proteins were removed by washing the beads six times with high ionic strength buffer. Bound protein was eluted with electrophoresis sample buffer containing 1% SDS, 2.5% 2-mercaptoethanol, 0.063 M Tris-HCl (pH 6.8) and 15% glycerol, and resolved by 6% or 8% SDS–PAGE. Western blot analysis was carried out with antisera against mPER2 (ref. 1), NPAS2 (ref. 10) and GAPDH (Sigma). Bound antibodies were detected by Supersignal West Pico Chemiluminescent Substrate (Pierce). Cyanocobalamin (335 μM) was added to the binding solution for competition studies.

Received 4 March; accepted 3 June 2004; doi:10.1038/nature02724.

1. Zheng, B. *et al.* Nonredundant roles of the *mPer1* and *mPer2* genes in the mammalian circadian clock. *Cell* **105,** 683–941 (2001).
2. Panda, S. *et al.* Coordinated transcription of key pathways in the mouse by the circadian clock. *Cell* **109,** 307–320 (2002).
3. Dioum, E. M. *et al.* NPAS2: a gas-responsive transcription factor. *Science* **298,** 2385–2387 (2002).
4. Gilles-Gonzalez, M. A. & Gonzalez, G. Signal transduction by haem-containing PAS-domain proteins. *J. Appl. Physiol.* **96,** 774–783 (2004).
5. Zheng, B. *et al.* The *mPer2* gene encodes a functional component of the mammalian circadian clock. *Nature* **400,** 169–173 (1999).
6. Yamaguchi, S. *et al.* Role of DBP in the circadian oscillatory mechanism. *Mol. Cell. Biol.* **13,** 4773–4781 (2000).
7. Rutter, J., Reick, M., Wu, L. C. & McKnight, S. L. Regulation of clock and NPAS2 DNA binding by the redox state of NAD cofactors. *Science* **293,** 510–514 (2001).
8. Yamaguchi, S. *et al.* The 5′ upstream region of *mPer1* gene contains two promoters and is responsible for circadian oscillation. *Curr. Biol.* **10,** 873–876 (2000).
9. Travnickova-Bendova, Z., Cermakian, N., Reppert, S. M. & Sassone-Corsi, P. Bimodal regulation of *mPeriod* promoters by CREB-dependent signaling and CLOCK/BMAL1 activity. *Proc. Natl Acad. Sci. USA* **99,** 7728–7733 (2002).
10. Garcia, J. A. *et al.* Impaired cued and contextual memory in NPAS2-deficient mice. *Science* **288,** 2226–2230 (2000).
11. Albrecht, U., Sun, Z. S., Eichele, G. & Lee, C. C. A differential response of two putative mammalian circadian regulators, mPer1 and mPer2, to light. *Cell* **91,** 1055–1064 (1997).
12. Tsutsui, K. & Mueller, G. C. Affinity chromatography of haem-binding proteins: an improved method for the synthesis of hemin–agarose. *Anal. Biochem.* **121,** 244–250 (1982).
13. Hashimoto, S. *et al.* Vitamin B12 enhances the phase-response of circadian melatonin rhythm to a single bright light exposure in humans. *Neurosci. Lett.* **220,** 129–132 (1996).
14. Nakamura, T., Uchida, K., Moriguchi, Y., Okamoto, N. & Morita, Y. Transient fluctuation of serum melatonin rhythm is suppressed centrally by vitamin B12. *Chronobiol. Int.* **14,** 549–560 (1997).
15. Hardin, P. E., Hall, J. C. & Rosbash, M. Feedback of the *Drosophila* period gene product on circadian cycling of its messenger RNA levels. *Nature* **343,** 536–540 (1990).
16. Kume, K. *et al.* mCRY1 and mCRY2 are essential components of the negative limb of the circadian clock feedback loop. *Cell* **98,** 193–205 (1999).
17. Shearman, L. P. *et al.* Interacting molecular loops in the mammalian circadian clock. *Science* **288,** 1013–1039 (2000).
18. Tunctan, B. *et al.* Circadian variation of nitric oxide synthase activity in mouse tissue. *Chronobiol. Int.* **19,** 393–404 (2002).
19. Artinian, L. R., Ding, J. M. & Gillette, M. U. Carbon monoxide and nitric oxide: interacting messengers in muscarinic signaling to the brain's circadian clock. *Exp. Neurol.* **171,** 293–300 (2001).
20. Rubio, M. F., Agostino, P. V., Ferreyra, G. A. & Golombek, D. A. Circadian haem oxygenase activity in the hamster suprachiasmatic nuclei. *Neurosci. Lett.* **353,** 9–12 (2003).
21. Fu, L., Pelicano, H., Liu, J., Huang, P. & Lee, C. C. The circadian gene *Period2* plays an important role in tumor suppression and DNA damage response *in vivo*. *Cell* **111,** 41–50 (2002).
22. Matsuo, T. *et al.* Control mechanism of the circadian clock for timing of cell division *in vivo*. *Science* **302,** 255–259 (2003).
23. Zhu, Y., Lee, H. C. & Zhang, L. An examination of haem action in gene expression: haem and haem deficiency affect the expression of diverse genes in erythroid k562 and neuronal PC12 cells. *DNA Cell Biol.* **21,** 333–346 (2002).
24. McLean, G. R. *et al.* Cobalamin analogues modulate the growth of leukemia cells *in vitro*. *Cancer Res.* **57,** 4015–4022 (1997).
25. Levitman, M. K. *et al.* Antitumor effect of combined treatment with ionizing radiation and vitamin B12–C complex. *Radiat. Biol. Radioecol.* **42,** 511–514 (2002).
26. Scagliotti, G. V. *et al.* Phase II study of pemetrexed with and without folic acid and vitamin B12 as front-line therapy in malignant pleural mesothelioma. *J. Clin. Oncol.* **21,** 1556–1561 (2003).
27. Norman, P. Pemetrexed disodium (Eli Lilly). *Curr. Opin. Invest. Drugs* **2,** 1611–1622 (2001).
28. Lincoln, D. W. II, Hrushesky, W. J. & Wood, P. A. Circadian organization of thymidylate synthase activity in normal tissues: a possible basis for 5-fluorouracil chronotherapeutic advantage. *Int. J. Cancer.* **88,** 479–485 (2000).
29. Freedman, M. L., Geraghty, M. & Rosman, J. Hemin control of globin synthesis. Isolation of a hemin-reversible translational repressor from human mature erythrocytes. *J. Biol. Chem.* **249,** 7290–7294 (1974).

Supplementary Information accompanies the paper on **www.nature.com/nature**.

