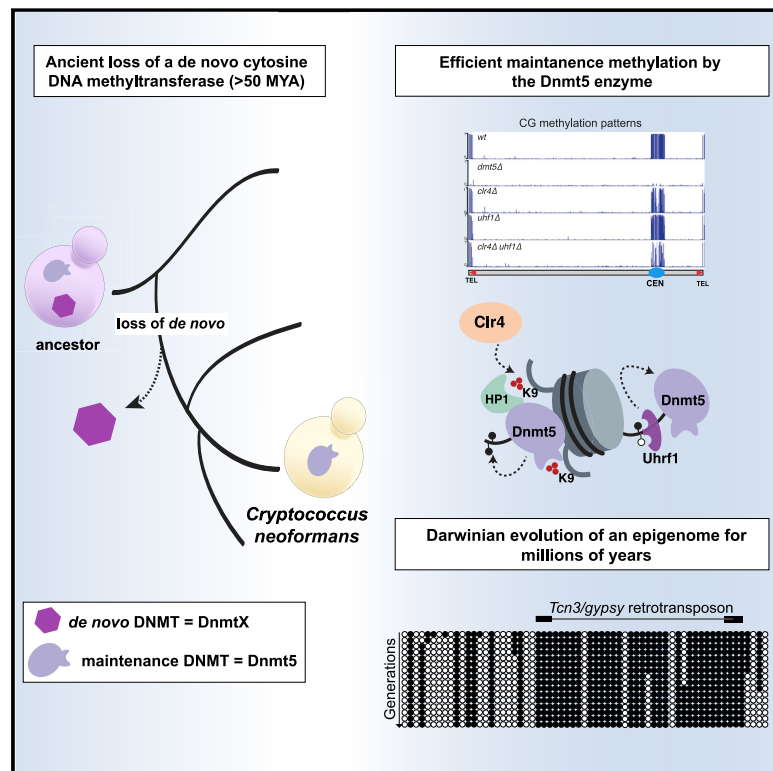# Evolutionary Persistence of DNA Methylation for Millions of Years after Ancient Loss of a *De Novo* Methyltransferase

Graphical Abstract

Authors

Sandra Catania, Phillip A. Dumesic, Harold Pimentel, ..., Christina A. Cuomo, Jonathan K. Pritchard, Hiten D. Madhani

Correspondence

hitenmadhani@gmail.com

In Brief

The pathogenic fungus *Cryptococcus neoformans* encodes a functional maintenance methyltransferase for 5mC but no *de novo* enzyme; 5mC has been maintained for millions of years by a combintation of random epimutation and natural selection.

Highlights

- Exquisitely specific maintenance methylase enzyme drives all 5mC in *C. neoformans*

- Once lost, methylation is not efficiently restored mitotically or meiotically

- The *de novo* enzyme DnmtX was lost in an ancestral species ∼50–150 mya

- Persistence of 5mC for millions of years explained by Darwinian epigenome evolution

CellPress

Article

Evolutionary Persistence of DNA Methylation for Millions of Years after Ancient Loss of a *De Novo* Methyltransferase

Sandra Catania,[1] Phillip A. Dumesic,[1,9] Harold Pimentel,[5,9] Ammar Nasif,[6,7] Caitlin I. Stoddard,[1] Jordan E. Burke,[1] Jolene K. Diedrich,[8] Sophie Cook,[1] Terrance Shea,[3] Elizabeth Geinger,[3] Robert Lintner,[3] John R. Yates III,[8] Petra Hajkova,[6,7] Geeta J. Narlikar,[1] Christina A. Cuomo,[3] Jonathan K. Pritchard,[4,5] and Hiten D. Madhani[1,2,10,*]

[1]Department of Biochemistry and Biophysics, University of California, San Francisco, San Francisco, CA 94158, USA
[2]Chan-Zuckerberg Biohub, San Francisco, CA 94158, USA
[3]Infectious Disease and Microbiome Program, Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA
[4]Department of Biology, Stanford University, Stanford, CA 94305, USA
[5]Howard Hughes Medical Institute, Stanford University, Stanford, CA 94305, USA
[6]MRC London Institute of Medical Sciences (LMS), Reprogramming and Chromatin Group, Du Cane Road, W12 0NN London, UK
[7]Institute of Clinical Sciences, Imperial College Faculty of Medicine, Du Cane Rd, W12 0NN London, UK
[8]Department of Molecular Medicine, The Scripps Research Institute, La Jolla, CA 92037, USA
[9]These authors contributed equally
[10]Lead Contact
*Correspondence: hitenmadhani@gmail.com
https://doi.org/10.1016/j.cell.2019.12.012

SUMMARY

Cytosine methylation of DNA is a widespread modification of DNA that plays numerous critical roles. In the yeast *Cryptococcus neoformans*, CG methylation occurs in transposon-rich repeats and requires the DNA methyltransferase Dnmt5. We show that Dnmt5 displays exquisite maintenance-type specificity *in vitro* and *in vivo* and utilizes similar *in vivo* cofactors as the metazoan maintenance methylase Dnmt1. Remarkably, phylogenetic and functional analysis revealed that the ancestral species lost the gene for a *de novo* methylase, DnmtX, between 50–150 mya. We examined how methylation has persisted since the ancient loss of DnmtX. Experimental and comparative studies reveal efficient replication of methylation patterns in *C. neoformans*, rare stochastic methylation loss and gain events, and the action of natural selection. We propose that an epigenome has been propagated for >50 million years through a process analogous to Darwinian evolution of the genome.

INTRODUCTION

Methylation of cytosine on its fifth carbon (5mC) in DNA is found in all domains of life (Bird, 2002). The DNA of all known vertebrates and land plants harbor this modification, likely because of its critical function in genome defense (Zemach et al., 2010). 5mC is initially deposited by *de novo* DNA methyltransferases (DNMTs) that act on unmethylated substrates. In the context of palindromic sequence arrangements (e.g., CG and CHG) in which cytosines on both strands are methylated, this initial activity can then be "remembered" by maintenance enzymes that function on hemimethylated DNA produced by DNA replication (Law and Jacobsen, 2010). One example of this division of labor is found in the mammalian DNA methylation system. Here, *de novo* methylation is principally catalyzed by the Dnmt3a/b-Dnmt3L complex (Du et al., 2015). Once the DNA methylation mark is established during germline and embryonic development, it is then maintained by the Dnmt1 enzyme in association with UHRF1, a protein that recognizes H3K9me and hemimethylated 5mC (Bostick et al., 2007; Du et al., 2015; Law and Jacobsen, 2010). This division of labor between *de novo* and maintenance enzymes extends to land plants (Bronner et al., 2019), although Dnmt3 appears to have been lost in angiosperms and replaced by other *de novo* enzymes (Yaari et al., 2019). DNA methylation is essential for development and plays critical roles in silencing of transposable elements, chromosome stability, monoallelic gene expression, and gene silencing (Jones, 2012). Moreover, Dnmt3a is a commonly mutated driver gene in acute myeloid leukemia (Brunetti et al., 2017) and mutations in Dnmt3b cause the autosomal recessive disease immunodeficiency, centromeric region instability, facial anomalies (ICF) syndrome (Walton et al., 2014).

While nearly all opisthokonts that have DNA methylation encode at least two DNMT homologs in their genomes, some, curiously, have only one (Bewick et al., 2017, 2019; Rošić et al., 2018). One such species is the human fungal pathogen *Cryptococcus neoformans*, a basidiomycetous yeast that possesses symmetric CG methylation at transposon-rich centromeric and subtelomeric regions (Huff and Zilberman, 2014). This methylation is dependent on the predicted cytosine DNA methyltransferase (DNMT) Dnmt5, encoded by the *DMT5* gene (Huff and Zilberman, 2014), a gene found previously to be required for fitness of *C. neoformans* during infection (Liu et al., 2008). The Dnmt5 family is characterized by an N-terminal chromodomain (CD) followed by a cytosine methyltransferase catalytic domain, a RING

Cell
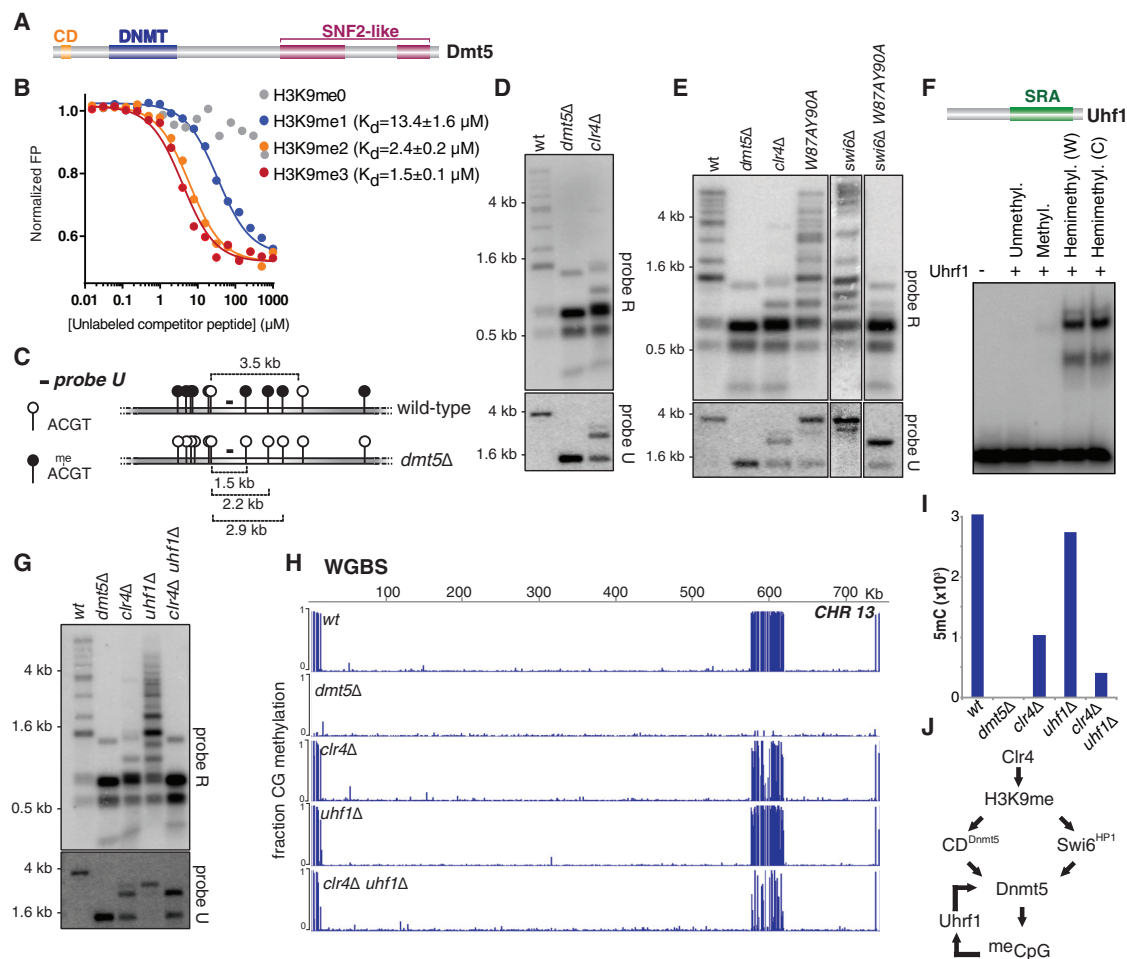
Figure 1. Multiple Mechanisms Promote Efficient DNA Methylation in *C. neoformans*

(A) Architecture of Dnmt5. CD, chromodomain; DNMT, DNA methyltransferase domain; SNF2-like, Swi/Snf ATPase domain.

(B) Binding affinity of recombinant Dnmt5 CD to the indicated peptides as determined by competition fluorescence polarization.

(C) Sites for the methylation-sensitive restriction endonuclease HpyCH4IV (ACGT) within a region of centromere 13. Filled circles, methylated sites; open circles, unmethylated sites.

(D) 5mC levels of wild-type, *dmt5Δ*, and *clr4Δ* assessed by Southern hybridization of HpyCH4IV-digested genomic DNA using probe corresponding to a repetitive sequence (probe R) or a unique sequence (probe U; B).

(E) Southern analysis of 5mC in cells carrying CD mutant of Dnmt5 (*W87AY90A*), deletion of Swi6 (*swi6Δ*), or a combination of the two (*swi6Δ W87AY90A*).

(F) Mobility shift assay testing the binding of recombinant Uhrf1 to indicated DNA probes. W and C indicate methylation of Watson versus Crick strands.

(G) Southern analysis of 5mC in wild-type, *dmt5Δ*, *clr4Δ*, *uhf1Δ*, and double mutant *clr4Δ uhf1Δ* strains.

(H) Whole-genome bisulfite sequencing (WGBS) analysis of wild-type, *dmt5Δ*, *clr4Δ*, *uhf1Δ*, and *clr4Δ uhf1Δ* strains. Shown are the data for chromosome 13.

(I) Number of methylated sites in the mutants analyzed in (H) as determined by WGBS.

(J) Regulatory circuitry of DNA methylation in *C. neoformans*.

See also Figures S1, S2, and S3 and Tables S1 and S4.

finger, and a domain related to those of SNF2-type ATPases (Figure 1A). This putative enzyme is widespread in fungi and in green algae, several of which have been shown to have CG methylation that impacts nucleosome positioning (Huff and Zilberman, 2014). 5mC in *C. neoformans* is concentrated at transposable element-rich repeats; it likely functions in transposon silencing as a sister species, *Cryptococcus deuterogattii*, has lost all active transposons, and concomitantly lost 5mC and displays inactivating mutations in the *DMT5* gene (Yadav et al., 2018a). Below we describe our investigations of *C. neoformans* Dnmt5.

RESULTS

Factors that Promote 5mC in *C. neoformans*

In *C. neoformans*, regions of the genome decorated with 5mC coincide with those we reported to display H3K9me (Dumesic et al., 2015; Huff and Zilberman, 2014). Therefore, we tested whether the CD of Dnmt5 recognizes this mark. Binding of a purified fragment containing the CD to a commercial array of modified human histone peptides yielded signals with H3K9me-containing peptides as well as those containing H3K27me (Figures

Cell

S1A and S1B). However, because H3 in *C. neoformans* differs from the human H3 sequence around lysine 27, we assessed the binding capability of the chromodomain using fluorescence polarization and peptides corresponding to *C. neoformans* H3 sequences (Figures 1B and S1C). Binding to H3K9me (K_d = 1.5 μM for H3K9me3) was substantially stronger than binding to H3K27me (K_d > 250 μM), suggesting that H3K9me is the principal binding partner of the CD.

To assess whether H3K9me affects 5mC, we deleted the gene encoding the only H3K9 methyltransferase in *C. neoformans*, Clr4 (*clr4Δ*) (Dumesic et al., 2015). 5mC was assayed by digestion of genomic DNA with a methylation-sensitive enzyme (HpyCH4IV; ÂCGT) followed by Southern hybridization. Based on published maps of 5mC in wild-type *C. neoformans*, we designed a probe directed to a unique region of centromere 13 (Figure 1C, probe U). In addition, another probe recognizing repetitive centromeric sequences (probe R) was created to analyze the distribution of 5mC across several centromeric sites at once.

In the regions probed, 5mC levels are markedly reduced in the *clr4Δ* strain (Figure 1D), while the *dmt5Δ* strain as well as strains harboring mutations in catalytic residues in the either the DNMT catalytic domain or ATPase domain display no 5mC (Figures 1D and S2A). Chromatin immunoprecipitation sequencing (ChIP-seq) analysis of FLAG-Dnmt5 demonstrated that its recruitment to H3K9me domains is dramatically reduced (Figure S2B). In contrast, a mutation of the Dnmt5 chromodomain that abolishes its binding to H3K9me (*W87A,Y90A*) has only a minor effect on 5mC, suggesting that other H3K9me-dependent factors contribute to 5mC levels (Figures 1E and S2C). The heterochromatin protein 1 (HP1) family includes several conserved proteins that, by binding to H3K9me, participate in heterochromatin formation and can contribute to 5mC methylation (Bannister et al., 2001; Lachner et al., 2001; Lewis et al., 2010). In *C. neoformans*, we identified a single HP1 ortholog (CNAG_03458), which we named Swi6 based on its characterized ortholog in fission yeast (Figure S2D). While the *swi6Δ* mutant marginally impacts 5mC, a *swi6Δ dmt5-W87A,Y90A* double mutant reduces the levels of 5mC to those seen in the *clr4Δ* strain (Figure 1E). Immunoprecipitation-mass spectrometry and coimmunoprecipitation analysis revealed that Swi6 and Dnmt5 physically associate (Figures S2E and S2F) suggesting that H3K9me promotes 5mC by recruiting Dnmt5 in two ways: via the CD of Dnmt5 and via HP1 (Figure S2G).

The residual 5mC seen in cells lacking H3K9me led us to seek additional factors involved in DNA methylation. *C. neoformans* possesses one Uhrf1-like protein that contains an SRA domain (CNAG_00677, Figure 1F) but lacks the Tudor H3K9me reader and RING E3 ligase domains found in its human ortholog. Using native gel electrophoresis, we tested the ability of recombinant *C. neoformans* Uhrf1 (Figure S3A) to bind labeled DNA that is unmethylated, hemimethylated, or symmetrically methylated at six CG dinucleotides. This experiment demonstrated that Uhrf1 selectively binds hemimethylated DNA (Figure 1F; multiple complexes seen likely reflects multiple molecules of Uhrf1 bound), a result confirmed by competition experiments (Figure S3B). The absence of Uhrf1 (*uhf1Δ*) has almost no detectable impact on 5mC *in vivo* but modestly alters 5mC patterns at sites probed when combined with a loss of Clr4 (*clr4Δ uhf1Δ*) (Figure 1G).

To assay these genotypes genome-wide, we performed whole-genome bisulfite-sequencing (WGBS) of DNA extracted from wild-type, *dmt5Δ*, *clr4Δ*, *uhf1Δ*, and *clr4Δ uhf1Δ* strains (Figure 1H). The total number of symmetrically methylated CG dinucleotides, which is reduced to 31% in *clr4Δ*, drops to 11% in the *clr4Δ uhf1Δ* double mutant (Figures 1I and 1J; n.b.: no non-CG methylation was observed). This analysis indicates that Uhrf1 and Clr4 promote 5mC through two parallel pathways (Figure 1J).

Purified Dnmt5 Displays Exquisite Maintenance-Type Specificity

C. neoformans Dnmt5 functions with a Uhrf1 homolog and itself recognizes H3K9me, which are features reminiscent of the mammalian maintenance methylation system. We therefore tested whether Dnmt5 has the substrate specificity expected of a maintenance enzyme. We expressed and purified full-length Dnmt5 in *S. cerevisiae*. We performed methyltransferase assays using a S-[³H-methyl]-adenosyl-methionine donor and synthetic 20 bp double-stranded DNA oligonucleotide substrates that harbored six CG sites. The substrate was either unmethylated, hemimethylated at all CG sites on the Watson strand, hemimethylated at all CG sites on the Crick strand, or symmetrically methylated on both stands. In the absence of ATP, no activity was observed on any substrate, as compared to control reactions that lacked Dnmt5 (Figure 2A). However, in the presence of ATP, methyltransferase activity was observed only on hemimethylated substrates with apparent catalytic rates comparable to that of human Dnmt1 (Figures 2A and 2B) (Pradhan et al., 1999). Endpoint assays with the unmethylated substrate did not reveal any activity above background on unmodified DNA; under the same conditions, we observed methyltransferase activity on the hemimethylated substrate that was ~1,300-fold over background (Figures S3C and S3D). In addition, Dnmt5 activity is specific for CG dinucleotides because no activity was detected on DNA substrates containing CHG motifs instead of CG motifs, regardless of cytosine methylation state (Figure S3E).

Because the chromodomain of Dnmt5 recognizes H3K9me, we asked whether *de novo* activity could be detected when trimethylated H3K9 peptide was added to the *in vitro* reaction. Rates on hemimethylated substrates were comparable in the presence of methylated or unmethylated H3K9 peptide, and no activity was detected on unmethylated DNA (Figures 2C and 2D). We also tested whether recombinant Uhrf1 and Swi6 could trigger a *de novo* activity when added to the reactions, but none could be detected (Figure S3F). We conclude that Dnmt5 is a maintenance enzyme *in vitro* with activity on hemimethylated substrates and no detectable activity on unmethylated substrates under a range of conditions.

