Centromere deletion in *Cryptococcus deuterogattii* leads to neocentromere formation and chromosome fusions

Klaas Schotanus 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

Running title: Neocentromere formation in *Cryptococcus deuterogattii*

Abstract

The human fungal pathogen *Cryptococcus deuterogattii* is RNAi-deficient and lacks active transposons in its genome. *C. deuterogattii* has regional centromeres that contain only transposon relics that are smaller than those of closely related RNAi-proficient species. To investigate the fate of a chromosome upon loss of a centromere, the centromere of chromosome 10 was deleted in *C. deuterogattii*. This deletion resulted in neocentromere formation in two non-repetitive genomic regions that were smaller than the native centromere. Interestingly, these neocentromeres cover two or three genes and maintained expression levels similar to wild-type. *cen10*Δ mutant strains exhibited growth defects and were aneuploid for chromosome 10. At an elevated growth temperature (37°C), *cen10*Δ mutants produced heterogeneous colonies. The larger colonies displayed wild-type growth rates as the *cen10*Δ chromosome was found to have stabilized by fusing with another native chromosome. Following chromosomal fusion, the neocentromere was inactivated, and the relative stronger native centromere of the fused chromosome serves as the active centromere. Chromosomal fusion occurred in subtelomeric regions and resulted in both telomere and gene loss. Together, these studies illustrate that centromere loss results in neocentromere formation but at elevated temperatures the neocentric chromosome can be fused to another native chromosome, resulting in loss of the neocentromere and formation on an altered karyotypes with a reduction in chromosome number. Events similar to those that we observed in *C. deuterogattii* may underlie the changes in genome organization within the *Cryptococcus* pathogenic species complex and may have contributed to speciation throughout the eukaryotic domain.

Introduction

Eukaryotic organisms have linear chromosomes with specialized regions: telomeres that cap the ends, origins of replication, and centromeres that are critical for chromosome segregation. During cell division, the centromere binds to a specialized protein complex known as the kinetochore [1]. Most centromeres are sequence-independent regional centromeres and are defined by the replacement of the canonical histone H3 by the histone homolog CENPA (CenH3 or CSE4) [2]. Human centromeres contain higher order α-satellite DNA that spans from 0.1 to 4.8 Mb [3], which is in contrast to most fungal centromeres, which contain transposable elements and repetitive sequences [4]. Fungal regional centromeres range from the small centromeres of *Candida albicans*, (the CENPA enriched regions range from 2 to 3.3 kb and are located in 4 to 18 kb open-reading frame [ORF]-free regions), to the large fungal regional centromeres described in *Neurospora crassa*, (which range from 174 to 287 kb and consist mainly of truncated transposable elements [TEs]) [5,6]. Similar to mice, some fungi have pericentric regions [7]. The most prominent examples are the centromeres of *Schizosaccharomyces pombe,* which have a CENPA-enriched region comprised of a central core flanked by heterochromatic pericentric regions divided into outer and inner repeats [8,9].

Infrequently, centromeres can be spontaneously inactivated, resulting in neocentromere formation (i.e., evolutionarily new centromeres) [10]. Neocentromere formation can occur either while the native centromeric sequence is still present on the chromosome or when the native centromere has been mutated or deleted (e.g., from chromosomal rearrangements or γ irradiation damage to the native centromere [10–12]). In addition, several studies have described neocentromere formation after deletion of native centromeres by molecular genetic engineering in fungi, chickens, and *Drosophila* [9,13–15]. In some organisms, the formation of neocentromeres can be deleterious, leading to disease, cancer, or infertility [12,16–19]. For example, human neocentromeres are often identified in liposarcomas [19]. However, neocentromere formation also can be beneficial, leading to speciation [10].

Fungal neocentromeres are well described in the diploid yeast *C. albicans* and the haploid fission yeast *S. pombe* [9,13,20]. Deletion of *C. albicans* native centromere 5 or 7 induced neocentromere formation and did not result in chromosome loss [13,20]. In these cases, the neocentromeres conferred chromosomal stability similar to the native centromere [13,21]. Deletion of a native centromere in *S. pombe* led to two types of isolates that survived via either neocentromere formation or chromosome fusion [8,20]. *S. pombe* neocentromeres formed in telomere-proximal regions near heterochromatin, and neocentromere organization featured a CENPA-enriched core domain and heterochromatin at the subtelomeric (distal) side. Interestingly, neocentromere formation occurred at the same regions in both wild-type and heterochromatin-deficient strains, suggesting that heterochromatin is dispensable for neocentromere formation in *S. pombe*, although the rate of survival by chromosome fusion was significantly increased in heterochromatin-deficient mutants [9].

In some cases, neocentromeres span genes. For example, *C. albicans* neocentromeres contain genes that are silenced due to the neocentromere formation. However the pathways that leads to silencing of neocentromeric genes are unknown in *C. albicans,* as proteins that are necessary for heterochromatin formation and gene silencing in other species (HP1, CLR4, and DNA methyltransferase) are absent in *C. albicans* [13]. As in *C. albicans*, neocentromeres of *S. pombe* can span genes. In wild-type cells, these genes are upregulated during nitrogen starvation and expressed at low levels during stationary growth. When spanned by neocentromeres, these genes were silenced under all conditions tested. In addition to neocentromeric genes, genes located within native centromeres have been identified in other fungi, as well as in centromeres (and neocentromeres) of rice and chicken [15,22,23].