Acknowledgements We thank S. L. McKnight for the *Npas2*$^{m/m}$ mice and NPAS2 antibody; M. Reick for technical comments; H. Okamura for the *mPer1–luc* and *Clock* expression constructs, P. Sassone-Corsi for the *mPer2–luc* expression plasmid; Z. Sun for suggesting the use of haem–agarose; and P. Hastings, R. Kellems and J. Lever for comments. This work was supported, in part, by a grant from the NIH (to C.C.L.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to C.C.L. (cheng.c.lee@uth.tmc.edu).

Role of transposable elements in heterochromatin and epigenetic control

Zachary Lippman[1]*, Anne-Valérie Gendrel[2]*, Michael Black[3]†, Matthew W. Vaughn[1], Neilay Dedhia[1], W. Richard McCombie[1], Kimberly Lavine[1], Vivek Mittal[1], Bruce May[1], Kristin D. Kasschau[4], James C. Carrington[4], Rebecca W. Doerge[3], Vincent Colot[2] & Rob Martienssen[1]

[1]*Watson School of Biological Sciences and Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA*
[2]*Unité de Recherche en Génomique Végétale (URGV), INRA/CNRS/UEVE, 2 Rue Gaston Crémieux, 91057 Evry Cedex, France*
[3]*Department of Statistics, Purdue University, West Lafayette, Indiana 47907, USA*
[4]*Center for Gene Research and Biotechnology, Oregon State University, Corvallis, Oregon 97330, USA*

* These authors contributed equally to this work
† Present address: Department of Statistics, The University of Auckland, Private Bag 92019, Auckland, New Zealand

Heterochromatin has been defined as deeply staining chromosomal material that remains condensed in interphase, whereas euchromatin undergoes de-condensation[1]. Heterochromatin is found near centromeres and telomeres, but interstitial sites of heterochromatin (knobs) are common in plant genomes and were first described in maize[2]. These regions are repetitive and late-replicating[3]. In *Drosophila*, heterochromatin influences gene expression, a heterochromatin phenomenon called position effect variegation[4]. Similarities between position effect variegation in *Drosophila* and gene silencing in maize mediated by "controlling elements" (that is, transposable elements) led in part to the proposal that heterochromatin is composed of transposable elements, and that such elements scattered throughout the genome might regulate development[2]. Using microarray analysis, we show that heterochromatin in *Arabidopsis* is determined by transposable elements and related tandem repeats, under the control of the chromatin remodelling ATPase DDM1 (Decrease in DNA Methylation 1). Small interfering RNAs (siRNAs) correspond to these sequences, suggesting a role in guiding DDM1. We also show that transposable elements can regulate genes epigenetically, but only when inserted within or very close to them. This probably accounts for the regulation by DDM1 and the DNA methyltransferase MET1 of the euchromatic, imprinted gene *FWA*, as its promoter is provided by transposable-element-derived tandem repeats that are associated with siRNAs.

Arabidopsis heterochromatin is organized into chromocentres in interphase nuclei[5]. It is localized mostly to the pericentromeric and nucleolar organizing regions (NOR), and two domains of interstitial heterochromatin (knobs) on chromosomes 4 (*hk4S*) and 5 have been completely sequenced[5–7]. These regions are rich in transposable elements, but in each case the most conspicuous feature is a tandem array of ~2-kilobase (kb) satellite repeats. The arrays are unrelated and not found elsewhere, although they have limited homology with CACTA and *Mutator-like* (MULE) DNA transposons, respectively. *hk4S* is part of a duplicated segment found on the long arm of chromosome 4 (ref. 7) and encompasses 8 of the 33 genes found in conserved order and orientation (Fig. 1). However, although the long arm segment is virtually free of transposable elements, the heterochromatic segment is interrupted by at least 34 retrotransposons and 40 DNA class transposons[8], often inserted into one another. All but one of these insertions arose after the duplication event, indicating that heterochromatin was derived

