# Supporting Information

## Rodrigues et al. 10.1073/pnas.1306164110

### SI Materials and Methods

**Nipponbare and Kitaake Rice Cultivars.** Nipponbare and Kitaake seeds were germinated in Petri dishes at 28 °C with light for 1 wk before transplantation to field soil in an environment-controlled green house at the University of California Davis Core facility. Greenhouse conditions during the period of growth included a temperature between 27 and 30 °C, humidity at 50%, and natural sunlight.

The two *japonica* cultivars of rice originate from Japan, with Kitaake growing in the harsher conditions of the northern Hokkaido prefecture and Nipponbare flourishing in the more temperate lower latitudes of Japan. Part of Kitaake's adaptation to the harsher climate is its fairly rapid life cycle (~9 wk from seed to seed), cold tolerance, and insensitivity to photoperiod, humidity, and wind agitation when flowering. Nipponbare is a high-yielding modern cultivar and does well with high fertilizer treatment; it is also grown outside of Japan, in regions with temperate climate such as Europe. Both Kitaake and Nipponbare are primarily grown to provide grain for food.

**Resequencing of the Kitaake Rice Cultivar.** Fourteen-day-old greenhouse-grown Kitaake seedlings were used to prepare genomic DNA (gDNA) using sucrose-based extraction buffer for extraction of nuclei and guanidine-based lysis buffer. gDNA was broken into smaller fragments via nebulization and ligated with adapters. Sequencing was performed using the Illumina Genome Analyzer II (GA-II) at the Department of Energy Joint Genome Institute. Analysis was performed using maq-0.7.1 mapping and assembly software (1). The *Oryza sativa* ssp. *japonica* cv. Nipponbare genome sequence MSU version 6.0 was used as reference to map Kitaake reads. Bases were called using Bustard 14.0 (80 × 1). After filtering, 169,819 SNPs were discovered between the Kitaake and Nipponbare genomes. The sequence data have been submitted to the NCBI Sequence Read Archive under accession number SRX037797.

**Isolation of Rice Endosperm and Embryos.** Reciprocal crosses were performed between Nipponbare and Kitaake cultivars of rice through manual emasculation of rice flowers followed by artificial pollination. Self-fertilized seeds from both parent varieties and $F_1$ seeds from crosses were harvested 7–8 d after pollination. The palea and lemma were separated from the seed coat, and a razor blade was used to slice open the seed coat. Milky stage endosperm was pipetted out from inside the seed coat and stored either in TRIzol for preparing small RNA libraries or 2% (wt/vol) hexadecyltrimethylammonium bromide (CTAB) extraction buffer for bisulfite sequencing libraries. Embryos were isolated after the endosperm had been collected and were washed individually through vigorous agitation in 0.5 mL of 1× PBS solution. Individually isolated $F_1$ seeds were verified for heterozygosity with a PCR-based assay using microsatellite marker RM1 (2).

**Preparation of Rice Seedling Tissue for sRNA Libraries.** Nipponbare rice seeds were grown in sterile flask culture with Gamborg's B-5 medium with sucrose (Caisson Laboratories). At day 21 after germination, seedlings were rinsed in sterile water, and root and shoot tissue was separated and snap-frozen in liquid nitrogen for future RNA extraction with TRIzol.

**Bisulfite Sequencing Library Construction and Sequencing.** Paired-end bisulfite sequencing libraries for Illumina sequencing were constructed as described previously (3) with minor modifications. In brief, ~1 μg of gDNA was fragmented by sonication, end repaired, and ligated to custom-synthesized methylated adapters (Eurofins MWG Operon) according to the manufacturer's (Illumina) instructions for gDNA library construction. Adapter-ligated libraries were subjected to two successive treatments of sodium bisulfite conversion using the EpiTect Bisulfite kit (Qiagen) as outlined in the manufacturer's instructions. One-quarter of each bisulfite-converted library was PCR amplified using the following conditions: 2.5 U of ExTaq DNA polymerase (Takara Bio), 5 μL of 10× Extaq reaction buffer, 25 μM dNTPs, 1 μL Primer 1.1, and 1 μL Primer 2.1 (50 μL final). PCR reactions were carried out as follows: 95 °C for 3 min and then 12–14 cycles of 95 °C for 30 s, 65 °C for 30 s, and 72 °C for 60 s. The enriched libraries were purified twice with the solid phase reverse immobilization (SPRI) method using AM-Pure beads (Beckman Coulter), before quantification with a Bioanalyzer (Agilent). Sequencing performed on the Illumina HiSeq 2000 platform by the Vincent J. Coates Genomic Sequencing Laboratory at University of California at Berkeley generated 100-bp single-end reads.

