Divergence among rice cultivars reveals roles for transposition and epimutation in ongoing evolution of genomic imprinting

Jessica A. Rodrigues[a,1], Ping-Hung Hsieh[a,1,2], Deling Ruan[b], Toshiro Nishimura[a], Manoj K. Sharma[b,3], Rita Sharma[b,4], XinYi Ye[a], Nicholas D. Nguyen[a], Sukhranjan Nijjar[a], Pamela C. Ronald[b,c], Robert L. Fischer[a,5], and Daniel Zilberman[a,d,5]

[a]Department of Plant and Microbial Biology, University of California, Berkeley, CA 94720; [b]Department of Plant Pathology, University of California, Davis, CA 95616; [c]The Genome Center, University of California, Davis, CA 95616; and [d]Department of Cell and Developmental Biology, The John Innes Centre, Norwich NR4 7UH, United Kingdom

Contributed by Robert L. Fischer, June 3, 2021 (sent for review March 11, 2021; reviewed by Sarah N. Anderson and Claudia Köhler)

Parent-of-origin–dependent gene expression in mammals and flowering plants results from differing chromatin imprints (genomic imprinting) between maternally and paternally inherited alleles. Imprinted gene expression in the endosperm of seeds is associated with localized hypomethylation of maternally but not paternally inherited DNA, with certain small RNAs also displaying parent-of-origin–specific expression. To understand the evolution of imprinting mechanisms in *Oryza sativa* (rice), we analyzed imprinting divergence among four cultivars that span both *japonica* and *indica* subspecies: Nipponbare, Kitaake, 93-11, and IR64. Most imprinted genes are imprinted across cultivars and enriched for functions in chromatin and transcriptional regulation, development, and signaling. However, 4 to 11% of imprinted genes display divergent imprinting. Analyses of DNA methylation and small RNAs revealed that endosperm-specific 24-nt small RNA–producing loci show weak RNA-directed DNA methylation, frequently overlap genes, and are imprinted four times more often than genes. However, imprinting divergence most often correlated with local DNA methylation epimutations (9 of 17 assessable loci), which were largely stable within subspecies. Small insertion/deletion events and transposable element insertions accompanied 4 of the 9 locally epimutated loci and associated with imprinting divergence at another 4 of the remaining 8 loci. Correlating epigenetic and genetic variation occurred at key regulatory regions—the promoter and transcription start site of maternally biased genes, and the promoter and gene body of paternally biased genes. Our results reinforce models for the role of maternal-specific DNA hypomethylation in imprinting of both maternally and paternally biased genes, and highlight the role of transposition and epimutation in rice imprinting evolution.

chromatin modification | epigenome | transcriptome | bisulfite sequencing | siren loci

Genomic imprinting results in biased expression from one parentally inherited allele, even though both alleles are present in the same cell and may be genetically identical. Here, maternally and paternally inherited alleles are differentiated postfertilization by a mark that is epigenetic or, in other words, not directly related to DNA sequence. This epigenetic mark is referred to as the "imprint" (1). To date, locus-specific imprinting of largely nonimprinted chromosomes is only observed in species with a "placental habit," namely placental mammals and flowering plants (2). The phenomenon appears to have evolved independently in both clades, around the same time as the placenta in mammals (3) and its equivalent, the endosperm, in plants (2). Imprinting still predominantly occurs in placental tissue (4–9) and seems crucial for offspring development (4, 10, 11). Several imprinted genes are important regulators of development (1, 4, 12).

Mammalian imprinting appears to have evolved via the spread of long terminal repeat (LTR) retrotransposons (3). Though the origins of plant imprinting are less clear, current evidence also suggests a role for transposable elements (TEs) in imprint initiation (1, 13–16). Not only do TEs create novel genetic variation at genes and promoters but TE-targeted chromatin modifications can affect transcription of neighboring genes (17, 18). Chromatin modification of nearby TE-related sequences is required for parent-of-origin–specific expression of several imprinted genes (7, 19–23). Additionally, allele-specific imprinting variation among *Arabidopsis thaliana* ecotypes has been associated with epigenetic variation at TEs when the underlying genetic sequences show little sequence divergence, suggesting that epimutation plays an additional role in the ongoing evolution of imprinting (7, 24).

Studies in *A. thaliana*, maize, and rice indicate that DNA methylation and histone 3 lysine 27 trimethylation (H3K27me3)

Significance

Plant gene imprinting (allele-specific chromatin modification) results in parent-of-origin–dependent expression in endosperm, a nutritive tissue essential for embryo viability. Imprinted genes function in species reproductive isolation by causing embryo abortion in crosses. We analyzed the evolution of rice gene imprinting using four cultivars spanning domestication and the divergence of two subspecies 9,000 and 300,000 y ago, respectively. Most imprinted genes are imprinted across cultivars and enriched in regulatory functions. However, approximately 10% of imprinted genes have lost or gained imprinting, often associated with stably inherited epigenetic and genetic variation, suggesting a role in rice diversification. Our results highlight the role of transposable elements and epigenetic variation in shaping heritable changes in gene expression during rice evolution.

Author contributions: J.A.R., P.C.R., R.L.F., and D.Z. designed research; J.A.R., P.-H.H., D.R., X.Y., N.D.N., and S.N. performed research; J.A.R., P.-H.H., T.N., M.K.S., and R.S. analyzed data; and J.A.R., P.-H.H., R.L.F., and D.Z. wrote the paper.

Reviewers: S.N.A., Iowa State University; and C.K., Max-Planck-Institut fur Molekulare Pflanzenphysiologie.

The authors declare no competing interest.

This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

[1]J.A.R. and P.-H.H. contributed equally to this work.

[2]Present address: The US Department of Energy Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA 94720.

[3]Present address: School of Biotechnology, Jawaharlal Nehru University, New Delhi 110067, India.

[4]Present address: Department of Biological Sciences, Birla Institute of Technology & Science, Pilani 333031, India.

[5]To whom correspondence may be addressed. Email: rfischer@berkeley.edu or daniel.zilberman@jic.ac.uk.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2104445118/-/DCSupplemental.

Published July 16, 2021.

Downloaded at Institute Science Tech Austria on August 11, 2021

are the major imprints for endosperm-expressed genes (5, 16, 20, 25–27). Plant endosperm is a triploid tissue, formed by fusion of a homodiploid female central cell and haploid male sperm cell in a fertilization event independent of but concurrent with that which creates the embryo (28, 29). DNA methylation-based imprints are established prefertilization by glycosylase-mediated DNA demethylation in the central cell and DNA methylation maintenance in developing sperm cells (30–35). Thus, they are observed as site-specific hypomethylation of maternally inherited but not paternally inherited endosperm DNA (26, 36, 37) and considered to be a cause rather than a consequence of imprinting (38). In mammals, differential DNA methylation is considered the canonical cause of imprinting (39, 40), with H3K27me3 serving as a non-canonical mechanism at a subset of paternally expressed genes (41). Both mechanisms serve to maintain mammalian imprints postfertilization (40, 42).

Like mammalian H3K27me3 imprints, plant H3K27me3 imprints are initiated and maintained by Polycomb group (PcG) proteins and are more often associated with paternally expressed than maternally expressed imprinted genes (5, 25, 43–46). While some endosperm H3K27me imprints are triggered by DNA hypomethylation (22, 25, 47, 48), others are observed in its absence, suggesting either that the causal imprint is H3K27me3 (49) or a yet undiscovered epigenetic mark, or that the genes are regulated by unknown long-distance regulatory elements, or that their endosperm epigenetic state does not reflect gametic imprints (50). Small RNAs (sRNAs) are also associated with imprint formation at some loci in the endosperm of *A. thaliana*, maize, and rice, such that maternally or paternally biased sRNA expression overlaps the silenced allele (12, 26, 43, 51, 52).

Here, we further dissect mechanisms central to plant imprinting and evolution in rice, a staple food crop for much of the world. In order to identify the contributions of genetic and epigenetic variation in ongoing divergence of rice cultivars, we identified divergence in imprinted expression across Kitaake, Nipponbare, 93-11, and IR64 rice cultivars of *Oryza sativa* L., using reciprocal crosses and cultivar-specific single-nucleotide polymorphisms (SNPs). By exploring differences in DNA methylation and sRNA production among cultivars, we find that genes with diverged imprinting differ in DNA methylation of key regulatory regions, and that these changes in DNA methylation may or may not be associated with genetic changes such as the transposition of class I or class II elements.

Results

Approximately 10% of Genes Expressed in Rice Endosperm Are Imprinted, with About 4 to 11% of Those Genes Displaying Imprinting Divergence across Four Rice Cultivars. We performed one reciprocal set of intra-*japonica* crosses (i.e., Nipponbare × Kitaake and Kitaake × Nipponbare), one set of intra-*indica* crosses (i.e., IR64 × 93-11 and 93-11 × IR64), and two sets of intersubspecies crosses (Nipponbare × IR64 and IR64 × Nipponbare, and Nipponbare × 93-11 and 93-11 × Nipponbare). Genome-wide comparison of identified SNPs revealed that Nipponbare and Kitaake *japonica* cultivars are the most closely related of the four, with the next most closely related, 93-11 and IR64 *indica* cultivars, differing at 3 times as many nucleotide positions while *indica* and *japonica* cultivars differ at 10 times as many positions (*SI Appendix*, Table S1 and Dataset S1). F1 seeds dissected at day 7 to 8 after pollination were genotyped to verify hybridity prior to endosperm and embryo RNA extraction for strand-specific RNA-sequencing (RNA-seq) libraries. Duplex-specific nuclease treatment of endosperm libraries enhanced transcript detection without altering the ratio of maternal SNPs to paternal SNPs (*SI Appendix*, Fig. S1).

