Extensive Demethylation of Repetitive Elements During Seed Development Underlies Gene Imprinting

Mary Gehring, **Kerry L. Bubb**, and **Steven Henikoff**[*]
Howard Hughes Medical Institute, Fred Hutchinson Cancer Research Center, 1100 Fairview Ave N, Seattle, WA 98109

Abstract

DNA methylation is an epigenetic mark associated with transposable element silencing and gene imprinting in flowering plants and mammals. In plants, imprinting occurs in the endosperm, which nourishes the embryo during seed development. We have profiled Arabidopsis DNA methylation genome-wide in the embryo and endosperm and find that large-scale methylation changes accompany endosperm development and endosperm-specific gene expression. Transposable element fragments are extensively demethylated in the endosperm. We discovered new imprinted genes by identifying candidates associated with regions of reduced endosperm methylation and preferential expression in endosperm relative to other parts of the plant. These data suggest that imprinting in plants evolved from targeted methylation of transposable element insertions near genic regulatory elements followed by positive selection when the resulting expression change was advantageous.

Cytosine DNA methylation is a stable epigenetic modification that has roles in transposable element silencing and gene imprinting in plant and animals. In plants, gene imprinting occurs in the endosperm during seed development (1). At fertilization, one sperm fertilizes the haploid egg cell, which becomes the diploid embryo, and the other sperm fertilizes the diploid central cell, generating the triploid endosperm. In Arabidopsis, the 5-methylcytosine DNA glycosylase *DEMETER* (*DME*) demethylates maternal alleles of imprinted genes in the central cell before fertilization, thus establishing methylation asymmetry between embryo and endosperm. Similarly, in maize an imprinted gene is less methylated in the central cell than in the egg cell or sperm (2). The asymmetry between embryo and endosperm represents a unique opportunity to characterize DNA methylation in parallel genomes established simultaneously at fertilization within the same seed.

To compare tissue-specific methylation patterns within developing seeds, we dissected embryo and endosperm from torpedo-stage seeds of two *Arabidopsis thaliana* accessions, Col-*gl* and L*er* (Fig. S1A). Methylated DNA was immunoprecipitated with a 5-methylcytosine antibody, sequenced using Illumina GA technology, and aligned to the reference Col-0 genome. Methylation profiles were created using high-quality reads that mapped to only one position in the genome (Table S1). Embryo and endosperm methylation profiles were highly correlated (Pearson's $R = 0.91$ for Col-*gl* and 0.89 for L*er*) and share similar features with other whole-genome methylation profiles that have been generated for Arabidopsis using other platforms (Fig. 1 and Fig. S1B) (3,4). Methylation levels are relatively high in gene-poor heterochromatic regions around centromeres, and decrease in gene-rich chromosome arms (Fig. 1A). Transposable element genes (a set of 3900 elements with open reading frames) are more heavily methylated than protein-coding genes, and genes are more methylated within their bodies than at their 5′ and 3′ ends (Fig. 1B). However, transposable element genes and regions flanking

[*]author for correspondence: steveh@fhcrc.org.

genes are on average less methylated in the endosperm than embryo (Fig. 1B). If the average methylation profile 5′ and 3′ of protein-coding genes is calculated by excluding methylation from transposable elements or their fragments, the methylation difference between embryo and endosperm in regions flanking genes almost entirely disappears, indicating that repetitive elements are hypomethylated in the endosperm (Fig. 1C). These results suggest there is a genome-wide decrease in methylation in the endosperm compared to the embryo. Reduction in methylated DNA in the endosperm is also observed when immunoprecipitated methylated DNA is hybridized to genomic tiling arrays (Fig. S2). This small genome-wide reduction is consistent with the report that maize endosperm has 13% less 5-methylcytosine than embryos or leaves (5).

