of these patterns may result in substantial variation in the targets of imprinting in different species. Here we present deep sequencing of RNAs isolated from reciprocal crosses of four diverse maize genotypes, providing a comprehensive analysis that allows evaluation of imprinting at more than 95% of endosperm-expressed genes. We find that over 500 genes exhibit statistically significant parent-of-origin effects in maize endosperm tissue, but focused our analyses on a subset of these genes that had >90% expression from the maternal allele (69 genes) or from the paternal allele (108 genes) in at least one reciprocal cross. Over 10% of imprinted genes show evidence of allelic variation for imprinting. A comparison of imprinting in maize and rice reveals that 13% of genes with syntenic orthologs in both species exhibit conserved imprinting. Genes that exhibit conserved imprinting between maize and rice have elevated nonsynonymous to synonymous substitution ratios compared with other imprinted genes, suggesting a history of more rapid evolution. Together, these data suggest that imprinting only has functional relevance at a subset of loci that currently exhibit imprinting in maize.

Imprinting describes a biased expression of alleles that depends upon the parent of origin. Imprinting is observed in both flowering plants and mammals (1–3) but there are differences in the mechanisms and organization of imprinted genes in these organisms (1, 4). In plants, imprinting is most prevalent in the endosperm, a triploid tissue that contains two maternal genomes and a single paternal genome (5). The endosperm provides an energy source for germinating seeds and, as the majority of harvested grain consists of endosperm tissue, a major source of calories in the human diet. A better understanding of imprinting will shed further light on epigenetic gene regulation and endosperm development and could provide an avenue for altering plant reproductive processes or seed quality.

Despite a widespread interest in imprinting and its potential importance, the function of most imprinted genes is not well characterized in plants, and imprinting has only recently been assayed on a genome-wide level. Imprinting is reflected in parentally biased allele-specific expression in the endosperm tissue of intraspecific reciprocal hybrids. A quantitative method for detecting the relative expression of two alleles that have nearly identical sequences is required to find such an effect, traditionally limiting analysis to a handful of imprinted genes identified based on phenotype or through targeted analyses (6–8). The implementation of deep sequencing of RNA molecules (RNA-seq) has allowed detection of additional imprinted genes (9–14). In each of these studies, allele-specific expression levels were monitored for a single cross of two parents in *Arabidopsis*, maize, or rice. This allowed for the analysis of imprinting in 50–58% of genes expressed in endosperm tissue. In each species there is evidence for several hundred imprinted genes with similar numbers of maternally expressed genes (MEGs) and paternally expressed genes (PEGs), but comparisons among flowering plants (8, 11, 12, 15) have revealed limited overlap in the genes that are imprinted among species.

There has been considerable speculation on the mechanisms that might lead to the origin of imprinted expression as well as the evolutionary mechanisms that would lead to the maintenance of imprinting (8, 16–18). Recent studies suggest that imprinting may arise due to programmed release of silencing marks in specific nuclei of male and female gametophytes (19, 20). Plant gametophytes are multinucleate structures. The male gametophyte includes a vegetative nucleus and two sperm nuclei. The female gametophyte has multiple cells including the haploid egg cell (which is fertilized by a sperm nuclei to generate the embryo) and the diploid central cell (which is fertilized by a sperm cell to generate the endosperm) (21). The loss of DNA methylation before fertilization leads to an epigenetic asymmetry in the endosperm because the maternal genomes (from the central cell)

Significance

In many eukaryotes, reproduction involves contributions of genetic material from two parents. At some genes there are parent-of-origin differences in the expression of the maternal and paternal alleles of a gene and this is referred to as imprinting. The analysis of allele-specific expression in several maize hybrids allowed the comprehensive detection of imprinted genes. By comparing allelic expression patterns in multiple crosses, it was possible to observe allelic variation for imprinting in maize. The comparison of genes subject to imprinting in multiple plant species reveals limited conservation for imprinting. The subset of genes that exhibit conserved imprinting in maize and rice may play important, dosage-dependent roles in regulation of seed development.

Author contributions: A.J.W., M.G., and N.M.S. designed research; A.J.W. and S.R.E. performed research; M.W.V. contributed new reagents/analytic tools; A.J.W., P.B., S.R.E., J.R.-I., M.G., and N.M.S. analyzed data; and A.J.W. and N.M.S. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The RNA-sequence reads reported in this paper have been deposited at the Sequence Read Archive (accession no. SRP031872).

[1]To whom correspondence should be addressed. E-mail: springer@umn.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1309182110/-/DCSupplemental.

have been demethylated, whereas the paternal genome (from a sperm nucleus) retains normal levels of methylation. Programmed DNA demethylation might result in the generation of siRNAs that could reinforce transposon silencing in adjacent cell types (egg and sperm cells) that contribute genetic material to the next generation (22). It has been hypothesized that this process, although targeted to transposons, could inadvertently influence nearby genes, resulting in imprinted expression (17). In support of this idea, several well-characterized imprinted genes contain transposon sequences in adjacent regions (6, 23–25,). The potential for transposons to contribute to the origin of imprinted expression of nearby genes may result in examples of imprinting that do not provide a selective advantage and would not be expected to persist over evolutionary time. Because imprinting at such loci would be of limited functional relevance and dependent on the presence of a transposable element, this could lead to substantial allelic variation for imprinting within a species. Indeed, several of the first characterized examples of imprinting in maize exhibit allelic variation such that certain alleles are imprinted whereas others are not (26, 27).

