Supporting Information

Lin et al. 10.1073/pnas.1716758114

SI Materials and Methods

Plant Growth and Tissue Collection. Soybean plants [*Glycine max* (L.) cv. Williams 82] were grown at 22 °C under long-day conditions with a 16-h light to 8-h dark cycle in the University of California, Los Angeles Plant Growth Center. Seeds were staged based on length, weight, and embryonic characteristics, such as shape and color (19). Hand-made drawings of soybean seeds and their corresponding EMBs at all developmental stages used in our experiments are shown in Fig. 1. glob, cot, em, mm, lm, pd1, and pd2 seeds had lengths of 1.0–1.5 mm, 3.0–3.5 mm, 6.0–7.0 mm, 11.0–12.0 mm, 12.0–14.0 mm, 8.0–10.0 mm, and 8–9 mm, respectively. In addition, mm, lm, pd1, pd2, and dry seeds weighed 150–250 mg, 230–350 mg, 150–260 mg, 120–150 mg, and 155 mg, respectively. mm EMBs had green COTL and a yellow AX tip, while lm EMBs were completely yellow (Fig. 1). sdlg were collected 6 d after imbibition, having primary roots, hypocotyls [average length 8 cm, flat green COTL (average length 19 mm)], and unifoliate leaves (Fig. 1). Postgermination COTL were dissected from sdlg 6 d after imbibition and weighed 250–400 mg. AX, COTL, and SC were manually separated from em and mm seeds (Fig. 3*B*). AX and COTL were harvested without the PL. Whole seeds and seed parts were harvested, frozen in liquid nitrogen, ground to a fine powder, and stored at −80 °C before DNA or RNA isolation.

Arabidopsis thaliana of ecotype Wassilewskija (Ws-0) were grown, and seeds were collected at glob, lcot, mg, pmg, and dry stages from 3-mo-old plants, as described in detail previously (38). Artistic renditions of *Arabidopsis* seeds at all stages of development are shown in Fig. 6*A*. EMB and SC were hand-dissected from mg-stage seeds (Fig. 6*D*). Leaves were collected from 4-wk-old plants. *A. thaliana* ecotype Columbia (Col-0) pmg seeds were collected from 3-mo-old *ddcc* mutant (15) and wild-type plants, and leaves were collected from 3-wk-old postgermination *ddcc* mutant and wild-type plants. Whole seeds and seed parts were harvested, frozen in liquid nitrogen, ground to a fine powder, and stored at −80 °C before DNA or RNA isolation.

LCM of Soybean Seed Regions and Tissues. em seeds were harvested, cut in half transversely, fixed in ethanol:acetic acid [3:1 (vol/vol)], dehydrated, infiltrated, and embedded in paraffin solution containing Paraplast-X-Tra tissue embedding medium (Fisher Scientific) according to the methods of Kerk et al. (41). Ten-micrometer paraffin cross-sections were prepared for each seed half using a Reicher-Jung 4 Rotary Histocut 820 Microtome, and floated in diethylpyrocarbonate-treated water to stretch ribbons containing seed serial sections. Seed sections were placed on PEN-foil slides (Leica Microsystems) and deparaffinized with two consecutive 2-min xylene treatments before LCM. Tissue sections were captured using a Leica LMD6000 microdissection scope into a PCR tube cap containing DNA isolation solution included in the FFPE DNA isolation kit (Qiagen).

Isolation of Endoreduplicated and Nonendoreduplicated COTL Tissues. The endoreduplication studies of Li and Nielsen (25) demonstrated that cells of the em-stage COTL ABPY tissue have undergone endoreduplication, whereas the COTL ADPY cells have not. To isolate parenchyma ABPY and ADPY tissues, ∼350 μm from the em-stage cotyledon ends was excluded from each cross-section, and only the middle sections of em seeds with COTL sizes 2,900–3,600 μm were used for LCM. The epidermis layer of the COTL was excluded from the ABPY collection, but was captured with the ADPY layer. Four ABPY cell layers and three ADPY cell layers were captured, respectively (Fig. 4*A*).

Isolation of SC Tissue Layers. To compare the methylation levels of different SC layers, SC-PY and SC-PA tissues were captured using LCM from the same em-stage paraffin sections used to capture the ABPY and ADPY tissue layers (Fig. 3*E*).

Isolation of cot-Stage Seed Parts. cot-stage seeds were harvested, fixed as described for em-stage seeds, and sectioned longitudinally. SC, AX, and COTL were captured from all sections using LCM (Fig. 3*A*). To avoid endosperm contamination, we used the laser to destroy the aleurone layer before capturing the SC.

Isolation of em-Stage Seed Parts, and AX Subregions and Tissues. em-stage seeds were harvested, fixed, and sectioned longitudinally. The entire SC and COTL were captured from all sections using LCM (Fig. 3*D*). PL, PA, PY, and RT were isolated from the same AX sections using LCM (Fig. S4*A*).

BS-Seq Library Construction. Approximately 100–1,000 ng of genomic DNA was used for BS sequencing library preparation following the methods of Hsieh et al. (10), with modifications. Specifically, 3 ng of unmethylated *cI857 Sam7* λ DNA (GenBank Accession no. NC_001416; Promega) was spiked-in with genomic DNA before DNA sonication to serve as an internal control for estimating BS conversion efficiency. Adapter-ligated genomic DNA was subjected to two rounds of BS treatment using the EpiTect Bisulfite Conversion kit (Qiagen). BS-treated DNA was purified and amplified for 10 cycles using ExTaq (Takara) DNA polymerase. PCR-amplified DNA fragments were size-selected using the AMpure XP beads (Beckman).