**Small RNA Library Construction and Sequencing.** Total RNA was prepared from snap-frozen rice seedling root and shoot tissue and snap-frozen embryos and endosperm using TRIzol reagent (Invitrogen). Ten micrograms of total RNA was loaded on a 15% (wt/vol) polyacrylamide, 7 M urea gel, and small RNA (sRNA) in the 17- to 30-nt size range were excised. The 3′ miRNA cloning linker 1 (Integrated DNA Technologies, Inc.) was ligated to the gel-excised small RNA using truncated T4 RNA ligase 2 (New England BioLabs), and ligation products were purified on a 15% (wt/vol) polyacrylamide, 7 M urea gel. The 5′ Illumina RNA linker (5′-rArCrArCrUrCrUrUrUrCrCrCrUrArCrArCrGrArCrGrCrUrCrUrUrCrCrGrArUrCrU-3′) was added using T4 RNA ligase (New England BioLabs), and ligation products were again purified on a 15% (wt/vol) polyacrylamide, 7 M urea gel. Purified ligated sRNAs were reverse transcribed using SuperScriptIII (Invitrogen) and PCR amplified with Phusion High-Fidelity DNA Polymerase (New England BioLabs) for 25 cycles (forward primer: 5′-AATGATACGGCGACCACCGAACACTCTTTCCCTACACGACG-3′, reverse primer: 5′-CAAGCAGAAGACGGCATACGATTGATGGTGCCTACAG-3′). PCR products were gel-purified on a 2% agarose gel to obtain fragments between 100 and 120 bp in length for sequencing in either 36- or 50-bp reads on the Illumina HiSeq 2000 (crossed $F_1$ embryo and endosperm) and GA-II (self-fertilized embryo, endosperm, seedling root, and seedline shoot) platforms.

**Rice Genomic Sequences and Annotations.** All analyses were performed with either the Nipponbare rice reference genome (MSU 6.1) or a Kitaake pseudogenome built using the Nipponbare reference and the list of SNPs we identified in Kitaake, as described (1). The gene annotations we used were the MSU version 6.1 gene annotations, whereas repeats were annotated using RepeatMasker with the Viridiplantae Repbase database of repetitive sequences. To survey the distribution of differentially methylated regions (DMRs) and sRNAs within various genomic features, we assigned all 50-bp windows in the genome to one of five broad categories: exons, introns [excluding transposable element (TE)-derived repeat sequences that occur in gene bodies], intergenic regions, repeats (RepeatMasker-identified TE-derived repeat sequences, including those that occur in gene bodies), and regions at the boundaries of gene and repeat annotations (edge).

**Allele-Specific Mapping of Reads.** For bisulfite-sequenced DNA, reads were sorted to either the Nipponbare reference genome or Kitaake pseudogenome, and DNA methylation of cytosines within sorted reads was calculated as described (3, 4). For sRNA libraries, adapter sequences were trimmed from the reads, and resultant trimmed reads were sorted into size classes from 17 to 30 nt using custom Python scripts. Bowtie (5) was used to independently align each size class to both the Nipponbare genome and Kitaake pseudogenome, and custom Perl scripts (6) were used to sort reads to one or the other parental genome. Reads were assigned to a parental genome only in instances where a read aligned to both genomes in the same position, with zero mismatches to one and a single mismatch to the other. Abundances of total and parent-specific sRNAs across the genome were calculated in reads per million (RPM).

