Supplementary Information

Supplementary Methods

mRNA-seq analysis

Preliminary filtering and mapping

Paired-end mRNA-seq reads were trimmed to remove adapters and poor quality ends using trim-galore version 0.4.0, with --stringency 3 and quality cutoff -q 25. Quality of reads both before and after trimming was determined using fastqc. Fastqc indicated that the first 8 bases of each read were of poorer quality, and these were also trimmed off using --clip_R1 8 and --clip_R2 8. Only read pairs in which both reads were at least 32 bp long after trimming were kept for mapping. Reads were aligned to the MN47 reference genome [1] using Tophat v2.0.13 [2] with minimum intron length -i 20 and maximum intron length -I 2000. The parameters for maximum mismatches, read edit distance, segment length, and mate inner distance varied based on library read length (see **Supplementary Table 1)**. A GTF file containing known gene annotations was provided using the -G option to improve alignment speed. The resulting SAM file was filtered to remove reads that mapped ambiguously to multiple locations in the genome by removing all alignments with MAPQ < 5, and PCR duplicates were removed using MarkDuplicates from Picard tools. All other alignments were kept for downstream analyses.

Updating gene annotations

After processing the libraries as described above, all reads from the 8 MN x MN libraries were pooled into a single dataset. Cufflinks v2.2.1 [3] was used to obtain a list of predicted transcripts from these pooled alignments, using the existing annotations as a guide (-g option), with maximum intron length -I 50000, and all other parameters left at their default values. In parallel, Trinity v2.0.6 was used in genome-guided mode to assemble transcripts based on these same alignments, with --genome_guided_max_intron 50000 [4]. The FASTA file of assembled transcripts from Trinity and the GTF file of predicted transcripts from Cufflinks were fed into the

PASA pipeline [5], which used this information to predict novel transcripts and update the existing annotations. Three rounds of updating with PASA were used. We also obtained annotations from a recent *A. lyrata* annotation update, which used RNA-seq data from aerial vegetative tissues [6]. We predicted which of the three possible gene models (the original annotation, the PASA updates, and new annotations from Rawat *et al*. [6]) was best supported by our data at each locus. Cuffcompare was used to map the three sets of annotations to each other for each locus. The best supported model for each locus was defined according to the following: (1) if only one of the sets of annotations had an annotated model at a locus, it was chosen automatically, (2) if expression at the locus was very low, defined as (counts^2)/len(CDS) < 10 for all models, the original annotation was chosen, (3) if one model had more than 1.5x higher (counts^2)/len(CDS) than either other model, that model was chosen, and finally if none of the other criteria were satisfied, the model with the longest coding sequence was chosen. This was repeated for all loci to obtain the final set of annotations, representing 36,732 putative protein-coding genes used for all subsequent analysis. Of the 36,732 genes, 19,648 were unchanged from the original version, 8,842 genes had altered UTRs but unchanged coding sequences, while 6,323 genes had altered coding sequences after the update. Finally, 1,919 genes were not found in the original annotations, most of which (1,861) were instead obtained from the annotations by Rawat et al [6]. The remainders were novel genes identified by PASA and were not present in either the original annotations or Rawat *et al*. [6] and likely represent genes specifically expressed in the embryo or endosperm.

Updating gene homology information

Since the updated annotations included a number of novel genes and altered the coding sequences of some existing genes, we also updated a list of putative *A. thaliana* homologues of *A. lyrata* genes obtained from Phytozome [7]. To obtain preliminary new homology information, we performed a reciprocal tblastx of all *A. lyrata* genes (using the updated annotations from

above) to the *A. thaliana* genome, and all *A. thaliana* genes to the *A. lyrata* genome. Only alignments with E-values below 0.0001 were reported. We first identified all pairs of *A. lyrata* and *A. thaliana* genes that were reciprocal hits to each other (defined as both alignments passing E-value cutoff, and additionally both alignments covering > 50% of the query gene). Then, for each *A. lyrata* gene ALY, we defined the likely homologue ATH: (1) if ALY and ATH are each other's reciprocal best hits; else if (2) ALY and ATH are each other's only reciprocal hit (even if not both highest scoring); else if (3) there are multiple pairs of reciprocal hits that include ALY, but ALY and ATH are the overall highest scoring according to E-value; else if (4) there are no pairs of reciprocal hits that include ALY, but either ALY has a high scoring alignment to ATH covering > 75% of ALY, or ATH has a high scoring alignment to ALY covering > 75% of ATH (this often occurs if one gene is a fragment of the other). We also kept the Phytozome homologue if the coding sequence of a gene remained unchanged in the new annotations, unless the new homology analysis revealed a different reciprocal best hit. In the final set of *A. thaliana* homologues for the 36,732 *A. lyrata* genes, 32,891 were unchanged from Phytozome (this includes cases where both versions had no identified *A. thaliana* homologue), 996 had a different homologue than before, 530 lost a homologue and 2315 gained a homologue.

Identifying SNPs

To identify SNPs between MN47 and Karhumäki (**Supplementary Table 7**) for allele-specific expression analysis, RNA-seq data from all datasets corresponding to either MN47 or Kar (**Supplementary Table 1**) were pooled into a single dataset containing all MN47-derived reads (the reference strain) and another containing all Kar-derived reads (the alternate strain). PCR duplicates were removed using the Picard tools MarkDuplicates function before pooling. Preliminary SNP info was obtained using SAMtools mpileup with default parameters. SNP calls were refined using vcf-annotate to exclude all sites with fewer than 20 overlapping reads, and

with fewer than 10 reads with the alternate allele. SNPs where the MN47 allele was not consistent with the published genome were also removed. To filter out likely heterozygous SNPs, only SNPs with a PLDiff > 20 were kept, where PLDiff is the difference between the Phred-scaled genotype likelihoods (PL) of a homozygous call and a heterozygous call. The remaining 182,256 SNPs were used for all RNA-seq data analyses.

To identify additional non-genic SNPs for analysis of bisulfite-seq data we also sequenced Kar leaf genomic DNA (40 bp single end). Reads were trimmed using trim-galore version 0.4.0, with --stringency 3 and quality cutoff -q 25. Additionally, 8 bp of lower quality at the 5' end were trimmed off using --clip_R1 8. Reads were aligned to the MN47 reference genome using bowtie2 (v. 2.2.5), with –N 0 and –L 22. Reads with mapping quality (MAPQ) greater than 5 were kept, and presumed PCR duplicates were removed using MarkDuplicates from the picard-tools suite. The remaining 44,789,279 reads were used to call SNPs between Kar and MN, requiring at least 10 overlapping reads at the SNP position, at least 10 reads with the alternate allele, and a PLDiff > 20. This resulted in 381,796 SNPs. We also called SNPs after pooling the DNA-seq reads with the RNA-seq reads used previously, using the parameters depth >=20, alt allele >=10, PLDiff > 20, resulting in 190,527 SNPs. The union of these two sets of SNPs, after removing SNPs where the MN47 allele was not consistent with the published genome or which were no longer called after the DNA-seq data was added, resulted in 487,939 SNPs used for all analyses with bisulfite-sequencing data (**Supplementary Table 7**).

