

14. N. Baumberger, D. C. Baulcombe, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 11928 (2005).
15. P. Brodersen *et al.*, *Science* **320**, 1185 (2008); published online 15 May 2008 (10.1126/science.1159151).
16. C. Himber, P. Dunoyer, G. Moissiard, C. Ritzenthaler, O. Voinnet, *EMBO J.* **22**, 4523 (2003).
17. S. Y. Berezina, L. Supekova, F. Supek, P. G. Schultz, A. A. Deniz, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 7682 (2006).
18. O. Voinnet, P. Vain, S. Angell, D. C. Baulcombe, *Cell* **95**, 177 (1998).
19. D. S. Schwarz *et al.*, *Cell* **115**, 199 (2003).
20. D. S. Schwarz, G. Hutvagner, B. Haley, P. D. Zamore, *Mol. Cell* **10**, 537 (2002).
21. Z. Havelda, C. Hornyik, A. Crescenzi, J. Burgyn, *J. Virol.* **77**, 6082 (2003).
22. P. Dunoyer *et al.*, *EMBO J.*, published online 22 April 2010 (10.1038/emboj.2010.65).
23. Research in O.V.'s laboratory is funded by a prize from the Bettencourt Foundation for Life Science Research and a starting grant from the European Research Council "Frontiers of RNAi" ERC 210890. P.D. and G.S. are supported by a research grant from Agence National pour la Recherche (ANR-08-JCJC-0063-01), J.C.C. is supported by grants from the NIH (AI43288) and the NSF (MCB-0618433), and A.T. is supported by a fellowship from the Japanese Society for the Promotion of Science. We thank J. Mutterer for help with image

acquisition, R. Wagner's team for plant care, and members of O.V.'s laboratory for critical reading of the manuscript.

Supporting Online Material

www.sciencemag.org/cgi/content/full/science.1185880/DC1
SOM Text
Figs. S1 to S5
References

14 December 2009; accepted 28 January 2010
Published online 22 April 2010;
10.1126/science.1185880
Include this information when citing this paper.

Genome-Wide Evolutionary Analysis of Eukaryotic DNA Methylation

Assaf Zemach, Ivy E. McDaniel, Pedro Silva, Daniel Zilberman*

Eukaryotic cytosine methylation represses transcription but also occurs in the bodies of active genes, and the extent of methylation biology conservation is unclear. We quantified DNA methylation in 17 eukaryotic genomes and found 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 that exclusion of histone H2A.Z from methylated DNA is conserved between plants and animals, and we present evidence for RNA-directed DNA methylation of fungal genes. 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* (1). Transposons and repeats are virtually uniformly methylated in all four, which suggests that transposon defense is an ancient function of methylation (1, 2). The expression of imprinted genes is regulated by DNA methylation in *A. thaliana* and mammals (3), and in both cases the bodies of active genes are methylated (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 deep sequencing of cDNA (10). The results reveal a complex evolutionary history of the DNA meth-

ylation pathway and allow us to reconstruct a plausible ancestral state.

Rice (*Oryza sativa*) 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 (1) (figs. S1 to S5 and table S4). Our data show that CG methylation is lowest from 100 base pairs (bp) upstream of the transcriptional start site (TSS) of rice genes to 500 bp within the transcript, plateauing around 1.5 kb after the TSS (Fig. 1A). Non-CG methylation is essentially absent from genes, whereas 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, which suggests 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, whereas genes at either transcriptional extreme are least likely to be methylated (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 have little methylation, 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. 1, G to I, and table S1): Genes are methylated virtually without exception (fig. S10), with a sharp drop of methylation at the promoter (Fig. 1G). Promoter methylation correlates negatively with gene expression (Fig. 1I), which suggests that TSS-proximal methylation represses transcription, as it does in land plants. Although CMT genes have thus far been described only in land plants (1), *Chlorella* contains a CMT homolog (figs. S1 to S3) and a substantial amount of CHG methylation (table S1), which is concentrated in repetitive elements and excluded from genes (Fig. 1, 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 displays 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 genome of the puffer fish *Tetraodon nigroviridis* 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, although 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

Department of Plant and Microbial Biology, 211 Koshland Hall, University of California, Berkeley, CA 94720, USA.

