



## Genome-Wide Demethylation of *Arabidopsis* Endosperm

Tzung-Fu Hsieh *et al.*

*Science* **324**, 1451 (2009);

DOI: 10.1126/science.1172417

---

*This copy is for your personal, non-commercial use only.*

---

If you wish to distribute this article to others, you can order high-quality copies for your colleagues, clients, or customers by [clicking here](#).

Permission to republish or repurpose articles or portions of articles can be obtained by following the guidelines [here](#).

**The following resources related to this article are available online at [www.sciencemag.org](http://www.sciencemag.org) (this information is current as of October 16, 2013):**

**Updated information and services**, including high-resolution figures, can be found in the online version of this article at:

<http://www.sciencemag.org/content/324/5933/1451.full.html>

**Supporting Online Material** can be found at:

<http://www.sciencemag.org/content/suppl/2009/06/11/324.5933.1451.DC1.html>

This article **cites 17 articles**, 4 of which can be accessed free:

<http://www.sciencemag.org/content/324/5933/1451.full.html#ref-list-1>

This article has been **cited by** 24 article(s) on the ISI Web of Science

This article has been **cited by** 53 articles hosted by HighWire Press; see:

<http://www.sciencemag.org/content/324/5933/1451.full.html#related-urls>

This article appears in the following **subject collections**:

Molecular Biology

[http://www.sciencemag.org/cgi/collection/molec\\_biol](http://www.sciencemag.org/cgi/collection/molec_biol)

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 gene dosage balance in the triploid endosperm.

#### References and Notes

- J. H. Huh, M. J. Bauer, T. F. Hsieh, R. L. Fischer, *Cell* **132**, 735 (2008).
- J. F. Gutierrez-Marcos *et al.*, *Nat. Genet.* **38**, 876 (2006).
- Materials and methods are available as supporting material on Science Online.

- D. Zilberman, M. Gehring, R. K. Tran, T. Ballinger, S. Henikoff, *Nat. Genet.* **39**, 61 (2007).
- X. Zhang *et al.*, *Cell* **126**, 1189 (2006).
- M. Lauria *et al.*, *Plant Cell* **16**, 510 (2004).
- M. Gehring *et al.*, *Cell* **124**, 495 (2006).
- T. Kinoshita *et al.*, *Science* **303**, 521 (2004).
- G. Makarevich, C. B. Villar, A. Erilova, C. Kohler, *J. Cell Sci.* **121**, 906 (2008).
- P. E. Jullien, T. Kinoshita, N. Ohad, F. Berger, *Plant Cell* **18**, 1360 (2006).
- S. W. Chan, I. R. Henderson, S. E. Jacobsen, *Nat. Rev. Genet.* **6**, 351 (2005).
- N. Buisine, H. Quesneville, V. Colot, *Genomics* **91**, 467 (2008).
- R. Lister *et al.*, *Cell* **133**, 523 (2008).
- J. Pennerman *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 6752 (2007).
- P. E. Jullien *et al.*, *PLoS Biol.* **6**, e194 (2008).
- R. C. Day, R. P. Herridge, B. A. Ambrose, R. C. Macknight, *Plant Physiol.* **148**, 1964 (2008).
- M. Nakamura *et al.*, *Plant Physiol.* **141**, 1363 (2006).

- D. P. Barlow, *Science* **260**, 309 (1993).
- R. K. Slotkin, R. Martienssen, *Nat. Rev. Genet.* **8**, 272 (2007).
- We thank the FHCRC Genomics Shared Resources for performing microarray hybridizations and Illumina sequencing, J. Henikoff for assistance with computational analyses, R. Deal and J. Cooper for comments on the manuscript, and D. Zilberman and R. Fischer for sharing unpublished data. M.G. is a HHMI Fellow of the Life Sciences Research Foundation. Illumina and NimbleGen data are deposited in the Genome Expression Omnibus (GEO) (accession number GSE14570).