To identify regions of the genome subject to the largest changes in DNA methylation [Differentially Methylated Regions (DMRs)], we calculated an embryo-endosperm difference score in overlapping 300-bp segments and set the cutoff to include the top 0.5% of differences (top DMRs) (Fig. 2A). This cutoff detects the previously described methylation differences 5′ and 3′ of the imprinted *MEA* gene (6) whereas those 5′ of *FWA* (7) fall just below it (+1.18 vs. a cutoff of 1.20) (Fig. 2B). This cutoff also detects previously hypothesized methylation differences (8,9) at the *PHE1* and *FIS2* imprinted genes (Fig. 2B). About 90% of the top DMRs have a positive score, indicating greater methylation in the embryo than endosperm (Fig. S3). In plants, DNA methylation is actively targeted by small RNAs, which often arise from and target repetitive elements (10). The top DMRs were almost 3-fold enriched in regions of the genome corresponding to transposable elements (11) and small RNAs (12) (Table S2), indicating that demethylation occurs at regions that are actively targeted for DNA methylation in other tissues. The distribution of the DMRs along the chromosomes parallels the distribution of transposable elements (Fig. S4).

Bisulfite sequencing around fifteen different regions that fell above and below the top 0.5% cutoff largely validated the predictions from the deep-sequencing analysis (Fig. S5). For DMRs with a positive score, methylation of individual bisulfite clones from the endosperm was more variable than from the embryo. Often two distinct subpopulations of clones were observed in the endosperm, including clones with no methylation, which were never observed in the embryo. Therefore, unmethylated clones in the endosperm might represent specific demethylation of the maternal genome by DME in the central cell before fertilization. In support of this possibility, clones with no methylation were nearly eliminated in *dme* mutant endosperm.

More than half of the top positive DMRs (endosperm less methylated than embryo) occur within 2-kb upstream or 2-kb downstream of genes (Fig. S3) (6,13). We identified all protein-coding genes where a top positive DMR fell within the body of the gene or 1 kb 5′ or 3′. This yielded 1276 genes for the Col-*gl* embryo-endosperm comparison (Table S3) and 1163 for L*er* embryo-endosperm (Table S4). Embryo methylation profiles for these genes display prominent peaks centered at ~700 bp both upstream and downstream that are largely absent in the endosperm (Fig. 3 and Fig. S6). This change represents loss of methylation in the endosperm and not gain of methylation in the embryo because these genes are similarly methylated in embryos and adult plants (Figs. S1C and S6C). Upstream and downstream methylation peaks are partially restored in *dme* mutant endosperm (Fig. 3), despite the fact that *dme* endosperm is overall hypomethylated compared to wild type (see Supporting Online Material [SOM]). Thus much of the methylation 5′ and 3′ of genes that is depleted in the endosperm is likely lost due to active demethylation by DME in the central cell before fertilization, although other mechanisms probably also contribute (14).

DNA methylation of promoters inhibits transcriptional initiation. The most prominent losses of gene-associated methylation in endosperm occur well upstream of the transcriptional start

site (Fig. 3) and, for most genes, the presence of nearby DNA methylation apparently has little effect on gene expression (Fig. S7). However, we found that genes with endosperm-preferred expression (15) are less methylated at 5′ sequences in the endosperm than embryo (Fig. S8), suggesting that 5′ loss of methylation is associated with increased expression of a subset of genes in the endosperm.

Known imprinted genes are less methylated in the endosperm than embryo (Fig. 2B) and exhibit endosperm-preferred expression (15). To identify novel imprinted genes we chose genes with top DMRs in comparisons between embryo and endosperm and between wild type and *dme* endosperm for Col-*gl* and L*er* datasets. A set of 113 genes have top DMRs in 3 of the 4 comparisons, including the known imprinted gene *MEA* (Table S5). Two of these genes, *HDG3* and *HDG9*, belong to the 16-member class IV homeodomain leucine zipper transcription (HD-ZIP) factor gene family that also includes the known imprinted gene *FWA*. Like *FWA*, three HD-ZIP genes are expressed primarily or exclusively in siliques, *HDG3*, *HDG8*, and *HDG9* (16).

To test parent-of-origin specific expression of putative imprinted genes, we performed reciprocal crosses between L*er* and Col-*gl* and assayed expression patterns by RT-PCR. Only the maternal *HDG9* allele was expressed in the endosperm (Fig. 4A). We confirmed that methylation was lost around the 5′ end of the gene specifically on maternal alleles in the endosperm, a region annotated as overlapping the remnant of a Helitron transposable element (Fig. 4C). *HDG8* is also primarily, but not exclusively, expressed from the maternal allele (Fig. 4A and SOM). *HDG3* is reciprocally imprinted --expression is predominantly paternal (Fig. 4B and SOM). Methylation is lost from maternal alleles on a 1.4 kb Helitron remnant that begins 100 bp 5′ of the gene (Fig. 4D). We confirmed allele-specific expression of two other genes in the endosperm, the ATMYB3R2 transcription factor, which is maternally expressed, and AT5G62110, a gene annotated as containing a homeodomain-like domain, which is predominantly paternally expressed (Fig. 4 and SOM).

