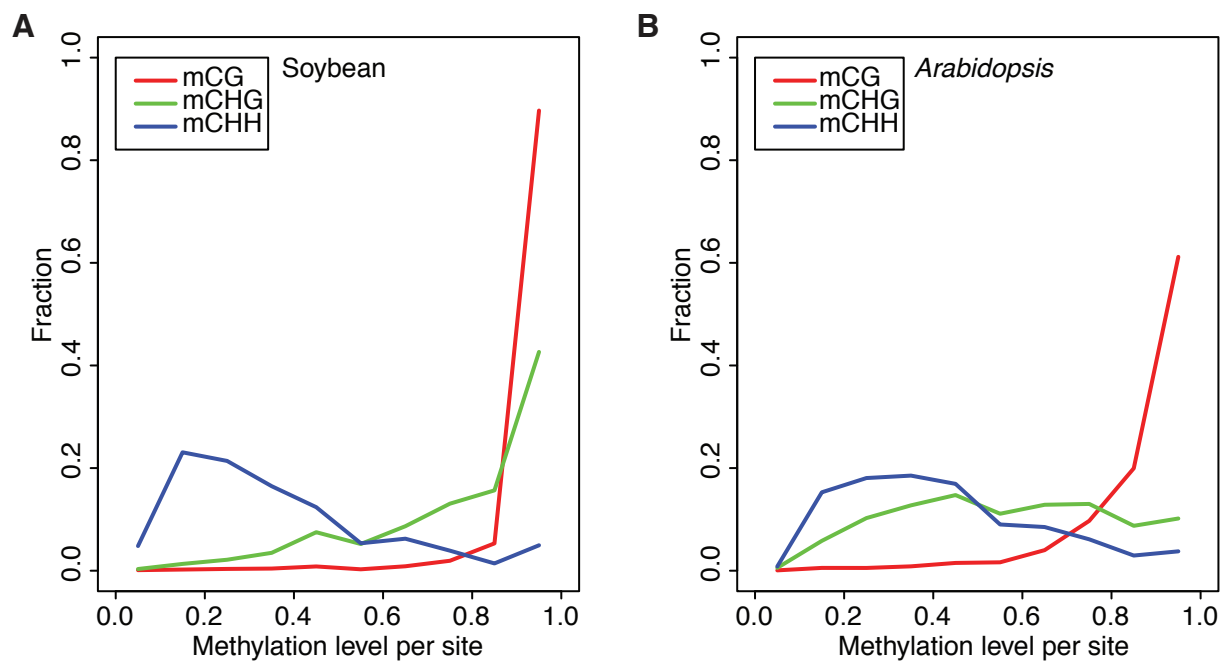
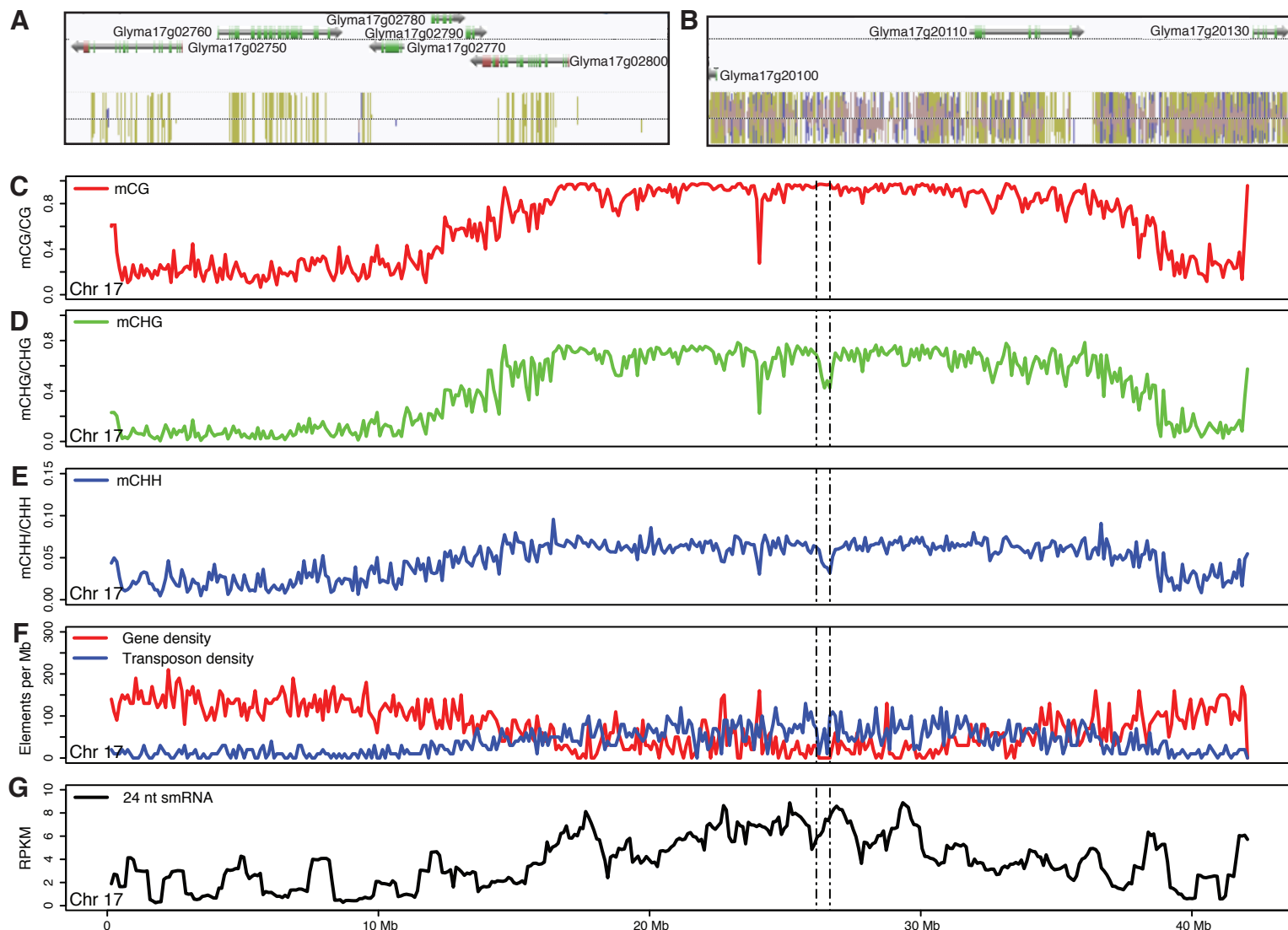
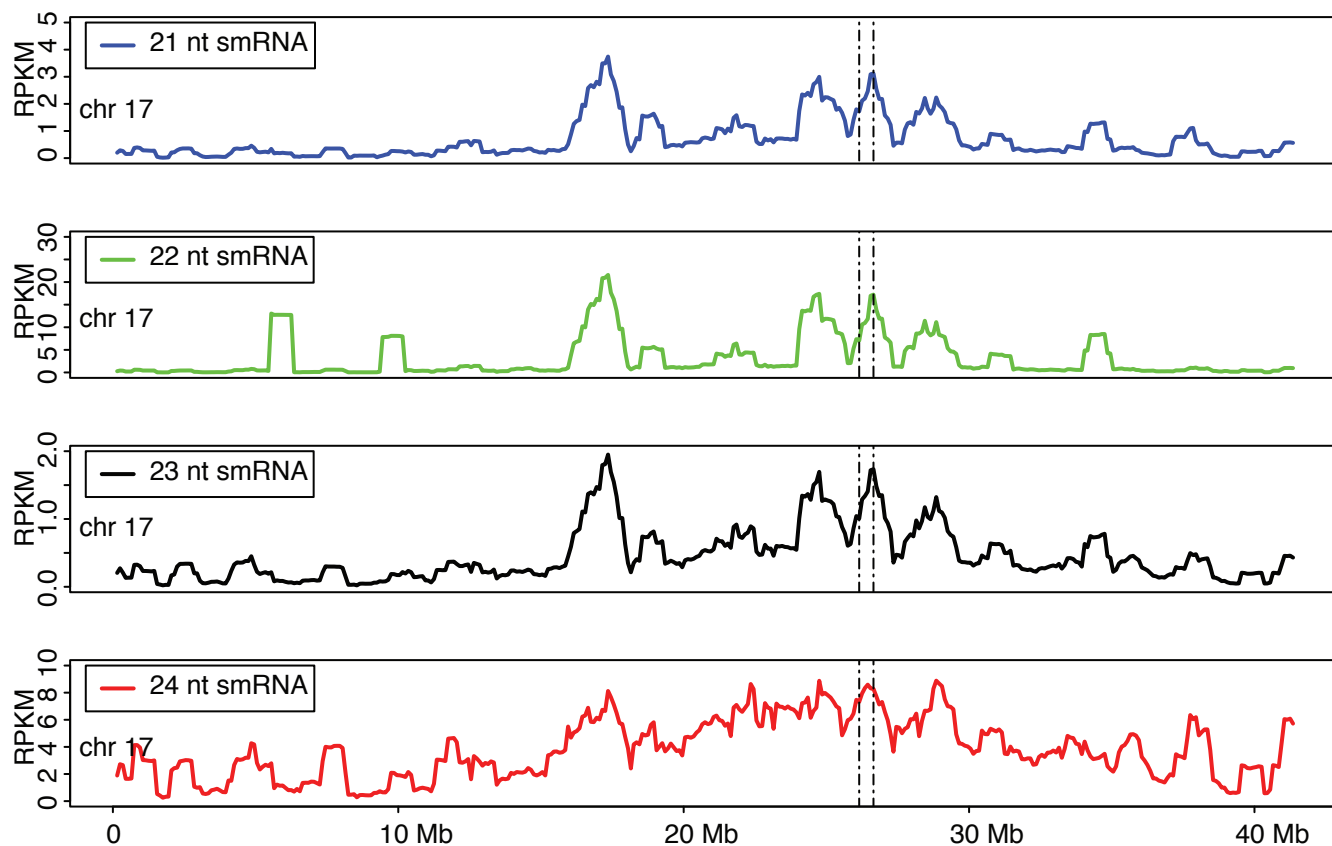




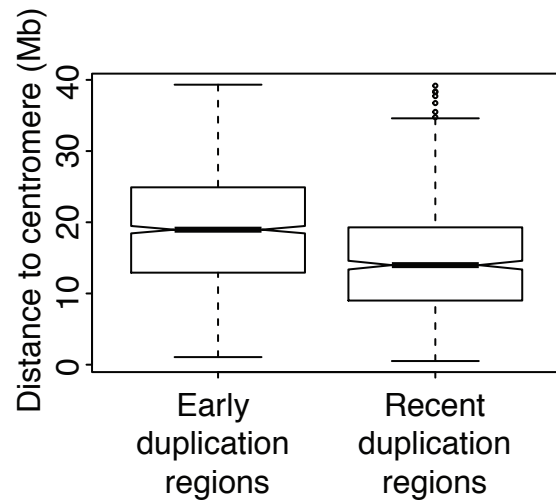
Supplementary Fig. 1. (A and B) The distribution of methylation levels at each mCG, mCHG, and mCHH site reveals similar patterns between *Arabidopsis thaliana* and soybean.



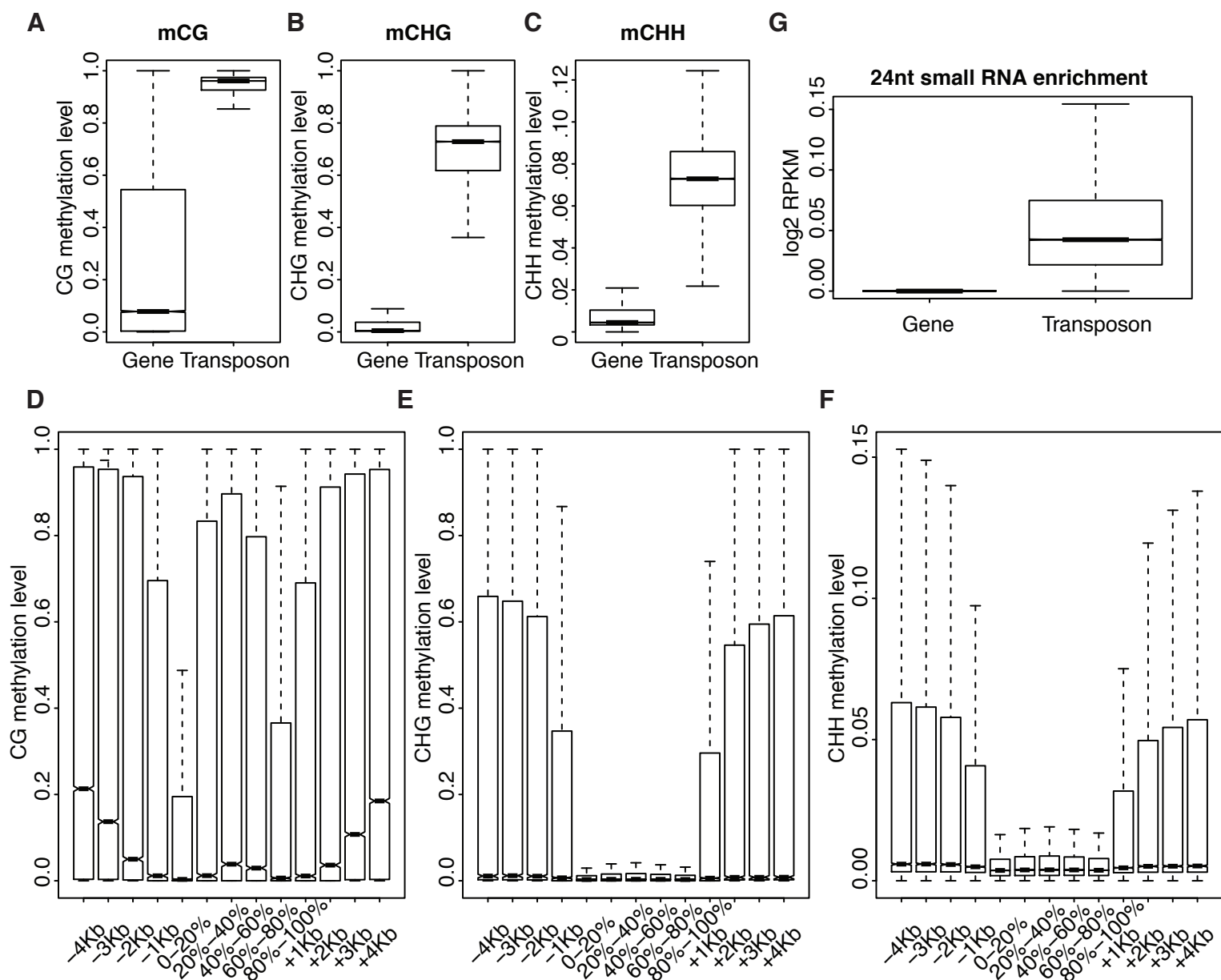
Supplementary Fig. 2. (A and B) Representative DNA methylation patterns in **(A)** euchromatic and **(B)** heterochromatic regions of the chromosome. Gold lines indicate mCG, purple lines indicate