5mC Is Propagated by a Maintenance Mechanism *In Vivo*

Although Dnmt5 displays the *in vitro* substrate specificity expected for a maintenance-type enzyme, it may function *in vivo* with cofactors that could change its properties. A way to address this question is to remove Dnmt5 from a cell, allow methylation to be lost, and then re-introduce the enzyme. If Dnmt5 behaves as a maintenance enzyme in cells, DNA methylation should not be

Cell
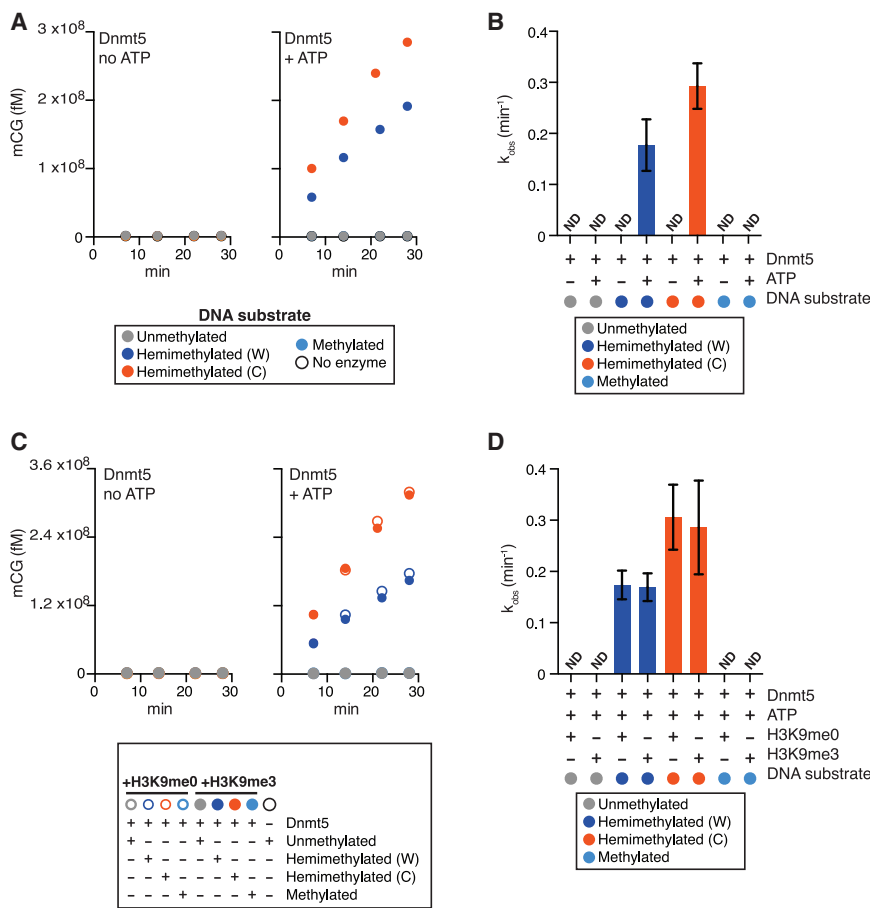
Figure 2. Hemimethylated DNA but Not Unmethylated DNA Is a Substrate of Purified Dnmt5

(A) Reaction curves of incorporation of label from S-[methyl-^{3}H]-adenosyl-L-methionine into the indicated DNA oligonucleotide substrates using 30 nM Dnmt5 and 5 μM DNA substrate, with and without 1 mM ATP.

(B) Initial rates of Dnmt5 activity on DNA substrates described in (A). n = 3–4; error bar represents SD. ND, not detected.

(C) Reaction curves of incorporation of label from S-[methyl-^{3}H]-adenosyl-L-methionine into the indicated DNA oligonucleotide substrates using 30 nM Dnmt5 and 5 μM DNA. Reactions were performed in the presence or absence of H3K9me0/3 peptide (5 μM) and ATP (1 mM).

(D) Initial rates of Dnmt5 activity on DNA substrates described in (C). n = 3–4; error bars represent SD. ND, not detected.

See also Figure S3.

was observed only after 25 generations, which correlated with the depletion of the mRNA produced by the pGAL-$DMT5$ overexpression construct (Figure S4A). However, upon transfer to galactose-containing inducing medium, we did not observe restoration of 5mC (Figure 3C; corresponding quantification shown in Figure 3D).

We performed an analogous analysis using a different approach. We first deleted a portion of the $DMT5$ gene followed by the re-introduction of the missing DNA sequence (Figure 4; the re-introduced allele is termed RI-$DMT5$; see STAR Methods). Although the expression of RI-$DMT5$ was comparable to that of a wild-type allele, we were not able to detect 5mC by Southern hybridization (Figure 4A). WGBS sequencing of the $dmt5\Delta$ strain revealed no methylation on any CG sites genome-wide (Figure 4B). Analysis of two independently derived RI-$DMT5$ strains revealed that methylation is also globally absent save for two sites detected in each strain (Figures 4B and S4B). To examine methylation using an independent global method, we employed attomole-sensitivity mass spectrometry (LC-MS/MS) to measure 5mC. Analysis of 5mC levels present in gDNA extracted from wild-type, $dmt5\Delta$, and RI-$DMT5$ revealed 0.33% 5mC (5mC/dG) in wild-type, none detectable in DNA extracted from the $dmt5\Delta$ strain, and trace amounts (detectable but below quantification limits) in DNA extracted from the RI-$DMT5$ strain (Figure 4C). This trace activity suggests that Dnmt5 may able to act at a very low level on unmodified DNA, an event that would be then "remembered" by its maintenance activity. Investigations of methylation dynamics in wild-type cells are described further below.

We observed that H3K9me is still globally maintained in both the $dmt5\Delta$ strain as well as in the RI-$DMT5$ strain, although the distribution is altered compared to wild-type depending on the genomic region with an increase of the signal at subtelomeric
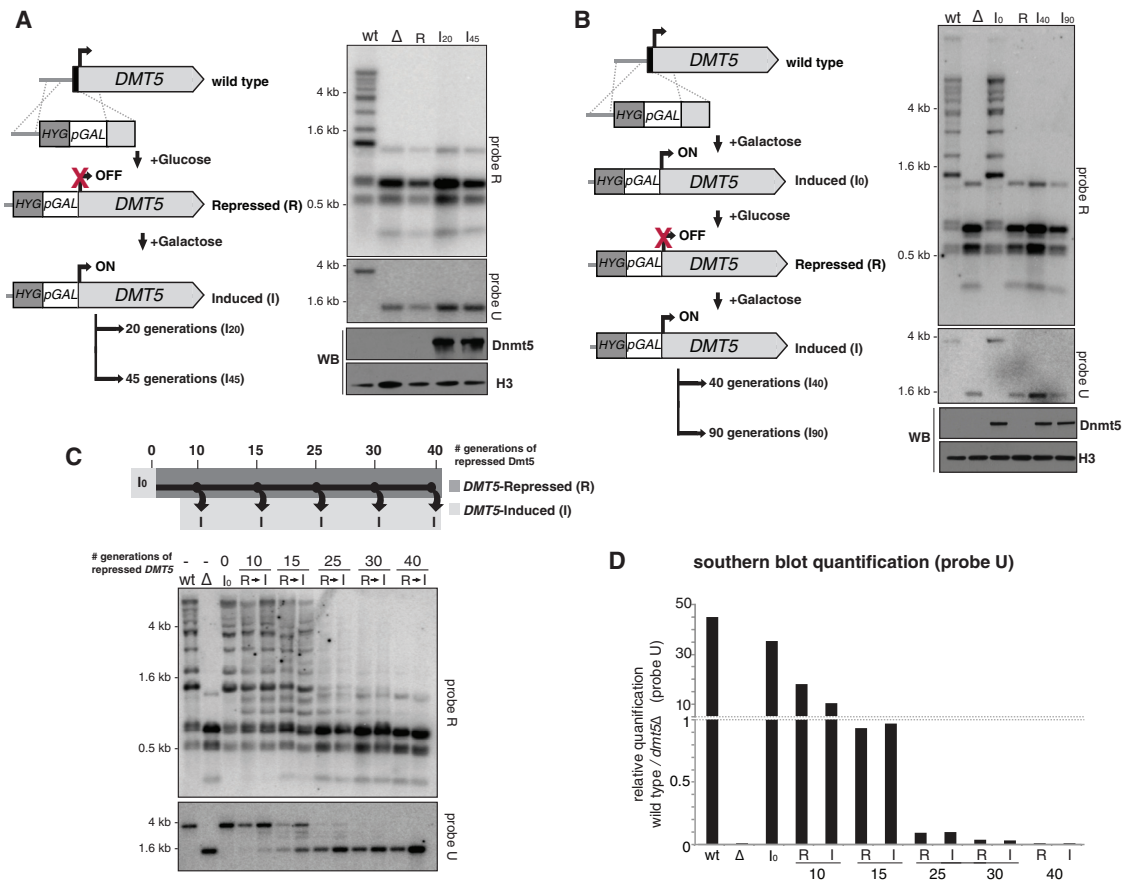
globally restored because its hemimethylated substrate would be lost.

We placed the gene for Dnmt5 under the control of a galactose-inducible promoter (Figure 3A; pGAL-$DMT5$). This allele was tagged at its N terminus with a 2XFLAG epitope tag. In the first series of experiments, the corresponding targeting construct was introduced by transformation and recipient cells were selected under conditions repressive for the p$GAL7$ promoter (R). After induction of the Dnmt5 protein using the inducer galactose for 20 (I$_{20}$) or 45 generations (I$_{45}$), we were not able to detect 5mC by Southern hybridization, although expression of the construct was evident (Figure 3A). To assess if the pGAL-$DMT5$ allele is functional, the targeting construct was transformed into wild-type cells, but this time recipient cells were selected under conditions that induce p$GAL7$ (Figure 3B). The 5mC pattern of the strain under these inducing conditions (I$_0$) is indistinguishable from that of wild-type indicating that the inducible, tagged allele is functional. Subsequently, upon repression of pGAL-$DMT5$, 5mC was lost. However, again, re-induction of Dnmt5 for 40 (I$_{40}$) or 90 generations (I$_{90}$) did not restore detectable 5mC. To determine whether 5mC could be restored if it was only partially depleted from cells, for example by a nucleation-spread mechanism, we set up a time course experiment in which we repressed pGAL-$DMT5$ for limited number of generations (R) and re-induced (I) it. In repressing conditions, substantial decrease of 5mC signal

Figure 3. Transient Repression of Dnmt5 Results in Sustained Loss of 5mC

(A) OFF-ON scheme. Knock in of a galactose-regulated *GAL7* promoter and epitope tag upstream of the Dnmt5 coding sequence is shown. Strains were selected under repressing (glucose) conditions (R). Dnmt5 was induced by addition of galactose for 20 (I_{20}) or 45 generations (I_{40}). Strains were analyzed by Southern hybridization as in Figure 1. WB, western blot indicating levels of FLAG-Dnmt5 and histone H3.

(B) ON-OFF-ON scheme. Knock in of a galactose-regulated *GAL7* promoter and FLAG epitope-tag upstream of the Dnmt5 coding sequence is shown. In contrast to (A), selection for the knock in allele was carried in presence of galactose (induced-I_0). Dnmt5 was then repressed to produce a loss of DNA methylation (R) and subsequently induced for 40 (I_{40}) or 90 generations (I_{90}). Strains were analyzed by Southern hybridization.

(C) ON-OFF-ON scheme with variable OFF times. Strain produced under inducing conditions as in (B) was subject to the indicated number of generations of repression (R) followed by 20 generations of induction. Southern analysis was performed as in (A).

(D) Quantification of the ratios of the band at 3.5 kb (wild-type) divided by the band at 1.5 kb (*dmt5Δ*) for probe U calculated for each sample in (C) using ImageJ. See also Figure S4.

regions (Figures 4D and S4C). Furthermore, ChIP-seq experiments demonstrate that the RI-Dnmt5 protein still localizes to heterochromatic regions (Figures 4E and S4D).

To test explicitly if *RI-DMT5* is capable of efficiently maintaining DNA methylation, a wild-type strain whose centromere 13 was marked with a linked drug resistance marker (*CEN13::natR*) and crossed to the *RI-DMT5* (Figure 4F). From the cross, meiotic progeny bearing the *RI-DMT5* allele and the marked centromere were selected. 5mC was assayed by Southern hybridization using a probe specific for *CEN13*. All progeny tested displayed a wild-type 5mC pattern (Figure 4F), indicating that the *RI-DMT5* allele is functional when provided with a methylated substrate.

In many species, sexual reproduction is required for the establishment of 5mC by one or more *de novo* enzymes. It has been reported that a tandemly repeated *URA5*-marked transgene

can be heritably silenced after mating and meiosis/sporulation in *C. neoformans*, but this silencing occurs via RNAi and not via DNA methylation (Wang et al., 2010). Nonetheless, to determine whether an unmethylated centromere can be methylated during sexual reproduction, we performed two genetic crosses and analyzed the meiotic progeny. In a control cross, we crossed a *CEN13::natR* strain with a *dmt5Δ* strain and analyzed three progeny that contained both the marked *CEN13* and wild-type *DMT5* allele by Southern hybridization using a probe specific for *CEN13*. As expected, 5mC was maintained at the methylated centromere (Figure 4G). In the experimental cross, we crossed a wild-type strain to one harboring a *dmt5Δ* allele that also carries *CEN13::natR.* We again analyzed three progeny that express the wild-type *DMT5* gene and the marked centromere, which entered the cross in an unmethylated state. We observed that

Cell
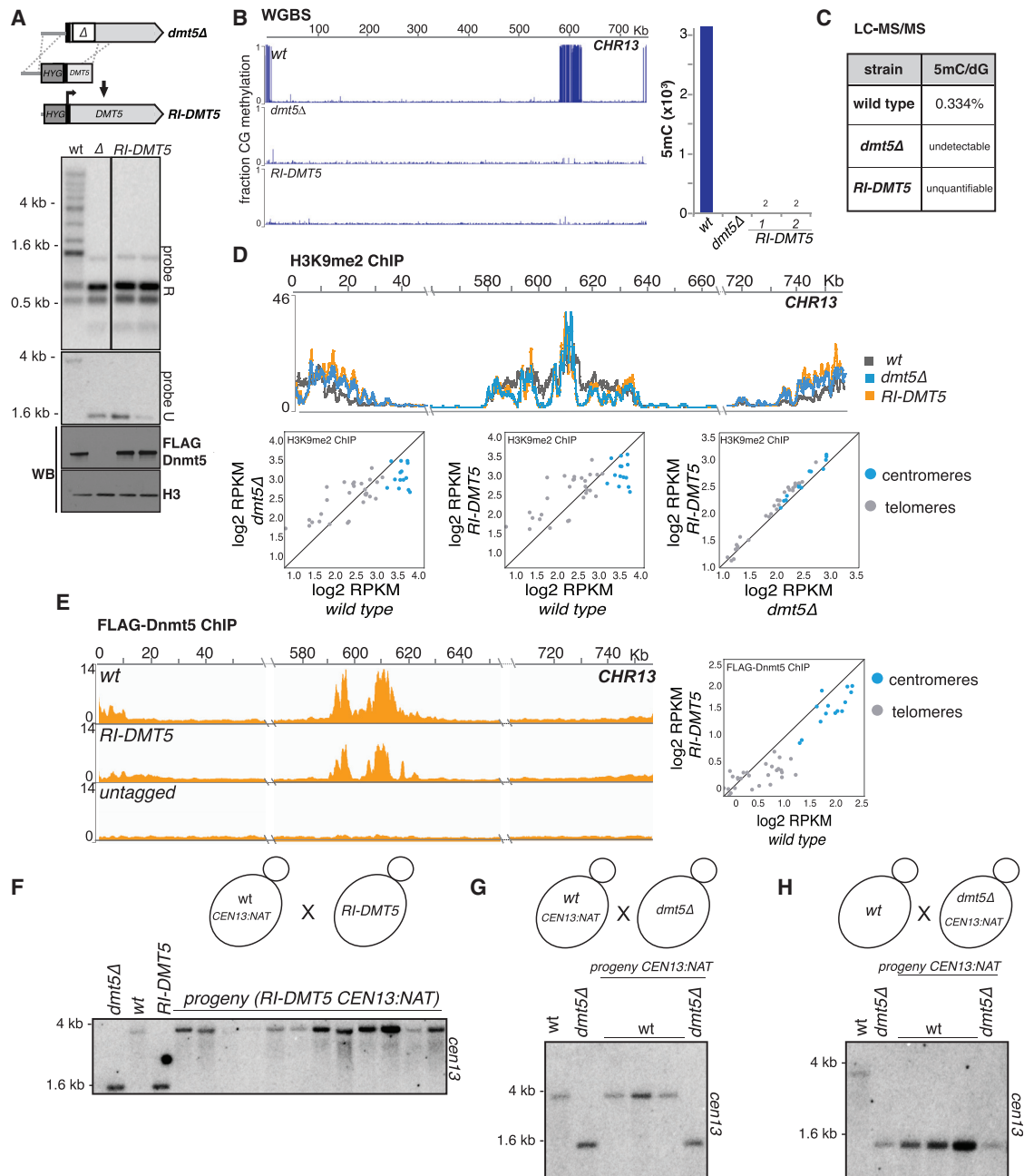
Figure 4. Genetic Loss of DNA Methylation Is Mitotically and Meiotically Irreversible

(A) Re-introduction (RI) scheme. A partial knock out of the *DMT5* locus was produced and then corrected by homologous recombination and N-terminally tagged as in Figure 2 to produce the *RI-DMT5* strain. A hygromycin resistance marker (*hygR*) is inserted upstream of the *DMT5* gene during this process. 5mC was assessed as in Figure 1.

(B) WGBS analysis of *RI-DMT5* strain. Shown are data for chromosome 13 for the indicated genotypes. Bar graph shows number of called symmetrically methylated CpGs for the indicated genotypes. Only 2 CpGs are detected in each RI-DMT5 isolate.

(C) Quantification through mass spectrometry (LC-MS/MS) of 5mC in DNA of wild-type, *dmt5Δ*, and *RI-DMT5* strains.

(D) Analysis of H3K9me2 signals in wild-type, *dmt5Δ*, and *RI-DMT5* strain. ChIP-seq signal for H3K9 methylation shown as normalized to signal for the H3 ChIP. Shown are the data for centromere 13 for the indicated genotypes (top) and the pairwise comparison of the total RPKM counts for each centromere (blue) and telomere (gray) (scatterplot, bottom).

(E) Analysis of Dnmt5 recruitment in the *RI-DMT5* strain. ChIP-seq signal for FLAG-Dnmt5 were normalized to signal for the input samples. Shown are the data for centromere 13 for the wild-type, *RI-DMT5*, and a strain not expressing FLAG-tagged proteins.

(legend continued on next page)

such progeny remained unmethylated, indicating that Dnmt5 is able to maintain methylated DNA but does not re-establish on a centromere that had completely lost methylation (Figure 4H). Thus, sexual reproduction does not lead to the efficient restoration of 5mC. Finally, we examined cells exposed to several stress conditions to determine whether they could reestablish methylation in the *RI-DMT5* strain, and we did not observe restoration of methylation (Figure S5A). To test whether there was a restoration of fitness, we sought a facile plate phenotype for *dmt5Δ* cells. Hypomethylation of centromeric regions is associated with chromosome instability in ICF syndrome patients harboring mutations in the *de novo* methylase Dnmt3B (Hansen et al., 1999; Okano et al., 1999). Therefore, we tested our strains for sensitivity to the microtubule-depolymerizing drug thiabendazole (TBZ), a phenotype of chromosome instability mutants. Indeed, both *dmt5Δ* and *RI-DMT5* strains are TBZ-sensitive, indicating a role for 5mC in cellular fitness (Figure S5B).

To assess whether Dnmt5 is able to maintain DNA methylation independently on non-endogenous sequences, we methylated *in vitro* a targeting construct containing a DNA fragment rich in HpaII sites with the HpaII methyltransferase (that produces symmetric CmCGG) and integrated it into the *C. neoformans* genome at the right border of centromere 4 in order to place it in proximity to H3K9me-modified chromatin (Figures 5A–5C). DNA from the transformants was digested with the 5mC-sensitive HpaII restriction endonuclease, and qPCR was performed using a primer pair that spanned the cleavage sites to measure the fraction of a particular site that was methylated (Figure 5D). Analysis across seven different HpaII sites revealed that all were maintained *in vivo* (Figure 5D). Thus, there appear to be no obligate sequence requirements beyond CG dinucleotides for methylation to be maintained *in vivo*, a finding consistent with our biochemical results.