We tested several other genes that were less methylated in the endosperm than embryo for imprinting (Table S6). CYCA1;1 is a differentially-methylated A-type cyclin that exhibits endosperm-preferred expression but is also expressed at many other stages of development (Fig. S8C) and is biallelically expressed in the endosperm. All of the differentially methylated genes with endosperm-preferred expression that were *not* imprinted were expressed in other tissues throughout development, whereas the five genes with top positive DMRs that we confirmed as imprinted were expressed at very low levels or not at all in other tissues (Fig. S9). This suggests that genes with low levels of expression in other plant tissues require loss of methylation for gene expression in the endosperm. Thus the best candidates for imprinted genes are those that are less methylated in the endosperm than embryo, exhibit endosperm-preferred expression (15), and are expressed at low levels in other parts of the plant (Table S7). Based on these considerations, we estimate that there are ~50 imprinted genes in Arabidopsis. Gene ontology analysis of molecular function indicates that these genes are enriched in the term "DNA or RNA binding", which includes transcription factors and genes with chromatin-related functions (Fig. S9).

Our identification and verification of five new imprinted genes, all of which are flanked by or harbor repetitive elements in coding sequences, doubles the number of known imprinted genes in Arabidopsis. Four of the ten imprinted genes are class IV homeodomain transcription factors, which offers the opportunity to study the evolution of imprinting (SOM). Our results support the theory that imprinting arose as a by-product of silencing invading foreign DNA (17). Transposable elements (TEs) have demonstrated toxic effects on genomes, but they are also a source of genetic and epigenetic material that can be utilized by the host (18). Insertion of a TE near a gene will have little functional impact in most instances, and strongly deleterious

TE insertions will be selected against. However, a subset of genes, perhaps depending on promoter strength, are susceptible to epigenetic regulation by TEs. Regulation of gene expression by DNA methylation could be selected for if imprinting of these genes is adaptive in the context of parental conflict or gene dosage balance in the triploid endosperm.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Huh JH, Bauer MJ, Hsieh TF, Fischer RL. Cell 2008;132:735. [PubMed: 18329361]