#### Supporting Online Material

www.sciencemag.org/cgi/content/full/324/5933/1447/DC1

Materials and Methods

SOM Text

Figs. S1 to S9

Tables S1 to S9

References

29 January 2009; accepted 24 April 2009

10.1126/science.1171609

# 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 reinforces 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 inter-

ference (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, 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* nuclear 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, CHG 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 (2.8% 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 RNAi-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 discrete 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

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

\*These authors contributed equally to this work.

†To whom correspondence should be addressed. E-mail: rfischer@berkeley.edu (R.L.F.), daniel.zilberman@nature.berkeley.edu (D.Z.)

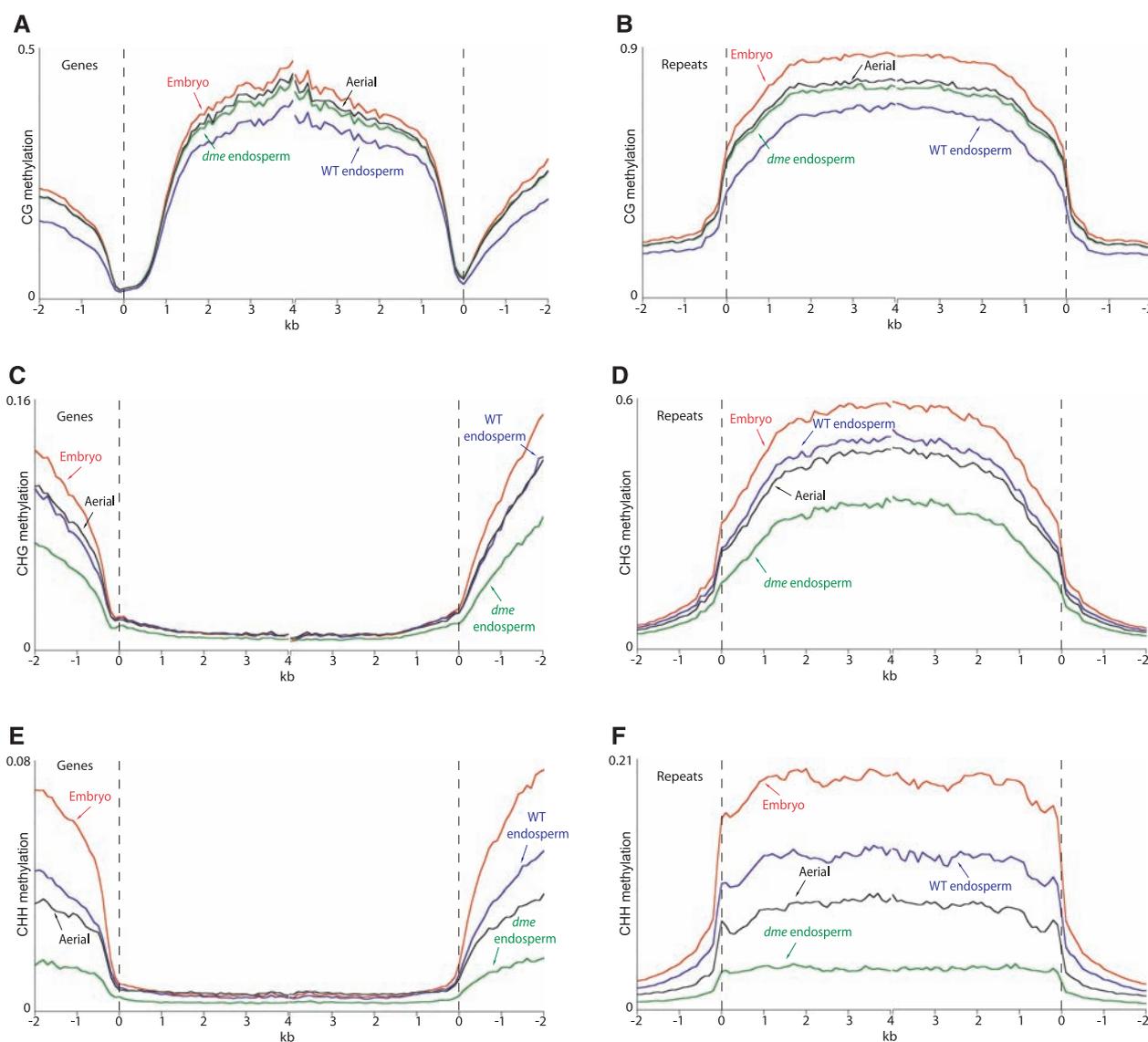
differences between embryo and endosperm. Notably, ~10% of identified loci were hypermethylated at CHG and CHH sites in the endosperm, compared with <1% hypermethylated at CG sites. Moreover, non-CG hypermethylated loci were strongly enriched in siRNAs (13) (Fig. 2A), further indicating that RNAi drives a substantial re-configuration of the seed methylation landscape.

