Conservation and divergence of methylation patterning in plants and animals

Suhua Feng[a,b,1], Shawn J. Cokus[b,c,1], Xiaoyu Zhang[d], Pao-Yang Chen[b,c], Magnolia Bostick[b], Mary G. Goll[e], Jonathan Hetzel[b], Jayati Jain[e], Steven H. Strauss[f], Marnie E. Halpern[e], Chinweike Ukomadu[g], Kirsten C. Sadler[h], Sriharsa Pradhan[i], Matteo Pellegrini[b,c,2], and Steven E. Jacobsen[a,b,2]

[a]Howard Hughes Medical Institute, [b]Department of Molecular, Cell and Developmental Biology, and [c]Institute for Genomics and Proteomics, University of California, Los Angeles, CA 90095; [d]Department of Plant Biology, University of Georgia, Athens, GA 30602; [e]Department of Embryology, Carnegie Institution for Science, Baltimore, MD 21218; [f]Department of Forest Ecosystems and Society, Oregon State University, Corvallis, OR 97331; [g]Division of Gastroenterology, Department of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115; [h]Department of Medicine, Division of Liver Disease and Department of Developmental and Regenerative Biology, Mount Sinai School of Medicine, New York NY 10029; and [i]New England BioLabs, Ipswich, MA 01938

Edited* by James C. Carrington, Oregon State University, Corvallis, OR, and approved March 19, 2010 (received for review March 4, 2010)

Cytosine DNA methylation is a heritable epigenetic mark present in many eukaryotic organisms. Although DNA methylation likely has a conserved role in gene silencing, the levels and patterns of DNA methylation appear to vary drastically among different organisms. Here we used shotgun genomic bisulfite sequencing (BS-Seq) to compare DNA methylation in eight diverse plant and animal genomes. We found that patterns of methylation are very similar in flowering plants with methylated cytosines detected in all sequence contexts, whereas CG methylation predominates in animals. Vertebrates have methylation throughout the genome except for CpG islands. Gene body methylation is conserved with clear preference for exons in most organisms. Furthermore, genes appear to be the major target of methylation in Ciona and honey bee. Among the eight organisms, the green alga Chlamydomonas has the most unusual pattern of methylation, having non-CG methylation enriched in exons of genes rather than in repeats and transposons. In addition, the Dnmt1 cofactor Uhrf1 has a conserved function in maintaining CG methylation in both transposons and gene bodies in the mouse, Arabidopsis, and zebrafish genomes.

BS-Seq | epigenetic profiling | DNA methylation | gene body methylation | UHRF1

Cytosine DNA methylation is an epigenetic mark important in many gene regulatory systems, including genomic imprinting, X-chromosome inactivation, silencing of transposons and other repetitive DNA sequences, as well as expression of endogenous genes. Methylation is conserved in most major eukaryotic groups, including many plants, animals, and fungi, although it has been lost from certain model organisms such as the budding yeast *Saccharomyces cerevisiae* and nematode worm *Caenorhabditis elegans* (1–3). DNA methylation can be categorized into three types according to the sequence context of the cytosines, namely CG, CHG, and CHH (H = A, C, or T). CG methylation is maintained by conserved Dnmt1 DNA methyltransferase enzymes. CHH methylation, and, to some extent CHG methylation, is generally maintained by the activity of the conserved Dnmt3 methyltransferases, whereas high levels of CHG methylation seen in the model plant Arabidopsis are maintained by the plant-specific methyltransferase CMT3 (2, 3). Generally speaking, DNA methylation is thought to occur "globally" in vertebrates, with CG sites being heavily methylated genome-wide except for those in CpG islands, whereas invertebrates, plants, and fungi have "mosaic" methylation, characterized by interspersed methylated and unmethylated domains (4). These differences are an interesting starting point for studying divergence in methylation pathways and regulatory mechanisms; however, determining precise genome-scale methylation patterns has been a challenge for complex genomes until the recent development of high-throughput sequencing technology. In this paper, we generated shotgun bisulfite sequencing data to profile DNA methylation in eight eukaryotic organisms. These organisms display wide variations in methylation levels and targets. Interestingly, methylation in genes, exons, repetitive DNA, and transposons shows distinct conservation and divergence among the tested organisms, helping to define the evolution of DNA methylation patterning and function.

Results and Discussion

To compare methylation patterns across species, we shotgun sequenced bisulfite-treated DNA from eight eukaryotic organisms, including *Arabidopsis thaliana* (Arabidopsis), *Oryza sativa* (rice), *Populus trichocarpa* (poplar), *Chlamydomonas reinhardtii* (green algae), *Ciona intestinalis* (sea squirt), *Apis mellifera* (honey bee), *Danio rerio* (zebrafish), and *Mus musculus* (mouse), using a previously described approach named BS-Seq. (5) (Fig. 1). Because of the very large genome sizes of some of these organisms (Tables S1 and S2), we chose a sequencing depth that permits accurate assessment of the level of methylation of major genomic elements including genes, transposons, and repeats (Table S3), but not sufficient for quantification of the level of methylation of individual cytosines. We previously analyzed methylation in the Arabidopsis genome at a high sequence depth (5), and for comparison we included here a lower coverage Arabidopsis BS-Seq dataset. We found that conclusions drawn from the low-coverage Arabidopsis BS-Seq data were similar to those of the previously published deeply sequenced Arabidopsis methylomes (5–7), validating the lower coverage approach for the analysis of major classes of genomic elements. For each organism, in addition to the nuclear chromosomes, we also report the observed overall methylation levels in the scaffolds and contigs that are yet not assembled onto nuclear chromosomes, as well as in chloroplast genomes where appropriate (Fig. 1, Figs. S1 and S2, and Table S2).