**Generation of Kernel Density Plots.** Kernel density plots of paternal-maternal differences in fractional methylation were generated using 50-bp windows where at least four cytosines were informative, and the fractional methylation of at least one of the parental genomes was greater than the defined threshold for a particular context. Thresholds were 0.7 for the CG context, 0.4 for the CHG context, and 0.01 for the CHH context. Kernel density plots of embryo-endosperm differences in fractional CG methylation were generated using 50-bp windows where at least 19 cytosines contributed to the overall methylation value in each cross and in which at least one of the tissues had a methylation value >0.7, and the methylation values of reciprocal crosses were within 0.1 of each other.

**Identification of sRNA-Producing Loci and Their Imprinting Status.** Tissue-specific loci of 24-nt sRNA expression were defined using two biological replicate libraries of Nipponbare (Nip) seedling shoot, two biological replicate libraries of Nip seedling root, embryo libraries from reciprocal crosses (one Nip × Kitaake library and one Kitaake × Nip library), and endosperm libraries from reciprocal crosses (using the same seeds as for embryo). Loci of 24-nt small RNA expression were defined by blocking together 50-bp windows with more than five read counts in both libraries from the same tissue type, when the windows were within 500 bp of each other. Where data from reciprocal crosses were available, the number of maternally sorted and paternally sorted 24-nt reads that mapped within these loci was recorded. Loci with 15 or more informative reads were evaluated for significant deviations from the expected maternal-paternal ratio using Fisher's exact test. Imprinted loci were only identified in the endosperm; maternally expressed loci were defined as loci with a significant bias ($p < 0.001$) and maternal-paternal ratio >3.3 in both crosses, whereas paternally expressed loci were defined as loci with a significant bias and maternal-paternal ratio <1 in both crosses. Loci that were not defined as maternally or paternally expressed were either defined to be biallelic on the basis of having a maternal-paternal ratio between 1.6 and 2.5 in both crosses or assigned to an "unclear" category of loci that showed variety-specific bias or were parentally biased but did not pass our significance criteria (Dataset S3).


1. Li H, Ruan J, Durbin R (2008) Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res* 18(11):1851–1858.
2. Panaud O, Chen X, McCouch SR (1996) Development of microsatellite markers and characterization of simple sequence length polymorphism (SSLP) in rice (*Oryza sativa* L.). *Mol Gen Genet* 252(5):597–607.
3. Hsieh T-F, et al. (2009) Genome-wide demethylation of *Arabidopsis* endosperm. *Science* 324(5933):1451–1454.
4. Zemach A, McDaniel IE, Silva P, Zilberman D (2010) Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science* 328(5980):916–919.
5. Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10(3):R25.
6. Hsieh T-F, et al. (2011) Regulation of imprinted gene expression in *Arabidopsis* endosperm. *Proc Natl Acad Sci USA* 108(5):1755–1762.