To determine how methylation changes in the endosperm affect gene expression, we identified genes with reduced DNA methylation (at a cutoff of  $P < 1 \times 10^{-7}$ ) within 1 kb of either the 5' or 3' end and compared their gene expression between endosperm and embryo based on available microarray data (13) (table S4) (genes demethylated near both ends were analyzed in the 5' category). Genes exhibiting reduced methylation upstream of the start of transcription were preferentially

expressed in the endosperm to a modest but significant degree ( $P = 0.0005$ , Wilcoxon rank-sum test) (Fig. 2B), whereas genes demethylated near the 3' end did not show a significant change in expression ( $P = 0.33$ ). Reduced methylation of the maternal endosperm genome has been implicated in allele-specific expression of all five known *Arabidopsis* imprinted genes (3, 5–8), so genes with reduced methylation and greater expression in endosperm than embryo are potentially imprinted.

To visualize methylation differences between tissues, we plotted the distribution density of windows for wild-type endosperm subtracted from embryo (Fig. 3, A to C, blue trace), *dme* endosperm subtracted from embryo (Fig. 3, A to C, red trace), and aerial tissues subtracted from embryo (fig. S4), showing only those windows that were methylated in at least one of the tissues being compared (13).

We also aligned all *Arabidopsis* annotated genes, which include some pseudogenes and transposable elements, at their 5' ends, stacked them from the top of chromosome 1 to the bottom of chromosome 5, and displayed fractional embryo methylation (left panels of Fig. 3, D and E) and the difference between embryo and wild-type endosperm methylation (right panels of Fig. 3, D and E) as heat maps. We performed a similar analysis for annotated transposons and other repeats (fig. S5). Virtually all sequences methylated in embryo in the CG context were less methylated in the endosperm (Fig. 3A). Gene bodies, gene adjacent sequences, and transposable elements were all similarly demethylated (Fig. 1, A and B, Fig. 3D, and figs. S5 and S6), with transposons demethylated to a somewhat greater extent than genes and shorter transposons on average demethylated more than longer ones (fig. S6).



**Fig. 1.** Profiles of DNA methylation in embryo, wild-type endosperm, and *dme* endosperm. (A to F) TAIR8-annotated genes [(A), (C), and (E)] or transposons [(B), (D), and (F)] were aligned at the 5' end (left panel) or the 3' end (right panel), and average methylation levels for each 100-bp interval are plotted from 2 kb away from the gene (negative numbers) to 4 kb into the

gene (positive numbers). Embryo methylation is represented by the red trace, wild-type (WT) endosperm by the blue trace, *dme* endosperm by the green trace, and aerial tissues by the black trace. The dashed line at zero represents the point of alignment. CG methylation is shown in (A) and (B), CHG in (C) and (D), CHH in (E) and (F).

