Neocentromeres lack epigenetic memory and hallmarks of native centromeres in *Cryptococcus deuterogattii*

Klaas Schotanus, Vikas Yadav, and Joseph Heitman*

Department of Molecular Genetics and Microbiology, Duke University Medical Center, 322 CARL Building, Box 3546, Research Drive, Durham, NC 27710 USA

*Corresponding author

Email: heitm001@duke.edu, Phone: +1-919-684-2824, Fax: +1-919-684-5458

Abstract

Deletion of native centromeres in the human fungal pathogen *Cryptococcus deuterogattii* leads to neocentromere formation. Native centromeres span truncated transposable elements, while neocentromeres span actively expressed genes. Neocentromeres in *cen10*Δ mutants are unstable and chromosome-chromosome fusions occur. After chromosome fusion, the neocentromere is silenced and the native centromere of the chromosome fusion partner remains as the sole active centromere. In the present study, the active centromere of a fused chromosome was deleted to investigate if epigenetic memory promoted re-activation of a silenced neocentromere. Our results show that the silenced neocentromere is not re-activated and instead a novel neocentromere forms directly adjacent to the deleted centromere of the fused chromosome. To explore the epigenetic organization of neocentromeres, we characterized the distribution of the heterochromatic histone modification H3K9me2 and 5mC DNA methylation. Native centromeres were enriched for both H3K9me2 and 5mC DNA methylation marks, while neocentromeres lacked these specific histone and DNA modifications. To study centromere dynamics, the actively expressed *URA5* gene was introduced into a native centromere. Introduction of the *URA5* gene led to loss of CENP-A from the native centromere, and a neocentromere formed directly adjacent to the native centromere location. Remarkably, the silenced native centromere remained enriched for heterochromatin, yet the integrated gene was expressed and devoid of H3K9me2. Analysis of multiple CENP-A distribution profiles revealed centromere drift in *C. deuterogattii*, a previously unknown phenomenon in fungi. The CENP-A-enriched region shifted within the pericentric regions, and a truncated transposable element in centromere 5 acted as a barrier between the CENP-A-associated regions of chromatin. Interestingly, this truncated transposable element was devoid of CENP-A binding or H3K9me2 modification and was instead marked by 5mC DNA methylation. Taken together, our findings reveal novel aspects about the epigenetic mechanisms that distinguish native centromeres and neocentromeres.

Introduction

To undergo proper cell division, chromosomes require a functional centromere that mediates binding of the kinetochore and microtubules to the chromosome [1]. In most organisms, the kinetochore assembles on a specific centromeric histone H3 (known as CENP-A, CENH3, or Cse4 depending on the species), which replaces the canonical histone H3 within the centromere [2]. Species lacking CENP-A are either holocentric or differ in their kinetochore structural organization [3–5]. The majority of studied centromeres are regional centromeres that are epigenetically regulated; there are however some exceptions, such as point centromeres that are sequence dependent [6,7].

The deletion of centromeres frequently results in neocentromere formation in genomic hotspots, and these are often in the vicinity of the native centromere [8]. This has been observed in model organisms such as *Candida albicans* and also in chicken DT40 cells, and might be attributable to the CENP-A cloud [9–13]. Unlike native centromeres in *C. albicans,* neocentromeres are not enriched for flanking repeats and genes spanned by CENP-A are silenced, although neocentromeres respond similarly to stress as native centromeres [12]. All native pericentric regions of *C. albicans*, except for chromosome 7, harbor either inverted repeat or long-terminal repeat sequences, which might function as pericentric domains in some chromosomes [14]. The pericentric regions are enriched for the euchromatic histone marks H3K9Ac and H4K16Ac, but are hypomethylated at H3K4, a feature of heterochromatic chromatin [15]. Histone H3 is completely depleted at the native centromeres of *C. albicans* and H3K9 methylation is absent in the *C. albicans* genome, as this pathway was lost during evolution [16,17]. The mixture of euchromatin and heterochromatin results in reduced expression of genes located in close proximity to the centromeres.

The majority (76%) of neocentromeres in chicken DT40 cells form in close proximity to the native centromere [10]. Chickens have a diploid genome, but the Z sex chromosome is present in only one copy in DT40 cells (ZW/female cells). Both the native centromere and neocentromeres of chromosome Z lack repeats, do not contain a specific DNA motif, and are similar in length (~35-45 kb). Based on chromatin immunoprecipitation sequencing (ChIP-seq) and immunofluorescence microscopy, native centromeres of chromosomes (Chr) 5, 27, and Z, as

well as all neocentromeres, are not enriched for H3K9me3. This is in contrast to the native centromeres of Chr 1 and 2, which span repeats and are enriched for H3K9me3. Despite the lack of H3K9me2, the expression level of a gene (*MAMDC2*) within a neocentromere is reduced by 20- to 100-fold relative to wild-type expression levels [10]. These findings suggest heterochromatin is not essential for centromere function in DT40 cells and that heterochromatin may serve only as a repressor of repetitive regions.

Native centromeres of *Schizosaccharomyces pombe* have a higher order structure with a CENP-A-enriched central core and heterochromatic inner and outer repeats that serve as pericentric regions [18]. The pericentric regions are enriched for H3K9me2/3, and this modification is RNAi-dependent and essential for CENP-A localization to the central core [19,20]. Neocentromeres in *S. pombe* form adjacent to heterochromatic regions, close to the telomeres, where pre-existing repeats mimic the higher-order repeat architecture of native centromeres [21]. Similar to *C. albicans*, *S. pombe* neocentromeres span genes that are silenced due to CENP-A binding [21]. Interestingly, deletion of native centromeres in wild-type *S. pombe* leads to either neocentromere formation or chromosome fusion, while centromere deletion in H3K9me2-deficient cells results in only chromosome fusion, suggesting that heterochromatic regions are essential for neocentromere function [21,22].

Similar to many fungal centromeres, the native centromeres of *Cryptococcus neoformans* are enriched for both full-length and truncated transposable elements [23,24]. Previous studies have shown that these elements are silenced via RNAi [25–27]. *Cryptococcus* centromeres are also enriched for H3K9me2 and 5mC DNA methylation, which is interesting because a recent study showed that the pathogenic *Cryptococcus* species complex has lost the *de novo* DNA methyltransferase and retained only the maintenance DNA methyltransferase (Dnmt5) [28,29]. In contrast to heterochromatin formation in *S. pombe*, heterochromatin formation in *C. neoformans* is not mediated by small RNAs [25].

Cryptococcus deuterogattii is the only member of the pathogenic species complex that is RNAi-deficient, and loss of RNAi has been correlated with the presence of only truncated transposable elements in centromeres [24,30,31]. Previously we showed that deletion of a native centromere in *C. deuterogattii* leads to neocentromere formation [32]. Neocentromeres of both

Chrs 9 and 10 spanned genes, which continued to be expressed at wild-type levels. This wild-type expression of neocentromeric genes suggests neocentromeres may not be heterochromatic in *C. deuterogattii*. However, neocentromeres of Chr 10 were unstable and centromere deletion often led to chromosome-chromosome fusions through non-homologous recombination between sub-telomeric regions. After chromosome fusion, the neocentromeres were silenced and only the native centromere of the fusion partner chromosome served as the centromere of the fused chromosome [32].

Here we show that deletion of the native centromere in a fused chromosome results in neocentromere formation in close proximity to the native centromere, not at the previously formed neocentromere location. This observation suggests that the CENP-A cloud determines the location for neocentromere formation, rather than epigenetic memory of the position of the silenced neocentromere. We find that, in contrast to native centromeres, neocentromeres are not enriched for the heterochromatin mark H3K9me2 or 5mC DNA methylation. Additionally, native centromeres that have lost CENP-A binding as a consequence of integration of a selectable marker gene remain enriched for H3K9me2 and neocentromeres form adjacent to the native centromere. Overall, our results demonstrate that neocentromeres in *C. deuterogattii* lack epigenetic marks that are present at native centromeres and do not retain any epigenetic memory when silenced.

Results

Silenced neocentromeres lack epigenetic memory

In a previous study we observed that *cen10*Δ mutant isolates with a neocentromere maintaining Chr 10 exhibited aneuploidy for Chr 10 and produced both large and small colonies at 37°C. Isolation and analysis of larger colonies at 37°C revealed that chromosome-chromosome fusions had occurred involving Chr 10 fusion with other chromosomes [32]. After chromosome-chromosome fusion, the neocentromere was silenced and the native centromere of the fusion partner chromosome acted as the active centromere. As the neocentromere was silenced in the fused chromosome, centromeric proteins were displaced from the neocentromere locus, and we experimentally confirmed this by showing that the neocentromere lacked CENP-A. However, we hypothesized that the silenced neocentromere might still bear epigenetic marks, and if so these marks might render this a preferred site for subsequent neocentromere formation.

To investigate whether such epigenetic memory existed at the silenced neocentromeric location, we deleted the active centromere of the fused chromosome and performed ChIP-seq analysis for CENP-A. For this experiment, we focused on an isolate (KS123) in which Chr 10 was fused with Chr 4. Using CRISPR-Cas9, we deleted native *CEN4* and obtained two independent *cen10*Δ *cen4*Δ mutants. In addition, we also deleted *CEN4* in the wild type (WT) as a control, and two independent *cen4*Δ mutants were isolated. All four mutants lacking *CEN4* were subjected to CENP-A ChIP-seq analysis and reads were mapped to the genome assembly. Interestingly, in all four mutants, the neocentromeres formed close to the location of native *CEN4* (Figure 1). In three cases, neocentromeres formed in the same chromosomal location, directly flanking the native centromere on the 3' side. These neocentromeres were smaller in size (~5 kb) than the native *CEN4* (~10 kb) and spanned three genes (CNBG_0196, CNBG_0195, and CNBG_0194). CNBG_0196 encodes a serine/threonine-protein kinase, CNBG_0195 encodes a purine-specific oxidized base lesion DNA N-glycosylase, and CNBG_0194 encodes a protein predicted to localize to the endoplasmic reticulum. In the fourth case, the neocentromere formed next to the native centromere on the 5' side. This neocentromere was much smaller in size (~1.3 kb) than the native *CEN4* as well as the other three neocentromeres and spanned only the first two exons of the CNBG_0197 gene, which encodes topoisomerase II alpha-4. Unlike *cen10*Δ mutants, *cen4*Δ

mutants in the otherwise wild-type background had uniform colony sizes, as was observed with *cen9*Δ mutants [32].