We found that the methylation patterns of the three flowering plants, the monocot, rice, and the dicots, Arabidopsis and poplar, were largely similar. Methylation was detected in all three cytosine contexts, with CG sites methylated at the highest level, CHG sites at a medium level, and CHH sites at the lowest level (Figs. 1–3, Figs. S1 and S3, and Table S2). All three types of methylation were highly enriched in repetitive DNA and transposons (Fig. 3 and

Author contributions: S.H.S., M.E.H., C.U., K.C.S., S.P., M.P., and S.E.J. designed research; S.F., M.B., M.G.G., J.H., and J.J. performed research; S.F., S.J.C., X.Z., P.-Y.C., M.P., and S.E.J. analyzed data; and S.F., S.J.C., M.P., and S.E.J. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

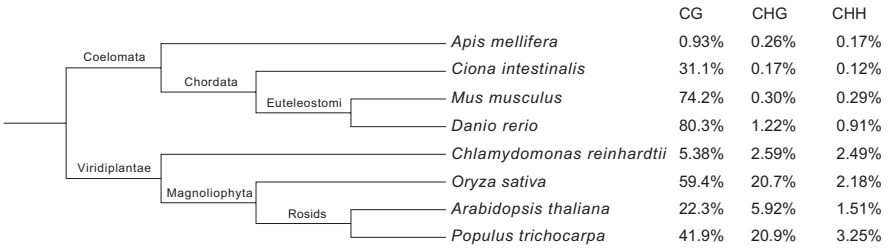
Data deposition: The sequence reads reported in this article have been deposited in the NCBI Gene Expression Omnibus (GEO) database, www.ncbi.nlm.nih.gov/geo (accession no. GSE21152).

[1]S.F. and S.J.C. contributed equally to this work.

[2]To whom correspondence may be addressed. E-mail: matteop@mcdb.ucla.edu or jacobsen@ucla.edu.

This article contains supporting information online at www.pnas.org/cgi/content/full/1002720107/DCSupplemental.

GENETICS



	CG	CHG	CHH
Apis mellifera	0.93%	0.26%	0.17%
Ciona intestinalis	31.1%	0.17%	0.12%
Mus musculus	74.2%	0.30%	0.29%
Danio rerio	80.3%	1.22%	0.91%
Chlamydomonas reinhardtii	5.38%	2.59%	2.49%
Oryza sativa	59.4%	20.7%	2.18%
Arabidopsis thaliana	22.3%	5.92%	1.51%
Populus trichocarpa	41.9%	20.9%	3.25%

Fig. 1. Overall methylation levels in eight eukaryotic organisms. Tree topology is from NCBI Taxonomy (http://www.ncbi.nlm.nih.gov/guide/taxonomy/). Methylation levels shown are from main chromosomes/linkage groups/scaffolds of each organism (Tables S1 and S2). All tissues are wild type. Data from Arabidopsis shoots and mouse embryos are shown here; for other tissues used are described in *Materials and Methods.* In the case of poplar, more than one third of the genome sequence exists in scaffolds not placed on any of the main linkage groups. These scaffolds have higher methylation levels (Table S2), as well as increased repeat density, increased sequence ambiguity, and low mapability (Fig. S2). This is consistent with the notion that the sequences in these scaffolds are of a highly repetitive nature that prevents them from being assembled properly, and methylation is enriched in repetitive DNA.

Fig. S3). Methylation also exhibits peaks in pericentromeric heterochromatin regions except for CHH methylation in rice, which might reflect the fact that repeats are widely spread across the rice genome, in contrast to the more localized pattern of repeats close to centromeres in Arabidopsis and poplar (Fig. S1). These data suggest that the CMT3 pathway in Arabidopsis, as well as the RNA-directed DNA methylation pathway (RdDM) that operates through the plant Dnmt3 homolog, DRM2, is likely conserved in other flowering plants, consistent with the presence of CMT3 and DRM2 homologs in poplar and rice (8, 9) (Fig. S4). Interestingly, CHG methylation levels in transposons and repeats were much higher in poplar than in Arabidopsis and rice (Fig. 3 and Fig. S3), suggesting that the CMT3 pathway is more active in poplar. CG methylation, but not CHG or CHH methylation, in all three flowering plants exhibited a characteristic peak in the body of protein-coding genes (Fig. 2), a phenomenon first observed genome-wide in Arabidopsis (5, 10, 11). Although the function of this gene body methylation remains unknown, it has been proposed to suppress spurious transcription from cryptic promoters that might otherwise interfere with gene regulation (11).