Recently, the centromeres of the human pathogenic fungus *Cryptococcus deuterogattii* were characterized and compared to those of the closely related species *Cryptococcus neoformans,* revealing dramatically smaller centromeres in *C. deuterogattii* [24,25]. *C. deuterogattii* is responsible for an ongoing outbreak in the Pacific Northwest regions of Canada and the United States [26]. In contrast to the sister species *C.*

neoformans, *C. deuterogattii* commonly infects immunocompetent patients [26]. *C. deuterogattii* is a haploid basidiomycetous fungus with 14 chromosomes [24,27,28]. The dramatic reduction in centromere size in *C. deuterogattii* may be attributable to loss of the RNAi pathway and the *DNMT5* gene, which encodes a DNA methyl transferase [24,28]. The centromeres of *C. deuterogattii* consist of truncated transposable elements, and active transposable elements are missing throughout the genome [24]. This is in stark contrast to *C. neoformans,* which has active transposable elements in centromeric regions [24,25,29].

Neocentromeres are frequently formed near genomic repeats, yet *C. deuterogattii* lacks active transposons that might seed neocentromere formation. Thus, *C. deuterogattii* is a unique organism in which to study centromere structure and function. To elucidate centromeric organization, the native centromere of chromosome 10 was deleted, leading to characterization of the first neocentromeres in the *Basidiomycota* phylum of the fungal kingdom.

Results

Deletion of centromere 10 results in neocentromere formation

To determine if neocentromere formation occurs in the *C. deuterogattii* reference strain R265, biolistic transformation was used to replace centromere 10 (*CEN10*) with either the *NAT* or *NEO* dominant drug-resistance gene via homologous recombination. Viable transformants with the correct integration and deletion were obtained and confirmed by 5' junction, 3' junction, loss of deleted regions, and spanning PCRs as well as Southern blot analysis (Figure S1, S2). Multiple independent *cen10*Δ deletion mutants (*cen10*Δ-1 to -7) were obtained from independent biolistic transformations. PFGE confirmed that the *cen10*Δ mutants had a wild-type karyotype and that chromosome 10 remained linear, because a circular chromosome 10 would not have entered the gel (Figure S3).

Chromatin immunoprecipitation of mCherry-CENPA and high-throughput sequencing (ChIP-seq) for three *cen10*Δ mutants (-1, -2 and -3) were performed. Prior to the ChIP-seq experiment, all of the *cen10*Δ mutants were streak purified eight times from single colonies to verify chromosomal stability. Sequence reads were mapped to a complete, whole genome assembly [24]. To quantify the ChIP-seq data, the CENPA-enriched regions were compared with the centromeres previously identified based on CENPC enrichment. Both the CENPA- and CENPC-enriched peaks were congruent for all of the native centromeres (Figure S4) [24]. This analysis identified 13 of 14 native centromeres (*CEN1-9* and *CEN11-14*), indicating that, as expected, chromosome 10 lacked its native centromere in all three *cen10*Δ deletion mutants. Instead, neocentromeres were observed on chromosome 10.

Neocentromeres formed at two unique locations on chromosome 10 (neocen1 and neocen2). Two of the independent mutants (*cen10*Δ-1 and -3) contained two CENPA-enriched regions, suggesting four possible models: 1) aneuploidy in which cells harbor two chromosomes, 2) a dicentric chromosome with two neocentromeres (neodicentric), 3) instability between two different neocentromere states (neocentromere switching), 4) or

only one CENPA enriched region functions as centromere and the second CENPA enriched region is not bound by the kinetochore (Figure 1).