mCHG and pink lines indicate mCHH. Chromosome-wide density of **(C)** CG DNA methylation, **(D)** CHG DNA methylation, and **(E)** CHH DNA methylation. **(F)** The chromosome-wide distribution of transposon and gene density. **(G)** The chromosome-wide density of 24 nt smRNAs.



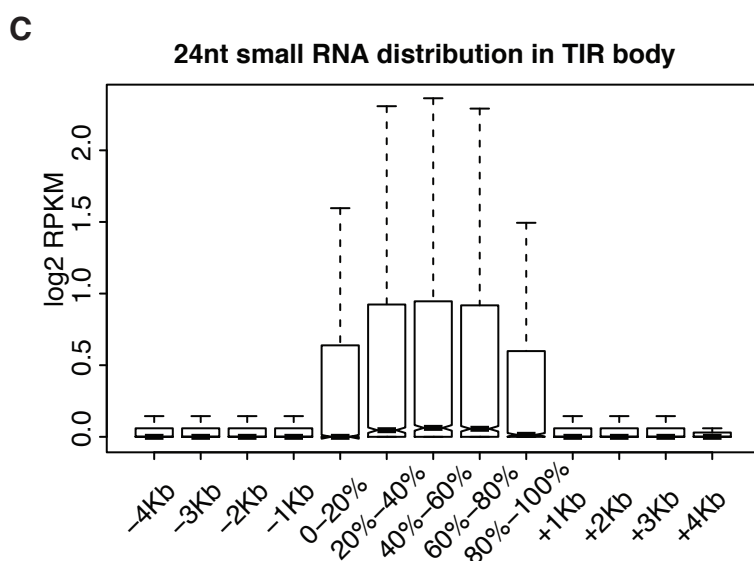
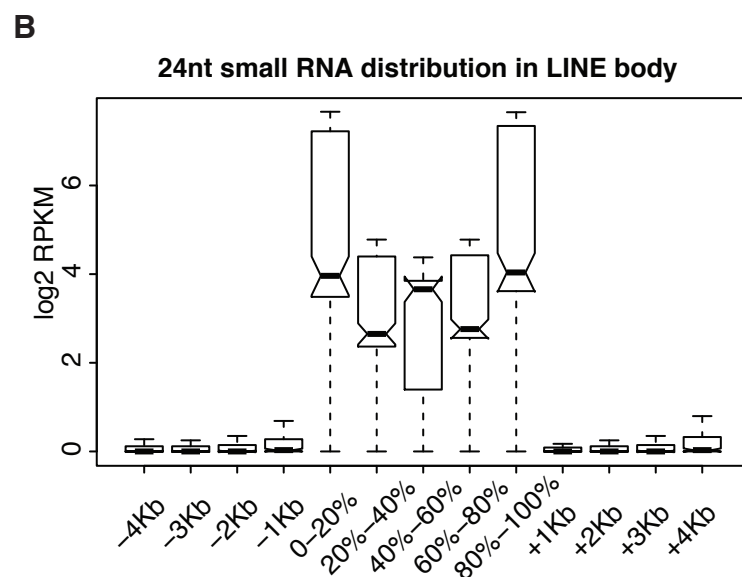
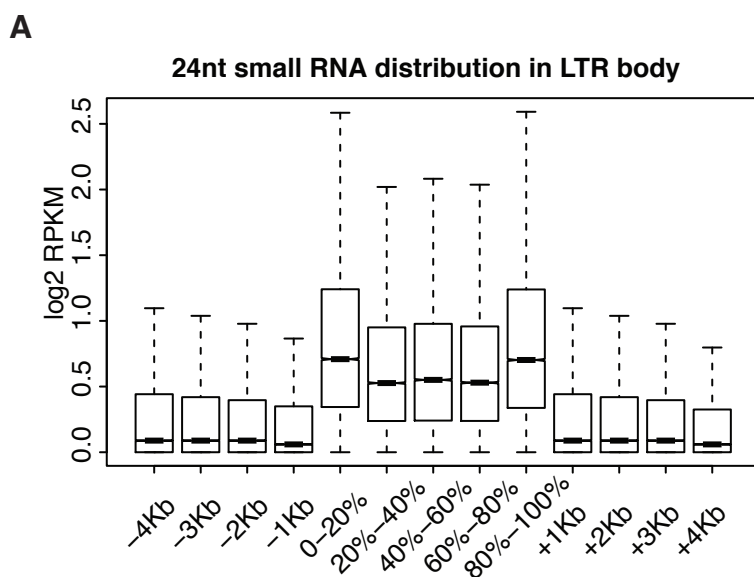
Supplementary Fig. 3. Chromosome-wide distribution of 21-24 nt smRNA levels for chromosome 17.

A

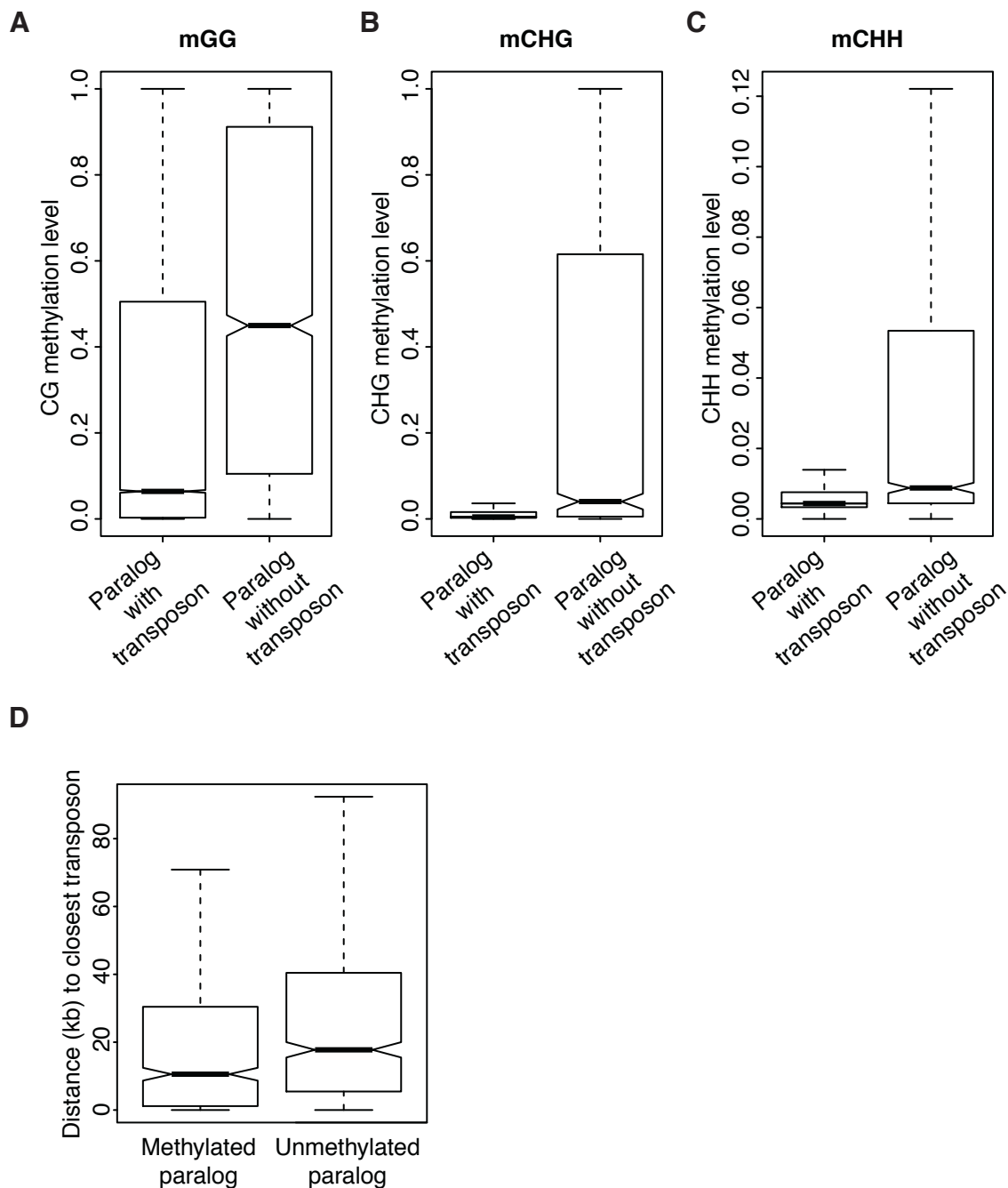
