Eukaryotic cytosine methylation represses transcription, but also occurs in the bodies of active genes, and the extent of methylation biology conservation is unclear. Here we quantify DNA methylation in seventeen eukaryotic genomes, and find that gene body methylation is conserved between plants and animals, whereas selective methylation of transposons is not. We show that methylation of plant transposons in the CHG context extends to green algae, and present evidence for RNA-directed DNA methylation of fungal genes. Exclusion of histone H2A.Z from methylated DNA is conserved between plants and animals. Our data demonstrate that extant DNA methylation systems are mosaics of conserved and derived features, and indicate that gene body methylation is an ancient property of eukaryotic genomes.

Our knowledge about cytosine methylation has been derived mainly from four species: humans, mouse, Arabidopsis thaliana, and Neurospora crassa (I). Transposons and repeats are virtually uniformly methylated in all four, suggesting that transposon defense is an ancient function of methylation (I, 2). Expression of imprinted genes is regulated by DNA methylation in A. thaliana and mammals (3), and both methylate bodies of active genes (4–6). However, these analogies do not necessarily imply conservation. Regulation of imprinting has evolved independently in mammals and flowering plants (3). Many transposons are not methylated in the tunicate Ciona intestinalis (7), and the transposon-rich silk moth Bombyx mori genome contains low levels of 5-methylcytosine (8), putting into question the conservation of transposon methylation (9). To understand how eukaryotic DNA methylation has evolved, we quantified DNA methylation by deep bisulfite sequencing in the genomes of five plants, seven animals and five fungi (Figs. S1 to S6 and tables S1 to S3). We also profiled transcription in these species by RNA-seq (10). The results revealed a complex evolutionary history of the DNA methylation pathway, and allowed us to reconstruct a plausible ancestral state.

Rice is a model monocot that contains Dnmt1 (CG methyltransferase), Dnmt3 (de novo methyltransferase responsible for plant CHH methylation (H = A, C or T)) and CMT (plant-specific CHG methyltransferase) orthologs (I) (Figs. S1 to S5 and table S4). Our data show that CG methylation is lowest from 100 bp upstream of the transcriptional start site (TSS) of rice genes to 500 bp within the transcript, plateauing around 1.5 kb following the TSS (Fig. 1A). Non-CG methylation is essentially absent from genes, while methylation in all contexts is abundant in transposable elements (TEs; Fig. 1, A and B), with short TEs particularly enriched in CHH methylation (Fig. S7).

Consistent with functioning to repress transcription, gene expression varies inversely with methylation of the TSS-proximal region (Fig. 1C). The same pattern is present at the 3’ end of genes, suggesting that lack of methylation around the transcription termination site is also important for gene expression (Fig. 1C). Within the gene body plateau, rice methylation exhibits a parabolic relationship with transcription – modestly expressed genes are most likely to be methylated, while genes at either transcriptional extreme are least likely (Fig. 1C and fig. S8). In all salient features, rice methylation patterns closely resemble those of A. thaliana (5, 6).

We examined methylation in two early diverging land plants: Selaginella moellendorffii and Physcomitrella patens (Fig. S1) (11), both possessing Dnmt1, Dnmt3 and CMT orthologs (Figs. S1 to S5). S. moellendorffii TEs are methylated in the CG, CHG and CHH contexts, but unlike the angiosperms, genes are not significantly methylated, and there is essentially no methylation around the TSS regardless of transcription (Fig. 1, D to F, and table S1). P. patens methylation patterns closely resemble those of S. moellendorffii (table S1 and fig. S9) (10). DNA methylation in these plants is strictly segregated away from genes.