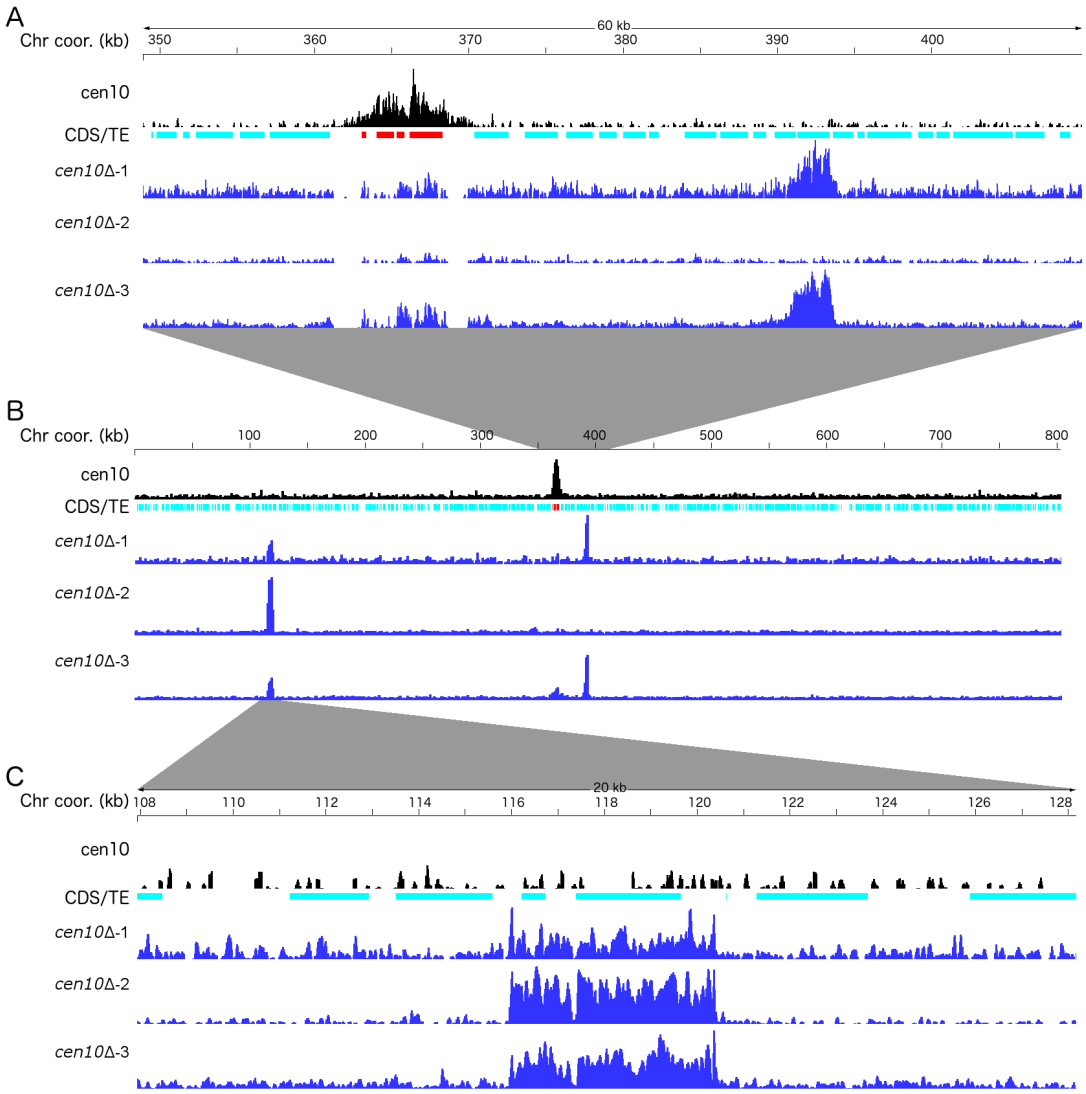


Figure 1. Centromere deletion leads to neocentromere formation.

For each panel, the chromosome coordinates are indicated. The native centromere is shown in black, transposable elements (TE) are shown in red, and genes are shown in light blue. For each *cen10*Δ mutant, the neocentromeric region is shown by enrichment of CENPA (dark blue). Grey regions indicate enlarged regions of chromosome 10. (A) Detailed view of the native centromere and the right neocentromere location (neocen2). Neocen2 is located 21 kb away from the native centromere and is smaller (2.85 kb) than the native centromere (6.77 kb). Neocen2 includes three genes that are bound by CENPA. (B) Chromosome-wide overview of chromosome 10. (C) Detailed view of the left neocentromeric location (neocen1). Neocen1 spans 4.46 kb and is located 242 kb away from the native centromere. All three *cen10*Δ mutants have CENPA enrichment in this region, and two genes are bound by CENPA.

One neocentromere (neocen1) formed at the same location in all three of the *cen10*Δ mutants analyzed. This neocentromere (neocen1) spanned 4.46 kb, was located 242 kb away from the 3' region of native *CEN10*, and was located 115 kb from the telomere. Two genes were enriched for CENPA at neocen1: the first encodes CENPC and the second a hypothetical protein (Table 1). The second neocentromere (neocen2) spanned 2.85 kb and was located closer to the native centromere 10 (21 kb from the native centromere) than neocentromere 1. Neocentromere 2 covered three genes that encodes a serine/threonine-protein phosphatase 2A activator 2 (RRD2) and two hypothetical proteins. In the isolates with two neocentromeres (*cen10*Δ-1 and -3), there was a primary CENPA peak (neocen2) and a secondary peak (neocen1) with reduced levels of CENPA (1.3- to 1.75-fold lower) compared to the primary CENPA peak. Neocentromere 1 was the only CENPA-enriched region in the *cen10*Δ-2 isolate.

Table 1. Genes located inside neocentromeres.

The chromosomal locations, sizes, and GC% for the native centromere, neocen1, and neocen2 are shown. For the neocentromeres, gene ID, predicted function, and the amount of CENPA coverage are indicated.

	*cen10*Δ mutant #	Chr coor (bp)	Size (kb)	GC%	Genes spanned by neocentromere	Gene ID	% covered by Neocentromere	Exons inside neocentromere
Native centromere 10	-	Chr10:362,876-369,657	6.77	42.6	-	-	-	-
Neocentromere 1	1,2,3	Chr10:115,954-120,422	4.46	46.9	*MIF2*	CNBG_4461	88.3	1, 2, 3, 4 (only 5th is outside)
					Hypothetical protein	CNBG_4462	100	
Neocentromere 2	1,3	Chr10:391,090-393,946	2.85	48.9	*RRD2*	CNBG_9459	10.6	last exon (5th)
					Hypothetical protein	CNBG_4366	100	
					Hypothetical protein	CNBG_4365	23.4	last exon (3th)

Previously generated RNA sequencing data was remapped to the *C. deuterogattii* reference strain R265 and analyzed to determine if the regions where neocentromeres formed in the *cen10*Δ mutants are transcribed in the wild-type (Figure S5)(Table S3) [30]. In the wild-type strain, all genes located in neocen1 and neocen2 regions of *cen10*Δ mutants were expressed. The expression levels of the neocentromeric genes in three *cen10*Δ mutants (*cen10*Δ- 1, -2 and -3) were tested by qPCR. The expression levels of the CENPA-associated neocentromeric genes were all found to be similar to the wild-type strain (Figure 2).

The neocentromeres were located in unique, nonrepetitive sequences and were not flanked by repetitive regions. The GC content of both neocentromeres was similar to the overall GC content of chromosome 10 whereas the native centromere 10 has a lower

GC content (Table 1). Comparing the reference genome with *de novo* genome assemblies confirmed that transposable elements did not enter either genomic region during neocentromere formation (Table S4).