2. Gutierrez-Marcos JF, et al. Nat Genet 2006;38:876. [PubMed: 16823380]

3. Zilberman D, Gehring M, Tran RK, Ballinger T, Henikoff S. Nat Genet 2007;39:61. [PubMed: 17128275]

4. Zhang X, et al. Cell 2006;126:1189. [PubMed: 16949657]

5. Lauria M, et al. Plant Cell 2004;16:510. [PubMed: 14729913]

6. Gehring M, et al. Cell 2006;124:495. [PubMed: 16469697]

7. Kinoshita T, et al. Science 2004;303:521. [PubMed: 14631047]

8. Makarevich G, Villar CB, Erilova A, Kohler C. J Cell Sci 2008;121:906. [PubMed: 18303047]

9. Jullien PE, Kinoshita T, Ohad N, Berger F. Plant Cell 2006;18:1360. [PubMed: 16648367]

10. Chan SW, Henderson IR, Jacobsen SE. Nat Rev Genet 2005;6:351. [PubMed: 15861207]

11. Buisine N, Quesneville H, Colot V. Genomics 2008;91:467. [PubMed: 18343092]

12. Lister R, et al. Cell 2008;133:523. [PubMed: 18423832]

13. Penterman J, et al. Proc Natl Acad Sci U S A 2007;104:6752. [PubMed: 17409185]

14. Jullien PE, et al. PLoS Biol 2008;6:e194. [PubMed: 18700816]

15. Day RC, Herridge RP, Ambrose BA, Macknight RC. Plant Physiol 2008;148:1964. [PubMed: 18923020]

16. Nakamura M, et al. Plant Physiol 2006;141:1363. [PubMed: 16778018]

17. Barlow DP. Science 1993;260:309. [PubMed: 8469984]

18. Slotkin RK, Martienssen R. Nat Rev Genet 2007;8:272. [PubMed: 17363976]
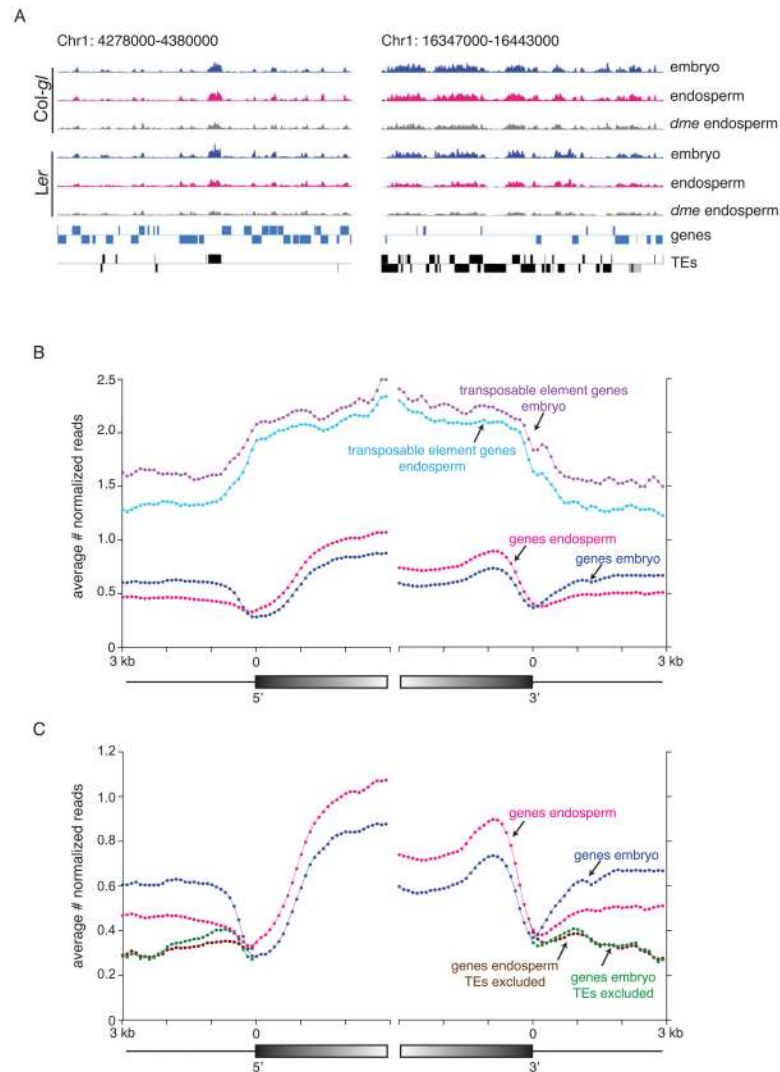
Figure 1. Endosperm is less methylated than embryo