Differential expression analysis

Differential expression analysis of *A. lyrata* genes between samples was performed with DESeq2.0 [8]. Genes were considered significantly differentially expressed between the two conditions if abs($\log_2$foldchange) > 1 and the adjusted p-value was < 0.05. The regularized log (rlog) transformation from DESeq2 was used to normalize data before performing PCA.

Identifying imprinted genes

Imprinted genes were identified using our previously described method [9,10]. Briefly, for each library in a pair of reciprocal crosses (MN x Kar compared to Kar x MN), reads overlapping a known SNP were assigned to the parent of origin, and htseq-count [11] was used to count the number of reads from the MN47 allele and the Kar allele found in each gene. Fisher's exact test was used to test the null hypothesis that the proportion of maternal to paternal reads was 1:1 (embryo) or 2:1 (endosperm) in both directions of the cross. Genes were considered imprinted if they had a Benjamini-Hochberg corrected p-value less than 0.01, as well as a minimum imprinting factor of 2 and a maximum cis-effect factor of 10 [9]. In addition, endosperm MEGs were required to have at least 85% maternally-derived reads in each pair of reciprocal crosses, and endosperm PEGs were required to have less than 50% maternally-derived reads in each pair of reciprocal crosses. For embryo, these cutoffs were > 70% maternal (MEGs) and < 30% maternal (PEGs) [10]. This analysis was performed separately for all 12 possible comparisons between the 3 MN x Kar endosperm libraries and the 4 Kar x MN endosperm libraries, and for all 9 possible comparisons between the 3 MN x Kar and 3 Kar x MN embryo libraries. A gene was considered imprinted if at least 5 of the possible comparisons could be evaluated for imprinting (defined as both reciprocal crosses being compared having at least 10 total allele-specific counts), and at least 40% of reciprocal cross comparisons positively identified that gene as imprinted. Under these criteria, 12,633 genes could be evaluated for imprinting. Additionally, we used DEseq2 to identify and filter out 3,449 genes with an estimated $\log_2$ fold change of more than 1.5 in MN x MN seed coat compared to MN x MN endosperm, since these genes could appear as MEGs due to potential seed coat contamination of endosperm samples. Finally, genes with a homologous mitochondrial or chloroplast *A. thaliana* gene were excluded from the final list of imprinted genes.

Validation of allelic bias

Allelic bias was validated by pyrosequencing from MN x MN, Kar x Kar, MN x Kar and Kar x MN embryo and endosperm cDNA. Cloning and Sanger sequencing of genomic DNA confirmed SNPs between MN and Kar. For pyrosequencing, DNase-treated RNA was reverse transcribed using SuperScript III and an oligo(dT) primer (Invitrogen). cDNA was amplified with primers listed in **Supplementary Table 8** using the PyroMark PCR kit (Qiagen). The University of Michigan DNA Sequencing Core performed pyrosequencing with the indicated sequencing primer.

Calculating average % maternal reads

For **Figure 1D** and **Supplementary Figure 4**, the average endosperm % maternal reads in *A. lyrata* was obtained by averaging together the 3 MN x Kar libraries and the 4 Kar x MN libraries separately. These two values (average % maternal in MN x Kar and average % maternal in Kar x MN) were then averaged together to obtain the overall average. This avoids giving undue weight to the cross with more replicates. Since there were an equal number of replicates for the two reciprocal crosses for both *A. thaliana* comparisons (3 replicates each for Col x Cvi and Cvi x Col, same for Col-L*er*), the average was simply taken over all 6 samples. Data was from Pignatta *et al*. 2014 [10]. In **Figure 1d**, average % maternal for *A. thaliana* represents the average over the 3 Col x Cvi and 3 Cvi x Col endosperm replicates.

Differential expression analysis between *A. lyrata* and *A. thaliana*

We used DESeq2 to calculate the $\log_2$ fold change of expression in *A. lyrata* over *A. thaliana* (positive values indicate *A. lyrata* is more highly expressed than *A. thaliana*, negative values indicate the reverse) (see **Figure 1e**). Counts from all 13 *A. lyrata* endosperm samples were used, as were counts from all 6 Col-Cvi and 6 Col-L*er* endosperm libraries from Pignatta *et al.*

(2014)[10]. Estimated $\log_2$ (*A. lyrata/A. thaliana*) values were plotted in **Figure 1e** for all conserved and non-conserved MEGs and PEGs for which there were homologues.

Whole genome bisulfite sequencing data analysis

8 bases were trimmed from the 5' end of the forward and reverse reads, as recommended by the EpiGnome kit protocol. Reads were further trimmed for quality and adapter contamination using trim-galore, with a quality cutoff of -q 25 and --stringency 3. Only read pairs in which both reads were at least 32 bp after trimming were kept for mapping. For samples sequenced from MN47, Bismark was used to align the reads to the "bisulfite treated" *Arabidopsis lyrata* JGI v1.0 (MN47) genome, including the chloroplast and mitochondrial scaffolds [12]. To reduce mapping bias in favor of reads from the MN allele, samples sequenced from Kar x MN47 crosses were aligned to a bisulfite treated "metagenome" containing the full JGI v1.0 genome (MN47) and the Kar "pseudogenome", in which the 487,939 Kar SNPs were introduced into the MN47 sequence. For 40 x 40 bp libraries, 1 mismatch was allowed within the length of the 40 nt seed region, and for 100 x 100 bp libraries, 2 mismatches were allowed within the length the 80 nt seed region. After mapping, PCR duplicates were removed by a script provided with Bismark (deduplicate_bismark_alignment_output.pl), which randomly chooses which single representative of a pool of presumed PCR duplicates to keep in order to avoid biases in methylation calls. Per-site methylation information was extracted using the bismark_methylation_extractor and converted to BED-like format, and cytosines covered by at least 5 reads were kept for further analyses.

Identifying differentially methylated regions

The genome was divided into consecutive 300 bp windows that overlapped by 200 bp. Each window was assessed as a potential DMR between two samples (e.g. embryo and endosperm) using the method described in Pignatta *et al.* [10]. Briefly, the weighted average methylation in

each window was computed for each library [13]. A window was required to contain at least 3 (CpG or CHG) or 10 (CHH) cytosines with at least 5 overlapping reads each in both samples in order to perform a comparison. The null hypothesis of no difference in methylation was tested using Fisher's exact test. All windows with a Benjamini-Hochberg corrected p-value < 0.01, and with a difference in average percent methylation of at least 35 (CpG and CHG) or 10 (CHH), were considered DMRs. For Kar x MN allele-specific DMRs (comparing the maternal allele of embryo to the maternal allele of endosperm, for example), all parameters were the same except we required an average % methylation difference of at least 20 (CpG and CHG) or 10 (CHH). Overlapping DMRs were merged together into single intervals using bedtools merge.

Plots of average methylation profiles across features

For each gene or TE, 50 bp windows were created beginning 2 kb 5' of the gene/TE start site, and ending 2 kb into genes and 1 kb into TEs. The same number of 50 bp windows were also created symmetrically around the gene/TE transcriptional start site (TSS), starting 2 kb 5' of the gene TSS or 1 kb 5' of the TE TSS, and ending 2 kb 3' of the gene/TE TSS. Neither set of windows was permitted to go beyond the midpoint of a particular gene/TE. Average methylation levels[13] were calculated for each window in each gene separately using the per-site methylation information, and then these values were averaged for each 50 bp window across all genes or TEs. For plots over imprinted genes, if a 50 bp window only had data for a single gene, that point was omitted from the plot.