We tested next whether 5mC could spread *in vivo* by using the *in vitro* methylated sites as nucleation sites. We digested the genomic DNA with the 5mC-sensitive enzymes BstUI (CGCG) and Hpy99I (CGWCG) and performed qPCR across these sites. We did not observe a signal above background, indicating no detectable spread of 5mC to these nearby CG dinucleotides (Figure 5E). We obtained similar results *in vitro*: we incubated purified Dnmt5 with a dsDNA substrate containing either no CG sites, or two unmethylated CG sites, or two ummethylated CG sites accompanied by a symmetrically methylated CG site (Figure 5F). Although methylation products were detected upon incubation with a control bacterial DMT M.SssI, no signal was detected upon incubation with Dnmt5, indicating that the enzyme does not display detectable *de novo* activity even when presented with a nearby methylated site (Figure 5F).

The Gene for a *De Novo* Methylase Was Lost in an Ancestral Species

Given that Dnmt5 is the sole DNMT predicted by the *C. neoformans* genome, and seems to behave as a maintenance DNMT, how was 5mC established? To approach this question, we examined the genomes in the family of organisms to which *Cryptococcus neoformans* belongs, the *Tremellaceae* (Figure 6, Findley et al., 2009; Guerreiro et al., 2013). The genomes of species closely related to *C. neoformans*, including the human pathogen *Cryptococcus gattii* and the nonpathogens *Cryptococcus amylolentus* and *Cryptococcus wingfieldii*, harbor a single predicted DNA methyltransferase, which is orthologous to Dnmt5 (Figure 6A). The exception is *C. deuterogattii*, which has inactivated all transposons and lost both RNAi and 5mC (Yadav et al., 2018a). The next closest species, *Cryptococcus depauperatus*, lacks predicted DNA methyltransferases, also indicating loss of *DMT5* during its evolution (Figure 6A). Strikingly, more distant species encode both Dnmt5 and an uncharacterized predicted DNA methyltransferase, which we term DnmtX (genetic locus: *DMX1*; Figure 6A). This predicted protein contains a bromo-associated homology (BAH) domain and a Dnmt catalytic domain. This phylogenetic pattern indicates that the *DMX1* gene was lost in the common ancestor to *C. neoformans* and *C. depauperatus*, and *C. depauperatus* subsequently lost the *DMT5* gene. Given the fact that Dnmt5 is a maintenance methylase, we hypothesized that DnmtX is a *de novo* methyltransferase.

To test this hypothesis, we cloned the genes for DnmtX from *Kwoniella mangroviensis*, *Kwoniella bestiolae*, and *Kwoniella pini*. Each of these was placed under the control of the *pGAL7* galactose-inducible promoter, tagged at the N terminus with a hemagglutinin epitope tag, and then introduced as a transgene into a *C. neoformans* strain in which the gene for Dnmt5 was disrupted and then restored with a FLAG-tagged allele (*RI-DMT5*). We chose this approach because we predicted that establishment of 5mC by a DnmtX enzyme might be inefficient, especially if cofactors important for *de novo* methylation were not also introduced.

We induced expression of DnmtX in each strain using medium containing galactose (Figure S5C) and then extracted DNA. The levels of 5mC were assessed genome-wide by two technically orthogonal methods, methylated DNA immunoprecipitation followed by high-throughput sequencing (MeDIP-seq) and WGBS. In both assays, we observed the broad accumulation of 5mC in each of the three strains, primarily at centromeric regions (Figures 6B and S6). These data indicate that all three DnmtXs act as *de novo* methylases *in vivo*, providing strong evidence that the ancestral DnmtX also had this function.

(F) Genetic cross testing the functionality of RI-Dnmt5 protein. Centromere 13 of a wild-type strain was tagged by insertion of a nourseothrycin resistance gene 3.5 kb from its left end (*CEN13::natR*), and the resulting strain was crossed to a strain harboring the *hygR*-marked *RI-DMT* allele. Progeny strains doubly resistant to hygromycin and nourseothrycin were selected. Strains were analyzed as described in Figure 1 using probe U, which is specific for chromosome 13.

(G) Control cross. Dnmt5-harboring cells containing *CEN13::natR* were crossed to *dmt5Δ* cells. Progeny harboring the Dnmt5 and *CEN13::natR* were analyzed as in (C).

(H) Experimental cross testing whether sexual reproduction enables *de novo* methylation. *dmt5Δ* cells harboring *CEN13::natR* were crossed to wild-type cells. Progeny harboring a wild-type *DMT5* gene and *CEN13::natR* were analyzed as in (C).

See also Figures S4 and S5.

Cell
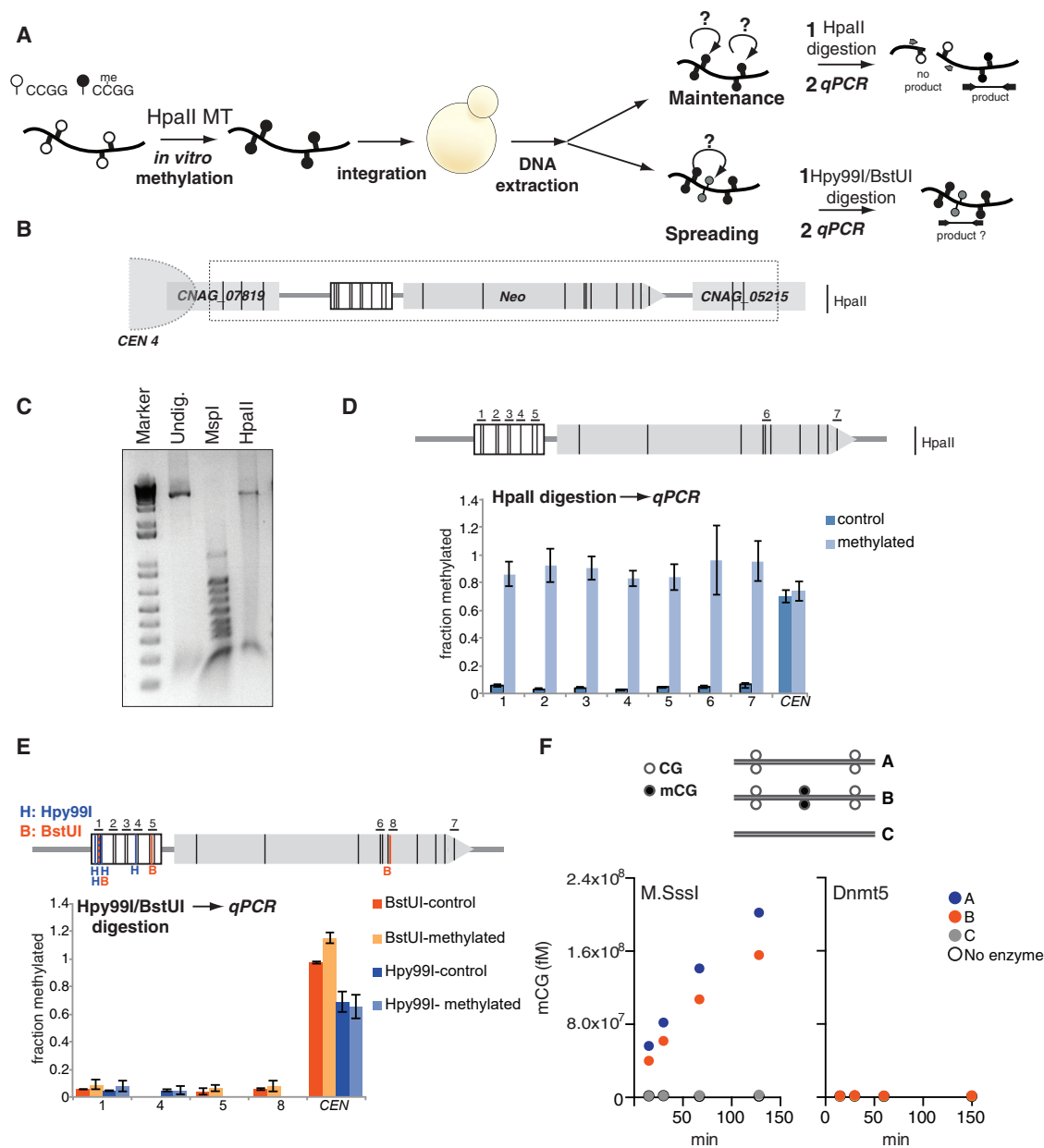
Figure 5. Dnmt5 Maintains *In Vitro* Methylated DNA

(A) A DNA fragment was methylated *in vitro* using the HpaII DNA methyltransferase (HpaII MT) and subsequently integrated into *C. neoformans*. gDNA extracted from isolates that correctly integrated the fragment, was digested with the restriction enzyme HpaII. To test the maintenance of the *in vitro*-methylated 5mC, qPCR was performed on region spanning the restriction sites. To test the ability 5mC to spread, gDNA was digested with either Hpy99 or BstUI and qPCR was performed across these sites. White and black lollipops represent unmethylated and methylated sites, respectively.

(B) Schematic representation of the construct and its integration adjacent to centromere 4 (*CEN4*). Vertical black lines represent the site (CCGG) recognized by the HpaII methyltransferase.

(C) Gel of undigested *in vitro* methylated DNA (Undig.) digested with the methylation insensitive restriction enzyme MspI (MspI) or its methylation-sensitive isoschizomer HpaII (HpaII). M, marker.

(D) qPCR performed on gDNA extracted from yeast transformed with the control unmethylated (control, dark blue) and *in vitro* methylated construct (methylated, light blue). gDNA was digested with HpaII and qPCR was performed as described in A on regions 1–7. qPCR was normalized over a region of the genome that is not cleaved by HpaII. *CEN*, positive control-region on centromere 13 that contains a methylated HpaII site. Black vertical lines indicate site of HpaII cleavage. (n = 3, error bars represented as SD).

(E) gDNA digested with either the restriction enzymes Hpy99I or BstUI. qPCR was performed on gDNA containing the unmethylated control (dark blue, Hpy99I-control and orange, BstUI-control) or the *in vitro* methylated construct (light blue, Hpy99I-methylated and yellow, BstUI-methylated). qPCR was normalized over a

(legend continued on next page)

Cell

Experimental Evolution Reveals Fidelity of 5mC Maintenance

The loss of the ancestral DnmtX gene occurred prior to the divergence of *C. neoformans* and *C. depauperatus*, but after the divergence of *Tremella mesenterica* and *Kwoniella heveanensis* (Figure 6A). The divergence time for *C. neoformans* and *C. gattii* is estimated to be between 34 and 49 mya (D'Souza et al., 2011; Ngamskulrungroj et al., 2009). Given the phylogenetic relationships, the divergence time of *C. neoformans* and *C. depauperatus* would be considerably more ancient. It has been estimated that the common ancestor of *C. neoformans* and *T. mesenterica* lived 153 mya (Floudas et al., 2012). Thus, the *DMX1* gene was likely to have been lost between roughly 150 and 50 mya. Taken together, these data indicate that 5mC has been maintained by Dnmt5 over a remarkably long period of time in the population that gave rise to the pathogenic *Cryptococcus* species complex.

To begin to understand how 5mC has persisted since the ancient loss of DnmtX, we performed an experimental evolution experiment in *C. neoformans*. A single colony from wild-type cells (generation 0, or g0) was grown in liquid rich media for 120 generations without bottlenecking the population. Dilutions of the final culture were plated to form single colonies on agar plates. Two independent clones were selected, and replicate WGBS analysis was performed on their genomic DNA as well as DNA extracted from a portion of the parental strain (Figure 6C). We detected 50 and 38 loss events, respectively, (Figures 6D and 6E), indicating that 99.0% and 99.2% of 5mC sites were maintained after 120 generations. Only two sites were lost in both colonies, an overlap expected by chance (Figure 6E; Table S1). Assuming all sites are equally maintained and assuming first order kinetics, this would correspond to a per-generation loss rate of 9.3×10^{-5} for each 5mC site. Consistent with the reintroduction experiments described above, we observed five gain events in each experiment (nonoverlapping; Figure 6E; Table S1). Given the number of unmethylated sites, this corresponds to a per-generation gain rate of 4.5×10^{-6}. Because the gain events occur at a rate that is ~20-fold below that of the loss events, the system is not at equilibrium under laboratory conditions. Supporting this conclusion, Wald binomial analysis of the experimental evolution data yielded a 95% confidence interval for the value of (methylation gains/gains + losses) of (0.0424, 0.1657) that allows us to confidently reject a model with a higher gain rate because the upper bound corresponds to only ~17% of events being gains. Thus, selection on methylation levels (across centromeres or specific regions) likely operates in natural populations to maintain the steady state of methylation (see Discussion).

Evolutionary Analysis of 5mC Patterns Reveals Conservation

To investigate further the inheritance of 5mC patterns, we examined its evolutionary conservation. To accomplish this, we first sought to align centromeric regions from isolates that had diverged by a significant period of time. *C. neoformans var. grubii*, the species of the reference H99 strain, has four well-characterized clades, VNI, VNII, VNBI, and VNBII, which shared a common ancestor ~4.7 mya (Ngamskulrungroj et al., 2009) (Figure 7A). Short-read sequences are available for several hundred strains (Desjardins et al., 2017), however, centromeric regions, where the bulk of 5mC resides, are not assembled as they contain repetitive sequences, including six families of *gypsy* and *copia* retrotransposons, Tcn1–Tcn6 (Janbon et al., 2014). Therefore, we used long read Oxford nanopore (ONT) sequencing to assemble near-complete genomes of two isolates from each of the four clades (Figure 7B). Dot plot analysis of centromeres revealed widespread relative rearrangements of sequence (insertions, deletion, inversions), precluding end-to-end multiple sequence alignment (Figure S7A). We instead adopted a local alignment approach to obtain a series of pairwise alignments between segments of centromeres of the eight assembled genomes. On average, 88.5% of centromeric sequence from a given strain could be aligned to the homologous centromere of at least one other strain, with a median alignment length of 607 bp (Figures S7B and S7C). We performed replicate WGBS analysis on each strain and overlaid these data onto the alignments. An example of this analysis is shown in Figure 7C, which displays the 5mC status of CG sites that can be aligned between a centromere of an arbitrarily chosen strain (A1_35_8) and the homologous centromere on other seven strains. Next, for each centromeric CG dinucleotide in each of the strains, we determined the number of strains (N) to which a given CG nucleotide could be identified by alignment and then calculated the fraction of CGs that were methylated in all N strains (Figure 7D). Statistical analysis revealed a significant excess of sharing of methylated sites compared to the null model, regardless of the strain used as a reference and the number of strains to which a centromeric CG dinucleotide could be aligned (see STAR Methods).

Given the concomitant loss of 5mC and intact transposons in *C. deuterogattii*, it seems likely that important targets of DNA methylation are the Tcn elements described above. We tested whether these elements display higher levels of 5mC than non-Tcn centromeric sequences and whether these patterns are conserved. In all eight strains, we identified sequences at centromeres that are highly homologous to Tcn1–Tcn6, including a fraction harboring two LTRs, which are candidates for active transposons (Figure S7D). Due to their length and repetitiveness, Tcn elements are not well-covered by WGBS; hence we supplemented the WGBS data using methylation calls from ONT long-read data, which we found to be highly correlated with those produced by WGBS (Figures S7E and S7F). Consistent with selection acting on 5mC in Tcn elements, we observed a significantly higher fraction of methylated CGs on Tcn-related sequences than non-Tcn sequences within centromeres (Figure 7E).

region of the genome not cleaved by Hpy99I or BstUI. In the scheme (top), black vertical lines indicate sites of HpaII cleavage, blue vertical lines indicate sites of Hpy99I, and orange vertical lines indicate the sites for BstUI. (n = 3, error bars represented as SD).

(F) Double-stranded DNA containing two unmethylated CG (A) or unmethylated CG surrounding a central methylated CG site (B) or without any CG dinucleotide (C) were used as substrate of DNA methylation kinetics with M.SssI (left) or 30 nM Dnmt5 (right).

See also Figures S5 and S6.

Cell
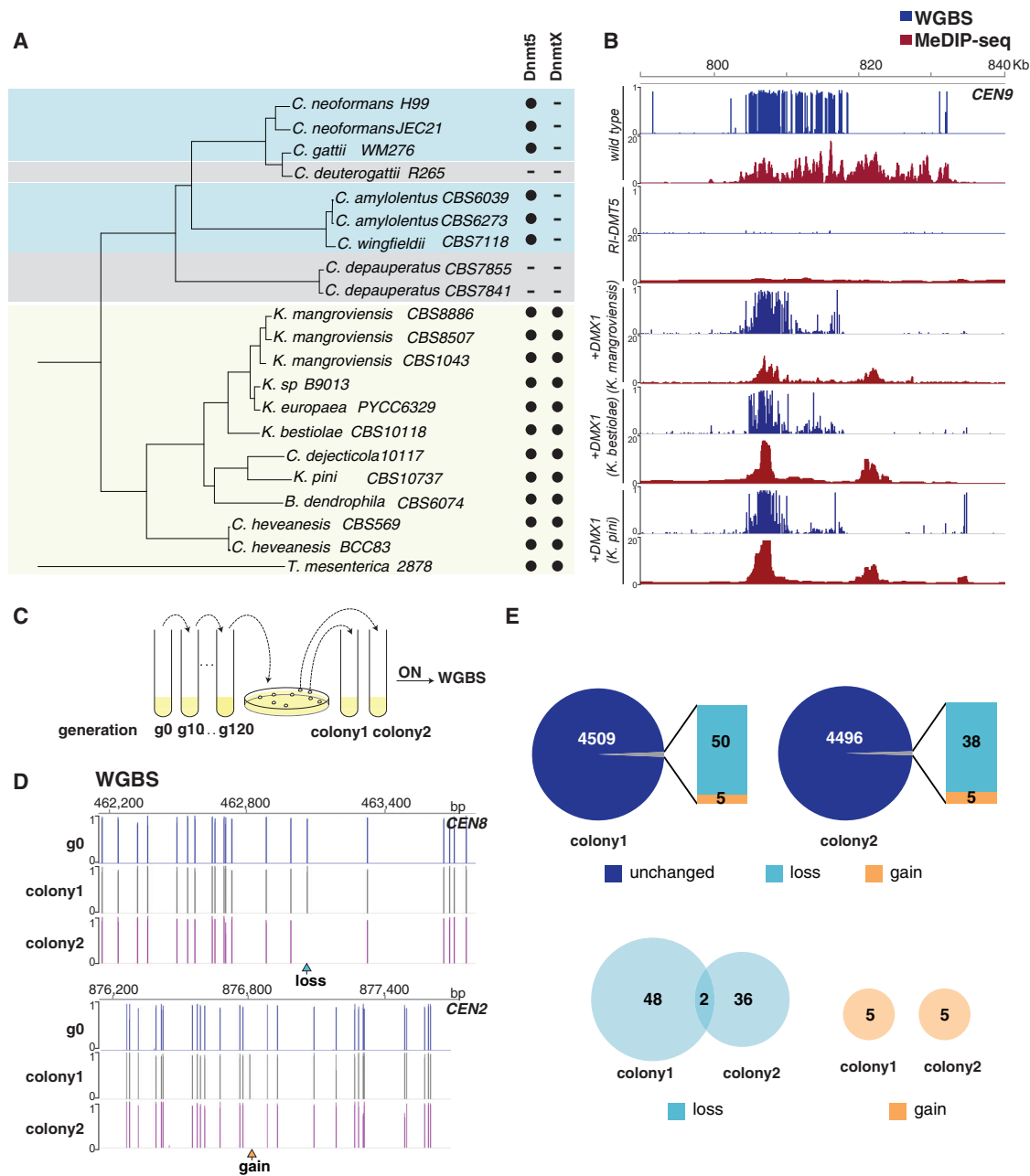
Figure 6. Ancient Loss of Gene Coding for a *De Novo* Dnmt in an Ancestor of *C. neoformans* and Laboratory Evolution Experiments

(A) Phylogenetic analysis of whole-genome sequences. Phylogeny is based on the analysis of whole-genome sequences (C.A.C., unpublished data). The genomic presence of a gene coding for orthologs of Dnmt5 and of a second DNMT (DnmtX) is indicated. Gene for DnmtX was lost prior to the divergence of *C. neoformans* and *C. depauperatus*, and gene for Dnmt5 was subsequently lost in the *C. depauperatus* and *C. deuterogattii* lineages.

(B) Expression of extant DnmtX-encoding genes in an *RI-DMT5 C. neoformans* strain triggers *de novo* 5mC. Three DnmtXs from indicated organisms were expressed in a *RI-DMT5* strain. 5mC was assessed by MeDIP-seq (red) and WGBS (blue). MeDIP signal is shown as ratio over *dmt5Δ*. Shown are data for centromere 9.

(C) Scheme for laboratory evolution assay. Wild-type cells (g0, generation 0) were propagated in rich media and for 120 generations and then plated for single colonies. Two independent isolates (colony 1 and colony 2) were grown overnight and the gDNA extracted processed in technical duplicates for bisulfite sequencing (BS-seq).