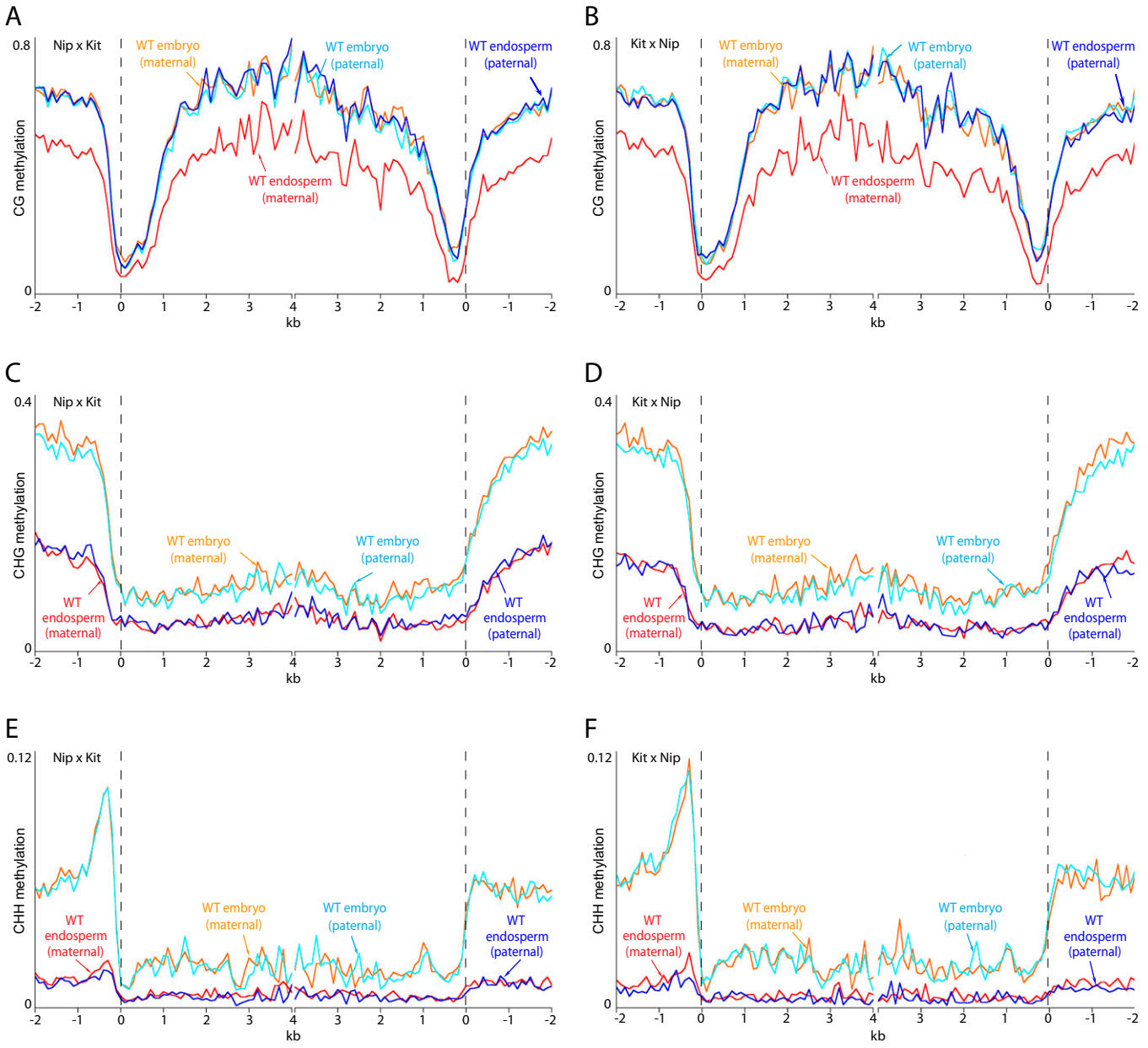


**Fig. S1.** (*A*–*F*) Genes were aligned at the 5′ end (*Left*) or the 3′ end (*Right*), and average methylation levels for each 100-bp interval were plotted for maternal and paternal genomes, from 2 kb away from the annotated region (negative numbers) to 4 kb into the annotated region (positive numbers). Dashed lines at zero represent the 5′ or 3′ point of alignment. CG methylation is shown in *A* and *B*, CHG in *C* and *D*, and CHH in *E* and *F*, with results from the Nipponbare × Kitaake cross in *A*, *C*, and *E* and those from the Kitaake × Nipponbare cross in *B*, *D*, and *F*.