Though 35 to 42% of annotated rice genes were detected in endosperm libraries, only ~20% of annotated genes had sufficient read depth at informative SNPs for reliable detection of expression biases in intersubspecies crosses (Fig. 1 *A* and *B*). Only 2 to 6% of annotated genes (5 to 17% of detectable transcripts) could be evaluated in intrasubspecies crosses. Cultivar-specific expression bias prevented clear imprinting inference at a further third of the transcripts. Of the remaining genes, which are the genes truly assessable for imprinting, parental biases varied along a continuum, as previously noted in rice (6). We found that roughly 10% of assessable genes appear to be imprinted in a moderate, strong, or complete manner, as defined by significant deviation (Fisher's exact test, $P < 0.01$) of the maternally derived fraction of sorted reads from the expected ratio of 0.67 (Fig. 1*A*). Having identified

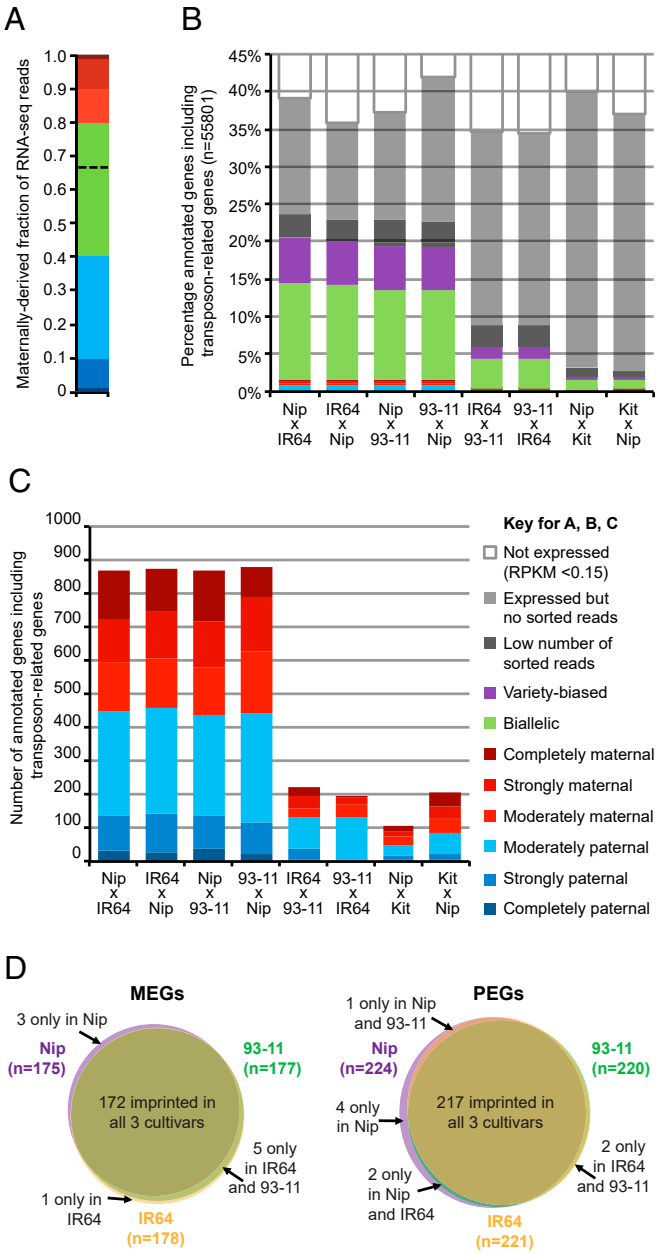

Fig. 1. Detection of imprinted genes in reciprocal crosses. (*A*) Imprinted gene expression was defined as significant deviation (Fisher's exact test, $P < 0.01$) from the biallelic parental ratio for endosperm (dashed line), and classified as moderate, strong (over 90% uniparental), or complete (over 99% uniparental). (*B*) Genes were grouped based on assessment of parental biases. (*C*) Hundreds of genes exhibited significant parental biases. (*D*) For MEGs and PEGs where imprinting could be assessed across Nipponbare, 93-11, and IR64 cultivars, imprinting was more often conserved than diverged.

Downloaded at Institute Science Tech Austria on August 11, 2021

several hundred candidate imprinted genes (Fig. 1*C*), we inferred imprinting conservation and divergence from reciprocal cross data as described in *SI Appendix*, Fig. S2, resolving many initial candidates for imprinting divergence to technical artifacts. Finally, we defined a set of 217 conserved paternally expressed genes (cPEGs), 175 conserved maternally expressed genes (cMEGs), and 19 genes with diverged imprinting (Fig. 1*D*, *SI Appendix*, Tables S2 and S3, and Dataset S2). Our list of conserved imprinted rice genes overlaps the majority of those from other groups (*SI Appendix*, Fig. S3). Differences are likely due to variation in SNPs, cultivars, analytical pipelines, developmental stages, and growth conditions. Of the 19 genes with divergent imprinting, variation between imprinted and biallelic states is the most common form of divergence, with only one gene varying between maternally biased and paternally biased expression.

Functional Gene Ontology Term Enrichment of Imprinted Genes Reveals Regulatory Functions. cMEGs and cPEGs were functionally annotated with the agriGO toolkit (53), and gene ontology (GO) term enrichment was compared with the pool of assessable endosperm genes. cPEGs were similar to those in both maize (54) and *A. thaliana* (7) in being enriched for functions in regulation, including that of chromatin and transcription (*SI Appendix*, Fig. S4*A*). cMEGs were enriched for functions in transcriptional regulation, development, and signaling (*SI Appendix*, Fig. S4*B*), and thus similar to those in maize (54) but not to those in *A. thaliana*, where only a weak enrichment for transcription factors was seen (7). However, the lack of GO annotations for 38% of cMEGs, compared with just 16% of cPEGs, suggests that the enrichment analysis may not be as illustrative of the general roles of cMEGs. Twenty of the 175 MEGs (11%) lacked GO annotations because they were TE genes (11 retrotransposons, 4 Pong MITE DNA transposons, 2 MuDR DNA transposons, 1 CACTA DNA transposon, and 2 unclassified). By contrast, only one of the 217 PEGs encoded an annotated TE protein (an LTR retrotransposon). Although the list of genes with diverged imprinting was too small to draw definitive conclusions about GO term enrichment, both MEGs and PEGs appeared enriched for similar functions compared with their conserved counterparts, with roles in transcription, chromatin regulation, protein modification, and protein–protein interactions for PEGs, and transposition and chromatin and transcriptional regulation for MEGs (*SI Appendix*, Tables S2 and S3).

sRNA Populations in Rice Endosperm Are Similar across Cultivars and Are Four Times More Likely to Be Imprinted than Genes. Having previously described a unique sRNA population in the endosperm of Kitaake and Nipponbare *japonica* cultivars (26), we performed strand-specific sequencing of sRNA (*SI Appendix*, Tables S4 and S5) from dissected endosperm and embryos of 7- to 8-d-old seeds of IR64 and 93-11 rice cultivars, as well as the reciprocal crosses used for detection of imprinted genes. Endosperm sRNA populations of IR64 and 93-11 *indica* cultivars also predominantly originate from the relatively small number of euchromatic loci we identified in Kitaake and Nipponbare. These "siren" loci (small-interfering RNA in endosperm) are often associated with high sRNA abundance in all four cultivars (*SI Appendix*, Figs. S5–S7), but a minority show divergent levels of expression among cultivars (Fig. 2*A*). Of 801 defined siren loci (Dataset S3), imprinting was observed for 41% of the 186 with significant read depth at informative SNPs and a lack of confounding variety-specific bias in crosses of at least three cultivars (*SI Appendix*, Fig. S8). More siren loci were paternally biased than maternally biased, with 27% compared with 11% (Fig. 2*B*).

Although they appear to originate from distinct noncoding transcripts, most siren loci reside within 5 kb of genes (Fig. 2*C* and *SI Appendix*, Figs. S5–S7 and S9), with the 801 imprinted and nonimprinted siren loci overlapping the bodies of 519 genes. Of these 519 genes, we assessed parental bias at the 112 with good read depth over SNPs lacking variety-specific bias and found 40 (36%) were imprinted (22 MEGs and 18 PEGs). This is higher than expected by chance as only 10% of endosperm-expressed genes are estimated to be imprinted. Wherever informative SNPs indicated conserved imprinting of both siren loci and the genes they overlapped, we observed perfectly opposed parental biases of siren loci and overlapping genes (Fig. 2*D*), supporting our previous results (26). Only six siren loci (3.2%) appeared to differ in parental bias among rice cultivars (Fig. 2*B* and *SI Appendix*, Fig. S8 and Table S6). Of these six siren loci, three overlapped genes but only two overlapped genes with detectable transcripts. Imprinting of the siren locus at MEG Os02g15350 correlated with gene imprinting while that of the siren locus at MEG Os08g28960 did not (*SI Appendix*, Table S6), though it is possible that the lack of correlation at Os08g28960 is due to cultivar-specific bias or higher mismapping of 24-nt sRNA reads compared with RNA-seq reads.

