bioRxiv preprint doi: https://doi.org/10.1101/149385; this version posted June 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Epigenetic maintenance of DNA methylation after evolutionary loss of the *de novo* methyltransferase

Sandra Catania[1], Phillip A. Dumesic[1], Caitlin Stoddard[1], Sophie Cooke[1], Jordan Burke[1], Christina A. Cuomo[3], Geeta J. Narlikar[1], Hiten D. Madhani[1,2,*]

[1]Department of Biochemistry and Biophysics
University of California
San Francisco, California 94158
USA

[2]Chan-Zukerberg Biohub
San Francisco, California 94158

[3]Infectious Disease and Microbiome Program
Broad Institute of MIT and Harvard
Cambridge, Massachusetts 02142
USA

[*]Corresponding author
hitenmadhani@gmail.com
tel (415) 514-0594
fax (415) 514-0859

bioRxiv preprint doi: https://doi.org/10.1101/149385; this version posted June 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

ABSTRACT

After the initial establishment of symmetric cytosine methylation patterns by *de novo* DNA methyltransferases (DNMTs), maintenance DNMTs mediate epigenetic memory by propagating the initial signal. We find that CG methylation in the yeast *Cryptococcus neoformans* is dependent on a purely epigenetic mechanism mediated by the single DNMT encoded by the genome, Dnmt5. Purified Dnmt5 is a maintenance methyltransferase that strictly requires a hemimethylated substrate, and methylation lost by removal of Dnmt5 *in vivo* is not restored by its mitotic or meiotic reintroduction. Phylogenetic analysis reveals that the ancestral species had a second methyltransferase, DnmtX, whose gene was lost between 50 and 150 Mya. Expression of extant DnmtXs in *C. neoformans* triggers *de novo* methylation. These data indicate that DNA methylation has been maintained epigenetically by the Dnmt5 system since the ancient loss of the DnmtX *de novo* enzyme, implying remarkably long-lived epigenetic memory.

Single sentence summary*: Epigenetic information can be inherited over geological timescales*

bioRxiv preprint doi: https://doi.org/10.1101/149385; this version posted June 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

INTRODUCTION

Methylation of cytosine on its fifth carbon (5mC) in DNA is found in all domains of life. In eukaryotes, it plays a role in epigenetic memory(*1*). 5mC can be deposited by *de novo* enzymes that act on an unmethylated substrate. This modification can be 'remembered' by maintenance enzymes that function primarily on hemimethylated DNA produced by DNA replication(*2*). One example of this division of labor is found in the human DNA methylation system. Here, *de novo* methylation on CG dinucleotides is principally catalyzed by the Dnmt3a/b-Dnmt3L complex. Once the DNA methylation mark is established, it is then maintained primarily by the Dnmt1 enzyme in association with UHRF1, a protein that recognizes H3K9me and hemimethylated 5mC(*2, 3*). DNA methylation patterns are essential for development and play critical roles in silencing of transposable elements, chromosome stability, monoallelic gene expression, and gene silencing(*4*). In principle, the ability of 5mC patterns to be copied enables this type of epigenetic information to be inherited transgenerationally. Indeed, in plants there are examples of gene silencing mediated by DNA methylation that can be inherited through the germline(*5*). Hence, it has been suggested that such epigenetic variation in plant populations might be sufficiently stable to be subject to natural selection(*6, 7*). These observations raise a general question: how long can 5mC-mediated epigenetic memory be maintained after its initial establishment? Here, we investigate this question using a yeast DNA methylation system.

FACTORS THAT PROMOTE 5mC IN *CRYPTOCOCCUS NEOFORMANS*

The human fungal pathogen *Cryptococcus neoformans* is a basidiomycetous yeast that harbors symmetric CG methylation at centromeric and subtelomeric regions(*8*). This methylation is

bioRxiv preprint doi: https://doi.org/10.1101/149385; this version posted June 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

dependent on a single predicted cytosine DNA methyltransferase (DNMT), Dnmt5, encoded by the *DMT5* gene(*8*). The Dnmt5 family of proteins is characterized by an N-terminal chromodomain (CD) followed by a cytosine methyltransferase catalytic domain, a RING finger, and domain related to those of SNF2-type ATPases (Fig.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(*8*).

The regions of the genome decorated with 5mC in *C. neoformans* coincide with those we reported to display H3K9me(*8, 9*). 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 (Fig S1, A and B) yielded signals with H3K9me-containing peptides as well as those containing H3K27me. 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 (Fig. 1B and Fig S1C). Binding to H3K9me (Kd=1.5 μM for H3K9me3) was substantially stronger than binding to H3K27me (Kd>250μM), suggesting that H3K9me is the principal binding partner of the Dnmt5 chromodomain.

To assess whether H3K9me affects 5mC, we deleted the gene encoding the only H3K9 methyltransferase in *C. neoformans*, Clr4 (*clr4*Δ)(*9*). 5mC was assayed by digestion of genomic DNA with a methylation-sensitive enzyme (HpyCH4IV; A^CGT) followed by Southern hybridization. Based on published maps of 5mC in wild-type *C. neoformans*, we designed a hybridization probe directed to a unique region of centromere 13 (Fig. 1C, probe U). In addition, another probe recognizing repetitive centromeric sequences (probe R) was created to analyse the distribution of 5mC across several centromeric sites at once.

bioRxiv preprint doi: https://doi.org/10.1101/149385; this version posted June 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

In the regions probed, 5mC levels are markedly reduced in the *clr4*Δ strain (Fig 1D) and ChIP-seq analysis of FLAG-Dnmt5 demonstrated that its recruitment to H3K9me domains is dramatically reduced (Fig. S2A). In contrast, a mutation of the Dnmt5 chromodomain that abolishes its binding to H3K9me (*W87AY90A*) has only a minor effect on 5mC, suggesting that other H3K9me-dependent factors contribute to 5mC levels (Fig. 1E and Fig. S2B). The heterochromatin protein 1 (HP1) family comprises several conserved proteins that, by binding to H3K9me, participate in heterochromatin formation and in some organisms contribute to 5mC methylation(*10-12*). In *C. neoformans*, we identified a single HP1 ortholog (CNAG_03458), which we named Swi6 based on its characterized ortholog in fission yeast (Fig. S2C). While the *swi6*Δ mutant does not detectably impact 5mC, a *swi6*Δ *dmt5-W87AY90A* double mutant reduces the levels of 5mC to those seen in the *clr4*Δ strain (Fig. 1E, Fig S2D). This result suggests that H3K9me promotes 5mC by recruiting Dnmt5 in two ways: via the CD of Dnmt5 and via HP1.