Regardless of the mechanisms that give rise to imprinted expression, parent-of-origin expression could, in some instances, provide a selective advantage. The kinship theory (16) suggests that MEGs should restrict growth or limit the flow of resources to offspring whereas PEGs might function to promote offspring growth. There are examples of imprinted genes that appear to exhibit these functions (28), but there is no clear evidence for these predicted functions in the annotations of the full set of previously identified MEGs or PEGs (3). Genes that are subject to parental conflict might be expected to exhibit signatures of positive selection (18, 29). For some imprinted genes, such as the *Arabidopsis* locus *MEDEA*, potential evidence of positive selection has been found in some cases (30, 31) but not others (32).

The presence of imprinting for a particular gene is often assumed to have functional relevance. Although this may be the case for a subset of genes, the potential for inadvertent acquisition of imprinting as a result of nearby transposon influences could result in numerous examples of imprinting that have limited functional relevance and thus show intra- or interspecific variation in imprinting. To distinguish between these possibilities and evaluate the functional importance of imprinting, we analyzed imprinting in multiple diverse genotypes of maize. Reciprocal crosses among four genotypes allowed for the surveying of imprinting at over 95% of the genes expressed in endosperm tissue. We find that only a subset of imprinted genes shows conserved imprinting in maize and rice and that these genes show evidence of distinct selective pressures. Comparison of imprinting in different haplotypes within maize reveals allelic variation for imprinting, further suggesting that imprinting may have limited functional consequence for many maize genes.

Results

Deep sequencing of RNA isolated from 14 days-after-pollination (DAP) endosperm tissue of five reciprocal hybrid pairs was performed to identify imprinted genes. This intermediate stage of endosperm development was selected because it is before major starch accumulation but after endosperm cellularization and because it reduces the potential for observation of stable transcripts contributed by the gametes. However, the analysis of one stage of endosperm development does not allow for assessment of transiently imprinted genes (33) or imprinting in embryo tissue (34). The five reciprocal hybrids we assayed included one previously analyzed dataset for the cross of inbred lines B73xMo17 (12), as well as four additional reciprocal hybrids generated by crossing inbred lines Ki11 and Oh43 with both B73 and Mo17 (Table 1). These additional genotypes were selected because whole-genome resequencing provided detailed SNP calls (35) and because they represent diverse genotypes (36).

A large number of reads (180–210 million) were recovered for each of the 10 genotypes and analyzed to study gene and allelic expression patterns (see *Methods* and Fig. S1 for details). The number of reads that mapped to each allele was summed across all SNPs for a transcript. Only transcripts that had at least 10 reads that could be assigned to a particular allele in each direction of the reciprocal cross were analyzed, resulting in allelic expression data for between 5,851 and 13,478 genes in each cross (Table 1 and Fig. S1). In total, 18,284 genes (95% of genes expressed in 14-DAP endosperm) had allele-specific expression data in at least one of the five reciprocal hybrid pairs (Table 1, Fig. 1*A*, and Fig. S2). Imprinted MEGs preferentially express the maternal allele in both directions of a cross, whereas imprinted PEGs express low levels of the maternal allele in both directions of the cross. Genes that exhibit consistent bias for the allele from one genotype, independent of parent of origin, reflect *cis*-regulatory allelic variation.

Comprehensive Discovery of Maize-Imprinted Genes. A combination of statistical significance and proportion filters was implemented to identify and classify MEGs and PEGs (Fig. S1). We assigned different levels of imprinting to parentally biased genes to compare imprinting strength within and between species in a more nuanced manner. Moderate MEGs/PEGs were defined as having significant allelic bias ($\chi^2 < 0.05$) and >80% of transcripts from the maternal allele (MEGs) or >60% of the transcripts from the paternal allele (PEGs) (red shaded areas in Fig. 1*A*) in both directions of a reciprocal cross. Strong MEGs and PEGs were defined as having significant allelic bias ($\chi^2 < 0.01$) and >90% of transcripts from the maternal allele (MEGs) or paternal allele (PEGs) (blue area in Fig. 1*A*). Complete MEGs or PEGs have >99% of the transcripts derived from the maternal or paternal allele, respectively. Genes with strong allelic bias (at least 95% reads from one allele) in one direction of the cross, but not in the reciprocal hybrid, are potentially representative of allelic variation for imprinting (green box in Fig. 1*A*).

Table 1. Discovery of maize-imprinted genes

Criteria	B73/Mo17	B73/Ki11	Mo17/Ki11	B73/Oh43	Mo17/Oh43	All	NR
No. of genes with ≥10 reads	11,856	10,531	5,851	13,478	9,434	2,087	18,284
Maternal bias	81	58	77	180	134	6	394
Moderate MEGs	75	42	22	118	44	4	198
Strong MEGs	31	28	9	39	25	4	69
Complete MEGs	13	9	3	16	12	3	37
Paternal bias	432	563	403	724	487	24	1,750
Moderate PEGs	171	192	74	191	120	18	367
Strong PEGs	56	55	24	76	45	6	108
Complete PEGs	8	17	3	15	5	0	31