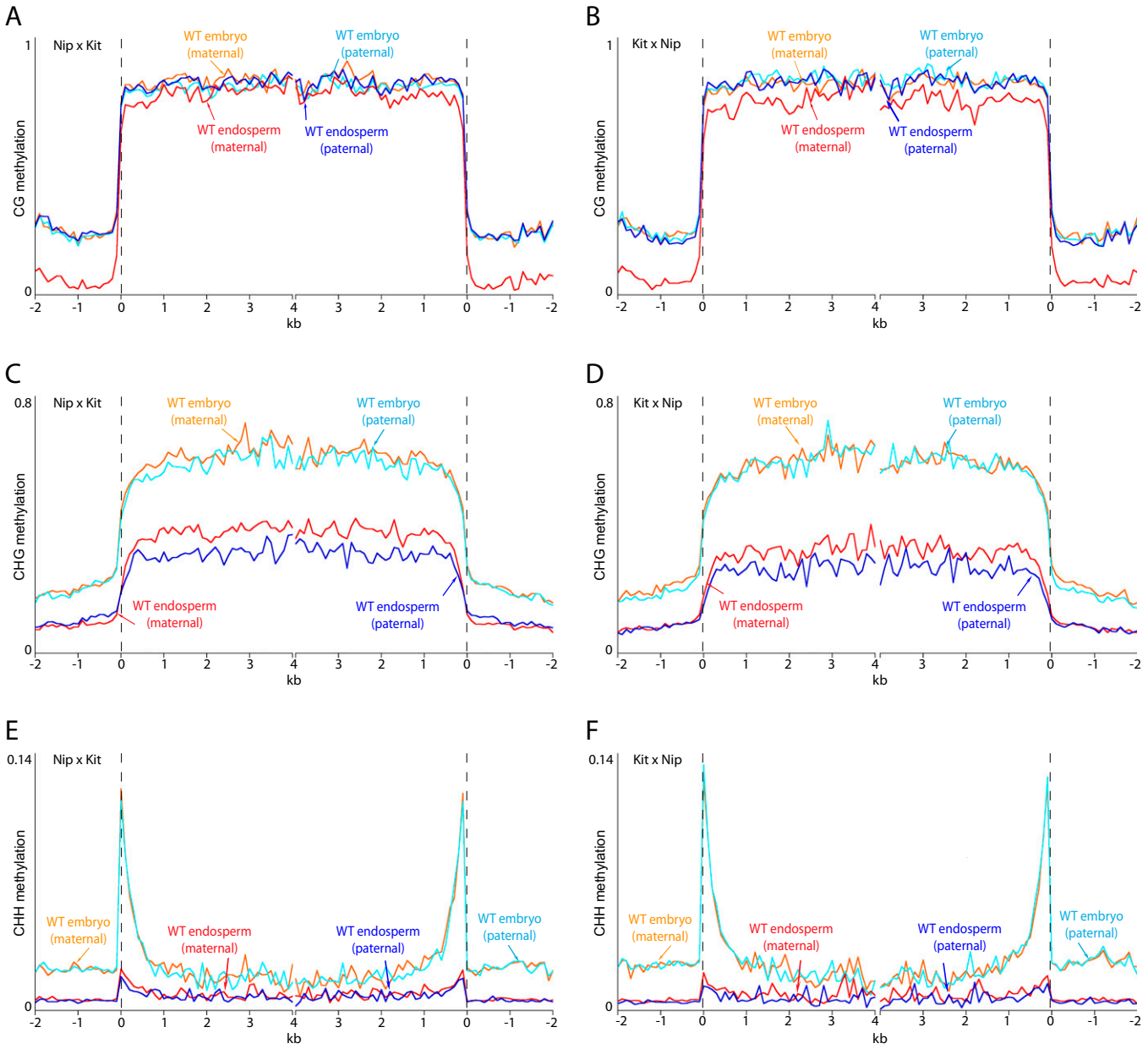


**Fig. S2.** TE repeats identified by RepeatMasker were aligned at the 5′ end (*Left*) or the 3′ end (*Right*), and average methylation levels for each 100-bp interval were plotted for maternal and paternal genomes, from 2 kb away from the annotated region (negative numbers) to 4 kb into the annotated region (positive numbers). Dashed lines at zero represent the 5′ or 3′ point of alignment. CG methylation is shown in *A* and *B*, CHG in *C* and *D*, and CHH in *E* and *F*, with results from the Nipponbare × Kitaake cross in *A*, *C*, and *E* and those from the Kitaake × Nipponbare cross in *B*, *D*, and *F*.