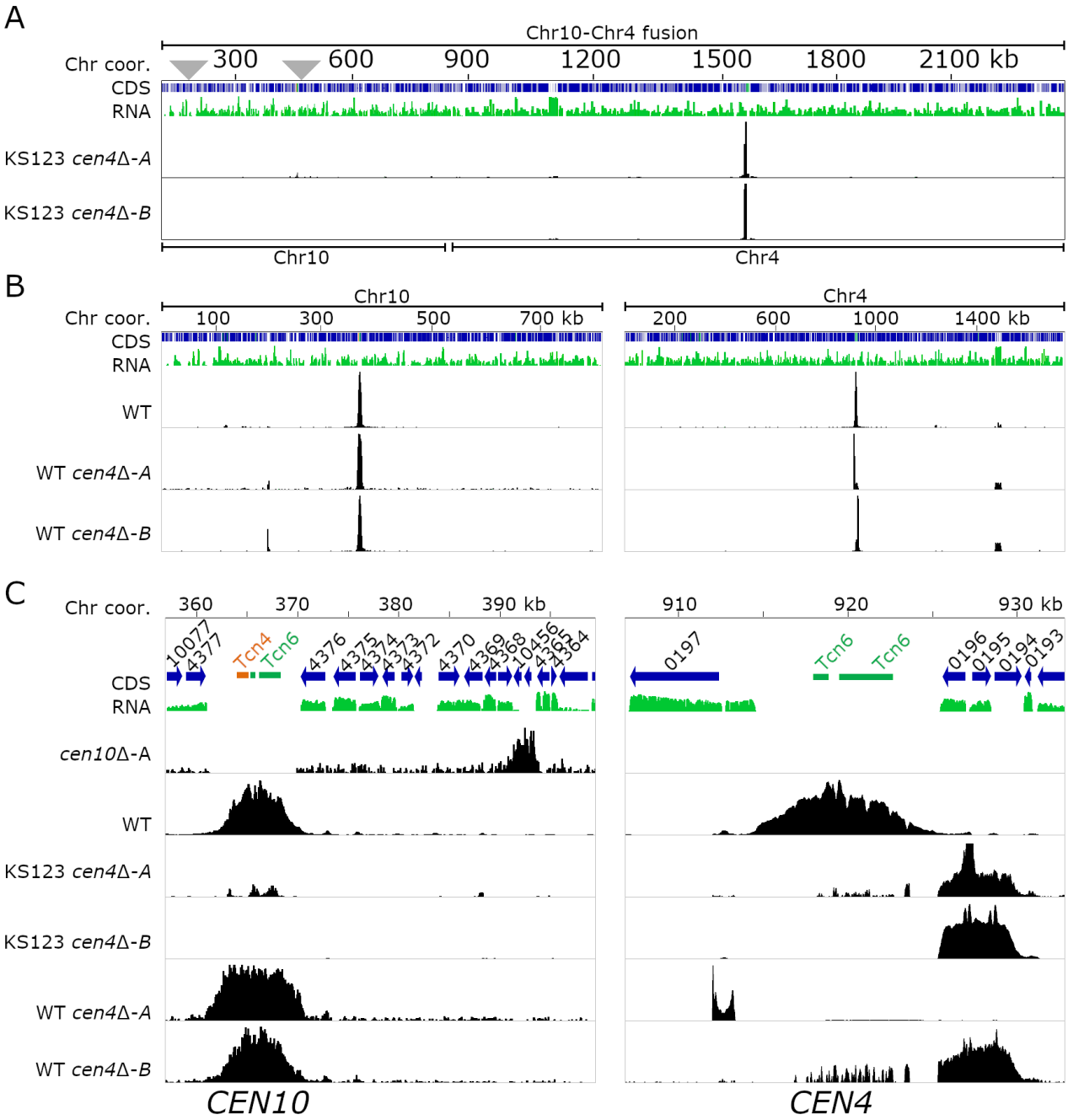

Figure 1. Neocentromeres in *C. deuterogattii* lack epigenetic memory.

(A) A full-length chromosomal view of two *cen4*Δ mutants in a Chr 4-10 fusion derived from a *cen10*Δ mutant background (KS123) is shown. The initial *cen10*Δ mutant had two CENP-A-enriched neocentromeric regions, and these silenced neocentromeres are indicated with gray triangles. **(B)** Full-length chromosome view for Chrs 10 and 4. The *cen4*Δ mutation in the wild type did not affect *CEN10*, and as shown in the second panel, neocentromeres of chromosome 4 formed proximal to the original chromosomal location of native *CEN*4. The additional low-abundance CENP-A peak at 1490 kb colocalized with the rDNA cluster. **(C)** Detailed view of native *CEN*10 and *CEN*4 regions. *cen10*Δ-*A* is the initial neocentromere mutant and the

neocentromere position is indicated by the CENP-A enrichment in *cen10*Δ-*A*. As shown with two independent *cen4*Δ deletions isolated in the strain harboring the fused chromosome, the neocentromere was silenced after chromosome fusion. The panel with *CEN*4 shows that neocentromeres formed proximal to the native *CEN*4, and this was independent of the presence of a silenced neocentromere.

Taken together, these results provide evidence that silenced neocentromeres in *C. deuterogattii* do not possess an epigenetic memory and suggest that *de novo* neocentromere formation may instead occur due to the presence of a CENP-A cloud.

To study the impact of chromosome fusion on centromeres in further detail, we generated a fusion of Chrs 4 and 10 to model the previously isolated Chr 4-10 fusion, but without the *CEN10* deletion to result in a presumptive dicentric Chr 4-10 fusion chromosome. Pulsed-field gel electrophoresis (PFGE) confirmed two isolates in which the two chromosomes had fused, as indicated by the absence of bands corresponding to Chrs 4 and 10 (Figure S2). However, from the EtBr stained banding pattern, it appears that the dicentric chromosome in one of the transformants (dicentric isolate 1) underwent chromosomal-fusion, followed by breakage and a new chromosomal band appeared at ~890 kb. The second transformant (dicentric isolate 2) did not show any new band suggesting that either the dicentric chromosome might have been stable or the PFGE was not able to resolve the new chromosomes. To confirm the chromosome fusion we performed Nanopore sequencing for both putative dicentric isolates. For dicentric isolate 2, Nanopore sequencing confirmed the presence of a stable chromosome fusion. For dicentric isolate 1, a scaffold (443 kb) with chromosome 10 fused to a 51 kb region of chromosome 4 was identified. Interestingly, we noticed that the same 51 kb region was duplicated in the genome and still present on a scaffold (433 kb) which corresponds to chromosome 4 (Figure S3). Short read sequencing confirmed that the 51 kb region was duplicated as the ploidy level of this regions is two instead of one. Thus in this isolate, the dicentric chromosome broke, and two new chromosomal ends were produced resulting in a 51 kb segmental duplication shared by the new Chr 4 and 10.

To test if either centromere 4 or 10 was silenced after chromosome fusion in the dicentric isolate 2 in which no new chromosomes were apparent by PFGE analysis, ChIP-qPCR analyses targeting *CEN4* and *CEN10* of the putative dicentric chromosome were performed (Figure S2).

Compared to the positive control (*CEN6*), the CENP-A enrichment for *CEN4* was 2-fold higher, while the CENP-A enrichment for *CEN10* was 0.5-fold lower. This could suggest that some cells in the population have an active *CEN4* whereas a smaller subset of cells have *CEN10* as the active centromere.

Neocentromeres are not enriched for the heterochromatic mark H3K9me2

Because the RNAi pathway is absent in *C. deuterogattii* and the genome contains only truncated transposable elements, it is possible that other pathway(s) are involved in silencing/suppression of transposable elements; for example, heterochromatin or DNA methylation, might have evolved differently in this species. To test if heterochromatin is still functional in *C. deuterogattii*, we performed ChIP followed by high-throughput sequencing (ChIP-seq) with an antibody specific to H3K9me2.

Despite the lack of active transposable elements, the native centromeres were still enriched for H3K9me2 that spanned the entire ORF-free region (Figure 2A and S5). Further analysis revealed that the CENP-A-bound region was embedded within the H3K9me2-enriched region and displayed a similar pattern as was previously shown in the closely-related species *C. neoformans* (Figure S5) [28]. Telomeres and sub-telomeric regions were also enriched for H3K9me2. In addition to the expected regions, we identified several short chromosomal regions that were enriched for H3K9me2; however, compared to the centromeres, the H3K9me2 enrichment was at least five-fold lower in these regions (Table S3). Some of these H3K9me2-enriched regions spanned genes whose products are predicted to be involved in sugar metabolism or drug transport.

The presence of a heterochromatic mark on a gene can lead to epimutation-mediated drug resistance. To test whether H3K9me2 is involved in such a phenomenon, we deleted the gene encoding the Clr4 methyltransferase and tested the strains for resistance to multiple drugs. Fluctuation assays with two antifungal drugs (5-Flourocytosine and Rapamycin+FK506) showed that, compared to the wild-type, *clr4*Δ mutants generated fewer or no resistant colonies (Figure S4). These results suggest Clr4 may be involved in regulating drug resistance in this species and this will be the subject of future studies.