(D) WGBS of a portion of the starting culture (g0) and the two different isolates grown for >120 generations (colony 1 and colony 2). Shown are the reproducible sites between technical replicates on centromere 9. Sites that are present in wild-type and absent after >120 generations are marked as "loss," while sites absent in wild-type but occurring in one of the isolates are marked as "gain."

(legend continued on next page)

Cell

Because Tcn3-related sequences displayed the longest median length and the most LTR-flanked elements (Figure S7D), we performed additional analysis on these elements. We aligned each CG dinucleotide from the longest (top quartile) Tcn3-related sequences and overlaid the 5mC data. Strikingly, for the vast majority of CG dinucleotides aligned (Figure 7F, vertical columns), we observed high sharing of methylation (mean fraction = 0.7, $p < 10^{-16}$) compared to a control model for Tcn3 of independently permuted 5mC sites (mean fraction = 0.4149) indicating sharing of 5mC patterns across strains above expectation. Note that in several cases we observed similar patterns of 5mC on different Tcn3 elements within a single strain (Figure 7F); this is more likely the result of recombination events such as unequal crossing over rather than retrotransposition as the latter would be expected to erase methylation patterns. The small number of Tcn3s that have little methylation may have undergone retrotransposition at some point that would have erased 5mC; alternatively these could be inactive elements upon which selection is no longer operative.

DISCUSSION

A Maintenance DNA Methylation System in the Fungal Kingdom

A fungal *de novo*-maintenance pair has not been functionally or biochemically identified. Our studies identified one-half of this pair, Dnmt5, in the human pathogenic yeast *C. neoformans*. To our surprise, Dnmt5 is a maintenance enzyme *in vitro* with extraordinarily high specificity for hemimethylated substrates. Dnmt5 is not closely related to the conserved Dnmt1 maintenance methylase, and unlike Dnmt1, Dnmt5 requires ATP for activity and displays considerably higher specificity for hemimethylated DNA than Dnmt1 (Bashtrykov et al., 2014). Nonetheless, the chromatin signals that promote maintenance methylation in *C. neoformans* are shared with mammals and plants.

Ancient Loss of a *De Novo* Enzyme yet Persistence of 5mC

Given that Dnmt5 is a highly specific maintenance enzyme *in vitro*, it is unexpected that Dnmt5 is the only predicted cytosine methylase encoded by the *C. neoformans* genome. One possibility would be that Dnmt5 acts differently *in vivo*. However, we observed that methylation loss after depleting cells of Dnmt5 was not restored upon re-introduction of the enzyme. This was true even if Dnmt5 was introduced through a genetic cross. Tandem repeats are silenced in *C. neoformans* during sexual reproduction in a process called sex-induced silencing (SIS) (Wang et al., 2010). However, SIS is an unstable RNAi-based phenomenon that does not result in methylation of repeats (Wang et al., 2010). Thus, it appears that there is no efficient *de novo* methylation mechanism in this organism. Consistent with these observations, we determined that the ancestor of

the pathogenic *Cryptococcus* species complex harbored a second DNMT homolog. Our functional experiments demonstrate that extant DnmtX from three different species are each capable of triggering widespread DNA methylation *de novo* in *C. neoformans*, indicating that DnmtX is a *de novo* enzyme. Thus, we conclude that 5mC has endured in *C. neoformans* for millions of years through the action of Dnmt5.

### Inheritance, Variation, and Selection in an Epigenome

Although we did not observed restoration of 5mC globally when Dnmt5 was removed and re-introduced, we did observe two apparent *de novo* events in two independent experiments. A caveat of this experiment is that cells that lack methylation to begin with would not be able to recruit Uhrf1 efficiently and thus may not well represent wild-type. Thus, we performed an experimental evolution experiment in wild-type cells. This work revealed preservation of 5mC patterns with high fidelity. We observed loss events, which occurred at a low rate, and gain events, which occurred at a considerably lower rate. The gain events may be produced by the action of Dnmt5 on unmethylated DNA, the activity of some other cellular methylase or gene conversion events that produce hemimethylated heteroduplex DNA that would then be a favored substrate for Dnmt5. The latter is plausible as the nonreciprocal transfer of 5mC has been observed during meiosis in *A. immersus* (Colot et al., 1996). Regardless of their source, these gain events, which seem to be rare and randomly distributed, provide a mechanism to inject new sites of methylation.

Because the rate of methylation loss is ∼20-fold higher than that of gain, we infer that the system is not at equilibrium. Our estimate of the loss rate for 5mC at a given site (∼9.3×10^{-5}/ generation) implies a half-life of methylation at a given site to be ∼7,500 generations. Even if we assume that *C. neoformans* divides 100-fold more slowly in the wild than in the laboratory, this would correspond to half-life of only ∼130 years. Our analysis of eight isolates of *C. neoformans var. grubii* that span the four major clades of this pathogen and diverged ∼5 mya (Figure 7), demonstrate significant sharing of 5mC sites across centromeres and across annotated dual LTR-harboring Tcn3 elements. Given the half-lives of methylated CG sites predicted by the laboratory evolution experiments, the observed sharing of patterns over million-year timescales cannot not be explained by shared ancestry without selection (i.e., identity-by-descent).

Instead, we propose that natural selection maintains methylation to higher levels than would be expected from a neutral loss-gain process. Supporting this model, our analysis of *C. neoformans var. grubii* epigenomes demonstrated higher fraction of methylated CGs (median = 0.59) in dual LTR-harboring centromeric retrotransposon (Tcn) sequences than in non-transposon centromeric sequences (median = 0.24). A simple population genetics model of the evolution of codon bias can

---

(E) Differential methylation analysis (DMS) was performed to compare g0 with the propagated isolates. The majority of reproducible sites remained unchanged (dark blue) and only a small fraction showed variation compared to g0. In light blue is shown the number of sites that have been lost and in orange are shown novel sites arising after propagation.

(F) Number and relationship of shared lost (left) and gained (right) sites between colony 1 and 2.

See also Figures S5 and S6 and Table S1.

Cell
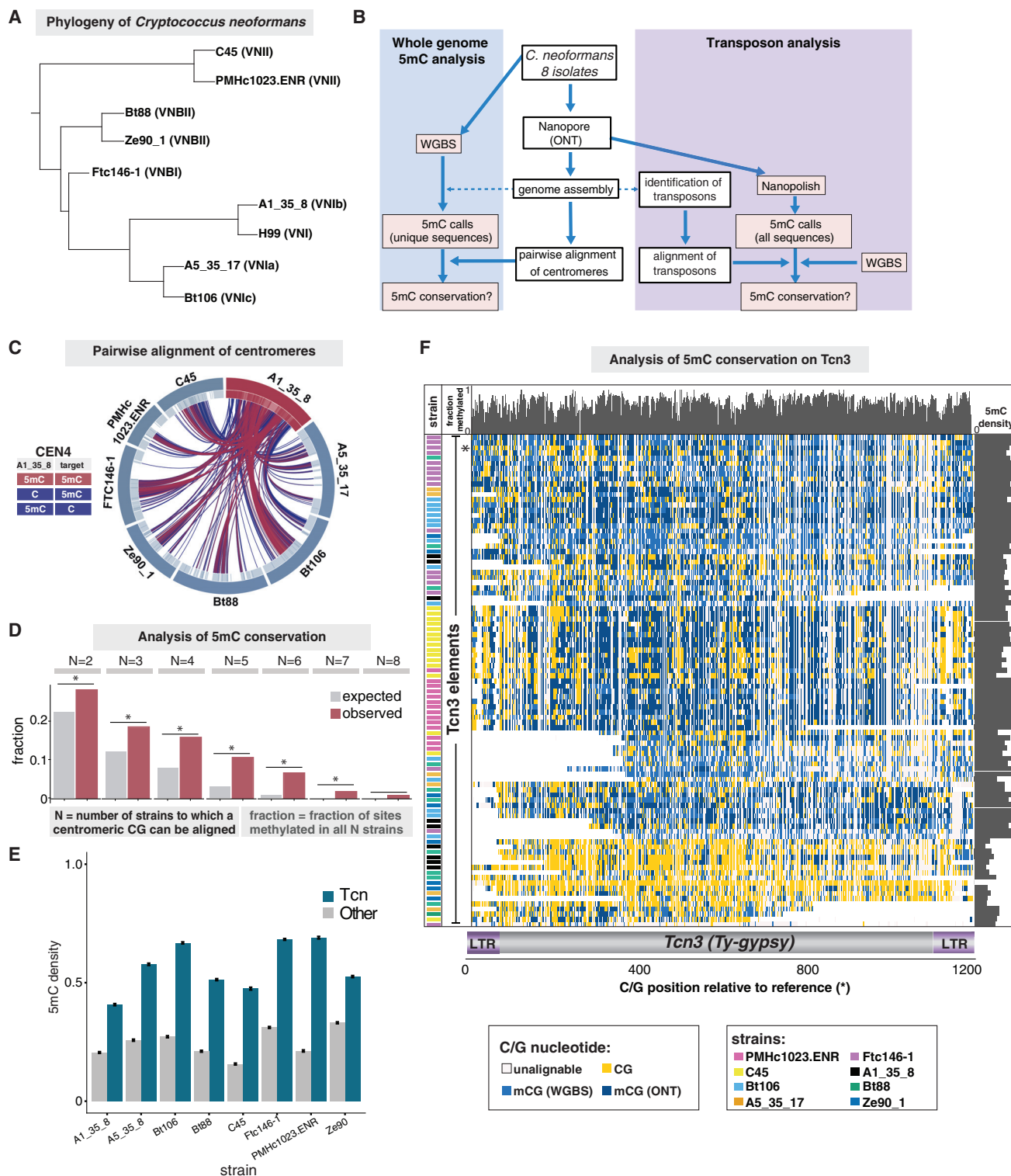
Figure 7. Cytosine Methylation Sites Are Conserved between Isolates of *C. neoformans*

(A) Phylogenetic relationships between isolated of *C. neoformans*. Phylogeny is based on the analysis of whole-genome sequences.

(B) Pipeline for the analysis on the evolutionary conservation of 5mC.

(C) Circos plot representing the pairwise alignment of *CEN4* in A1_35_8 (red segment) with the centromeres of the other seven isolates (blue segments). The red lines represent CG dinucleotides that are methylated in both A1_35_8 and in target centromeres. The blue lines correspond to CG sites that lost 5mC in either A1_35_8 or the compared centromere. Light gray shading represents regions that are both alignable and display coverage in WGBS data.

(legend continued on next page)

Cell

be adapted to estimate the expected fractions with and without selection (modified from Hershberg and Petrov, 2008),

$$P(methylated) = \frac{1}{1 + \left(\frac{\mu loss}{\mu gain}\right) * e^{-2N_e s}},$$

where P(methylated) is the fraction of sites that display the methylated state, μ_{gain} and μ_{loss} are the rates of epimutation from unmethylated and methylated states to the opposite states, respectively, N_e is the effective population size, and s is the selection coefficient for the methylated state. Based on the observed rates of methylation loss and gain, the null model ($s = 0$) would predict an equilibrium fraction of methylated CGs of 0.047. This is much lower than the observed fraction for Tcn elements (0.59), consistent with a role for purifying selection. This calculation assumes that the relative rates of methylation loss and gain measured in the laboratory generally reflect those of wild populations and that this ratio is not dramatically different across different centromeric intervals. We note that non-Tcn CG sequences of the centromeres have a higher fraction of methylation than predicted by a neutral model, which may reflect a role for CG methylation in centromere function as inferred from the TBZ sensitivity of cells lacking 5mC. Together, our experimental and observational evolutionary studies support with a model in which 5mC is inherited with high fidelity and that 5mC loss and gain events in *C. neoformans* act in a manner akin to mutation in DNA sequence by producing random variation upon which natural selection acts. As with the evolution of codon bias, selection is likely to be weak with no requirement that any specific CG site be in the methylated state.

We did not observe sequence specificity beyond CG dinucleotides for 5mC occurrence in *C. neoformans*: an alignment of methylated sites across the eight genomes described in Figure 7 reveals only a CG motif (Figure S7G). Consistent with this finding, we found that introduced *in vitro* methylated DNA of bacterial origin is well-maintained by Dnmt5 *in vivo*, and purified Dnmt5 methylates artificial DNA sequences *in vitro* harboring hemimethylated CG dinucleotides. These findings argue that sharing of methylation sites across phylogeny we observe is not explained by a detectable requirement by Dnmt5 for particular sequences beyond hemimethylated CG dinucleotides.

Perspective

As Dnmt5 orthologs are widespread in the fungal and protist kingdoms, we speculate that the unique properties this enzyme endow organisms that harbor it to also undergo epigenome evo-

lution over long timescales. We anticipate that the availability of a tractable haploid yeast system for investigating a highly specific maintenance cytosine methylation system will continue to provide insights into the mechanistic underpinnings and biological roles of epigenetic memory.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- LEAD CONTACT AND MATERIALS AVAILABILITY
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Yeast growth and manipulations
- METHOD DETAILS
  - Genetic crosses
  - DNA extraction and Southern analysis
  - Stress conditions experimental procedures
  - Protein extraction and western analysis
  - Affinity purification and mass spectrometry analysis
  - Bisulfite conversion and library construction
  - Chromatin immunoprecipitation
  - Methylated DNA immunoprecipitation (MeDIP)
  - *In vitro* methylation and 5mC-qPCR assay
  - Experimental evolution
  - Recombinant protein expression and purification
  - Histone peptide array binding assay
  - Fluorescence polarization binding assay
  - Gel mobility shift assay
  - DNA methyltransferase assay
  - Liquid chromatography mass spectrometry (LC-MS/MS) method to quantify 5mC
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - ChIP-seq analysis
  - WGBS analysis
  - Experimental evolution 5mC analysis
  - Nanopore sequencing of eight C. neoformans var grubii isolates
  - WGBS analysis of eight C. neoformans var grubii isolates
  - Analysis of 5mC conservation
  - Transposon analysis
  - Phylogenetic methods
  - Mass Spectrometry analysis
- DATA AND CODE AVAILABILITY

(D) 5mC conservation analysis. Each CG dinucleotide was aligned and sorted into seven different categories (N) according to the number of strains in which the CG could be aligned (N = 2 to 8; N includes the reference strain). The fraction of CGs in which all N alignable CGs are methylated (red) is higher than that expected by chance (gray) (*p < 1 × 10^{-4}).

(E) Comparison of the mean methylation levels on centromeric sequence containing annotated full-length (LTR-LTR) Tcn elements (in blue) and other centromeric sequences (gray) from different isolates of *C. neoformans*.

(F) Alignment of 5mC in Tcn3 from the different isolates. Longest (top fifth) Tcn3 sequences were aligned to a reference Tcn3 from the strain Ftc146_1 (indicated with an asterisk) and clustered according to their 5mC status. Bottom: schematic representation of Tcn3 (gray box) flanked by two long terminal repeats (LTR-purple boxes). x axis, each box represents a C/G position relative to the reference; white, sequence could not be aligned; cream, C/G nucleotide not conserved; yellow box, CG is not methylated; light blue, CG dinucleotides is methylated (data from WGBS analysis); dark blue, 5mC from Nanopore. Faction of CGs methylated for each position is display across the top (mean: 0.7039, p < 1 × 10^{-16}, test of independence with null generated by permutations). Fraction of CGs methylated for each Tcn3 is shown on the right histogram (mean, 0.7113).

See also Figure S7 and Table S5.

Cell

AUTHOR CONTRIBUTIONS

S. Catania coordinated the work and designed, performed, and analyzed all *in vivo* experiments including computational analysis of WGBS data and computational identification of full-length transposons. P.A.D. designed and performed all *in vitro* experiments described in this study including protein expression and purification work as well as enzymological experiments that were performed with C.I.S. under the guidance of G.J.N. H.P. developed and performed the computational analysis of cytosine methylation evolution in collaboration with S. Catania under the guidance of J.K.P. R.L. and E.G. performed nanopore sequencing, and T.S. and C.A.C. performed genome assembly, genome comparisons to identify rearrangements and centromeric regions, and computational identification of methylated sites using DNA provided by S.Catania. C.A.C. and H.D.M. identified DnmtX and its evolutionary loss. A.N. and P.H. performed attomole-sensitive DNA mass spectrometry using DNA provided by S. Catania. J.K.D. and J.R.Y. performed protein mass spectrometric analysis on samples provided by S. Catania. J.E.B. generated scripts to analyze whole-genome data. S. Cooke performed stress induction experiments. S. Catania and H.D.M. conceptualized the study. H.D.M. oversaw the work. S. Catania and H.D.M. wrote the manuscript with input from all of the authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

Bannister, A.J., Zegerman, P., Partridge, J.F., Miska, E.A., Thomas, J.O., Allshire, R.C., and Kouzarides, T. (2001). Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. Nature *410*, 120–124.

Bashtrykov, P., Jankevicius, G., Jurkowska, R.Z., Ragozin, S., and Jeltsch, A. (2014). The UHRF1 protein stimulates the activity and specificity of the maintenance DNA methyltransferase DNMT1 by an allosteric mechanism. J. Biol. Chem. *289*, 4106–4115.

Bewick, A.J., Vogel, K.J., Moore, A.J., and Schmitz, R.J. (2017). Evolution of DNA Methylation across Insects. Mol. Biol. Evol. *34*, 654–665.

Bewick, A.J., Hofmeister, B.T., Powers, R.A., Mondo, S.J., Grigoriev, I.V., James, T.Y., Stajich, J.E., and Schmitz, R.J. (2019). Diversity of cytosine methylation across the fungal tree of life. Nat. Ecol. Evol. *3*, 479–490.

Bird, A. (2002). DNA methylation patterns and epigenetic memory. Genes Dev. *16*, 6–21.

Bostick, M., Kim, J.K., Estève, P.-O., Clark, A., Pradhan, S., and Jacobsen, S.E. (2007). UHRF1 plays a role in maintaining DNA methylation in mammalian cells. Science *317*, 1760–1764.

Bronner, C., Alhosin, M., Hamiche, A., and Mousli, M. (2019). Coordinated Dialogue between UHRF1 and DNMT1 to Ensure Faithful Inheritance of Methylated DNA Patterns. Genes (Basel) *10*, E65.

Brunetti, L., Gundry, M.C., and Goodell, M.A. (2017). DNMT3A in Leukemia. Cold Spring Harb. Perspect. Med. *7*, a030320.

Capella-Gutiérrez, S., Silla-Martínez, J.M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics *25*, 1972–1973.

Chun, C.D., and Madhani, H.D. (2010). Applying genetics and molecular biology to the study of the human pathogen Cryptococcus neoformans. Methods Enzymol. *470*, 797–831.

Colot, V., Maloisel, L., and Rossignol, J.L. (1996). Interchromosomal transfer of epigenetic states in Ascobolus: transfer of DNA methylation is mechanically related to homologous recombination. Cell *86*, 855–864.

D'Souza, C.A., Kronstad, J.W., Taylor, G., Warren, R., Yuen, M., Hu, G., Jung, W.H., Sham, A., Kidd, S.E., Tangen, K., et al. (2011). Genome variation in Cryptococcus gattii, an emerging pathogen of immunocompetent hosts. mBio *2*, e00342-10.

Desjardins, C.A., Giamberardino, C., Sykes, S.M., Yu, C.H., Tenor, J.L., Chen, Y., Yang, T., Jones, A.M., Sun, S., Haverkamp, M.R., et al. (2017). Population genomics and the evolution of virulence in the fungal pathogen *Cryptococcus neoformans*. Genome Res. *27*, 1207–1219.

Du, J., Johnson, L.M., Jacobsen, S.E., and Patel, D.J. (2015). DNA methylation pathways and their crosstalk with histone methylation. Nat. Rev. Mol. Cell Biol. *16*, 519–532.

Dumesic, P.A., Homer, C.M., Moresco, J.J., Pack, L.R., Shanle, E.K., Coyle, S.M., Strahl, B.D., Fujimori, D.G., Yates, J.R., 3rd, and Madhani, H.D. (2015). Product binding enforces the genomic specificity of a yeast polycomb repressive complex. Cell *160*, 204–218.

Edgar, R.C. (2004). MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics *5*, 113.

Findley, K., Rodriguez-Carres, M., Metin, B., Kroiss, J., Fonseca, A., Vilgalys, R., and Heitman, J. (2009). Phylogeny and Phenotypic Characterization of Pathogenic *Cryptococcus* Species and Closely Related Saprobic Taxa in the Tremellales. Eukaryotic cell *8*, 353–361.