Overall, the correlation seen at conserved imprinted siren loci, along with the finding that biallelic siren loci tend to overlap poorly expressed genes (Fig. 2*D*), indicates that siren locus expression associates with gene silencing. Observation of endosperm-specific RNA-directed DNA methylation (RdDM) due to siren-derived sRNA is complicated by low levels of endosperm non-CG DNA methylation relative to other tissues (55) as well as endosperm-specific maternal hypomethylation (*SI Appendix*, Figs. S9 and S10). After attempting to resolve such complications, we find that CHG and CHH methylation vary differently across genomic features of seed tissues (*SI Appendix*, Figs. S11 and S12) but CHH methylation is significantly increased at siren loci and decreased at other endosperm 24-nt sRNA-producing loci compared with embryo (Fig. 2*E*), suggestive of siren-mediated RdDM. Surprisingly, CHH methylation of imprinted siren loci shows an opposite direction of parent-of-origin–specific bias compared with sRNAs, with paternally biased sRNAs associated with maternal hypermethylation and maternally biased sRNAs associated with paternal hypermethylation, though data fell short of statistical significance (Fig. 2*E* and *SI Appendix*, Fig. S12).

DNA Methylation Profiles Are Highly Similar among Cultivars, Associating Imprinting with Maternal-Specific DNA Hypomethylation. We previously obtained DNA methylation data from dissected endosperm and embryos of 7- to 8-d-old seeds of Nipponbare, Kitaake, and their reciprocal crosses (26). Using the same methods for IR64 and 93-11 rice cultivars (*SI Appendix*, Table S7), we find that, like sRNA populations, DNA methylation features are largely conserved among the four rice cultivars (Fig. 3*A* and *SI Appendix*, Figs. S6, S9, S10, and S13). Furthermore, patterns of endosperm-specific maternal hypomethylation, indicated by differential methylation between embryo and endosperm (26), are conserved (*SI Appendix*, Fig. S14) even more so than general DNA methylation (Fig. 3*A* and *SI Appendix*, Figs. S6, S9, and S13).

We further analyzed our previous finding that PEGs are enriched for endosperm-specific maternal DNA hypomethylation in the gene body and promoter, while MEGs are enriched for hypomethylation in the promoter, transcription termini, and 3′ downstream region (26) (*SI Appendix*, Figs. S15 and S16). As many endosperm transcripts differed in transcription start site (TSS) compared with reference gene annotations, we generated a curated set of endosperm-expressed transcription termini, updated using de novo assembled transcripts (*SI Appendix*, *Materials and Methods* and Dataset S5). We then used Nipponbare data to categorize individual cMEGs, cPEGs, and conserved biallelic genes into five embryo–endosperm DNA hypomethylation patterns: TSS, gene body, promoter, 3′ downstream, and a final pattern where none of the sequence within 1 kb of the gene is hypomethylated (Fig. 3*B*). Though all five patterns were observed within each subgroup, cMEGs and cPEGs show significantly greater overlap with regions of embryo–endosperm DNA hypomethylation than do biallelic genes (Fisher's

PLANT BIOLOGY

Downloaded at Institute Science Tech Austria on August 11, 2021

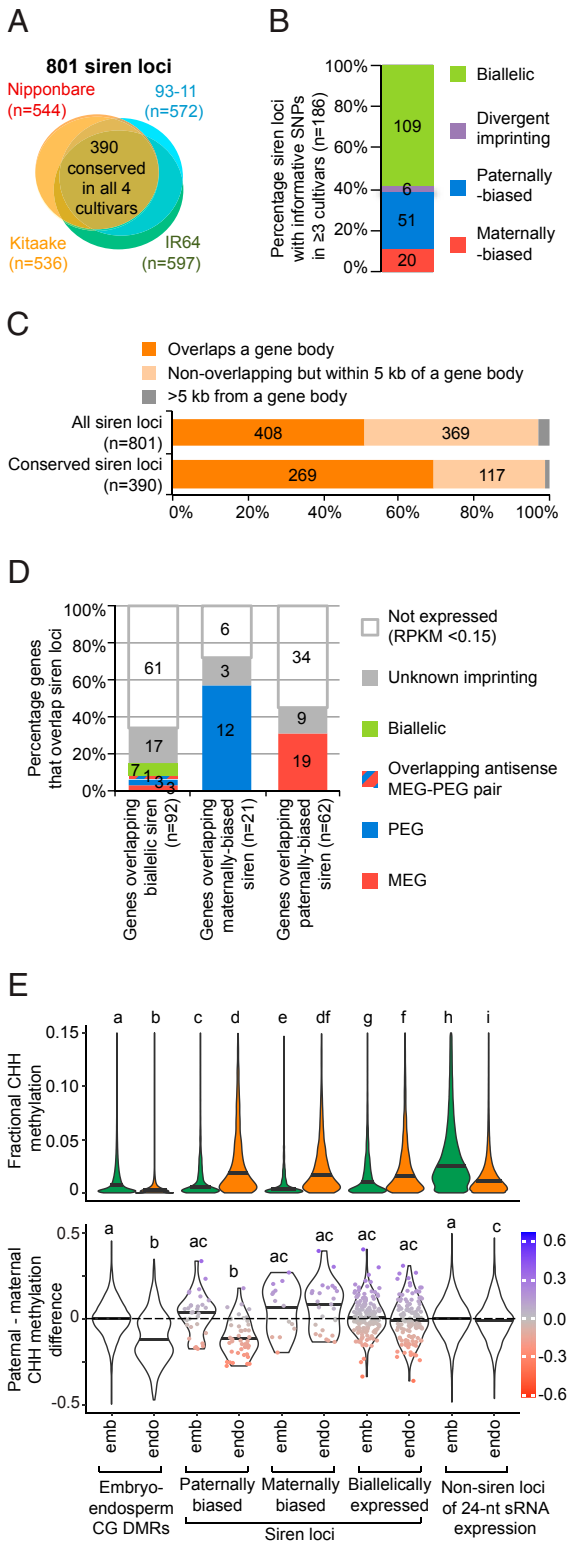

Fig. 2. Siren loci, which are loci associated with a relatively high production of 24-nt sRNAs in rice endosperm, can be imprinted and associate with gene silencing. (*A*) Venn diagram showing the overlap of significantly high 24-nt sRNA expression in siren loci of different rice cultivars. (*B*) Where imprinting could be assessed, a larger proportion of siren loci are imprinted compared with genes. (*C*) Most siren loci overlap genes or nearby intergenic regions and this trend is more marked for siren loci conserved among cultivars. (*D*) Where imprinted siren loci overlap genes, 24-nt sRNAs are expressed from the silenced parental allele of the gene. (*E*) Siren loci show greater CHH methylation in endosperm (endo; orange) than embryo (emb; green), unlike other 24-nt sRNA-producing regions of endosperm. Average fractional methylation of intersubspecies reciprocal crosses and the corresponding difference between parental alleles are shown for 50-bp windows, using embryo–endosperm CG differentially methylated regions as a control. Parental alleles differ most at imprinted siren loci compared with other endosperm sRNA-producing regions, though a paucity of data points reduces statistical power. sRNA-producing windows were defined where the sum of 24-nt reads across the four intersubspecies crosses was ≥30. Window methylation is based on ≥15 cytosines per cross, with differences between parental alleles only calculated where at least one allele is significantly methylated (>0.15) in each direction of the reciprocal cross. Data points are displayed in colors based on parent-of-origin–specific bias (blue, paternal; red, maternal; gray, biallelic) when <1,000 per violin. Different letter groupings indicate Bonferroni-corrected $P < 0.01$ in pairwise Wilcoxon rank-sum tests. Markers indicate the median. To resolve differences between genomic features, sRNA-expressing windows exclude CG DMRs and those of siren loci additionally exclude repeats.

exact test, $P = 10^{-35}$ and 10^{-30}, respectively), with the TSS pattern dominant among cMEGs and the gene body pattern dominant among cPEGs (Fig. 3*B* and *SI Appendix*, Fig. S14). A role for the TSS hypomethylation pattern in endosperm-specific gene activation is supported by findings that the pattern is enriched in embryo-silenced cMEGs compared with embryo-expressed cMEGs (Fig. 4*A*) and that, in general, expression levels are lowest for cMEGs with the TSS pattern compared with cPEGs and other cMEGs in the embryo (Fig. 4*B*) as well as other nonendosperm tissues (*SI Appendix*, Fig. S17).