We also profiled the green alga Chlamydomonas, a unicellular organism in the plant kingdom distantly related to flowering plants such as Arabidopsis, poplar, and rice. Chlamydomonas was previously shown to exhibit heavy methylation of chloroplasts during mating, which contributes to uniparental inheritance of chloroplast DNA (12), but methylation of the nuclear genome has not been characterized. We used vegetative cells not containing gametes in our analysis and found only low levels of methylation in the chloroplast genome (Fig. S1 and Table S2), similar to levels previously reported in vegetative cells (13). We found that the Chlamydomonas nuclear genome had low levels of methylation (~5.4% for CG methylation, and ~2.5% each for CHG and CHH methylation; Fig. 1 and Table S2) with unique distribution in genes and transposons that was very different from all other organisms studied here (Figs. 2 and 3 and Figs. S1 and S3). First, although we found a slight enrichment of CG methylation in the bodies of genes in Chlamydomonas, this methylation was much lower than in flowering plants (Fig. 2), consistent with reduced levels of CG methylation overall. Nevertheless, we found evidence that genes (and exons in particular) are preferentially methylated, as discussed later here (Fig. 4). Second, we found that transposon sequences were preferentially methylated (especially LINE elements with methylation levels as high as ~30%), which was true only for CG methylation (Fig. S3). In contrast, CHG and CHH methylation were rather uniformly distributed along the chromosomes and showed little enrichment on either repeats or transposons (Fig. 3 and Fig. S3). In Chlamydomonas, CG, CHG, and CHH methylation are possibly all mediated by Dnmt1/MET1 homolog(s), as it has been shown that CrMET1/DMT1, one of the Chlamydomonas Dnmt1/MET1-like DNA methyltransferase enzymes, has little preference for any particular cytosine sequence context (13, 14). Furthermore, the lack of CHG and CHH methylation targeted to transposons is consistent with the apparent absence of a Dnmt3/DRM2 homolog in Chlamydomonas (8, 9), and that the closest CMT3 homolog is highly diverged relative to flowering plants (Fig. S4). Thus, although Chlamydomonas clearly methylates transposon sequences, the mechanisms involved appear to be different from those in flowering plants.

Ciona is a primitive chordate that is distantly related to vertebrates (15), and, like vertebrates, it possesses Dnmt1 and Dnmt3 homologs (8). Previously, via targeted DNA methylation analysis and computational methods, it was shown that Ciona protein-coding genes show high levels of methylation but that promoters, intergenic DNA, and transposons were not preferentially methylated (16, 17). It was also shown that, like in Arabidopsis, body methylated genes tend to be those with moderate levels of gene

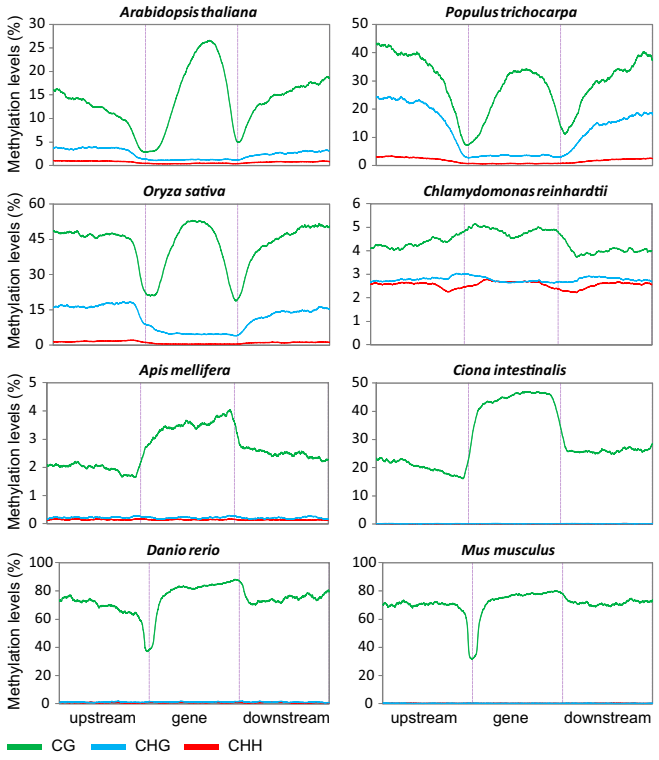


Fig. 2. Distribution of methylation along protein-coding genes. Upstream and downstream regions are the same length as the gene. Only BS-Seq data up to halfway to the next nonoverlapping gene are used in this analysis. Two vertical purple lines mark the gene boundaries. Tissues are the same as in Fig. 1. BS-Seq data annotated to at least one repeat and/or transposon are withheld from this analysis.

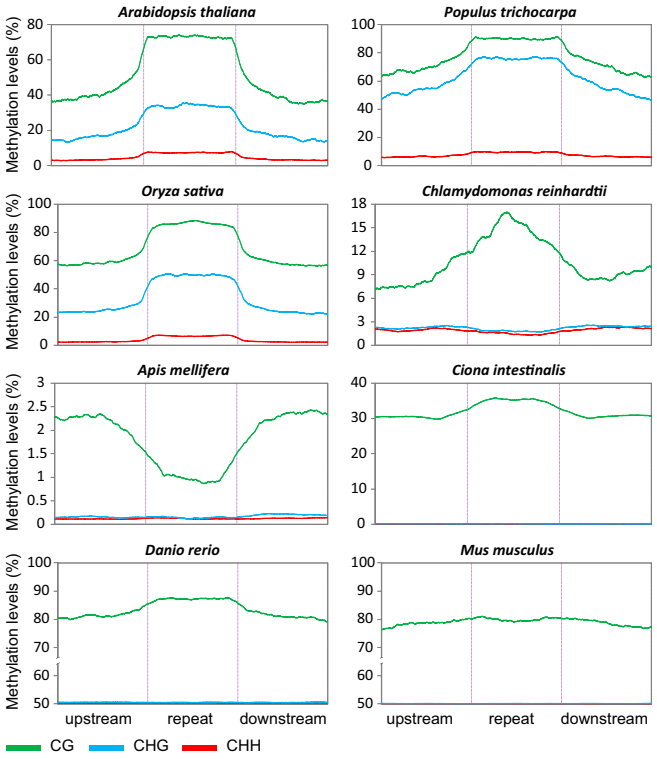


Fig. 3. Distribution of methylation along repetitive DNA. Upstream and downstream regions are the same length as the repeat. Only BS-Seq data up to halfway to the next nonoverlapping repeat are used in this analysis. Two vertical purple lines mark the repeat boundaries. Chlamydomonas repeats shown here are interspersed repeats only. Tissues are the same as in Fig. 1.