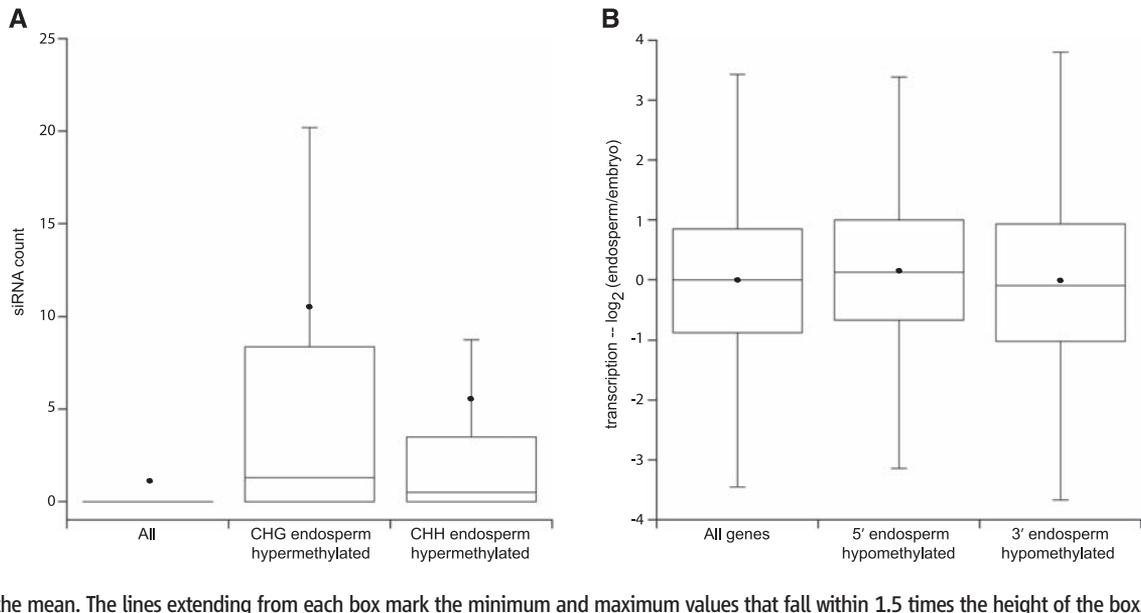
CHG and CHH methylation of most sequences was also higher in embryo (Fig. 1, C to F; Fig. 3, B, C, and E; and fig. S5). The *dme* mutation uniformly restored CG methylation, while uniformly reducing CHG and CHH methylation (Fig. 1 and Fig. 3, A to C). Methylation in all contexts was higher in embryo than in aerial tissues (Fig. 1 and fig. S4), with particularly extensive CHH hypermethylation: We identified 10,858 loci covering 21.88 million bp

with an absolute change in CHH methylation of at least 5% ( $P < 0.0001$ , Fisher's exact test), 96.8% of which (10,510) were more methylated in embryo (table S3). Virtually, genome-wide CG demethylation of the maternal endosperm genome is thus accompanied by similarly extensive CHH hypermethylation in the embryo.