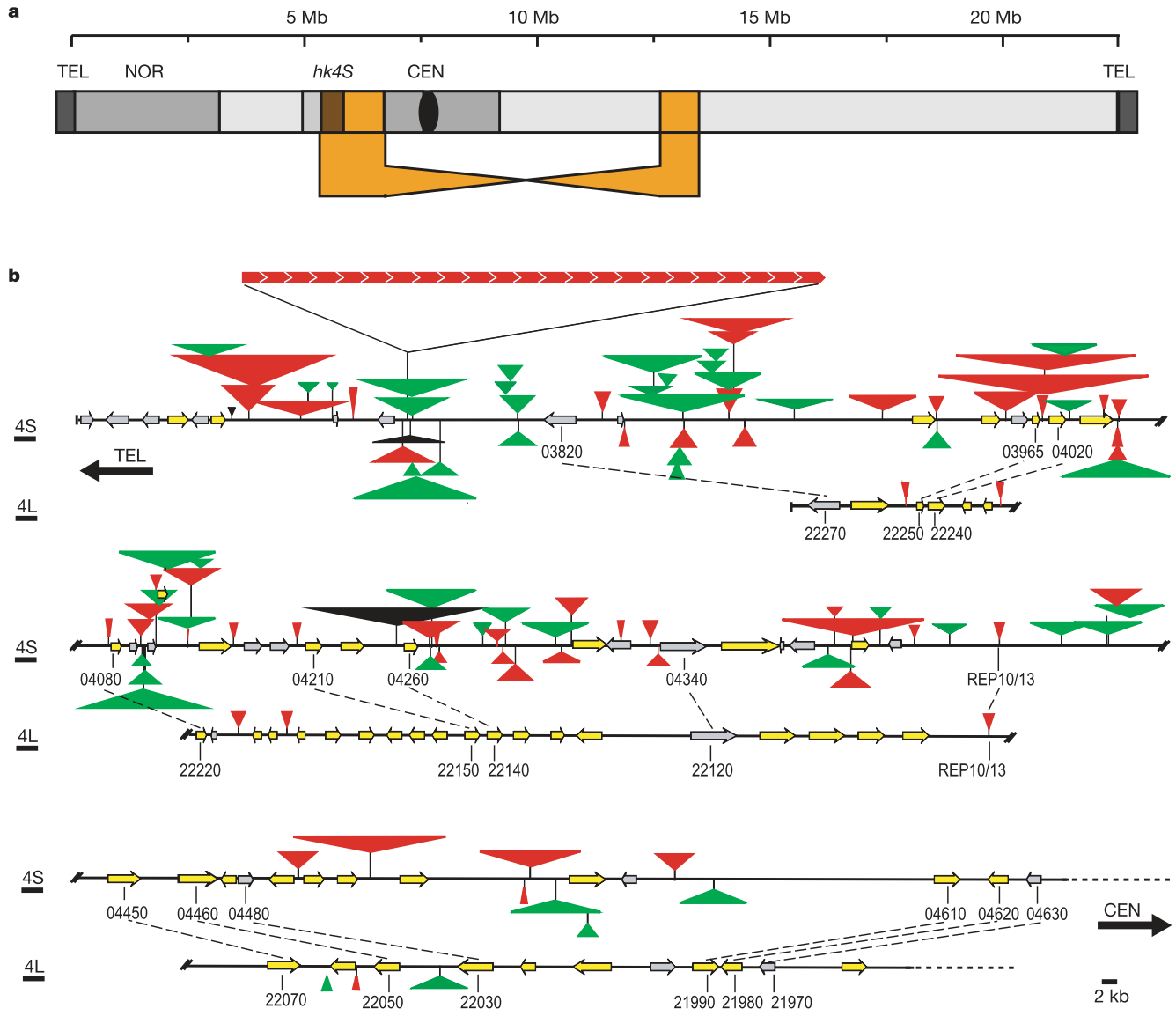

Figure 1 The heterochromatic knob (*hk4S*) on chromosome 4. **a**, The chromosomal position of the segmental duplication between 4S and 4L (orange) is indicated relative to heterochromatin (grey), and includes part of *hk4S* (brown). The segments are inverted around the centromere (CEN). NOR, nucleolar organizing region; TEL, telomere. **b**, Genome organization of the segmentally duplicated regions. Transposable elements are shown as nested insertions between known genes (yellow arrows) and hypothetical genes (grey arrows), and arose after the region was duplicated. DNA transposons and retrotransposons are indicated by red and green triangles, respectively. Black triangles represent complex transposable element insertions that we were unable to resolve. The orientation of each region relative to the telomere and centromere is shown. Homologous genes between the duplicated regions are in collinear order and are indicated by locus designators starting with At4g03820 (03820); they are joined by dashed lines.

Figure 2 Expression and chromatin profiling of wild type and *ddm1* using genomic tiling microarrays. **a**, The entire 1.5-Mb region surrounding *hk4S*. The region between 1,200–2,670 kb from the NOR at the tip of chromosome 4 is displayed along with coordinates (in kb). *hk4S* lies between 1,600–2,330 kb (black bar). Relative transcription levels (mRNA) are indicated for wild-type (green) and *ddm1* mutant seedlings (red). Annotation includes four tracks (from top to bottom): 1, ORFs; 2, repeats; 3, gene trap insertions; and 4, small RNA. ORFs (TIGR v.2.0) are annotated as known genes (yellow), hypothetical genes (grey), retrotransposons (green) and DNA transposons (red) (track 1). Repeats are from RepBase[8] with long terminal repeats (LTRs) indicated by pink boxes (see Methods) (track 2). Gene trap insertions are indicated by blue lines (track 3); whereas small RNA matches are indicted by black lines (track 4). DNA methylation (5mC), histone H3 methylation of lysine 9 (mK9), and histone H3 methylation of lysine 4 (mK4) that was significantly above or below the average level found in euchromatic features are highlighted in brown (5mC), blue (mK9) and green (mK4), respectively, for wild type (WT) and *ddm1* (Methods). Those features with euchromatic levels of each ▶ modification are coloured grey. Blank tiles reflect failure to amplify and print DNA. Examples of gene islands are highlighted by boxes at 1,770, 1,865, 1,930, 2,020 and 2,115 kb. These regions are expressed equally in *ddm1* and wild type, and have reduced 5mC and mK9. The BAC T27D20, which decorates a euchromatic loop in interphase cells[5], spans the region from 2,001 to 2,081 kb. **b**, Epigenetic activation of heterochromatin. Expression profiling of *ddm1* mutants (top) is compared to *ddm1*/+ plants (bottom) after backcrosses, compared with wild-type plants in each case. Similar profiles indicate that epigenetically activated heterochromatin is inherited. The *ATGP1* element at 1,620 kb is silenced in this backcross. **c**, Genes are insulated in gene islands. A close-up of the gene island at 1,930 kb and specific PCR amplification of reverse-transcribed cDNA (RT–PCR) for select genes and transposable elements (+). Mock RT–PCR was performed without reverse transcriptase (−). LINE, long interspersed nuclear element.

from euchromatin by the insertion of transposable elements (Fig. 1b).

Typically, euchromatin is associated with acetylated histones and with histone H3 di-methylation on lysine 4 (H3mK4), whereas heterochromatin is associated with histone H3 di-methylation on lysine 9 (H3mK9)[9,10]. In many eukaryotes, heterochromatic DNA is also heavily methylated[11]. Mutations in the *Arabidopsis* chromatin remodelling gene *DDM1* affect both DNA and histone H3 methylation, perhaps because chromatin-modifying enzymes act in multiprotein complexes[12]. *DDM1* is conserved between yeast, animals and plants, and mutants have comparable phenotypes in the mouse[13].

To determine the sequences responsible for heterochromatic modifications in *Arabidopsis*, we profiled histone and DNA modifications in *hk4S*, as well as expression, using DNA microarrays. Silent heterochromatic elements are poorly represented on commercial microarrays; therefore, we took a tiling approach by amplifying sequential 1-kb segments over a 1.5-megabase (Mb) region centred on *hk4S*. In addition, 36 unlinked 10-kb regions were included, either as controls or because of their potential regulation by *DDM1*. Microarrays were printed on glass slides and hybridized with labelled complementary DNA, or with labelled genomic DNA that had been enriched for sequences recovered by chromatin immunoprecipitation (ChIP) with antibodies raised against H3mK4 or H3mK9 (Supplementary Information), or depleted of methylated sequences (Supplementary Fig. S1). Control hybridizations with total genomic DNA allowed DNA methylation and histone H3 methylation to be quantified for each feature on the array.

A linear model approach was used to estimate technical and biological effects, allowing statistical detection of features significantly enriched for cytosine methylation and histone modifications, and to detect differential gene expression (see Methods). Individual results were validated by polymerase chain reaction (PCR) with specific primer pairs to account for cross-hybridization with closely related sequences elsewhere in the genome (Supplementary Fig. S4). The microarray data were displayed graphically in the context of genomic annotation (Fig. 2) using a customized version of Generic Genome Browser (http://chromatin.cshl.org/ddm1/).