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, Fig.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 to bind DNA that is unmethylated, hemimethylated or symmetrically methylated. This experiment demonstrates that Uhrf1 selectively binds hemimethylated DNA (Fig. 1F), a result confirmed by competition experiments (Fig. S3). The absence of Uhrf1 (*uhf1*Δ) has almost no detectable impact on 5mC *in vivo*, but modestly alters 5mC patterns at sites probed when combined with a loss of Clr4 (*clr4*Δ *uhf1*Δ*)* (Fig. 1G).

bioRxiv preprint doi: https://doi.org/10.1101/149385; this version posted June 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

To assay these genotypes genome-wide, we performed whole-genome bisulfite-sequencing (WGBS) of DNA extracted from wild-type, *dmt5*Δ, *clr4*Δ, *uhf1*Δ and *clr4*Δ *uhf1*Δ strains (Fig. 1H). The total number of symmetrically methylated CG dinucleotides, which is reduced to 31% in *clr4*Δ, drops to 11% in the *clr4*Δ *uhf1*Δ double mutant (Fig. 1I). This analysis indicates that Uhrf1 and Clr4 promote 5mC through two parallel pathways (Fig 1J).

PURIFIED DNMT5 IS A MAINTENANCE METHYLASE

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-[^{3}H-methyl]-adenosyl-methionine donor and synthetic 60bp 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 methylated on both stands. In the absence of ATP, no activity was observed on any substrate, as compared to control reactions that lacked Dnmt5 (Fig. 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 (Fig. 2A, 2B)(*13*). Endpoint assays with the unmethylated substrate did not reveal any activity on unmodified DNA; under these conditions we observed signal only with hemimethylated substrate that is ~1300-fold over background. (Fig. S4B). In addition, Dnmt5 activity is specific for CG dinucleotides since no activity was detected on DNA substrates containing CHG motifs instead of CG motifs, regardless of cytosine methylation state (Fig. S4C).

bioRxiv preprint doi: https://doi.org/10.1101/149385; this version posted June 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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 (Fig. 2C and D). We also tested whether recombinant Uhrf1 and Swi6 could trigger a *de novo* activity when added to the reactions, but none could be detected (Fig. S4D).

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 MAINTAINED EPIGENETICALLY IN *CRYPTOCOCCUS NEOFORMANS*

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 globally restored as its hemimethylated substrate would be lost.

We placed the gene for Dnmt5 under the control of a galactose-inducible promoter (Fig. 3A-*pGAL7- 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 *pGAL7* promoter (*R*). After induction of the Dnmt5 protein using the inducer galactose for 20 (I_{20}) or 45 generations (I_{45}), we

bioRxiv preprint doi: https://doi.org/10.1101/149385; this version posted June 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

were not able to detect 5mC by Southern hybridization, although expression of the construct was evident (Fig. 3A). To assess if the *pGAL7-DMT5* allele is functional, the targeting construct was transformed into wild-type cells but this time recipient cells were selected under conditions that induce *pGAL7* (Fig. 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 indeed functional. Subsequently, upon repression of *pGAL7-DMT5*, 5mC was lost. However, again, re-induction of Dnmt5 for 40 (I_{40}) or 90 generations (I_{90}) did not restore any 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 *pGAL7-DMT5* for limited number of generations (*R*) and re-induced (*I*) it. Again, we failed to observe restoration of 5mC to wild-type levels (Fig. 3C).

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. (Fig. 4; the re-introduced allele is termed *RI-DMT5;* see materials and methods). Although the expression of *RI-DMT5* was comparable to that of a wild type allele, we were not able to detect any 5mC by Southern hybridization (Fig. 4, A and B). WGBS sequencing of the *dmt5*Δ strain revealed no methylation on any CG sites genomewide (Fig 4B). Analysis of two independently-derived *RI-DMT5* strains revealed that methylation is also globally absent save for two sites detected in each strain (Fig. 4B, S5). This trace activity suggests that Dnmt5 does not have infinite specificity for hemimethylated DNA. H3K9me is still globally maintained in both the *dmt5*Δ strain as well as in the *RI-DMT5* strain, although the distribution is altered compared to wild-type depending on the genomic region with a notable increase of the signal at subtelomeric regions (Fig. 4C, Fig. S6). Furthermore, ChIP-seq experiments demonstrate that the *RI-DMT5* allele still localizes to heterochromatic regions (Fig. 4D). To test explicitly if *RI-DMT5* is capable of

bioRxiv preprint doi: https://doi.org/10.1101/149385; this version posted June 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

efficiently maintaining DNA methylation, wild-type strains whose centromere 13 was marked with a linked drug resistance marker (*CEN13::natR*) was crossed to the *RI-DMT5* (Fig. 4E). 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, indicating that the *RI-DMT5* allele is highly 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 that this silencing occurs via RNAi and not via DNA methylation(*14*). Nonetheless, to determine whether an unmethylated centromere can be methylated during sexual reproduction, we performed two genetic crosses and analysed the meiotic progeny. In a control cross, we crossed a *CEN13:natR* strain with a *dmt5*Δ strain and analysed three progeny that harbored 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 (Fig. 4F). In the experimental cross, we crossed a wild-type strain to one harboring a *dmt5*Δ allele that also carries *CEN13::natR.* We again analysed three progeny that express the wild-type *DMT5* gene and the marked centromere, which entered the cross in an unmethylated state. We observed that 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 its methylated state (Fig. 4G). Thus, sexual reproduction does not lead to the efficient restoration of 5mC. Finally, we tested several stress conditions to determine whether they could reestablish methylation in the *RI-DMT5* strains. Again, the results were negative (Fig. S7).