RNA-Seq Library Construction. Approximately 100 ng of soybean poly-A+ RNA was used for RNA-Seq library preparation according to the Illumina RNA-Seq Sample Prep Kit (Illumina). For *Arabidopsis* pmg wild-type and *ddcc* seeds, 25 ng total RNA was used to generate double-stranded cDNA using Ovation RNA-Seq System v2 (Nugen), and then 1 μg of double-stranded cDNA was used for RNA-Seq library preparation with the Illumina TruSeq DNA Sample Prep Kit (Illumina).

Small RNA-Seq Library Construction. Total RNA was isolated from em AX, COTL, and SC using the Concert Plant RNA Reagent (Invitrogen), according to the manufacturer's instructions. Appropriately 250 ng total RNA was used for the TruSeq Small RNA Sample Preparation kit (Illumina), and 15 PCR cycles were used for the final PCR enrichment step.

Illumina Next-Generation Sequencing. Single-end 50-bp reads were generated for the RNA-Seq and small RNA-Seq libraries, whereas 100-bp reads were generated for BS-Seq library using the Illumina Genome Analyzer IIx or HiSeq 2000 sequencing machines in the University of California, Los Angeles Genome Sequencing Center or Broad Stem Cell Research Center High Throughput Sequencing Core. A ϕX174 DNA control was spiked into each library by the sequencing facility before cluster formation and sequencing.

BS-Seq Data Processing and Analysis. Sequences were aligned to a reference genome using the BS Seeker program (45) allowing up to two mismatches. We used soybean genome build version Wm82.a1 (https://www.soybase.org) as a reference, which consists of scaffold sequences, including the 20 nuclear chromosomes, mitochondrial DNA, and unanchored sequences (43). Scaffolds containing chloroplast sequences were replaced with the 152,218-bp fully sequenced soybean chloroplast genome sequence (DQ317523) (49). For *Arabidopsis*, we used both TAIR10 Columbia (Col-0) (https://arabidopsis.org/index.jsp) and Ws-0 (mtweb.cs.ucl.ac.uk/mus/www/19genomes/) as reference genomes (44, 50). We compared the Col-0 and Ws-0 genomes and found that: (*i*) only 0.4% of cytosines were affected by single nucleotide polymorphisms that could affect their methylation status, and (*ii*) similar CG-, CHG-, and CHH-context bulk methylation results across seed development were obtained with both *Arabidopsis* ecotypes. That is, the results shown in Fig. 6 are valid for both Col-0 and Ws-0 ecotypes. Finally, we used the 48,502-bp *cI857 Sam7* λ genome (NC_001416) (51) and the 5,386-bp ϕX174 genome (NC_001422) (52) to map reads from our λ and ϕX174 DNA controls.

Only reads that mapped uniquely to the reference genomes were retained for detailed analysis. The BS-Seeker output was subjected to postprocessing that consisted of two steps: (*i*) clonal reads, or reads containing identical 5′ mapped positions and exact nucleotide sequences, were collapsed and all but one read was retained to reduce PCR amplification bias for each library; and (*ii*) reads containing three or more consecutive cytosines in the CHH context were removed as they are likely not bisulfite converted (32).

Analysis of BS Conversion Efficiency. During BS conversion experiments, unmethylated cytosines might not be converted to uracil (thymine) and will appear erroneously as methylated cytosines in the final sequencing reads. To assess the extent of nonconversion during BS treatment, we spiked in unmethylated λ DNA into each genomic DNA sample during the library preparation step. We obtained at least 300× coverage of the λ genome, and detected up to 99.98% of the genomic cytosines, representing comprehensive coverage of the λ genome. Overall, we obtained excellent BS conversion efficiency of unmethylated cytosine to uracil (thymine), averaging 99.55% over 49 BS-Seq libraries, which were similar to conversion rates obtained by others using the same BS kits (42) (Dataset S1).

Estimate of Genome Sequencing Coverage. Because the soybean genome is an ancient polyploid with >60% of the genome represented by repetitive sequences (43), we asked what fraction of the nuclear genome sequences we could detect and map uniquely with 100-bp sequencing reads. Using a sliding-window approach, we generated >950 million 100-bp reads with a 99-bp overlap covering the entire genome. We next collapsed and removed identical redundant reads leaving 805 million (∼85%) reads that are unique. The 805 million reads were aligned to the genome using BS-Seeker to mimic our data processing pipeline allowing for no mismatch. Approximately 782 million reads aligned to the genome, uniquely covering 857 million bases, or 90% of the nuclear genome, including ∼325 million cytosines or 90% of the genomic cytosines. These results suggest that although soybean is an ancient polyploid with a large repeat-rich genome, most of the sequences have diverged significantly and can be distinguished clearly from 100-bp reads using BS-Seeker, indicating that our whole-genome BS sequencing can interrogate most, if not all, of the genome sequences. We used 90% of cytosines in the soybean genome as the basis for the calculation of cytosines detected in Dataset S1 by our BS sequencing. We didn't carry out an analogous simulation for *Arabidopsis*, as it contains much fewer repetitive sequences in its genome, and used the total number of genomic cytosines as the basis for the numbers in Dataset S1 (44).