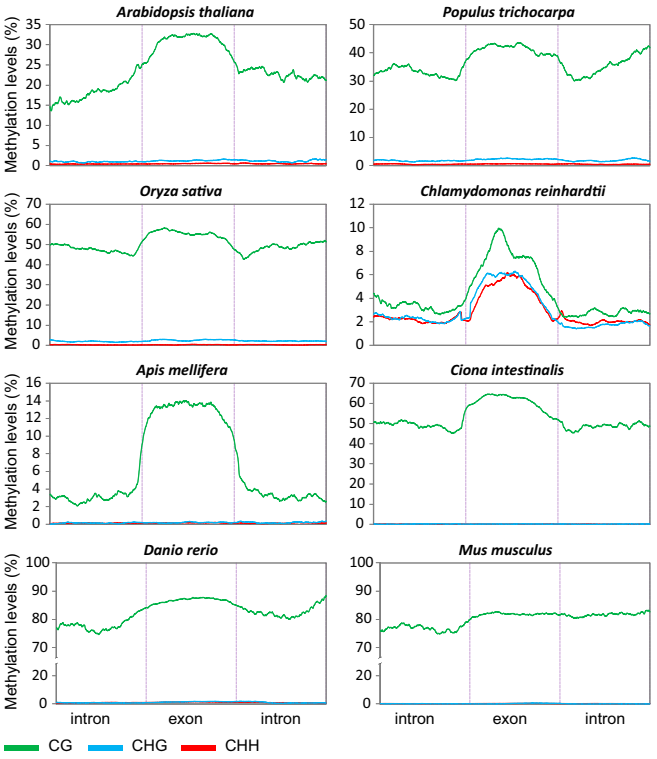


Fig. 4. Comparison of methylation levels across exons and introns. Only internal exons (flanked by introns on both ends) that do not contain any 5′- or 3′-UTR bases are used. Upstream and downstream regions are the same length as the exon. Only BS-Seq data up to halfway to the next exon are used in this analysis. Two vertical purple lines mark the intron–exon and exon–intron boundaries. Tissues are the same as in Fig. 1. BS-Seq data annotated to at least one repeat and/or transposon are withheld from this analysis.

expression (16). We found that methylation in Ciona was nearly exclusively in CG context (Fig. 1 and Table S2). In addition, consistent with previous findings, CG methylation increased sharply in the body of protein-coding genes, a phenomenon reminiscent of gene body CG methylation in flowering plants, except that the methylation level difference between genic and intergenic regions in Ciona was much larger, being almost two times higher in genes (Fig. 2). In Ciona, transposons and other repeats were also moderately methylated. We observed ~10% higher methylation on transposons and repeats relative to immediate flanking regions; but this preference was less pronounced than for genes, supporting the idea that genes are a major target of methylation in Ciona (Fig. 3 and Fig. S3).

Previous studies of methylation in insects suggest that, although the model organism *Drosophila melanogaster* does not have a conserved methylation system, some other related insects such as honey bee have apparently functional Dnmt1 and Dnmt3 DNA methyltransferases (18). Studies of individual honey bee genes have shown evidence of genic methylation (19–21). We found that the honey bee genome contains very low levels of CHG or CHH methylation and only ~1% CG methylation (Fig. 1 and Table S2). Interestingly, this CG methylation was concentrated in genes, with a pattern in the body of genes that roughly resembles that of Ciona and flowering plants (Fig. 2). In contrast, repetitive sequences were not preferentially methylated (Fig. 3 and Fig. S3), suggesting that genes are the main targets of DNA methylation in honey bee.

We profiled methylation of the embryos of two vertebrate species, mouse and zebrafish, which revealed very high levels of CG methylation (~74% and ~80% respectively) and low levels of CHG and CHH methylation (Fig. 1 and Table S2). Many previous mammalian methylation studies (using a variety of techniques, including shotgun bisulfite sequencing in human) have shown that CG methylation generally covers the entire nuclear genome, with the exception of so-called CpG rich islands near the promoters of many active genes, which tend to be unmethylated (4, 22). Our results confirmed these earlier conclusions for mouse, and also showed that the zebrafish genome has a roughly similar pattern of global methylation. Interestingly, in both mouse and zebrafish, there was slightly higher CG methylation in the body of genes than in intergenic regions and depletion of methylation around transcriptional start sites, coinciding with CpG islands (Fig. 2 and Fig. S5). These patterns bear some similarity to gene body methylation in the other organisms profiled in this study, although this conclusion is more difficult to draw given the high global level of CG methylation in the vertebrate genomes.

Together, the finding of gene body methylation in a variety of organisms in the plant and animal kingdoms suggests that this phenomenon may have an ancient origin predating the last common ancestor of plants and animals.