As expected, sequences from *hk4S* had high levels of H3mK9 and DNA methylation, but were significantly depleted of H3mK4 relative to the surrounding euchromatin (Fig. 2a). Notably, individual features associated with high levels of H3mK9 were almost always methylated, indicating a genome-wide relationship between these heterochromatic marks. Moreover, H3mK9 and DNA methylation were not distributed uniformly, but coincided with transposable elements and related repeats. DNA methylation was markedly reduced in *ddm1* mutants, whereas H3mK9 was lost from heterochromatin. Erasure of heterochromatic marks was accompanied by a rather uniform replacement with H3mK4 to levels typical of euchromatin. H3mK9 and H3mK4 are both retained in *ddm1* mutants (ref. 14), suggesting that they are redistributed rather than lost, consistent with a role for DDM1 in heterochromatic histone variant exchange[10].

Transposable elements were associated with heterochromatic modification both within and outside *hk4S*. For example, *VANDAL2* (At4g03300) and *ATENSPM4* (At4g03310) are inserted in euchromatin and are associated with H3mK9 and DNA methylation (at 1,450 and 1,459 kb, respectively, in Fig. 2a). Thus, these elements constitute 'cryptic' heterochromatin, defined biochemically but not cytologically. In wild type, transposable elements were silent or expressed at low levels, but in *ddm1* more than half were strongly expressed (Fig. 3). Often, transposable elements that were disrupted by other transposable elements permanently lost their activity (Supplementary Fig. S3). Expressed transposable elements were inserted more recently and included both low- and high-copy elements such as CACTA transposons and Athila gypsy-like retrotransposons, respectively. Thus copy number did not account for high levels of expression in *ddm1*. In wild type, CACTA and gypsy-like transposable elements were the most uniformly methylated and associated with H3mK9 (Supplementary Fig. S3 and Table S1).

ddm1-2 is strictly recessive, but, when crossed to wild-type, hypomethylated centromeric and rDNA sequences were not remethylated[15]. To investigate whether activated heterochromatin is inherited in subsequent generations, *ddm1* was out-crossed to wild type, and messenger RNA was extracted from the resulting *ddm1*/+ plants. Expression profiles were similar to *ddm1*, indicating that transposable elements activated in *ddm1* plants retained their activity in *ddm1*/+ plants even though *DDM1* function was restored (Fig. 2b). The gypsy class element *ATGP1* was exceptional in that it was partially silenced in the backcross (at 1,620 kb in Fig. 2b). Interestingly, DNA methylation, H3mK9 and siRNA were retained at *ATGP1* in *ddm1* plants (ref. 12).

One hundred and six known genes are present in the 1.5-Mb region centred on *hk4S*; 16 of these lie in the knob, along with 15 hypothetical open reading frames (ORFs). Most genes were expressed, together with contiguous unannotated features, representing 5′ and 3′ untranslated regions. Unlike transposable elements, genes were expressed at similar levels in wild-type and *ddm1* plants, regardless of location. Genes in the knob are found in gene islands (Fig. 2a, c), which may correspond to interphase 'loops'

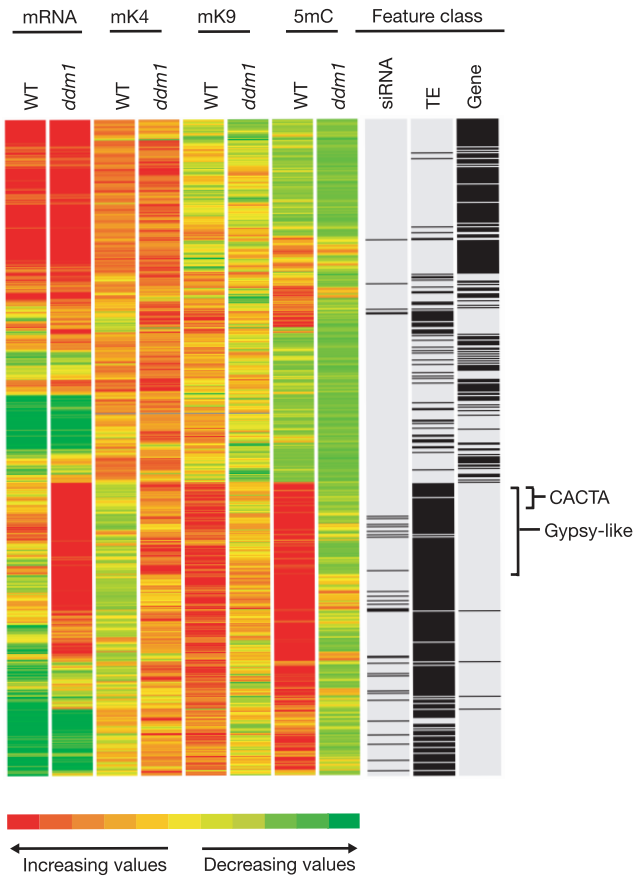

Figure 3 Cluster analysis. Tiles corresponding to ORFs were clustered according to the levels of expression (mRNA), H3mK4 (mK4), H3mK9 (mK9) and DNA methylation (5mC) in wild-type and *ddm1* plants using GeneSpring 6.0, and subsequently scored as matching known genes (Gene), transposable elements (TE) and siRNA (siRNA). Genes and transposable elements were sharply distinguished from each other. Genes were expressed in wild type and associated with gene traps and with H3mK4, whereas transposable elements were silent, and associated with siRNA, H3mK9 and DNA methylation. In *ddm1*, transposable elements adopted gene-like chromatin properties and the majority were expressed.

of euchromatin that emanate from the chromocentre[5]. Most genes on the array were unmethylated, although low levels of methylation were detected in some genes, regardless of their expression (Supplementary Table S1). For example, two closely related callose synthase genes (at 1,572 and 2,536 kb in Fig. 2) were methylated at their 3′ ends (Supplementary Fig. S2). Unlike transposable elements, methylated genes were not associated with significant levels of H3mK9 in wild type, and their expression (Fig. 2a) and DNA methylation was usually unchanged in *ddm1* (Supplementary Fig. S2). We conclude that genes are insulated from DDM1-mediated heterochromatic silencing.

Gene and enhancer traps are sensitive indicators of chromatin structure[6]. Among the 3,000 insertions mapped to the short arm of chromosome 4, there is a pronounced bias for recovery of insertions in gene islands (http://genetrap.cshl.org). For example, there are six gene trap insertions in At4g04020 (at 1,931 kb in Fig. 2a) but none in the surrounding 100 kb, indicating that genes within the knob are more permissive than transposable elements for marker expression.