A) Typical methylation profiles from a gene-rich and gene-poor region of the genome showing highly similar methylation patterns in embryo, endosperm, and *dme* endosperm from the L*er* and Col-*gl* accessions. B) Average methylation in 100-bp windows of protein-coding genes and transposable element genes aligned at either their 5′ or 3′ ends in Col-*gl* embryo and endosperm. C) Methylation profiles of protein-coding genes when a set of 31,076 transposable element fragments (11) are and are not excluded from the averaging.
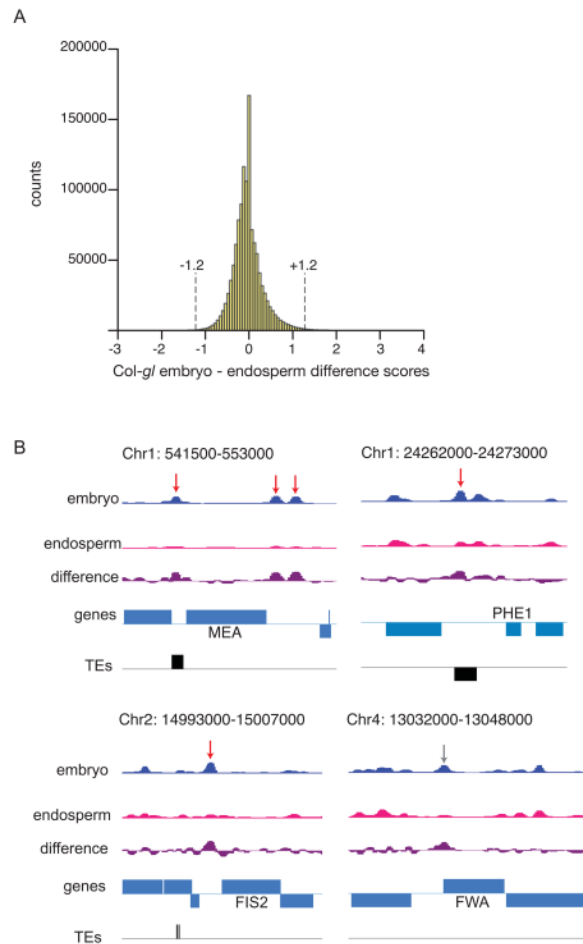
Figure 2. Known imprinted genes are associated with top DMRs
A) Histogram of Col-*gl* embryo - endosperm difference scores for 1.2 million overlapping 300-bp segments. Dashed lines represent the cutoff for the top 0.5% of methylation differences (~6000 300-bp segments). B) Embryo and endosperm methylation profiles of known imprinted genes. Red arrows, regions within the top 0.5% of methylation differences; gray arrow, region below the cutoff.
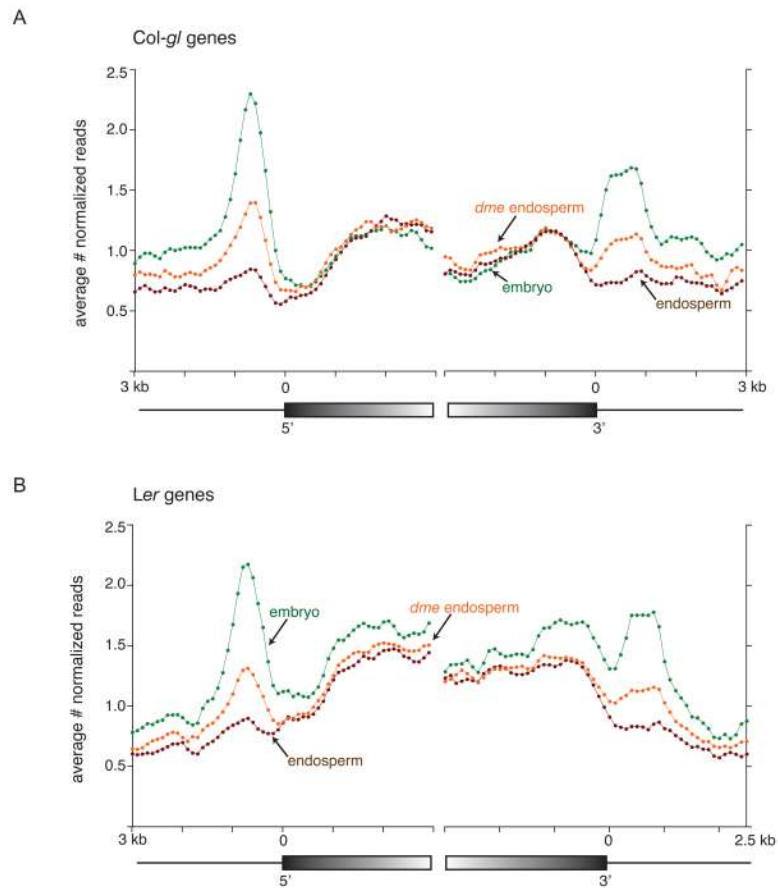
Figure 3. Methylation is lost 5′ and 3′ of genes in the endosperm

