sequences from all 11 human G serotypes shows that most of the variability is in residues on the outward-facing surface of the VP7 trimer (fig. S4). Because there is no extensive conserved patch on this surface, any potential cellular receptor that binds VP7 on the virion must have a very small binding footprint, or its identity and site of interaction must vary among serotypes or isolates. It has been suggested that VP7 may interact with eXb2 integrins after attachment (28). The site proposed as an integrin-binding motif (GPR, resides 253 to 255) is on the inward-facing surface of the trimer and would only be available to interact with integrins after uncoating.

We have designed a disulfide-linked variant of the VP7 trimer by substituting cysteines for Thr276 and Gln305, which face each other across the subunit contact. The resulting disulfide is largely buried at the interface. The secreted product of insect cell expression is indeed stably trimeric (fig. S5). The arm-grip mode of association with VP6 thus allows the VP7 trimer to bind and to lock VP4 in place, but cross-linking of the core subunits retards dissociation, probably by several orders of magnitude. These observations support the notion that withdrawal of Ca2+ is the uncoating trigger in an endosome.

Rotavirus infection and parental immunization with virions both induce a strong VP7-specific neutralizing antibody response (1). Recombinant VP7 elicits neutralizing antibodies only inefficiently, probably because the free trimer dissociates; the response can be enhanced by adding a C-terminal membrane anchor, which presumably increases trimer stability by immobilizing the subunit in two dimensions on the cell surface (5, 29). The disulfide-cross-linked VP7 trimer, which binds neutralizing antibodies such as mAb 159, is a good first candidate for a more effective, stable, structurally engineered subunit immunogen.

References and Notes
30. We thank M. Babayevshev for help with protein preparations, the staff of the Northeastern Collaborative Access Team (Advanced Photon Source, Argonne National Laboratory) for assistance with x-ray data collection, R. Stanfield for the library of Fab structures, and J. Patton for help with sequence variability analysis. Supported by NIH grant CA-13202 (S.C.H.), by a VA Merit Award and NIH grants AI-21362 and DK-56339 (H.B.G.), and by an Ellison Medical Foundation New Investigators in Infectious Diseases Award (P.R.D.). S.C.H. is an Investigator of the Howard Hughes Medical Institute. Atomic coordinates and structure factor amplitudes for the VP4-148 Fab complex have been deposited in the Protein Data Bank [www.rcsb.org] with accession code 3FMG.

Supporting Online Material
www.sciencemag.org/cgi/content/full/324/5933/1444/DC1
Materials and Methods
Figs. S1 to S5
Tables S1 and S2
References
5 January 2009; accepted 28 April 2009
10.1126/science.1170481

Extensive Demethylation of Repetitive Elements During Seed Development Underlies Gene Imprinting

Mary Gehring, Kerry L. Bubb, Steven Henikoff*

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 found 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 the identification of 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 an 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-0 and Ler (fig. S1A) (3). Methylated DNA was immunoprecipitated with an antibody to 5-methylcytosine, sequenced with Illumina Genome Analyzer technology (Illumina, San Diego, CA), and aligned to the reference Col-0 genome. We created methylation profiles 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-0 and 0.89 for Ler) and share similar features with other whole-

Howard Hughes Medical Institute (HHMI), Fred Hutchinson Cancer Research Center (FHCRC), 1100 Fairview Avenue North, Seattle, WA 98109, USA.

*To whom correspondence should be addressed. E-mail: stever@fhcrc.org

www.sciencemag.org SCIENCE VOL 324 12 JUNE 2009
1447

Downloaded from http://science.sciencemag.org/ on January 8, 2020
genome methylation profiles that have been generated for *Arabidopsis* with other platforms (Fig. 1 and fig. S1B) (4, 5). Methylation levels are relatively high in gene-poor heterochromatinic 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 genes are on average less methylated in the endosperm than embryo (Fig. 1B). If the calculation of average methylation profiles 5' and 3' of protein-coding genes excludes methylation from transposable elements or their fragments, the methylation difference between embryo and endosperm in regions flanking genes almost entirely disappears, which indicates that repetitive elements are hypomethylated in the endosperm (Fig. 1C). These results suggest that there is a genome-wide decrease in methylation in the endosperm as compared with 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 do embryos or leaves (6).