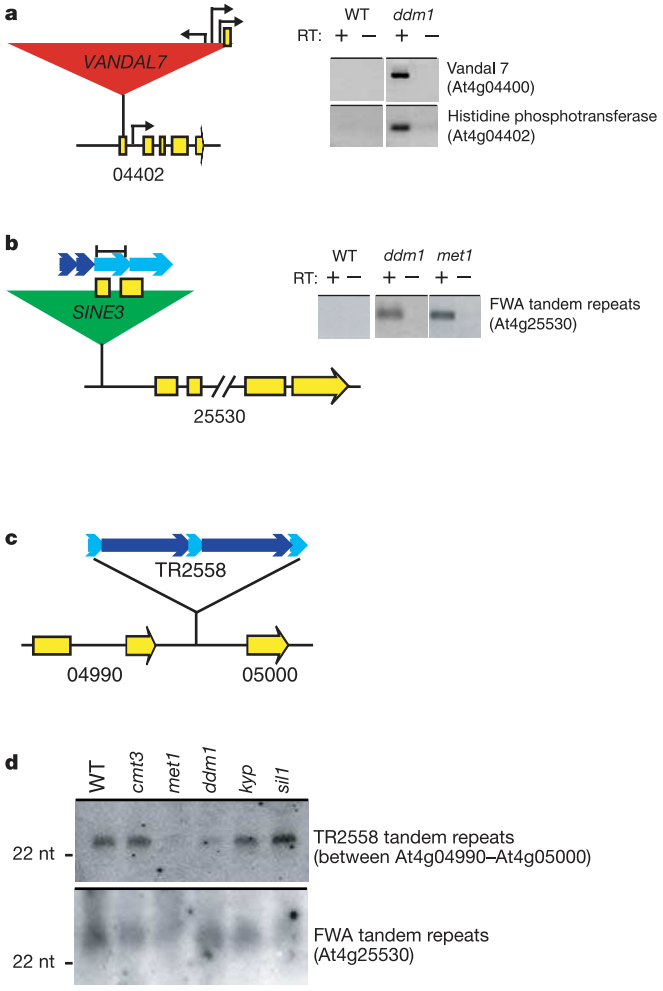

Figure 4 DDM1-dependent gene regulation. **a**, The histidine phosphotransferase gene At4g04402. This gene is interrupted by a *MuDR*-like *VANDAL* transposon and is silenced in wild type. RT–PCR using specific primers indicates that both the gene and the transposon are activated in *ddm1*. Rapid amplification of cloned cDNA ends (5′ RACE) was used to map transcription start sites (arrows). The VANDAL element provides two of the three sites detected for AT4g04402. **b**, *FWA* and its associated repeats. The promoter and upstream region of the late-flowering gene *FWA* are provided by a SINE retrotransposon and associated tandem repeats (a detailed analysis is shown in Supplementary Fig. S5). These repeats are silent in wild type, but transcripts accumulate in *ddm1* and *met1*. Tandem repeat transcripts are multimeric (data not shown). The primer set used for RT–PCR is shown (brackets). **c**, Tandem repeats corresponding to siRNA. The arrangement of intergenic tandem repeats at coordinate 2,558 kb (*TR2558*) is shown. **d**, siRNA corresponding to tandem repeats were reduced in *ddm1* and/or *met1*. For comparison, chromatin mutants with little effect on siRNA accumulation are shown: the non-CG DNA methyltransferase *cmt3* (*chromomethylase3*), the histone lysine 9 methyltransferase *kyp* (*kryptonite*), and the histone deacetylase *sil1* (*silencing locus1*). A 22-oligonucleotide (nt) size marker is indicated.

Although genes and transposable elements are sharply distinguished by *DDM1*, they resemble each other in most respects, such as G+C content and coding capacity. Thus, the mechanism by which they are differentiated is probably sequence specific. One possibility is that transposable elements are recognized by RNA interference (RNAi). RNAi silences transposable elements in many eukaryotes, generating 21–24-nucleotide siRNA derived from double-stranded (ds)RNA by the RNase-helicase Dicer[16,17]. siRNAs from *Arabidopsis* are currently being cloned and sequenced (http://cgrb.orst.edu/smallRNA/db/), and so far 93 features from the 1.5-Mb region match one or more siRNA (Figs 2 and 3; http://chromatin.cshl.org/ddm1/).

To determine the role of DDM1 and siRNA in targeting transposable elements, microarray data were subjected to unsupervised cluster analysis and then compared to annotation (Supplementary Methods). Features annotated as transposable elements and siRNA clustered together, whereas genes were more similar to each other (Fig. 3). Although cluster analysis is a useful exploratory tool, confidence levels cannot be easily assigned to any given association. Therefore, significantly enriched or depleted features from the linear model analysis were also summarized (Fig. 2a; see also Supplementary Table S1). Transposable elements and siRNAs were associated with H3mK9 and DNA methylation, and transposable elements that corresponded to siRNA were typically expressed in *ddm1* mutants (Fig. 3; see also Supplementary Table S1). This observation suggests that transcription of transposable elements targets their heterochromatic modification, consistent with a mechanism whereby siRNA guides DNA and histone methylation[18]. dsRNA can arise from transposable elements by read-through transcription, and transcripts were detected from both strands of many transposable elements in *ddm1*, probably due to integration into other transposable elements (Supplementary Fig. S3). This mechanism can account for much of the siRNA specific to transposable elements[16,17].

Tandem repeats are conspicuous features of heterochromatin, and theoretically they can amplify siRNA through re-iterative RNA replication and cleavage[19]. The knob sequence includes 22.5 copies of a 1,850-base-pair (bp) tandem repeat, *AtENSAT1* (1,700–1,750 kb in Fig. 2a), of which the first 300 bp correspond to the terminal inverted repeats of the CACTA transposon *ATENSPM4* (refs 6, 8). In addition, there are several euchromatic tandem repeats represented on the array including *TR2558* (at 2,558 kb in Fig. 2a) (Fig. 4c). Both *ATENSAT1* and *TR2558* accumulate a large number of siRNAs. The repeats are methylated and associated with H3mK9. These modifications are under the control of *DDM1*, and siRNAs from *TR2558* are severely reduced in both *ddm1* and *met1* mutants (Fig. 4d). Thus siRNAs are derived from both transposable elements and tandem repeats, and are regulated by *DDM1*.

The high resolution provided by the microarray (~1 kb) indicated that DDM1 sharply distinguishes genes and transposable elements (Fig. 2); however, a handful of genes found in the knob were silenced by *DDM1*. At4g04402 is related to histidine phosphotransferase *AtHP2* genes found on chromosomes 3 and 5. Although these paralogues were strongly expressed and associated with H3mK4 (Supplementary Fig. S4), At4g04402 was silent, heavily methylated and associated with H3mK9 (at 2,163 kb in Fig. 2a). A 13-kb *VANDAL7* transposon is inserted 134 bp within the silent gene, but in *ddm1* mutants both the transposon and the gene were transcriptionally activated (Fig. 4a). Furthermore, this activation was inherited in *ddm1*/+ plants (data not shown).