Floudas, D., Binder, M., Riley, R., Barry, K., Blanchette, R.A., Henrissat, B., Martínez, A.T., Otillar, R., Spatafora, J.W., Yadav, J.S., et al. (2012). The Paleozoic origin of enzymatic lignin decomposition reconstructed from 31 fungal genomes. Science *336*, 1715–1719.

Guerreiro, M.A., Springer, D.J., Rodrigues, J.A., Rusche, L.N., Findley, K., Heitman, J., and Fonseca, A. (2013). Molecular and Genetic Evidence for a Tetrapolar Mating System in the Basidiomycetous Yeast *Kwoniella mangrovensis* and Two Novel Sibling Species. Eukaryotic Cell *12*, 746–760.

Guo, W., Fiziev, P., Yan, W., Cokus, S., Sun, X., Zhang, M.Q., Chen, P.Y., and Pellegrini, M. (2013). BS-Seeker2: a versatile aligning pipeline for bisulfite sequencing data. BMC Genomics *14*, 774.

Guo, W., Zhu, P., Pellegrini, M., Zhang, M.Q., Wang, X., and Ni, Z. (2018). CGmapTools improves the precision of heterozygous SNV calls and supports allele-specific methylation detection and visualization in bisulfite-sequencing data. Bioinformatics *34*, 381–387.

Hansen, R.S., Wijmenga, C., Luo, P., Stanek, A.M., Canfield, T.K., Weemaes, C.M., and Gartler, S.M. (1999). The DNMT3B DNA methyltransferase gene is mutated in the ICF immunodeficiency syndrome. Proc. Natl. Acad. Sci. USA *96*, 14412–14417.

Harris, B.Z., Hillier, B.J., and Lim, W.A. (2001). Energetic determinants of internal motif recognition by PDZ domains. Biochemistry *40*, 5921–5930.

He, L., Diedrich, J., Chu, Y.Y., and Yates, J.R., 3rd. (2015). Extracting Accurate Precursor Information for Tandem Mass Spectra by RawConverter. Anal. Chem. *87*, 11361–11367.

Cell

Hershberg, R., and Petrov, D.A. (2008). Selection on codon bias. Annu. Rev. Genet. *42*, 287–299.

Homer, C.M., Summers, D.K., Goranov, A.I., Clarke, S.C., Wiesner, D.L., Diedrich, J.K., Moresco, J.J., Toffaletti, D., Upadhya, R., Caradonna, I., et al. (2016). Intracellular Action of a Secreted Peptide Required for Fungal Virulence. Cell Host Microbe *19*, 849–864.

Huff, J.T., and Zilberman, D. (2014). Dnmt1-independent CG methylation contributes to nucleosome positioning in diverse eukaryotes. Cell *156*, 1286–1297.

Janbon, G., Ormerod, K.L., Paulet, D., Byrnes, E.J., 3rd, Yadav, V., Chatterjee, G., Mullapudi, N., Hon, C.C., Billmyre, R.B., Brunel, F., et al. (2014). Analysis of the genome and transcriptome of Cryptococcus neoformans var. grubii reveals complex RNA expression and microevolution leading to virulence attenuation. PLoS Genet. *10*, e1004261.

Jones, P.A. (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat. Rev. Genet. *13*, 484–492.

Koren, S., Walenz, B.P., Berlin, K., Miller, J.R., Bergman, N.H., and Phillippy, A.M. (2017). Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. Genome Res. *27*, 722–736.

Krueger, F., and Andrews, S.R. (2011). Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics *27*, 1571–1572.

Kurtz, S., Phillippy, A., Delcher, A.L., Smoot, M., Shumway, M., Antonescu, C., and Salzberg, S.L. (2004). Versatile and open software for comparing large genomes. Genome Biol. *5*, R12.

Lachner, M., O'Carroll, D., Rea, S., Mechtler, K., and Jenuwein, T. (2001). Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. Nature *410*, 116–120.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. *10*, R25.

Law, J.A., and Jacobsen, S.E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. Nat. Rev. Genet. *11*, 204–220.

Leitch, H.G., McEwen, K.R., Turp, A., Encheva, V., Carroll, T., Grabole, N., Mansfield, W., Nashun, B., Knezovich, J.G., Smith, A., et al. (2013). Naive pluripotency is associated with global DNA hypomethylation. Nat. Struct. Mol. Biol. *20*, 311–316.

Lewis, Z.A., Adhvaryu, K.K., Honda, S., Shiver, A.L., Knip, M., Sack, R., and Selker, E.U. (2010). DNA methylation and normal chromosome behavior in Neurospora depend on five components of a histone methyltransferase complex, DCDC. PLoS Genet. *6*, e1001196.

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv, arXiv:1303.3997.

Li, L., Stoeckert, C.J., Jr., and Roos, D.S. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. *13*, 2178–2189.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009a). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078–2079.

Li, M., Hays, F.A., Roe-Zurz, Z., Vuong, L., Kelly, L., Ho, C.-M., Robbins, R.M., Pieper, U., O'Connell, J.D., Miercke, L.J.W., et al. (2009b). Selecting optimum eukaryotic integral membrane proteins for structure determination by rapid expression and solubilization screening. J. Mol. Biol. *385*, 820–830.

Lindsley, J.E., and Wang, J.C. (1993). On the coupling between ATP usage and DNA transport by yeast DNA topoisomerase II. J. Biol. Chem. *268*, 8096–8104.

Liu, O.W., Chun, C.D., Chow, E.D., Chen, C., Madhani, H.D., and Noble, S.M. (2008). Systematic genetic analysis of virulence in the human fungal pathogen Cryptococcus neoformans. Cell *135*, 174–188.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal *17*, 10–12.

Ngamskulrungroj, P., Gilgado, F., Faganello, J., Litvintseva, A.P., Leal, A.L., Tsui, K.M., Mitchell, T.G., Vainstein, M.H., and Meyer, W. (2009). Genetic Diversity of the Cryptococcus Species Complex Suggests that Cryptococcus gattii Deserves to Have Varieties. PLoS One *4*, e5862.

Okano, M., Bell, D.W., Haber, D.A., and Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell *99*, 247–257.

Pack, L.R., Yamamoto, K.R., and Fujimori, D.G. (2016). Opposing Chromatin Signals Direct and Regulate the Activity of Lysine Demethylase 4C (KDM4C). J. Biol. Chem. *291*, 6060–6070.

Pause, A., and Sonenberg, N. (1992). Mutational analysis of a DEAD box RNA helicase: the mammalian translation initiation factor eIF-4A. EMBO J. *11*, 2643–2654.

Peng, J., Elias, J.E., Thoreen, C.C., Licklider, L.J., and Gygi, S.P. (2003). Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: the yeast proteome. J. Proteome Res. *2*, 43–50.

Pradhan, S., Bacolla, A., Wells, R.D., and Roberts, R.J. (1999). Recombinant human DNA (cytosine-5) methyltransferase. I. Expression, purification, and comparison of de novo and maintenance methylation. J. Biol. Chem. *274*, 33002–33010.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

Reményi, A., Good, M.C., Bhattacharyya, R.P., and Lim, W.A. (2005). The role of docking interactions in mediating signaling input, output, and discrimination in the yeast MAPK network. Mol. Cell *20*, 951–962.

Rošić, S., Amouroux, R., Requena, C.E., Gomes, A., Emperle, M., Beltran, T., Rane, J.K., Linnett, S., Selkirk, M.E., Schiffer, P.H., et al. (2018). Evolutionary analysis indicates that DNA alkylation damage is a byproduct of cytosine DNA methyltransferase activity. Nat. Genet. *50*, 452–459.

Simpson, J.T., Workman, R.E., Zuzarte, P.C., David, M., Dursi, L.J., and Timp, W. (2017). Detecting DNA cytosine methylation using nanopore sequencing. Nat. Methods *14*, 407–410.

Stamatakis, A. (2006). RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics *22*, 2688–2690.

Tabb, D.L., McDonald, W.H., and Yates, J.R., 3rd. (2002). DTASelect and Contrast: tools for assembling and comparing protein identifications from shotgun proteomics. J. Proteome Res. *1*, 21–26.

Urich, M.A., Nery, J.R., Lister, R., Schmitz, R.J., and Ecker, J.R. (2015). MethylC-seq library preparation for base-resolution whole-genome bisulfite sequencing. Nat. Protoc. *10*, 475–483.

Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q., Wortman, J., Young, S.K., and Earl, A.M. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS ONE *9*, e112963.

Walton, E.L., Francastel, C., and Velasco, G. (2014). Dnmt3b Prefers Germ Line Genes and Centromeric Regions: Lessons from the ICF Syndrome and Cancer and Implications for Diseases. Biology (Basel) *3*, 578–605.

Wang, X., Hsueh, Y.-P., Li, W., Floyd, A., Skalsky, R., and Heitman, J. (2010). Sex-induced silencing defends the genome of Cryptococcus neoformans via RNAi. Genes Dev. *24*, 2566–2582.

Wyszynski, M.W., Gabbara, S., and Bhagwat, A.S. (1992). Substitutions of a cysteine conserved among DNA cytosine methylases result in a variety of phenotypes. Nucleic Acids Res. *20*, 319–326.

Xu, T., Park, S.K., Venable, J.D., Wohlschlegel, J.A., Diedrich, J.K., Cociorva, D., Lu, B., Liao, L., Hewel, J., Han, X., et al. (2015). ProLuCID: An improved SEQUEST-like algorithm with enhanced sensitivity and specificity. J. Proteomics *129*, 16–24.

Yaari, R., Katz, A., Domb, K., Harris, K.D., Zemach, A., and Ohad, N. (2019). RdDM-independent de novo and heterochromatin DNA methylation by plant CMT and DNMT3 orthologs. Nat. Commun. *10*, 1613.

Yadav, V., Sun, S., Billmyre, R.B., Thimmappa, B.C., Shea, T., Lintner, R., Bakkeren, G., Cuomo, C.A., Heitman, J., and Sanyal, K. (2018a). RNAi is a critical determinant of centromere evolution in closely related fungi. Proc. Natl. Acad. Sci. USA *115*, 3108–3113.

Zemach, A., McDaniel, I.E., Silva, P., and Zilberman, D. (2010). Genome-wide evolutionary analysis of eukaryotic DNA methylation. Science *328*, 916–919.

Cell

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
anti-FLAG antibody	Millipore-Sigma	Cat#F3165; RRID:AB_259529
anti-HA antibody	Thermo Fisher	Cat#26183; RRID:AB_10978021
anti-Histone H3 Polyclonal Antibody	Thermo Fisher	Cat#PA5-16183; RRID:AB_10985434
Anti-Histone H3 (di methyl K9) antibody ChIP grade	Abcam	Cat#ab1220; RRID:AB_449854
Anti-Histone H3 antibody - Nuclear Loading Control and ChIP Grade	Abcam	Cat#ab1791; RRID:AB_302613
Goat Anti-Rabbit IgG-HRP Conjugate	Bio-Rad	Cat#1706515; RRID:AB_11125142
Goat Anti-Mouse IgG-HRP Conjugate	Bio-Rad	Cat#1706516; RRID:AB_11125547
anti-5-methylcytosine antibody	Epigentek	Cat#A-1014-050; RRID:AB_2819207
anti-GST antibody	Millipore-Sigma	Cat#G7781; RRID:AB_259965
Goat Anti-Rabbit IgG-HRP Conjugate	Santa Cruz Biotechnology	Cat#sc-2004; RRID:AB_631746
Chemicals, Peptides, and Recombinant Proteins		
3X FLAG Peptide	Millipore-Sigma	Cat#F4799
Critical Commercial Assays		
Qubit dsDNA BR Assay Kit	Thermo-Fisher	Cat#Q32853
EZ-DNA methylation Gold Kit	Zymo Research	Cat#D5005
DNeasy Plant Mini Kit	QIAGEN	Cat#69106
Genomic DNA Clean & Concentrator	Zymo Research	Cat#D4065
Deposited Data		
Raw and analyzed WGBS and ChIP-seq data	This paper	GEO: GSE134684
Raw Oxford nanopore sequence and genome assemblies	This paper	NCBI BioProject: PRJNA517966
Experimental Models: Organisms/Strains		
List of strains	This paper	See Table S2
Oligonucleotides		
Oligonucleotides used	This paper	See Table S3
Recombinant DNA		
pBH4-Swi6	This paper	BHM2226
pMAL -Dnmt5(1-150)	This paper	BHM2172
pMAL -Dnmt5(1-150) W87A Y90A	This paper	BHM2192
pETARA-Dnmt5(1-150)	This paper	BHM2170
pBH4-Uhrf1	This paper	BHM2211
83ν- Dnmt5	This paper	BHM2244
Software and Algorithms		
Bismark	Krueger and Andrews, 2011	https://www.bioinformatics.babraham.ac.uk/projects/bismark/
BS-Seeker2	Guo et al., 2013	https://github.com/BSSeeker/BSseeker2
BEDTools	Quinlan and Hall, 2010	https://bedtools.readthedocs.io/en/latest/
bowtie1	Langmead et al., 2009	http://bowtie-bio.sourceforge.net/index.shtml
Pysam	Li et al., 2009a	https://github.com/pysam-developers/pysam
CGmap tools	Guo et al., 2018	https://cgmaptools.github.io/
Nanopolish	Simpson et al., 2017	https://github.com/jts/nanopolish
Canu	Koren et al., 2017	https://canu.readthedocs.io/en/latest/
NUCmer 3	Kurtz et al., 2004	http://mummer.sourceforge.net/

(Continued on next page)

Cell

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Pilon	Walker et al., 2014	http://software.broadinstitute.org/software/pilon/
Samtools	Li et al., 2009a	http://samtools.sourceforge.net/
RAxML	Stamatakis, 2006	https://github.com/stamatak/standard-RAxML
Cutadapt	Martin, 2011	http://cutadapt.readthedocs.io/en/stable/

LEAD CONTACT AND MATERIALS AVAILABILITY

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Hiten D. Madhani (hitenmadhani@gmail.com).

All reagents generated in this study are available from the Lead Contact without restriction.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Yeast growth and manipulations

Yeast manipulation and protocols were performed as described in Chun and Madhani (2010). Tagging, deletions and insertions of genes were obtained through homologous recombination by transforming about 10 μg of either a PCR product or a PmeI-digested plasmids bearing the sequence to be transformed. Mutations of Dnmt5 on its catalytic methyltransferase domain *(C440A)* and ATP-ase domain *(K1469A)* were created based on prior work (Pause and Sonenberg, 1992; Wyszynski et al., 1992)

The *RI-DMT5* strain used for WGBS was produced by deleting the 5′ portion of the *DMT5* gene and then restoring it by homologous replacement. To be certain that no residual DNA methylation was present in *dmt5Δ* prior to introduction of the targeting construct that restores the gene, the *dmt5Δ* strain was first propagated in YPAD for at least 90 generations.

To express DnmtX in an *RI-DMT5* strain, genomic DNA from *K. pini, K. mangroviensis* and *K. bestiolae* was used as a template to amplify the genes putative DnmtXs (*K. mangroviensis* I203_05465, *K. pini* I302_00877, *K. bestiolae* I206_02865). DnmtX was cloned downstream a *GAL7* promoter and the entire construct was then inserted at the *URA5* locus. Upon transformation, cells were selected in YPAD+nourseothrycin and then propagated for two sequential single-colony streaks in presence of galactose to induce DnmtX.

Other strains generated by transformation were colony-purified, patched, verified, and frozen. Liquid cultures were inoculated directly from the frozen stock. A list of strains used in this study can be found on Table S2.

METHOD DETAILS

Genetic crosses

Crosses were carried out using strains of the KN99 background. Cells of different mating types were grown overnight in YPAD, spotted on mating plates (4% Agar, 10% 10X Murashige and Skoog basal medium, 0.1mg/ml myo-inositol) and kept in the dark for 1 to 2 weeks. Spores were collected with a toothpick and resuspended in 500 μl of water. Dilutions of this suspension were then spread on selective media.

DNA extraction and Southern analysis

DNA from *C. neoformans, K. pinus, K. mangroviensis* and *K. bestiolae* was extracted as previously described (Chun and Madhani, 2010) with some modifications. In brief, 100 OD of yeast (100 ml, OD1) were harvested, frozen in liquid nitrogen, resuspended in 5 mL CTAB buffer (100 mM Tris pH 7.6, 1M NaCl, 10 mM EDTA, 1% Cetyltrimethyl ammonium bromide, 1% β-mercaptoethanol) and incubated at 65°C for at least 1 hr. 3 mL of chloroform were added to the lysate, mixed and spun for 5 min at 3000 g. The aqueous phase was then precipitated by adding the same volume of isopropanol. The dried pellet was resuspended in 300 μl TE containing 1 μg RNase A (Thermo Scientific) and incubated 30 min at 37°C followed by addition of 3 μl Proteinase K (20 mg/ml) and incubation at 50°C for 1 hr. Phenol-chloroform and chloroform extraction was carried out and the DNA is finally precipitated by addition of 1/10 volume NaOAc and three volumes of ethanol. The DNA was additionally purified with Genomic DNA Clean & Concentrator (D4065, Zymo Research). For Southern analysis, 10 μg were then digested with the meCpG sensitive enzyme HpyCH4IV (NEB), separated by electrophoresis on a 1% agarose gel and transferred to nylon membrane (Hybond NX, Amersham) by capillary action. DNA was crosslinked to the membrane using UV light. PCR products were used as template to incorporate radioactive ^{32}P-dCTP (Perkin Elmer) using the High Prime kit (Roche) according to manufacturer's protocol. Primer sequences for the probes are listed in Table S3.

Stress conditions experimental procedures

To assess if specific stress conditions induce *de novo* DNA methylation activity, wild-type and RI-Dnmt5 cells were grown in the presence of stressors. 2.3 M NaCl at 30°C for 3 hr, 140 mM/ 2.3mM NaNO$_2$/succinic acid at 30°C for 3 hr, 0.14% H$_2$O$_2$ at 30°C for 3 hr,

Cell

0.9% SDS at 30°C for 3 hr, 1.9 mg/ml caffeine at 37°C for 20 hr, DMEM at 37°C, 5% CO_2, for 24 hr. The concentrations of stressors used represent the maximum concentration that allows growth in the conditions tested.

Protein extraction and western analysis

Two milliliters of culture at OD600 = 1 were collected by centrifugation and frozen in liquid nitrogen, resuspended in 10% TCA and incubated on ice for 10 min. Cells were washed once with acetone and air-dried for 10 min. The pellet was then resuspended in 150 μl 2x Laemmli buffer adjusted with 80 μl Tris-HCl (pH 8.0) and bead-beated 2 × 90 s. The lysate was boiled for 5min and centrifuged to remove residual cells. Western analysis was performed using anti-FLAG antibody (1:3000- F3165, sigma), anti-HA (1:10000-26183, Thermo Fisher) or anti-H3 antibody (1:1000- PA5-16183, Thermo Fisher) diluted with 5% milk in TBS-T (10 mM Tris-HCl pH 7.6, 150 mM NaCl, 0.1% Tween 20) for 1 hr followed by 2 × 10 min washes in TBS-T. The membrane was then incubated with antibody anti-rabbit or anti-mouse conjugated to HRP (BIO-RAD) (1:8000 in 5% milk+TBS-T) for 45 min followed by 2 × 10 min washes in TBS-T. The membrane was then incubated in SuperSignal West pico luminol (Thermo Scientific) and visualized using film.

Affinity purification and mass spectrometry analysis

Dynabeads M-270 Epoxy (Cat n. 14301) were conjugated following the manual with the following modifications. 150 μg anti-FLAG antibody (Sigma-F3165) was conjugated with $1x10^9$ beads overnight at 37°C. The beads were washed once with 100 mM Glycine HCL pH 2.5, once with 10 mM Tris pH 8.8, 4x with 1x PBS for 5 min, once with PBS + 0.5% Triton X-100 for 5 min, once PBS +0.5% Triton X-100 for 15 min. Beads were the resuspended in the PBS and kept at 4°C before using them.

4 L of *C. neoformans* cells were grown on YPAD at 30°C until OD = 2, harvested, resuspended in 10 mL Lysis buffer (25 mM HEPES-KOH, pH 7.9, 2 mM $MgCl_2$, 100 mM KCl, 16% glycerol, 0.1% Tween 20) and dripped into liquid nitrogen to form 'popcorn.' Cell were lysed by cryo-grinding in a SPEX Sample Prep 6870 instrument for 3 cycles of 2 min at 10 cps with 2 min of cooling between each cycle. The extract was then thawed at 4°C, the final volume was brought to 80 mL with Lysis Buffer+2 mM $CaCl_2$ and incubated with 1000U of RQ1 DNase (Promega- PRM6101) for 30min at RT. At this point, the extract was clarified by centrifugation at 40Kxg for 40 min at 4°C and FLAG-conjugated magnetic beads were added to the supernatant and incubated for 4h at 4°C with rotation. The beads were then washed 3x10 min with lysis buffer and using a magnetic rack. Bound proteins were eluted three times in 150 μl elution buffer (25 mM HEPES-KOH, pH 7.9, 2 mM $MgCl_2$, 300 mM KCl, 20% glycerol) with 1 mg/ml 3X FLAG peptide (Sigma). For the co-immunoprecipitation experiments, 1 L culture was used in the assay and the final elution step was performed by boiling the beads with 2x Laemmli buffer for 10 min.