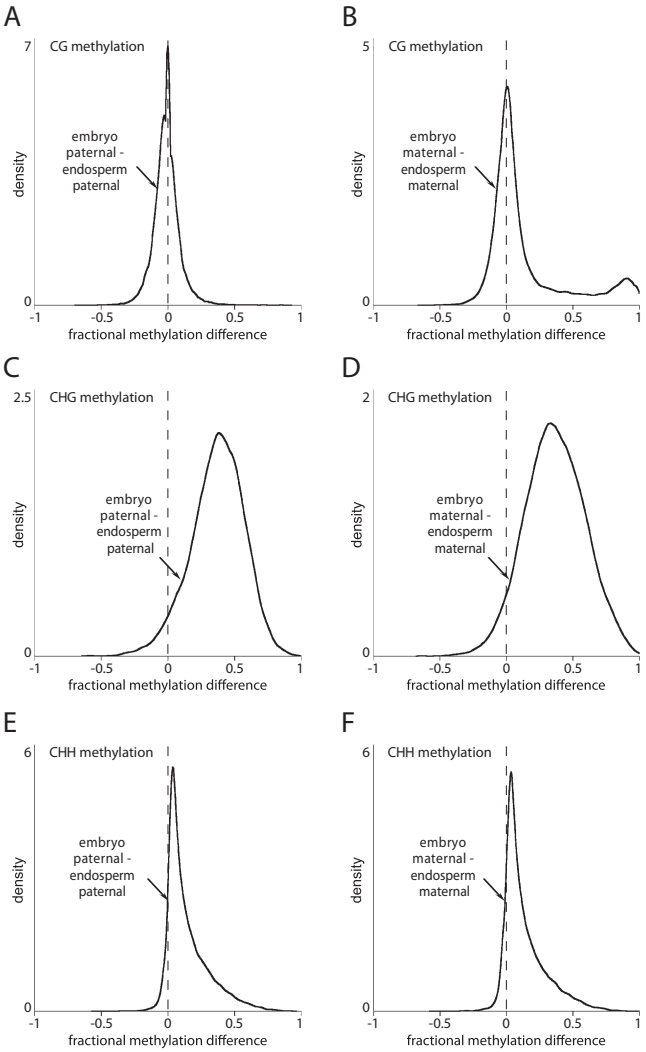


**Fig. S3.** Kernel density plots showing the frequency distribution of DNA methylation differences between embryo and endosperm across the paternal genome (*A*, *C*, and *E*) and maternal genome (*B*, *D*, and *F*). Differences were only plotted for 50-bp windows containing at least four informative sequenced cytosines and where fractional methylation of at least one of the parental genomes was >0.7 in the CG context (*A* and *B*), 0.4 in the CHG context (*C* and *D*), and 0.01 in the CHH context (*E* and *F*).