VANDAL elements belong to the *MuDR* superfamily, which have

been shown to bring genes under epigenetic control in maize[20]. As in maize, At4g04402 transcripts arose from an outward-reading promoter provided by the transposable element (Fig. 4a). In contrast, a similar *VANDAL7* element integrated less than 500 bp upstream of At4g04320 (at 2,112 kb in Fig. 2a) did not bring the gene under its control. Three other genes in the knob (At4g03950, At4g04080 and At4g04110) were silenced by *DDM1*, and had repetitive insertions relative to euchromatic paralogues (data not shown). However, activation was not observed in every biological replicate, and was not explored further. We conclude that epigenetic gene silencing mediated by transposable elements requires insertion within or very near the gene.

Heritable changes in the expression of transposable-element-regulated genes provide a potential mechanism for the 'syndrome' of sporadic phenotypes that arise in inbred *ddm1* mutants[15]. One such phenotype is a delay in flowering owing to inappropriate expression of *FWA*, which encodes a homeodomain protein normally expressed in the seed[21]. *FWA* expression is imprinted in the endosperm under the control of *MET1* (ref. 22). Although it is not located in the knob, we included *FWA* on the array because of its regulation by *DDM1* (ref. 21).

The *FWA* promoter and 5′ untranslated region comprise two sets of tandem repeats of 38 bp and 198 bp, respectively, which are methylated under *DDM1* control[21]. Because of the involvement of transposable elements in *DDM1*-mediated gene silencing, we inspected this region and discovered that the repeats are encompassed within a 422-bp *SINE3* retrotransposon, distantly related to *Arabidopsis AtSN2* (Supplementary Fig. S5). Short interspersed nuclear element (SINE) insertion and additional rearrangements involving a 10-bp target site duplication result in the retroelement providing the first two exons of the gene (Supplementary Fig.S5), including the start site of *FWA* transcription[21]. Microarray analysis revealed that this retrotransposon is associated with H3mK9 in wild type but not in *ddm1* (http://chromatin.cshl.org/ddm1/). These data indicate that *FWA* is under transposable element control, and implicate transposable elements in imprinting. To explore this mechanism further, we searched for siRNAs corresponding to the 198-bp repeat. These were detected in wild type and *ddm1*, but were reduced in *met1* (Fig. 4d). Low levels of transcripts corresponding to both copies of the repeat could be readily detected in *ddm1* and *met1*, and are probably the source of these siRNAs (Fig. 4b).

We have found that heterochromatin, whether cytologically visible or not, is defined by transposable elements, consistent with the view expressed by B. McClintock more than 50 years ago[2]. Genes are insulated from heterochromatin within local euchromatic environments[5]. A similar organization is found within a rice centromere[23], and may also apply to the bulk of the maize genome, given that gene islands are unmethylated and separated by a sea of methylated transposable elements[24].

In fission yeast and in *Drosophila*, which mostly lack DNA methylation, RNA interference and H3mK9 are required for position effect variegation associated with heterochromatic repeats[10,18,25]. Position effect variegation silences nearby genes, but the silencing is relatively unstable, resulting in variegation. In *Arabidopsis*, both DNA methylation and H3mK9 are maintained by *DDM1*, and heterochromatic silencing is heritable. *DDM1* distinguishes genes from transposable elements and repeats, and siRNAs probably contribute to this distinction. Some siRNAs corresponding to transposable elements are lost in *ddm1* mutants, suggesting that siRNAs are engaged in a stabilizing interaction with DDM1 (ref. 12). In fission yeast, the H3K9 methyltransferase clr4 + is similarly required for siRNA accumulation[26], possibly mediated by the RITS complex[27]. However, the *Arabidopsis* RNAi genes *AGO1* and *AGO4* have only limited roles in transposable element silencing[12,28]. One explanation is that siRNAs are only needed to target heterochromatic modifications, which once established, are maintained by DDM1. This would also explain why silencing cannot be re-established on out-crossing of *ddm1* once H3mK9, DNA methylation and siRNA are lost[12].

We show that transposable elements silence genes epigenetically, but only when they integrate within or very near them, resembling suppressible insertions in plants, animals and fungi[11]. In the mouse, for example, the insertion of an IAP retrotransposon immediately upstream of the agouti gene provides new transcripts and leads to mutant phenotypes only when unmethylated.

Late-flowering epimutants of the euchromatic gene *FWA* arise in *ddm1* and even more frequently in *met1* (refs 15, 21). Furthermore, *FWA* expression is imprinted in the endosperm under the control of *MET1* (ref. 22). The promoter and first two exons of *FWA* are provided by a SINE, which brings the gene under the control of DDM1. The SINE contains tandem repeats that are methylated and correspond to siRNA. As with heterochromatic tandem repeats, these siRNAs are reduced in *met1*, and longer transcripts accumulate in both *met1* and *ddm1*. Loss of tandem repeat siRNA in the female germ line might account for loss of heterochromatic modifications and specific expression from the maternal allele in the developing seed. Notably, differentially methylated regions of imprinted genes in mammals are also transcribed and contain tandem repeats, some of which correspond to microRNA[29]. Thus, in repeat-rich genomes, transposable elements and related repeats are likely to have major regulatory roles in development and disease. *Note added in proof*: While under review, additional reports[30,31] have indicated that RNAi is required to establish *FWA* transgene silencing in *Arabidopsis*, in agreement with our results. □

Methods

Plant material

All wild-type and homozygous *ddm1-2 Arabidopsis* plants used in this study were of the Columbia ecotype. Individual *ddm1-2* mutant lines were self-pollinated for three generations and then pooled for analysis[14]. Plants were grown *in vitro* on 0.5 × Murashige and Skoog (MS) media, with sucrose (0.7%). Nine-day-old seedlings were then collected for chromatin immunoprecipitation, DNA and RNA extraction. Endogenous siRNAs were identified from cDNA libraries prepared using small RNA isolated from inflorescence tissues. All siRNAs can be viewed at http://cgrb.orst.edu/smallRNA/db/. Small RNA detection in chromatin mutants was carried out as previously described[12] using soil-grown plants in the Landsberg *erecta* background. The blot shown in Fig. 4d was repeated twice using microRNA miR171 as a loading control (not shown). It was unaffected in all genotypes[12].