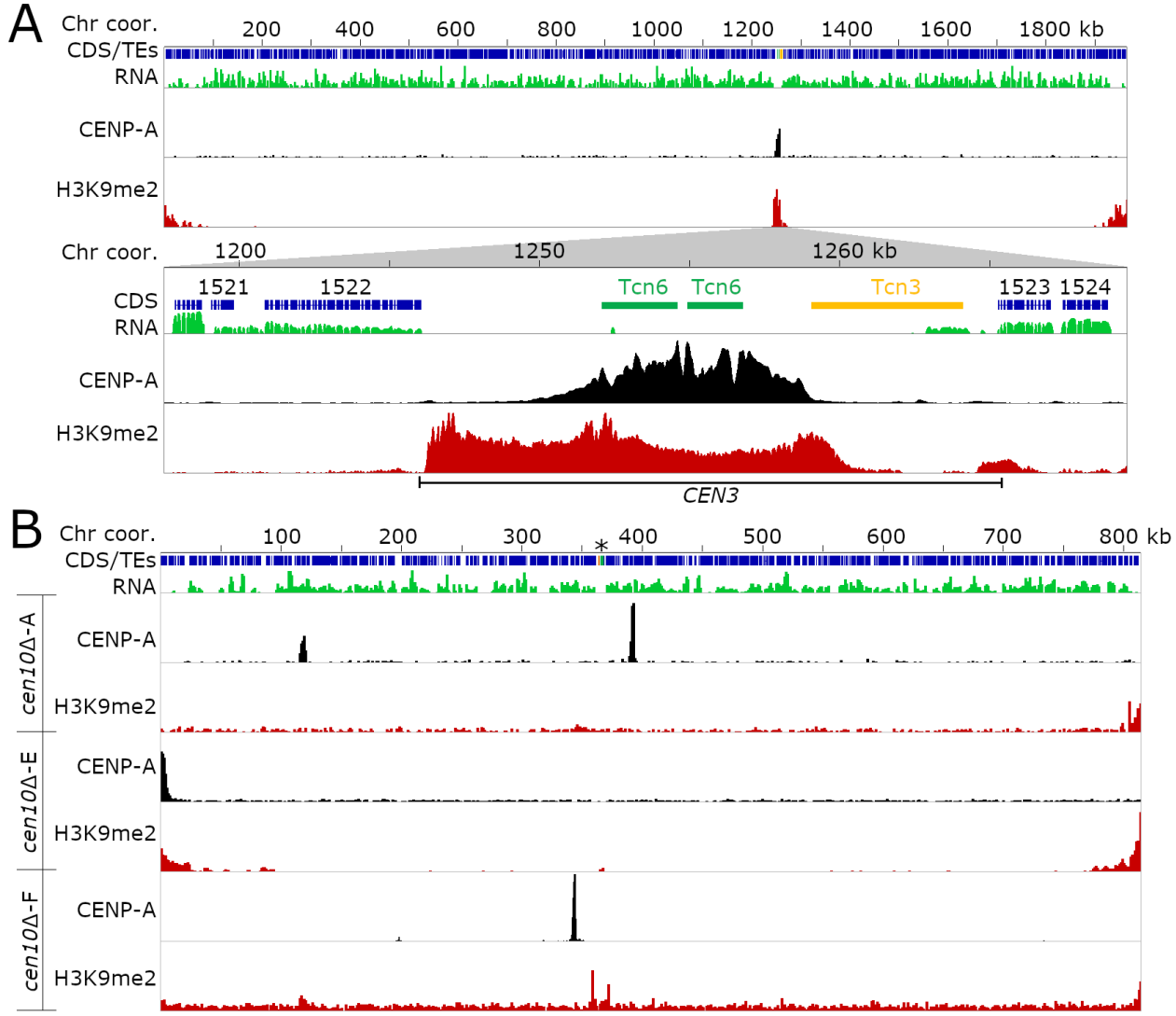

Figure 2. *C. deuterogattii* native centromeres, but not neocentromeres, are enriched for H3K9me2.

For each panel, the chromosome coordinates are indicated. Genes are shown in blue, TEs are presented as per the specific TE family color, and RNA-seq is shown in green. CENP-A ChIP-seq is shown in black and ChIP-seq for H3K9me2 is shown in red. **(A)** H3K9me2 distribution of Chr 3 is shown as an example. The top panel shows the full-length Chr 3 and the lower panel shows an expanded view of the centromeric region. **(B)** CENP-A and H3K9me2 ChIP-seq with previously generated *cen10*Δ mutants. For three *cen10*Δ mutants the CENP-A enrichment and the H3K9me2 profile is shown. Mutant *cen10*Δ-*F* shows a slight H3K9me2 enrichment at the native centromere location (indicated with an asterisk) due to IGV viewer limitations, and detailed analysis shows that this is an artifact.

Previously, we deleted native *CEN10* and obtained multiple isolates with neocentromeres [32]. To test if neocentromeres are enriched for H3K9me2, ChIP-seq with an H3K9me2 antibody was performed for three *cen10*Δ strains containing neocentromeres (Figure 2B). Isolate *cen10*Δ-*A* has two CENP-A-enriched regions of which one is the primary peak and the second is a less

abundantly enriched peak [32]. Isolate *cen10*Δ-*E* has a neocentromere directly flanking the left telomere and isolate *cen10*Δ-*F* has a neocentromere proximal to the deleted native centromere [32]. After mapping the H3K9me2 ChIP-seq reads to the genome, we observed that the neocentromeres of *cen10*Δ-*A* and *cen10*Δ-*F* lacked H3K9me2 enrichment. The sub-telomeric regions of all 14 chromosomes and the native centromeres of the other 13 chromosomes were still enriched for H3K9me2 as expected. As the neocentromere of *cen10*Δ-*E* is telocentric, it was enriched with H3K9me2. However, the sub-telomeric regions were modestly enriched for H3K9me2 in the wild type; thus, we think the enrichment in this isolate was not due to neocentromere formation but rather to its location at a subtelomeric region.

Integration of a *URA5* transgene into *CEN2* inactivated the native centromere and resulted in neocentromere formation

C. deuterogattii neocentromeres are present in gene-rich regions and span genes that are actively expressed despite CENP-A binding. This prompted us to test the impact of insertion of an actively transcribed gene into native centromeres on centromeric chromatin. We introduced the *URA5* gene into the CENP-A-binding region of *CEN*2 by homologous recombination. We hypothesized four possible outcomes from this experiment: 1) CENP-A would cover the *URA5* gene, making the CENP-A-bound region larger than the native centromere; 2) the *URA5* gene might divide the CENP-A enrichment into two independent regions; 3) only one of the regions flanking *URA5* would be enriched for CENP-A, generating a smaller centromere; or 4) *URA5* integration might abolish *CEN*2 function leading to neocentromere formation (Figure S6).

Through homologous recombination, the *URA5* gene was introduced into the CENP-A-enriched region of *CEN*2 in an R265 *ura5*- strain. Following transformation, isolates were selected for growth on SD-ura medium. Isolates were further screened by inoculating the mutants on SD-ura and 5-FOA media. PCR analysis confirmed the desired integration of the *URA5* gene into *CEN*2. Prior to ChIP-seq, we verified that the *URA5* gene introduced into *CEN*2 was expressed based on growth on SD-URA medium (Figure S6). Southern blot analysis was performed to confirm a single insertion event of the *URA5* gene in the genome at the desired location in *CEN*2, and that integration had occurred in a single copy and was not a tandem integration. A probe

corresponding to the *URA5* gene hybridized to only two locations in the genome corresponding to the *ura5*- gene at its native location and the *URA5* gene integrated into *CEN*2 (Figure S7).

Subsequently, we performed CENP-A ChIP-seq to investigate which of the four hypotheses might be correct. After mapping the ChIP-seq reads to the wild-type reference genome, we found that the CENP-A distribution shifted dramatically and co-localized with a region slightly enriched for H3K9me2, flanking one side of the native *CEN*2 (Figure 3). Despite the formation of a *de novo* CENP-A peak flanking *CEN*2, there was still a small amount of residual CENP-A binding at the native location. The H3K9me2-enriched region spanned ~6 kb, and included three genes with low expression levels: CNBG_4498 which encodes a pyridoxal kinase, CNBG_10170 which encodes the DNA mismatch repair protein Msh4, and CNBG_4497 which encodes a hypothetical protein.

The shift in CENP-A binding after *URA5* insertion into *CEN*2 provided a unique opportunity to test if H3K9me2 enrichment remained at the native location. To characterize H3K9me2 enrichment in this strain (KS174), we performed ChIP-qPCR and quantified H3K9me2 enrichment at four loci within *CEN*2 (Figure 3C). The first qPCR primer pair was located outside of the *URA5* allele used for homologous recombination. The second and fourth primer pairs were located inside the upstream flank (UF) and downstream flank (DF) of the *URA5* homologous recombination allele, respectively. Primer pair three was specific for the *URA5* gene, and this region lacked enrichment for H3K9me2, consistent with the observation of continued expression of the *URA5* gene despite its integration within the original *CEN*2. As a control, a primer pair located inside *CEN*6 was also employed. All data were normalized to the *ACT1* gene, which lacks H3K9me2. All ChIP-qPCRs showed significant enrichment for H3K9me2, indicating that regions within *CEN*2 were still enriched for heterochromatin, despite the loss of CENP-A binding. Overall, these results suggest that H3K9me2 localization can be maintained independently of CENP-A enrichment or the presence of active transposable elements.