Determination of Whether a Genomic Site Is Methylated or Not. Whole seeds were used for most of our BS-Seq experiments and, therefore, the BS-Seq reads represent the genomes of different cells within the seeds. If the methylation status of a given cytosine site in the different cell types differs, then BS-Seq will detect both the methylated and unmethylated cytosines. That is, not all equivalent genomic sites within different seed cells might be methylated. For example, it is possible that within a seed only a fraction of the cells might be methylated at a specific cytosine location resulting in a mixture of methylated and unmethylated cytosines in the BS-Seq data for that genomic site. This also applies to tissues captured by LCM, as not all cells within a tissue might have the same developmental equivalency and methylation states.

We used the following approach to assign the methylation status of a given seed genomic site. For each cytosine site in the reference genome, the total number of cytosine reads (representing methylated cytosines) and thymine reads (representing unmethylated cytosines) were summarized using custom scripts. We then eliminated cytosine sites that had only one mapped read (cytosine or thymine). Next, we used a statistical test (binominal test, $P < 0.05$) and only cytosines that passed this filter were used for downstream analysis to ensure that they were not false positives because of: (*i*) nonconversion of unmethylated cytosines to thymines (i.e., they remain cytosines after BS treatment and scored falsely as methylated in the BS-Seq analysis) and (*ii*) the amount of sequence coverage (53). We used 0.5% as the value for nonconversion of unmethylated cytosines in the binomial test, which was similar that obtained in our λ genome spiked-in controls (Dataset S1). In general, genomic sites with two or more cytosine reads passed our statistical filter and were considered methylated. By contrast, genomic sites were considered unmethylated if there was at least one thymine BS-Seq read, indicating that BS converted an unmethylated cytosine to uracil (thymine) at that site. By using these criteria, we were able to assign the methylation status of each cytosine site in the genome regardless of which seed cell from which it was derived.

Calculation of Average or Bulk Methylation Levels Across the Entire Seed. Bulk methylation is the average cytosine methylation percentage for all filtered reads, irrespective of their genomic sites, and represents the average methylation levels across all genomes within the entire seed or tissue. Bulk methylation levels for all cytosines (C), and those in CG, CHG, and CHH contexts, were calculated by using the formula $[C/(C + T) \times 100]$, where C and T represent total cytosine (methylated) or thymine (unmethylated) reads (10). For example, 10 methylated cytosines of 100 detected cytosines in all reads mapped within a specific genomic element [e.g., 500-kb genomic windows (soybean), 100-kb windows (*Arabidopsis*), genes, or TEs] is a 10% bulk cytosine methylation level. Bulk methylation levels were used for construction of methylation box plots and pair-wise seed-stage comparisons presented in this paper (e.g., see Fig. 2*A*). Each box plot represents the middle 50% of methylation levels. The black bar in Fig. 2*A* indicates the median methylation level, while the whiskers indicate 1.5 times the box length.

Visualization of Methylation Levels. The methylation levels of cytosine sites were converted to bigwig format and viewed using the Integrative Genomics Viewer (54) (e.g., Fig. 5*B*). The heat maps in Fig. 2*B* and Fig. 6C were drawn using Circos (55). In the heat maps, chromosome, centromere, and pericentromere coordinates were obtained from Phytozome (phytozome.net) for soybean or from TAIR10 (www.arabidopsis.org) for *Arabidopsis*. Tracks representing the densities of genes and TEs in 500- or 100-kb windows along the genome were based on soybean version W82.a1.v1.1 annotations from SoyBase (https://www.soybase.org), the SoyTEdb database (56), or the TAIR10 genome release (www.arabidopsis.org).

RNA-Seq Data Processing and Analysis. Illumina raw reads were first filtered using the Illumina purity filter, and then trimmed at their 5′ and 3′ ends based on positions with error rates >0.1%. rRNA reads were identified by mapping trimmed reads against a rRNA database using Bowtie (v0.12.7) and then removed from further analysis (46). The remaining high-quality reads were mapped to either the soybean or *Arabidopsis* reference genomes and cDNA models using Bowtie, allowing up to two mismatches. Only uniquely mapped reads (i.e., reads that map to one unique genomic locus) were used for subsequent analysis. Read counts for each gene model were computed using a customized script. RNA-Seq expression values for each gene within a dataset were normalized as RPKM (57). Genes and TEs were identified as differentially expressed in the *ddcc* mutant relative to wild-type with the EdgeR package (v3.18.1) (47) using the following filtering parameters: (*i*) genes: a corrected *P* value < 0.001 and >fivefold difference; and (*ii*) TEs: a corrected *P* value < 0.01 and detection in both replicates. We used BEDTools (58) to obtain normalized coverage per base from the alignments of RNA-Seq data, converted to BAM format, and viewed using the Integrative Genomics Viewer (54). VirtualPlant 1.3 (59) was used for *Arabidopsis* gene Gene Ontology enrichment analysis with a cut-off FDR value < 0.01. Proteins encoded by the 106 up-regulated and de-repressed TE RNAs were determined using the corresponding reads as query sequences in translated BLAST (blastx) against *Arabidopsis* TE proteins in the National Center for Biotechnology Information protein database (https://www.ncbi.nlm.nih.gov/protein/) with 1E-4 as a cut-off criterion.

Small RNA Sequence Processing and Analysis. Quality filtered small RNA sequences were trimmed to remove the adapter sequences. Trimmed reads that mapped to soybean ribosomal RNA (rRNA) and transfer RNA sequences were removed, and then 18- to 24-nt reads were kept for further analysis. These reads were aligned to the soybean genome (version Wm82.a1) using Bowtie (v0.12.7) with no mismatches, but allowing for matches to multiple positions within the genome (60). The filtered reads were then sorted into 21-, 22-, and 24-nt siRNAs (Fig. 3*G*). We used BEDTools (58) to obtain normalized counts within TEs from all mapped sequences for 21-, 22-, and 24-nt small RNAs in em-stage AX, COTL, and SC seed parts (Fig. 3*G*).