Microarrays

Bacterial artificial chromosome (BAC) templates covering the chromosome 4 heterochromatic knob and flanking sequences were used to generate the fragments printed on microarrays by PCR amplification using primers selected at 1-kb intervals. BACs come from the 'tiling path' used for sequencing of the genome of *Arabidopsis thaliana* cv. Columbia (http://www.mips.biochem.mpg.de/proj/thal/db/). The 20 BACs used were: T10P11, T5J8, T4I9, F4C21, F9H3, T5L23, T5H22, T7M24, T25H8, T24M8, T24H24, T27D20, T19B17, T26N6, F4H6, T19J18, T4B21, T1J1, T32N4 and C17L7. All PCR products were quantified on agarose gels, before printing in duplicate or triplicate onto glass slides. Polyadenylated RNA from wild-type and *ddm1* seedlings was labelled indirectly with Cy3 and Cy5 and hybridized to pairs of slides in dye-swap experiments. Two dye swaps were performed using the same RNA (that is, four technical replicates), and one dye swap was carried out using seedlings grown under similar conditions (two biological replicates). In addition to labelled cDNA, the arrays were also hybridized with total genomic DNA extracted from seedlings and the same genomic DNA depleted for methylated sequences. DNA was sheared to a constant size using nebulization and then divided into two equal samples, one of which was digested with the methylation-dependent restriction enzyme McrBC, which cuts the sequence A/G 5mC. Depleted and undepleted DNA samples were then size-fractionated on agarose gels to recover fragments greater than 1 kb, closely matching the resolution provided by the tiling array. The intensity ratios gave a quantitative indication of cytosine methylation. We performed chromatin immunoprecipitation with antibodies raised against H3 dimethyl K4 or H3 dimethyl K9 (Upstate Biotechnology). Whole seedlings were treated with formaldehyde to fix histone–DNA complexes, chromatin was extracted, sonicated and immunoprecipitated before amplification, labelling and hybridization to the array.

A linear model approach[32] was used to estimate experimental effects, determine the level of variability, and detect features undergoing statistically significant changes in fluorescence intensity between the conditions investigated by each experiment. The linear models procedure partitions the sources of variation such that global and feature-specific array and dye effects are removed, creating a corrected signal for each feature. This signal was then used to assess changes in fluorescence intensity for each feature solely due to differences between the conditions of interest. In the case of DNA methylation the average ratio of hybridization intensities for selected DNA and total DNA was calculated for each feature in the dye-swap experiments. In both wild-type and *ddm1* samples, unmethylated

features gave a ratio of close to 1.0, whereas methylated features gave a ratio greater than 1.0. Similarly, for chromatin immunoprecipitation, the ratio of selected and total DNA was calculated for each feature. In this case the ratio found in euchromatin was arbitrarily set to 1.0 in both mutant and wild type, as recovery of immunoselected DNA is always much less than 100%. This is the equivalent of using euchromatic genes such as actin as a control. Statistical tests were used to detect features undergoing significant changes, based on control of both the family-wise error rate (FWER) and the false discovery rate (FDR)[32].

This statistical analysis allows widely differing profiles to be compared such that the resulting data were highly reproducible, with less than 1% of the features displaying different expression ratios under biological replication (not shown). Because of the repetitive nature of the sequences involved, cross-hybridization between repeats and gene family members was likely to account for some of the signal observed. However, PCR validation with specific primer pairs indicated that, except in rare instances, hybridization provided an accurate measure of expression as well as DNA and histone methylation patterns across the 1.5-Mb region represented on the array (Supplementary Methods).

Received 2 January; accepted 7 May 2004; doi:10.1038/nature02651.

1. Heitz, E. Das heterochromatin der Moose. *Jehrb. Wiss. Botanik* **69,** 762–818 (1928).
2. McClintock, B. Chromosome organization and genic expression. *Cold Spring Harb. Symp. Quant. Biol.* **16,** 13–47 (1951).
3. Hennig, W. Heterochromatin. *Chromosoma* **108,** 1–9 (1999).
4. Schotta, G., Ebert, A., Dorn, R. & Reuter, G. Position-effect variegation and the genetic dissection of chromatin regulation in *Drosophila*. *Semin. Cell Dev. Biol.* **14,** 67–75 (2003).
5. Fransz, P., De Jong, J. H., Lysak, M., Castiglione, M. R. & Schubert, I. Interphase chromosomes in *Arabidopsis* are organized as well defined chromocenters from which euchromatin loops emanate. *Proc. Natl Acad. Sci. USA* **99,** 14584–14589 (2002).
6. CSHL/WUGSC/PEB, *Arabidopsis* Sequencing Consortium. The complete sequence of a heterochromatic island from a higher eukaryote. *Cell* **100,** 377–386 (2000).
7. *Arabidopsis* Genome Initiative, Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408,** 796–815 (2000).
8. Jurka, J. Repbase update: a database and an electronic journal of repetitive elements. *Trends Genet.* **16,** 418–420 (2000).
9. Jenuwein, T. & Allis, C. D. Translating the histone code. *Science* **293,** 1074–1080 (2001).
10. Vermaak, D., Ahmad, K. & Henikoff, S. Maintenance of chromatin states: an open-and-shut case. *Curr. Opin. Cell Biol.* **15,** 266–274 (2003).
11. Martienssen, R. A. & Colot, V. DNA methylation and epigenetic inheritance in plants and filamentous fungi. *Science* **293,** 1070–1074 (2001).
12. Lippman, Z., May, B., Yordan, C., Singer, T. & Martienssen, R. Distinct mechanisms determine transposon inheritance and methylation via small interfering RNA and histone modification. *PLoS Biol.* **1,** E67 (2003).
13. Verbsky, M. L. & Richards, E. J. Chromatin remodeling in plants. *Curr. Opin. Plant Biol.* **4,** 494–500 (2001).
14. Gendrel, A. V., Lippman, Z., Yordan, C., Colot, V. & Martienssen, R. A. Dependence of heterochromatic histone H3 methylation patterns on the *Arabidopsis* gene DDM1. *Science* **297,** 1871–1873 (2002).
15. Kakutani, T. Epi-alleles in plants: inheritance of epigenetic information over generations. *Plant Cell Physiol.* **43,** 1106–1111 (2002).
16. Sijen, T. & Plasterk, R. H. Transposon silencing in the *Caenorhabditis elegans* germ line by natural RNAi. *Nature* **426,** 310–314 (2003).
17. Aravin, A. A. *et al.* Double-stranded RNA-mediated silencing of genomic tandem repeats and transposable elements in the *D. melanogaster* germline. *Curr. Biol.* **11,** 1017–1027 (2001).
18. Volpe, T. A. *et al.* Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science* **297,** 1833–1837 (2002).
19. Martienssen, R. A. Maintenance of heterochromatin by RNA interference of tandem repeats. *Nature Genet.* **35,** 213–214 (2003).
20. Barkan, A. & Martienssen, R. A. Inactivation of maize transposon Mu suppresses a mutant phenotype by activating an outward-reading promoter near the end of Mu1. *Proc. Natl Acad. Sci. USA* **88,** 3502–3506 (1991).
21. Soppe, W. J. *et al.* The late flowering phenotype of fwa mutants is caused by gain-of-function epigenetic alleles of a homeodomain gene. *Mol. Cell* **6,** 791–802 (2000).
22. Kinoshita, T. *et al.* One-way control of FWA imprinting in *Arabidopsis* endosperm by DNA methylation. *Science* **303,** 521–523 (2004).
23. Nagaki, K. *et al.* Sequencing of a rice centromere uncovers active genes. *Nature Genet.* **36,** 138–145 (2004).
24. Rabinowicz, P. D. *et al.* Genes and transposons are differentially methylated in plants, but not in mammals. *Genome Res.* **13,** 2658–2664 (2003).
25. Pal-Bhadra, M. *et al.* Heterochromatic silencing and HP1 localization in *Drosophila* are dependent on the RNAi machinery. *Science* **303,** 669–672 (2004).
26. Schramke, V. & Allshire, R. Hairpin RNAs and retrotransposon LTRs effect RNAi and chromatin-based gene silencing. *Science* **301,** 1069–1074 (2003).
27. Verdel, A. *et al.* RNAi-mediated targeting of heterochromatin by the RITS complex. *Science* **303,** 672–676 (2004).
28. Zilberman, D., Cao, X. & Jacobsen, S. E. ARGONAUTE4 control of locus-specific siRNA accumulation and DNA and histone methylation. *Science* **299,** 716–719 (2003).
29. Seitz, H. *et al.* Imprinted microRNA genes transcribed antisense to a reciprocally imprinted retrotransposon-like gene. *Nature Genet.* **34,** 261–262 (2003).
30. Chan, S. W. *et al.* RNA silencing genes control *de novo* DNA methylation. *Science* **303,** 1336 (2004).
31. Xie, Z. *et al.* Genetic and functional diversification of small RNA pathways in plants. *PLoS Biol.* **2,** E104 (2004).
32. Craig, B. A., Black, M. A. & Doerge, R. W. Gene expression data: the technology and statistical analysis. *J. Agric. Biol. Environ. Stat.* **8,** 1–28 (2003).