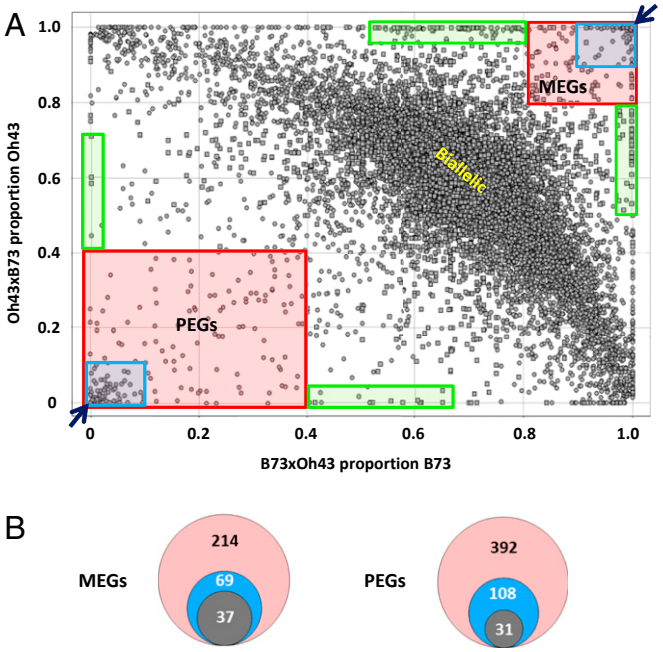

Fig. 1. Discovery of imprinted genes in maize. (*A*) Allele-specific expression analysis for the reciprocal F1 genotypes generated by crossing B73 and Oh43. The proportion of maternal transcripts in both reciprocal hybrids is plotted for the 13,478 genes that had at least 10 allelic reads in both directions of the cross of B73 and Oh43 (plots for other genotypes are in Fig. S2). The shaded areas indicate moderate (pink), strong (blue), or complete (arrows) MEGs (upper right) or PEGs (lower left). The green shaded areas indicate genes with potential allelic variation for imprinting. (*B*) The number of nonredundant moderate (pink), strong (blue), and complete (gray) MEGs and PEGs that were detected in at least one of the five reciprocal crosses are shown.

The number of genes classified as moderate, strong, or complete MEGs and PEGs varied for each genotype (Table 1 and Fig. 1*B*) in large part due to differences in the number of genes with polymorphisms. For the subsequent analyses, only the genes that were classified as strong or complete MEGs/PEGs were used. There are a total of 108 nonredundant, strong PEGs, including 31 examples (28%) that were classified as complete PEGs in at least one genotype (Table 1 and Dataset S1). Additional filtering criteria were applied to MEGs to remove genes that might exhibit maternal bias due to contamination of maternally derived tissues. RNA-seq data from a B73 expression atlas (37) were used to identify MEGs that may be the result of maternal contamination, resulting in a filtered list of 69 nonredundant, strong MEGs, with a larger number (37 or 54%) showing complete imprinting than seen in PEGs (Table 1 and Dataset S2).

Quantitative SNP assays designed using the Sequenom MassArray platform were used to validate imprinting for 13 MEGs and 13 PEGs (Tables S1 and S2). These assays are based on a single SNP for each gene and could only be used to assess imprinting in the crosses that were polymorphic for the targeted SNP. The analysis of allele-specific expression in a different 14-DAP endosperm sample for the same set of five reciprocal crosses confirmed imprinting in the majority of samples for both MEGs (23/24) and PEGs (28/28). The one allele that was not validated showed imprinted expression in one direction of the cross but biallelic expression in the reciprocal hybrid. The same quantitative SNP assays were also used to assess whether imprinting for these genes was also detected in several other genotypes (NC358, Ms71, and M162W) that were reciprocally crossed with B73 and Mo17. Most of these genes were imprinted in each of the other genotypes that were tested, with the exception of one locus (GRMZM2G020302) (Tables S1 and S2). Finally, the quantitative SNP assays were also used to assess whether imprinted expression was maintained at earlier and later stages of endosperm development. Imprinting was consistently observed for 26/26 MEGs and 25/26 PEGs at 12-DAP, 14-DAP, 16-DAP, and 20-DAP samples of B73xMo17, B73xNC358, and Mo17xNC358 (Tables S1 and S2). These data confirm that our RNA-seq data and subsequent imprinting analysis pipeline are highly reproducible.

The Expression of Imprinted Genes Is Endosperm Specific. Several plant-imprinted genes have expression that is restricted to the endosperm (18). This endosperm-specific expression could be because these genes have specific functions in the endosperm, or because it is beneficial to silence these genes in somatic tissues. Only a subset of MEGs and PEGs exhibited preferential expression in endosperm relative to other tissues in maize (Fig. S3 *A* and *B*). The majority of MEGs (68%) were preferentially expressed in endosperm, whereas only 26% of PEGs are preferentially expressed in endosperm (Fig. S3 *A* and *B*). Many MEGs exhibited increasing levels of expression during endosperm development, suggesting that these genes were actively transcribed in endosperm tissue as opposed to being stable, maternally inherited transcripts. There was no evidence that MEGs or PEGs exhibit unusually high or low expression levels; instead MEGs and PEGs exhibited a range of expression levels in endosperm tissue (Fig. S3*C*).