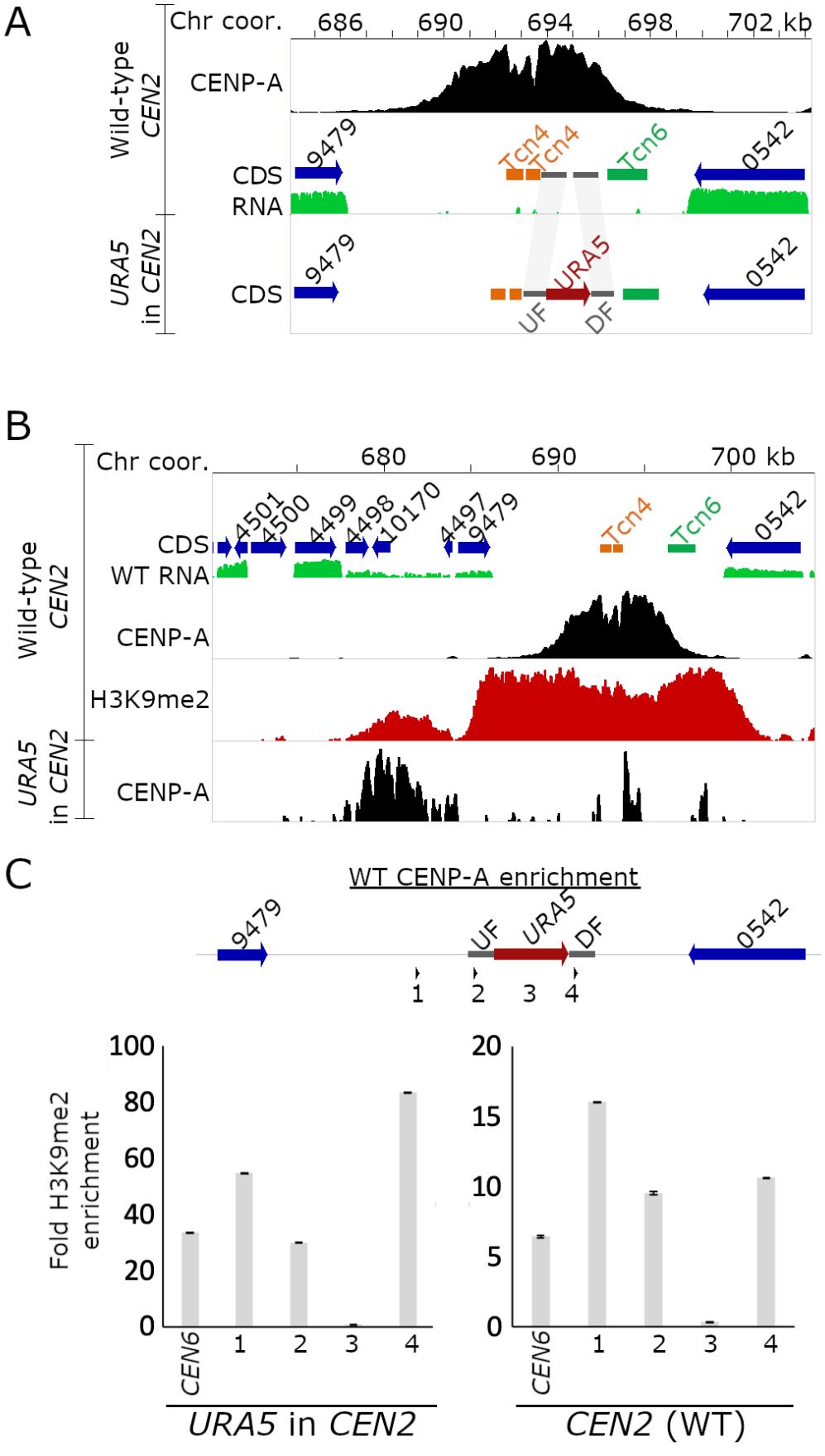

Figure 3. Insertion of a transgene into a native centromere leads to a shift in the CENP-A-binding region.

(A) Schematic representation of the introduction of the *URA5* gene into *CEN*2 by homologous recombination. The top panel (black) shows the CENP-A-enriched region of the native *CEN*2. The middle panel shows the wild-type genome annotation. The sequences used as flanking regions for homologous recombination of the *URA5* integration construct are depicted in dark gray. The third panel shows a hypothetical map after the integration of the *URA5* gene into native *CEN*2. **(B)** ChIP-seq of CENP-A enrichment in the wild-type and *URA5* insertion strain. Genes are shown with blue arrows and RNA-seq is shown in green. The first and second panel show CENP-A and H3K9me2 enrichment, respectively, for wild-type native *CEN*2. The third panel shows the CENP-A enrichment for the *URA5* integration in *CEN*2 (*URA5* in *CEN*2). The CENP-A in '*URA5* in *CEN*2' co-localizes with a secondary H3K9me2 peak directly flanking the native *CEN*2. **(C)** Enrichment for H3K9me2 in the silenced native *CEN*2 of '*URA5* in *CEN*2' was tested by ChIP-qPCR. Four regions were tested for H3K9me2 enrichment at the modified *CEN*, these are numbered 1 to 4, and the chromosomal location is shown at the top. Genes flanking the native *CEN*2 are shown with blue arrows and the region enriched for CENP-A in the wild type is indicated with a black line. ChIP-qPCRs were normalized to actin and show the fold enrichment of H3K9me2 in the four chromosomal regions tested by qPCR for *URA5* in the *CEN*2 strain and the wild type. As a positive control, *CEN*6 was included in the qPCR analysis. Error bars show SD.

Correlation between 5mC DNA methylation and H3K9me2 across centromeres

Previously we reported that the DNA methyltransferase gene *DMT5* was truncated in the *C. deuterogattii* R265 reference genome, and 5mC methylation was thought to be entirely absent based on PCR-based assays with methylation-sensitive restriction enzymes [24]. An updated gene annotation of the R265 reference genome and re-mapping of RNA-seq data revealed that the *DMT5* gene was previously misannotated and encodes a putative fully functional gene [33] Figure S8). To corroborate this finding, bisulfite sequencing was performed with DNA isolated from the wild-type strain and a *dmt5*Δ deletion mutant. Bisulfite sequencing data analysis showed that centromeres are enriched for 5mC in *C. deuterogattii*; however, methylation was significantly reduced compared to the *C. neoformans* wild-type reference strain H99 (Figure 4 and S5). As expected, 5mC methylation was abolished in the R265 *dmt5*Δ mutant strain. The bisulfite sequencing analysis confirmed that 5mC was lacking in the R265 genomic regions that were previously analyzed by PCR following methylation-sensitive restriction enzyme digestion [24].

Figure 4

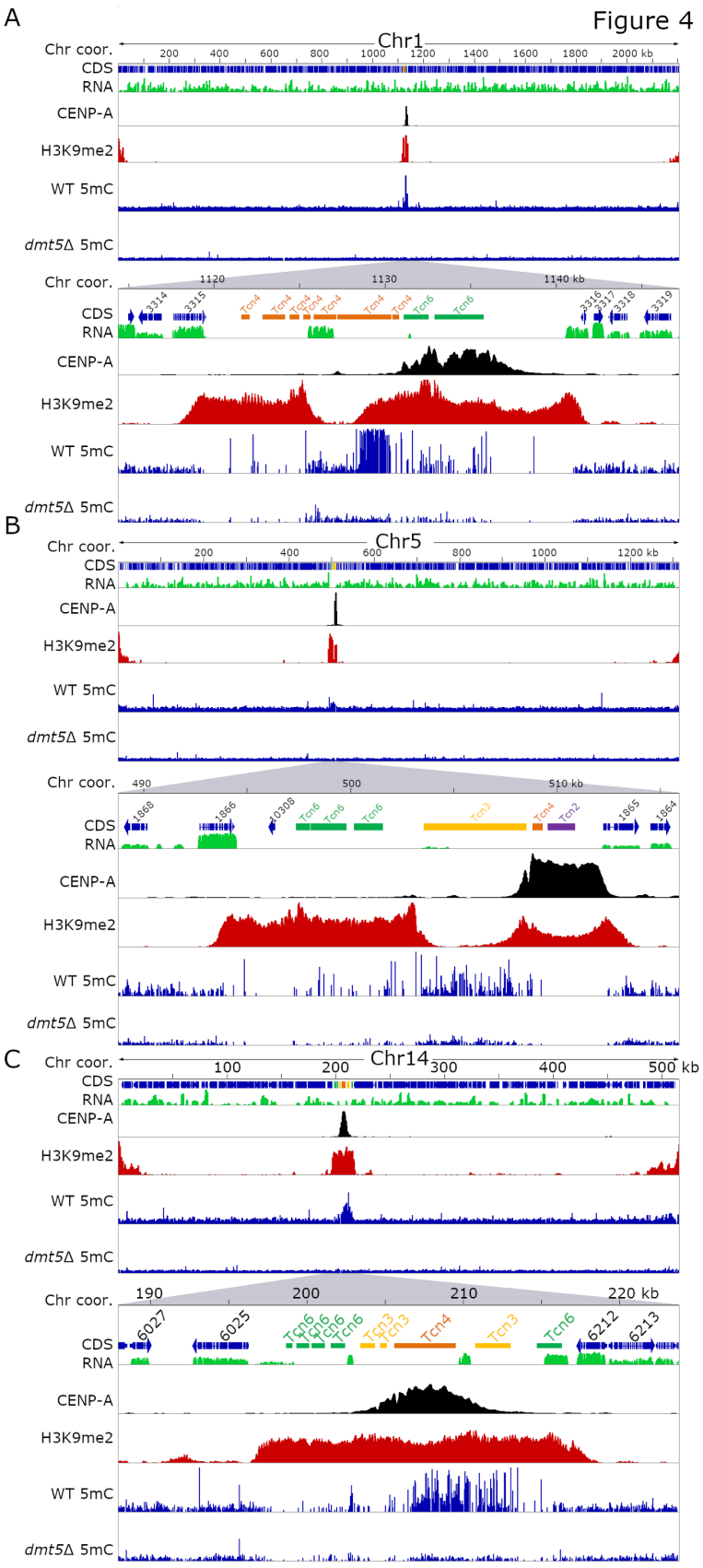

Figure 4. Heterochromatin marks H3K9me2 and 5mC DNA methylation do not always overlap with each other in *C. deuterogattii* centromeres.

For each panel, the full-length chromosome, is on the top and the lower section shows a region spanning the native centromere with one or two flanking genes. Each panel shows the chromosome coordinates, with arrows indicating the cds/TEs, RNA-seq shown in green, CENP-A reads indicating the native centromere shown in black, H3K9me2 enrichment is in red, and 5mC enrichment is shown in blue (for both the wild type and *dmt5*Δ). Panel **(A)** shows Chr 1, panel **(B)** shows Chr 5, and panel **(C)** shows Chr 14.

5mC DNA methylation in *C. neoformans* is co-localized with H3K9me2 and was also found to be dependent on the presence of H3K9me2 [32]. However, our results showed that these two modifications were not well-correlated with each other in *C. deuterogattii*. Specifically, the two longest Tcn3 elements (~5 kb), located in *CEN*3 and *CEN*5, which lacked CENP-A and H3K9me2 enrichment, were among the most abundant 5mC-enriched regions in the *C. deuterogattii* genome. In total, there are six Tcn3 elements, ranging from 0.4 to ~5 kb in length. The four shorter Tcn3 elements (~0.4 to 2.2 kb) were enriched for H3K9me2 and CENP-A. Re-mapping of previously generated RNA-seq to the *C. deuterogattii* reference genome showed that the two longest Tcn3 elements are weakly expressed (Table S4) [34]. Transposable elements Tcn4 and Tcn6 were also enriched with 5mC but showed binding of H3K9me2. Furthermore, *CEN*4, *CEN*6, and *CEN*7 completely lacked any 5mC enrichment but were significantly enriched with H3K9me2. Based on these results, we conclude that even though 5mC methylation is present in *C. deuterogattii*, the overall level is significantly reduced compared to *C. neoformans*. Furthermore, the 5mC localization pattern in *C. deuterogattii* differs significantly from the closely related species *C. neoformans*.