Supplementary Information accompanies the paper on **www.nature.com/nature**.

Acknowledgements We thank E. Richards and our colleagues T. Osborn, L. Comai, J. Chen and J. Birchler for their comments and advice. We also thank P. Rabinowicz for advice on ChIP microarray experiments. V.C. thanks M. Caboche for laboratory space and continuous support. Z.L. is an Arnold and Mabel Beckman graduate fellow in the Watson School of Biological Sciences. A.V.G. is supported by a graduate studentship from the French Ministry of Research. M.V. is a National Science Foundation Bioinformatics postdoctoral fellow. This work was supported by a grant from the NSF Plant Genome Program (to R.W.D. and R.M.), as well as grants from Genopole and the CNRS (to V.C.), grants from NSF and NIH to J. C., and NIH to R.M.

Competing interests statement The authors declare competing financial interests: details accompany the paper on **www.nature.com/nature**.

Correspondence and requests for materials should be addressed to R.M. (martiens@cshl.edu) or V.C. (colot@evry.inra.fr).

Periodic cycles of RNA unwinding and pausing by hepatitis C virus NS3 helicase

Victor Serebrov[1] **& Anna Marie Pyle**[1,2]

[1]*Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06520, USA*
[2]*Howard Hughes Medical Institute, New Haven, Connecticut 06520, USA*

The NS3 helicase is essential for cytoplasmic RNA replication by the hepatitis C virus[1–4], and it is a representative member of helicase superfamily 2 (SF2). NS3 is an important model system for understanding unwinding activities of DExH/D proteins[5–7], and it has been the subject of extensive structural and mutational analyses[8–11]. Despite intense interest in NS3, the molecular and kinetic mechanisms for RNA unwinding by this helicase have remained obscure. We have developed a combinatorial, time-resolved approach for monitoring the microscopic behaviour of a helicase at each nucleotide of a duplex substrate. By applying this analysis to NS3, we have independently established the 'physical' and 'kinetic' step size for unwinding of RNA (18 base pairs, in each case), which we relate to the stoichiometry of the functional, translocating species. Having obtained microscopic unwinding rate constants at each position along the duplex, we demonstrate that NS3 unwinds RNA through a highly coordinated cycle of fast ripping and local pausing that occurs with regular spacing along the duplex substrate, much like the stepping behaviour of cytoskeletal motor proteins[12].

Replication by the hepatitis C virus (HCV) is performed by a protein complex that includes the multifunctional NS3 helicase, the NS5B polymerase and other nonstructural proteins[1–4]. Although there have been numerous studies on the protease, ATPase and helicase activities of the protein[13–16], a quantitative analysis of its RNA unwinding mechanism has never been conducted. Helicase activity is traditionally monitored on short, 'tailed' substrates that consist of a duplex flanked by single stranded nucleic acid. Information about helicase mechanism (such as kinetic parameters and processivity) can be obtained by monitoring the extent and efficiency of unwinding for a series of duplexes that sequentially increase in length. Although this approach has provided a wealth of information about numerous helicases[15,17–20], it requires laborious synthesis of multiple substrates, lacks single-nucleotide resolution and the results are potentially influenced by end-effects at the duplex termini. Recent advances in single molecule technology have provided information about overall unwinding rate constants and processivity[21–24], however, they lack the resolution to report on microscopic behaviour at the single nucleotide level.

To address these issues and to build a mechanistic framework for


a
mRNA
Annotation
WT 5mC
WT mK9
WT mK4
ddm1 5mC
ddm1 mK9
ddm1 mK4
1,200
1,300
1,400
1,500
1,600
1,700
1,800
1,900
2,000
2,100
2,200
2,300
2,400
2,500
2,600
(kb)
b
WT versus ddm1
WT versus ddm1/+
Annotation
1,500
1,600
1,700
1,800
1,900
2,000
(kb)
c
mRNA
Annotation
WT 5mC
WT mK9
WT mK4
ddm1 5mC
ddm1 mK9
ddm1 mK4
1
2
3
1,920
1,930
1,940
1,950
(kb)
RT–PCR
WT
ddm1
RT: + − + −
Fibrillin (At4g04020)
LINE-like repeat (At4g04030)
Phosphofructokinase (At4g04040)
Actin