PEGs Are Often Targets of H3K27 Methylation. We previously (38) documented genome-wide H3K27me3 levels for five tissues of maize, including endosperm. We assessed the presence of H3K27me3 for the MEGs and PEGs identified in this study. Consistent with previous work (13, 38), we found that PEGs were more likely to be targets for histone methylation than MEGs. Only 5 of the 69 MEGs exhibited H3K27me3 in endosperm tissue (Dataset S2), in contrast to 87 of the 108 PEGs (Dataset S1). The 87 PEGs that were marked with H3K27me3 in endosperm tissue included 64 genes with expression in vegetative tissues and 23 genes with preferential expression in endosperm (Dataset S1). Only 8% of the 64 PEGs that were expressed in vegetative tissues exhibited H3K27me3 in the four vegetative tissues analyzed, whereas 65% of the 23 PEGs with preferential expression in endosperm were marked by H3K27me3 in at least three of the four vegetative tissues that were analyzed (Dataset S1).

PEGs Exhibit Higher Levels of Sequence Conservation. MEGs and PEGs also differ in their conservation between species and their annotation. The frequency of PEGs with syntenic orthologs in rice (39) was much higher (83%) than MEGs (46%) (Datasets S1 and S2). Similarly, the proportion of PEGs with high sequence similarity (E < 1E-50) to an *Arabidopsis* gene (61%) was higher than the proportion of MEGs (36%) (Datasets S1 and S2). Overrepresentation of functional categories of Gene Ontology (GO) annotations were investigated for the 66 PEGs and 23 MEGs that had high sequence similarity to *Arabidopsis* (E score <1E-50) using the Biological Networks Gene Ontology tool, BiNGO (40). PEGs exhibited significant ($P < 0.05$) enrichment for GO terms including flower development, chromatin modification, regulation of biological process, and DNA binding (Fig. S3*D*). MEGs exhibited significant ($P < 0.05$) enrichment for terms including DNA binding, transcription-factor activity, developmental process, and responses to stimulus (Fig. S3*E*).

Allelic Variation for Imprinting. Several of the earliest examples of imprinted loci exhibited imprinting for alleles from some genotypes but not others (26, 27). Our analysis of multiple maize genotypes provides an opportunity to comprehensively assess

allelic variation in imprinting (Fig. 2*A*). In general, when data were available for multiple crosses, many genes (88%) that exhibited imprinting in one cross were also imprinted in the other crosses, but there were examples in which genes imprinted in one cross displayed allelic variation for imprinting in another cross (Fig. 2 and Fig. S4). We identified 17 genes (8 PEGs and 9 MEGs) that showed consistent patterns of allelic variation in imprinting (Fig. 2*B* and Dataset S3). Gene GRMZM2G384780, for example, showed complete maternal allele expression in the Mo17/Oh43 cross, the B73 allele was not silent when it was inherited paternally (Fig. 2*C*). Similar variation was observed for other MEGs (Fig. S4) and PEGs (Fig. 2*D* and Fig. S4). The PEG GRMZM2G106222 showed expression of the maternal allele only when Oh43 was the maternal parent (Fig. 2*D*). A quantitative SNP assay was used to confirm the allele-specific imprinting for this gene (Fig. 2*D*). Overall, these data suggest standing allelic variation for imprinting is about 12% (17/144, the total is the number of genes with data in at least two sets of reciprocal crosses) of the imprinted genes even though we only assayed at most four haplotypes for each locus.

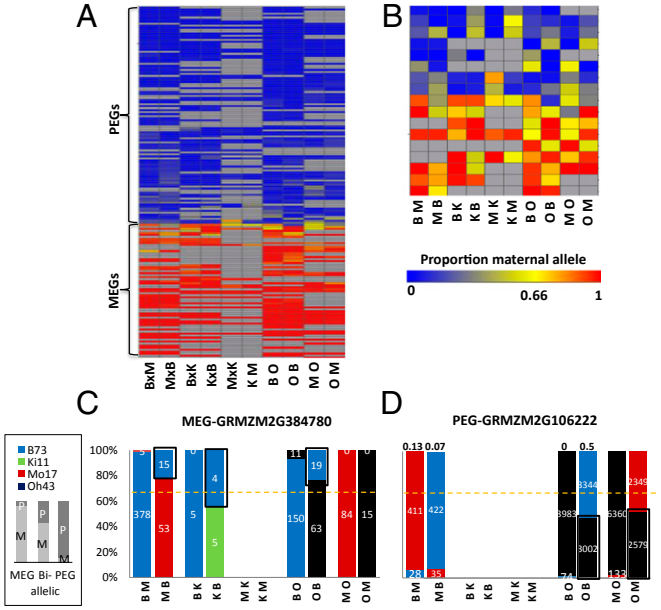

Fig. 2. Conservation of imprinting among maize haplotypes. (*A*) The proportion of expression from the maternal allele using a heat map (blue = 0; red = 1; yellow = 0.66; gray = missing data) is shown for all 10 genotypes for each of the nonredundant imprinted genes. The genotypes are abbreviated using the first letter of each parent, and the maternal parent is listed first. (*B*) A similar heat map is shown for the 17 genes with allelic variation for imprinting. (*C*) The expression patterns for one of the allele-specific imprinted MEGs (GRMZM2G384780) is shown. For this gene, the B73 allele is not silenced when paternally inherited, but alleles from the other haplotypes are silenced when inherited from the paternal parent. For each bar, the upper portion represents the proportion of paternal expression and the lower portion represents the proportion of maternal expression (see gray bars in key for expectations for MEGs, biallelic expression, and PEGs). The colors represent the four alleles assessed (see key for descriptions), and the values listed inside the bars are the number of maternal (M, upper portion of each bar) or paternal (P, lower portion of each bar) reads. The orange dashed line across the plot represents the expected biallelic ratio of 66% maternal reads. Black boxes highlight the nonimprinted allele. (*D*) Allele-specific imprinting pattern for the PEG GRMZM2G106222, which exhibits a failure to silence the Oh43 when it is maternally inherited. This gene was also validated by a quantitative SNP assay. SNPs were available to distinguish B73-Mo17 and B73-Oh43 alleles. The values listed above the bars are the proportion of the maternal allele determined from the quantitative SNP assay.