*To whom correspondence should be addressed. E-mail: danielz@berkeley.edu

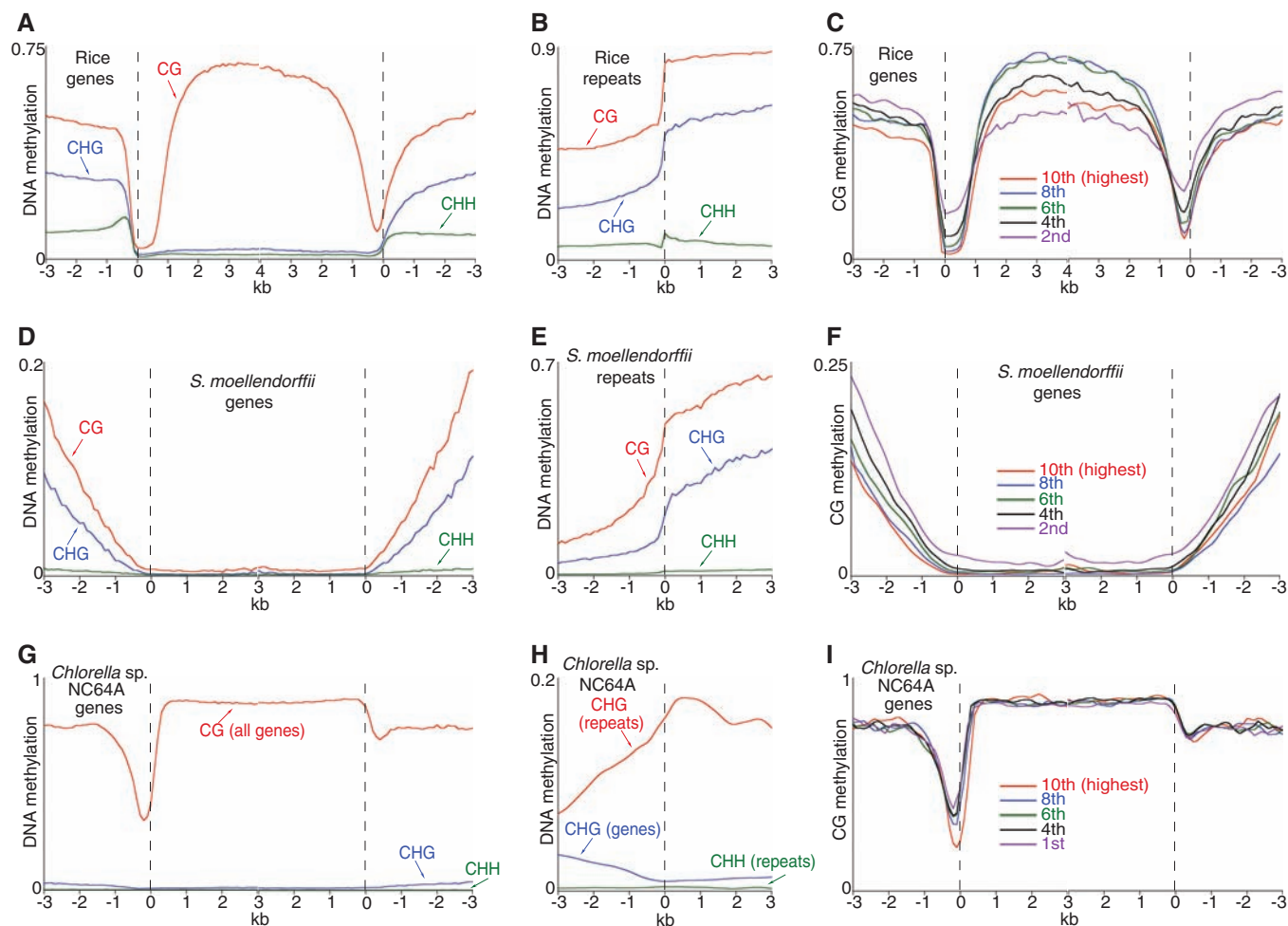
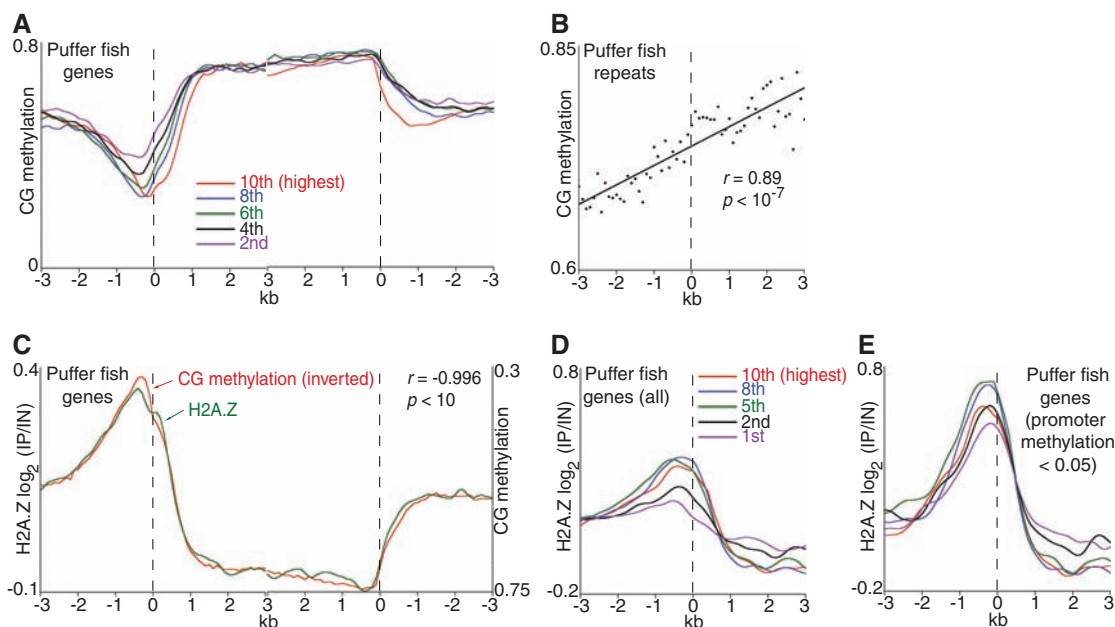