In rice, there are two orthologs of the *A. thaliana* CG methylation maintenance gene *MET1*, *OsMET1-1* and *OsMET1-2* (56). A previous study has shown that *OsMET1-2* is the major gene responsible for the maintenance of CG methylation in rice: Loss of *OsMET1-2* expression due to retrotransposon insertion into a key exon results in abnormal seed development and seedling lethality (56). To further assess the effects of CG methylation on imprinting gene expression, we compared the expression levels of cPEGs and cMEGs in wild-type and *OsMET1-2*$^{-/-}$ null mutant seedlings. cMEGs repressed in wild-type rice seedlings are activated in *OsMET1-2*$^{-/-}$ seedlings (Fig. 4*C* and *SI Appendix*, Fig. S18), with the greatest effects on those with endosperm-specific TSS hypomethylation. Though 12 of the 77 cMEGs with apparent endosperm-specific TSS hypomethylation were unexpectedly expressed (reads per kilobase of transcript per million reads mapped [RPKM] > 0.5) in wild-type seedlings (Fig. 4*C*), 9 of them were predominantly transcribed from an alternative unmethylated TSS in seedlings (*SI Appendix*, Fig. S19 and Table S8) and tissue-specific TSS definition was likely imprecise at the remaining 3 cMEGs due to very low or very high expression (*SI Appendix*, Table S8).

A role for the gene body hypomethylation pattern in endosperm-specific gene silencing is supported by the highest embryo expression for cPEGs compared with cMEGs either with or without the TSS pattern (Fig. 4*B*). However, though the prevalent model suggests that cPEG gene body hypomethylation recruits silencing machinery, cPEGs with endosperm-specific gene body demethylation are not suppressed in *OsMET1-2*$^{-/-}$ seedlings (Fig. 4*C*). This result might reflect the specificity of PEG silencing machinery to the endosperm. The greater relevance of gene body methylation to regulation of PEGs than regulation of MEGs is highlighted by the overrepresentation of cPEGs with significant gene body methylation (the CG methylation levels of 66% of cPEGs are between 0.3 and 0.7; *SI Appendix*, Fig. S20) as compared with cMEGs (only 34% of cMEGs show gene body methylation between 0.3 and 0.7; *SI Appendix*, Fig. S20).

***cis*-Acting Genetic and Epigenetic Variation Is Associated with Imprinting Divergence.** We next assessed the relationship between cultivar-specific DNA sequence variation and epigenetic variation at the 19 loci with diverged imprinting by Sanger sequencing of cultivar

Downloaded at Institute Science Tech Austria on August 11, 2021

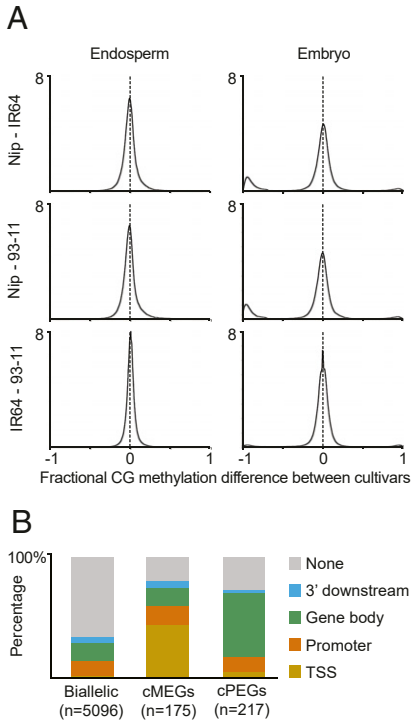

Fig. 3. DNA methylation profiles tend to be conserved among cultivars, with DNA demethylation enriched at both maternally expressed genes and paternally expressed genes. (*A*) Kernel density plots quantify the difference in fractional DNA methylation between cultivars of different subspecies (Nipponbare–IR64 and Nipponbare–93-11) and cultivars of the same subspecies (IR64–93-11) across all 50-bp windows of the genome with significant CG methylation based on ≥19 cytosines. Significant CG methylation for embryo was defined as embryo fractional methylation level >0.7 for at least one cultivar in the comparison, while that for endosperm required endosperm fractional methylation >0.7 for at least one cultivar in the comparison as well as embryo fractional methylation >0.7 for both cultivars. The centering of the main peak at zero indicates that most methylated windows were methylated in both cultivars. Shoulders indicating differences between cultivars are greater in the embryo panel and are reflective of general DNA methylation while shoulders are smaller in the endosperm panel and are reflective of endosperm-specific demethylation. (*B*) Nipponbare embryo–endosperm DNA hypomethylation patterns of 175 cMEGs, 217 cPEGs, and 5,096 conserved biallelic genes were assigned to one of five categories: a TSS pattern, gene body pattern, promoter pattern, 3′ downstream pattern, and a pattern where none of the sequence within 1 kb of the gene is hypomethylated. For detailed category definitions, refer to *SI Appendix, Materials and Methods*.

alleles. For 17 of the 19 loci, we assembled contigs of Nipponbare, 93-11, and IR64 alleles, verifying known SNPs and identifying novel SNPs and insertion/deletion events (indels) up to 5 kb away from transcription termini (*SI Appendix*, Table S9). Contigs of the remaining 2 loci were not assembled due to insurmountable difficulties in amplification and sequencing. To confirm associations of imprinting divergence with variation in maternal DNA hypomethylation, we mapped bisulfite-sequencing (BS-seq) libraries of endosperm and embryo DNA from each cultivar to the Sanger-sequenced loci and used embryo–endosperm CG hypomethylation as a proxy for maternal hypomethylation.

Of the 17 loci, 9 display gene-proximal correlative endosperm DNA hypomethylation among cultivars (Fig. 5). Of the 8 loci lacking correlating methylation features (*SI Appendix*, Fig. S21), 4 (3 PEGs and 1 MEG) possess cultivar-specific TEs in the promoter or gene body that mostly correlate with imprinted expression, except for one correlated with biallelic expression at PEG Os09g24220. The cultivar-specific TE in the gene body of PEG Os11g09329 is an intact autonomous Copia LTR retrotransposon, suggesting that its relatively recent transposition is associated with the emergence of imprinting at a previously biallelically expressed locus. It is possible that the epigenetic changes associated with these insertions act in a manner unrelated to DNA cytosine methylation; however, it is also possible that our estimation of DNA methylation over these repeats is obscured by the methylation of near-identical repeats elsewhere in the genome, or that the insertions themselves are not directly related to imprinting divergence but instead associated with more distal changes that have not yet been surveyed.

Of the nine genes with correlative hypomethylation differences, Os02g57200 displays direct association of genetic and epigenetic variation: Paternally expressed Nipponbare and IR64 alleles possess characteristic endosperm-specific gene body hypomethylation, but this hypomethylation is disrupted in the biallelically expressed 93-11 allele by intronic insertion of a 5.6-kb Copia LTR retrotransposon (Fig. 5*A*). We posit that this insertion prevents PcG-mediated repression of the 93-11 maternal allele, leading to biallelic expression. Similarly, paternal expression of the peptidyl-prolyl *cis–trans* isomerase Os11g38990 (Fig. 5*B*) in Nipponbare correlates with promoter and gene body hypomethylation. In the 93-11 and IR64 alleles, the promoter and gene body are less hypomethylated and the gene is biallelically expressed, possibly due to the disruption of demethylation recognition sites by two small promoter indels (5 and 8 bp, indicated by the black dashed box in Fig. 5*B*).

On the other hand, maternal expression of the Os11g31630 Nipponbare allele correlates with promoter and TSS DNA methylation in the embryo and maternal-specific hypomethylation in the endosperm (Fig. 5*C*). Lack of promoter and TSS DNA methylation in the embryo and endosperm of 93-11 and IR64 alleles associates with biallelic expression. This *indica*-specific epiallele may result from an upstream 300-bp DNA transposon PIF/Harbinger insertion that either forms a boundary that restricts spreading of DNA methylation from the promoter to the TSS or creates alternative transcription factor binding sites whose transcription prevents promoter and TSS DNA methylation. The TSS of the endosperm-expressed transcript variant for retrotransposon Os11g45295 (Fig. 5*D*) also displays endosperm-specific hypomethylation at the maternally expressed Nipponbare allele and lack of methylation at biallelically expressed 93-11 and IR64 alleles, with the epiallele associating with an adjacent small indel.

Imprinting at the homeodomain-containing protein Os01g57890 appears to effect maternal expression of the Nipponbare allele and paternal expression of IR64 and 93-11 *indica* alleles (*SI Appendix*, Table S3). DNA methylation variation between Nipponbare and the *indica* cultivars spans the promoter, TSS, and part of the gene body, such that the Nipponbare allele is methylated and the 93-11 and IR64 alleles are unmethylated in embryos (Fig. 5*E*). Nipponbare-specific embryo methylation correlates with an ∼250-bp LINE L1 element insertion. Based on similar methylation profiles of *indica* cultivars and wild rice species (Fig. 6), Os01g57890 was likely silenced by the PcG complex in the last common ancestor of the cultivars, with promoter methylation of the paternal allele protecting it from silencing. In Nipponbare, the silencing effect of TSS methylation caused by the LINE L1 insertion appears to be epistatic to the effect of promoter methylation in preventing PcG-mediated silencing, so the overall effect of the insertion-associated methylation is silencing of the paternal Nipponbare allele. Additionally, the Nipponbare LINE L1 insertion may erase genetic or epigenetic signals for PcG recruitment to unmethylated alleles and result in activation of the maternal Nipponbare allele by hypomethylation.