Locus-specific bisulfite PCR

Loci of interest were identified based on DMR data. MN x MN and Kar x MN seeds at the walking stick/bent cotyledon stage were dissected into embryo and endosperm, DNA was extracted using a CTAB method, RNase-treated, and genomic DNA from multiple dissections was pooled. 214-500 ng of gDNA was bisulfite treated with the MethylCode Bisulfite Conversion

kit (Invitrogen), according to manufacturer's instructions. PCR products were amplified using ExTaq for 40 cycles, with annealing temperatures varying from 50 – 55°C, and were cloned into either pJET1.2/blunt (Thermo Fisher) or pCR2.1-TOPO TA (Thermo Fisher) vectors. Primer sequences are in **Supplementary Table 8**. Individual bisulfite clones were Sanger sequenced and DNA methylation analysis was performed using CyMATE [14].

Testing the relationship between endosperm CHG methylation and gene expression

Of the 36,732 genes in our *A. lyrata* annotations, 23,809 had sufficient coverage in both MN x MN embryo and endosperm to evaluate differential expression using DESeq2 (all replicates of MN x MN embryo and endosperm RNA-seq data were used). 4057 genes were significantly less expressed in endosperm than embryo (defined as having a DESeq2 adjusted p-value < 0.05 and $\log_2$ fold change (endosperm/embryo) < -1). Of the 1606 genes that gained endosperm CHG methylation, 1225 had sufficient coverage to evaluate differential expression; 338 were significantly less expressed in endosperm than embryo. We evaluated whether an overlap of 338 genes between the 4057 genes less expressed in endosperm than embryo and the 1225 endosperm CHG methylated genes was significantly greater than expected by chance using a hypergeometric test in R:

```
phyper(337,4057,23809-4057,1225,lower.tail = FALSE, log.p = FALSE)
```

This is the probability of observing 338 or more genes shared between these two groups. We obtained $P(x \geq 338) = 1.766 \times 10^{-21}$, suggesting that genes that gain CHG methylation in endosperm are significantly more likely to be less expressed in endosperm compared to embryo than would be expected by chance. Similarly, we identified 4671 genes that were significantly more highly expressed in endosperm than in embryo (adjusted p-value < 0.05 and $\log_2$ fold change (endosperm/embryo) > 1), 159 of which also gained CHG methylation:

```
phyper(158,4671,23809-4671,1225,lower.tail = FALSE, log.p = FALSE)
```

We obtained $P(x \geq 159) \approx 1$, or equivalently $P(x \leq 159) = 2.04 \times 10^{-10}$ (phyper(159,4671,23809-4671,1225,lower.tail = TRUE, log.p = FALSE)), suggesting that there are significantly fewer genes more expressed in endosperm compared to embryo among the endosperm CHG hypermethylated genes than would be expected by chance.

RT-qPCR

RNA was isolated from dissected embryo and endosperm using the RNAqueous Micro Kit (Ambion) and treated with DNase I (Invitrogen). RNA was reverse transcribed with SuperScript III using an oligo(dT) primer (Invitrogen). qPCR was performed with Fast SYBR Green Master Mix (Applied Biosystems) using primers listed in **Supplementary Table 8**. Relative enrichment was calculated using the $\Delta\Delta$Ct method [15].

Supplementary References

1. Hu, T. T. *et al.* The Arabidopsis lyrata genome sequence and the basis of rapid genome size change. *Nat. Genet.* **43,** 476–481 (2011).

2. Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14,** R36 (2013).

3. Trapnell, C. *et al.* Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature Protocols* **7,** 562–578 (2012).

4. Haas, B. J. *et al.* De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols* **8,** 1494–1512 (2013).

5. Haas, B. J. *et al.* Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Research* **31,** 5654–5666 (2003).

6. Rawat, V. *et al.* Improving the Annotation of Arabidopsis lyrata Using RNA-Seq Data. *PLoS ONE* **10,** e0137391 (2015).

7. Goodstein, D. M. *et al.* Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Research* **40,** D1178–86 (2012).

8. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11,** R106 (2010).

9. Gehring, M., Missirian, V. & Henikoff, S. Genomic Analysis of Parent-of-Origin Allelic Expression in Arabidopsis thaliana Seeds. *PLoS ONE* **6,** e23687 (2011).

10. Pignatta, D. *et al.* Natural epigenetic polymorphisms lead to intraspecific variation in Arabidopsis gene imprinting. *Elife* e03198 (2014).

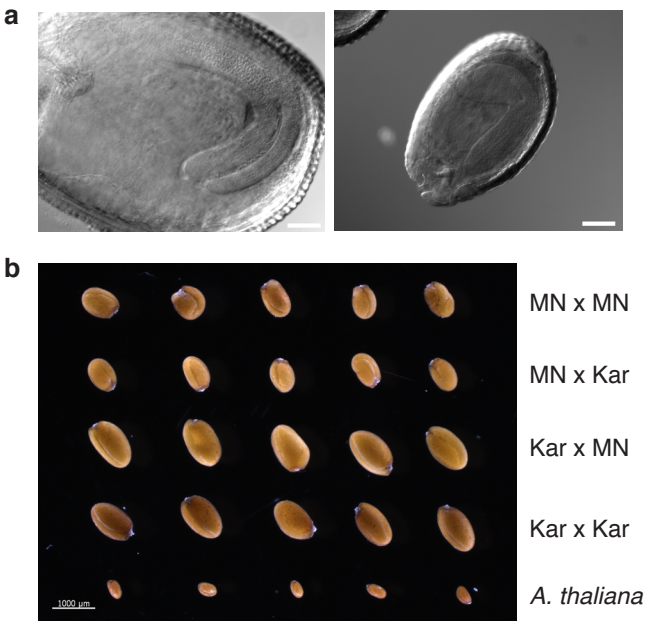
11. Anders, S., Pyl, P. T. & Huber, W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31,** 166–169 (2015).