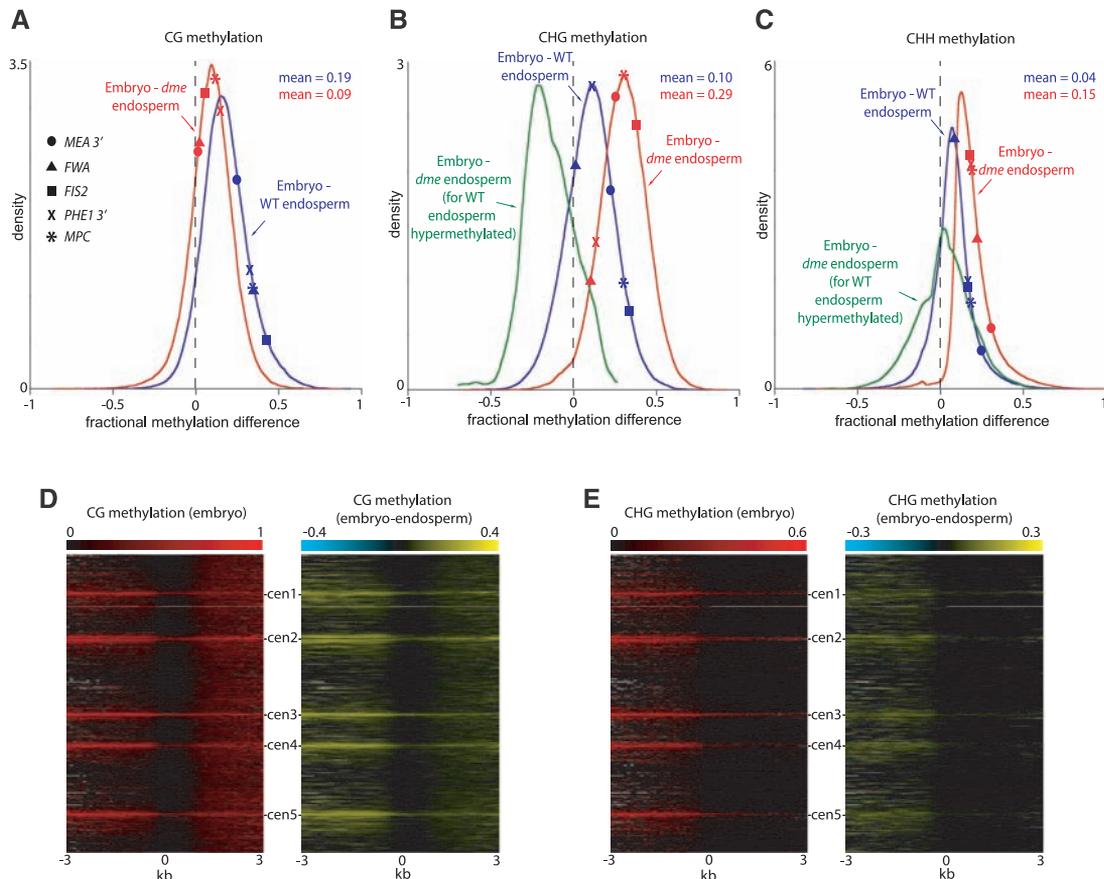
We investigated the source of the substantial non-CG hypermethylation in wild-type endosperm

compared with the embryo (table S3) by examining methylation differences between embryo and *dme* endosperm of sequences that were more methylated in wild-type endosperm than in embryo (Fig. 3, B and C, green trace). If endosperm hypermethylation were random, we would expect to see no correlation between hypermethylation in wild-type and *dme* endosperm. Our analysis showed that for both CHG and CHH contexts, loci hypermethylated in

**Fig. 2.** Associations between endosperm methylation, siRNAs, and expression. **(A)** Box plots showing siRNA abundance within 50-bp windows in the entire *Arabidopsis* genome (All) and in sequences hypermethylated in WT endosperm compared with the embryo in the CHG and CHH contexts. **(B)** Box plots showing differences in gene expression between embryo and endosperm for all genes ( $n = 21,021$ ), genes with 5' hypomethylation in endosperm ( $n = 1097$ ), and genes with 3' hypomethylation in endosperm ( $n = 505$ ). Each box encloses the middle 50% of the distribution, with the horizontal line marking the median and the dot marking the mean. The lines extending from each box mark the minimum and maximum values that fall within 1.5 times the height of the box.



**Fig. 3.** Genome-wide demethylation of endosperm. **(A to C)** Kernel density plots of the differences between embryo and WT endosperm methylation (blue trace) and the differences between embryo and *dme* endosperm methylation (red trace). The green trace in (B) and (C) represents methylation differences between embryo and *dme* endosperm for windows with absolute fractional methylation increase in WT endosperm compared with embryo of at least 0.4 in the CHG context (B) ( $n = 135$ ) or at least 0.2 in the CHH context (C) ( $n = 6168$ ). Methylation differences for the 3' *MEA* repeats, *FWA*, *FIS2*, *PHE1*, and *MPC* are indicated; specifics are listed in table S2. **(D and E)** All TAIR8-annotated genes (28,244) were aligned at the 5' end and stacked from the top of chromosome 1 to the bottom of chromosome 5. Embryo methylation is displayed as a heat map in the left panel, differences between embryo and WT endosperm in the right panel. CG methylation is shown in (D), CHG in (E).



wild-type endosperm had a strong tendency to be hypermethylated in *dme* endosperm as well (Fig. 3, B and C, green trace), despite the overall reduction of non-CG methylation caused by the *dme* mutation. Endosperm hypermethylation is thus a highly specific, RNAi-targeted process.