A 5mC-enriched Tcn3 element generates centromere drift

All *C. deuterogattii* centromeres have a pericentric region enriched with H3K9me2, which harbors the CENP-A enriched region. Comparison of the CENP-A profiles of *cen4*Δ, *cen9*Δ, *cen10*Δ, and wild-type strains revealed that the CENP-A-bound regions of most centromeres are co-linear and coincident within the strains sequenced (Figure 5) [24]. However, for *CEN*5 and *CEN*14, CENP-A binding was observed to drift within the ORF-free, heterochromatin-enriched centromeric regions. For *CEN*5, centromeric drift was observed to be strongly associated with the presence of

one of the two longest Tcn3 elements in the middle of the centromere. None of the strains analyzed in ChIP-Seq experiments exhibited CENP-A binding to this Tcn3 element, and instead the binding was localized to either the 5' or 3' end of the Tcn3 element. Interestingly this Tcn3 element was also not enriched for H3K9me2, but was enriched for 5mC DNA methylation. The centromeric drift in *CEN*14 was less distinct and the CENP-A-enriched region shifted within the pericentric region; therefore, no definitive barrier could be identified. In addition to the CENP-A profiles, we re-mapped the earlier obtained CENP-C ChIP-seq data and this analysis further supported the presence of centromeric drift in the pericentric region.

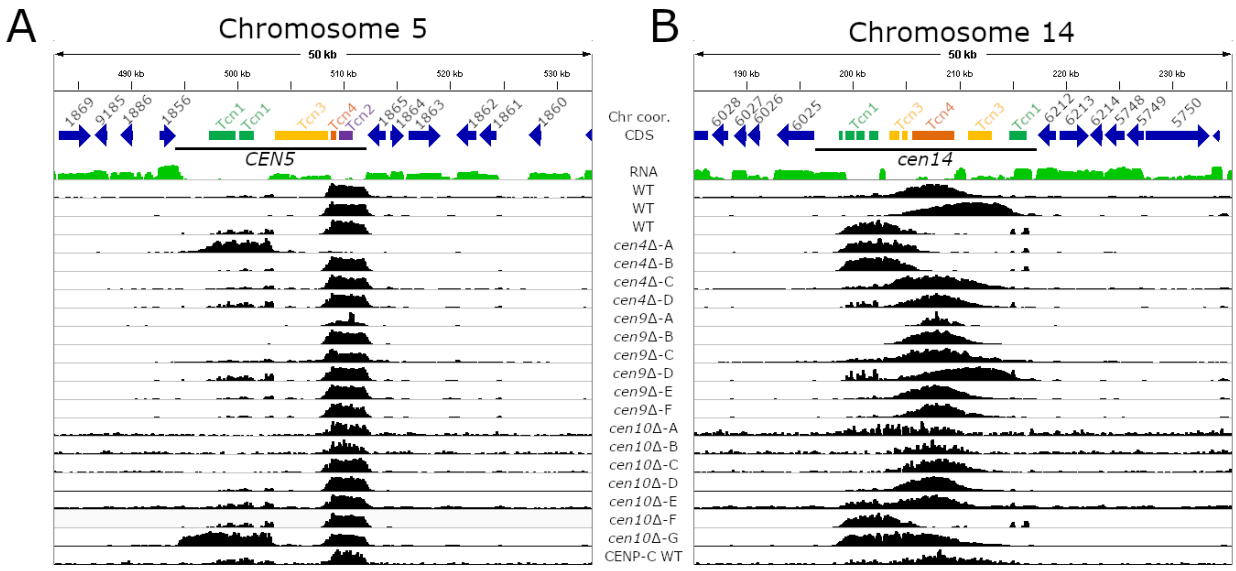

Figure 5. Centromeric drift occurs in pericentric regions of *C. deuterogattii*.

(A) Centromeric drift in *CEN*5 is correlated with the presence of a large Tcn3 element. The truncated Tcn3 element functions as a barrier and for all isolates with the exception of *cen10*Δ-G, CENP-A is present on only one side of the truncated Tcn3 element. **(B)** Centromere drift occurs in *CEN*14; however, in this case there is no Tcn element acting as a barrier and the CENP-A peaks drift within the ORF-free region of *CEN*14 and lack a defining feature.

Discussion

Neocentromere formation

Previously, we have shown that most neocentromeres of *C. deuterogattii* chromosomes 9 and 10 form in close proximity to the deleted native centromere [32]. In *S. pombe*, neocentromeres form near heterochromatic repeats in sub-telomeric regions [21], whereas in *C. albicans* and chicken DT40 cells, neocentromeres typically form in close proximity to native centromeres, likely due to the seeding of CENP-A near the native centromere (i.e. the CENP-A cloud) [10–12,35]. The CENP-A cloud not only influences the chromosomal regions directly flanking the centromere, but also chromosomal regions that are in close vicinity to the centromere due to chromosome folding, and these regions can be several kilobases away from the native centromeric chromosomal location [36]. While *C. albicans* has clustered centromeres, neocentromere sites in chicken cells were found to be connected to heterochromatin that might have seeded them with CENP-A. Several studies employing HiC analysis have shown that clustering of centromeres and pericentric regions occurs in *S. pombe*, plants, and chicken cells [9,37,38]. Fluorescence microscopy studies have shown that the native centromeres of *C. deuterogattii* and *C. neoformans* are organized as multiple nuclear foci during interphase, but cluster together into one focus during mitosis [39]. This stage-specific clustering has been shown to be dependent on the association of centromeres with the nuclear envelope, which involves proteins other than kinetochore proteins [40]. Such a connection might aid in clustering of the DNA next to a deleted centromere along with other native centromeres during mitosis, thus enabling seeding of *CEN*-proximal regions with CENP-A [24,39].

Naturally occurring dicentric B-chromosomes in maize (*Zea mays*) are stabilized by silencing of one of the centromeres [38]. Interestingly the silenced centromere can be reactivated due to chromosome recombination, which suggests there is an epigenetic memory or at least a preference for the centromere location and chromosomal content. Dicentric chromosomes in *S. pombe* have been obtained by chromosomal fusion of chromosomes 2 and 3 [41]. This study revealed multiple fates for these dicentric chromosomes: 1) epigenetic silencing of a centromere, resulting in a stable, fused chromosome; 2) stabilization of the chromosome fusion by deletion of a centromeric region; or 3) chromosome breakage resulting in three chromosomes with

functional centromeres. Furthermore, silencing of a centromere was found to be associated with spreading of heterochromatin into the CENP-A-bound central core.

All of these examples describe chromosome fusions resulting in stable chromosomes with only one active centromere. Stable dicentric chromosomes in human cell lines have been engineered by displacement of the telomere protein Trf2 [42]. Half of the isolates obtained were *bona fide* dicentrics and stable for >150 cell divisions. In the other half of the isolates, the chromosome was stabilized either by partial deletion of centromeric repeats in one centromere or via epigenetic silencing.

Native centromeres, but not neocentromeres, are heterochromatic in *C. deuterogattii*

Like the centromeres of the closely related species *C. neoformans* [28], our studies show that the centromeres and pericentric regions of *C. deuterogattii* are enriched for the heterochromatic mark H3K9me2. This feature is also observed in several other fungi, including *Neurospora crassa* and *Fusarium fujikuroi*, in which centromeres are enriched for H3K9me2/3 [43,44]. Similar observations have been made in the oomycete *Phytophthora sojae* [45]. This is in contrast to *S. pombe*, in which the CENP-A-enriched regions lack heterochromatin, but pericentric regions are enriched for heterochromatin [20–22]. A few other fungi, such as *C. albicans* and *S. cerevisiae*, lack the H3K9me2/3 modification in their genomes. The regional centromeres of the fungal plant pathogen *Zymoseptoria tritici* lack heterochromatic marks and span actively expressed genes, while transposable elements are silenced by this heterochromatic mark [46].

Beyond fungi, neocentromeres have been studied in a number of additional species. The native centromeres of chickens can be repetitive or non-repetitive, and heterochromatin enrichment depends upon the presence of centromeric repeats [9]. Similar to their non-repetitive native centromeres, chicken neocentromeres lack repeats and are not enriched for any of the histone marks tested [9]. Centromere repositioning occurs in plants when evolutionarily new centromeres (ENCs) are formed; however, the majority of ENCs are formed in heterochromatic regions flanking the native centromere [47]. These studies suggest that while

heterochromatin might play a role in defining the location of neocentromeres in some species, it is mostly dispensable for neocentromere function.

The centromeric *URA5* transgene lacks heterochromatin silencing

Some species have native centromeres that span actively expressed genes, and human artificial chromosomes have centromeres that are only sparsely enriched for heterochromatin [46,48]. Neocentromeres in *C. deuterogattii* can also span actively expressed genes, whereas neocentromeres in *C. albicans* and *S. pombe* silence the underlying genes [10,12,21,32]. Along similar lines, when a transgene is inserted into the native centromeres of *C. albicans* or *S. pombe*, the gene is silenced. In *S. pombe*, the extent of transgene silencing depends on the location of the insertion and the transgene is not silenced when integrated into a CENP-A-binding region [49]. On the other hand, when the *URA3* gene is inserted into a centromere of *C. albicans*, the gene is expressed conditionally and exhibits reversible silencing [50]. Moreover, the transgene in *C. albicans* is bound by CENP-A when it is silenced but not when it is expressed. We found that neocentromeric genes are expressed in *C. deuterogattii* [32]. Integration of a transgene into a native centromere in *C. deuterogattii* also led to a shift of CENP-A binding, similar to observations in *C. albicans*. The *URA5* gene integrated into *CEN*2 is expressed based on complementation of the auxotrophic uracil requirement and the *URA5* gene was not decorated with H3K9me2. Our ChIP-seq experiments for CENP-A and the ChIP-qPCR experiments for H3K9me2 conducted to characterize this phenomenon were performed with cells isolated from overnight cultures without selection for *URA5* expression, providing further evidence for stability of the transgene and its expression status. However, we cannot exclude the possibility that CENP-A binding might revert back to the native *CEN*2 position and that H3K9me2 would span the *URA5* gene if the strain were grown for many generations in media without selection.

The genomic distributions of H3K9me2 and 5mC DNA methylation differ in *C. deuterogattii*

5mC DNA methylation and H3K9me2 are present at centromeres in many fungal and plant species. The pericentric regions of *N. crassa* are enriched for 5mC, while H3K9me3 is found at both pericentric and CENP-A-enriched regions [43]. *S. pombe* lacks 5mC DNA methylation but

H3K9me2 is present in the pericentric regions. As in *C. neoformans*, there is a link between DNA methylation and H3K9me2 enrichment in tomatoes, and it was shown that the 5mC DNA methylation enrichment level at transposable elements in plants depends on the age of the specific transposable element [51].