Os08g20500, Os12g35590, Os11g06650, and Os01g05510 are the four genes where DNA hypomethylation variation in critical regulatory regions for imprinting was observed without significant indels (<20 bp; Fig. 5 *F*–*I*). Os08g20500 (Fig. 5*F*), Os12g35590

Downloaded at Institute Science Tech Austria on August 11, 2021

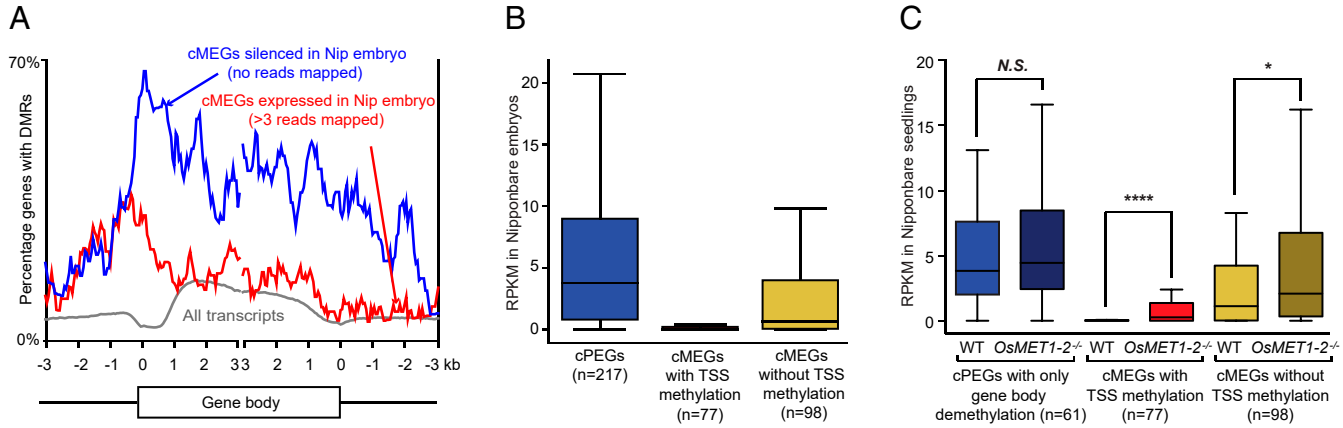

Fig. 4. Conserved maternally expressed genes with DNA methylation at the transcription start site are repressed in somatic tissues. (*A*) Curated endosperm genes were aligned at the TSS and transcription termination site (56), and the proportion of genes with a DMR present is plotted for each 50-bp interval within 3 kb of the alignment sites. Silencing in Nipponbare embryos was defined as no mapped RNA-seq reads while expression was defined as >3 mapped RNA-seq reads (0.31 reads per million reads mapped; the median embryo read count for maternally expressed genes). (*B*) Boxplots showing expression in Nipponbare embryos of cPEGs, cMEGs with TSS embryo–endosperm DMRs (i.e., with TSS DNA methylation in embryos), and cMEGs without TSS embryo–endosperm DMRs. (*C*) Boxplots showing expression in Nipponbare wild-type (WT) and *OsMET1-2*$^{-/-}$ mutant seedlings of cPEGs overlapped only with gene body embryo–endosperm DMRs within 1 kb of the gene, cMEGs with TSS embryo–endosperm DMRs, and cMEGs without TSS embryo–endosperm DMRs. *N.S.*, not significant; $*P < 0.05$, $****P < 10^{-4}$; Mann–Whitney test.

(Fig. 5*G*), and Os11g06650 (Fig. 5*H*) are maternally expressed in 93-11 and IR64, and biallelically expressed in Nipponbare. Their imprinting divergence may be explained by epialleles at the TSS (Os08g20500 and Os12g35590) or promoter (Os11g06650), where maternally expressed alleles possess embryo CG methylation and low but significantly higher non-CG methylation (*SI Appendix*, Figs. S22 and S23) while biallelically expressed alleles are unmethylated. Although no large indel events were found at these loci, the density of SNPs between Nipponbare and *indica* alleles at these loci (~1%) is higher than the genome average (~0.45%). For Os01g05510 (Fig. 5*I*), paternal bias of Nipponbare and IR64 alleles correlates with endosperm-specific hypomethylation in the middle of the gene body. The 93-11 allele possesses additional DNA methylation in this region that is not targeted for hypomethylation and may prevent

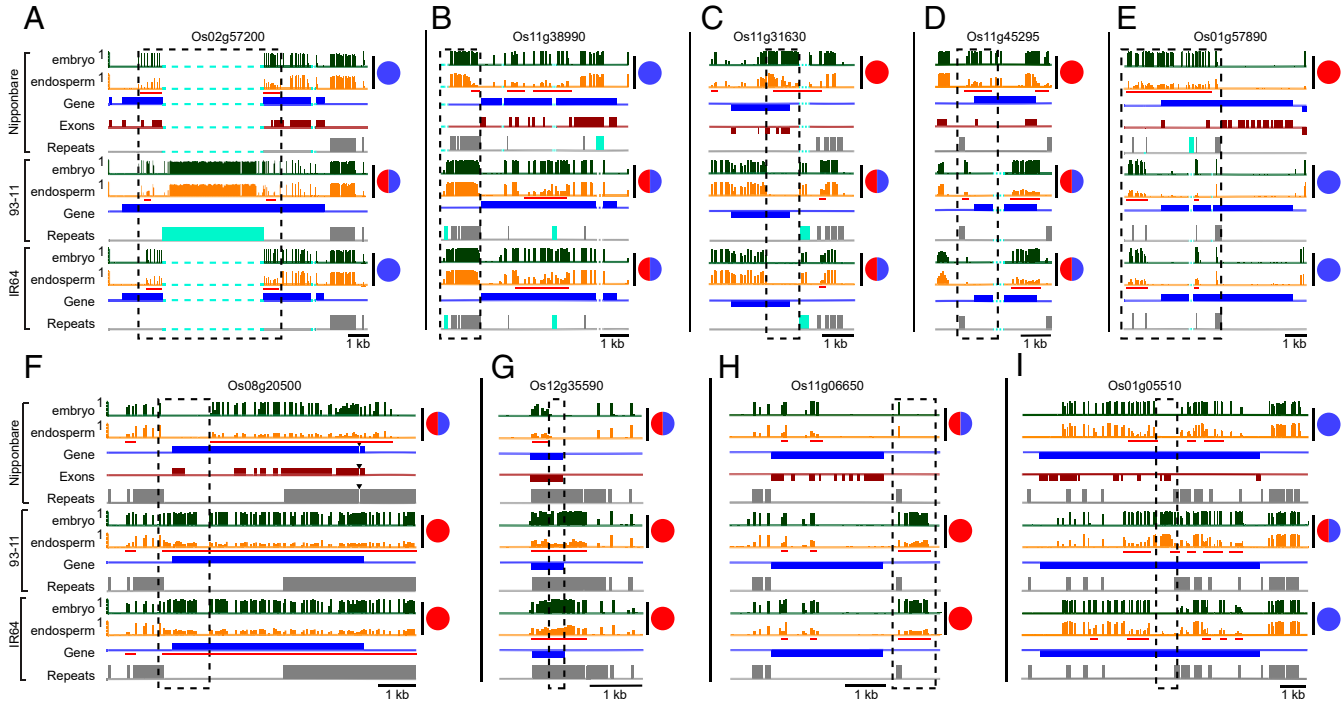

Fig. 5. Snapshots of CG methylation in Nipponbare, 93-11, and IR64 cultivars across gene contigs showing imprinting divergence. (*A*) Os02g57200. (*B*) Os11g38990. (*C*) Os11g31630. (*D*) Os11g45295. (*E*) Os01g57890. (*F*) Os08g20500. (*G*) Os12g35590. (*H*) Os11g06650. (*I*) Os01g05510. Paternal expression bias of cultivar alleles is indicated by a blue circle, maternal bias by a red circle, and biallelic expression by a half-red half-blue circle. Red lines underlying endosperm tracks indicate regions of significant CG hypomethylation compared with embryo (Fisher's exact test, $P < 10^{-3}$; fractional methylation difference ≥ 0.3). Black dashed boxes indicate variation in embryo–endosperm DNA hypomethylation that is associated with imprinting divergence. Teal-colored boxes and lines highlight nearby genetic variation in *A*–*E*. Only 50-bp windows with >10 bisulfite-seq reads are shown.