To examine gene body methylation further, we compared methylation levels across introns and exons (Fig. 4). We found that methylation in Arabidopsis was higher in exons than in introns, which was also confirmed by reanalyzing a published microarray-based Arabidopsis methylome dataset (10) (Table 1). Rice, poplar, and Ciona also showed clear enrichment of methylation in exons, although of a smaller magnitude than Arabidopsis. Honey bee showed the largest increase of methylation in exons compared with introns. The overall low average level of methylation in the honey bee genome (Figs. 1–3, Figs. S1 and S3, and Table S2) suggests that CG methylation is located predominantly in exons, consistent with a previous report on several individual honey bee genes (20). Strikingly, in Chlamydomonas, all three types of methylation increased

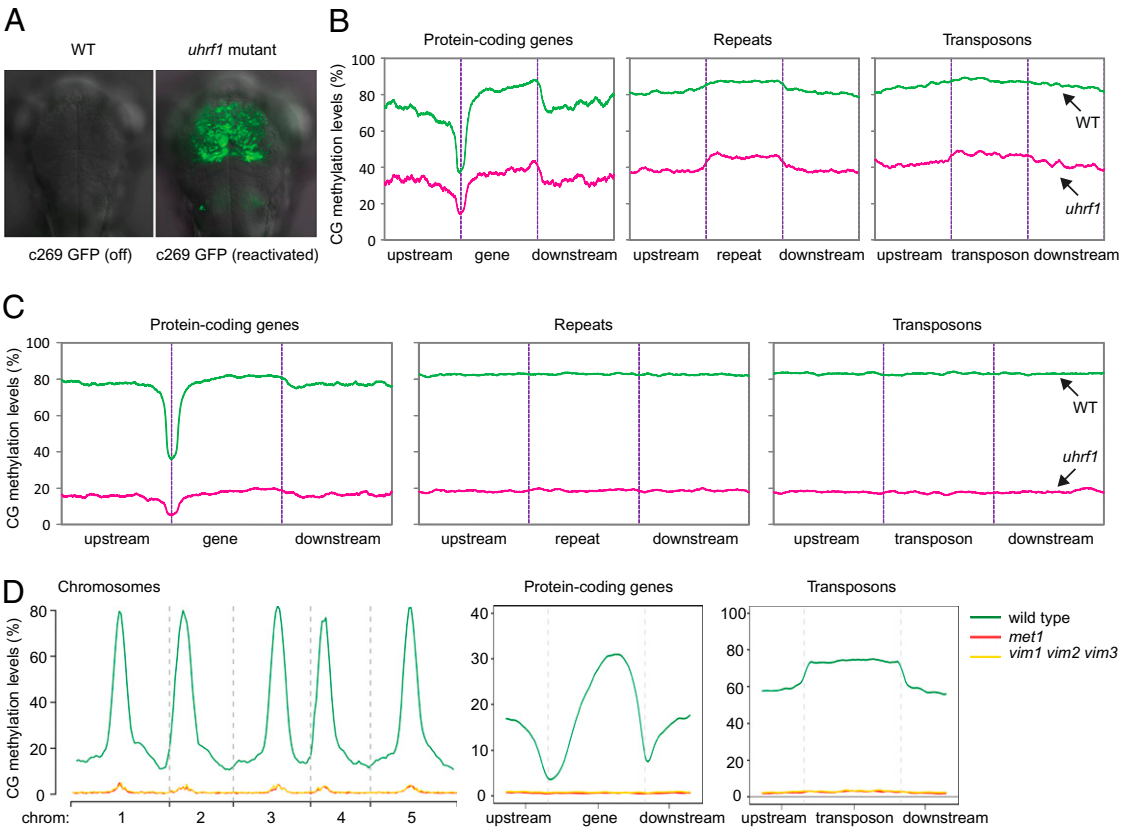


Fig. 5. Decrease of CG methylation in mutants lacking Uhrf1. (*A*) Reactivation of a transcriptionally silenced transgene in zebrafish *uhrf1* homozygous mutants. WT larvae (3 days postfertilization) carrying the silenced allele of the brain-specific transgene Gt(Gal4-VP16;UAS:EGFP) c269 GFP (off) do not show any GFP labeled cells in the brain (*Left*) due to methylation of CG sites in the multicopy UAS (33). On the contrary, GFP labeled cells resulting from reactivation of the silenced c269 transgene are readily detected in the brains of *uhrf1* homozygous mutant larvae (*Right*). (*B–D*) Comparison of CG methylation levels between wild-type and *uhrf1* mutant lines from zebrafish (*B*), mouse (*C*), and Arabidopsis (*D*). For Arabidopsis, *met1* mutant is used as a control for virtual total loss of CG methylation. Tissues used are mouse embryonic stem cells, Arabidopsis flowers, and zebrafish embryos.

in exons, in contrast to repeats and transposons that were enriched only for CG methylation (Figs. 3 and 4 and Fig. S3). This further suggests that, although the enrichment of methylation is not as clear as in other organisms (Fig. 2), gene bodies do appear to be a preferential target of methylation in Chlamydomonas. Furthermore, exons are the only features we found where non-CG methylation is enriched in Chlamydomonas (Figs. 2–4 and Fig. S3). In the two vertebrates, zebrafish and mouse, there was only a slight increase of CG methylation in exons, consistent with the overall high genome-wide levels of CG methylation. Overall, the preference of methylation for exons in the various plant and animal species suggests that, like gene body methylation, exon methylation may be an ancestral condition. It has been recently shown that exons are enriched in nucleosome content relative to introns (23–25), so it is tempting to speculate that nucleosomes might act to guide DNA methyltransferases resulting in higher methylation levels on exons. A further speculative idea is that certain histone marks present on gene bodies might act to specifically target methylation to genes.