In the closely related species *C. neoformans*, 5mC DNA methylation is enriched in heterochromatic regions and deletion of *CLR4*, which encodes the enzyme that generates H3K9me2, leads to reduced 5mC levels but does not abolish them completely [29]. The authors of that study further showed that H3K9me2 regulates DNA methylation levels through two different pathways: the first pathway depends on binding of the N-terminal chromodomain of the DNA methyltransferase Dnmt5 directly to H3K9me2, and the second pathway that directs 5mC DNA methylation involves the HP1 (Heterochromatin protein 1) ortholog Swi6, which recruits Dnmt5 to H3K9me2-enriched regions. In the double mutant, in which *SWI6* is deleted and the N-terminal chromodomain of Dnmt5 is mutated, 5mC levels are similar to those in the *clr4*Δ mutant [29]. These results and our finding that not all 5mC-enriched regions are enriched for H3K9me2, suggest that H3K9me2 is not the only determining factor for DNA methylation in the *Cryptococcus* species complex. Swi6 is also present in *C. deuterogattii,* and future studies could focus on the interaction of these proteins in this species and how this may differ from their interactions in *C. neoformans*. Moreover, DNA methylation levels in *C. deuterogattii* are very diminished and restricted compared to *C. neoformans*, suggesting a different and possibly less prominent role for this modification in *C. deuterogattii*.

Centromeric drift in *C. deuterogattii*

Several studies have shown that CENP-A peaks drift (or shift) within a defined chromosomal region. For example, CENP-A profiles of several maize isolates showed that centromeric drift occurs in centromeres 5 and 8 [52]. Centromeres 4 and 7 in donkeys have two distinct CENP-A peaks; however, hybrid mules derived from donkeys and horses have only one CENP-A-enriched region in both centromeres [52]. Compared to donkey centromeres 4 and 7, the CENP-A-enriched regions of mule centromeres are slightly shifted and not completely homologous to the parental donkey CENP-A locations. ChIP-seq of five independent horses

revealed five unique CENP-A-bound regions for the non-repetitive centromere 11 [53]. Additionally, the CENP-A-enriched region of the non-repetitive centromere Z in chicken DT40 cells drifts in the population, and this drift occurs more frequently in CENP-U and CENP-S mutants [54].

We observed a similar phenomenon of centromere drift in *C. deuterogattii*. We observed drifting of CENP-A binding in two different centromeres: *CEN*5 and *CEN*14. Centromere drift has not been observed in any fungal species yet, and this is possibly due to a lack of multiple ChIP-seq datasets for a single species. While we observed that drift in *CEN*5 was associated with the presence of fragments of a specific transposon, *CEN*14 lacked this feature. Based on this and other studies, centromere drift seems to happen more frequently with centromeres lacking active transposable elements compared to centromeres harboring active transposable elements. This model is further supported by previous studies where centromere drift occurred in the neocentromeres of species with repeat-rich native centromeres, suggesting that accumulation of repeats or satellite DNA occurs after neocentromere formation and possibly restricts centromere drift.

Material and methods

Strains, primers, and culture conditions

Primers are listed in Supplementary Table 1. Strains used in this study are listed in Supplementary Table 2. All strains were stored in glycerol stocks at -80°C, inoculated on solid YPD (yeast extract, peptone, and dextrose) medium, and grown for two days at 30°C. Liquid YPD cultures were inoculated with single colonies from solid medium and grown, while shaking, at 30°C overnight.

Genetic manipulations

Prior to biolistic transformation or CRISPR-Cas9-mediated transformation, PCR products with homologous DNA sequences (1 to 1.5 kb) flanking the deleted region were PCR-amplified with Phusion High-Fidelity DNA Polymerase (NEB, Ipswich MA, USA). These flanking PCR products were fused to both sides of either a *NEO* or *NAT* dominant selectable marker via overlap PCR, conferring G418 or nourseothricin resistance, respectively. For the deletion of *CEN4*, two guide RNAs (gRNAs) flanking *CEN4* were designed. The chromosome fusion of Chrs 4 and 10 was also mediated by two gRNAs targeting the sub-telomeric region of Chr 4 and the sub-telomeric region of Chr 10, respectively. Biolistic transformation and CRISPR-Cas9-mediated transformation were performed as previously described, and transformants were selected on YPD medium containing G418 (200 μg/mL) or nourseothricin (100 μg/mL) [32,55,56]. All mutations were introduced into a previously generated *C. deuterogattii* strain that expresses mCherry-CENP-A. To confirm the correct replacement of the region of interest by the appropriate drug resistance marker, PCR analysis was conducted for the 5' junction, 3' junction, and spanning the target locus.

To delete the *DMT5* gene, 1-kb homologous flanking sequences were utilized to make a deletion allele with a *NEO* selectable marker in a similar approach as described above. Two independent gRNAs cleaving the *DMT5* gene at two different locations were used along with the recombination template for the CRISPR-Cas9-mediated transformation. Transformants obtained were screened and confirmed with 5' and 3' junction PCRs.

Prior to the integration of *URA5* in centromere 2 (*CEN2*), an mCherry-CENP-A tagged wild-type strain was plated on medium containing 5-FOA, and spontaneously-derived resistant

colonies were isolated. These colonies were further screened for their lack of growth on SD-URA drop-out medium and normal growth on 5-FOA-containing medium. Colonies growing on 5-FOA but not on SD-ura were identified as *ura5* mutant colonies and utilized for further experiments. One such 5-FOA resistant, uracil auxotrophic isolate (KS174) was shown to be auxotrophic for uracil and sequence analysis of the *URA5* gene revealed a single base-pair deletion mutation resulting in the lack of an adenine at position 156 resulting in a frameshift. This *ura5* mutant then served as the background strain in which the *URA5* gene was integrated into *CEN2*, and transformants were selected on SD-ura medium. The integration of *URA5* at the correct location was then confirmed by PCR analysis and Southern blot hybridization (as shown in Figures S1 and S7).

For the dicentric isolates, chromosome fusion was mediated by CRISPR-Cas9 with two guide RNAs targeting the sub-telomeric regions of Chrs 4 and 10 and the resulting double-stranded breaks were repaired using a linear overlap PCR product that contained a 1.5 kb region homologous to Chr 4, a *NAT* selection marker, and a 1.5 kb region homologous to Chr 10.

Chromatin immunoprecipitation (ChIP) followed by high-throughput sequencing or qPCR

ChIP analyses were performed as previously described [32]. A polyclonal antibody against mCherry (ab183628, Abcam) was used to identify the CENP-A-enriched regions. To identify histone H3K9me2-enriched regions, ChIP analysis was conducted with a murine monoclonal antibody against histone H3K9me2 (ab1220, Abcam). Subsequently, the samples were analyzed by qPCR or subjected to library preparation and Illumina sequencing at the Duke University Sequencing and Genomic Technologies Shared Resource facility. Sequencing was performed with a NovaSeq 6000 sequencer and 50-bp PE reads were obtained. Reads were mapped to the *C. deuterogattii* R265 reference genome, similar to whole-genome sequencing reads. To analyze the ChIP-seq data, the ChIP sample was normalized with the input sample and visualized with IGV viewer. qPCRs were performed in triplicate with Brilliant III Ultra-Fast SYBR® Green qPCR Master Mix (Agilent Technologies) on an ABI 1900HT qPCR machine.

Fluctuation assays

Fluctuation assays were performed as previously described [57].

Pulsed-field gel electrophoresis (PFGE)

Isolation of whole chromosomes and conditions for PFGE analysis were performed as previously described [58].

Bisulfite sequencing

Genomic DNA was isolated following the CTAB method, checked for quality, and submitted to the Duke University Sequencing and Genomic Technologies Shared Resource facility for whole-genome bisulfite sequencing [59]. After sequencing, the reads were analyzed and mapped to the *C. deuterogattii* R265 reference genome using Bismark (https://www.bioinformatics.babraham.ac.uk/projects/bismark/). The default parameters for alignment and methylation extraction were used for both the wild-type strain and the *dmt5*Δ mutant. Methylation extraction results were obtained as a .bed file that was then imported into IGV for visualization.

Deposited data

ChIP sequences have been deposited under NCBI BioProject Accession ID: #####.

Acknowledgements

We thank Shelby Priest for comments on the manuscript. These studies were supported by NIH/NIAID grants R01 AI050113-16 and R37 MERIT award AI039115-23 to JH. JH is co-director and fellow of the CIFAR program Fungal Kingdom: Threats & Opportunities.

References

1. Barra V, Fachinetti D. The dark side of centromeres: types, causes and consequences of structural abnormalities implicating centromeric DNA. Nat Commun. 2018;9: 4340. doi:10.1038/s41467-018-06545-y
2. Freitag M. The kinetochore interaction network (KIN) of ascomycetes. Mycologia. 2016;108: 485–505. doi:10.3852/15-182
3. Navarro-Mendoza MI, Pérez-Arques C, Panchal S, Nicolás FE, Mondo SJ, Ganguly P, et al. Early diverging fungus *Mucor circinelloides* lacks centromeric histone CENP-A and displays a mosaic of point and regional rentromeres. Curr Biol. 2019;29: 3791-3802.e6. doi:10.1016/j.cub.2019.09.024
4. Drinnenberg IA, DeYoung D, Henikoff S, Malik HS. Recurrent loss of CenH3 is associated with independent transitions to holocentricity in insects. eLife. 2014;3: 1–19. doi:10.7554/eLife.03676
5. Hooff JJ, Tromer E, Wijk LM, Snel B, Kops GJ. Evolutionary dynamics of the kinetochore network in eukaryotes as revealed by comparative genomics. EMBO Rep. 2017;18: 1559–1571. doi:10.15252/embr.201744102
6. Friedman S, Freitag M. Centrochromatin of fungi. Centromeres and Kinetochores. Cham, Switzerland: Springer International Publishing; 2017. pp. 85–109. doi:10.1016/S0092-8674(03)00115-6
7. Kobayashi N, Suzuki Y, Schoenfeld LW, Müller CA, Nieduszynski C, Wolfe KH, et al. Discovery of an unconventional centromere in budding yeast redefines evolution of point centromeres. Curr Biol. 2015;25: 2026–2033. doi:10.1016/j.cub.2015.06.023
8. Scott KC, Sullivan BA. Neocentromeres: a place for everything and everything in its place. Trends Genet. 2013;30: 66–74. doi:10.1016/j.tig.2013.11.003
9. Nishimura K, Komiya M, Hori T, Itoh T, Fukagawa T. 3D genomic architecture reveals that neocentromeres associate with heterochromatin regions. J Cell Biol. 2019;218: 134–149. doi:10.1083/jcb.201805003
10. Shang W-HH, Hori T, Martins NMCC, Toyoda A, Misu S, Monma N, et al. Chromosome engineering allows the efficient isolation of vertebrate neocentromeres. Dev Cell.