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-base pair (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 (7), whereas those 5' of *FWA* (8) fall just below it (+1.18 versus a cutoff of 1.20) (Fig. 2B). This cutoff also detects previously hypothesized methylation differences (9, 10) at the *PHE1* and *FIS2* imprinted genes (Fig. 2B). About 90% of the top DMRs have a positive score, which indicates 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 (11). The top DMRs were almost threefold enriched in regions of the genome corresponding to transposable elements (12) and small RNAs (13) (table S2), which indicates 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 15 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 kilobase (kb) upstream or 2 kb downstream of genes (fig. S3) (7, 14). We identified all protein-coding genes in which 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 Ler embryo-endosperm comparison (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...
Fig. 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 indicate regions within the top 0.5% of methylation differences; the gray arrow indicates a region below the cutoff.

Fig. 3. Methylation is lost 5′ and 3′ of genes in the endosperm. Shown are 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 was determined every 100 bp.
overall hypomethylated as compared with wild type [supporting online material (SOM) text]. Thus, much of the methylation 5′ and 3′ of genes that is depleted in the endosperm is probably lost because of active demethylation by DME in the central cell before fertilization, although other mechanisms probably also contribute (15).

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 (16) are less methylated at 5′ sequences in the endosperm than embryo (fig. S8), which suggests 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 (16). To identify previously unknown 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 Lar data sets. A set of 113 genes have top DMRs in three of the four 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. As with FWA, three HD-ZIP genes are expressed primarily or exclusively in siliques: HDG3, HDG8, and HDG9 (17).

To test parent-of-origin–specific expression of putative imprinted genes, we performed reciprocal crosses between Ler and Col-gl and assayed expression patterns by means of reverse transcriptase polymerase chain reaction (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 text). HDG3 is reciprocally imprinted; expression is predominantly paternal (Fig. 4B and SOM text). 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 text).

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). It is bi-
of genetic and epigenetic material that can be utilized by the host (19). 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, is susceptible to epigenetic regulation by TEs. Regulation of gene expression by means of DNA methylation could be selected for if imprinting of these genes is adaptive in the context of parental conflict or genetic dosage balance in the triploid endosperm.

References and Notes
3. Materials and methods are available as supporting material on Science Online.

Genome-Wide Demethylation of Arabidopsis Endosperm

Tzung-Fu Hsieh,* Christian A. Ibarra,* Pedro Silva,* Assaf Zemach, Leor Eshed-Williams, Robert L. Fischer,*†Daniel Zilberman†

Parent-of-origin-specific (imprinted) gene expression is regulated in Arabidopsis thaliana endosperm by cytosine demethylation of the maternal genome mediated by the DNA glycosylase DEMETER, but the extent of the methylation changes is not known. Here, we show that virtually the entire endosperm genome is demethylated, coupled with extensive local non-CG hypermethylation of small interfering RNA–targeted sequences. Mutation of DEMETER partially restores endosperm CG methylation to levels found in other tissues, indicating that CG demethylation is specific to maternal sequences. Endosperm demethylation is accompanied by CHH hypermethylation of embryo transposable elements. Our findings demonstrate extensive reconfiguration of the endosperm methylation landscape that likely represses transposon silencing in the embryo.

Gene imprinting, the differential expression of alleles of the same gene depending on parent-of-origin, independently evolved in mammals and in flowering plants (1). Imprinting occurs in the placenta of mammals and the endosperm of plants, structures that nourish the developing embryo. Maternal allele expression in the central cell, the diploid maternal plant cell that is fertilized to give rise to the triploid endosperm, is activated by the DEMETER (DME) DNA glycosylase, which excises 5-methylcytosine, resulting in imprinted expression of several genes in the endosperm (2–8). Although important for imprinting, DNA methylation in flowering plants primarily silences transposons, retrotransposons, and repeated sequences (9). In addition to methylation in the CG sequence context, plant DNA methylation occurs at CHG (H is A, C or T) and CHH sites, with CHH and to a lesser extent CHG methylation mediated through active targeting by RNA interference (RNAi) machinery (9). Arabidopsis gene bodies are commonly methylated in the CG context, whereas all types of methylation are present in repeats (10, 11). A given CG site is generally methylated over 80% or not at all, whereas methylation of a CHG site is typically 30 to 80%, and methylation of a CHH site tends to be below 30% (10, 11).