bioRxiv preprint doi: https://doi.org/10.1101/149385; this version posted June 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

THE GENE FOR *DE NOVO* DNA METHYLTRANSFERASE WAS LOST IN A *C. NEOFORMANS* ANCESTOR

If Dnmt5 is a pure maintenance enzyme, how was 5mC established? A clue came from an analysis of genomes in the family of organisms to which *Cryptococcus neoformans* belongs, the *Tremellaceae* (Fig. 5). The genomes of species closely related to *C. neoformans*, including the human pathogen *Cryptococcus gattii* and the nonpathogens *Cryptococcus amylolentus* and *Tsuchiyaea wingfieldii*, harbor a single predicted DNA methyltransferase, which is orthologous to Dnmt5 (Fig. 5A). The next closest species, *Filobasidella depauperata*, lacks predicted DNA methyltransferases, indicating loss of *DMT5* during its evolution (Fig. 5B). Strikingly, more distant species encode both Dnmt5 and an uncharacterized predicted DNA methyltransferase, which we term DnmtX (genetic locus: *DMX1*; Fig. 5B). This predicted protein contains a bromo-associated homology (BAH) domain and a Dnmt catalytic domain (Fig 5B). This phylogenetic pattern indicates that the *DMX1* gene was lost in the common ancestor to *C. neoformans* and *F. depauperata* and that *F. depauperata* subsequently lost the *DMT5* gene. Given the fact that Dnmt5 is a maintenance methylase, we hypothesized that DnmtX is the *de novo* methyltransferase that produced the original precursor to the hemimethylated substrate for Dnmt5, which has since maintained 5mC.

To assess if DnmtX is a *de novo* methyltransferase, genes for DnmtX were cloned from *Kwoniella mangroviensis, Cryptococcus bestiolae* and *Cryptococcus pinus*. 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-DMT*5). We chose this approach because we predicted that establishment of 5mC by a

bioRxiv preprint doi: https://doi.org/10.1101/149385; this version posted June 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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 (Fig. S8A) and DNA was then extracted. 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. By both assays, we observed the broad accumulation of 5mC in each of the three strains, primarily at centromeric regions (Fig. 5C and Fig. S8B). These data indicate that all three DnmtXs act as *de novo* methylase *in vivo*, and provides strong evidence that the ancestral DnmtX also had this function.

EPIGENETIC MEMORY OVER GEOLOGICAL TIMESCALES

The loss of the ancestral DnmtX gene occurred prior to the divergence of *C. neoformans* and *F. depauerata* but after the divergence of *Tremella mesenterica* and *Cryptococcus hevanensis* (Fig. 5). The divergence time for *C. neoformans* and *C. gattii* is estimated to be between 34 and 49 Mya(*15, 16*). Given the phylogenetic relationships, the divergence time of *C. neoformans* and *F. depauperata* would be considerably more ancient. It has been estimated that the common ancestor of *C. neoformans* and *T. mesenterica* lived 153 Mya. Thus the *DMX1* gene was likely to have been lost between roughly 150 and 50 Mya. This largely overlaps with the Cretaceous period (145-66 Mya). Taken together, these data indicate that 5mC has been epigenetically maintained by Dnmt5 over a remarkably long period of time in the population that gave rise to the pathogenic *Cryptococcus* species complex.

bioRxiv preprint doi: https://doi.org/10.1101/149385; this version posted June 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

This conclusion raises a number of questions. The first concerns the accuracy of inheritance of 5mC mediated by Dnmt5 and its cofactors. Although the error rate of replicating DNA methylation patterns is unknown, the apparently redundant role of the CD of Dnmt5, Swi6, Uhrf1 and additional factors likely conspire to produce a high fidelity of inheritance of 5mC. Given the fact that Dnmt5 is required for fitness in some environments (*17*), natural selection for its function (which may include roles in centromere function and transposon repression) provides an additional mechanism for the long-term maintenance of 5mC. In other words, if a loss of methylation at a CG dinucleotide due to an error in inheritance produced in a reduction in fitness, individuals that experienced this loss would be purified out of the population by selection. Our phylogenetic analysis indicates that *F. depauperata* lost the Dnmt5 gene after the loss of the DnmtX gene. This sequence of events suggests that that the epigenetic maintenance of 5mC in this lineage likely failed at some point after the loss of DnmtX, which would then make Dnmt5 superfluous. Phylogenetic studies of the 5mC patterns themselves across *C. neoformans* isolates or related species will likely further inform our understanding of the evolutionary dynamics of its inheritance.

Organisms that harbour DnmtX (Fig. 5) may provide tractable yeast systems for studying the mechanistic basis for *de novo* DNA methylation. Indeed, a complete sexual cycle has been demonstrated for *C. heaveanesis*, providing the beginnings of genetics in this system(*18*). Whether the loss of DnmtX was compensated by a change in the activity of another system with overlapping functions is an intriguing question as is how the *F. depauperata* lineage has survived in the absence of both enzymes. Elucidating these dynamics will likely inform how other lineages have survived the loss of 5mC, such as those that gave rise to the model organisms *C. elegans* and *D. melanogaster*.

bioRxiv preprint doi: https://doi.org/10.1101/149385; this version posted June 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

The ability of 5mC to be maintained without the DnmtX *de novo* system over million-year timescales raises the possibility that there exist correspondingly stable epialleles in populations, particularly in the fungi and algae that harbour the Dnmt5 system. If they exist, they could represent a large cache of inherited information on which natural selection might operate. Such variation would be invisible to approaches that assume a basis in DNA sequence for traits that are heritable over geological timescales.