Copy Number Analysis. The sequencing read depth method of Yoon et al. (61) was used to detect copy number variation between: (*i*) soybean seed endoreduplicated and nonendoreduplicated ABPY and ADPY COTL regions (Fig. 4*C*) and (*ii*) *Arabidopsis* TEs in *ddcc* and wild-type pmg seeds (Fig. 9*E*). We used BEDTools (58) to obtain normalized read depth per base from the alignments of BS-Seq data. The average normalized read depth in a soybean genomic window (500 kb), or an *Arabidopsis* TE, was calculated from single base normalized read depth.

***Arabidopsis* Nuclear Size Assay.** Nuclear size assays were performed according to Moissiard et al. (62) with the following modifications. Two hundred COTL from hand-dissected *Arabidopsis* pmg seeds, or 0.5 g of leaves from 2-mo-old plants were fixed in Tris buffer (10 mM Tris pH 7.5, 10 mM EDTA, 100 mM NaCl) containing 4% paraformaldehyde for 20 min and washed twice in Tris buffer without paraformaldehyde. Samples were ground in 45 μL lysis buffer (15 mM Tris pH 7.5, 2 mM EDTA, 0.5 mM spermine, 80 mM KCl, 20 mM NaCl, 0.1% Triton X-100) using a glass grinder and filtered through a 35-μm cell strainer. The suspension containing nuclei was added to sorting buffer (100 mM Tris pH 7.5, 50 mM KCl, 2 mM $MgCl_2$, 0.05% Tween-20, 20.5% sucrose) and transferred to slides to air dry for 1 h. Slides were postfixed in a PBS (10 mM sodium phosphate, pH 7, 143 mM NaCl) containing 4% paraformaldehyde for 20 min followed by three washes with PBS alone. The mounting medium Vectashield containing DAPI (Vector Laboratories H-1200) was added to the slides to hold nuclei in place between the coverslip and the slide and to stain the nuclei. Nuclei were observed using a Zeiss Imager D2 microscope and more than 110 nuclei were analyzed for each genotype.

***Arabidopsis* Seed Morphology and Germination Assays.**

Seed development. The development of *Arabidopsis* wild-type and *ddcc* seeds was characterized using the procedures of Yadegari et al. (63) with the following modifications. Siliques were fixed in ethanol:acetic acid [9:1 (vol/vol)] overnight followed by two washes in 90% and 70% ethanol for 1 h, respectively. Siliques were cleared with a chloral hydrate/glycerol/water solution [8:1:2, (w/vol/vol)] for 1 h, and seeds were visualized using a Zeiss Imager D2 microscope equipped with Nomarski optics.

Seed germination analysis. *Arabidopsis* dry seeds were washed with 70% ethanol twice, 50% bleach twice, and sterilized water four times, then resuspended in sterilized water and put in refrigerator for 3 d. Fifty seeds were paced in a grid alignment on five replicate agar plates (Phytoblend; Caisson Labs) containing Murashige and Skoog medium, pH 5.7. Plates were put in a growth chamber at 22 °C with a 16-h light to 8-h dark cycle, and sdlg numbers were counted after 4 d.