Label free mass spectrometry was performed as follows. Samples were precipitated by methanol/ chloroform. Dried pellets were dissolved in 8 M urea/100 mM TEAB, pH 8.5. Proteins were reduced with 5 mM tris(2-carboxyethyl)phosphine hydrochloride (TCEP, Sigma-Aldrich) and alkylated with 10 mM chloroacetamide (Sigma-Aldrich). Proteins were digested overnight at 37°C in 2 M urea/ 100 mM TEAB, pH 8.5, with trypsin (Promega). Digestion was quenched with formic acid, 5% final concentration.

The digested samples were analyzed on a Fusion Orbitrap tribrid mass spectrometer (Thermo). The digest was injected directly onto a 25 cm, 100 μm ID column packed with BEH 1.7 μm C18 resin (Waters). Samples were separated at a flow rate of 300 nl/min on a nLC 1000 (Thermo). Buffer A and B were 0.1% formic acid in water and 0.1% formic acid in acetonitrile, respectively. A gradient of 1%–25% B over 110 min, an increase to 40% B over 10 min, an increase to 90% B over 10 mins and held at 90%B for a final 10 min was used for 140 min total run time. Column was re-equilibrated with 20 μl of buffer A prior to the injection of sample. Peptides were eluted directly from the tip of the column and nanosprayed directly into the mass spectrometer by application of 2.5 kV voltage at the back of the column. The Orbitrap Fusion was operated in a data dependent mode. Full MS scans were collected in the Orbitrap at 120K resolution with a mass range of 400 to 1500 m/z and an AGC target of $4e^5$. The cycle time was set to 3 s, and within this 3 s the most abundant ions per scan were selected for CID MS/MS in the ion trap with an AGC target of $5e^4$ and minimum intensity of 5000. Maximum fill times were set to 50 ms and 100 ms for MS and MS/MS scans respectively. Quadrupole isolation at 1.6 m/z was used, monoisotopic precursor selection was enabled and dynamic exclusion was used with exclusion duration of 5 s.

Bisulfite conversion and library construction

Bisulfite library construction was performed as previously described (Urich et al., 2015). Briefly, 100 OD units of yeast were harvested, lyophilised overnight and the DNA was extracted using DNeasy Plant Mini Kit (QIAGEN). The DNA was additionally purified with a Genomic DNA Clean & Concentrator (D4065- Zymo Research). Genomic DNA was quantified with Qubit dsDNA BR Assay Kit (ThermoFisher). 5 ng of lambda DNA (Promega) was added to ~1 ug yeast DNA and fragmented using a Bioruptor Pico instrument (Diagenode- 13 cycles, 30 s/ 30 s On/Off). The DNA was end-repaired and ligated to methylated adapters. Two rounds of bisulfite conversion of the adaptor-ligated DNA were performed using EZ-DNA methylation Gold Kit (D5005- Zymo Research), and used as template for three separate PCR reactions (eight cycles of amplification) as described in Urich et al. (2015). DNA obtained from the three PCR reactions were then combined and analyzed for size on a High Sensitivity DNA chip (Agilent Technologies) and sequenced as SE100.

Chromatin immunoprecipitation

ChIP and library preparation were performed as previously described (Homer et al., 2016) with the following modifications. 50 OD were lysed in ChIP lysis buffer (50 mM HEPES-KOH pH 7.5, 140 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% sodium deoxycholate,

1mM PMSF and 1x protease inhibitors) using a bead beater (8 cycles x 90 s) and lysate was clarified by centrifugation 10 min at 6800 g. The chromatin pellet was sonicated in a Bioruptor Pico instrument (Diagenode; 25 cycles, 30 s/ 30sec on/off) and centrifuged for 20 min at 20000 g. The cleared lysate was incubated overnight with 6 μg anti-histone H3K9me2 (ab1220, Abcam), 4 μl anti-FLAG antibodies (F3165, sigma) or 3 μl anti-H3 (ab1791, Abcam) with the addition of 30 μl of protein A (for anti-H3K9me) or protein G (for anti-FLAG, anti-H3) Dynabeads (Thermo Fisher) overnight at 4°C. Beads were then washed once with ChIP lysis buffer, once with ChIP lysis buffer with 500 mM NaCl, once with wash buffer (10 mM Tris-Cl pH 8.0, 0.25 M LiCl, 0.5% NP-40, 0.5% sodium deoxycholate, 1 mM EDTA), and once with 1 mL TE buffer (10 mM Tris-Cl pH 8.0, 1 mM EDTA). Each wash was performed for 10 min at room temperature. Beads were eluted with 100 μL elution buffer (50 mM Tris-Cl pH 8.0, 10 mM EDTA, 1% SDS) at 70°C for 15 min, followed by a second elution at room temperature for 5 min with 150 μL TE buffer with 0.67% SDS and 334 μg/ml proteinase K. 200 μL TE buffer with 1% SDS and 250 μg/ml proteinase K were added to the input samples. All samples were incubated overnight at 65°C. The DNA was purified with NucleoSpin PCR clean-up columns (Macherey-Nagel) using NTB buffer (Macherey-Nagel).

For the ChIP library preparation, the DNA was end repaired with Fast DNA End Repair Kit (Thermo Fisher Scientific) and then A-tailed with Klenow (exo-) (NEB) according to manufacturers' instructions. Illumina adapters were ligated using Rapid T4 DNA Ligase (QIAGEN Beverly Inc) at room temperature for 30 min. The DNA was purified using NucleoSpin PCR clean-up columns (Macherey-Nagel) between each step of the library preparation. The library was then amplified with 1U Phusion DNA polymerase (NEB), 2 M betaine, 75 μM dNTPs, and 0.4 μM each primer in a 50 μL reaction. Cycling conditions were: 98°C for 3 min; 15 cycles of 98°C for 1 min 20 s, 65°C 30 s, 72°C 30 s; 72°C 5 min.

Methylated DNA immunoprecipitation (MeDIP)

The initial steps are the same as described for the bisulfite library protocol (Urich et al., 2015) with the exception that for MeDIP, unmethylated adapters were ligated to repaired, dA-tailed DNA. Upon ligation and purification, DNA was boiled for 10 min and kept on ice for 5 min. Ice cold binding buffer (10 mM sodium phosphate buffer, pH 7, 140 mM NaCl 0.05% Triton X-100) was added to the DNA together with 1 μl of anti-5-methylcytosine antibody (A-1014-050, Epigentek) and incubated on a nutator overnight at 4°C. 20 μl Protein G Dynabeads (Thermo Fisher) were added and incubated on a Nutator for further 2 hr at 4°C. Beads were washed 10 × 5 min in binding buffer and the DNA was eluted by incubating the beads with Elution Buffer (TE, 0.25 mg/ml Proteinase K, 0.25% SDS) for 2 hr at 55°C. The DNA was purified with NucleoSpin PCR clean-up columns (Macherey-Nagel) using NTB buffer (Macherey-Nagel) and PCR amplified as for ChIP. Bioinformatics analyses were performed as for ChIP-seq.

In vitro methylation and 5mC-qPCR assay

The construct was *in vitro* methylated with HpaII methyltransferase (NEB), which recognizes the CCGG sequence and modifies the internal cytosine residue. At least 10 μg of the targeting construct were incubated with 20U HpaII methyltransferase for 2 h at 37°C followed by inactivation of the enzyme at 65°C for 20 min. The DNA was then purified with NucleoSpin Gel and PCR Clean-up Columns (Macherey-Nachel, 740609) and the efficiency of the *in vitro* methylation reaction was assayed by digesting 100 ng of the *in vitro* methylated product with 1U of the HpaII endonuclease (run: undigested, unmethylated+digestion, methylated+digestion, Figure 5C). Transformation was carried out with the *in vitro* methylated or with the unmodified fragment and the transformants were selected in the presence of G418.

To assess the ability of cells to maintain 5mC at the *in vitro* methylated CCGG sites, we calculated the methylation fraction for specific sites as ratio of qPCR values across the CCGG-site of interest (5mC protects from restriction enzyme digestion) to a region that does not contain the CCGG site. To test the capability of Dnmt5 to promote the spread 5mC, adjacent sites to HpaII sequence were used and in specific CGCG sites (BstUI) or CGWCG sites (Hyp99I). Genomic DNA was prepared as previously described and cleaned up with a Genomic DNA Clean & Concentrator kit (D4065- Zymo Research). 50 ng of genomic DNA was digested with the appropriate restriction endonuclease and purified using Genomic DNA Clean & Concentrator kit (D4065- Zymo Research). 0.6 ng was used as template for qPCR analysis using PowerUp SYBR Green Master Mix (Life Technologies) with primers listed in Table S3.

Experimental evolution

Wild-type cells from a starter culture (g0) were propagated at 30°C in rich media for 120 generations maintaining the logarithmic growth by diluting the culture as soon as it reached an OD600 of 1. After 120 generations, the cells were plated on YPAD agar plates to form single colonies at 30°C. Two independent isolates (colony 1 and colony 2) were grown overnight in rich media to OD600 of 1 and the genomic DNA extracted processed in technical duplicates for WGBS. WGBS libraries and analysis were performed as described above.

Recombinant protein expression and purification

A codon-optimized DNA sequence encoding *C. neoformans* Dnmt5 (residues 1-150) was cloned into pETARA or pMAL vectors and used to transform *E. coli* strain BL21(DE3) (Harris et al., 2001; Reményi et al., 2005). Transformed cells were grown to OD_{600} = 0.8 in 2x YT medium, then induced with 1 mM IPTG overnight at 18°C. Recombinant GST-Dnmt5(1-150)-6xHis was purified with Ni-NTA

Cell

agarose resin (QIAGEN), measured by A_{280} (ε = 66,030 cm^{-1} M^{-1}), and used for histone peptide array binding assays. Recombinant MBP-Dnmt5(1-150)-6xHis was purified with Ni-NTA agarose resin (QIAGEN), measured by A_{280} (ε = 89,590 cm^{-1} M^{-1}), and used for fluorescence polarization binding assays.

A codon-optimized DNA sequence encoding full-length *C. neoformans* Uhrf1 was cloned into pBH4 and expressed in *E. coli* as above. Recombinant 6xHis-Uhrf1 was purified with Ni-NTA agarose resin (QIAGEN), measured by A_{280} (ε = 40,920 cm^{-1} M^{-1}), and used for electrophoretic mobility shift assays.

Full-length cDNA encoding Swi6 was cloned from *C. neoformans* and expressed in *E. coli* as above using the pBH4 vector. Recombinant 6xHis-Swi6 was purified with Ni-NTA agarose resin (QIAGEN), measured by A_{280} (ε = 45,330 cm^{-1} M^{-1}), and used for fluorescence polarization binding assays.

For expression in *S. cerevisiae*, full-length cDNA encoding Dnmt5 was cloned from *C. neoformans*, inserted into the 83ν vector (Li et al., 2009b), and used to transform the *S. cerevisiae* strain JEL1 (Lindsley and Wang, 1993). Starter cultures were grown overnight in SC -ura medium (2% glucose), then used to inoculate 2 L cultures of YPGL medium (1x YEP, 1.7% lactic acid, 3% glycerol, 0.12% glucose, 0.15 mM adenine) to a starting OD_{600} of 0.03. After growth at 30°C to an OD_{600} of 1.0, expression was induced by addition of 2% galactose. After 6 hr of continued growth at 30°C, cells were harvested, washed once in TBS (50 mM Tris-Cl pH 7.6, 150 mM NaCl), and snap frozen. Frozen cells were lysed in a ball mill (6 × 3 min at 15 Hz), resuspended in Ni-NTA lysis buffer (50 mM NaH$_2$PO$_4$ pH 8, 300 mM NaCl, 10% glycerol, 10 mM imidazole, 2 mM β-mercaptoethanol, 0.02% NP-40, 1x CPI), and centrifuged 20,000 x g for 30 min at 4°C. Lysate was bound to Ni-NTA resin in batch format for 2 hr at 4°C, after which resin was washed in column format with five bed volumes Ni-NTA buffer followed by ten bed volumes Ni-NTA wash buffer (identical to Ni-NTA lysis buffer except 20 mM imidazole). Finally, bound protein was eluted with four bed volumes of elution buffer (identical to Ni-NTA lysis buffer except 300 mM imidazole and no detergent). Protein was dialyzed against storage buffer and applied to a HiTrap Q HP anion exchange column (GE Life sciences) pre-equilibrated in buffer (50 mM HEPES-KOH pH 7.9, 150 mM KCl, 10% glycerol, 2 mM β-mercaptoethanol). Fractions were collected across a 150-1000 mM KCl gradient, and those containing Dnmt5 were pooled, dialyzed against storage buffer, and frozen.

Histone peptide array binding assay

Modified histone peptide arrays (Active Motif) were blocked with 5% dried milk in TBS-T (10 mM Tris-HCl pH 7.6, 150 mM NaCl, 0.1% Tween 20) overnight at 4°C. Each slide was subsequently washed in TBS-T (4x5 min) and binding buffer (2 × 5 min). Binding buffer consisted of 50 mM HEPES-KOH pH 7.9, 200 mM NaCl. The slide was incubated with 0.5 μM GST-Dnmt5(1-150) in binding buffer for 4 hr at 23°C. It was then washed with TBS-T (4x 5 min). Next, the slide was incubated with anti-GST antibody (Sigma G7781; 1:13,000 dilution) in 5% dried milk in TBS-T for 1hr at 23°C. After washing in TBS-T (4 × 5 min), the slide was incubated with HRP-conjugated goat anti-rabbit secondary antibody (sc-2004, Santa Cruz Biotechnology; 1:5,000 dilution) in 5% milk in TBS-T for 1hr at 23°C. The slide was washed in TBS-T (4 × 5 min) and visualized using SuperSignal West Pico chemiluminescent substrate (Pierce) and a ChemiDoc MP System (BioRad). Signal intensity normalization was performed using Array Analyze Software (Active Motif).

Fluorescence polarization binding assay

Peptides were synthesized by GenScript (Piscataway, NJ) or Peptide 2.0 (Chantilly, VA) to > 95% purity. For study of H3K9 methylation, unlabeled peptides corresponded to residues 1-15 of *C. neoformans* histone H3. For study of H3K27 methylation, unlabeled peptides corresponded to residues 23-34 of *C. neoformans* histone H3 followed by a cysteine residue. Peptide concentrations were determined by A_{205} or, in the case of fluorescein-conjugated peptides, A_{495} (ε = 80,000 M^{-1} cm^{-1}).

For direct measurement of peptide binding, increasing concentrations of MBP-Dnmt5(1-150) were incubated with 10 nM labeled peptide in a solution of 20 mM HEPES pH 7.9, 120 mM KCl, 0.8 mM DTT, and 0.01% NP-40. Peptide fluorescence polarization was measured using a Spectramax M5e plate reader (Molecular Devices) and non-stick 384-well plates (Corning 3820). For competition assays, unlabeled competitor peptides were added at increasing concentration in the setting of 5 μM MBP-Dnmt5(1-150) or 30 μM 6xHis-Swi6 and 10 nM fluorescein-conjugated H3K9me3 peptide. Dissociation constants were calculated using a competition binding equation in Prism (GraphPad Software) (Pack et al., 2016):

$$FP_{obs} = \frac{K_i * (FP_{max} * [Dnmt5] + FP_{min} * K_d) + FP_{min} * K_d * [I]}{K_i * (K_d + [Dnmt5]) + K_d * [I]}$$

Dissociation constants for H3K9me3 peptide were comparable when measured directly (1.3 μM) or as a competitor (1.5 μM).

Gel mobility shift assay

Primer sequences are listed in Table S3 and were annealed to generate unmethylated, hemimethylated, or symmetrically methylated 20 bp dsDNA substrates. The substrates were then radiolabeled using the KinaseMax kit (Ambion) and ATPγ-^{32}P (Perkin Elmer), after which they were purified using a G-25 illustra microspin column (GE Life Sciences). For direct measurement of DNA binding, recombinant 6xHis-Uhrf1 was incubated with 0.2 nM labeled DNA probe in a 10 μL solution of 15 mM HEPES-KOH pH 7.9, 7.5% glycerol, 75 mM KCl, 0.075% NP-40, 0.05 μg/ul poly-dI:dC (Sigma), 0.5 μg/ul BSA, 1 mM DTT, and 5 mM MgCl$_2$. After 30 min equilibration at 23°C, samples were resolved in polyacrylamide gels (4.5% acrylamide:bis 29:1 (Bio-Rad), 1% glycerol, 0.25X TBE) at 4°C. Gels were

subsequently dried and imaged using a storage phosphor screen (Amersham) and a Typhoon 9400 imager (Amersham). Densitometry was performed using ImageJ. For competitive binding assays, conditions were as above except Uhrf1 protein was kept constant at 150 nM in the presence of 0.2 nM labeled hemimethylated probe and excess amounts of unlabeled DNA oligonucleotides.

DNA methyltransferase assay

DNA oligonucleotides were synthesized and annealed to generate 20 or 60 bp dsDNA substrates (Table S3). DNA methylation was monitored in multiple turnover conditions by incubating 30 nM Dnmt5 in a solution of 50 mM Tris pH 8, 25 mM NaCl, 10% glycerol, and 2 mM DTT, in the presence of 5 μM DNA substrate. When indicated, ATP and $MgCl_2$ were supplemented at 1 mM, histone tail peptides corresponding to histone H3 residues 1-15 were added at 5 μM, Uhrf1 was added at 4 μM, and Swi6 was added at 4 μM. Reactions were initiated by the addition of 4 μM ^{3}H-SAM (Perkin Elmer) and carried out at 23°C. Reaction aliquots were removed at indicated time points and quenched in a solution of 10 mM SAM in 10 mM H_2SO_4. The quenched solution was pipetted onto DE81 filter paper and allowed to dry for 10 min. Filter papers were subsequently washed three times in 200 mM ammonium bicarbonate (5 min each), then once with water (5 min). Filter papers were then rinsed twice in ethanol, after which they were dried for 20 min. Filters were added to scintillation fluid (Bio-Safe NA, Research Products International Corp.), and bound ^{3}H was detected in an LS 6500 scintillation counter (Perkin Elmer). Background signal was assessed using reactions that lacked Dnmt5 enzyme. Detectable signal was defined as reactions that exhibited cpm greater than 2-fold above background at each measured time point. Background signal was typically 50-100 cpm and signal for productive reactions ranged from ~1000 to ~100,000 cpm depending on conditions and time point. For reactions in which methylation was detected, rates were calculated over the first 15-20 min where reaction progress was linear and less than 10% of available hemimethylated sites had been acted upon. These initial rate values were divided by Dnmt5 concentration to obtain k_{obs}. DNA substrate was confirmed to be present at saturating concentration for these experiments. Serial dilutions confirmed that ^{3}H detection using DE81 was linear to the level of background signal.

Liquid chromatography mass spectrometry (LC-MS/MS) method to quantify 5mC

Genomic DNA was isolated as previously described and purified with Genomic DNA Clean & Concentrator kit (D4065 - Zymo Research). Enzyme mix provided (New England Biolabs) was used to digest the gDNA to nucleosides. Standard nucleosides were diluted in different ratios using LC-MS water. Identical amount of isotope labeled-internal standards (provided by T. Carell) were spiked in the dilutions of standards and in the samples. Chromatographic separation of nucleosides was achieved using an Agilent RRHD Eclipse Plus C18 2.1 × 100 mm 1.8 μm column and the analysis was performed using an Agilent triple quadrupole mass spectrometer, (1290) UHPLC hyphenated to a (6490) QQQ-MS. The LC-MS method used for the quantitation of 5mC was performed as previously described (Leitch et al., 2013). A standard curve was generated by dividing unlabeled over isotope-labeled nucleosides and used to covert the peak area values to quantities. The signal-to-noise (S/N) used for quantification (above 10) was calculated using the peak-to-peak method. Global 5mC level is quoted relative to total levels of deoxyguanine (dG). The LOD (level of detection) for 5mC in the experiments was 25 amol and LOQ (limit of quantification) was 250 amol.