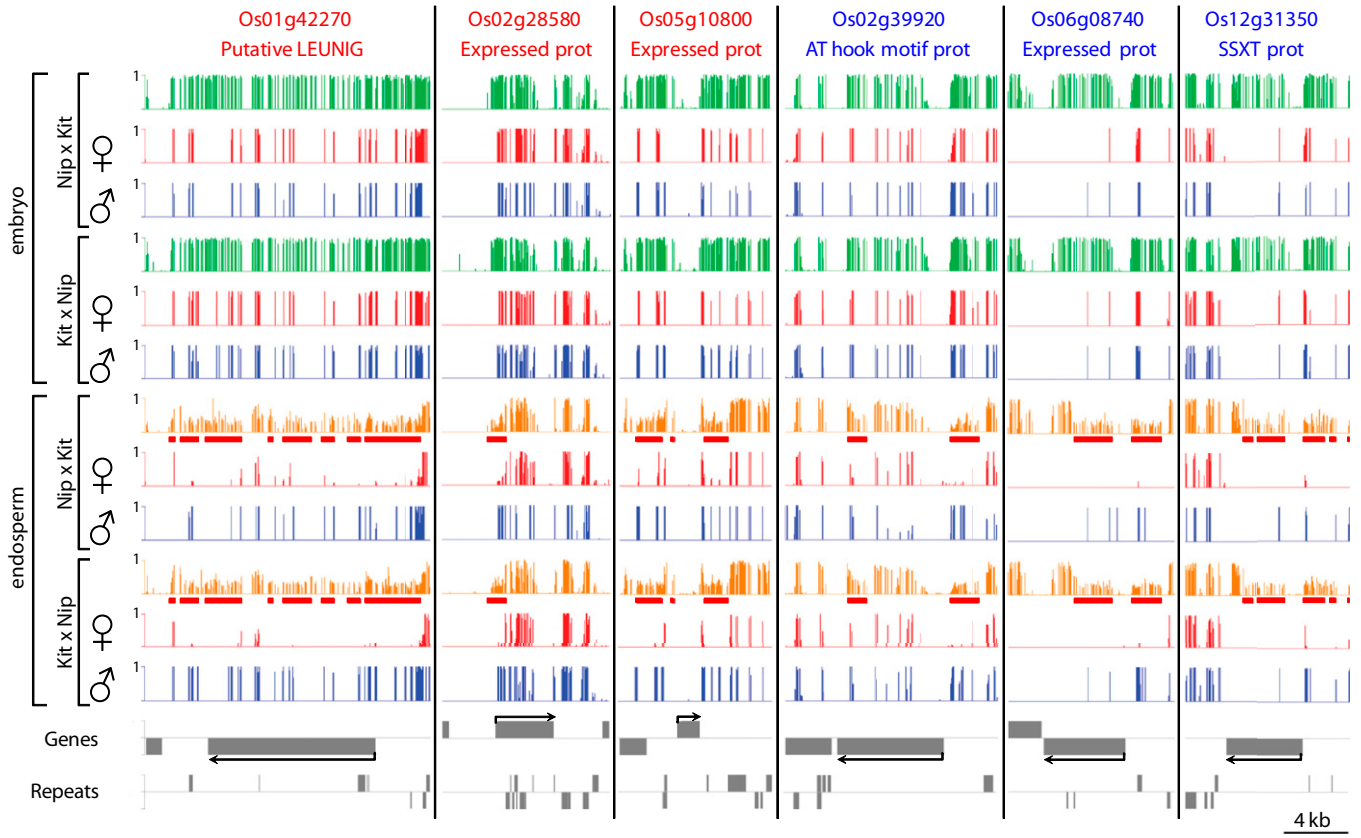


**Fig. S4.** Snapshots of CG methylation in indicated tissues near maternally expressed (red) and paternally expressed (blue) rice imprinted genes. Green bars represent embryo methylation, orange bars represent endosperm methylation, and red and blue bars represent methylation of the maternal and paternal genomes, respectively. DMRs identified between embryo and endosperm are underlined in red.

**Table S1. Coverage and mean DNA methylation in CG, CHG, and CHH contexts for libraries that were bisulfite-sequenced**

Table S1

Chloroplast CHH methylation is a measure of cytosine nonconversion and other errors. M/P, maternal/paternal; the expected ratio is 1 for embryo and 2 for endosperm.

**Table S2. Summary statistics for small RNA libraries**

Table S2

**Dataset S1. A total of 169,819 SNPs between Kitaake gDNA sequence and the Nipponbare MSU 6.1 reference genome**

Dataset S1

**Dataset S2. A total of 27,669 DMRs between rice embryo and endosperm**

Dataset S2

**Dataset S3. One hundred twenty-five defined loci of small RNA expression in endosperm with more than 15 parentally sorted reads from both rice reciprocal crosses**

Dataset S3

This table also includes information on the expression level of the locus, expression bias of the locus, and overlapping gene/s (if any).