2013;24: 635–648. doi:10.1016/j.devcel.2013.02.009

11. Thakur J, Sanyal K. Efficient neocentromere formation is suppressed by gene conversion to maintain centromere function at native physical chromosomal loci in *Candida albicans*. Genome Res. 2013; 638–652. doi:10.1101/gr.141614.112
12. Ketel C, Wang HSW, McClellan M, Bouchonville K, Selmecki A, Lahav T, et al. Neocentromeres form efficiently at multiple possible loci in *Candida albicans*. PLOS Genet. 2009;5: e1000400. doi:10.1371/journal.pgen.1000400
13. Burrack LS, Hutton HF, Matter KJ, Clancey SA, Liachko I, Plemmons AE, et al. Neocentromeres provide chromosome segregation accuracy and centromere clustering to multiple loci along a *Candida albicans* chromosome. PLOS Genet. 2016;12: e1006317. doi:10.1371/journal.pgen.1006317
14. Mishra PK, Baum M, Carbon J. Centromere size and position in *Candida albicans* are evolutionarily conserved independent of DNA sequence heterogeneity. Mol Genet Genomics. 2007;278: 455–65. doi:10.1007/s00438-007-0263-8
15. Freire-Benéitez V, Price RJ, Buscaino A. The chromatin of *Candida albicans* pericentromeres bears features of both euchromatin and heterochromatin. Front Microbiol. 2016;7: 1–11. doi:10.3389/fmicb.2016.00759
16. Price RJ, Weindling E, Berman J, Buscaino A. Chromatin profiling of the repetitive and nonrepetitive genomes of the human fungal pathogen *Candida albicans*. mBio. 2019;10: 1–17. doi:10.1128/mBio.01376-19
17. Freire-Benéitez V, Price RJ, Tarrant D, Berman J, Buscaino A. *Candida albicans* repetitive elements display epigenetic diversity and plasticity. Sci Rep. 2016;6: 1–12. doi:10.1038/srep22989
18. Partridge JF, Borgstrøm B, Allshire RC. Distinct protein interaction domains and protein spreading in a complex centromere. Genes Dev. 2000;14: 783–791. doi:10.1101/gad.14.7.783
19. Volpe TA, Kidner C, Hall IM, Teng G, Grewal SIS, Martienssen RA. Regulation of heterochromatic silencing and histone H3 Lysine-9 methylation by RNAi. Science. 2002;297: 1833–1837. doi:10.1126/science.1074973

20. Folco HD, Pidoux AL, Urano T, Allshire RC, Drosophila I. Heterochromatin and RNAi are required to establish CENP-A chromatin at centromeres. Science. 2008;1: 94–97. doi:10.1126/science.1150944

21. Ishii K, Ogiyama Y, Chikashige Y, Soejima S, Masuda F, Kakuma T, et al. Heterochromatin integrity affects chromosome reorganization after centromere dysfunction. Science. 2008;321: 1088–91. doi:10.1126/science.1158699

22. Ohno Y, Ogiyama Y, Kubota Y, Kubo T, Ishii K. Acentric chromosome ends are prone to fusion with functional chromosome ends through a homology-directed rearrangement. Nucleic Acids Res. 2016;44: 232–244. doi:10.1093/nar/gkv997

23. Janbon G, Ormerod KL, Paulet D, Byrnes EJ, Yadav V, Chatterjee G, et al. Analysis of the genome and transcriptome of *Cryptococcus neoformans* var. *grubii* reveals complex RNA expression and microevolution leading to virulence attenuation. PLOS Genet. 2014;10: e1004261. doi:10.1371/journal.pgen.1004261

24. Yadav V, Sun S, Billmyre RB, Thimmappa BC, Shea T, Lintner R, et al. RNAi is a critical determinant of centromere evolution in closely related fungi. Proc Natl Acad Sci. 2018; doi:10.1073/pnas.1713725115

25. Wang X, Hsueh Y-P, Li W, Floyd A, Skalsky R, Heitman J. Sex-induced silencing defends the genome of *Cryptococcus neoformans* via RNAi. Genes Dev. 2010;24: 2566–2582. doi:10.1101/gad.1970910

26. Wang X, Wang P, Sun S, Darwiche S, Idnurm A, Heitman J. Transgene induced co-suppression during vegetative growth in *Cryptococcus neoformans*. PLOS Genet. 2012;8: e1002885. doi:10.1371/journal.pgen.1002885

27. Janbon G, Maeng S, Yang D-H, Ko Y-J, Jung K-W, Moyrand F, et al. Characterizing the role of RNA silencing components in *Cryptococcus neoformans*. Fungal Genet Biol. 2010;47: 10. doi:10.1016/j.fgb.2010.10.005

28. Dumesic PA, Homer CM, Moresco JJ, Pack LR, Shanle EK, Coyle SM, et al. Product binding enforces the genomic specificity of a yeast polycomb repressive complex. Cell. 2015;160: 204–218. doi:10.1016/j.cell.2014.11.039

29. Catania S, Dumesic PA, Pimentel H, Nasif A, Stoddard CI, Burke JE, et al. Evolutionary

persistence of DNA methylation for millions of years after ancient loss of a de novo methyltransferase. Cell. 2020;180: 263-277.e20. doi:10.1016/j.cell.2019.12.012

30. Farrer RA, Desjardins CA, Sakthikumar S, Gujja S, Saif S, Zeng Q, et al. Genome evolution and innovation across the four major lineages of *Cryptococcus gattii*. mBio. 2015;6: e00868-15. doi:10.1128/mBio.00868-15

31. Feretzaki M, Billmyre RB, Clancey SA, Wang X, Heitman J. Gene network polymorphism illuminates loss and retention of novel RNAi silencing components in the *Cryptococcus* pathogenic species complex. PLOS Genet. 2016;12: e1005868. doi:10.1371/journal.pgen.1005868

32. Schotanus K, Heitman J. Centromere deletion in *Cryptococcus deuterogattii* leads to neocentromere formation and chromosome fusions. eLife. 2020;9: e56026. doi:10.7554/eLife.56026

33. Gröhs Ferrareze PA, Maufrais C, Streit RSA, Priest SJ, Cuomo C, Heitman J, et al. Application of an optimized annotation pipeline to the *Cryptococcus deuterogattii* genome reveals dynamic primary metabolic gene clusters and genomic impact of RNAi loss. bioRxiv. 2020; doi:https://doi.org/10.1101/2020.09.01.278374

34. Schneider RO, de Souza Süffert Fogaça N, Kmetzsch L, Schrank A, Vainstein MH, Staats CC. Zap1 regulates zinc homeostasis and modulates virulence in *Cryptococcus gattii* . PLoS One. 2012;7: e43773. doi:10.1371/journal.pone.0043773

35. Scott KC, Bloom KS. Lessons learned from counting molecules: How to lure CENP-A into the kinetochore. Open Biol. 2014;4. doi:10.1098/rsob.140191

36. Fukagawa T, Earnshaw WC. The centromere: Chromatin foundation for the kinetochore machinery. Dev Cell. 2014;30: 496–508. doi:10.1016/j.devcel.2014.08.016

37. Mizuguchi T, Fudenberg G, Mehta S, Belton JM, Taneja N, Folco HD, et al. Cohesin-dependent globules and heterochromatin shape 3D genome architecture in *S. pombe*. Nature. 2014;516: 432–435. doi:10.1038/nature13833

38. Han F, Gao Z, Birchler JA. Reactivation of an inactive centromere reveals epigenetic and structural components for centromere specification in maize. Plant Cell. 2009;21: 1929–1939. doi:10.1105/tpc.109.066662

39. Kozubowski L, Yadav V, Chatterjee G, Sridhar S, Yamaguchi M, Kawamoto S, et al. Ordered kinetochore assembly in the human-pathogenic basidiomycetous yeast *Cryptococcus neoformans*. mBio. 2013;4: 1–8. doi:10.1128/mBio.00614-13

40. Yadav V, Sanyal K. Sad1 spatiotemporally regulates kinetochore clustering to ensure high-fidelity chromosome segregation in the human fungal pathogen *Cryptococcus neoformans*. mSphere. 2018;3: 1–12. doi:10.1128/msphere.00190-18

41. Sato H, Masuda F, Takayama Y, Takahashi K, Saitoh S. Epigenetic inactivation and subsequent heterochromatinization of a centromere stabilize dicentric chromosomes. Curr Biol. 2012;22: 658–667. doi:https://doi.org/10.1016/j.cub.2012.02.062

42. Stimpson KM, Song IY, Jauch A, Holtgreve-Grez H, Hayden KE, Bridger JM, et al. Telomere disruption results in non-random formation of de novo dicentric chromosomes involving acrocentric human chromosomes. PLOS Genet. 2010;6: e1001061. doi:10.1371/journal.pgen.1001061

43. Smith KM, Phatale PA, Sullivan CM, Pomraning KR, Freitag M. Heterochromatin is required for normal distribution of *Neurospora crassa* CenH3. Mol Cell Biol. 2011;31: 2528–2542. doi:10.1128/MCB.01285-10

44. Wiemann P, Sieber CMK, von Bargen KW, Studt L, Niehaus EM, Espino JJ, et al. Deciphering the cryptic genome: genome-wide analyses of the rice pathogen *Fusarium fujikuroi* reveal complex regulation of secondary metabolism and novel metabolites. PLOS Pathog. 2013;9. doi:10.1371/journal.ppat.1003475

45. Fang Y, Coelho MA, Shu H, Schotanus K, Thimmappa, Bhagya C, Yadav V, et al. Long transposon-rich centromeres in an oomycete reveal divergence of centromere features in Stramenopila-Alveolata-Rhizaria lineages. PLOS Genet. 2019; 1–30. doi:doi.org/10.1371/journal.pgen.1008646