We analyzed two green algae, Chlorella sp. NC64A and Volvox carteri (Figs. S1 to S3). Chlorella has the greatest amount of CG methylation of any organism analyzed here (Fig. 2, G to I, and table S1): genes are methylated virtually without exception (Fig. S10), with a sharp drop of methylation at the promoter (Fig. 2G). Promoter methylation correlates negatively with gene expression (Fig. 2I), suggesting that TSS-proximal methylation represses transcription, as it does in land plants. Although CMT genes have thus far only been described in land plants (J), Chlorella contains a CMT homolog (Figs. S1 to S3) and a significant amount of CHG.
methylation (table S1), which is concentrated in repetitive elements and excluded from genes (Fig. 2, G to I). The *V. carteri* genome is much less methylated exclusively in the CG context, with a similar relationship between promoter methylation and transcription, and preferential methylation of TEs (table S1 and fig. S9). Methylation of both gene bodies and TEs thus appears to be an ancient property of plants.

The puffer fish *Tetraodon nigroviridis* genome is virtually devoid of transposable elements (12), yet we find heavy methylation exclusively at CG sites, the vast majority of which is found in genes, though TEs do show particularly dense methylation (Fig. 2, A and B, and table S1). Genes exhibit a prominent dip in methylation just upstream of the TSS, and reduced promoter methylation correlates with enhanced expression (Fig. 2A). Gene body methylation exhibits the same relationship with transcription that exists in *A. thaliana* and rice – a roughly parabolic curve with the most methylated genes around the 70th transcription percentile (fig. S8).

The similarities between plants and vertebrates prompted us to ask whether aspects of the interaction between DNA methylation and chromatin are conserved. We recently showed that the histone variant H2A.Z exhibits a strong antagonism with DNA methylation in *A. thaliana* (13). Utilizing ChIP-seq, we found that puffer fish H2A.Z peaks in the promoter and is depleted in gene bodies, with a pattern that is almost precisely opposite to that of DNA methylation (Fig. 2C). H2A.Z is enriched in unmethylated promoters regardless of transcription (Fig. 2, D and E), and is quantitatively depleted in methylated gene bodies (fig. S11). Within unmethylated promoters, H2A.Z exhibits a parabolic distribution relative to transcription – H2A.Z is most enriched around the 50th expression percentile (Fig. 2E and fig. S11). Within gene bodies, H2A.Z enrichment varies inversely with transcription (fig. S11). These features are essentially the same as in *A. thaliana* (13), indicating that the interactions between H2A.Z, DNA methylation and transcription are conserved between angiosperms and vertebrates.

We examined six invertebrate species belonging to three major groups that diverged 900-1000 Mya (fig. S1) (11). Flour beetle (*Tribolium castaneum*) adults and *Drosophila melanogaster* 0-3 hour embryos do not have detectable DNA methylation of the nuclear genome (table S1 and fig. S12) (10). The general methylation features of the other four species (*Ciona intestinalis*, honeybee (*Apis mellifera*), silk moth *Bombyx mori*, and anemone (*Nematostella vectensis*)) are quite similar (Fig. 3 and table S1). Gene bodies are methylated in roughly the same pattern as plant and fish genes, with highest methylation of moderately expressed genes, but no correlation between TSS methylation and transcription (Fig. 3 and fig. S8). TEs are hypomethylated, with methylation rising linearly with distance from the TE (Fig. 3 and fig. S10; we did not perform linear correlation analysis in honeybees due to low TE content). There is thus little evidence that methylation inhibits transcription or silences TEs in invertebrates. Our data indicate that gene body methylation is basal, predating the divergence of plants and animals around 1.6 billion years ago (fig. S1), while the anti-transposon function probably evolved independently in the vertebrate and plant lineages.