QUANTIFICATION AND STATISTICAL ANALYSIS

ChIP-seq analysis

Reads were trimmed for adaptor sequence ("GATCGGAAGA") and filtered for reads of at least 50 nt in length using Cutadapt (Martin, 2011) with were aligned to the C. neoformans genome using bowtie1 (modified parameters: -v2, -M1,–best) (Langmead et al., 2009). Alignment files were processed with SAMtools (Li et al., 2009a) and converted to bedgraph format for visualization with BEDTools (Quinlan and Hall, 2010). Bedgraphs were scaled by million aligned reads and normalized to whole cell extract at each genomic position, then smoothed using a 500 bp centered rolling mean using custom Python scripts.

For scatterplots, reads (were counted in each region (centromere or telomere) using the Pysam Python library (https://github.com/pysam-developers/pysam). RPKM was determined by scaling the reads in each region, *n*, to the total number of aligned reads in the sample, *M* (in millions), and the length of the region, *l* (in kb):

$$RPKM = \frac{n}{M * l}$$

Normalized enrichment in ChIP over whole cell extract, *E*, was calculated based on RPKM as follows:

$$E = RPKM_{ChIP}/RPKM_{WCE}$$

WGBS analysis

Reads from the bisulfite-treated library were trimmed using cutadapt (https://cutadapt.readthedocs.io/en/stable/) and analyzed with Bismark (Krueger and Andrews, 2011) using bowtie2 with the options:–score_min L,0,0. Unconverted reads were removed using

Cell

Bismark-filter_non_conversion and methylation was called using the standard setting. To quantify CG methylation levels globally, data were further filtered using a custom Python script for signals at CG dinucleotide that displayed a fractional methylation of 0.5 or greater on both strands.

Experimental evolution 5mC analysis
Differential methylation site (DMS) analysis was performed using CGmap tools (https://cgmaptools.github.io). Bismark outputs were converted into CGmap file using cgmaptools-convert-bismark2cgmap and then sorted and intersected with cgmaptools-sort and –intersect, respectively. Differential analysis was performed using cgmaptools-dms on the intersected files. Only sites of symmetric CG with > 50% methylation that were reproducible between technical replicates were considered in the analysis. DMS was considered significant when p value < 0.05.

Nanopore sequencing of eight C. neoformans var grubii isolates
Genomic DNA was extracted as described above with the only modification that the cells were incubated with lysis buffer for only 30 min at 65°C and other 30 min at room temperature.

Oxford Nanopore libraries were constructed using the 1D ligation kit (SQK-LSK108) and loaded on a FLO-MIN106 flow cell for each sample on a Minion. Basecalling was performed using Albacore v2.0.2. A total of 63- to 314-fold average genome coverage was generated for each of the eight samples. ONT fastq reads were assembled using Canu (Koren et al., 2017) release v1.5 with the following parameters: -nanopore-raw < input.fastq > , correctedErrorRate = 0.075, and stopOnReadQuality = false. The Canu1 assembly was corrected using nanopolish (v0.8.1); reads were aligned to the Canu assembly using bwa-mem2 (Li, 2013) (0.7.12) with the "-x ont2d" parameter and nanopolish (https://github.com/jts/nanopolish) v0.9 variants in–consensus mode was run on each contig to generate a polished consensus. Next, assemblies were manually inspected and edited as follows. Contig ends without telomere repeats were aligned using NUCmer3 (Kurtz et al., 2004) to unassembled contigs and extended where alignments were uniquely mapped and contained both contig ends. Redundant small contigs with high identity NUCmer alignments to larger contigs were examined for read support; while all appeared redundant, some of these regions had higher accuracy and were used to correct the large contigs in two assemblies. Two manual joins between contigs were also made where supported by alignments to unassembled contigs. Alignment to the H99 genome4 (Janbon et al., 2014) with NUCmer was also used to validate these assemblies. All 8 assemblies showed the same rearrangement between H99 chromosomes 3 and 11; while seven of these assemblies appeared co-linear, Bt106 showed a unique rearrangement between H99 chromosomes 4 and 10 that was validated by read alignments. Finally, Illumina paired reads were aligned to the resulting polished contigs using bwa mem (version 0.7.7-r441) followed by Pilon5 (Walker et al., 2014) (version 1.13) correction using the–fix all setting; this step was repeated for a second round of alignment and Pilon polishing.

WGBS analysis of eight C. neoformans var grubii isolates
WGBS library preparation was performed as described above and sequenced as SE150. Two technical replicates of the same genomic DNA were prepared for each sample. Reads from the bisulfite-treated library were analyzed in parallel with Bismark (Krueger and Andrews, 2011) and BSseeker2 (Guo et al., 2013). Bismark alignment: bowtie2 with the options:–score_min L,0,0. Unconverted reads were removed using Bismark-filter_non_conversion and methylation was called using the standard setting. BSseeker2 alignment: bowtie2,–mismatches = 0–XS = 0.5, 5, call methylation: -x .

Bismark outputs were converted into CGmap file using cgmaptools-convert-bismark2cgmap (https://cgmaptools.github.io). The CGmap files from the technical replicates from BS-Seeker2 and Bismark were then cross-compared and only symmetric methylated CGs with at least 50% methylation and covered by > = 2 reads were considered in further analysis.

5mC calls on the Nanopore long reads data was performed using Nanopolish (Simpson et al., 2017) call-methylation and scripts/calculate_methylation_frequency.py.

The final table obtained was then converted into a CGmap file to be compared with our WGBS analysis. Since the calling 5mC with Nanopolish cannot distinguish methylated status of those CGs within 10bp of each other, the same methylation status was assigned to the entire group in the final CGmap. This results in a modest overestimation of the overall methylation status.

Analysis of 5mC conservation
Centromeres were pairwise-aligned to a reference (A1-35-8) using NUCmer and only the longest increasing subset was kept for each alignment (delta-filter -q -r). Methylation calls from WGBS (for whole genome methylation analysis) or WGBS-combined with ONT-methylation calls (for methylation analysis on transposons) were then projected onto these coordinates.

For the whole genome methylation analysis (Figure 7D), we binned sites with conserved CG sequence in X number of strains (from N = 2 to 8) into CG conservation groups and asked the relative rates of methylation in each bin. A null distribution was derived by permuting within each group to mimic the heterogeneity in methylation expected given sequence conservation. We permuted only the CGs with at least one strain methylated (Figure 7D). For example, to generate the null distribution for strain Bt106 where 4 sites are shared, all the CGs that Bt106 shares with 3 other strains were permuted (conditional on at least one being methylated). This resulted in a much more conservative estimate than the fully random model.

Cell

Transposon analysis

Tcns were annotated by aligning the transposable sequences of *C. neoformans var*. *neoformans* obtained from RepBase (https://www.girinst.org/repbase) to the ONT assembled genomes. A custom script was utilized to BLAST the Tcn sequences and allowing a minimum of 10% identity between queries and targets followed by manual annotation and verification. The coordinates of each Tcn are described in Table S5.

For the Tcn methylation analysis, a class of transposable elements was chosen (e.g., Tcn3), then the longest full-length Tcn in that class was chosen as a reference. Each Tcn was then aligned to the reference using Bioconductor pairwiseAlignment in global mode. WGBS were used to evaluate the methylation status. In cases where WGBS was not available (e.g., repetitive regions), we used the ONT calls to fill in the missing information. This is justified by high concordance between the two methods where there is bisulfite data (calculated as Pearson coefficient -Figure S7E). Independence of methylation and sequence was tested by computing a goodness-of-fit at each site relative to the reference, conditioned on the aligned sequence. The procedure follows: 1) the expected value for each site was assumed to be the relative rate of methylation for the respective repeat (e.g., the random uniform model); 2) the statistic for each repeat at a specific C/G is (observed methylation status – repeat methylation rate)$\hat{}$2/ repeat methylation rate; 3) the per repeat and site statistic was then the sum across all repeats and sites. The null distribution was generated by permuting methylation status within each repeat then recomputing the test statistic 1000 times.

Phylogenetic methods

For phylogenetic analysis, orthologs were identified across sequenced Tremellales (C.A.C., unpublished data) based on BLASTP pairwise matches with expect < 1e-5 using ORTHOMCL v1.4 (Li et al., 2003). A phylogeny was estimated from 4080 single copy genes as follows. Individual proteins were aligned using MUSCLE (Edgar, 2004) and the individual alignments were concatenated and poorly aligning regions removed with trimal (Capella-Gutiérrez et al., 2009). This sequence was input to RAxML (Stamatakis, 2006) version 7.7.8 (raxmlHPC-PTHREADS-SSE3) and a phylogeny estimated in rapid bootstrapping mode with model PROTCATWAG and 1,000 bootstrap replicates.

Mass Spectrometry analysis

Protein and peptide identification were done with Integrated Proteomics Pipeline – IP2 (Integrated Proteomics Applications). Tandem mass spectra were extracted from raw files using RawConverter (He et al., 2015) and searched with ProLuCID (Xu et al., 2015)against *Cryptococcus neoformans* UniProt database. The search space included all fully-tryptic and half-tryptic peptide candidates. Carbamidomethylation on cysteine was considered as a static modification. Data was searched with 50 ppm precursor ion tolerance and 600 ppm fragment ion tolerance. Identified proteins were filtered using DTASelect (Tabb et al., 2002) and utilizing a target-decoy database search strategy to control the false discovery rate to 1% at the protein level (Peng et al., 2003).

Proteins that did not appear in the untagged control that displayed at least 10% sequence coverage in both duplicates were considered hits in our analysis. Common contaminants (e.g., ribosomal proteins) were removed from the analysis. The complete dataset can be found in Table S4.

DATA AND CODE AVAILABILITY

All raw Oxford nanopore sequence and genome assemblies are available in NCBI under BioProject PRJNA517966. The accession number for WGBS and ChIP-seq reported in this study is NCBI GEO: GSE134684
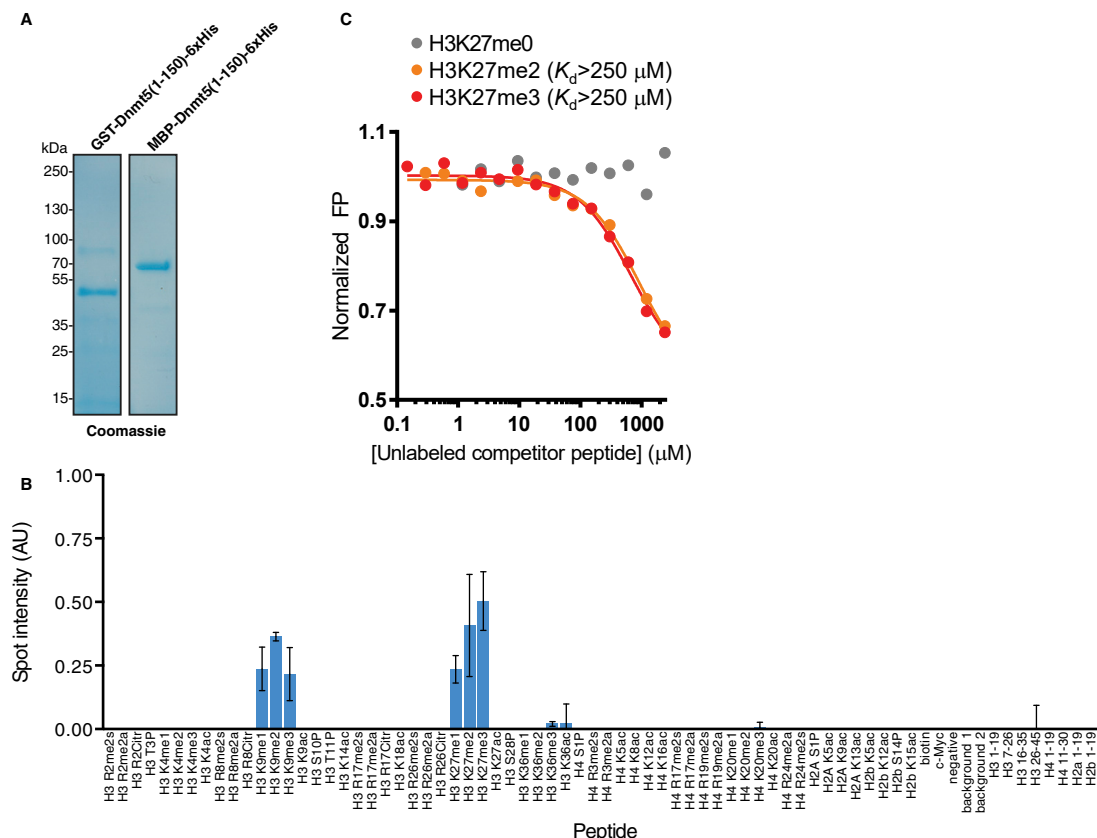
Figure S1. The Chromodomain of Dnmt5 Binds to H3K9me Peptides, Related to Figure 1

(A) Coomassie Brilliant Blue staining of GST-Dnmt5(1-150) and MBP-Dnmt5(1-150) after affinity purification from lysates of *E. coli* and resolution by SDS-PAGE. This truncation contains the Dnmt5 chromodomain.

(B) Dnmt5(1-150)-GST binding to a peptide microarray of human-derived histone sequences and modifications. Intensity represents background-subtracted signal at peptide array spots exhibiting the modifications indicated below the graph. Data are the average from two peptide arrays; error bars represent SE.

(C) Binding of MBP-Dnmt5(1-150) to H3K27 peptides, as assessed by a fluorescence polarization binding competition assay. Dnmt5(1-150) was bound to a fluorescently labeled H3K9me3 peptide in the presence of increasing concentrations of unlabeled H3K27me0/2/3 peptides corresponding to a region of *C. neoformans* H3. Polarization was normalized to that observed in the absence of competitor peptide; K_d calculated from 2-3 replicates.

A

wt | dmt5Δ | Dmt5_CAT | Dmt5_SNF

B

FLAG-Dnmt5 ChIP

100 200 300 400 500 600 700 Kb

wild-type

CHR13

ratio over WCE

clr4Δ

untag

C

- Wild-type (K_d=1.3±0.1 μM)
- W87A, Y90A

Normalized FP

[Dnmt5(1-150)] μM

D

CD   CS

HP1

6xHis-Swi6

kDa
185-
115-
80-
65-
50-
30-
25-

Coomassie

- H3K9me0
- H3K9me1 (K_d=20±6 μM)
- H3K9me2 (K_d=21±7 μM)
- H3K9me3 (K_d=11±3 μM)

Normalized FP

[Unlabeled competitor peptide] (μM)

E

Dnmt5 IP/MS

CNAG	peptide#	coverage	name
CNAG_07552	632	70.70%	Dnmt5
CNAG_03458	95	85.10%	Swi6
CNAG_04828	36	84.10%	H3.3
CNAG_00677	34	10.90%	Uhfr1
CNAG_01148	33	38.60%	Fkbp4
CNAG_05391	28	26.40%	Mif2
CNAG_03658	23	48.30%	
CNAG_08025	12	47.00%	Rnh1
CNAG_00083	12	39.50%	Ccc1
CNAG_07756	4	12.00%	Cdh1
CNAG_01636	2	10.30%	Cpr3

Swi6 IP/MS

CNAG	peptide#	coverage	name
CNAG_00998	118	42.10%	
CNAG_03458	56	56.10%	Swi6
CNAG_01292	39	30.50%	Mit1
CNAG_01306	32	21.70%	Mit2
CNAG_01563	29	18.30%	Clr3
CNAG_02922	18	18.10%	Clr2
CNAG_07807	13	36.90%	H4
CNAG_01648	13	36.90%	H4
CNAG_07564	11	12.80%	
CNAG_06746	10	37.40%	H2B
CNAG_01406	9	40.20%	
CNAG_04828	7	27.50%	H3.3
CNAG_06745	7	27.50%	H3
CNAG_06747	7	18.30%	H2A
CNAG_00063	4	11.30%	Cse4
CNAG_05221	3	15.90%	H2A.Z

Uhrf1 IP/MS

CNAG	peptide#	coverage	name
CNAG_00677	48	51%	Uhfr1
CNAG_04828	16	58.70%	H3.3

F

2xFLAG-Dmt5	+	-	+
Swi6-3xHA	-	+	+

FLAG — INPUT
HA — INPUT
FLAG — FLAG-IP
HA — FLAG-IP

lane 1 2 3

G

Clr4

K9

HP1

Dnmt5

K9

Uhrf1

Dnmt5

(legend on next page)

Figure S2. Interactions of Dnmt5 and Heterochromatin, Related to Figure 1

(A) Southern analysis of 5mC in cells carrying deletion of Dnmt5 (*dmt5Δ), a* mutation of Dnmt5 in its catalytic domain (*C440A*; Dnmt5-CAT) or in the ATPase domain (*K1469A*; Dnmt5-SNF). Southern hybridization was performed using a probe corresponding to a repetitive centromeric sequence (probe R).

(B) ChIP-seq analysis of FLAG-Dnmt5 in wild-type and *clr4Δ* backgrounds compared to untagged strain. Shown are the data for chromosome 13.

(C) Effect of chromodomain aromatic cage mutations on binding of MBP-Dnmt5(1-150) to H3K9 peptides. Wild-type or *W87AY90A* mutant MBP-Dnmt5(1-150) was incubated at increasing concentrations with H3K9me3 peptide. Polarization was normalized such that a value of 1 was equivalent to the polarization in the absence of MBP-Dnmt5(1-150) protein.

(D) Coomassie staining of 6xHis-Swi6/HP1 after purification from *E. coli* and resolution by PAGE. Left: chromodomain and chromoshadow domains of full-length Swi6 are indicated. Right: binding of Swi6 to H3K9 peptides, as assessed by fluorescence polarization binding competition assay. Swi6 was bound to labeled H3K9me3 peptide in the presence of increasing concentrations of unlabeled H3K9me0/1/2/3 peptides. Polarization was normalized to that observed in the absence of competitor peptide; K_d represents average ± 95% CI of 2-3 replicates.

(E) List of factors detected by Mass Spectrometry analysis. The proteins used as baits are marked in yellow.

(F) western blot of co-immunoprecipitation assay performed using FLAG antibody that recognize FLAG-Dnmt5. The presence of Swi6-HA in the immunoprecipitated proteins mix was detected by using anti-HA antibody.

(G) Schematic model of the factors involved in the maintenance of 5mC in *C. neoformans*.

Figure S3. Biochemical Analyses of Uhrf1 and Dnmt5, Related to Figures 1, 2, and 3

(A) Top: Domain structure of *C. neoformans* Uhrf1 ortholog. Bottom: Coomassie Brilliant Blue staining of Uhrf1-6xHis after purification from *E. coli* and resolution by SDS-PAGE. Predicted protein domains of full-length Uhrf1 are indicated.

(B) EMSA assessing competition between labeled hemimethylated DNA probe and excess concentrations unlabeled DNA (0.2, 1.0, 5.0 μM) of indicated methylation state. Where indicated, Uhrf1 protein was added at 150 nM. Graph indicates fraction of probe bound relative to condition in which no cold competitor was added; n = 2, error bars represent SD.

(legend continued on next page)

(C) Double stranded DNA substrates used in methyltransferase experiments. Each 20 bp substrate contains 6 CG sites that are either unmethylated, hemi-methylated, or symmetrically methylated.

(D) End point measurement of DNA methylation activities on methylated versus unmethylated substrates after a 6 hr reaction of 100 nM Dnmt5 and 5 μM DNA substrate. Background: no enzyme control. Error bars represent SD; n = 3.

(E) DNA methylation kinetics using 30 nM Dnmt5 and 5 μM DNA substrates containing either CHG or CG motifs in an unmethylated or hemimethylated state.

(F). Left: control DNA methylation reaction in which 30 nM Dnmt5 was incubated with 1 mM ATP and 5 μM of the unmethylated or hemimethylated DNA substrates. Center and right: DNA methylation reactions in which 30 nM Dnmt5 was incubated with 1 mM ATP and 5 μM unmethylated DNA in the presence or absence of 5 μM H3K9me0/3 peptide, 4 μM Uhrf1, and/or Swi6/HP1.