Fig. 1. DNA methylation in gene bodies and repeats of plants: (A to C) rice, (D to F) *S. moellendorffii*, and (G to I) *Chlorella* sp. NC64A. Genes or repeats were aligned at the 5' end (left dashed line) or the 3' end (right dashed line), and average methylation levels for each 100-bp interval are plotted. In (C), (F), and (I), genes were grouped into deciles by transcription, with five of the deciles shown for clarity. Only 5' alignments are shown in (B), (E), and (H).

Fig. 2. Puffer fish DNA methylation and H2A.Z are anti-correlated. (A to E) Genes or repeats were aligned as in Fig. 1, and average methylation levels [(A) to (C)] or H2A.Z enrichment [(C) to (E)] for each 100-bp interval are plotted. In (A), (D), and (E), genes were grouped into deciles by transcription. Only 5' alignments are shown in (B), (D), and (E). In (C) to (E), IP, immunoprecipitated; IN, input control.



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). Using chromatin immunoprecipitation coupled with deep sequencing, 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 among 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 million to 1 billion years ago (fig. S1) (11). Flour beetle (*Tribolium castaneum*) adults and *Drosophila melanogaster* embryos aged 0 to 3 hours 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 [*C. intestinalis*, honey bee (*Apis mellifera*), silk moth, 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 honey bees because of 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), whereas the antitransposon 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) more than 1 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 (1), demonstrates that this is an ancient methyltransferase family. We found 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, whereas active