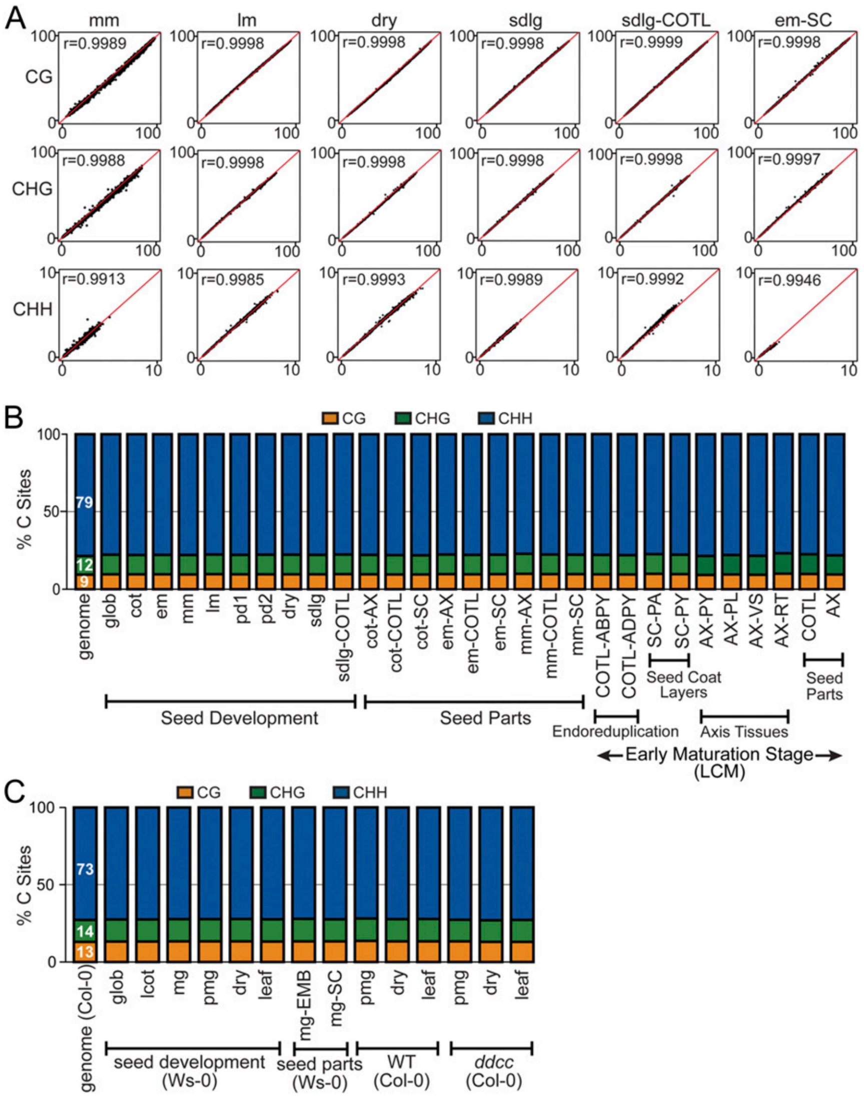

Fig. S1. Quality of BS-Seq methylome libraries. (*A*) Correlation coefficients between biological replicates of soybean seed BS-Seq libraries. The average methylation levels in 500-kb windows across the genome from biological replicates with similar sequencing depths were used to determine the correlation coefficients. Proportions of cytosine bases in CG, CHG, and CHH contexts for the soybean genome (*B*) and *Arabidopsis* genome (*C*) were determined from the BS-Seq results reported here, the soybean genome sequence (version Wm82.a1) (https://www.soybase.org) (43), and the *Arabidopsis* genome sequence (version TAIR10) (https://www.arabidopsis.org/index.jsp). See Table 1 for definition of abbreviations.

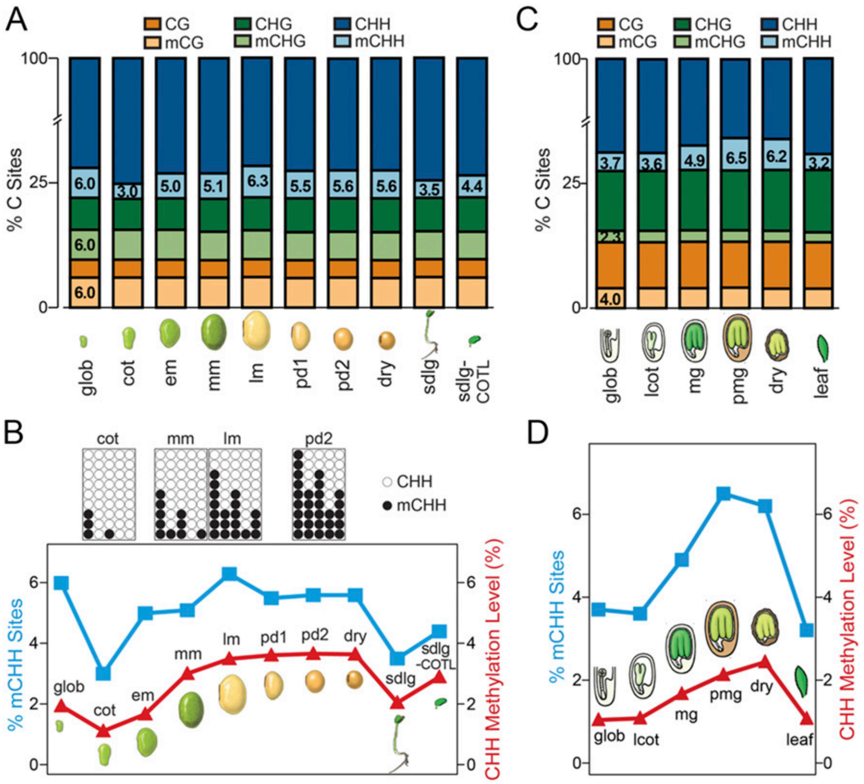

Fig. S2. DNA methylation changes during seed development and germination. (*A*) Changes in the proportion of methylated cytosine sites in CG, CHG, and CHH contexts during soybean seed development. CG, CHG, and CHH represent unmethylated cytosines, whereas mCG, mCHG, and mCHH indicate methylated cytosines. Numbers represent the proportions of methylated cytosine sites in each sequence context during seed development. (*B*) Comparison between the changes in methylated cytosine sites (blue) and bulk methylation percentages (red) in CHH context during soybean seed development. Methylated cytosine site data were taken from values in *A*, and those for bulk methylation percentages were taken from Dataset S2. The latter represents mean DNA methylation levels in 500-kb windows across the soybean genome. Boxes represent a heuristic model for the changes in methylated cytosine sites during seed development. Each circle along the bottom row represents one cytosine site in the genome, whereas the vertical circles represent cytosines in the same genomic positions in different seed cells. During seed development individual CHH-context sites across the genome become methylated (dark circles), resulting in an increase in the absolute number of methylated CHH sites. These same sites become methylated stochastically in different seed cells during development, resulting in an increase in bulk methylation percentage. (*C*) Changes in the proportion of methylated cytosine sites in CG, CHG, and CHH contexts during *Arabidopsis* seed development similar to those presented for soybean seeds in *A*. (*D*) Comparison between the changes in methylated cytosine sites (blue) and bulk methylation percentages (red) in CHH context during *Arabidopsis* seed development similar to that presented for soybean in *B*. See Table 1 for abbreviations of soybean and *Arabidopsis* seed stages.