Supplementary Fig. 4. (A) The distribution of distance of early and recently duplicated regions from the centromeres. For the presented box plots, non-overlapping confidence intervals represents that two medians are significantly different with a p-value < 0.05.



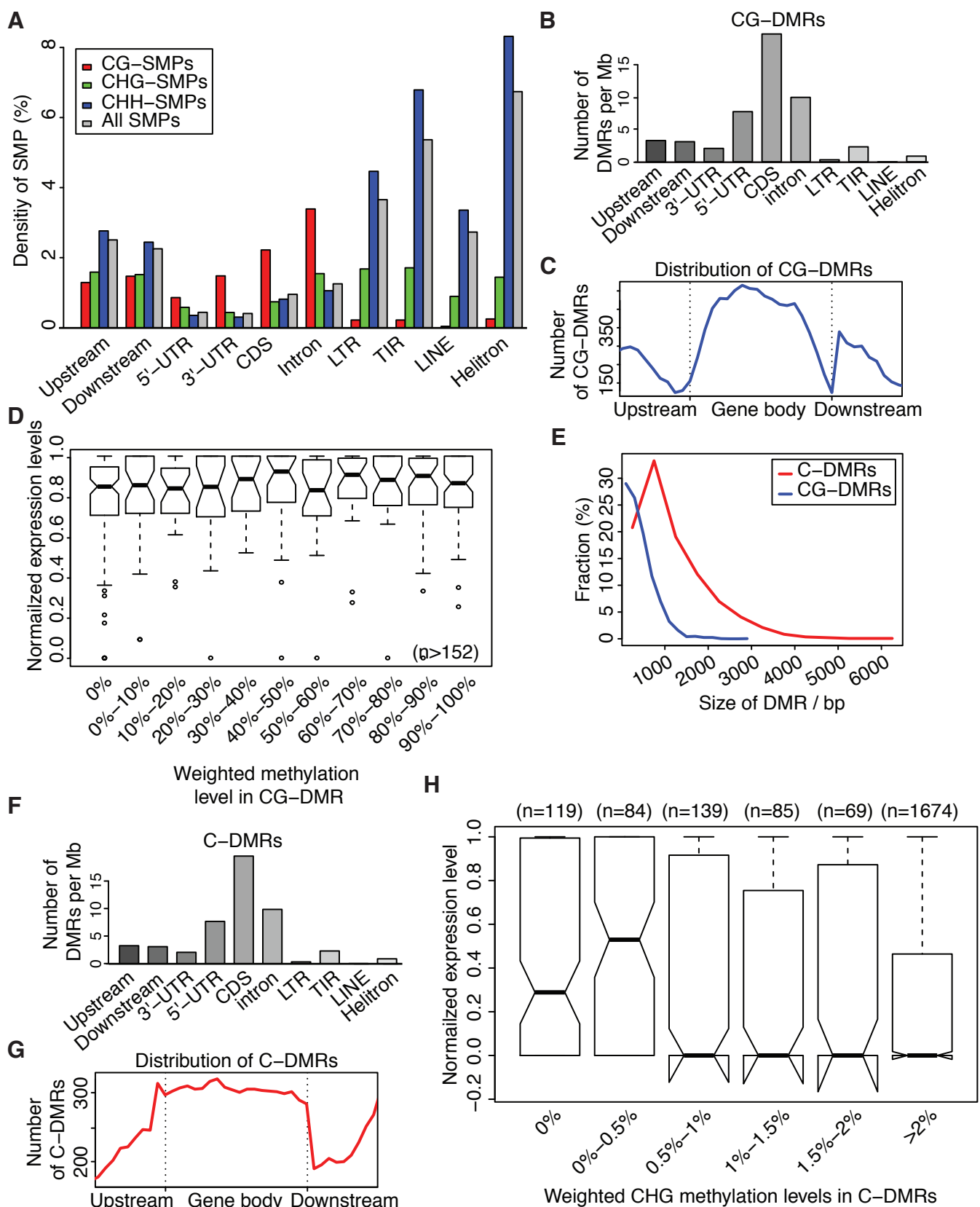
Supplementary Fig. 5. (A-C) The distribution of methylation levels in genes versus transposons for (A) CG, (B) CHG and (C) CHH methylation. (D-F) The distribution of DNA methylation levels from 4Kb upstream of the start codon to 4Kb downstream of the stop codon and in 20 percent intervals within gene bodies