Conservation of Imprinting Between Species. If imprinting plays a similar functional role in all flowering plant species regardless of differences in endosperm growth or development, then it might be expected that there would be strong conservation for the targets of imprinting. Previous work has found only 5–10 examples of conserved imprinting between species (3, 11, 12), but has had limited comparative power due to the use of only a single cross in which not all genes may show polymorphism. The availability of a comprehensive list of MEGs and PEGs analyzed in multiple crosses in maize and information on syntenic gene relationships in rice allowed us to investigate the conservation of imprinting in monocots in more detail. There were 58 maize PEGs and 27 maize MEGs that had syntenic orthologs in rice that were assessed for imprinting by Luo et al. (11) in a cross between two haplotypes. Of these, 9 PEGs and 3 MEGs showed imprinting for both the maize and rice syntenic orthologs (Fig. S5*C*) and an additional 2 PEGs and 1 MEG that had imprinting for a closely related rice gene not located at a syntenic genomic position (Fig. S5*A*). This is a relatively low level of conservation but is significantly higher than expected by chance (χ^2, $P < 0.001$). There were also 3 moderate MEGs and 8 moderate PEGs that showed imprinting of their corresponding syntenic rice gene (Fig. S5*D*), and a low but statistically significant (χ^2, $P < 0.001$) level of conservation for imprinting of related sequences (not necessarily syntenic) in maize and *Arabidopsis* (Fig. S5*B*). Genes with conserved imprinting in maize and rice included a variety of annotations that comprised many genes with putative roles in transcriptional regulation or signaling pathways (Fig. S5 *C* and *D*). Two of these, encoding an ARID/BRIGHT DNA binding domain protein and a flavin-binding monooxygenase protein, also showed imprinting for related sequences in *Arabidopsis*. A recent report (41) provides evidence that one of these genes with conserved imprinting (GRMZM2G091819) plays an important role in influencing endosperm development in maize. Many of the genes with conserved imprinting have not been analyzed functionally but could play important roles in controlling endosperm development.

Finally, we analyzed the conservation of imprinting between paralogs from the recent whole-genome duplication event in maize. Following an allopolyploid whole-genome duplication event 5–12 Mya (42), subsequent rearrangements and fractionation have resulted in varying patterns of retention and loss (39, 43) of syntenic paralogs. A larger proportion of PEGs (73 genes or 68%) than MEGs (31 genes or 45%) were found in one of the two syntenic blocks assigned to subgenomes (Fisher's exact test two-tailed P value = 0.005; Datasets S1 and S2) (Table S3). The larger number of MEGs outside of syntenic blocks may have been due to recent duplication: 17% (12/69) of MEGs showed greater than 95% homology via BLAST to another gene in the genome compared with only 6% (6/108) of PEGs (Fisher's exact test two-tailed $P = 0.0037$). For those MEGs and PEGs found in either subgenome, both groups showed similar ratios of genes with retained syntenic duplicates (7/31 MEGs and 18/73 PEGs, $P = 1.0$) (Table S3). Of the 7 MEGs with retained duplicates in both subgenomes, two of the duplicates exhibited moderate imprinting, two were not imprinted but were expressed in the endosperm, and three were not expressed in the endosperm (Dataset S2). Among the 18 PEGs with retained duplicates, 10 were imprinted, 7 were expressed in the endosperm but not imprinted, and 1 was not expressed in the endosperm (Dataset S1).

Conserved Imprinted Genes Show Evidence of Positive Selection. To further investigate the evolution of imprinted loci, we took advantage of recent whole-genome analyses of maize and teosinte (44) to compare patterns of genetic diversity in imprinted and nonimprinted genes. Although kernel traits (including endosperm) have likely been selected during recent maize evolution, we found no evidence that imprinted loci were enriched in

regions targeted by selection during domestication or subsequent improvement (Table S4). Moreover, imprinted genes themselves showed few signs of selection, with values of nucleotide and haplotype diversity generally similar to genome-wide trends (Table S5). The only exception to this trend can be found in the paucity of high-frequency derived mutations seen in MEGs (median normalized Fay and Wu H = 1.11, Wilcoxon rank sum test P value = 0.0028), perhaps suggesting weak purifying selection.