A

RT-qPCR (*DMT5*)

**B**

WGBS

C

D
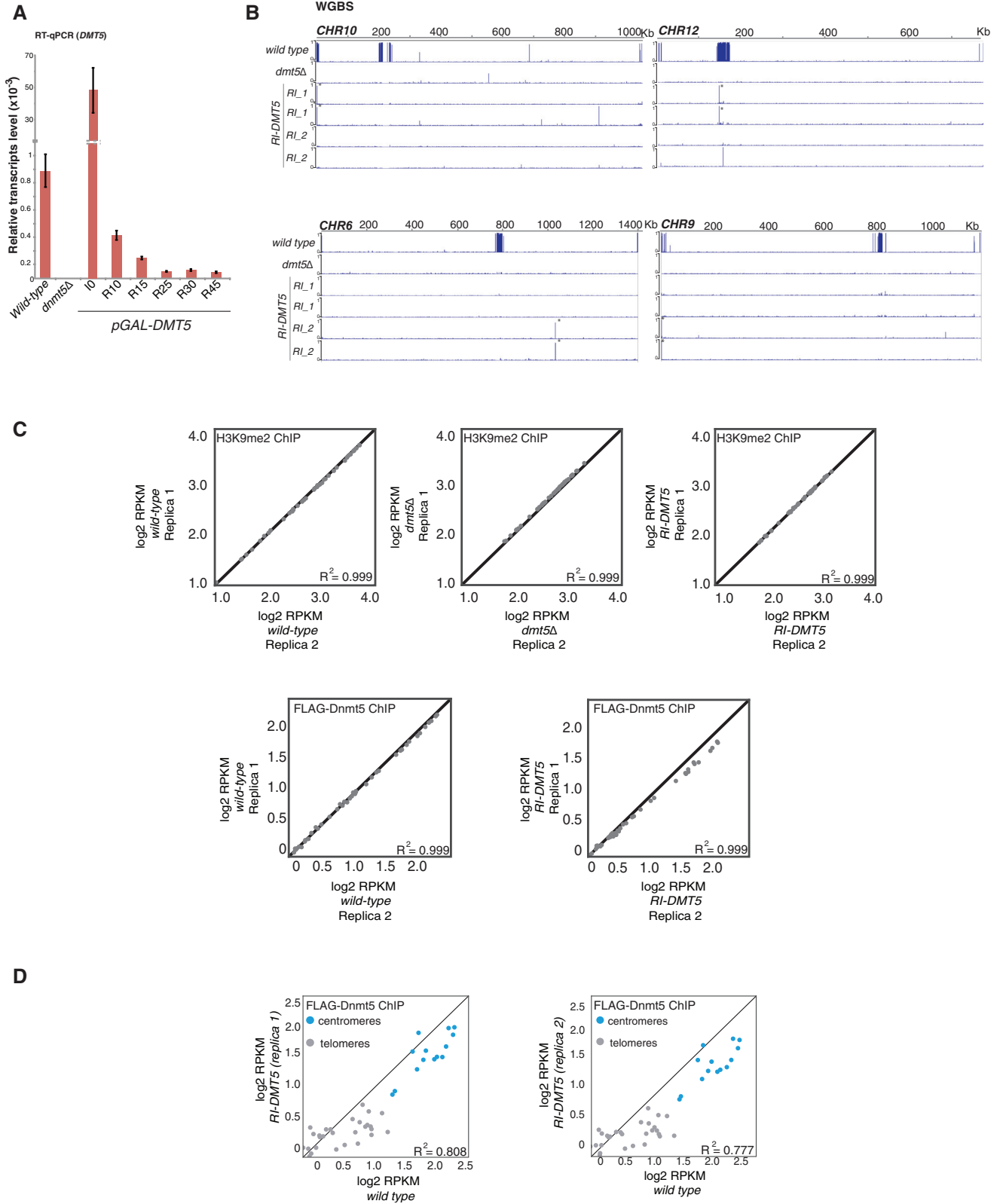
(legend on next page)

Figure S4. Analysis of Dnmt5 Function *In Vivo*, Related to Figures 3 and 4

(A) The promoter region of *DMT5* was replaced with a galactose-inducible promoter (as described in Figure 3C). *pGAL-DMT5* was repressed for 10, 15, 25, 30 and 45 generations (R10-R45) and mRNA levels of *pGAL-DMT5* were compared to *pGAL-DMT5* at the starting induced condition (I_0), wild-type cells (expressing *DMT5* under its own promoter), and cell lacking the methyltransferase gene (*dmt5Δ*). RT-qPCR was normalized over *ACT1*. n = 2, error bars represent SD.

(B) DNA from two independently derived strains (RI_1 and RI_2) were bisulfite treated in technical duplicate. Shown are the peaks present in both technical duplicates. Marked with asterisks are CG sites that are the symmetrically methylated and reproducible in both technical replicates.

(C) Scatterplot analysis of anti-H3K9me2 and anti-FLAG ChIP. Points represent the signals for each centromere and telomere (RPKM). Each sample was compared against a biological replicate.

(D) Scatterplot analysis of anti-FLAG ChIP of *RI-DMT5* compared to wild-type. Pairwise comparison of the total RPKM counts for each centromere (blue) and telomere (gray). Analyses of both biological replicates are shown.
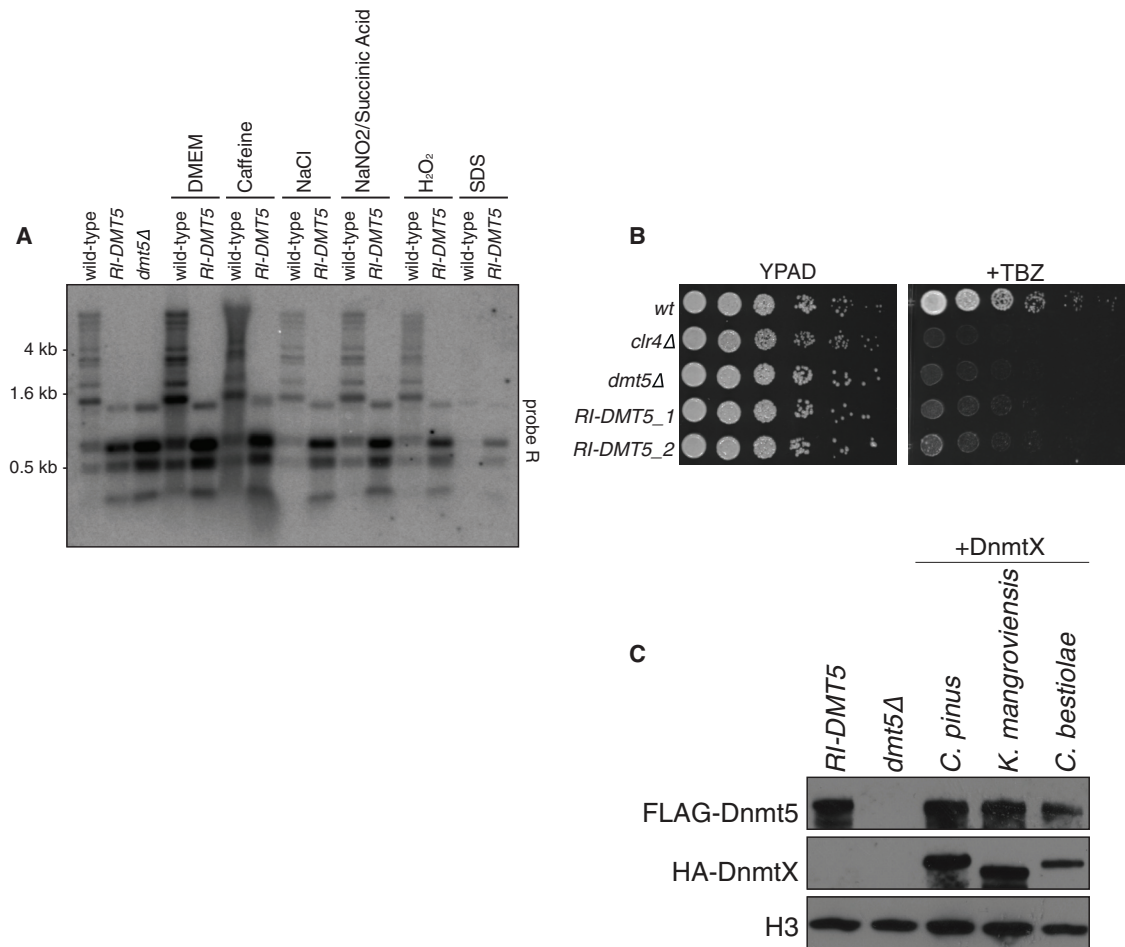
Figure S5. Impact of Stress on Dnmt5 Function and Phenotype and Analysis of DnmtX Expression *In Vivo*, Related to Figures 4 and 5

(A) To assess if specific stress conditions induce *de novo* DNA methylation activity, wild-type and RI-Dnmt5 cells were grown in the presence of stressors The genomic DNA was extracted, digested with the CG methylation sensitive endonuclease HpyCHIV and analyzed by Southern hybridization using a probe corresponding to a repetitive centromeric sequence (probe R).

(B) Spotting assay of wild-type, *clr4Δ*, *dmt5Δ* and *RI-DMT5* strains on rich-medium and in the presence of the microtubule-depolymerizing drug thiabendazole (TBZ; 11μg/ml).

(C) HA-tagged version of DnmtX from *K. pinus, K. mangroviensis* and *K. bestiolae* were expressed under the control of *GAL7* promoter in an *RI-DMT5* strain. Expression of the proteins was assessed by western blotting using indicated antibodies.
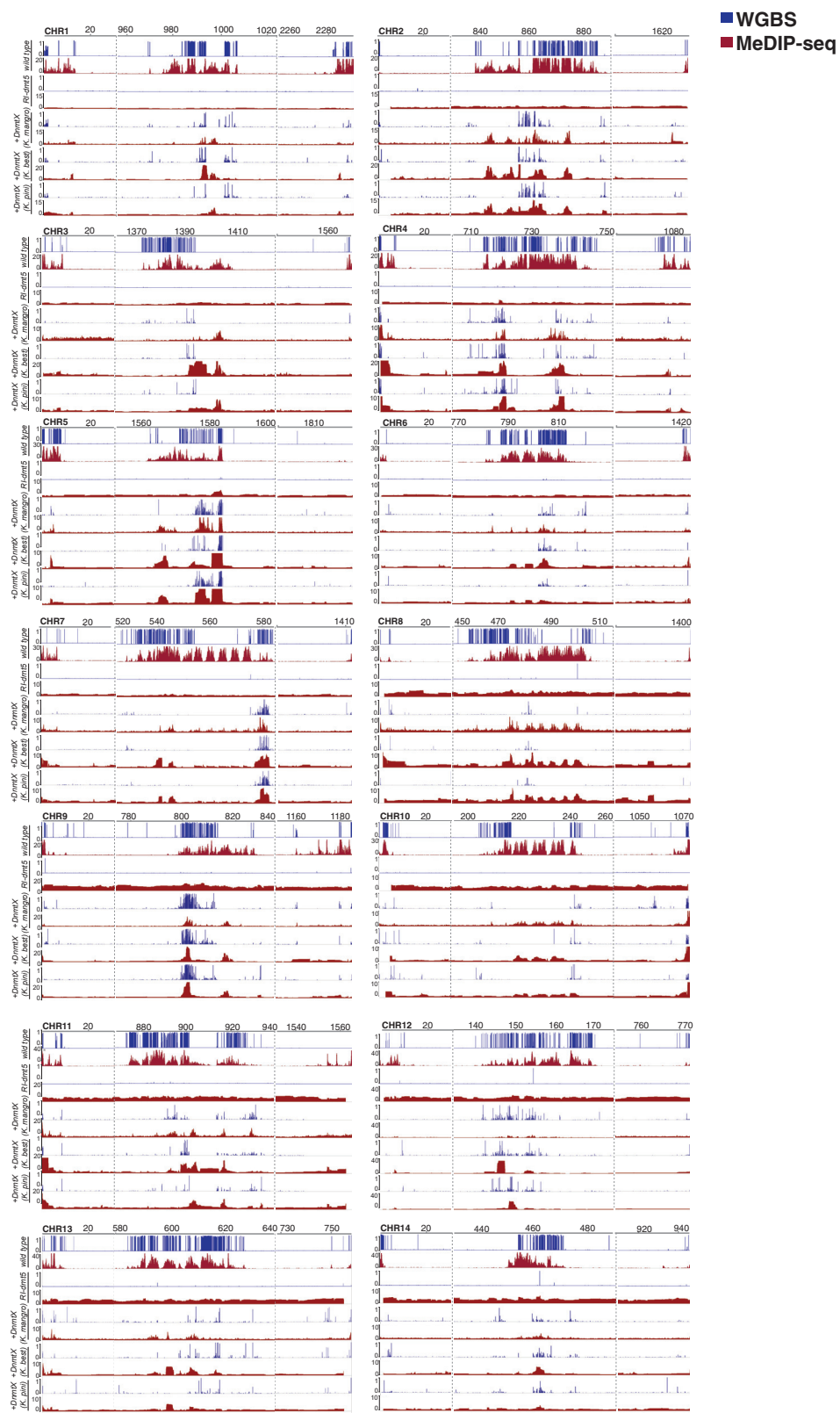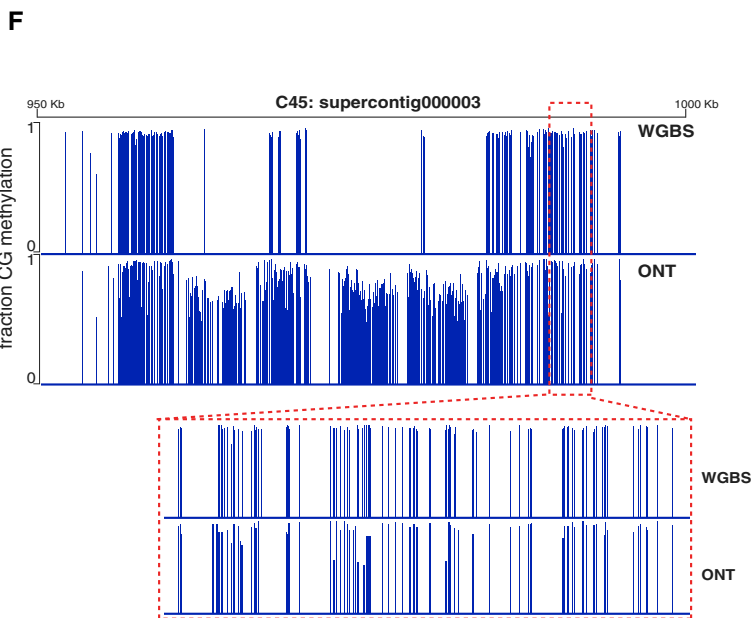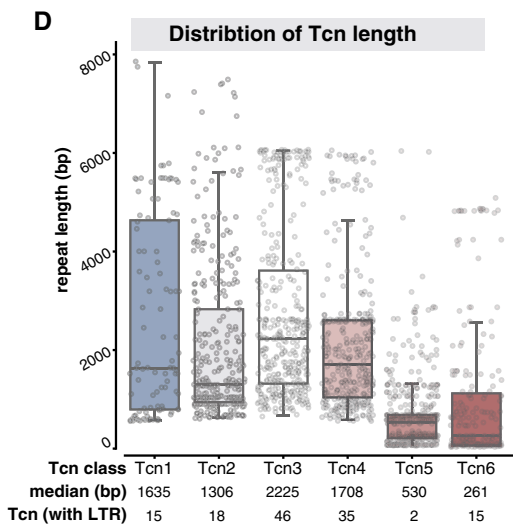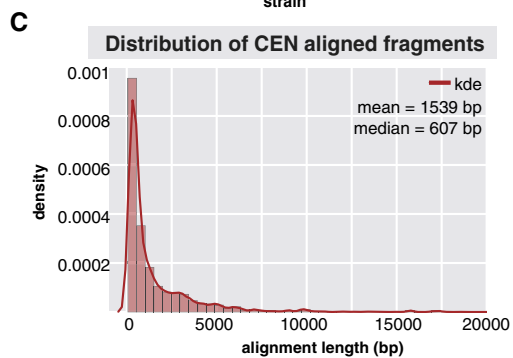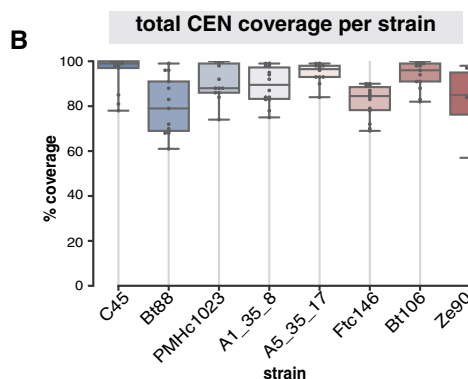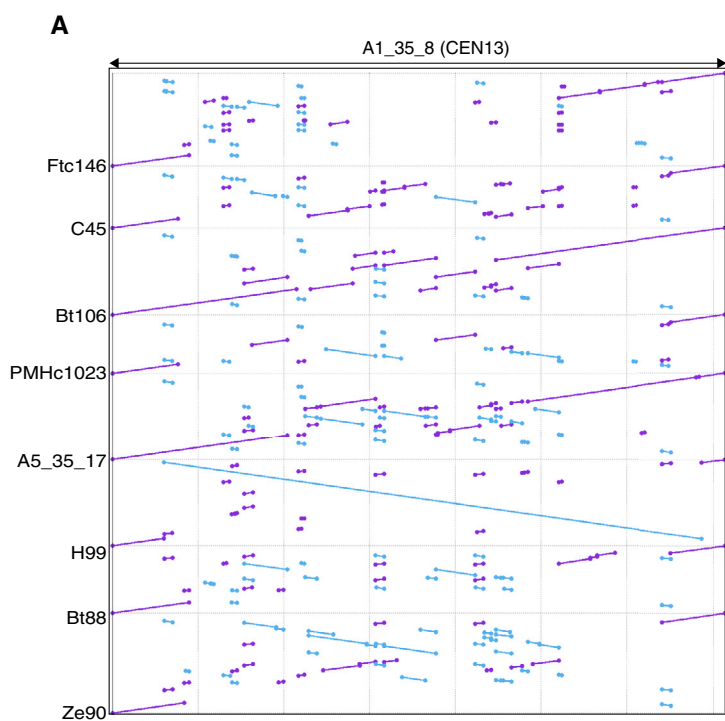
Cell

Figure S6. DnmtX Acts as a *De Novo* DNA Methyltransferase in Cells Expressing RI-Dnmt5, Related to Figure 5
BS-seq (blue) MeDIP (red) signals for centromeres and telomere (*CEN1-14, TELR/L1-14*) from cells expressing DnmtX (+DnmtX) compared to wild-type and the control without DnmtX (top). Strains were analyzed after 14 days of induction of DnmtXs on YP-galactose solid media. The signals shown normalized to that of a *dmt5Δ* strain.

A

A1_35_8 (CEN13)

Ftc146

C45

Bt106

PMHc1023

A5_35_17

H99

Bt88

Ze90

B

total CEN coverage per strain

C

Distribution of CEN aligned fragments

kde
mean = 1539 bp
median = 607 bp

D

Distribtion of Tcn length

Tcn class	Tcn1	Tcn2	Tcn3	Tcn4	Tcn5	Tcn6
median (bp)	1635	1306	2225	1708	530	261
Tcn (with LTR)	15	18	46	35	2	15

F

C45: supercontig000003

950 Kb   1000 Kb

WGBS

ONT

WGBS

ONT

E

WGBS vs ONT	
strain	Pearson coefficient
A1_35_8	0.839
A5_35_17	0.856
Bt106	0.894
Bt88	0.855
C45	0.791
Ftc146	0.885
PMHc1023	0.901
Ze90_1	0.877

G

WebLogo 3.7.4

(legend on next page)
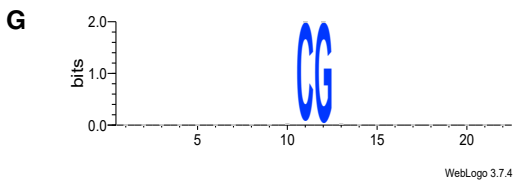
Figure S7. Evolutionary Conservation of 5mC, Related to Figure 7

(A) Example of dot plot analysis of centromere alignment. Here, *CEN13* from the A1_35_8 strain was aligned to the centromeres of the other 7 species. In purple, direct alignment and in light blue, inversions.

(B) Percentage of centromeric sequence covered by the pairwise alignment. Each dot in the box represents a centromere.

(C) Distribution of the fragment lengths produced by pairwise alignment of centromeres. Red line: kernel density estimation (kde).

(D) Six families of transposable elements (Tcn1-6) were identified in the analysis. Shown is the distribution of length of the identified Tcn for each family (SE approximated by Bernoulli distribution).

(E) Pearson correlations of methylation data from WGBS and Nanopore sequencing data (ONT) for the 8 strains analyzed.

(F) Centromeric region of contig000014 in strain C45, as an example of 5mC calls derived from WGBS and ONT analysis. Data missing from the WGBS but present in ONT, correspond to highly repetitive regions that are not available to the WGBS pipeline.

(G) WebLogo representation of the conserved motif identified from the alignment of methylated sites across the eight genomes described in Figure 7.