for (D) CG, (E) CHG and (F) CHH methylation. (G) The distribution of 24nt smRNA levels between genes and transposons. For all presented box plots, non-overlapping confidence intervals represents that two medians are significantly different with a p-value < 0.05.



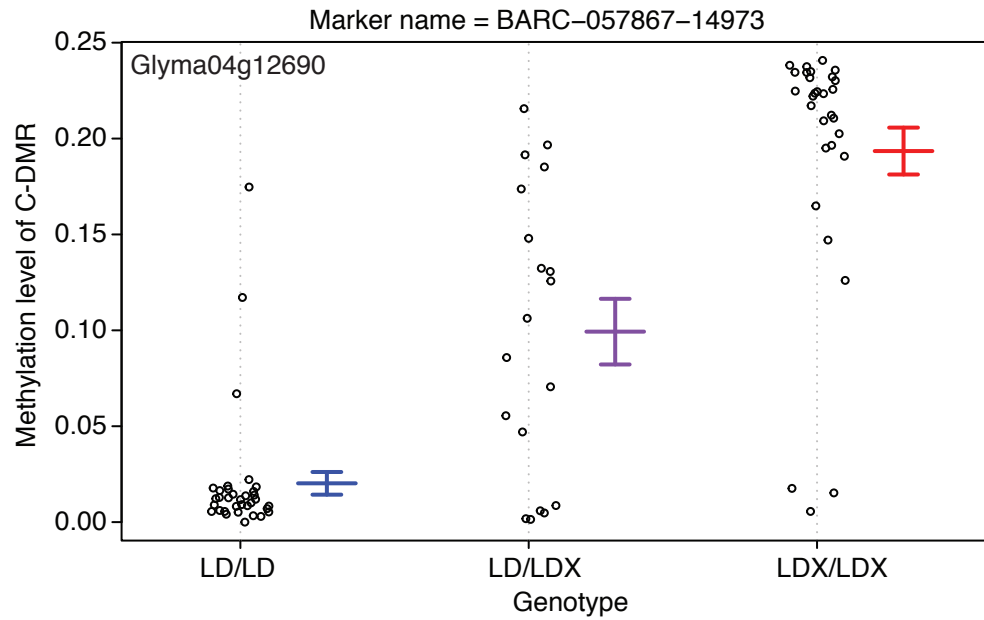
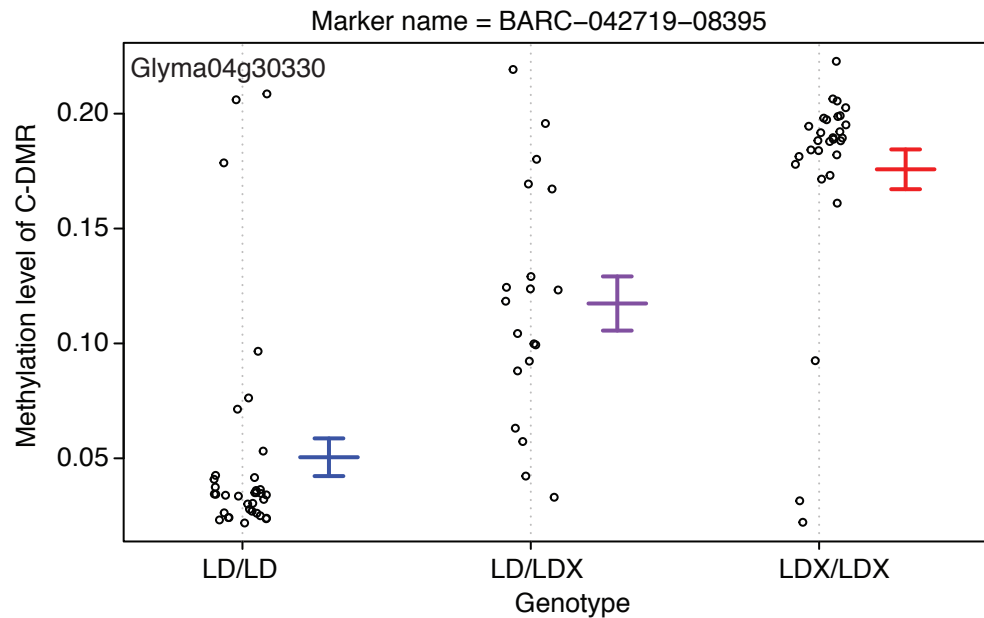
Supplementary Fig. 6. (A-C) The distribution of 24nt smRNA levels from 4Kb upstream of the start codon to 4Kb downstream of the stop codon and in 20 percent intervals within gene bodies for (A) LTR, (B) LINE and (C) TIR transposons. For