We further evaluated the evolutionary importance of imprinted genes by comparing the ratio of nonsynonymous to synonymous substitutions (dN/dS) between maize, rice, and sorghum (Fig. 3). Genes with conserved imprinting show higher dN/dS values than both nonconserved imprinted genes (Wilcoxon rank, $P < 0.01$) and all genes tested (Wilcoxon rank, $P < 0.01$), although nonconserved imprinted genes differ from other tested genes in maize-rice and maize-sorghum comparisons (Wilcoxon rank, $P < 0.01$) (Fig. 3). Codon-based analysis of dN/dS in both conserved and nonconserved imprinted loci revealed the predominant effects of purifying selection across both classes of loci (Fig. S6). Eight of the twelve conserved imprinted genes and three nonconserved imprinted genes showed evidence of positive selection on at least one codon (Fig. S6*A*); masking these codons had little effect on overall dN/dS values.

Discussion

Our analysis of allele-specific expression in multiple crosses of maize provides a comprehensive study of imprinted genes in maize. Over 95% of the genes that were expressed in endosperm could be tested for imprinting due to the presence of polymorphisms in at least one of the crosses. Several hundred genes showed consistent parent-of-origin effects in at least one of the crosses. The availability of a relatively complete set of imprinted genes for maize provides an opportunity to examine the conservation of imprinting within and between species.

Imprinted genes are often treated as a single class in the literature. However, there are differences between MEGs and PEGs that suggest maternally and paternally biased expression may reflect distinct processes. MEGs are much more likely to exhibit endosperm-specific expression than PEGs, more likely to lack predicted function, and much less frequently associated with H3K27me3. There are also differences in the conservation of MEGs and PEGs between species. The maize PEGs have fewer recent duplications and are more likely to have a retained a syntenic ortholog in rice and a highly similar sequence in *Arabidopsis*. In addition, there are more examples of conserved imprinting in maize and rice, or between paralogs, for the PEGs.

The initial discovery of imprinting was based on studies of the *R* locus in maize (7), which exhibits allelic variation for imprinting (26). There are several other examples of potential allelic variation for imprinting in maize (33), but there have been few studies that assess imprinting for multiple alleles within a species. Our data reveal that over 10% of the genes with strong imprinting show allelic variation among the four maize haplotypes surveyed. This rate would undoubtedly increase if additional haplotypes were tested: we found at least one example of a gene for which the four alleles tested by RNA-seq were all imprinted, but the allele in at least one additional genotype tested by a gene-specific assay was not (Tables S1 and S2). This allelic variation in imprinting may reflect differences in transposon content near maize genes, although we have not yet been able to assess this. Studies of haplotype structure variation in maize (45) provide evidence for substantial allelic variation in the type of repetitive elements surrounding genes. Additional study of the specific haplotypes present at alleles that vary for imprinting may shed further light on the genetic or epigenetic changes that contribute to imprinted expression.

We investigated the conservation of imprinting between two monocots with persistent endosperm, maize and rice. In total we identified 88 imprinted maize genes with a syntenic rice gene evaluated by Luo et al. (11) but only 12 exhibit conserved imprinting. Although higher than expected by chance alone, this limited number suggests that conservation of imprinting over longer periods of evolutionary time is not common. It is important to note, however, that there are likely more than 12 examples of conserved imprinting in maize and rice because a number of loci, such as the rice ortholog of the maize MEG *Mez1* (32), could not be tested in rice due to a lack of polymorphisms (11). Genes with conserved imprinting tended to show elevated dN/dS ratios, and conserved imprinted genes showed greater evidence of positive selection on individual codons. Although these results are consistent with a functional role or perhaps even their involvement in genomic conflict, masking positively selected codons had little effect on overall dN/dS values and we cannot rule out weaker purifying selection as an alternative explanation. The limited conservation of imprinting among species also may help to guide future functional studies. Genes with conserved imprinting among species may play important functional roles in regulating seed development and growth and would be useful targets for reverse-genetic analysis. Future experiments to test the functional role of these genes in seed development are necessary. In contrast, the genes with imprinting only in certain species may reflect unique reproductive strategies in those species or represent imprinting that is not functionally relevant but is simply the result of inadvertent imprinting due to allelic differences such as transposon or epigenetic variation.

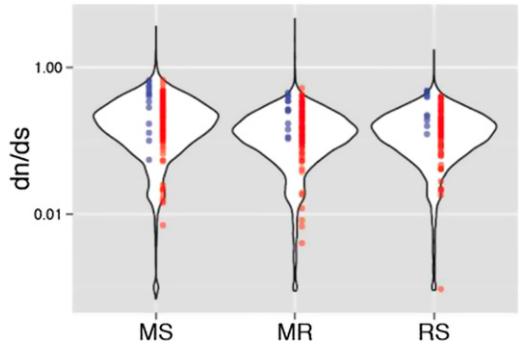

Fig. 3. Genes with conserved imprinting exhibit evidence for positive selection. Genes with conserved imprinting exhibit differential evidence of selection. dN/dS values for genome-wide comparisons of maize (M), rice (R), and sorghum (S). In each comparison, the width of the violin plot (white) represents the genome-wide distributions of dN/dS, red dots represent values for nonconserved imprinted genes in maize, and blue dots represent values for genes with conserved imprinting. Because imprinting data are not available in sorghum, the sorghum ortholog of maize-imprinted genes was used in the RS comparison.