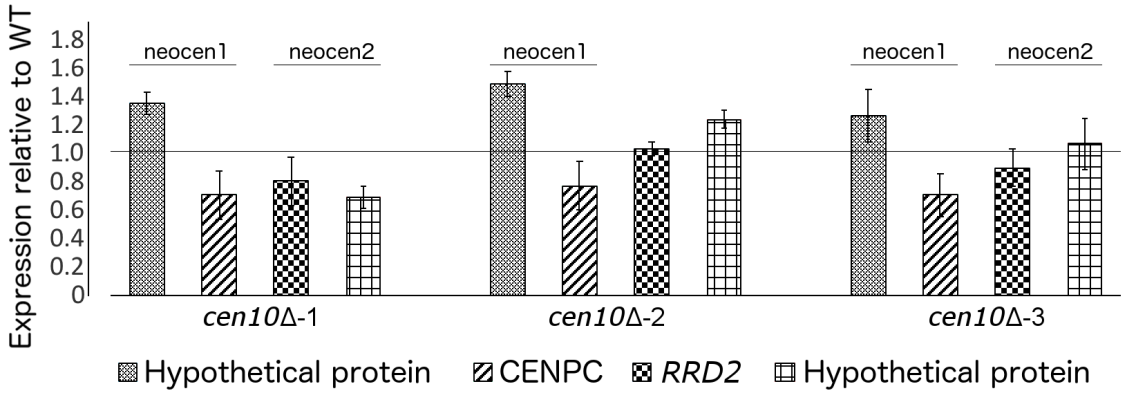


Figure 2. Expression of neocentromeric genes.

Expression of the neocentromeric genes was assessed by qPCR for three *cen10*Δ mutants. Two genes were selected from each neocentromeric region, and expression is shown as $2^{\Delta\Delta Ct}$. Wild-type gene expression levels are indicated with the black line and are normalized to a level of 1. Genes located within neocentromere 1 (neocen1) or neocentromere 2 (neocen2) are indicated. *cen10*Δ-2 has only one CENPA-enriched region (neocen1), and in this case, the genes located within neocen2 served as controls.

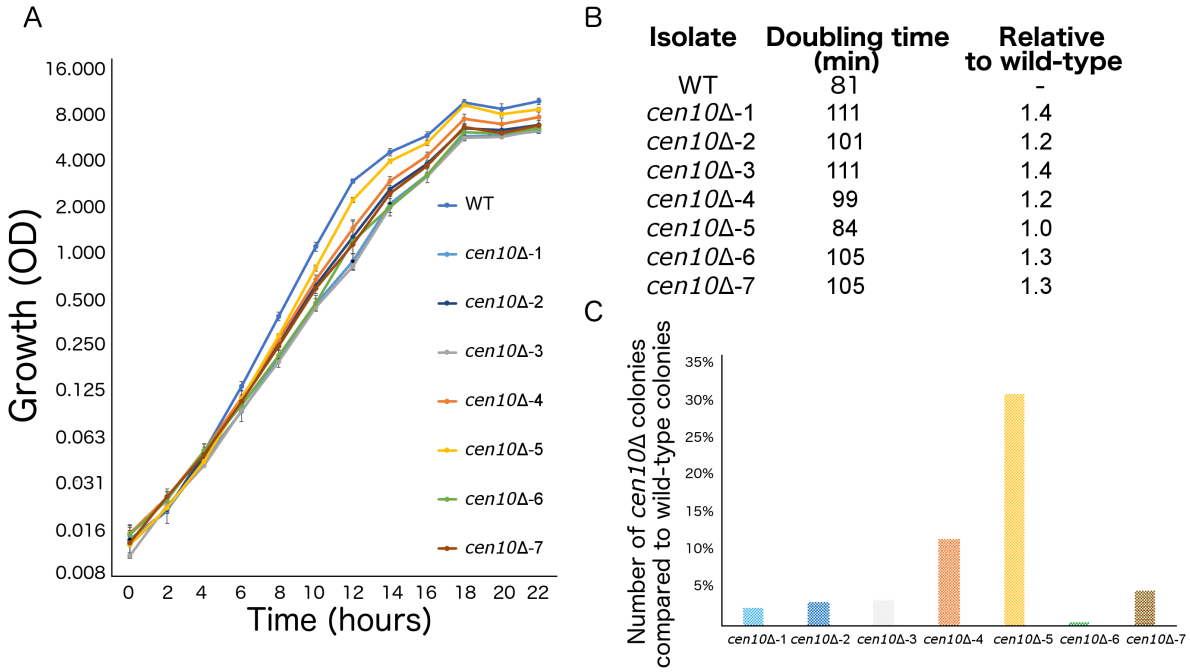
The formation of neocentromeres is an infrequent event. A total of 99 independent biolistic transformations resulted in only seven confirmed *cen10*Δ mutants, suggesting that *CEN10* deletion is lethal in most circumstances. In comparison, deletion of nonessential genes by homologous recombination in the *C. deuterogattii* R265 strain typically results in ~100 colonies with a high success rate (~80-90% correct replacement). We estimate that the likelihood of deleting a centromere and recovering a viable colony is at least 1000-fold lower than would be expected for deletion of a nonessential gene.

Neocentromere formation reduces fitness

The majority of *cen10*Δ mutants exhibited slower growth rates compared to the wild-type parental strain R265. Six of seven *cen10*Δ mutants exhibited significant fitness defects compared to the wild-type strain, with doubling times ranging from 101 to 111 minutes compared to 81 minutes for the wild-type strain (Figure 3). In contrast, one mutant (*cen10*Δ-5) grew similarly to the wild-type and had a similar doubling time (84 min

for the mutant vs 81 min for the wild-type strain). Compared to the wild-type, *cen10*Δ mutants with increased doubling times produced smaller colonies during growth on non-selective media.

To compare fitness, a competition assay was performed with 1:1 mixtures of wild-type and *cen10*Δ mutants grown in liquid YPD medium. With no growth defect, the expectation would be that the wild-type strain and *cen10*Δ mutants would grow at the same growth rate, resulting in a 1:1 ratio. In fact, fewer *cen10*Δ cells were found in the population after growth in competition with the wild-type strain, and this observation is consistent with their slower doubling time resulting in reduced fitness compared to wild-type (Figure 3).