In all eight of the organisms tested, an ortholog of the CG maintenance methyltransferase Dnmt1/MET1 is present, consistent with the detection of substantial levels of CG methylation in all cases (Fig. 1) (8, 9, 19, 26). Molecular genetic studies in mouse and Arabidopsis have shown that a PHD/SRA/RING protein called Uhrf1 (mouse) / VIM (Arabidopsis) is an important cofactor that works with Dnmt1 (mouse) or MET1 (Arabidopsis) to maintain CG DNA methylation (27–30). Uhrf1 acts in part by directly recognizing hemimethylated DNA that is produced after DNA replication, likely facilitating Dnmt1 activity at these sites (27, 28, 31). Intriguingly, like Dnmt1/MET1, Uhrf1 homologs are found in all eight of the species in this study, but are lacking in organisms such as yeast, *C. elegans*, and *Drosophila* that lack CG methylation (Fig. S6), suggesting a conserved role for Uhrf1 in eukaryotic methylation. To gain additional evidence for a conserved function for Uhrf1, we profiled DNA methylation in a zebrafish *uhrf1* hypomorphic mutant line (32). We observed a strong reduction of CG DNA methylation (from ~80% to ~40%) (Table S2), consistent with a function of zebrafish Uhrf1 in maintaining CG DNA methylation. Moreover, we detected reactivation of a methylated and silenced transgene (33) in *uhrf1* mutant zebrafish (Fig. 5), further supporting a conserved role of Uhrf1 and

Table 1. Microarray comparison of methylation levels between exons and introns in Arabidopsis

	Total (bp)	Not methylated (bp)	Methylated (bp)	% Methylated
Exons	28,162,672	21,739,481	6,423,191	22.81%
Introns	13,179,008	11,558,059	1,620,949	12.30%

Affymetrix tiling array data analyzed here are from a previous study (10). Methylation status was determined by TileMap HMM on unique probes. A threshold of 0.5 was applied to obtain binary methylation calls per 25-mer probe, crediting the call of "methylated" or "not methylated" to all bases of the probe. Subsequently, methylation level in exons or introns (% methylated) was calculated by number of base pairs methylated / (methylated + not methylated).

DNA methylation in gene silencing. We also compared methylation of mouse wild-type ES cells with an ES cell line containing a targeted disruption allele of *Uhrf1* (5, 27, 34) and found that loss of *Uhrf1* caused a reduction of CG methylation (from ~80% to ~18%) (Table S2). Finally, we compared wild-type Arabidopsis with a *vim1 vim2 vim3* triple mutant defective in the three functional Uhrf1 homologs in Arabidopsis (29). We found that the *vim* triple mutant had virtually lost CG methylation (Fig. 5). Together these genetic results strongly support a conserved role of Uhrf1 in the maintenance of CG DNA methylation.

To test whether Uhrf1 is differentially required for methylation of genes or repeats, we compared the methylation pattern of wild-type or mutant Arabidopsis, mouse, and zebrafish at these sites. In all cases, loss of Uhrf1 resulted in a strong loss of DNA methylation in both gene bodies and in repeats and transposons (Fig. 5 and Fig. S7). The *uhrf1* mutations also reduced CG methylation levels in both exons and introns (Fig. S7). In the case of the Arabidopsis *vim* triple mutant, consistent with results at individual genes (29), we found that CG methylation was virtually eliminated in all parts of the genome, similar to the effects of a null *met1* allele (Fig. 5). Together, these results suggest that Uhrf1 serves a relatively global and conserved role as an essential Dnmt1 cofactor.

In summary, our survey of methylation patterning in diverse organisms reveals that, in addition to the well-known phenomenon of transposon methylation, methylation of protein-coding genes is also common among different plant and animal groups and tends to favor exons. Intriguingly, despite its wide presence, neither the biological function of gene body methylation nor the mechanisms by which genes are recognized by DNA methylation systems are understood. In flowering plants, the absence of CHG and CHH methylation in gene bodies demonstrates that genes and transposons are differentially recognized, and in Ciona and honey bee the preferential methylation on genes further supports that genes are being targeted specifically. The conserved methylated DNA binding factor Uhrf1 is also required for the maintenance of gene body methylation in plants and animals, indicating that this methylation is in part maintained by DNA replication-coupled maintenance DNA methylation. Understanding how gene body DNA methylation is initially targeted and why exons are favored over introns are key questions for future research.

Materials and Methods

The Arabidopsis *vim1 vim2 vim3* triple mutant and *uhrf1* mutant lines for mouse and for zebrafish have been described previously (29, 32, 34). The tissues used for making BS-Seq libraries were shoots and flowers from Arabidopsis (strain Col-0), leaves from Japonica rice (strain Nipponbare), leaves from poplar (strain Nisqually-1), a single immature male honey bee (strain DH4), vegetative cells from Chlamydomonas (strain CC503), Ciona animals (collected from Half Moon Bay, CA), 5-day-old embryos from zebrafish, and E13.5 embryos (strain C57BL/6J) and ES cells (line E14) from mouse. For zebrafish, heterozygous *uhrf1-h272* (32) adult zebrafish were incrossed to produce homozygous mutant embryos. At least 20 mutants and their phenotypically wild-type siblings were pooled on 5 days postfertilization, and genomic DNA was isolated.