We calculated methylation levels of sequences either known or strongly inferred to cause imprinted expression of five *Arabidopsis* genes (3, 5–8): the *MEA* 3' repeats, the *FWA* promoter and start of transcription, the *FIS2* promoter, the *PHE1* 3' repeats, and the *MPC* gene and flanking regions (Fig. 3, A to C, and table S2). *MEA* methylation was reduced from 88% CG, 39% CHG, and 42% CHH in embryo to 63% CG, 16% CHG, and 17% CHH in wild-type endosperm. *MEA* CG methylation was restored to 87% in *dme* endosperm, whereas CHG (13%) and CHH (8%) methylation was further reduced. The other four genes behaved similarly (Fig. 3, A to C, and table S2), in line with the overall trends. Imprinted genes are thus not exceptional sequences specifically targeted for demethylation in the central cell but rather part of a nearly universal process that reshapes DNA methylation of the entire maternal genome in the endosperm (14). Imprinted expression of genes regulated by allele-specific DNA methylation could potentially arise whenever a transposable element insertion or a local duplication near a gene's regulatory sequences induces methylation and gene silencing in other tissues, including the paternal endosperm genome.

Genomic imprinting is a fast-evolving process driven by genetic conflict between parents (1). In mammals, which exhibit virtually global CG methylation (15), imprinting is orchestrated in part by differential methylation of specific sequences in the gametes (16). *Arabidopsis*, which targets methylation primarily to transposable elements (9), apparently adapted a radical implementation of imprinting by partially suspending its transposon suppression system and globally demethylating central cell DNA, resulting in a hypomethylated maternal endosperm genome. Because the endosperm genome is not transmitted to the next generation, transient transposon activation is likely to carry a fairly low cost, especially in an organism with few functional transposons, like *Arabidopsis*. Transposon activation and siRNA accumulation in the central cell might actually contribute to enhanced methylation and silencing of elements in the egg cell (and later the embryo) through siRNA transport (17), which could be the original selective force driving the evolution of central cell demethylation. An analogous mechanism has recently been proposed to operate between the vegetative and reproductive cells of pollen (18). It is an open question whether other plants, particularly those with more aggressive transposable elements, have adopted a similar strategy.

#### References and Notes

1. R. Feil, F. Berger, *Trends Genet.* **23**, 192 (2007).
2. Y. Choi *et al.*, *Cell* **110**, 33 (2002).

3. M. Gehring *et al.*, *Cell* **124**, 495 (2006).
4. J. H. Huh, M. J. Bauer, T.-F. Hsieh, R. L. Fischer, *Cell* **132**, 735 (2008).
5. P. E. Jullien, T. Kinoshita, N. Ohad, F. Berger, *Plant Cell* **18**, 1360 (2006).
6. T. Kinoshita *et al.*, *Science* **303**, 521 (2004).
7. G. Makarevich, C. B. R. Villar, A. Erilova, C. Kohler, *J. Cell Sci.* **121**, 906 (2008).
8. S. Tiwari *et al.*, *Plant Cell* **20**, 2387 (2008).
9. I. R. Henderson, S. E. Jacobsen, *Nature* **447**, 418 (2007).
10. S. J. Cokus *et al.*, *Nature* **452**, 215 (2008).
11. R. Lister *et al.*, *Cell* **133**, 523 (2008).
12. A. Meissner *et al.*, *Nature* **454**, 766 (2008).
13. Materials and methods are available as supporting material on Science Online.
14. P. E. Jullien *et al.*, *PLoS Biol.* **6**, e194 (2008).
15. M. G. Goll, T. H. Bestor, *Annu. Rev. Biochem.* **74**, 481 (2005).
16. A. Munshi, S. Duvvuri, *J. Genet. Genomics* **34**, 93 (2007).
17. Y. Z. Han, B. Q. Huang, S. Y. Zee, M. Yuan, *Planta* **211**, 158 (2000).
18. R. K. Slotkin *et al.*, *Cell* **136**, 461 (2009).
19. We thank L. Tonkin for performing Illumina sequencing, J. Shin for gene annotation, and S. Henikoff for sharing unpublished data. This work was partially funded by a NIH grant (GM69415) to R.L.F. A.Z. is a fellow of the Jane Coffin Childs Memorial Fund for Medical Research. Sequencing data are deposited in GEO with accession number GSE15922.