Isolate	Doubling time (min)	Relative to wild-type
WT	81	-
*cen10*Δ-1	111	1.4
*cen10*Δ-2	101	1.2
*cen10*Δ-3	111	1.4
*cen10*Δ-4	99	1.2
*cen10*Δ-5	84	1.0
*cen10*Δ-6	105	1.3
*cen10*Δ-7	105	1.3

Figure 3. *cen10*Δ mutant strains have reduced fitness compared to the wild-type strain.

(A) Six out of seven *cen10*Δ mutants had a longer doubling time and slower growth than the wild-type strain. In contrast *cen10*Δ-5 grows similarly to the wild-type. (B) Doubling times and fold change compared to wild-type are shown. (C) Competition assays with the wild-type and *cen10*Δ mutant strains. Mixed cultures (1:1) were grown overnight and plated with and without selection agents. After four days, the colonies were counted, and the percentage of *cen10*Δ mutants in each culture was plotted.

*cen10*Δ isolates are aneuploid

Because deletion of a centromere could lead to defects in chromosome segregation, *cen10*Δ mutants were assessed for aneuploidy (Figure 4). Overall, *cen10*Δ mutants exhibited a mixture of large and small colony sizes during growth on YPD medium at 37°C. Aneuploidy in *C. neoformans* often leads to a similar phenotype [31]. To exacerbate the aneuploidy-associated slow growth phenotype, four *cen10*Δ mutants were grown at elevated temperature (37°C), causing these isolates to produce smaller, growth-impaired and larger, growth-improved colonies (Figure S6).

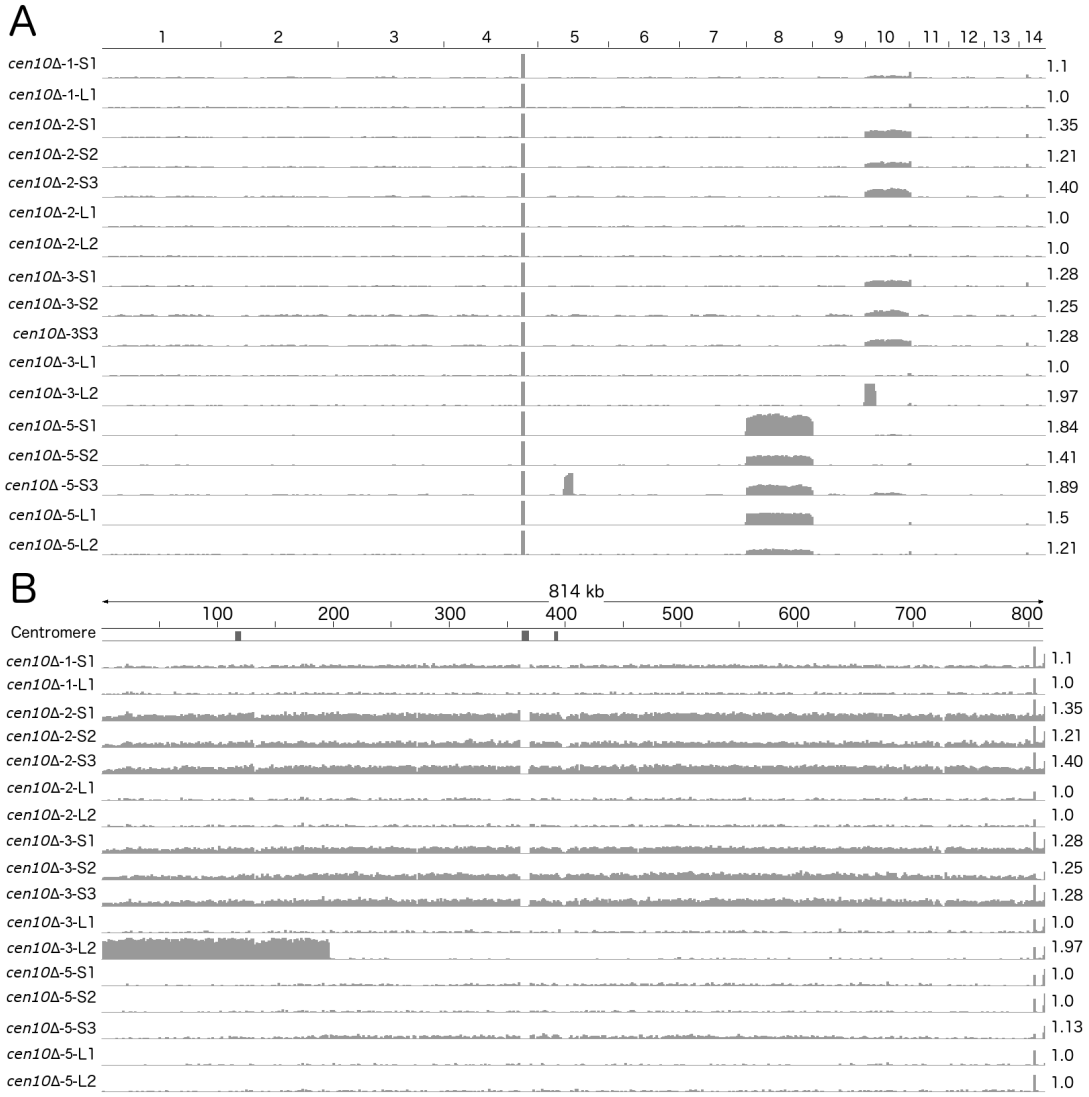


Figure 4. *cen10*Δ mutants are aneuploid.