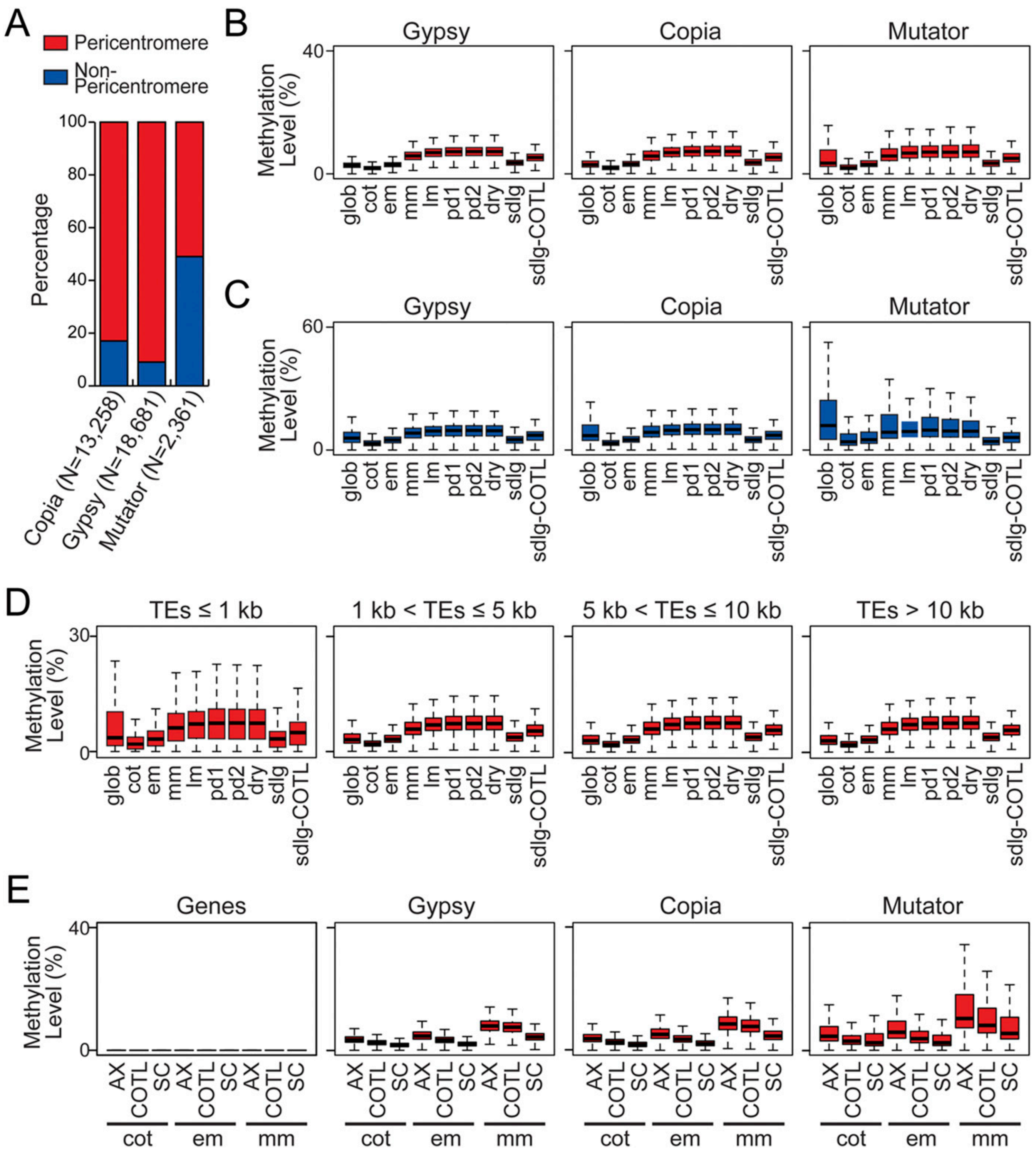

Fig. S3. Changes in CHH-context DNA bulk methylation levels in different TE classes during soybean seed development and germination. (*A*) Proportion of TE classes in chromosomal pericentromeric and arm regions. Box plots of methylation levels for different TE classes in pericentromere regions (*B*) and arm regions (*C*) during seed development. (*D*) Box plots of methylation levels for different-sized TEs across seed development. (*E*) Box plots of methylation levels for genes and TEs in different seed parts. See Table 1 for abbreviations of seed stages and parts.

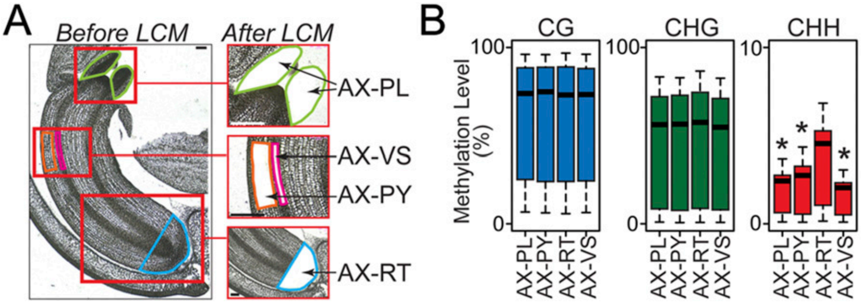

Fig. S4. Comparison of methylation levels within different soybean AX regions. (*A*) Paraffin sections of early maturation stage AX-PL, AX-PA, AX-VS, and AX-RT regions before and after capture by LCM. (Scale bar, 200 μm.) (*B*) Box plots of DNA methylation levels in 500-kb windows across the genome in different AX tissues. Asterisks indicate significant comparisons between RT and other AX tissues (*t* test, $P < 0.001$ and fold change > 1.5) (Dataset S3).