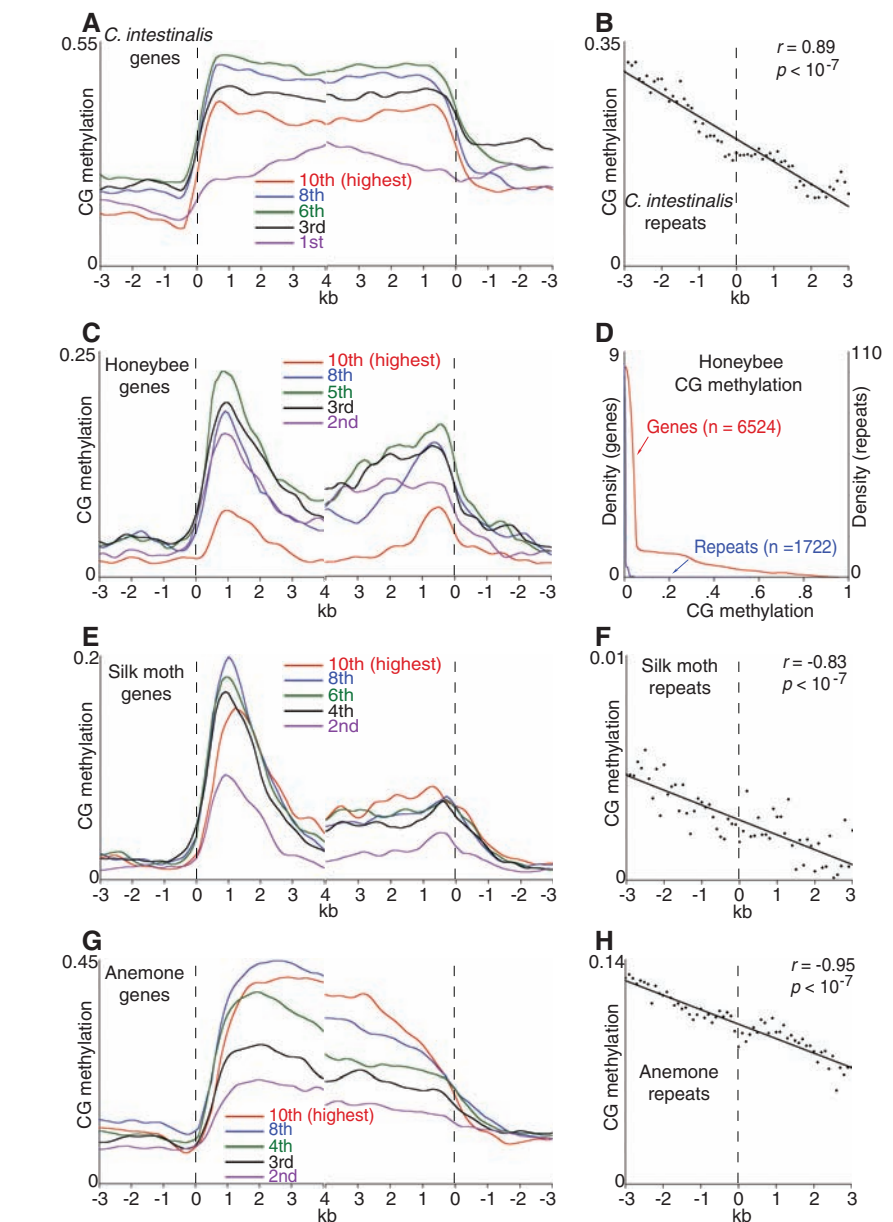


Fig. 3. Invertebrate gene bodies, but not TEs, are preferentially methylated: (A and B) *C. intestinalis*, (C and D) honey bee, (E and F) silk moth, (G and H) anemone. Genes or repeats were aligned as in Fig. 1. In (A), (C), (E), and (G), genes were grouped into deciles by transcription. Only 5' alignments are shown in (B), (F), and (H). (D) Kernel density plot, which has the effect of tracing the frequency distribution, of the average CG methylation within honey bee genes (red trace) and repeats (blue trace).