all presented box plots, non-overlapping confidence intervals represents that two medians are significantly different with a p-value < 0.05.



Supplementary Fig. 7. (A-C) The distribution of DNA methylation levels between differentially methylated paralogs that contain transposons within the gene body compared to paralogs that do not contain transposons within their gene body. **(D)** For differentially methylated paralogs, the distribution of distances away from the nearest transposon is closer for methylated paralogs versus unmethylated paralogs. For all presented box plots, non-overlapping confidence intervals represents that two medians are significantly different with a p -value < 0.05 .



Supplementary Fig. 8. Natural variation of DNA methylation in a soybean recombinant inbred lines population. Location and density of (A) CG-SMPs, CHG-SMPs, and CHH-SMPs, within genomic features per CG, CHG or CHH sites, respectively. (B) Number and type of genomic feature overlapping CG-DMRs. (C) The distribution of CG-DMRs within gene bodies including (± 4 kb) up- and down-stream. (D) The methylation levels of CG-DMRs within gene bodies are not associated with global changes in gene expression. Box plot representation of normalized gene expression levels (y-axis) across increasing levels of methylation (x-axis). (E) The distribution of sizes of CG- and C-DMRs. (F) Location and enrichment of C-DMRs within genomic features. (G) The distribution of C-DMR locations within and around gene bodies. The x-axis indicates the location of the midpoint of the C-DMR and the y-axis indicates the frequency. (H) Of the C-DMRs that overlapped genes, the genes containing the methylated form of the C-DMR are expressed at lower levels compared to the unmethylated form of the C-DMR.

A**B**

Supplementary Fig. 9. (A and B) The range of methylation levels for C-DMRs overlapping (A) Glyma04g12690 and (B) Glyma04g30330 and their respective genotypes. The y-axis indicates the methylation level and the genotypes determined from the SNP array are presented along the x-axis. Clear examples of outliers exist for each homozygous genotype, which could indicate potential epimutations or could also reflect incorrect genotypes as assessed from the SNP array. The error bars indicate the standard error around the mean.