46. Schotanus K, Soyer JL, Connolly LR, Grandaubert J, Happel P, Smith KM, et al. Histone modifications rather than the novel regional centromeres of *Zymoseptoria tritici* distinguish core and accessory chromosomes. Epigenetics Chromatin. 2015;8: 41. doi:10.1186/s13072-015-0033-5

47. Mandáková T, Hloušková P, Koch MA, Lysak MA. Genome evolution in *Arabideae* was

marked by frequent centromere repositioning. Plant Cell. 2020; doi:10.1105/tpc.19.00557

48. Alonso A, Hasson D, Cheung F, Warburton PE. A paucity of heterochromatin at functional human neocentromeres. Epigenetics and Chromatin. 2010;3: 1–12. doi:10.1186/1756-8935-3-6

49. Allshire RC, Javerzat JP, Redhead NJ, Cranston G. Position effect variegation at fission yeast centromeres. Cell. 1994;76: 157–169. doi:10.1016/0092-8674(94)90180-5

50. Sreekumar L, Jaitly P, Chen Y, Thimmappa BC, Sanyal A, Sanyal K. Cis- and trans-chromosomal interactions define pericentric boundaries in the absence of conventional heterochromatin. Genetics. 2019;212: 1121–1132. doi:10.1534/genetics.119.302179

51. Wang Z, Baulcombe DC. Transposon age and non-CG methylation. Nat Commun. 2020;11: 1–9. doi:10.1038/s41467-020-14995-6

52. Nergadze SG, Piras FM, Gamba R, Corbo M, Cerutti F, McCarter JGW, et al. Birth, evolution, and transmission of satellite-free mammalian centromeric domains. Genome Res. 2018;28: 789–799. doi:10.1101/gr.231159.117

53. Purgato S, Belloni E, Piras FM, Zoli M, Badiale C, Cerutti F, et al. Centromere sliding on a mammalian chromosome. Chromosoma. 2015;124: 277–287. doi:10.1007/s00412-014-0493-6

54. Hori T, Kagawa N, Toyoda A, Fujiyama A, Misu S, Monma N, et al. Constitutive centromere-associated network controls centromere drift in vertebrate cells. J Cell Biol. 2017;216: 101–113. doi:10.1083/jcb.201605001

55. Davidson RC, Blankenship JR, Kraus PR, De M, Berrios J, Hull CM, et al. A PCR-based strategy to generate integrative targeting alleles with large regions of homology. Microbiology. 2002;2: 2607–2615.

56. Billmyre RB, Clancey SA, Heitman J. Natural mismatch repair mutations mediate phenotypic diversity and drug resistance in *Cryptococcus deuterogattii*. eLife. 2017;6: e28802. doi:10.7554/eLife.28802

57. Priest SJ, Coelho MA, Mixão V, Clancey S, Xu Y, Sun S, et al. Factors enforcing the species boundary between the human pathogens *Cryptococcus neoformans* and *Cryptococcus*

deneoformans. bioRxiv. 2020;

58. Findley K, Rodriguez-Carres M, Metin B, Kroiss J, Fonseca A, Vilgalys R, et al. Phylogeny and phenotypic characterization of pathogenic *Cryptococcus* species and closely related saprobic taxa in the tremellales. Eukaryot Cell. 2009;8: 353–361. doi:10.1128/EC.00373-08

59. Pitkin JW, Panaccione DG, Walton JD. A putative cyclic peptide efflux pump encoded by the TOXA gene of the plant-pathogenic fungus *Cochliobolus carbonum*. Microbiology. 1996;142: 1557–1565. doi:10.1099/13500872-142-6-1557

Supplementary figures

Figure S1. PCR confirmation of gene deletion mutants.

PCR confirmations for *dmt5*Δ, *cen4*Δ, and *clr4*Δ mutants. For *dmt5*Δ mutants, the 5' and 3' junction PCRs are shown. For *cen4*Δ and *clr4*Δ mutants, the 5' and 3' junction PCRs and internal PCRs are shown. Panel **(A)** shows PCR confirmations for *dmt5*Δ, panel **(B)** shows PCR confirmations for *cen4*Δ, and panel (C) shows PCR confirmations for *clr4*Δ.

Figure S2. Formation of a dicentric chromosome.

(A) Schematic view of the formation of a dicentric chromosome. At the top, a sub-telomeric region of wild-type Chr 4 and 10 is shown. The chromosomal targets of the guide RNAs are depicted by red triangles. The double-stranded DNA breaks were repaired by homologous recombination that was mediated by an overlap PCR product containing regions homologous to both chromosomes flanking a selectable marker. The homologous regions are depicted by black lines and labeled "Flank". **(B)** Spanning PCR analysis confirmed chromosome fusion in mutant 1 ($G418^R$ 4) and 2 (Nat^R 6) are shown. **(C)** PFGE analysis with all mutants obtained after recovering mutants on selective media. Dicentric 1 and2 lack a wild-type size band for Chr 10 confirming the spanning PCR products in panel A. Instead of the wild-type Chr 10 band, mutant dicentric 1 has an additional band (~60 kb higher) than the wild-type Chr 10 band. Based on this PFGE, dicentric 1 has no additional resolvable bands. **(D)** PFGE analysis shows that wild-type Chr 10 is absent in dicentric strains 1 and 2, which confirms that Chrs 4 and 10 are fused in these two strains. However, an additional chromosomal band at 60 kb appears in dicentric 1, suggesting that the dicentric chromosome broke. **(E)** ChIP-qPCR analyses were performed for dicentric mutant 2 and wild type. CENP-A fold enrichment is shown for *CEN*4 and *CEN*10 and a positive control (*CEN*6). The fold-enrichment was compared to actin

Figure S3. Nanopore sequencing reveals that the fused dicentric chromosome underwent chromosomal breakage.

Panel (**A**) shows the two scaffolds that harbor the ~60 kb duplicated region. Scaffold 13 corresponds to chromosome 4 and scaffold 12 is the broken chromosomal fusion product of

chromosome 10 and 4. (**B**) Illumina sequencing indicated that a ~60 kb region of chromosome 4 is duplicated and has a ploidy level of 2. The panel shows short reads of dicentric 2 mapped to the wild-type genome. As a control, dicentric isolate 1 was sequenced and this strain has a ploidy level of 1 for chromosome 4. (**C**) Putative model explaining the chromosome fusion between chromosome 4 and 10, the intermediate state in which the fused chromosome is intact and the final karyotype after the chromosomal breakage.

Figure S4. Fluctuation assay with WT, *dmt5*Δ, and *clr4*Δ mutants.

Panels **(A & B)** show a fluctuation assay with *dmt5*Δ and *clr4*Δ mutants, respectively. As a positive control an *msh2*Δ hypermutator strain was included. Panels **(C & D)** show fluctuation analysis (N=10) to measure mutation rates to resistance to antifungal drug 5-fluorocytosine (5FC), which results from loss of function mutations in several possible genes. Panel **D** shows the same information as in panel **C**, but the data from the *msh2*Δ mutant is removed from the plot. Panels **(E & F)** show fluctuation analysis (N=10) to measure mutation rates on the combination of the antifungal drugs rapamycin and FK506, which selects for loss of function mutations in the *FRR1* gene encoding the common drug target FKBP12. Panel **F** shows the same information as in panel **E**, but the data from the *msh2*Δ mutant is removed from the plot. Error bars represent upper and lower 95% confidence intervals.

Figure S5. Whole-genome view of histone marker enrichment and 5mC DNA methylation.

All 14 chromosomes are depicted for both *C. neoformans* and *C. deuterogattii*. For each chromosome, plots presented show the chromosome coordinates, gene content, TE content, RNA-seq (for *C. deuterogattii* only), CENP-A enrichment, H3K9me2 enrichment, H3K27me3 enrichment (for *C. neoformans* only), and 5mC data [28].

Figure S6. *URA5* integration in *CEN*2.

(A) A schematic model of *CEN*2 is shown. The CENP-A peak is shown in black, genes flanking the pericentric region are shown in blue. The *URA5* gene is shown in red and the truncated transposable elements are shown in orange and green (Tcn4 and Tcn6). Hypothetical

outcomes for *URA5* integration into *CEN*2 include: (1) CENP-A would cover the *URA5* gene making the CENP-A-bound region larger than the native centromere. (2) The *URA5* gene might divide the CENP-A-enriched region into two independent regions. (3 & 4) Only one of the regions flanking *URA5* would be enriched for CENP-A, generating a smaller centromere. (5) *URA5* integration might abolish *CEN*2 function leading to neocentromere formation. **(B)** Prior to the ChIP-seq experiment for *URA5*ins*CEN*2, the strain was tested for *URA5* expression. For all three plates, the wild type was included as a control and *ura5*- is the parental strain in which the *URA5* gene was integrated into *CEN*2. As expected all three strains grew on the control medium (YPD). Only *ura5*- was able to grow on medium containing 5-FOA. On SD-URA medium, the wild-type and *URA5*ins*CEN*2 strains were able to grow due to the presence of an active *URA5* gene.

Figure S7. Southern blot analysis confirms *URA5* integration into *CEN*2 in the genome.

(**A**) Southern blot analysis with three independent restriction digest analyses of the *CEN*2ins*URA5* strain. The *URA5* gene was used as a probe. The Southern blot hybridization pattern confirms that a single copy of the *URA5* gene has been inserted at the desired targeted site in *CEN*2. (**B**) Table indicating the expected restriction digest product sizes. Because the *CEN*2ins*URA5* strain still has the *ura5*- gene present at its native locus, the pattern shows products from both the *ura5* native gene and the *CEN*2ins*URA5* transgene.

Figure S8. The *C. deuterogattii DMT5* gene encoding Dnmt5 is intact and expressed similarly to the *C. neoformans* ortholog.

(A) Alignment view with genomic regions surrounding the *DMT5* gene (CNAG_07752 and CNBG_3156) of *C. neoformans* and *C. deuterogattii*. Shown at the top of the panel are the chromosomal coordinates of *C. neoformans*. The genes are indicated with blue arrows with exons and introns marked. For *C. deuterogattii*, two gene annotations are shown. The old (FungiDB) annotation predicted that *DMT5* was truncated. The new genome annotation shows that the *DMT5* gene is full-length and has a similar length to the *C. neoformans* ortholog [33]. **(B)** Detailed view of *C. deuterogattii* introns and exons of the *DMT5* gene are supported by RNA-seq reads, which are shown to support the gene structure.