Methods

RNA-Seq Analysis. Two ears of reciprocal F1 hybrid crosses of B73xMo17, B73xKi11, Mo17xKi11, B73xOh43, and Mo17xOh43 (Fig. 1) were collected 14 DAP; endosperm tissue was isolated by dissection; and RNA was extracted, purified, and submitted for sequencing using the Illumina HiSeq-2500 platform. Reads were aligned to the 39,540 genes in the filtered gene set (version 5b.60) using Tophat aligner (46), from which fragments per kilobase per million reads and allele-specific expression rates were calculated. Potential false SNPs were removed by requiring each SNP be supported by at least 1% of the reads at that position in each pair of reciprocal hybrids. The RNA-seq reads have been deposited at the Sequence Read Archive under accession no. SRP031872. See *SI Methods* for additional information.

Allelic Variation Detection. Allele-specific read counts or Sequenom data for each set of reciprocal crosses were analyzed to discover genes that exhibit

allelic variation of imprinting. Genes with at least 20 RNA-seq reads were run through a pipeline that pulls the maize gene ID for genes that showed allelic variation of imprinting in at least two sets of reciprocal crosses, and identifies which alleles are not imprinted at the locus.

Quantitative SNP Assays. Quantitative SNP assays (Sequenom MassArray) were used to validate imprinted genes and assess imprinting across additional genotypes or over a time course of seed development (see *SI Methods* for additional details).

Annotation and Comparative Genomics of Imprinted Genes. Maize syntenic orthologs in rice and retained whole-genome duplicates were identified by Schnable et al. (39).

Diversity and Divergence Analyses. Population genetic data from Hufford et al. (44) for a total of 14,982 (all genes with allelic expression data in endosperm tissue for at least one reciprocal cross) genes were included in our analysis, including 90 PEGs and 51 MEGs. Pairwise comparisons of dN/dS were made between syntenic genes in the genomes of *Zea mays* (v2, ID 11266), *Oryza sativa* Japonica (v7, ID 16890 masked), and *Sorghum bicolor* (v1.4 ID 95 masked repeats 50×) using the software SynMap and SynFind available through CoGe (47). To identify differences in patterns of evolution across codons of conserved imprinted genes, we performed fast, unconstrained bayesian approximation (48) analyses (see *SI Methods* for additional details).

ACKNOWLEDGMENTS. We thank Peter Hermanson for assisting with the crosses and tissue collections for this study; Ming Luo, Anna Koltunow, and Jen Taylor for sharing rice imprinting data; the University of Minnesota Biomedical Genomics Center for performing the Illumina sequencing for this study; and the Minnesota Supercomputing Institute and the Texas Advanced Computing Center for providing computational support. This work was created using resources or cyberinfrastructure provided by iPlant Collaborative. The iPlant Collaborative is funded by National Science Foundation (NSF) Grant DBI-0735191 (www.iplantcollaborative.org), and the research was supported by NSF Grant MCB-1121952 (to N.M.S. and M.G.).