We analyzed five species belonging to the three major fungal groups. *Phycomyces blakesleeanus* diverged from the Dikarya (ascomycetes and basidiomycetes) over a billion years ago (fig. S1) (11), and has two methyltransferases: a Dnmt1-like and a Dim-2-like (figs. S1 to S3). The presence of Dim-2, previously described only in ascomycetes (I), demonstrates that this is an ancient methyltransferase family. We find a substantial amount of CG methylation and a small amount of non-CG methylation (table S1), both concentrated in transcriptionally silent, repetitive loci (Fig. 4, A and B, and fig. S13) – methylation in this fungus appears to be used to silence TEs and other repeats, while active genes are unmethylated. The three basidiomycete species we examined (*Coprinopsis cinerea*, *Laccaria bicolor* and *Postia placenta*) show very similar methylation patterns (Fig. 4, C and D, table S1, and figs. S13 and S14), indicating that this is an ancient state that was present in the last common ancestor of most, if not all, modern fungi.

*Uncinocarpus reesii* is an ascomycete that is distantly related to *Neurospora crassa* (fig. S1). We find that *U. reesii* methylation is generally similar to that of *N. crassa* (14): silent repeated loci are heavily methylated in all sequence contexts, with little methylation elsewhere (Fig. 4, E to G). There is no preference for CG sites, consistent with lack of Dnmt1 (Fig. 4, E to G, and table S1). In fact, *U. reesii* repeats have very little CG methylation (Fig. 4, E to G), because the repeats contain only 2% of the expected frequency of CG sites, with a corresponding increase of CA and TG dinucleotides (fig. S15), the expected products of C/G to T/A deamination. Our observations are consistent with repeat-induced point mutation (RIP), a DNA methylation-linked process that causes mutation of repeated fungal sequences (15). *N. crassa* RIP was suggested to be an evolutionary hindrance because the process limits evolution by gene duplication (15), but the presence of a severe version of RIP in an organism separated from *N. crassa* by over 600 million years would indicate that RIP is a stable evolutionary strategy.

Unlike all other examined fungi, *U. reesii* exhibits methylation of active genes, which is not accompanied by RIP, so CG methylation is abundant (Fig. 4, E, G, and H). There is a direct, linear relationship between RNA levels and methylation of exons, while introns tend to be unmethylated (Fig. 4H and fig. S16). These features strongly argue that
generic DNA methylation is directed by spliced mRNA. It is remarkable that *U. reesii*, which possesses none of the methyltransferases of plants, evolved to methylate both TEs and active genes, and, like plants, can distinguish between them.

Our data allow us to undertake a reconstruction of the DNA methylation machinery and functionality of the last common ancestor of plants, animals and fungi. This organism possessed Dnmt1, Dnmt3, and probably a CMT/Dim-2 enzyme – structural, functional and phylogenetic data strongly argue that CMT and Dim-2 are monophyletic (figs. S2, S3, and S17) (10). The methylation landscape resembled that of extant angiosperms: transcription-influenced CG methylation of gene bodies and extensive CG and non-CG methylation of TEs. H2A.Z was excluded from methylated DNA. The ancestral system then underwent varying degrees of change. A clue to this plasticity is suggested by the observation that all examined land plants and vertebrates retain extensive DNA methylation (1). These groups share two characteristics: they use methylation to repress TEs, and they reproduce primarily by sexual outcrossing. The degree of sexual outcrossing correlates with TE aggressiveness, because outcrossing partially breaks the link between host and TE fitness (16), so land plants and vertebrates are under strong selection to maintain effective TE suppression. Early unicellular animals and fungi that primarily reproduced asexually were more at liberty to lose TE methylation. In contrast, gene body methylation is likely a double-edged sword: it is mutagenic (1), so whatever benefits are derived by methylating exons come at a price (10). Loss of this pathway would thus be more common, and in fact has occurred early in fungal evolution, and later in several plant and animal lineages.

**References and Notes**

10. See supporting material on Science Online.
17. Acknowledgments are included in supporting material on Science Online. Sequencing data are deposited in GEO with accession number GSE19824.

**Supporting Online Material**

www.sciencemag.org/cgi/content/full/science.1186366/DC1

Materials and Methods

SOM Text

Figures S1 to S17

Tables S1 to S4

References

23 December 2009; accepted 5 March 2010

Published online 15 April 2010; 10.1126/science.1186366

Include this information when citing this paper.