The whole genomes of small and large colonies derived from four *cen10*Δ mutants were sequenced and read coverage (corresponding to ploidy levels) was plotted. Small colonies of *cen10*Δ mutants were partially aneuploid for

chromosome 10, while the large colonies are euploid. (A) Genome-wide read depth coverage for small and large colonies. On the right, the fold coverage for the highest ploidy level is indicated for each sample. For example, chromosome 10 of *cen10*Δ-2-S1 had an aneuploidy level of 1.35-fold compared to the wild-type strain. Chromosome 4 had a small region with increased read depth due to the ribosomal rDNA gene cluster and was excluded from the analysis. Chromosome 8 of *cen10*Δ-5 was duplicated, in both small and large colonies suggesting that the duplication most likely occurred during, or even before, the transformation. In addition, *cen10*Δ-5-S3 had an additional duplicated region of 162 kb of chromosome 5 that spans the sequence of native centromere 5.
(B) Detailed view of read depth of chromosome 10. As in panel A, read depth is indicated on the right. On the top of panel B, the native centromere 10, neocen1, and neocen2 are each indicated with a grey peak. Due to the deletion of centromere 10, the location of the native centromere lacks sequence reads for each sample.

Three small and two large colonies were selected from each isolate and whole genome analysis was performed based on Illumina sequencing. Sequences were mapped to the reference R265 genome, revealing that the small colonies were indeed aneuploid (Figure 4A). The small colonies of *cen10*Δ-2 and *cen10*Δ-3 had ploidy levels for chromosome 10 in the range of 1.25- to 1.36-fold higher compared to the other 13 chromosomes which suggested that only a proportion (25 to 36%), not all of the cells in a culture, were aneuploid. The remainder of the genome was euploid. Interestingly, chromosome 10 of the small colonies derived from *cen10*Δ-1 and *cen10*Δ-5 exhibited ploidy levels ranging from 1.1- to 1.14-fold. This was similar to, but slightly higher than, the read coverage of the other 13 chromosomes. The genomes of all of the large colonies were euploid, indicating that chromosome 10 of the large colonies was restored to euploid when large colonies were isolated at 37°C.

*cen10*Δ chromosome is rescued by chromosome fusion

Based on whole genome sequence and PFGE analysis, fusion of *cen10*Δ chromosome 10 to other chromosomes was a common event in the large colonies. Whole genome sequence analysis revealed that sequences corresponding to the 3' subtelomeric region of chromosome 10 (including 1 gene) were absent in the sequences obtained from all of the large colonies analyzed (Figure S7). In addition, the large colonies of *cen10*Δ-1 were missing sequences for two genes in the 5' subtelomeric region of chromosome 4. Large colonies of *cen10*Δ-2 were missing 18.5 kb at the 5' subtelomere of chromosome 7

(including eight genes). The large colonies of *cen10*Δ-5 lacked a small part of one gene in the 3' subtelomeric region of chromosome 1. In total, of the 14 subtelomeric genes that were lost in these three chromosome fusion isolates, ten encode hypothetical proteins and four encode proteins with predicted functions. Seven genes have homologs in *C. neoformans* and are present in *C. neoformans* deletion libraries (Table S5). This observation suggests that either subtelomeric deletions occurred, or that chromosomal fusions led to the loss of subtelomeric regions. Notably, sequences from the small colonies spanned the entire genome with no evidence of these subtelomeric deletions.

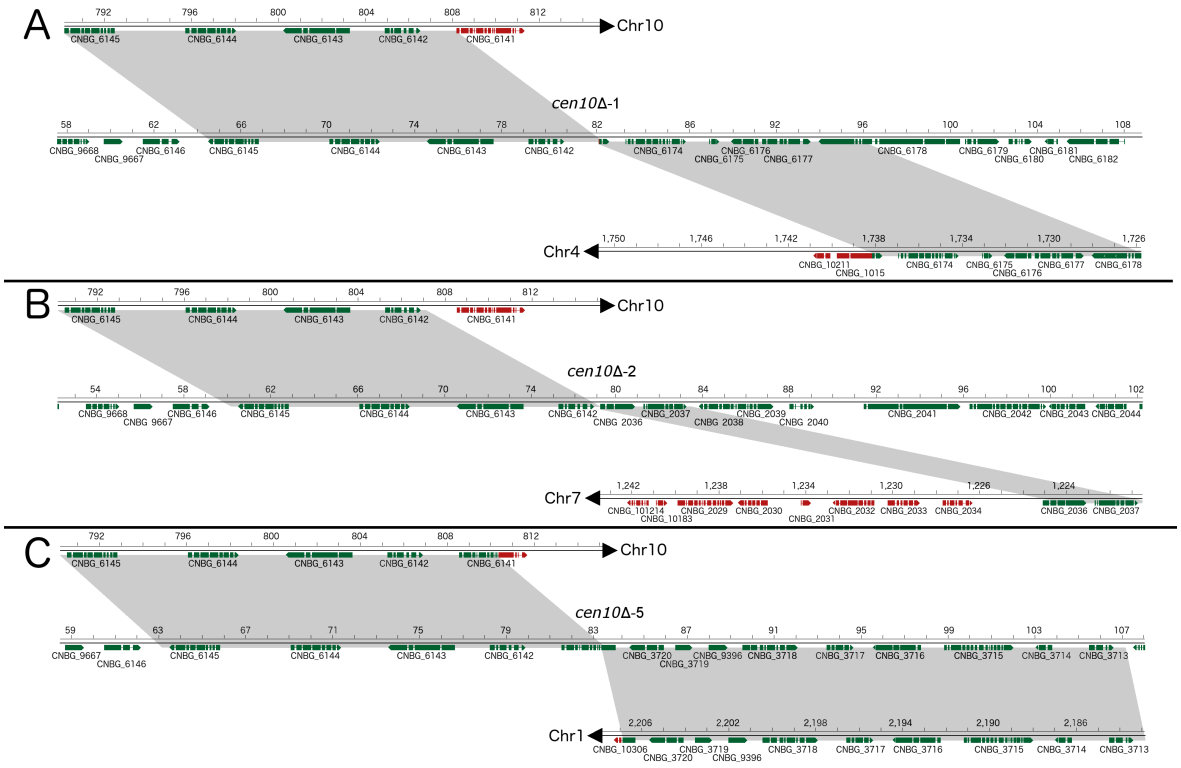


Figure 5. Impaired growth of *cen10*Δ mutants at 37°C is rescued by chromosomal fusion.