1. Bartolomei MS, Ferguson-Smith AC (2011) Mammalian genomic imprinting. *Cold Spring Harb Perspect Biol* 3(7):a002592.
2. Gutierrez-Marcos JF, Constância M, Burton GJ (2012) Maternal to offspring resource allocation in plants and mammals. *Placenta* 33(Suppl 2):e3–e10.
3. Pignatta D, Gehring M (2012) Imprinting meets genomics: New insights and new challenges. *Curr Opin Plant Biol* 15(5):530–535.
4. Das R, Hampton DD, Jirtle RL (2009) Imprinting evolution and human health. *Mamm Genome* 20(9-10):563–572.
5. Li J, Berger F (2012) Endosperm: Food for humankind and fodder for scientific discoveries. *New Phytol* 195(2):290–305.
6. Gehring M, Bubb KL, Henikoff S (2009) Extensive demethylation of repetitive elements during seed development underlies gene imprinting. *Science* 324(5933): 1447–1451.
7. Kermicle JL (1970) Dependence of the R-mottled aleurone phenotype in maize on mode of sexual transmission. *Genetics* 66(1):69–85.
8. Köhler C, Wolff P, Spillane C (2012) Epigenetic mechanisms underlying genomic imprinting in plants. *Annu Rev Plant Biol* 63:331–352.
9. Hsieh TF, et al. (2011) Regulation of imprinted gene expression in Arabidopsis endosperm. *Proc Natl Acad Sci USA* 108(5):1755–1762.
10. Gehring M, Missirian V, Henikoff S (2011) Genomic analysis of parent-of-origin allelic expression in Arabidopsis thaliana seeds. *PLoS ONE* 6(8):e23687.
11. 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(6):e1002125.
12. Waters AJ, et al. (2011) Parent-of-origin effects on gene expression and DNA methylation in the maize endosperm. *Plant Cell* 23(12):4221–4233.
13. Wolff P, et al. (2011) High-resolution analysis of parent-of-origin allelic expression in the Arabidopsis Endosperm. *PLoS Genet* 7(6):e1002126.
14. Zhang M, et al. (2011) Extensive, clustered parental imprinting of protein-coding and noncoding RNAs in developing maize endosperm. *Proc Natl Acad Sci USA* 108(50): 20042–20047.
15. Jiang H, Köhler C (2012) Evolution, function, and regulation of genomic imprinting in plant seed development. *J Exp Bot* 63(13):4713–4722.
16. Haig D (2004) Genomic imprinting and kinship: How good is the evidence? *Annu Rev Genet* 38:553–585.
17. Köhler C, Weinhofer-Molisch I (2010) Mechanisms and evolution of genomic imprinting in plants. *Heredity (Edinb)* 105(1):57–63.
18. Berger F, Vu TM, Li J, Chen B (2012) Hypothesis: Selection of imprinted genes is driven by silencing deleterious gene activity in somatic tissues. *Cold Spring Harb Symp Quant Biol* 77:23–29.
19. Feng S, Jacobsen SE, Reik W (2010) Epigenetic reprogramming in plant and animal development. *Science* 330(6004):622–627.
20. Bauer MJ, Fischer RL (2011) Genome demethylation and imprinting in the endosperm. *Curr Opin Plant Biol* 14(2):162–167.
21. Huh JH, Bauer MJ, Hsieh TF, Fischer RL (2008) Cellular programming of plant gene imprinting. *Cell* 132(5):735–744.
22. Rodrigues JA, et al. (2013) Imprinted expression of genes and small RNA is associated with localized hypomethylation of the maternal genome in rice endosperm. *Proc Natl Acad Sci USA* 110(19):7934–7939.
23. Zemach A, et al. (2010) Local DNA hypomethylation activates genes in rice endosperm. *Proc Natl Acad Sci USA* 107(43):18729–18734.
24. Kinoshita Y, et al. (2007) Control of FWA gene silencing in Arabidopsis thaliana by SINE-related direct repeats. *Plant J* 49(1):38–45.
25. Lippman Z, et al. (2004) Role of transposable elements in heterochromatin and epigenetic control. *Nature* 430(6998):471–476.
26. Kermicle JL, Alleman M (1990) Gametic imprinting in maize in relation to the angiosperm life cycle. *Dev Suppl* pp 9–14.
27. Chaudhuri S, Messing J (1994) Allele-specific parental imprinting of dzr1, a post-transcriptional regulator of zein accumulation. *Proc Natl Acad Sci USA* 91(11): 4867–4871.
28. Costa LM, et al. (2012) Maternal control of nutrient allocation in plant seeds by genomic imprinting. *Curr Biol* 22(2):160–165.
29. Hutter B, Bieg M, Helms V, Paulsen M (2010) Divergence of imprinted genes during mammalian evolution. *BMC Evol Biol* 10:116. Available at www.biomedcentral.com/1471-2148/10/116.
30. Spillane C, et al. (2007) Positive darwinian selection at the imprinted MEDEA locus in plants. *Nature* 448(7151):349–352.
31. Miyake T, Takebayashi N, Wolf DE (2009) Possible diversifying selection in the imprinted gene, MEDEA, in Arabidopsis. *Mol Biol Evol* 26(4):843–857.
32. Haun WJ, et al. (2007) Genomic imprinting, methylation and molecular evolution of maize Enhancer of zeste (Mez) homologs. *Plant J* 49(2):325–337.
33. Springer NM, Gutierrez-Marcos JF (2008) Imprinting in maize. *The Maize Handbook*, eds Hake S, Bennetzen J (Springer, New York).
34. Jahnke S, Scholten S (2009) Epigenetic resetting of a gene imprinted in plant embryos. *Curr Biol* 19(19):1677–1681.
35. Chia JM, et al. (2012) Maize HapMap2 identifies extant variation from a genome in flux. *Nat Genet* 44(7):803–807.
36. Yu J, Holland JB, McMullen MD, Buckler ES (2008) Genetic design and statistical power of nested association mapping in maize. *Genetics* 178(1):539–551.
37. Sekhon RS, et al. (2013) Maize gene atlas developed by RNA sequencing and comparative evaluation of transcriptomes based on RNA sequencing and microarrays. *PLoS ONE* 8(4):e61005.
38. Makarevitch I, et al. (2013) Genomic distribution of maize facultative heterochromatin marked by trimethylation of H3K27. *Plant Cell* 25(3):780–793.
39. Schnable JC, Freeling M, Lyons E (2012) Genome-wide analysis of syntenic gene deletion in the grasses. *Genome Biol Evol* 4(3):265–277.
40. Maere S, Heymans K, Kuiper M (2005) BiNGO: A Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* 21(16):3448–3449.
41. Bernardi J, et al. (2012) Impaired auxin biosynthesis in the defective endosperm 18 mutant is due to mutational loss of expression in the ZmYuc1 gene encoding endosperm-specific YUCCA1 protein in maize. *Plant Physiol* 160(3):1318–1328.
42. Gaut BS, Doebley JF (1997) DNA sequence evidence for the segmental allotetraploid origin of maize. *Proc Natl Acad Sci USA* 94(13):6809–6814.
43. Woodhouse MR, et al. (2010) Following tetraploidy in maize, a short deletion mechanism removed genes preferentially from one of the two homologs. *PLoS Biol* 8(6):e1000409.
44. Hufford MB, et al. (2012) Comparative population genomics of maize domestication and improvement. *Nat Genet* 44(7):808–811.
45. Messing J, Dooner HK (2006) Organization and variability of the maize genome. *Curr Opin Plant Biol* 9(2):157–163.
46. Trapnell C, et al. (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol* 28(5):511–515.
47. Lyons E and Freeling M (2008) How to usefully compare homologous plant genes and chromosomes as DNA sequences. *Plant J* 53(4):661–73.
48. Murrell B, et al. (2013) FUBAR: A fast, unconstrained bayesian approximation for inferring selection. *Mol Biol Evol* 30(5):1196–1205.