Downloaded at Institute Science Tech Austria on August 11, 2021

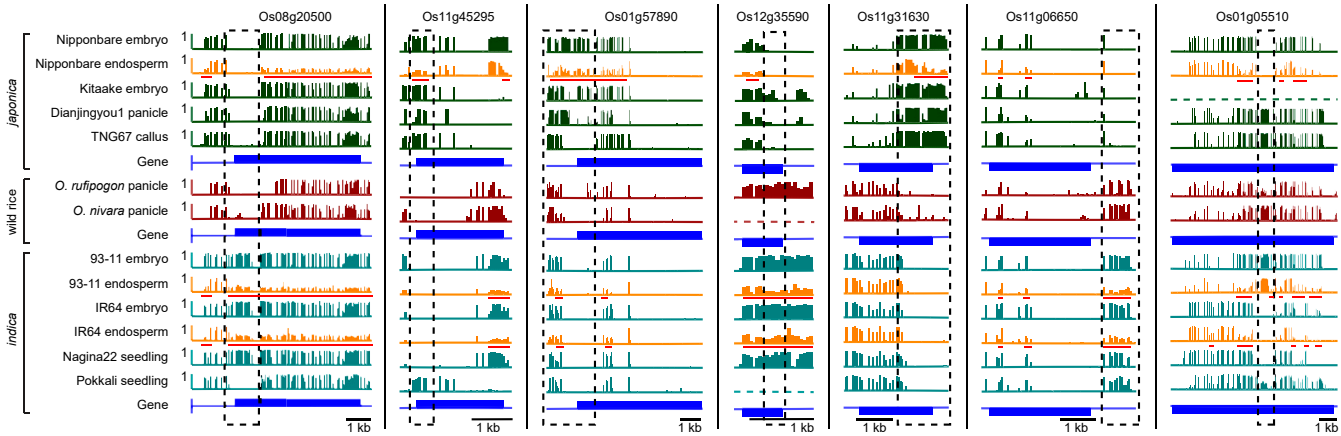

Fig. 6. Snapshots of CG methylation across *japonica*, *indica*, and wild rice reveal that epialleles associated with imprinting divergence are stable within lineages. Red lines underlying endosperm tracks indicate regions of significant CG hypomethylation compared with embryo (Fisher's exact test, $P < 10^{-3}$; fractional methylation difference ≥ 0.3). Black dashed boxes indicate epialleles associated with imprinting divergence. Tracks consisting of dashed lines represent cultivars without informative read mapping. Only 50-bp windows with >10 bisulfite-seq reads are shown.

PcG targeting despite hypomethylation at flanking sites. SNPs within this region only exist between Nipponbare and *indica* cultivars, so the epiallele between 93-11 and IR64 cannot be explained by sequence variation.

Epialleles Associated with Imprinting Divergence Are Usually Stably Inherited. As DNA methylation is known to be faithfully transmitted through generations with a low rate of epimutation, we examined the stability of the epialleles associated with imprinting divergence using publicly available rice BS-seq data from different cultivars of *japonica* and *indica* rice as well as wild rice species (57–59). In line with expectations, we found that cultivars within the same grouping (i.e., *japonica*, *indica*, or wild rice) mostly had very similar epialleles at six of the seven assessable loci, with two major types of epialleles identified for each locus (Fig. 6). Notable exceptions included the saline-tolerant *indica* cultivar Pokkali possessing a *japonica*-like unmethylated TSS for retrotransposon genes Os08g20500 and Os11g45295, while the TSS of homeobox domain–containing protein gene Os01g57890 was only methylated in a subset of *japonica* cultivars composed of Nipponbare and Kitaake. For the seventh locus, Os01g05510, we found three epialleles: the unmethylated state of Nipponbare, IR64, and Nagina22, the highly methylated state of Dianjingyou1, TNG67, *Oryza nivara*, and 93-11, and the intermediately methylated state of *Oryza rufipogon* and Pokkali. The lack of association between epialleles and subspecies groups may result from either occasional reversion events or introgressions during the long breeding histories of cultivars.

Discussion

Though rapid genome-wide detection of plant imprinted genes has demonstrated that gene imprinting evolves as cultivars and species diverge, little is known about the forces driving imprinting divergence. Using cultivars from both *indica* and *japonica* subspecies that carry importance in the rice community (37, 60, 61), we show that the extent of imprinting divergence among rice cultivars is similar to that among maize cultivars (54) and *A. thaliana* ecotypes (7), providing an opportunity to dissect mechanisms of imprinting divergence. At PEG Os11g09329, we identify the possible birth of a gene imprint due to a relatively recent retrotransposition event. We did not observe associated changes in parent-of-origin–specific DNA methylation at the locus, either because the imprint is formed at a distal location or by other chromatin modifications, or because analysis of DNA methylation at such TEs is obscured by near-identical sequences elsewhere in the genome. We also found imprinting divergence without local differences in parent-of-origin–specific DNA methylation to be associated with the presence of specific DNA transposons at Os09g24220 and Os05g47870. Here, the cut-and-paste mechanism of DNA transposons makes it difficult to distinguish between imprint birth and death. At other loci, genetic changes more often caused imprint death through reduced parent-of-origin–specific differences in DNA methylation, either by apparent deletion as seen in Os11g45295 and Os01g69910 or insertion as in Os02g57200, Os11g38990, and Os11g31630. Where TE insertion correlated with imprinting divergence, TE class or family did not predict imprinting state, suggesting that the surrounding chromatin environment plays a role in defining the targets of imprinting mechanisms. Epiallelic variation without genetic divergence is also a key contributor to imprint evolution, observed at four of nine genes where DNA methylation differences correlate with imprinting. We found that such DNA methylation epialleles are stably transmitted even as lineages diverge, producing consistent effects on gene expression as measured by imprinted expression.

These results reinforce models developed by identifying dominant DNA methylation and H3K27me3 histone methylation patterns at rice MEGs and PEGs, where DNA methylation is a causal imprint (25, 26, 43). Our finding that the *OsMET1-2* mutation derepresses MEGs in nonendosperm tissues such as rice seedlings is consistent with experiments using the DNA methylation inhibitor 5-aza-2′-deoxycytidine (43). The case study of Os01g57890, which is maternally expressed from the Nipponbare allele and paternally expressed from *indica* alleles, suggests that silencing by TSS methylation is epistatic to other regulatory mechanisms such as PcG-mediated silencing and, as such, is a stronger predictor of gene silencing. Data from Os01g57890 and Os11g45295 also support the previous finding that endosperm-specific TSS hypomethylation mediates switching between alternative TSSs (25). Our finding that the *OsMET1-2* mutation did not silence PEGs suggests either that DNA methylation alone does not form a causal imprint for PEGs or that PEG regulation in *OsMET1-2*$^{-/-}$ mutants is affected by a redistribution of other chromatin marks during genome-wide suppression of DNA methylation maintenance, as seen in *A. thaliana* (47). Unlike a recent study in *A. thaliana* (62), we did not find rice PEGs enriched for maternal-specific CHG methylation in the endosperm but it is possible that such CHG methylation and associated PcG-mediated marks serve as a transitory imprint in the central cell.

Another point of difference with *A. thaliana* is that rice endosperm siren loci appear to differ in genomic distribution from similar "siren-like" loci that dominate ovary and egg sRNA populations

PLANT BIOLOGY

Downloaded at Institute Science Tech Austria on August 11, 2021

(63). Siren sRNAs in *Brassica rapa* and *A. thaliana* are expressed via the canonical RNA POLYMERASE IV (Pol IV)–RNA-DEPENDENT RNA POLYMERASE 2 (RDR2) pathway from the same loci in maternal and seed tissues, relatively highly in ovules and seed coat, moderately in the endosperm, and lowly in embryos and mature seeds (64). We show that rice endosperm siren sRNAs, like those of Brassicaceous ovules (64, 65) but unlike those of rice eggs and ovary (66), are associated with greater CHH methylation compared with the relatively lowly expressed nonsiren 24-nt sRNAs, suggestive of siren-mediated RdDM. However, our finding that imprinted siren sRNAs mediate as much or more *trans* RdDM than *cis* RdDM, despite frequent imprinted gene expression from the *trans* allele of genes which overlap siren loci, suggests either that siren-mediated RdDM is ineffective at gene silencing in rice endosperm or that imprinted genes overlapped by siren loci progress toward silencing of both alleles during rice endosperm development. It is unclear whether Brassicaceous endosperm siren sRNAs also associate with RdDM but they and other endosperm 24-nt sRNAs are not required for proper seed development in *A. thaliana* (64, 67). In *B. rapa*, siren and/or other 24-nt sRNAs are essential for seed development (64) but endosperm effects are yet to be dissociated from those in maternal reproductive tissues.

Although imprinted sRNAs associate with parent-of-origin–specific gene silencing in the endosperm across *A. thaliana* (7, 51), maize (52), and rice (12, 26, 43), we find that siren locus expression is not a strong driver of rice gene imprinting divergence. Our observation that rice endosperm sRNAs are imprinted more often than coding transcripts suggests that epigenetically regulated DNA is imprinted more often than can be estimated based on coding regions. This concurs with indications that epigenetic regulation is essential for balancing parental genome dosage in the *A. thaliana* endosperm (68). Unlike in *A. thaliana* (51, 64), sRNAs associated with imprinted rice genes are more likely to be paternally biased than maternally biased. This may be due to general differences in epigenetic regulation between *A. thaliana* and rice (55) such that some imprinted rice genes may behave more like *A. thaliana* TEs, which show an overall paternal bias in associated sRNA (51). This may also be due to the overlap of endosperm and ovule siren loci in *A. thaliana* (64) but not rice (66).

Further work is necessary to establish whether PcG-mediated histone methylation acts as a causal imprint in rice, as it appears to at *A. thaliana* PEGs (62, 69). PcG activity may also explain imprinting of MEGs that do not display parent-of-origin–specific DNA methylation patterns (49). We expect technical advances in the detection of plant distal regulatory elements (70, 71) will help resolve the origin of imprinting at some loci for which the mechanism of imprinting is yet unknown.

Materials and Methods

See *SI Appendix, Materials and Methods* for full methods.