Chromosomal fusions were studied in detail for three *cen10*Δ mutants restored to wild-type growth levels at 37°C (large colonies). After chromosome fusion, the fused chromosomes of *cen10*Δ-1-L and *cen10*Δ-2-L lost the gene CNBG_6141, which is located in the 3' subtelomeric region of chromosome 10. Genes present in the fused chromosome are depicted in green, and genes lost after chromosome fusion are indicated in red. Gray highlights indicate regions present in both the parental and fused chromosomes. Each fusion occurred in a unique nonrepetitive region. (A) *cen10*Δ-1-L, the fusion occurred between chromosome 10 and chromosome 4. (B) In *cen10*Δ-2-L, chromosomal fusion occurred between chromosomes 10 and 7. (C) *cen10*Δ-5-L chromosomal fusion occurred between chromosomes 10 and 1.

We hypothesized that the subtelomeric gene loss was due to chromosomal fusion and tested this hypothesis with *de novo* genome assemblies and PFGE (Figure 5 and Figure 6). Based on *de novo* genome assemblies for the large colonies of *cen10*Δ-1, *cen10*Δ-2, and *cen10*Δ-5, chromosome 10 fused with chromosome 1, 4, or 7 respectively (Figure 5 and Table S5). In the large colony of *cen10*Δ-1 (*cen10*Δ-1-L), the fusion occurred between chromosome 10 and chromosome 4. Chromosomal fusion led to the loss of the CNBG_10211 gene (on chromosome 4), and the fusion junction was within the CNBG_6174 gene of chromosome 4. For *cen10*Δ-2-L, the chromosomal fusion occurred between chromosomes 10 and 7. Seven genes of chromosome 7 were lost in the fused chromosome. The chromosome fusion junction was intergenic on chromosome 7. *cen10*Δ-5-L was due to a chromosomal fusion between chromosomes 10 and 1. The fusion was intragenic for both chromosomes. The fusion point occurred in CNBG_6141 on chromosome 10 and CNBG_10308 on chromosome 1.

Because all of the large *cen10*Δ colonies had chromosome 10 fusions, we examined the fusion location on chromosome 10 in detail. The fusions occurred 1.7, 0.3, and 3.6 kb from the chromosome 10 gene CNBG_6142, respectively (Figure 5 and S8). The fusion occurred in unique DNA sequences and was not flanked by repetitive regions. The overlapping region between chromosome 10 and the fused chromosome was at most 6 bp, suggesting that these fusions occurred via microhomology-mediated end joining (MMEJ) (also known as alternative nonhomologous end-joining [Alt-NHEJ]).

Chromosome fusion may result in loss of the neocentromeres, and the kinetochore may bind to the native centromere of the fused chromosome and function as the active centromere. This hypothesis was tested by performing ChIP-qPCR for CENPA binding (Figure 6). For each neocentromere, CENPA enrichment was tested with four primer sets. For all analyzed *cen10*Δ chromosome 10 fusion isolates, the neocentromeres were not CENPA-associated and therefore no longer active (Figure 6). This suggests that the native centromere of the fusion partner of chromosome 10 (i.e., chromosome 1, 4, or 7) is the active centromere of the Chr10-Chr1, Chr10-Chr4, and Chr10-Chr7 fusions.

In addition to *cen10*Δ-1, -2, and -5 mutants, whole genome sequencing was performed for two large colonies of *cen10*Δ-3. Although it was not possible to identify the chromosome fusion based on whole genome sequencing data for either of the large colonies of the *cen10*Δ-3 mutant, PFGE analysis showed that *cen10*Δ-3-L1 had a fusion between chromosomes 10 and 13 (Figure 6b).

*cen10*Δ-3-L2 had read coverage of 1.99-fold for a region of ~200 kb of chromosome 10 (Figure 4). The rest of chromosome 10 was euploid, suggesting that the ~200 kb region is duplicated and is either a single chromosome or fused to another chromosome in this isolate. PFGE analysis suggested that this fragment is duplicated on chromosome 10, resulting in a larger chromosome (Figure 6). In contrast to the other fused chromosomes, this chromosomal fragment did not fuse to a chromosome with a native centromere, and thus the mutant still had a fitness defect. The larger chromosome is euploid, suggesting that the unstable neocentromere(s), rather than causing aneuploidy, resulted in a fitness cost in this isolate.