To determine the methylation landscape during Arabidopsis seed development, we isolated DNA from wild-type embryos, wild-type endosperm, and endosperm from seeds with a defective maternal allele of DME, and adult aerial tissues, and used the Illumina Genome Analyzer platform to quantify DNA methylation by high-throughput bisulfite sequencing (10–12) (bisulfite treatment converts unmethylated cytosine to uracil) (fig. S1). We aligned 2.5 billion bases for embryo, 2.2 billion bases for wild-type endosperm, 2.0 billion bases for dme endosperm, and 1.5 billion bases for aerial tissues, which corresponds to 21-fold, 18-fold, 16-fold, and 13-fold coverage of the Arabidopsis genome, respectively (13) (table S1). Our aerial tissue results closely matched previously published bisulfite sequencing data (table S2 and fig. S2).

Bulk methylation in wild-type endosperm (20.9% CG, 8.9% CHG, 2.8% CHH) was lower in all sequence contexts compared with the embryo (26.9% CG, 10.6% CHG, and 4.4% CHH) (Fig. 1 and fig. S3). CG methylation was reduced in both gene bodies and repeats (Fig. 1, A and B) and was partially restored in dme endosperm (23.1%). In the developing seed, DME is expressed only in the central cell before fertilization (2), indicating that we were primarily detecting demethylation of the maternal endosperm genome. In contrast to CG methylation, CHH methylation was decreased (8.9% to 5.8%) in dme endosperm (Fig. 1, C and D), whereas CHH methylation was reduced by a factor of 3.5 (2.6% to 0.8%) (Fig. 1, E and F). CG and CHG methylation in aerial tissues (25.7% and 9.4%, respectively) was somewhat lower than in embryos, and aerial CHH methylation (2.3%) was half of that found in embryos and even lower than that of endosperm (Fig. 1), indicating that small interfering RNA (siRNA)-mediated DNA methylation is enhanced in the seed. Reduced non-CG methylation in dme endosperm suggests that DME activity is necessary for up-regulating siRNA-mediated methylation, perhaps through activation of transposable elements by DNA demethylation.

To identify sequences that are differentially methylated in the endosperm compared with the embryo, we calculated fractional methylation in each context within 50 base pair (bp) windows and subtracted endosperm methylation from embryo methylation. We identified 36,749 discreet loci corresponding to 10.33 million bp with an absolute change in CG methylation of at least 10% (P < 0.0001, Fisher’s exact test), 99.4% of which (36,534) were more methylated in embryo (table S3). Using the same criteria, we found 5694 loci (2.87 million bp) with a change in CHG methylation, 91.3% of which (5200) were more methylated in embryo (table S3). We also identified 9749 loci (17.98 million bp) with an absolute change in CHH methylation of at least 5% (P < 0.0001, Fisher’s exact test), 89.9% of which (8760) were more methylated in embryo (table S3). Although the above values represent a substantial underestimate, they provide a clear indication of the extent of methylation...
Dynamic Imprinting

Gene imprinting—the silencing of either a maternally derived or paternally derived gene allele—is controlled in large part by DNA methylation. In plants, imprinting occurs in the endosperm, which nourishes the embryonic plant. Gehring et al. (p. 1447) and Hsieh et al. (p. 1451) analyzed the dynamics of DNA methylation in the endosperm and embryo of Arabidopsis and found extensive demethylation in the endosperm, suggesting that many imprinted genes are likely to exist. Gehring et al. characterized five imprinted genes in detail. Four of the 10 known imprinted genes are related homeodomain transcription factors. Furthermore, 5′ sequences demethylated in several of the genes were found to be derived from transposable elements, which supports the idea that imprinting arose as a by-product of silencing invading DNA.