bioRxiv preprint doi: https://doi.org/10.1101/149385; this version posted June 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

References and Notes:

1. A. Bird, in *Genes & Development*. (2002), vol. 16, pp. 6-21.

2. J. A. Law, S. E. Jacobsen, in *Nat Rev Genet*. (2010), vol. 11, pp. 204-220.

3. M. Bostick *et al.*, in *Science*. (2007), vol. 317, pp. 1760-1764.

4. P. A. Jones, in *Nat Rev Genet*. (2012), vol. 13, pp. 484-492.

5. R. J. Schmitz, J. R. Ecker, in *Trends in Plant Science*. (2012), vol. 17, pp. 149-154.

6. F. Johannes *et al.*, in *PLoS Genet*. (2009), vol. 5, pp. e1000530.

7. A. van der Graaf *et al.*, in *Proceedings of the National Academy of Sciences*. (2015), vol. 112, pp. 6676-6681.

8. J. T. Huff, D. Zilberman, in *Cell*. (2014), vol. 156, pp. 1286-1297.

9. P. A. Dumesic *et al.*, in *Cell*. (2015), vol. 160, pp. 204-218.

10. A. J. Bannister *et al.*, in *Nature*. (2001), vol. 410, pp. 120-124.

11. M. Lachner, D. O'Carroll, S. Rea, K. Mechtler, T. Jenuwein, in *Nature*. (2001), vol. 410, pp. 116-120.

12. Z. A. Lewis *et al.*, in *PLoS Genet*. (2010), vol. 6, pp. e1001196.

13. S. Pradhan, A. Bacolla, R. D. Wells, R. J. Roberts, in *J. Biol. Chem.* (1999), vol. 274, pp. 33002-33010.

bioRxiv preprint doi: https://doi.org/10.1101/149385; this version posted June 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

14. X. Wang *et al.*, in *Genes & Development*. (2010), vol. 24, pp. 2566-2582.

15. C. A. D'Souza *et al.*, in *MBio*. (2011), vol. 2, pp. e00342-00310.

16. P. Ngamskulrungroj *et al.*, in *PLoS ONE*. (2009), vol. 4, pp. e5862.

17. O. W. Liu *et al.*, in *Cell*. (2008), vol. 135, pp. 174-188.

18. B. Metin, K. Findley, J. Heitman, in *PLoS Genet*. (2010), vol. 6, pp. e1000961.

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

20. M. A. Urich, J. R. Nery, R. Lister, R. J. Schmitz, J. R. Ecker, in *Nat Protoc*. (2015), vol. 10, pp. 475-483.

21. W. Guo *et al.*, in *BMC Genomics*. (2013), vol. 14, pp. 774.

22. C. M. Homer *et al.*, in *Cell Host Microbe*. (2016), vol. 19, pp. 849-864.

23. B. Z. Harris, B. J. Hillier, W. A. Lim, in *Biochemistry*. (2001), vol. 40, pp. 5921-5930.

24. A. Reményi, M. C. Good, R. P. Bhattacharyya, W. A. Lim, in *Molecular Cell*. (2005), vol. 20, pp. 951-962.

25. M. Li *et al.*, in *Journal of Molecular Biology*. (2009), vol. 385, pp. 820-830.

26. J. E. Lindsley, J. C. Wang, in *J. Biol. Chem.* (1993), vol. 268, pp. 8096-8104.

bioRxiv preprint doi: https://doi.org/10.1101/149385; this version posted June 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

27. L. R. Pack, K. R. Yamamoto, D. G. Fujimori, in *J. Biol. Chem.*, vol. 291, pp. 6060-6070 (2016)

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

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

30. S. Capella-Gutierrez, J. M. Silla-Martinez, T. Gabaldon, trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* **25**, 1972-1973 (2009).

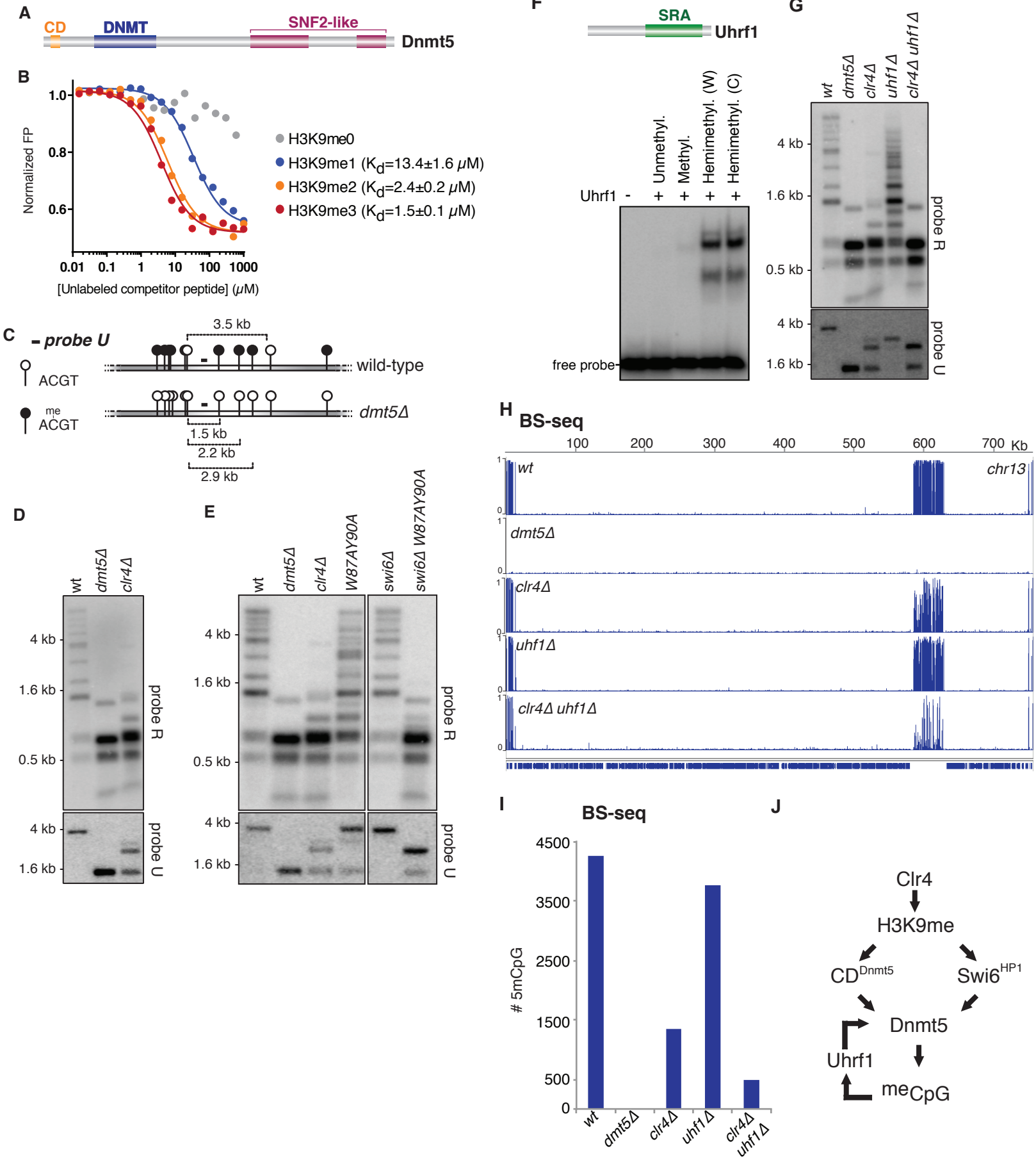
31. A. Stamatakis, RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688-2690 (2006).