12. Krueger, F. & Andrews, S. R. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27,** 1571–1572 (2011).

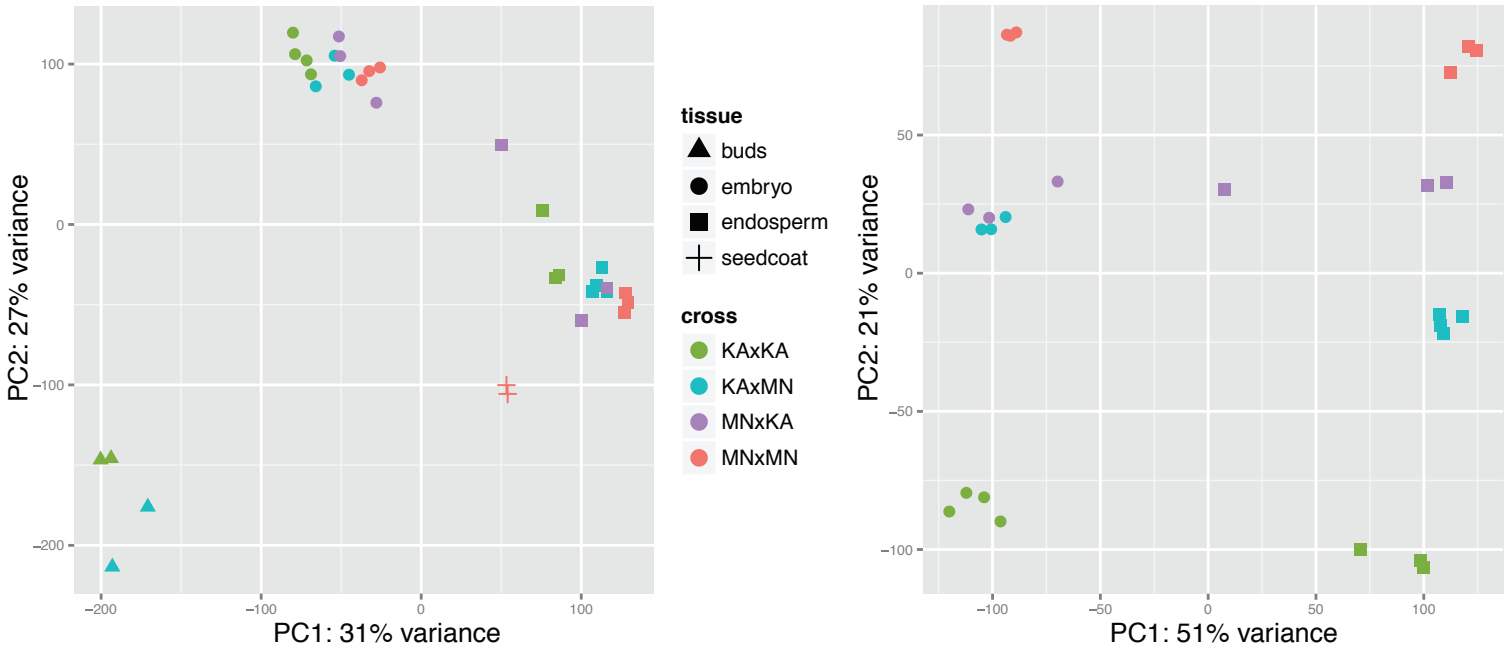
13. Schultz, M. D., Schmitz, R. J. & Ecker, J. R. 'Leveling' the playing field for analyses of single-base resolution DNA methylomes. *Trends Genet.* **28,** 583–585 (2012).

14. Hetzl, J., Foerster, A. M., Raidl, G. & Mittelsten Scheid, O. CyMATE: a new tool for methylation analysis of plant genomic DNA after bisulphite sequencing. *Plant J.* **51,** 526–536 (2007).

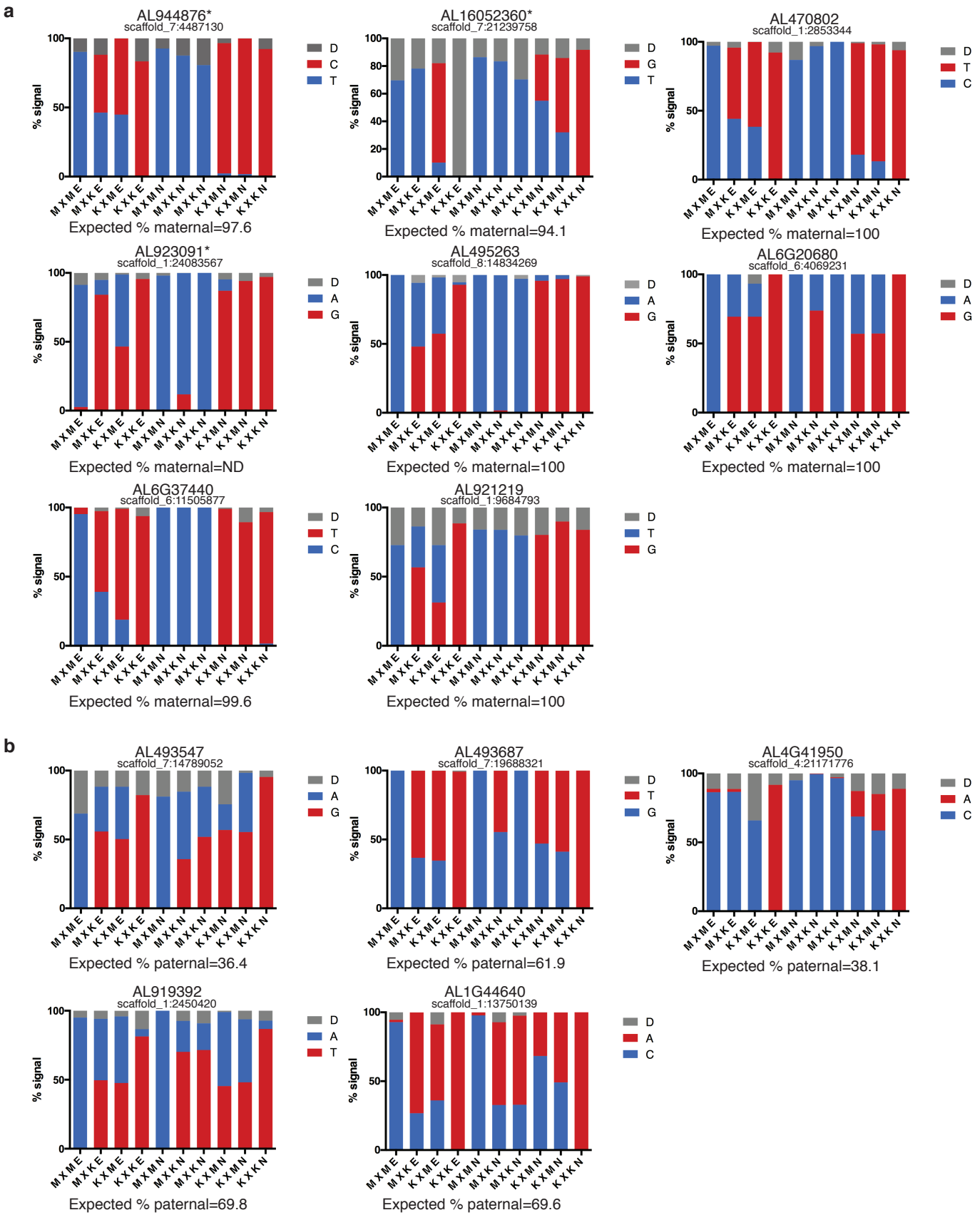
15. Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25,** 402–408 (2001).