Next-Generation Sequencing Data. Strand-specific RNA-sequencing libraries were constructed as previously described (72), using ribosomal RNA-depleted total RNA as input and duplex-specific nuclease treatment of the final library (73). Bisulfite- and sRNA-sequencing libraries were constructed as previously described (26). Using the Illumina HiSeq 2000 platform, RNA-seq and BS-seq libraries were sequenced as 100-bp single-end reads and sRNA-seq libraries as 50-bp single-end reads. We used previously published DNA methylation data for *japonica* cultivars TNG67 (58) and Dianjingyou1 (59), wild rice species *O. rufipogon* (59) and *O. nivara* (59), and *indica* cultivars Nagina22 (57) and Pokkali (57).

Data Processing. Resequencing identified SNPs between IR64 and Nipponbare (Dataset S1), while previously published lists were used for SNPs between Kitaake and Nipponbare (26) and between 93-11 and Nipponbare (74, 75). We used our previous methods for read trimming (26), allele-specific read mapping (5, 26), calculation of fractional DNA methylation (36), and definition of differentially methylated regions (DMRs) (Dataset S6) and sRNA-producing loci (26). Scaled Venn diagrams were generated using eulerAPE v3 (76). Kernel density plots were made using 50-bp windows of the genome.

Analysis of Imprinted Expression in Endosperm. Imprinted genes identified based on deviation of RNA-seq read counts from expected biallelic ratios (Fisher's exact test, $P < 0.01$; Fig. 1*A*) were defined as conserved or diverged among cultivars after resolving for variety-specific biases, read complexity, mismapping, erroneous SNPs, alternative transcript variants, developmental resetting, and potential contamination with maternal tissue (*SI Appendix, Materials and Methods* and Fig. S2). sRNA-producing siren loci were also classified as imprinted (maternal fraction >0.77 or <0.5 in both reciprocal crosses; Fisher's exact test, $P < 0.001$), biallelic (maternal fraction >0.617 and <0.717 in reciprocal crosses), or unclear, and assessed for imprinting conservation or divergence (*SI Appendix, Materials and Methods* and Fig. S7). Functional GO term enrichment of imprinted genes used the agriGO toolkit (53).

Sanger Sequencing and Assembly of Cultivar Alleles. Primers in *SI Appendix*, Table S10 amplified ~1.7-kb overlapping fragments of the gene body, plus flanking regions either up to 5 kb away or until neighboring transcription termini. PCRs used Phusion High-Fidelity DNA Polymerase (NEB; M0530), Q5 High-Fidelity DNA Polymerase (NEB; M0491), or LongAmp Taq DNA Polymerase (NEB; M0323) with melting temperature between 68 and 78 °C and optional addition of 1 or 2 M betaine. PCR amplicons purified by SPRI beads or the CloneJET PCR Cloning Kit (Fisher Scientific; K1232) were Sanger-sequenced with forward and reverse primers either from the amplicon or pJET1.2 and assembled into contigs (Dataset S4) using the CodonCode Aligner (https://www.codoncode.com/).

Data Availability. All data reported in this article have been deposited in the National Center for Biotechnology Information. IR64 genome-resequencing data are deposited by the Joint Genome Institute in the Sequence Read Archive (project accession nos. SRP006613 and PRJNA337128). All other DNA-sequencing, bisulfite DNA-sequencing, and RNA-sequencing raw data as well as key processed data files are deposited in the Gene Expression Omnibus (accession no. GSE130122).

ACKNOWLEDGMENTS. We thank W. Schackwitz, M. Joel, and the Joint Genome Institute sequencing team for generating the IR64 genome sequence and initial analysis; L. Bartley and E. Marvinney for genomic DNA preparation for IR64 resequencing; and the University of California (UC), Berkeley Sanger sequencing team for technical advice and service. This work was partially funded by NSF Grant IOS-1025890 (to R.L.F. and D.Z.), NIH Grant GM69415 (to R.L.F. and D.Z.), NIH Grant GM122968 (to P.C.R.), a Young Investigator Grant from the Arnold and Mabel Beckman Foundation (to D.Z.), an International Fulbright Science and Technology Award (to J.A.R.), and a Taiwan Ministry of Education Studying Abroad Scholarship (to P.-H.H.). This work used the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley, supported by NIH Instrumentation Grant S10 OD018174.

1. H. Jiang, C. Köhler, Evolution, function, and regulation of genomic imprinting in plant seed development. *J. Exp. Bot.* **63**, 4713–4722 (2012).
2. N. D. Pires, U. Grossniklaus, Different yet similar: Evolution of imprinting in flowering plants and mammals. *F1000Prime Rep.* **6**, 63 (2014).
3. M. B. Renfree, S. Suzuki, T. Kaneko-Ishino, The origin and evolution of genomic imprinting and viviparity in mammals. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **368**, 20120151 (2013).
4. J. M. Frost, G. E. Moore, The importance of imprinting in the human placenta. *PLoS Genet.* **6**, e1001015 (2010).
5. T. F. Hsieh *et al*., Regulation of imprinted gene expression in *Arabidopsis* endosperm. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 1755–1762 (2011).
6. M. Luo *et al*., A genome-wide survey of imprinted genes in rice seeds reveals imprinting primarily occurs in the endosperm. *PLoS Genet.* **7**, e1002125 (2011).
7. D. Pignatta *et al*., Natural epigenetic polymorphisms lead to intraspecific variation in *Arabidopsis* gene imprinting. *eLife* **3**, e03198 (2014).
8. A. J. Waters *et al*., Parent-of-origin effects on gene expression and DNA methylation in the maize endosperm. *Plant Cell* **23**, 4221–4233 (2011).
9. M. Zhang *et al*., Extensive, clustered parental imprinting of protein-coding and noncoding RNAs in developing maize endosperm. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 20042–20047 (2011).
10. M. J. Bauer, R. L. Fischer, Genome demethylation and imprinting in the endosperm. *Curr. Opin. Plant Biol.* **14**, 162–167 (2011).
11. Y. Choi *et al*., DEMETER, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed viability in *Arabidopsis*. *Cell* **110**, 33–42 (2002).
12. J. Yuan *et al*., Both maternally and paternally imprinted genes regulate seed development in rice. *New Phytol.* **216**, 373–387 (2017).