Average methylation profiles in embryo, endosperm, and *dme* endosperm of the A) 1276 Col-*gl* and B) 1163 Ler genes associated with more methylation in embryo than endosperm. Genes were aligned at their 5′ or 3′ end and the average methylation determined every 100 bp.
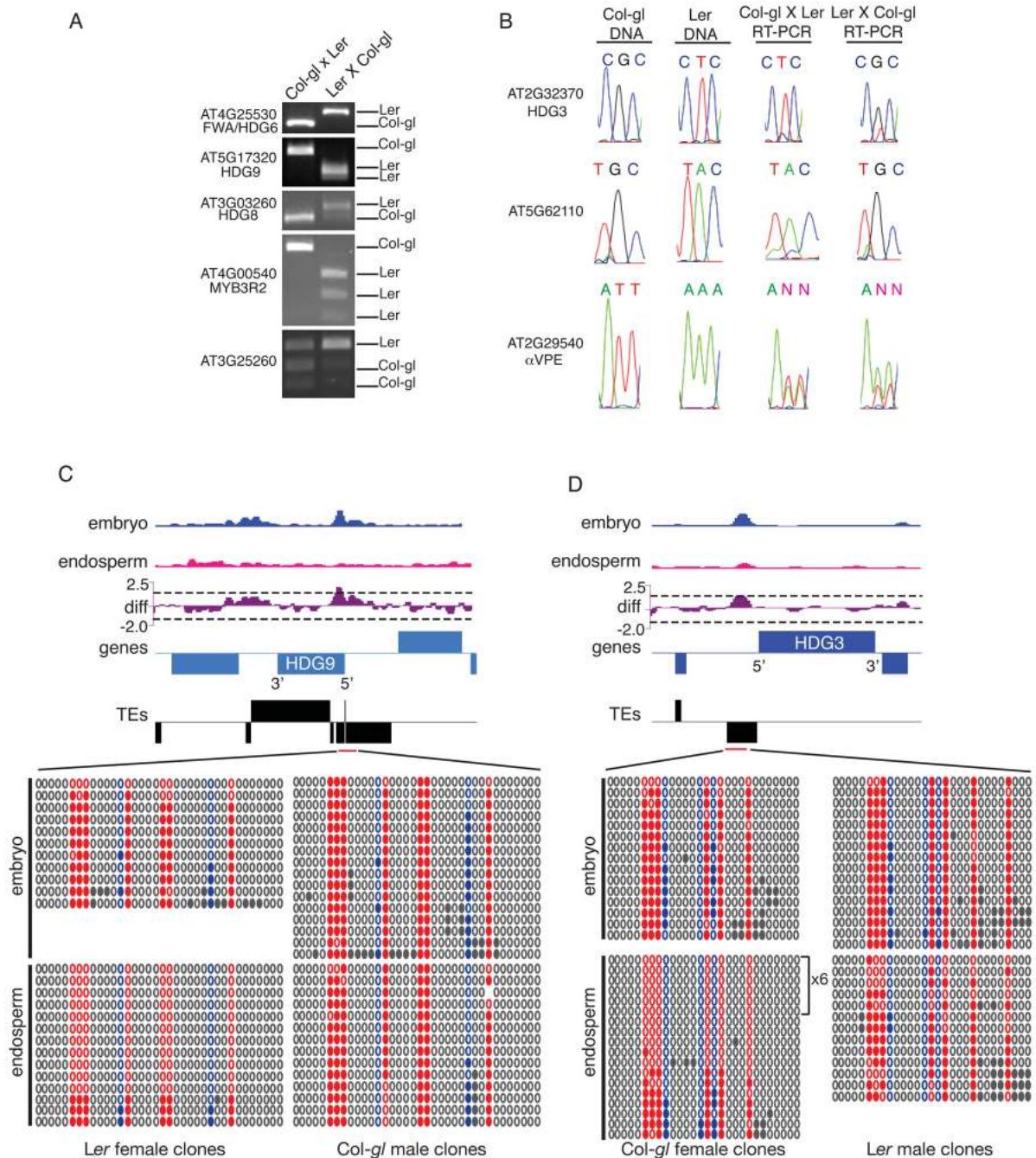
Figure 4. Expression and methylation analysis of new imprinted genes

A) RT-PCR allele-specific expression analysis from endosperm RNA of Col-*gl* females X
L*er* males and L*er* females X Col-*gl* males. *FWA* is a control imprinted gene; AT3G25260 is
biallelic. B) RT-PCR sequencing chromatograms for the paternally expressed genes
AT2G32370 and AT5G62110. The biallelically-expressed gene αVPE is shown as a control.
C) and D) Methylation profiling and allele-specific bisulfite sequencing analysis of new
imprinted genes. Dashed lines represent the cutoff for the top 0.5% of differences. Each line
of circles represent a bisulfite sequencing clone from embryo or endosperm. Filled circle,
methylated cytosine. Red, CG; blue, CHG; gray, CHH.