BS-Seq DNA library construction and high-throughput sequencing by Illumina/Solexa Genome Analyzer were performed according to published protocols (5), except that a total number of six PCR cycles were removed from the library generation procedure. NEBNext DNA sample preparation kit (New England BioLabs) and CpGenome DNA Modification Kit (Chemicon) were used during library construction. DNA methylation error rates were estimated by the level of apparent methylation in the unmethylated Arabidopsis chloroplast genome as described previously (5), which were between 0.28% and 0.5% as described in Table S2.

Tandem and inverted repeats were identified de novo using the software packages Tandem Repeat Finder and Inverted Repeat Finder, respectively (10, 35, 36). Dispersed repeats for honey bee were identified de novo using the software RECON (10, 37). Dispersed repeats in other genomes were identified using the software RepeatMasker (RepeatMasker Open-3.0.1996–2004, http://www.repeatmasker.org) with repeat libraries from Repbase.

Gene annotations were obtained from Gramene (http://www.gramene.org/) and TAIR (http://www.arabidopsis.org/) for Arabidopsis, rice, and poplar, from Ensembl (http://www.ensembl.org/) for honey bee, Ciona, zebrafish, and mouse, and from JGI (http://www.jgi.doe.gov/) for Chlamydomonas. Transposon annotations were obtained from the same databases, except for Arabidopsis and Chlamydomonas, the transposons of which were identified using RepeatMasker with repeat library from Repbase (see above).

Computational methods employed were identical to those described previously (5) except as follows. Gaussian mixture models for basecalling were estimated by a robust FAST-MCD method (38) rather than expectation maximization. Constant tag bases at the beginning of BS-Seq reads as a result of the library preparation protocol were used to inform per-read prior forward (FW) vs. reverse (RC) strand probabilities. Basecalling and mapping parameters were adjusted to compensate for Illumina instrument/reagent kit/instrument software pipeline changes. For each lane, some genomic bases are covered by unusually many uniquely-mapped BS-Seq reads (e.g., because of the incomplete nature of reference genomes leading to some repetitive sequences being classified as unique); uniquely mapping reads contributing to coverage of the 1% most highly covered bases by such were suppressed.

The methylation level (%) for any given genomic interval refers to the ratio of the number of BS-Seq "methylated" bases aligned to any genomic cytosine (in CG/CHG/CHH context) in that interval to the number of methylated or unmethylated bases aligned to the same. For methylation plots, the *x* axis is divided into 3,000 bins (1,000 bins for upstream region; 1,000 bins for gene/repeat/transposon/exon region; and 1,000 bins for downstream region). The methylation level shown for any given *x* axis bin is that from the genomic interval corresponding to a window centered around that bin extending ±50 or more bins. The number of bins varied to accommodate differences in genome sizes, annotation object interval lengths, and BS-Seq sequencing depth.

Note Added in Proof. Another paper by Zemach et al. (39), which examined additional species, also supports the conservation of gene body methylation.

ACKNOWLEDGMENTS. We thank Z. Chen and S.F. Nelson for sharing sequencing equipment; N. Jiang for assistance in the identification of repeats; and H. La, B. Ding, G.L. Wang, J. Kropat, S. Merchant, G. Robinson, E. Wagner, M. Levine, A. Girard, G. Hannon, A. Conway, A. Clark, C. Popp, W. Dean, W. Reik, and E. Richards for providing tissues and DNA samples. S.F. is a Howard Hughes Medical Institute Fellow of the Life Sciences Research Foundation. P.C. is funded by Eli & Edythe Broad Center of Regenerative Medicine & Stem Cell Research at University of California at Los Angeles. M.B. is funded by National Institutes of Health National Research Service Award (5F32CA126302). M.G.G. is a Fellow of the Damon Runyon Cancer Research Foundation (DRG-#1945-07). K.C.S. is supported by the Basil O'Connor Starter Scholar Award from the March of Dimes and a Young Investigator award from the Breast Cancer Alliance. This work was supported in part by grants from the National Science Foundation Plant Genome Research Program and the Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research at University of California at Los Angeles (to S.E.J and M.P.), the National Institutes of Health (GM60398 to S.E.J., R01HD058530 to M.E.H., and R01DK080789 to C.U. and K.C.S.), and the Department of Energy Office of Science (DE-FC02-02ER63421 to M.P. and DE-FG02-08ER64665 to S.H.S.). S.E.J. is an investigator of the Howard Hughes Medical Institute.

1. Henderson IR, Jacobsen SE (2007) Epigenetic inheritance in plants. *Nature* 447: 418–424.
2. Chan SW, Henderson IR, Jacobsen SE (2005) Gardening the genome: DNA methylation in Arabidopsis thaliana. *Nat Rev Genet* 6:351–360.
3. Goll MG, Bestor TH (2005) Eukaryotic cytosine methyltransferases. *Annu Rev Biochem* 74:481–514.
4. Suzuki MM, Bird A (2008) DNA methylation landscapes: Provocative insights from epigenomics. *Nat Rev Genet* 9:465–476.
5. Cokus SJ, et al. (2008) Shotgun bisulphite sequencing of the Arabidopsis genome reveals DNA methylation patterning. *Nature* 452:215–219.
6. Hsieh TF, et al. (2009) Genome-wide demethylation of Arabidopsis endosperm. *Science* 324:1451–1454.
7. Lister R, et al. (2008) Highly integrated single-base resolution maps of the epigenome in Arabidopsis. *Cell* 133:523–536.
8. Ponger L, Li WH (2005) Evolutionary diversification of DNA methyltransferases in eukaryotic genomes. *Mol Biol Evol* 22:1119–1128.
9. Pavlopoulou A, Kossida S (2007) Plant cytosine-5 DNA methyltransferases: Structure, function, and molecular evolution. *Genomics* 90:530–541.
10. Zhang X, et al. (2006) Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis. *Cell* 126:1189–1201.
11. Tran RK, et al. (2005) DNA methylation profiling identifies CG methylation clusters in Arabidopsis genes. *Curr Biol* 15:154–159.
12. Umen JG, Goodenough UW (2001) Chloroplast DNA methylation and inheritance in Chlamydomonas. *Genes Dev* 15:2585–2597.