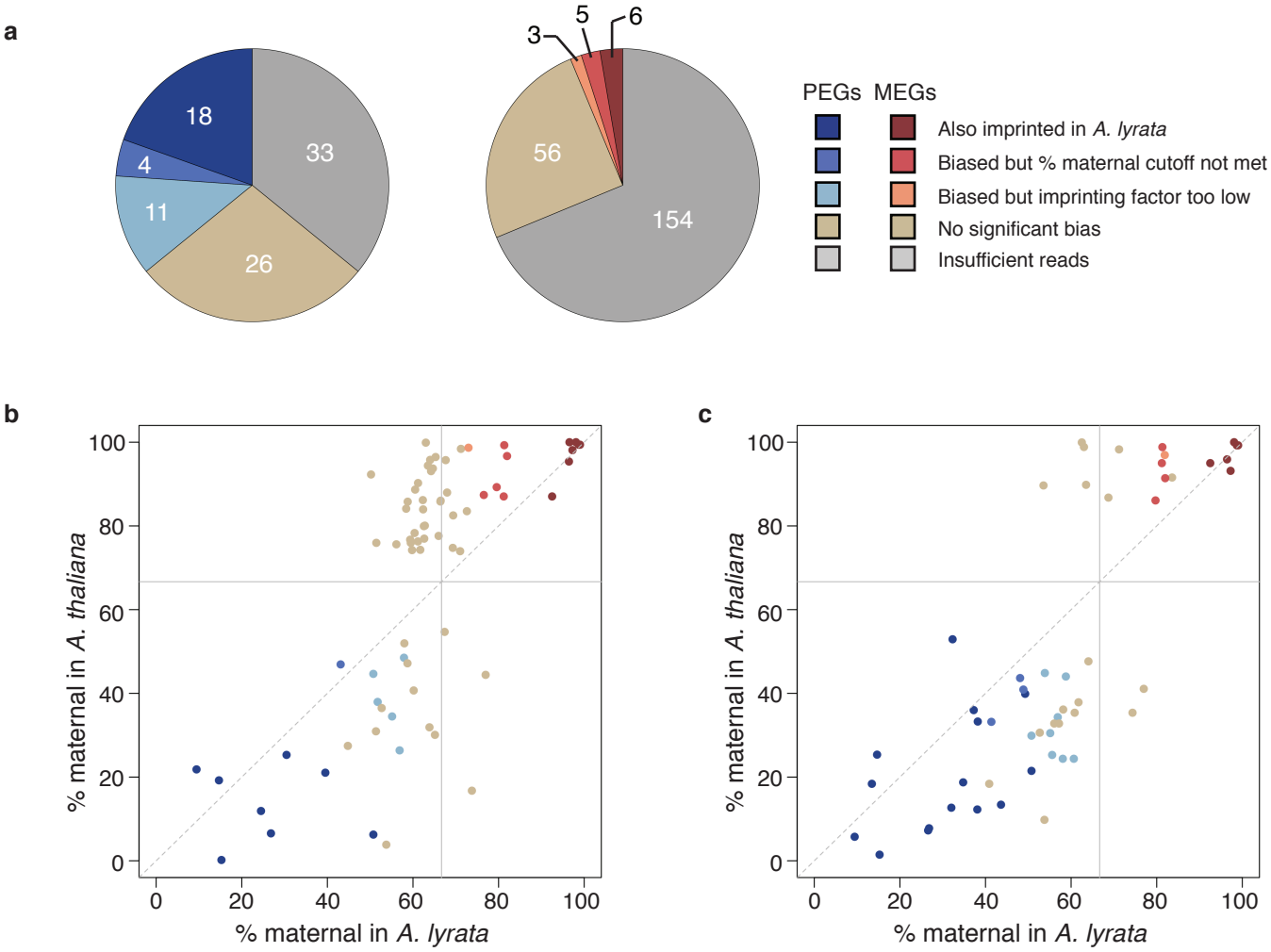
Supplementary Figure 1: *A. lyrata* seed structure. **a)** *A. lyrata* MN47 seed at the bent cotyledon stage, ~18 days after pollination (DAP) (left). *A. thaliana* Col-0 seed at the bent cotyledon stage, ~7 DAP (right). Scale bar is 100 μm. **b)** Mature dry seed of *A. lyrata* parental strains, their hybrids, and *A. thaliana* Col-0. Scale bar is 1000 μm.



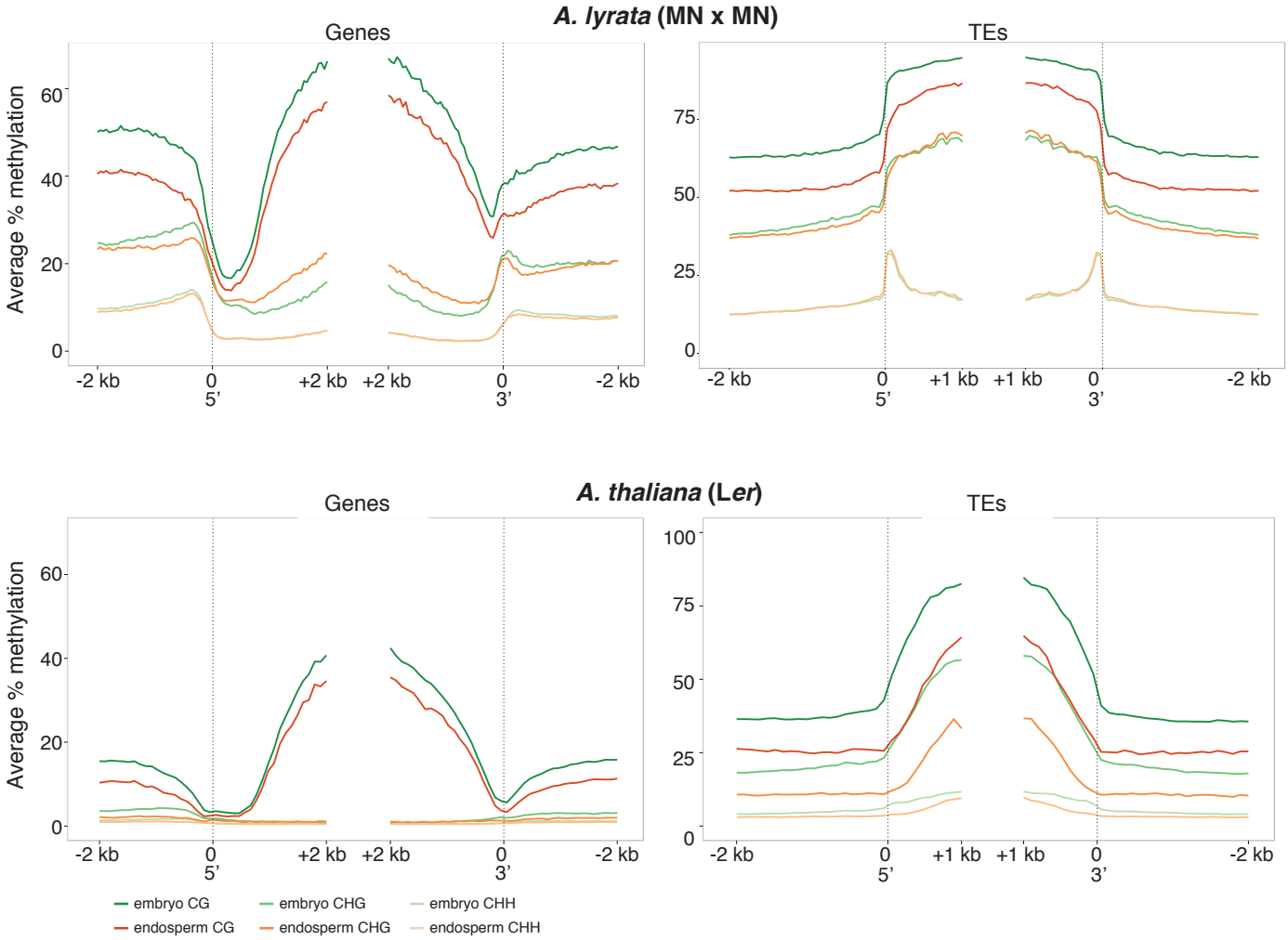
Supplementary Figure 2: Principal component analysis of *A. lyrata* gene expression. PCA of all mRNA-seq samples (left) or embryo and endosperm samples only (right). Female parent in the cross is listed first.