Acknowledgments

We thank the members of the Madhani lab for scientific discussions and support, Yin Shen (UCSF) for methylated adaptors and advice on WGBS, and Matteo Pelligrini (UCLA) for advice on bioinformatics analysis of WGB. S.C. was supported by an EMBO Postdoctoral Fellowship. Research in the Madhani laboratory is supported by grants from the U.S. National Institutes of Health. H.D.M. is a Chan-Zukerberg Biohub Investigator. .

bioRxiv preprint doi: https://doi.org/10.1101/149385; this version posted June 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 1. Catania et al.



bioRxiv preprint doi: https://doi.org/10.1101/149385; this version posted June 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Fig. 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, *dmt5*Δ and *clr4*Δ assessed by Southern hybridization of HpyCH4IV-digested genomic DNA using probe corresponding to a repetitive sequence (probe R) or a unique sequence (probe U; panel B).

(**E**) Southern analysis of 5mC in cells carrying CD mutant of Dnmt5 (*W87AY90A*), deletion of Swi6 (*swi6*Δ) or a combination of the two (*swi6*Δ *W87AY90A*).

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

(**G**) Southern analysis of 5mC in wild-type, *dmt5*Δ, *clr4*Δ, *uhf1*Δ and double mutant *clr4*Δ *uhf1*Δ strains.

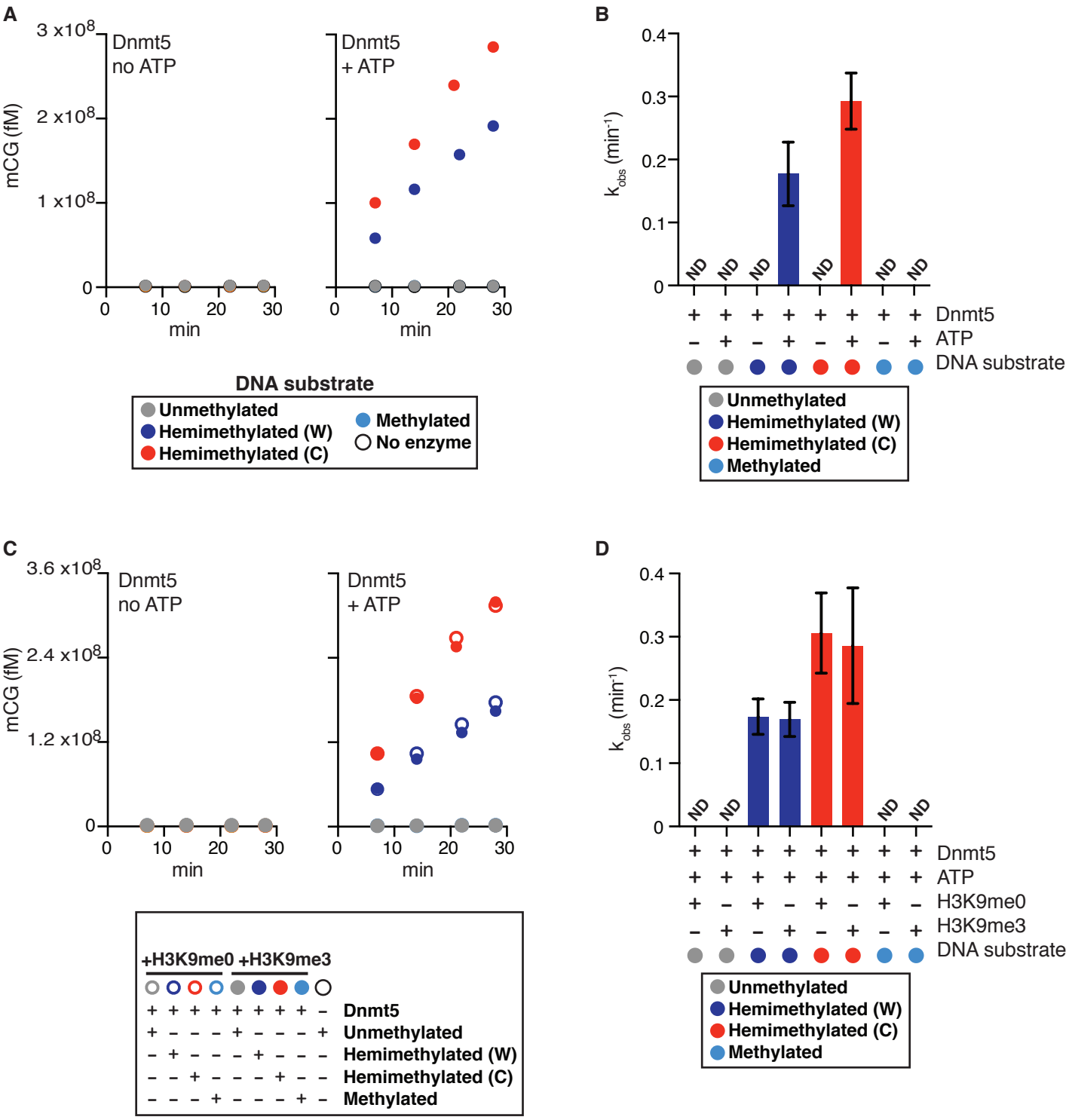
(**H**) Whole-genome bisulfite sequencing (WGBS) analysis of wild-type, *dmt5*Δ, *clr4*Δ, *uhf1*Δ and *clr4*Δ *uhf1*Δ strains. Shown are the data for chromosome 13.