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 found 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 else-

where (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

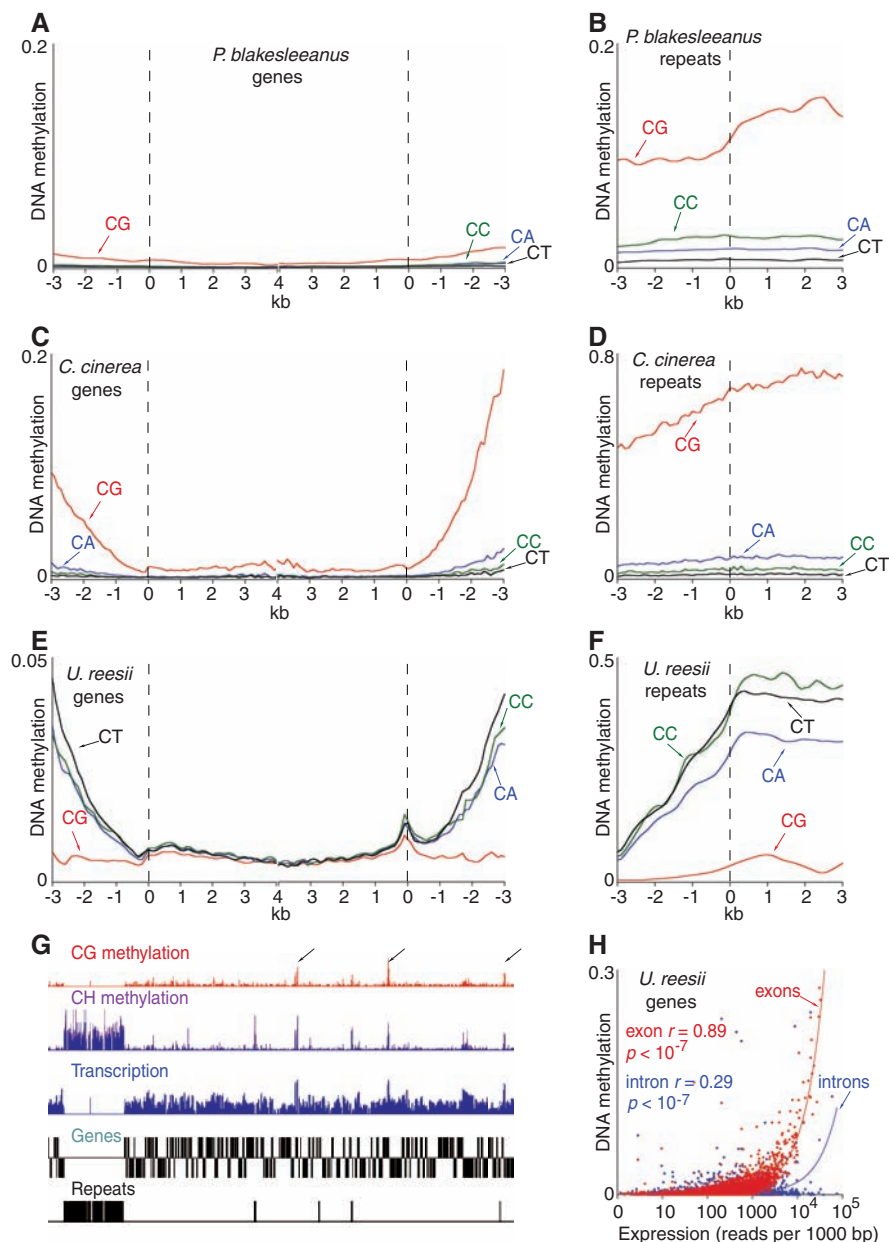


Fig. 4. DNA methylation patterns in fungi: (A and B) *P. blakesleeenanus*, (C and D) *C. cinerea*, (E and F) *U. reesii*. Genes or repeats were aligned as in Fig. 1. Only 5' alignments are shown in (B), (D), and (F). (G) A representative snapshot of *U. reesii* chromosome 1 (positions 3,931,000 to 4,424,000). Arrows highlight CG methylation peaks corresponding to highly transcribed genes. (H) Scatterplot of *U. reesii* DNA methylation in exons (red trace) and introns (blue trace) versus genic RNA levels.

by gene duplication (15), but the presence of a severe version of RIP in an organism separated from *N. crassa* by more than 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, whereas introns tend to be unmethylated (Fig. 4H and fig. S16). These features strongly argue that genic 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—transcription-influenced CG methylation of gene bodies and

extensive CG and non-CG methylation of TEs—resembled that of extant angiosperms. 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 did occur early in fungal evolution and later in several plant and animal lineages.