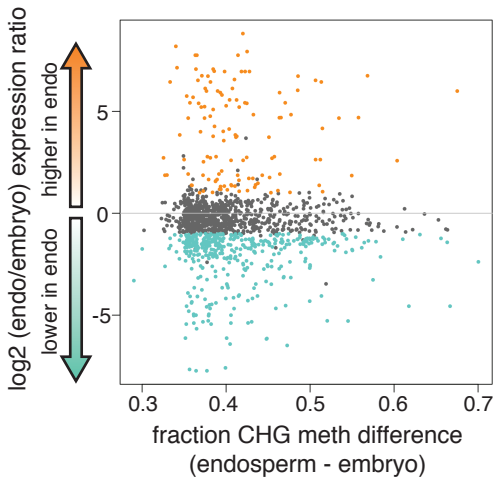
Supplementary Figure 3: Validation of mRNA-seq data by pyrosequencing. Validation of parent-of-origin results for selected SNPs with varying degrees of maternal (**a**) or paternal (**b**) bias. Signal for each SNP from parental or F1 embryo and endosperm is shown. Expected percent maternal or paternal expression was calculated using endosperm mRNA-seq data. Imprinted genes are marked with an asterisk. Blue, MN sequence; Red, Kar sequence; Gray, undetermined signal. M, MN; K, Kar; E, Embryo; N, Endosperm.



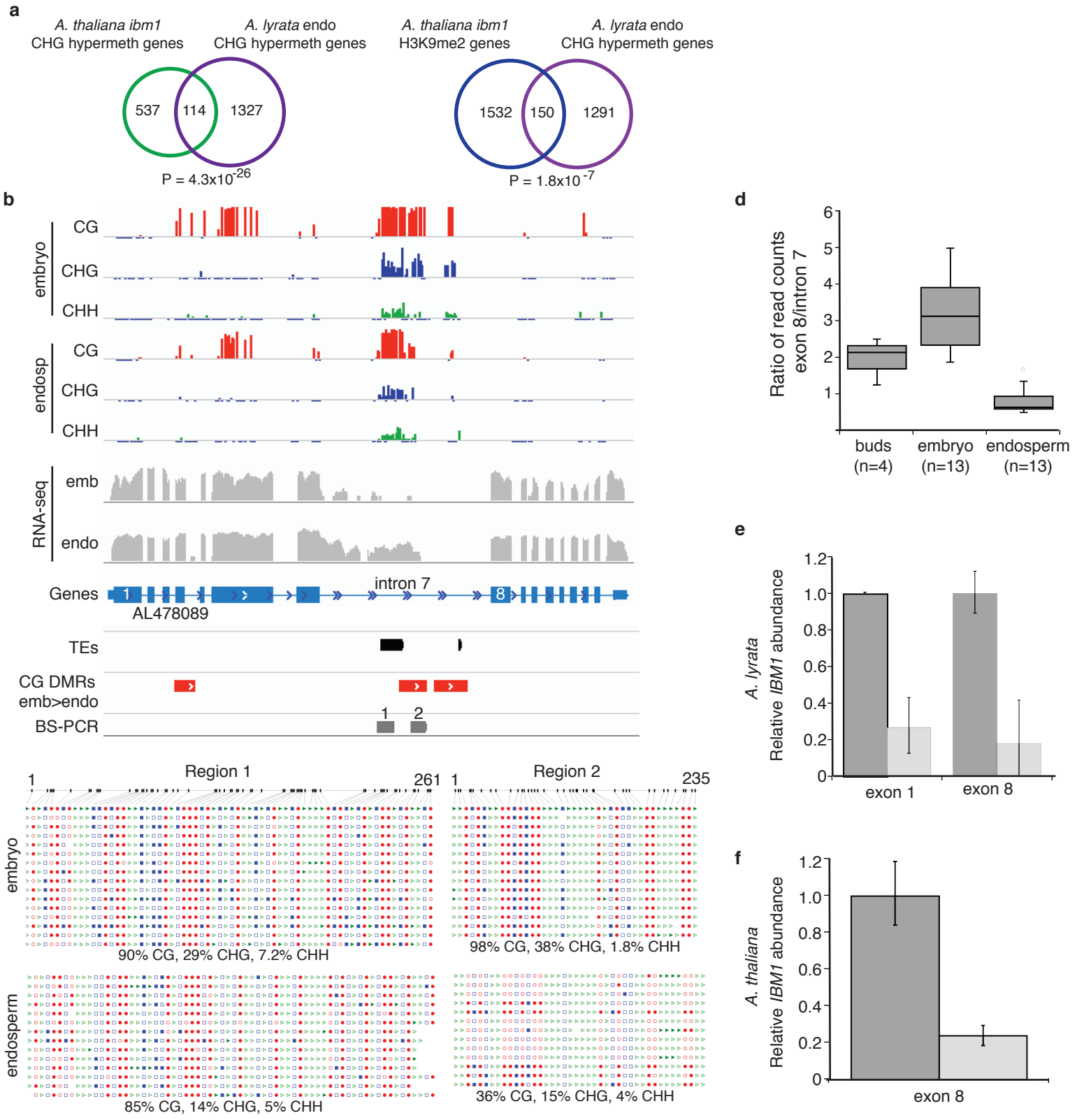
Supplementary Figure 4: Comparison of *A. thaliana* imprinted genes to *A. lyrata*. a) Conservation of imprinting between *A. thaliana* and *A. lyrata* endosperm. Each gene in the union set of *A. thaliana* imprinted genes in Pignatta *et al*. (2014) with an *A. lyrata* homologue was examined for imprinting or parental bias in *A. lyrata*. Reason for lack of imprinting in *A. lyrata* is given in the legend. **b)** Comparison of endosperm parental bias in *A. thaliana* and *A. lyrata* for all genes considered imprinted in Col-L*er* reciprocal crosses in *A. thaliana* from Pignatta *et al*. (2014). To obtain representative % maternal values for MN-Kar and Col-L*er*, reads for all replicates were pooled and % maternal was calculated using pooled counts for both reciprocal crosses in a pair separately. The average % maternal for the two reciprocal crosses was used as the final representative value for MN-Kar or Col-L*er*. Colors of points correspond to categories in (a). **c)** Same as (b), but % maternal for *A. thaliana* was calculated from Col-Cvi reciprocal cross data and plotted for all genes considered imprinted in Col-Cvi crosses in *A. thaliana* from Pignatta *et al.* (2014).