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

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

bioRxiv preprint doi: https://doi.org/10.1101/149385; this version posted June 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 2. Catania et al.



bioRxiv preprint doi: https://doi.org/10.1101/149385; this version posted June 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Fig. 2. **Hemimethylated DNA but not unmethylated DNA is a substrate of purified Dnmt5**

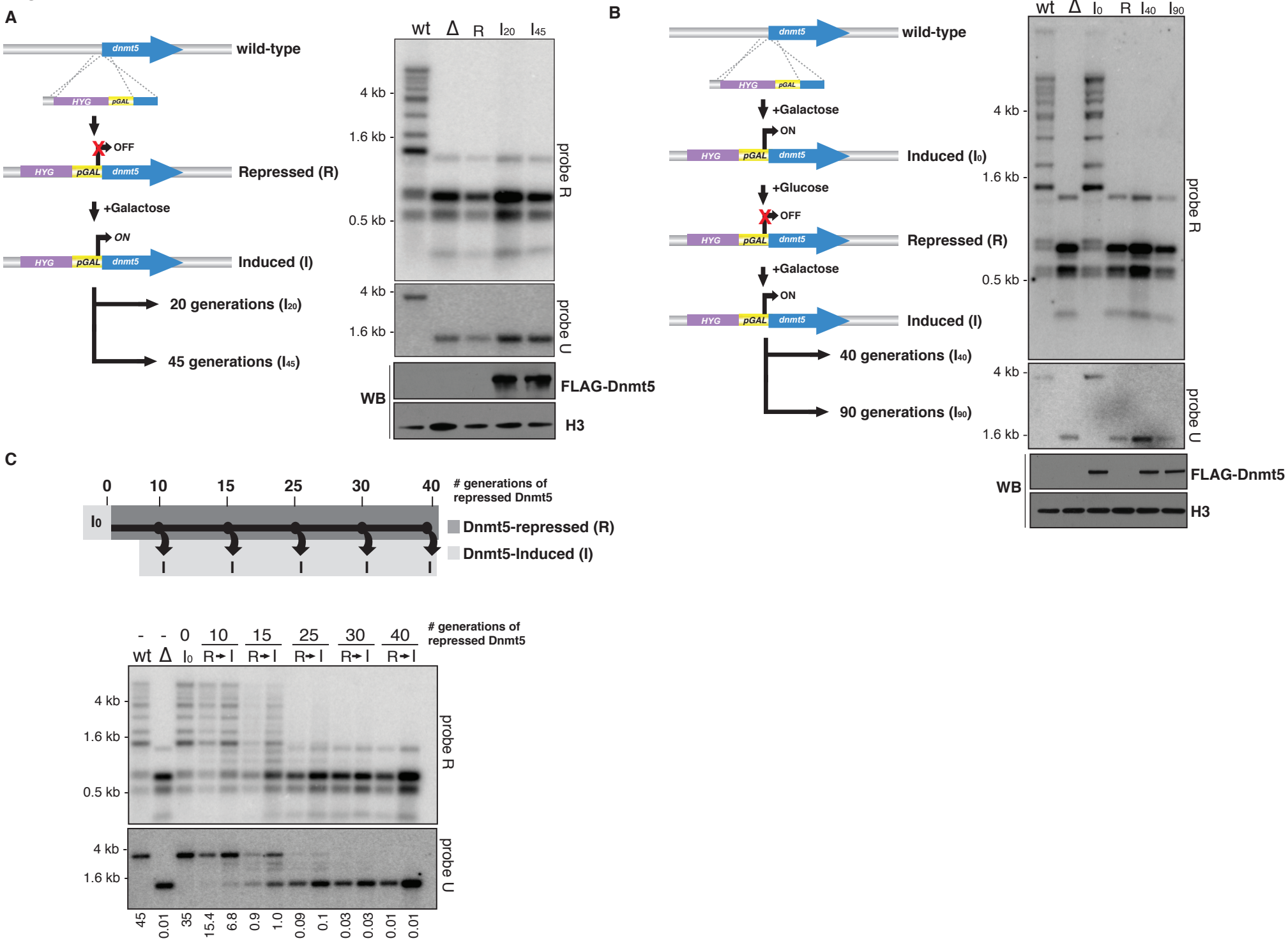
(**A**) Reaction curves showing incorporation of label from S-[methyl-^{3}H]-adenosyl-L-methionine into the indicated DNA oligonucleotide substrates using 30 nM Dnmt5 and 5 μM DNA substrate, in the presence or absence of 1 mM ATP.

(**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.