13. Nishiyama R, Ito M, Yamaguchi Y, Koizumi N, Sano H (2002) A chloroplast-resident DNA methyltransferase is responsible for hypermethylation of chloroplast genes in Chlamydomonas maternal gametes. *Proc Natl Acad Sci USA* 99:5925–5930.
14. Nishiyama R, et al. (2004) Role of a nonselective de novo DNA methyltransferase in maternal inheritance of chloroplast genes in the green alga, *Chlamydomonas reinhardtii*. *Genetics* 168:809–816.
15. Dehal P, et al. (2002) The draft genome of *Ciona intestinalis*: Insights into chordate and vertebrate origins. *Science* 298:2157–2167.
16. Suzuki MM, Kerr AR, De Sousa D, Bird A (2007) CpG methylation is targeted to transcription units in an invertebrate genome. *Genome Res* 17:625–631.
17. Simmen MW, et al. (1999) Nonmethylated transposable elements and methylated genes in a chordate genome. *Science* 283:1164–1167.
18. Consortium THGS, Honeybee Genome Sequencing Consortium (2006) Insights into social insects from the genome of the honeybee Apis mellifera. *Nature* 443:931–949.
19. Schaefer M, Lyko F (2007) DNA methylation with a sting: An active DNA methylation system in the honeybee. *Bioessays* 29:208–211.
20. Wang Y, et al. (2006) Functional CpG methylation system in a social insect. *Science* 314:645–647.
21. Kucharski R, Maleszka J, Foret S, Maleszka R (2008) Nutritional control of reproductive status in honeybees via DNA methylation. *Science* 319:1827–1830.
22. Lister R, et al. (2009) Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* 462:315–322.
23. Tilgner H, et al. (2009) Nucleosome positioning as a determinant of exon recognition. *Nat Struct Mol Biol* 16:996–1001.
24. Schwartz S, Meshorer E, Ast G (2009) Chromatin organization marks exon-intron structure. *Nat Struct Mol Biol* 16:990–995.
25. Andersson R, Enroth S, Rada-Iglesias A, Wadelius C, Komorowski J (2009) Nucleosomes are well positioned in exons and carry characteristic histone modifications. *Genome Res* 19:1732–1741.
26. Anderson RM, et al. (2009) Loss of Dnmt1 catalytic activity reveals multiple roles for DNA methylation during pancreas development and regeneration. *Dev Biol* 334:213–223.
27. Bostick M, et al. (2007) UHRF1 plays a role in maintaining DNA methylation in mammalian cells. *Science* 317:1760–1764.
28. Sharif J, et al. (2007) The SRA protein Np95 mediates epigenetic inheritance by recruiting Dnmt1 to methylated DNA. *Nature* 450:908–912.
29. Woo HR, Dittmer TA, Richards EJ (2008) Three SRA-domain methylcytosine-binding proteins cooperate to maintain global CpG methylation and epigenetic silencing in Arabidopsis. *PLoS Genet* 4:e1000156.
30. Johnson LM, et al. (2007) The SRA methyl-cytosine-binding domain links DNA and histone methylation. *Curr Biol* 17:379–384.
31. Hashimoto H, et al. (2008) The SRA domain of UHRF1 flips 5-methylcytosine out of the DNA helix. *Nature* 455:826–829.
32. Sadler KC, Krahn KN, Gaur NA, Ukomadu C (2007) Liver growth in the embryo and during liver regeneration in zebrafish requires the cell cycle regulator, uhrf1. *Proc Natl Acad Sci USA* 104:1570–1575.
33. Goll MG, Anderson R, Stainier DY, Spradling AC, Halpern ME (2009) Transcriptional silencing and reactivation in transgenic zebrafish. *Genetics* 182:747–755.
34. Muto M, et al. (2002) Targeted disruption of Np95 gene renders murine embryonic stem cells hypersensitive to DNA damaging agents and DNA replication blocks. *J Biol Chem* 277:34549–34555.
35. Benson G (1999) Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Res* 27:573–580.
36. Warburton PE, Giordano J, Cheung F, Gelfand Y, Benson G (2004) Inverted repeat structure of the human genome: The X-chromosome contains a preponderance of large, highly homologous inverted repeats that contain testes genes. *Genome Res* 14:1861–1869.
37. Bao Z, Eddy SR (2002) Automated de novo identification of repeat sequence families in sequenced genomes. *Genome Res* 12:1269–1276.
38. Rousseeuw PJ, van Driessen K (1999) A fast algorithm for the minimum covariance determinant estimator. *Technometrics* 41:212–223.
39. Zemach A, McDaniel IE, Silva P, Zilberman D (2010) Genome-wide evolutionary analysis of eukaryotic DNA methylation. Science, 10.1126/science.1186366.