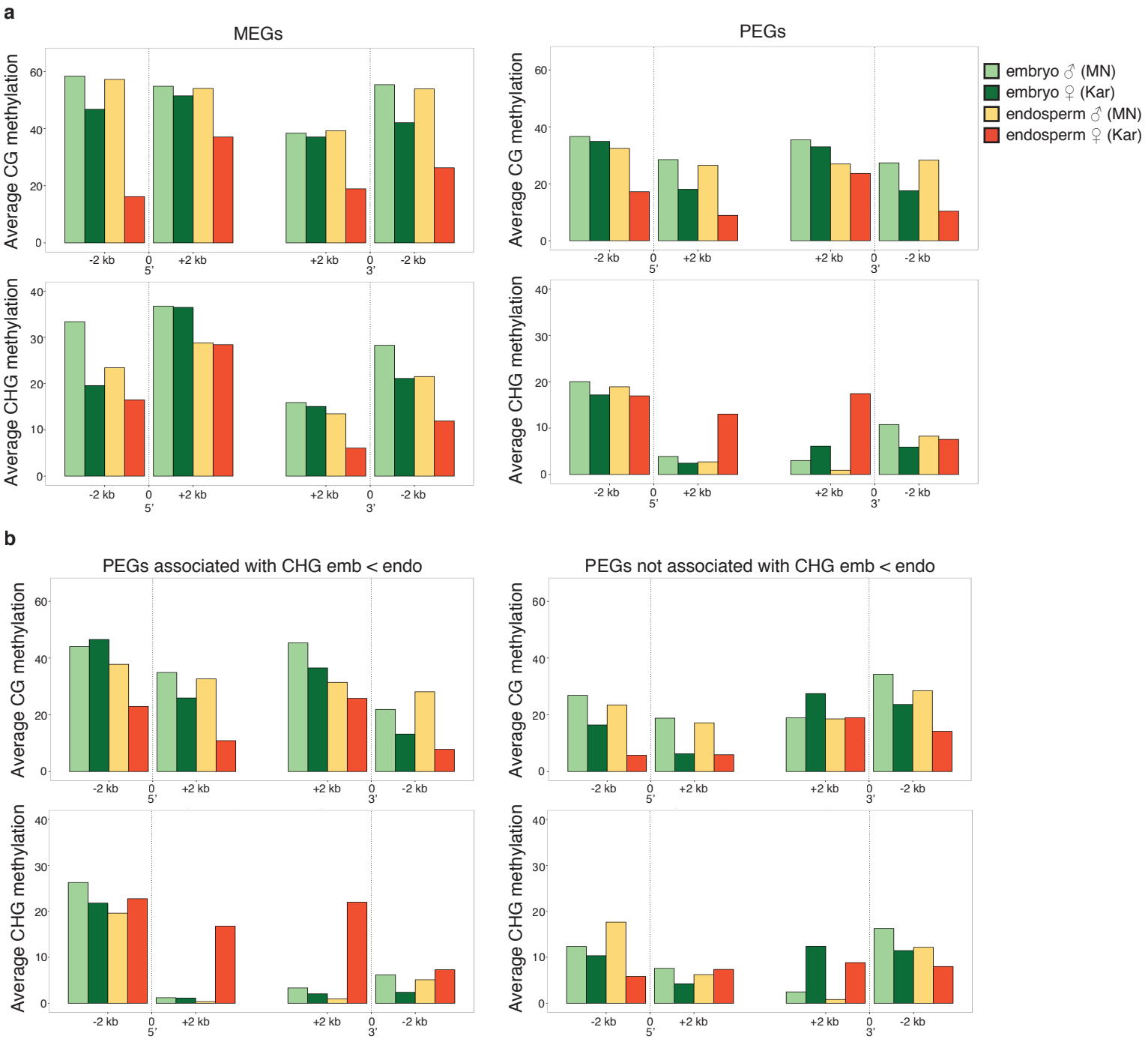
Supplementary Figure 5: Comparison of average methylation profiles in *A. lyrata* and *A. thaliana* seed tissues. Comparison of methylation over genes and TEs in *A. lyrata* (top) and *A. thaliana* (bottom) embryo and endosperm. Methylation data for *A. thaliana* were taken from Pignatta *et al.* 2014 for the L*er* strain. *A. thaliana* has lower levels of methylation in all contexts, and *A. thaliana* gene bodies do not gain CHG methylation in endosperm, unlike what is observed in MN x MN *A. lyrata*.



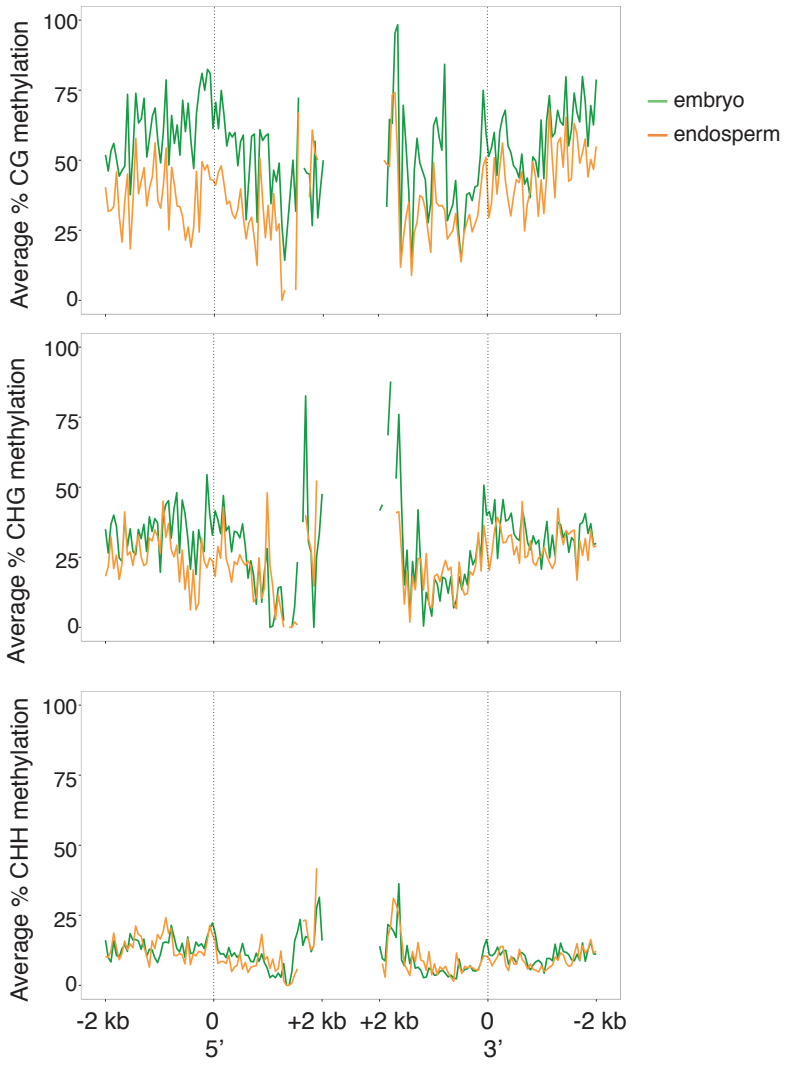
Supplementary Figure 6: Expression bias of genes that gain CHG methylation in endosperm relative to embryo. Scatterplot showing the average MN x MN endosperm *vs.* embryo gene expression ratio, calculated using DESeq2, for genes that overlap an endosperm CHG hypermethylated DMR. X-axis is the CHG methylation difference between endosperm and embryo for each gene. Genes with signficantly different expression are highlighted. Orange; higher expression in endosperm; cyan, lower expression in endosperm.