Figure 3. Catania et al.



bioRxiv preprint doi: https://doi.org/10.1101/149385; this version posted June 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Fig. 3. Transient repression of Dnmt5 results in sustained loss of 5mC

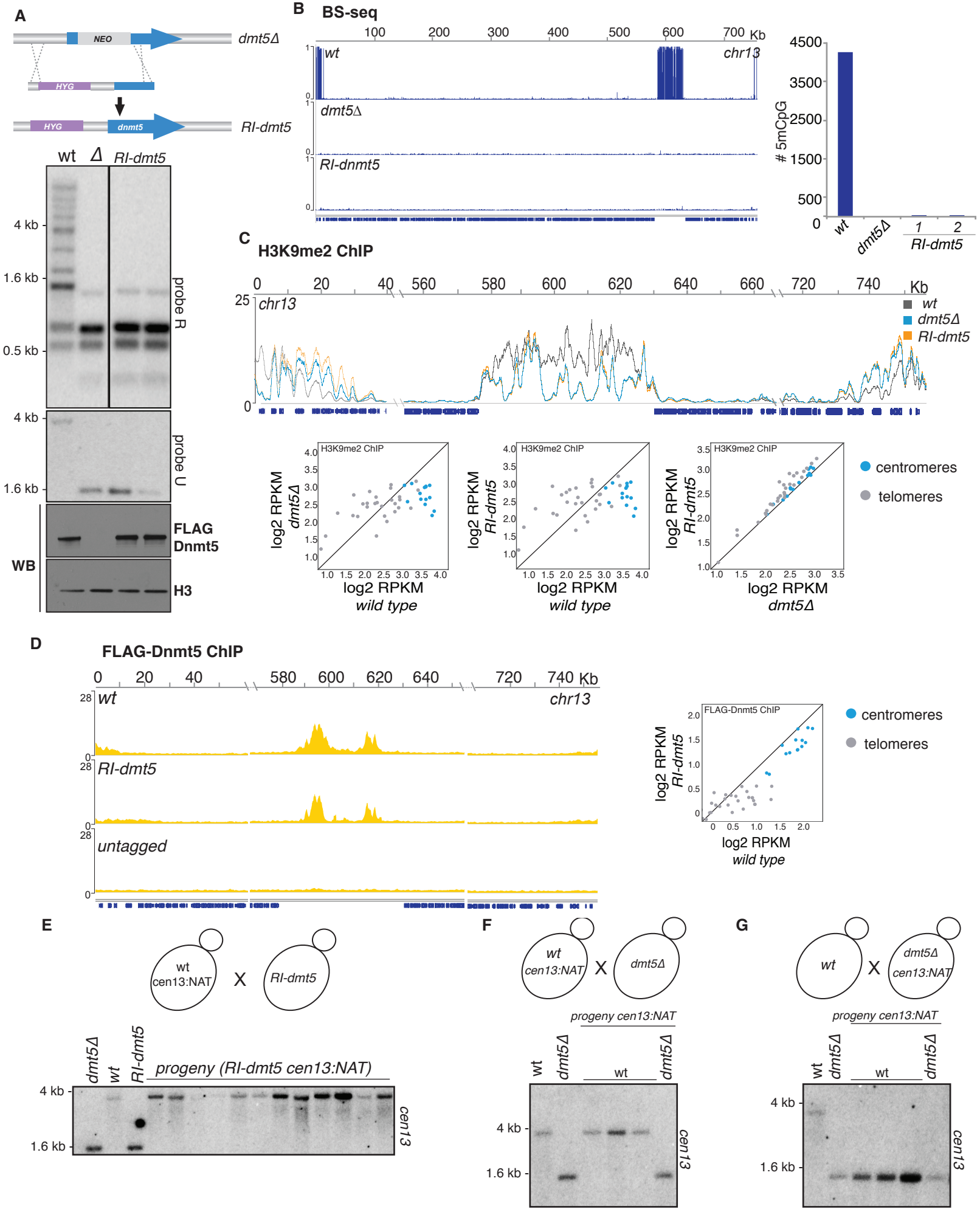
(**A**) OFF-ON scheme. Knock-in of a galactose-regulated *GAL7* promoter and epitope tag upstream of the Dnmt5 coding sequence was performed as shown. Strains were selected under repressing (glucose) conditions (R). Dnmt5 was induced by addition of galactose for 20 (I_{20}) or 45 generations (I_{40}). Strains were analysed 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 was performed as 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 analysed 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). Shown are the ratios of the intensities of 3.5 kb (fully methylated) band over the 1.5 kb (unmethylated) band, calculated for using ImageJ software.

bioRxiv preprint doi: https://doi.org/10.1101/149385; this version posted June 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 4. Catania et al.



bioRxiv preprint doi: https://doi.org/10.1101/149385; this version posted June 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Fig. 4. Genetic loss of DNA methylation is mitotically and meiotically irreversible

(**A**) Re-introduction (RI) scheme. A partial knockout of the *DMT5* locus was produced and then corrected by homologous recombination and N-terminally tagged as in Figure 2 to produce the RI-DNMT5 strain. A hygromycin resistance marker (*hygR*) is inserted upstream of the Dmt5 gene during this process. DNA methylation 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.

(**C**) Analysis of H3K9me2 signals in wild type, *dmt5*Δ and *RI-DMT5* strain. ChIP-seq signal for H3K9 methylation shown as normalised to signal for the input samples. 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 (grey) (scatter plot, bottom).

(**D**) Analysis of Dnmt5 recruitment in *the RI-DMT5* strain. ChIP-seq signal for FLAG-Dnmt5 were normalised 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.

(**E**) 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-dnmt5* allele. Progeny strains doubly resistant to hygromycin and nourseothrycin were selected. Strains were analysed for as described in Figure 1 using probe U, which is specific for chromosome 13.