#### Supporting Online Material

www.sciencemag.org/cgi/content/full/324/5933/1451/DC1

Materials and Methods

Figs. S1 to S6

References

17 February 2009; accepted 1 May 2009

10.1126/science.1172417

## Hyper-Recombination, Diversity, and Antibiotic Resistance in *Pneumococcus*

William Paul Hanage,<sup>1\*</sup> Christophe Fraser,<sup>1</sup> Jing Tang,<sup>2</sup> Thomas Richard Connor,<sup>1</sup> Jukka Corander<sup>2</sup>

*Streptococcus pneumoniae* is a pathogen of global importance that frequently transfers genetic material between strains and on occasion across species boundaries. In an analysis of 1930 pneumococcal genotypes from six housekeeping genes and 94 genotypes from related species, we identified mosaic genotypes representing admixture between populations and found that these were significantly associated with resistance to several classes of antibiotics. We hypothesize that these observations result from a history of hyper-recombination, which means that these strains are more likely to acquire both divergent genetic material and resistance determinants. This could have consequences for the reemergence of drug resistance after pneumococcal vaccination and also for our understanding of diversification and speciation in recombinogenic bacteria.

Many bacteria undergo homologous recombination, in which short tracts of DNA in the recipient are replaced by the corresponding tract from a donor strain, resulting in a mosaic of DNA from different ancestors (1). Although this occurs mainly within species and declines markedly with increasing sequence divergence between donor and recipient (2), occasional gene transfers between species

do occur. Such events have the potential to introduce new phenotypes, such as virulence or antibiotic resistance, into a new genetic background that may or may not be the same as the species of the donor strain (3–7) and may have considerable impacts on bacterial evolution and human health.

One group in which homologous recombination is frequent is the mitis group streptococci. This includes the major human pathogen *Streptococcus*

*pneumoniae*, the pneumococcus, which is responsible for at least 1 million deaths per year worldwide (8). The closely related species *S. oralis*, *S. mitis*, and *S. pseudopneumoniae* (among others) have a history of taxonomic confusion, which may be partly explained by genetic diversity within the mitis group (9, 10). Moreover, rare but important events have led to the acquisition of antibiotic resistance by pneumococcus as a result of the transfer of resistance determinants across species boundaries (4, 5). The high rates of recombination within the species have the potential to shuffle resistance determinants among pneumococcal genotypes. It is not known whether or not recombination, either at resistance loci or housekeeping genes, is equally likely for all members of the species or whether some strains are more likely to be involved in this process.

Although a vaccine is available for 7 of the more than 90 pneumococcal serotypes, this has not eliminated pneumococcal disease because the nonvaccine serotypes derive an ecological advantage from the removal of their competitors and have been increasing in carriage prevalence (11) and, concomitantly, in disease (12). Alongside

<sup>1</sup>Department of Infectious Disease Epidemiology, Imperial College London, Norfolk Place, London W2 1PG, UK. <sup>2</sup>Department of Mathematics, Åbo Akademi, FI-20500, Turku, Finland.

\*To whom correspondence should be addressed. E-mail: w.hanage@imperial.ac.uk