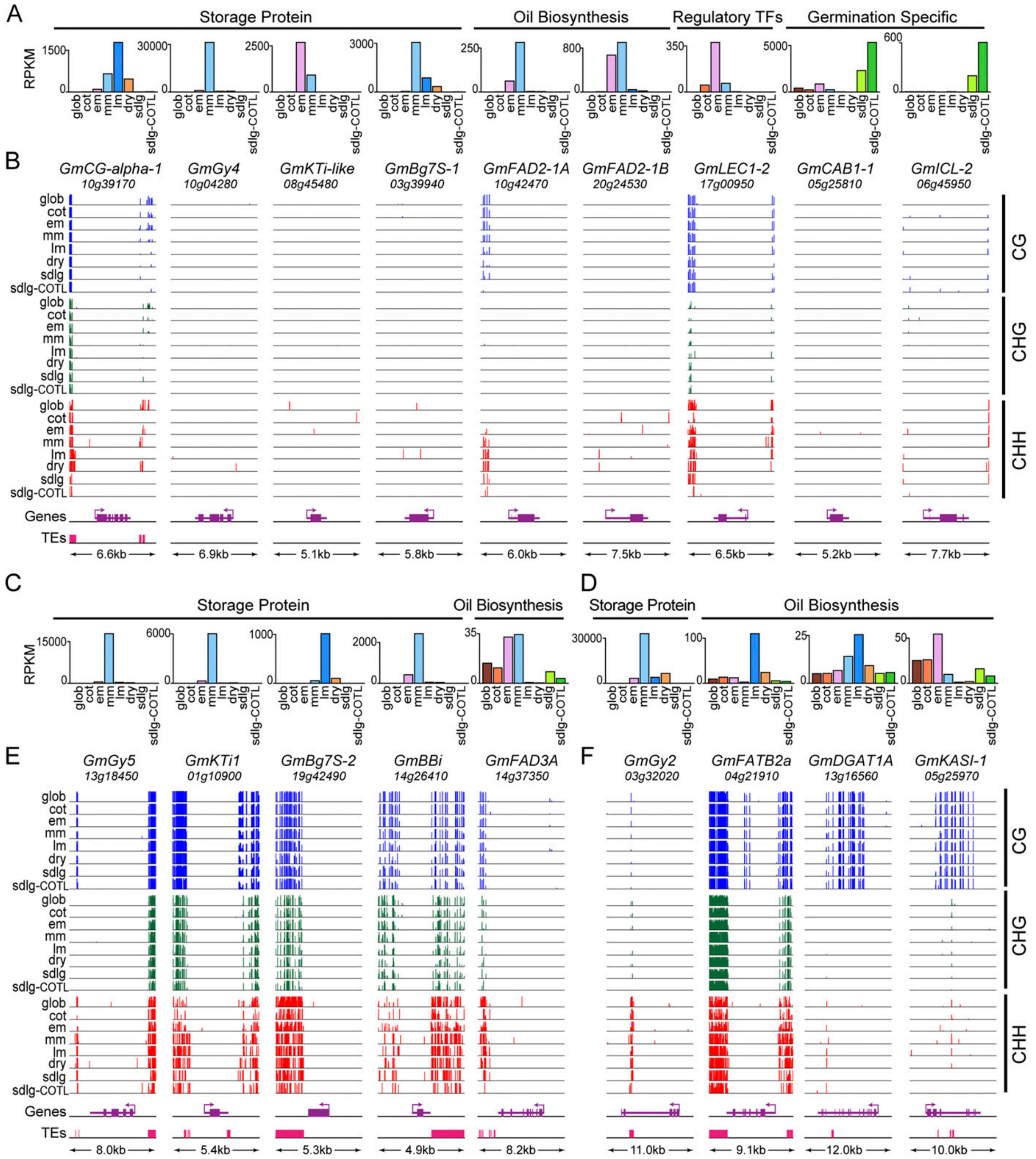

Fig. S5. Methylation levels and mRNA accumulation patterns of major soybean seed-specific gene classes during seed development and germination. (*A*, *C*, and *D*) mRNA accumulation levels for major seed and germination mRNAs. RPKM were taken from the Goldberg-Harada soybean (*i*) whole-seed RNA-Seq dataset GEO number GSE29163 (37), and (*ii*) cotyledon-specific RNA-Seq dataset GSE29134 (sdlg-COTL). (*B*, *E*, and *F*) Methylation levels of CG-, CHG-, and CHH-context cytosine sites are shown in genome browser view (vertical lines). Genes in *A* and *B* have either zero or <5% cytosine methylation in their gene bodies and at least 1 kb of upstream and downstream regions. Genes in *C* and *E* have no detectable cytosine methylation in their gene bodies, but have methylated TEs within 1 kb of flanking gene regions. Genes in *D* and *F* have methylated cytosines within their gene bodies. Additional soybean seed and germination genes with similar methylation patterns [described as classes (*i*), (*ii*), and (*iii*)] are listed in Dataset S4. Gene structures, transcription directions (arrows), and TEs are shown below each genome browser view. Adjacent genes are not shown. The size of each genomic region, including 2 kb of gene flanking region, is shown at the bottom of the browser views. *GmBBi*, Bowman Birk inhibitor; *GmBg7S-1*, basic 7S globulin-1; *GmBg7S-2*, basic 7S globulin-2; *GmCAB1-1*, Chlorophyll A/B Protein 1-1; *GmCG-α-1*, *Gm*β-conglycinin-α-1; *GmDGAT1A*, diacylglycerol acyl-transferase 1A; *GmFAD2-1A*, oleoyl desaturase 2-1A; *GmFAD2-1B*, oleoyl desaturase 2-1B*; Gm FAD3A*, linoleoyl desaturase 3A; *GmFATB2a*, fatty acyl-ACP thioesterase B2a; *GmGy2*, glycinin2; *GmGy4*, glycinin4; *GmGy5*, glycinin5; *GmICL-2*, isocitrate lyase-2; *GmKASI-1*, 3-ketoacyl-ACP synthase I-1; *GmKTi1*, Kunitz trypsin inhibitor 1; *GmKTi-like*, Kunitz trypsin inhibitor-like; *GmLEC1-2*, Leafy cotyledon 1-2. See Table 1 for developmental stage abbreviations.