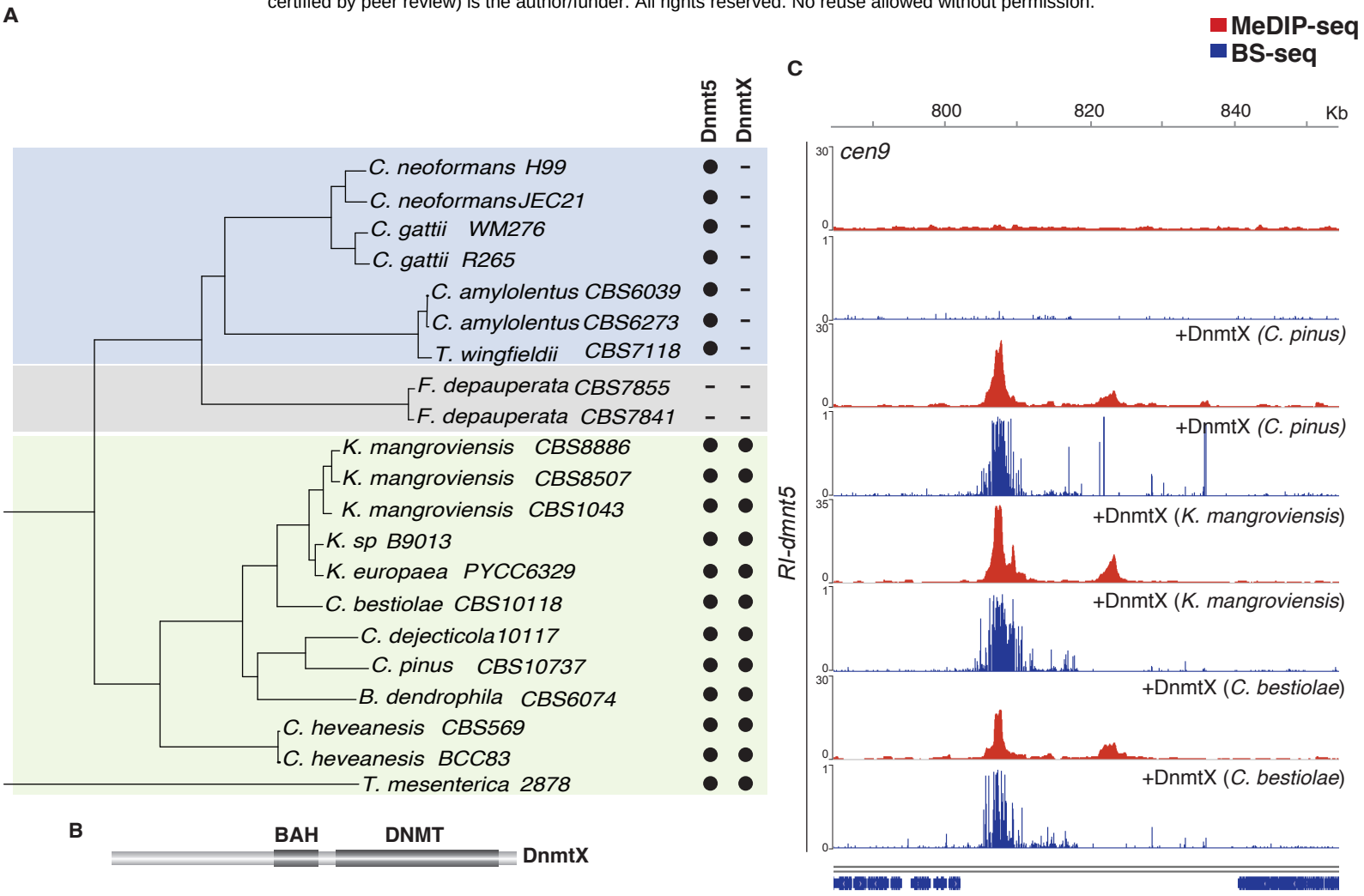
(**F**) Control cross. Dnmt5-harboring cells containing *CEN13::natR* were crossed to *dnmt5*Δ cells. Progeny harbouring the Dnmt5 and *CEN13::natR* were analysed as in (C).

bioRxiv preprint doi: https://doi.org/10.1101/149385; this version posted June 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

(**G**) Experimental cross testing whether sexual reproduction enables de novo methylation. *dmt5*Δ cells harboring *CEN13::natR* were crossed to wild-type cells. Progeny harbouring a wild-type Dnmt5 gene and *CEN13::natR* were analysed as in (C).

bioRxiv preprint doi: https://doi.org/10.1101/149385; this version posted June 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 5. Catania et al.



bioRxiv preprint doi: https://doi.org/10.1101/149385; this version posted June 13, 2017. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Fig. 5. **Ancient loss of gene coding for a *de novo* Dnmt in an ancestor of *C. neoformans*.**

(**A**) Phylogenetic analysis of whole-genome sequences. Shown are the phylogenetic relationships between *C. neoformans* and characterized *sensu stricto* and *sensu lato* groups. Phylogeny is based on the analysis of whole-genome sequences (C. Cuomo et al., manuscript in preparation). The genomic presence of gene coding for orthologs of Dnmt5 and of a second DNMT (DnmtX) are indicated by a filled circles. As indicated by the arrows, gene for DnmtX was lost prior to the divergence of *C. neoformans* and *F. depauperata*, and the gene for Dnmt5 was subsequently lost in the *F. depauperata* lineage.

(**B**) Predicted domain structure of DnmtX. DnmtX orthologs all have a canonical catalytic domain flanked by an N-terminal bromo-associated homology (BAH) domain.

(**C**) Expression of extant DnmtX-encoding genes in an *RI-DMT5 C. neoformans* strain triggers *de novo* 5mC. Three DnmtX from indicated organisms were expressed in a RI-DNMT5 strain. 5mC was assessed by MeDIP-seq (red) and BS-seq (blue). MeDIP signal is shown as ratio over *dmt5*Δ. Shown are data for centromere 9 which showed particularly efficient *de novo* methylation.