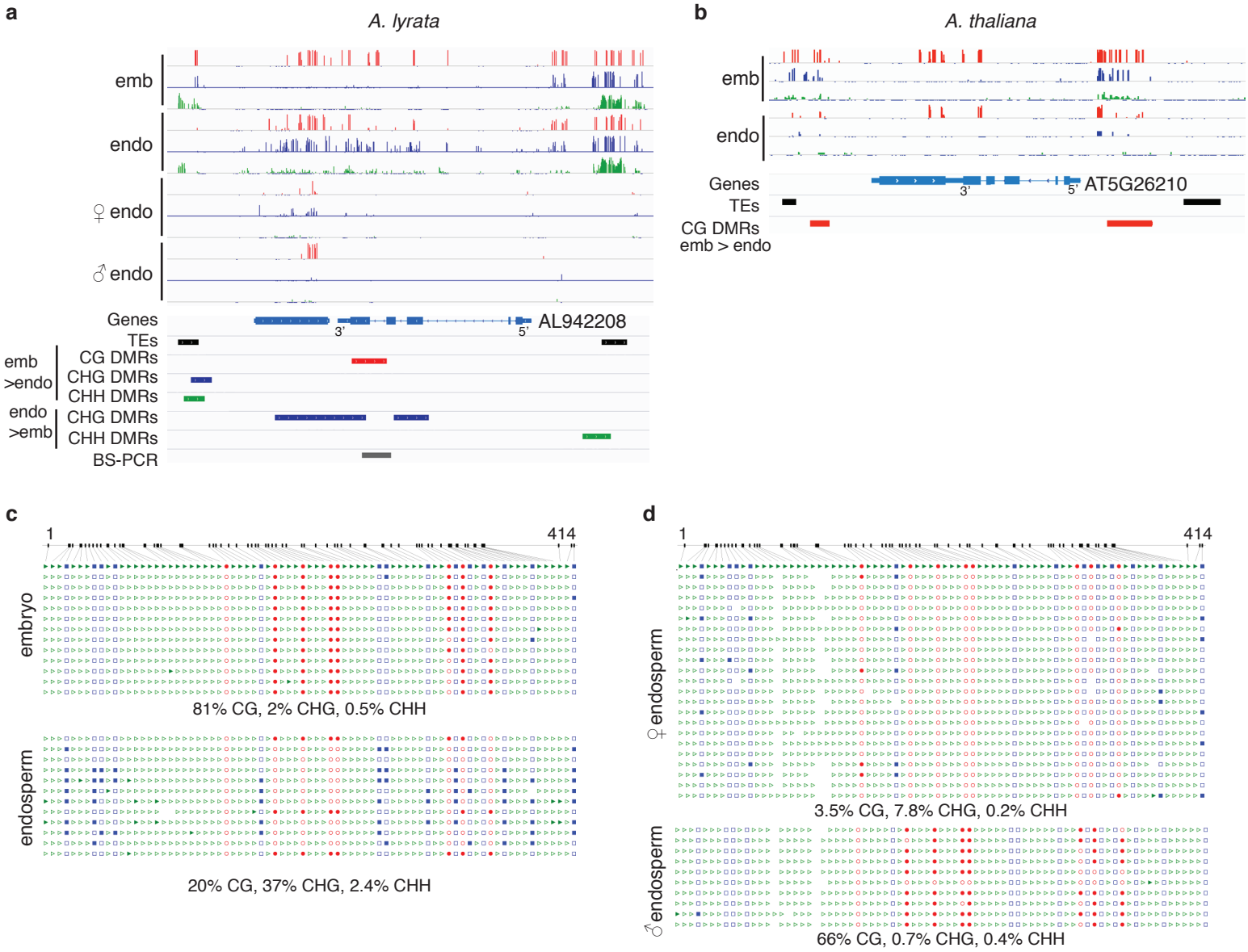
Supplementary Figure 7: Reduced *IBM1* mRNA in *A. lyrata* endosperm is correlated with reduced intronic methylation. **a)** Overlap between *A. thaliana* orthologues of *A. lyrata* endosperm CHG hypermethylated genes and genes identified in Miura *et al.* (2009) as strongly DNA hypermethylated in *ibm1* (left) and genes identified as gaining H3K9me2 in *ibm1* from Deleris *et al.* 2012 (right). P values for significance of overlap determined using the hypergeometric test. **b)** BS-seq and mRNA-seq of the MN x MN *A. lyrata IBM1* locus. RNA accumulates in intron 7 in the endosperm, correlated with reduced intronic methylation in that tissue. **c)** Locus-specific bisulfite PCR validation of embryo-endosperm CG DMR (region 2) in *IBM1* intron and lack of a DMR in region 1. **d)** Box plot of read count ratios in *IBM1* exon 8 (long transcript form) and intron 7 (short transcript form) in *A. lyrata* buds, embryo, and endosperm mRNA-seq datasets. **e)** Abundance of *IBM1* transcripts relative to actin in *A. lyrata* embryo (dark gray) and endosperm (light gray) determined by RT-qPCR. Data is from 2 biological replicates. Error bars show standard deviation. **f)** Abundance of *IBM1* transcripts relative to AT1G58050 in *A. thaliana* embryo (dark gray) and endosperm (light gray). Data is from 2 biological replicates of Col torpedo stage seeds. Error bars show standard deviation.



Supplementary Figure 8: Average methylation levels within and around imprinted genes. a) Average % methylation levels on the maternal (Kar) and paternal (MN) alleles of *A. lyrata* embryo and endosperm in the CG and CHG contexts across all MEGs and PEGs. Average methylation values are calculated over 4 separate regions (2 kb flanking the TSS and TTS). **b)** Average % methylation calculated over the 27 PEGs associated with endosperm CHG hypermethylation (left) and the 22 PEGs not associated with endosperm CHG hypermethylation (right).



Supplementary Figure 9: Methylation patterns of MEGs. Average methylation profiles in the CG, CHG and CHH contexts in MN x MN embryo and endosperm for MEGs aligned at their 5' and 3' ends.



Supplementary Figure 10: Distinct methylation profiles at the conserved PEG AT5G26210. a) *A. lyrata* bisulfite-seq methylation profile at AL942208, a conserved PEG homologous to AT5G26210. Methylation data from MN x MN embryo and endosperm are shown, along with allele-specific endosperm methylation data from Kar x MN. All MN x MN embryo-endosperm DMRs are indicated. Red tracks, CG methylation; blue, CHG methylation; green, CHH methylation. Tick marks below the line indicate Cs with sufficient coverage but no methylation. All tracks set at 100%. **b)** Methylation of the corresponding region in *A. thaliana*. Data is from L*er* x L*er*, published in Pignatta *et al.* (2014). **c)** BS-PCR validation of a CG hypomethylated and CHG hypermethylated DMR indicated in (a) in MN x MN embryo and endosperm. **d)** Allele-specific BS-PCR of the same region from Kar x MN endosperm.