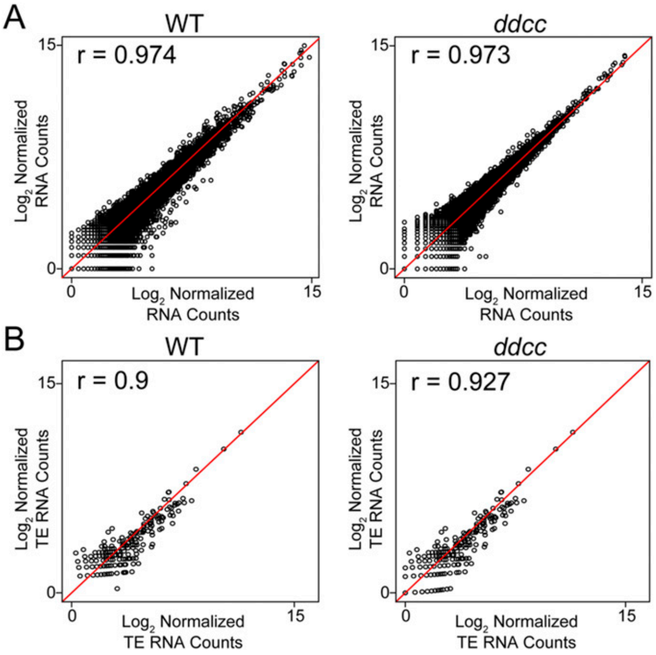

Fig. S6. Comparison between biological replicates of *Arabidopsis* wild-type (WT) and *ddcc* pmg seed gene (*A*) and TE (*B*) transcriptomes.

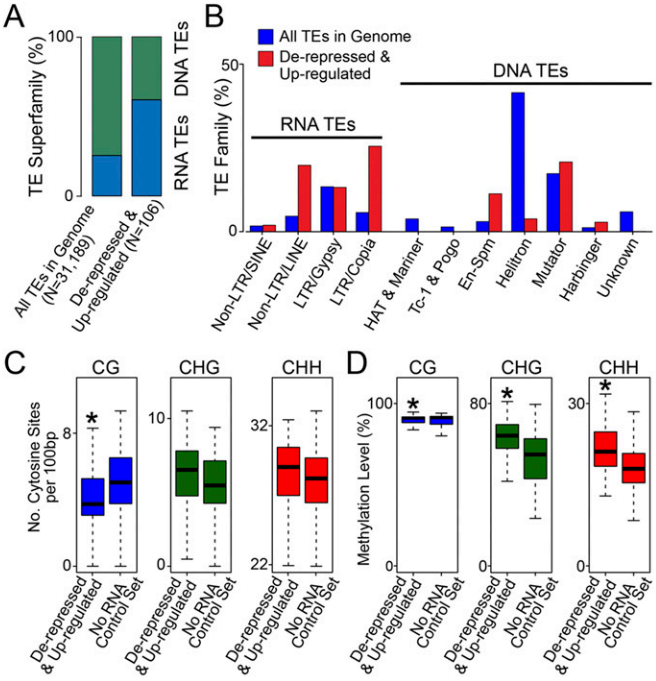

Fig. S7. De-repressed and up-regulated TEs in *Arabidopsis ddcc* pmg seeds. Representation of RNA and DNA TEs (*A*) and individual TE families (*B*) in *ddcc* seed de-repressed and up-regulated TEs. (*C*) Box plots of the number of cytosine sites per 100 nt for each sequence context in *ddcc* seed de-repressed and up-regulated TEs. (*D*) Bulk DNA methylation level box plots for each sequence context in *ddcc* seed de-repressed and up-regulated TEs. The no RNA control used in *C* and *D* represent 106 randomly selected TEs, which have (*i*) no detectable RNA wild-type reads and (*ii*) similar TE family distribution and lengths compared with the 106 de-repressed and up-regulated TEs (Fig. 9). The asterisks indicate statistically significant comparisons between the no RNA control TEs and de-repressed and up-regulated TEs (*t* test, $P < 0.01$).

Other Supporting Information Files

Dataset S1 (XLSX)
Dataset S2 (XLSX)
Dataset S3 (XLSX)
Dataset S4 (XLSX)
Dataset S5 (XLSX)
Dataset S6 (XLSX)
Dataset S7 (XLSX)
Dataset S8 (XLSX)
Dataset S9 (XLSX)