Note added in proof: Similar conclusions regarding the conservation of gene body methylation have been reached by Feng *et al.* (17).

References and Notes

1. M. G. Goll, T. H. Bestor, *Annu. Rev. Biochem.* **74**, 481 (2005).
2. I. R. Henderson, S. E. Jacobsen, *Nature* **447**, 418 (2007).
3. R. Feil, F. Berger, *Trends Genet.* **23**, 192 (2007).
4. R. Lister *et al.*, *Nature* **462**, 315 (2009).
5. X. Zhang *et al.*, *Cell* **126**, 1189 (2006).
6. D. Zilberman, M. Gehring, R. K. Tran, T. Ballinger, S. Henikoff, *Nat. Genet.* **39**, 61 (2006).
7. M. W. Simmen *et al.*, *Science* **283**, 1164 (1999).
8. C. V. Patel, K. P. Gopinathan, *Anal. Biochem.* **164**, 164 (1987).
9. M. M. Suzuki, A. Bird, *Nat. Rev. Genet.* **9**, 465 (2008).
10. See supporting material on Science Online.
11. S. B. Hedges, J. Dudley, S. Kumar, *Bioinformatics* **22**, 2971 (2006).
12. O. Jaillon *et al.*, *Nature* **431**, 946 (2004).
13. D. Zilberman, D. Coleman-Derr, T. Ballinger, S. Henikoff, *Nature* **456**, 125 (2008).
14. Z. A. Lewis *et al.*, *Genome Res.* **19**, 427 (2009).
15. J. E. Galagan, E. U. Selker, *Trends Genet.* **20**, 417 (2004).
16. T. H. Bestor, *Genetica* **107**, 289 (1999).
17. S. Feng *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 10173 (2010).
18. Acknowledgments are included in supporting material on Science Online. Sequencing data are deposited in Gene Expression Omnibus with accession number GSE19824.

Supporting Online Material

www.sciencemag.org/cgi/content/full/science.1186366/DC1
Materials and Methods
SOM Text
Figs. 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.

Genome-Wide Evolutionary Analysis of Eukaryotic DNA Methylation

Assaf Zemach, Ivy E. McDaniel, Pedro Silva and Daniel Zilberman

Science **328** (5980), 916-919.

DOI: 10.1126/science.1186366 originally published online April 15, 2010

Epigenetic Maps

Methylation of genomic DNA on cytosine bases provides critical epigenetic regulation of gene expression and is involved in silencing transposable elements (TEs) and repeated sequences, as well as regulating imprinted gene expression. **Zemach et al.** (p. 916, published online 15 April; see the Perspective by **Jeltsch**) analyzed DNA methylation in the genomes of five plants, five fungi, and seven animals by bisulfite sequencing. The data suggest that land plants and vertebrates, which have extensive DNA methylation, are under strong selective pressure to repress TEs, because of their sexual mode of reproduction. Unicellular animals and fungi that reproduce asexually are more likely to lose TE methylation. Although gene body methylation is evolutionarily ancient, it is also mutagenic, and so loss of this pathway has been relatively common and occurred early in fungal evolution and later in several plant and animal lineages.

ARTICLE TOOLS

<http://science.sciencemag.org/content/328/5980/916>

SUPPLEMENTARY MATERIALS

<http://science.sciencemag.org/content/suppl/2010/04/15/science.1186366.DC1>

RELATED CONTENT

<http://science.sciencemag.org/content/sci/328/5980/837.full>

REFERENCES

This article cites 15 articles, 2 of which you can access for free
<http://science.sciencemag.org/content/328/5980/916#BIBL>

PERMISSIONS

<http://www.sciencemag.org/help/reprints-and-permissions>

Use of this article is subject to the [Terms of Service](#)