Downloaded at Institute Science Tech Austria on August 11, 2021

13. S. N. Anderson, N. M. Springer, Potential roles for transposable elements in creating imprinted expression. *Curr. Opin. Genet. Dev.* **49**, 8–14 (2018).
14. M. Gehring, Genomic imprinting: Insights from plants. *Annu. Rev. Genet.* **47**, 187–208 (2013).
15. T. M. Vu *et al*., RNA-directed DNA methylation regulates parental genomic imprinting at several loci in *Arabidopsis*. *Development* **140**, 2953–2960 (2013).
16. P. Wolff *et al*., High-resolution analysis of parent-of-origin allelic expression in the *Arabidopsis* endosperm. *PLoS Genet.* **7**, e1002126 (2011).
17. G. Bourque *et al*., Ten things you should know about transposable elements. *Genome Biol.* **19**, 199 (2018).
18. C. D. Hirsch, N. M. Springer, Transposable element influences on gene expression in plants. *Biochim. Biophys. Acta. Gene Regul. Mech.* **1860**, 157–165 (2017).
19. M. Gehring, K. L. Bubb, S. Henikoff, Extensive demethylation of repetitive elements during seed development underlies gene imprinting. *Science* **324**, 1447–1451 (2009).
20. C. A. Ibarra *et al*., Active DNA demethylation in plant companion cells reinforces transposon methylation in gametes. *Science* **337**, 1360–1364 (2012).
21. Y. Kinoshita *et al*., Control of FWA gene silencing in *Arabidopsis thaliana* by SINE-related direct repeats. *Plant J.* **49**, 38–45 (2007).
22. C. B. R. Villar, A. Erilova, G. Makarevich, R. Trösch, C. Köhler, Control of PHERES1 imprinting in *Arabidopsis* by direct tandem repeats. *Mol. Plant* **2**, 654–660 (2009).
23. P. Wang *et al*., Genome-wide high-resolution mapping of DNA methylation identifies epigenetic variation across embryo and endosperm in maize (*Zea may*). *BMC Genomics* **16**, 21 (2015).
24. D. Pignatta, K. Novitzky, P. R. V. Satyaki, M. Gehring, A variably imprinted epiallele impacts seed development. *PLoS Genet.* **14**, e1007469 (2018).
25. M. Du, M. Luo, R. Zhang, E. J. Finnegan, A. M. Koltunow, Imprinting in rice: The role of DNA and histone methylation in modulating parent-of-origin specific expression and determining transcript start sites. *Plant J.* **79**, 232–242 (2014).
26. J. A. Rodrigues *et al*., Imprinted expression of genes and small RNA is associated with localized hypomethylation of the maternal genome in rice endosperm. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 7934–7939 (2013).
27. M. Zhang *et al*., Genome-wide high resolution parental-specific DNA and histone methylation maps uncover patterns of imprinting regulation in maize. *Genome Res.* **24**, 167–176 (2014).
28. A. Bleckmann, S. Alter, T. Dresselhaus, The beginning of a seed: Regulatory mechanisms of double fertilization. *Front. Plant Sci.* **5**, 452 (2014).
29. R. Yadegari, G. N. Drews, Female gametophyte development. *Plant Cell* **16**, S133–S141 (2004).
30. J. F. Gutiérrez-Marcos *et al*., Epigenetic asymmetry of imprinted genes in plant gametes. *Nat. Genet.* **38**, 876–878 (2006).
31. S. Jahnke, S. Scholten, Epigenetic resetting of a gene imprinted in plant embryos. *Curr. Biol.* **19**, 1677–1681 (2009).
32. T. Morales-Ruiz *et al*., DEMETER and REPRESSOR OF SILENCING 1 encode 5-methylcytosine DNA glycosylases. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 6853–6858 (2006).
33. K. Park *et al*., DNA demethylation is initiated in the central cells of *Arabidopsis* and rice. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 15138–15143 (2016).
34. J. Penterman *et al*., DNA demethylation in the *Arabidopsis* genome. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 6752–6757 (2007).
35. W. Xiao *et al*., Imprinting of the MEA Polycomb gene is controlled by antagonism between MET1 methyltransferase and DME glycosylase. *Dev. Cell* **5**, 891–901 (2003).
36. T. F. Hsieh *et al*., Genome-wide demethylation of *Arabidopsis* endosperm. *Science* **324**, 1451–1454 (2009).
37. H. W. Zhang *et al*., Comparison of differentially expressed genes involved in drought response between two elite rice varieties. *Mol. Plant* **5**, 1403–1405 (2012).
38. D. P. Barlow, Imprinting: A gamete's point of view. *Trends Genet.* **10**, 194–199 (1994).
39. A. C. Ferguson-Smith, Genomic imprinting: The emergence of an epigenetic paradigm. *Nat. Rev. Genet.* **12**, 565–575 (2011).
40. W. A. MacDonald, M. R. Mann, Epigenetic regulation of genomic imprinting from germ line to preimplantation. *Mol. Reprod. Dev.* **81**, 126–140 (2014).
41. A. Inoue, L. Jiang, F. Lu, T. Suzuki, Y. Zhang, Maternal H3K27me3 controls DNA methylation-independent imprinting. *Nature* **547**, 419–424 (2017).
42. Z. Chen, Q. Yin, A. Inoue, C. Zhang, Y. Zhang, Allelic H3K27me3 to allelic DNA methylation switch maintains noncanonical imprinting in extraembryonic cells. *Sci. Adv.* **5**, eaay7246 (2019).
43. C. Chen *et al*., Characterization of imprinted genes in rice reveals conservation of regulation and imprinting with other plant species. *Plant Physiol.* **177**, 1754–1771 (2018).
44. M. Gehring *et al*., DEMETER DNA glycosylase establishes MEDEA Polycomb gene self-imprinting by allele-specific demethylation. *Cell* **124**, 495–506 (2006).
45. P. E. Jullien, T. Kinoshita, N. Ohad, F. Berger, Maintenance of DNA methylation during the *Arabidopsis* life cycle is essential for parental imprinting. *Plant Cell* **18**, 1360–1372 (2006).
46. C. Köhler, D. R. Page, V. Gagliardini, U. Grossniklaus, The *Arabidopsis thaliana* MEDEA Polycomb group protein controls expression of PHERES1 by parental imprinting. *Nat. Genet.* **37**, 28–30 (2005).
47. A. Deleris *et al*., Loss of the DNA methyltransferase MET1 induces H3K9 hypermethylation at PcG target genes and redistribution of H3K27 trimethylation to transposons in *Arabidopsis thaliana*. *PLoS Genet.* **8**, e1003062 (2012).
48. I. Weinhofer, E. Hehenberger, P. Roszak, L. Hennig, C. Köhler, H3K27me3 profiling of the endosperm implies exclusion of Polycomb group protein targeting by DNA methylation. *PLoS Genet.* **6**, e1001152 (2010).
49. R. A. Batista, C. Köhler, Genomic imprinting in plants—Revisiting existing models. *Genes Dev.* **34**, 24–36 (2020).
50. M. T. Raissig, C. Baroux, U. Grossniklaus, Regulation and flexibility of genomic imprinting during seed development. *Plant Cell* **23**, 16–26 (2011).
51. R. M. Erdmann, P. R. V. Satyaki, M. Klosinska, M. Gehring, A small RNA pathway mediates allelic dosage in endosperm. *Cell Rep.* **21**, 3364–3372 (2017).
52. M. Xin *et al*., Dynamic parent-of-origin effects on small interfering RNA expression in the developing maize endosperm. *BMC Plant Biol.* **14**, 192 (2014).
53. Z. Du, X. Zhou, Y. Ling, Z. Zhang, Z. Su, agriGO: A GO analysis toolkit for the agricultural community. *Nucleic Acids Res.* **38**, W64–W70 (2010).
54. A. J. Waters *et al*., Comprehensive analysis of imprinted genes in maize reveals allelic variation for imprinting and limited conservation with other species. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 19639–19644 (2013).
55. A. Zemach *et al*., Local DNA hypomethylation activates genes in rice endosperm. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 18729–18734 (2010).
56. L. Hu *et al*., Mutation of a major CG methylase in rice causes genome-wide hypomethylation, dysregulated genome expression, and seedling lethality. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 10642–10647 (2014).
57. R. Garg, V. Narayana Chevala, R. Shankar, M. Jain, Divergent DNA methylation patterns associated with gene expression in rice cultivars with contrasting drought and salinity stress response. *Sci. Rep.* **5**, 14922 (2015).
58. F.-M. Hsu *et al*., Dynamics of the methylome and transcriptome during the regeneration of rice. *Epigenomes* **2**, 14 (2018).
59. X. Li *et al*., Single-base resolution maps of cultivated and wild rice methylomes and regulatory roles of DNA methylation in plant gene expression. *BMC Genomics* **13**, 300 (2012).
60. W. R. Coffman, T. R. Hargrove, "Modern rice varieties as a possible factor in production variability" in *Variability in Grain Yields: Implications for Agricultural Research and Policy in Developing Countries*, J. R. Anderson, P. B. R. Hazell, Eds. (International Food Policy Research Institute, 1989), pp. 133–146.
61. J. L. Wu *et al*., Chemical- and irradiation-induced mutants of indica rice IR64 for forward and reverse genetics. *Plant Mol. Biol.* **59**, 85–97 (2005).
62. J. Moreno-Romero, G. Del Toro-De León, V. K. Yadav, J. Santos-González, C. Köhler, Epigenetic signatures associated with imprinted paternally expressed genes in the *Arabidopsis* endosperm. *Genome Biol.* **20**, 41 (2019).
63. C. Li *et al*., Resetting of the 24-nt siRNA landscape in rice zygotes. *bioRxiv* [Preprint] (2021). https://doi.org/10.1101/2020.08.31.275958 (Accessed 1 May 2021).
64. J. W. Grover *et al*., Abundant expression of maternal siRNAs is a conserved feature of seed development. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 15305–15315 (2020).
65. M. Zhou *et al*., The CLASSY family controls tissue-specific DNA methylation patterns in Arabidopsis. *bioRxiv* [Preprint] (2021). https://doi.org/10.1101/2021.01.23.427869 (Accessed 1 May 2021).
66. C. Li *et al*., Genome-wide redistribution of 24-nt siRNAs in rice gametes. *Genome Res.* **30**, 173–184 (2020).
67. R. C. Kirkbride *et al*., Maternal small RNAs mediate spatial-temporal regulation of gene expression, imprinting, and seed development in *Arabidopsis*. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 2761–2766 (2019).
68. P. R. V. Satyaki, M. Gehring, Paternally acting canonical RNA-directed DNA methylation pathway genes sensitize *Arabidopsis* endosperm to paternal genome dosage. *Plant Cell* **31**, 1563–1578 (2019).
69. R. A. Batista *et al*., The MADS-box transcription factor PHERES1 controls imprinting in the endosperm by binding to domesticated transposons. *eLife* **8**, e50541 (2019).
70. E. Li *et al*., Long-range interactions between proximal and distal regulatory regions in maize. *Nat. Commun.* **10**, 2633 (2019).
71. B. Weber, J. Zicola, R. Oka, M. Stam, Plant enhancers: A call for discovery. *Trends Plant Sci.* **21**, 974–987 (2016).
72. L. Wang *et al*., A low-cost library construction protocol and data analysis pipeline for Illumina-based strand-specific multiplex RNA-seq. *PLoS One* **6**, e26426 (2011).
73. M. Matvienko *et al*., Consequences of normalizing transcriptomic and genomic libraries of plant genomes using a duplex-specific nuclease and tetramethylammonium chloride. *PLoS One* **8**, e55913 (2013).
74. X. Huang *et al*., High-throughput genotyping by whole-genome resequencing. *Genome Res.* **19**, 1068–1076 (2009).
75. D. Wang, Y. Xia, X. Li, L. Hou, J. Yu, The Rice Genome Knowledgebase (RGKbase): An annotation database for rice comparative genomics and evolutionary biology. *Nucleic Acids Res.* **41**, D1199–D1205 (2013).
76. L. Micallef, P. Rodgers, eulerAPE: Drawing area-proportional 3-Venn diagrams using ellipses. *PLoS One* **9**, e101717 (2014).

Downloaded at Institute Science Tech Austria on August 11, 2021