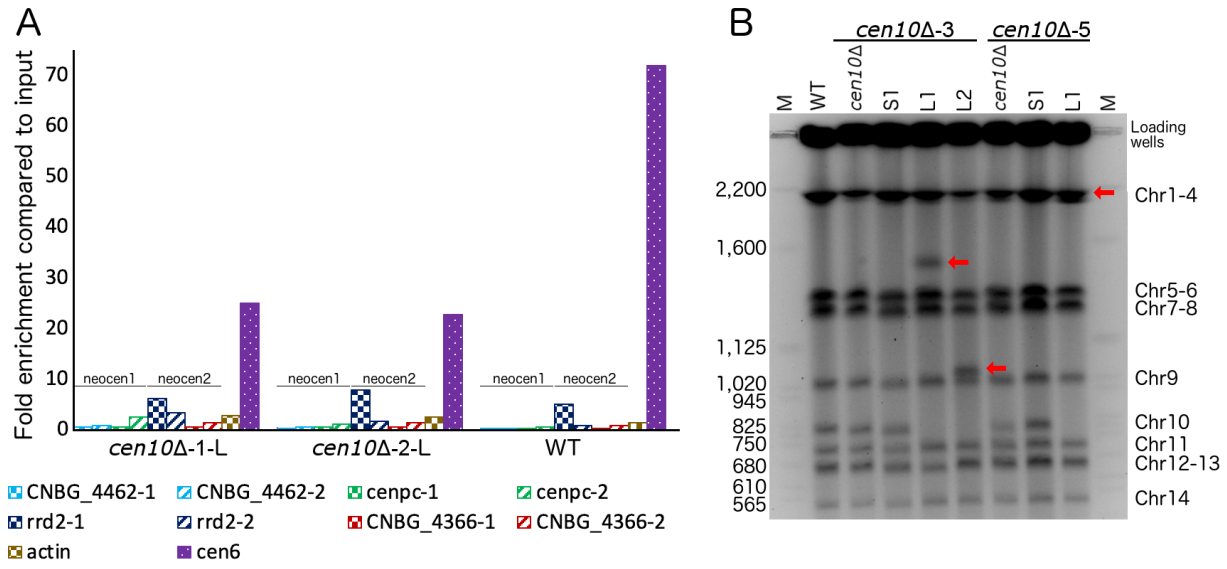


Figure 6. Chromosome fusion results in neocentromere inactivation and karyotype reduction.

(A) Neocentromeres are inactive after chromosomal fusion. For each neocentromeric gene, two ChIP-qPCRs were performed. Data are shown for *cen10*Δ-1-L, *cen10*Δ-2-L, and the wild-type strain. The neocentromeric regions of the large colonies have CENPA coverage similar to wild-type. cen6 was included as a positive control, and actin was

included as a negative control. (B) PFGE analysis shows that the band corresponding to chromosome 10 was in lost the large colonies and instead larger bands appear due to the fusion of chromosome 10 with other chromosomes. *cen10*Δ deletion mutants and small colonies derived from 37°C show a wild-type karyotype. Chromosome 10 of the large colonies was fused to chromosome 13, 10, or 1, respectively. Due to limitations of PFGE conditions the chromosome 10–chromosome 1 fusion did not separate from chromosomes 2, 3, and 4. The positions of the fused chromosomes are indicated with arrows.

Discussion

Composition of neocentromeres in *C. deuterogattii*

The native centromeres of *C. deuterogattii* are found in repetitive regions and are flanked by—but do not contain-protein-encoding genes [24]. By contrast, neocentromeres of *C. deuterogattii* span genes, lack repetitive elements, and like the native centromeres, are flanked by genes. We found that neocen2 was located 21 kb away from the native centromere of chromosome 10 and propose that this neocentromere could have formed due to CENPA seeding; neocen1 was located more distal. The neocentromeres of *C. deuterogattii* are significantly shorter than the native centromeres. Most neocentromere in other species have similar length as the native centromeres.

Native centromeres of *S. pombe* have a central core that is enriched with CENPA and flanked by repetitive pericentric regions [9]. While neocentromere formation in *S. pombe* favors repeats in the pericentric regions, neocentromere formation is possible without the repetitive pericentric regions [9]. The majority of the neocentromeres in *C. albicans* and chicken are formed close to native centromeres due to seeding of CENPA that is located near the native centromere (the so-called CENPA cloud) [13,15]. Although *C. deuterogattii* neocentromeres formed in the same location, there is no consensus between the two regions. Also, there is no similarity to neocentromere formation in other eukaryotes. Our results suggest that neocentromeres form by different mechanisms that do not rely on nearby transposable elements/repeats to initiate de novo centromere assembly.

Neocentromeric genes are expressed

Neocentromeres induced in several species sometimes span genes, resulting in silencing or reduced gene expression. For example, all genes were suppressed within five independent neocentromeres in *C. albicans* that spanned nine genes [32]. In *S. pombe,* neocentromeres span genes that are only expressed in response to nitrogen starvation in the wild-type strain, and neocentromere formation silences these genes during nitrogen starvation [9]. The native centromere 8 of rice contains an approximately 750 kb CENPA-

enriched region with four genes that are expressed in both leaf and root tissues of three closely related species [23,33]. Neocentromeres of rice span genes that are expressed similar to the wild-type [34]. Chicken neocentromeres have been induced on chromosome Z or 5 [15]. Chromosome Z neocentromeres span eight genes, but in wild-type cells only MAMDC2 is expressed during normal growth. The other seven genes were either not expressed at any detectable level in all tested developmental stages or only during early embryonic stages [15]. When a neocentromere formed, expression of the MAMDC2 encoding gene was reduced 20- to 100-fold. Chromosome 5 of chicken is diploid, and neocentromeres on this chromosome span genes that are expressed. The hypothesis is that one allele functions as a centromere, while the other allele codes for the genes.

Here, we showed that the neocentromeres in *C. deuterogattii* span two or three genes and that these genes are expressed at levels similar to the wild-type strain. The *cen10*Δ mutants of *C. deuterogattii* are aneuploid, and it is possible that the genes in the neocentromeres of *C. deuterogattii* are still expressed because only one allele functions as a neocentromere. However, the aneuploid level is at most 1.40-fold higher than background, and the *cen10*Δ-1 mutant is euploid for chromosome 10. The expression of genes enriched for CENPA chromatin is similar to that of wild-type, and if the allelic hypothesis were valid, the expectation would be a reduction by 60% in expression levels.

Of the five *C. deuterogattii* genes spanned by the neocentromere region, two have a predicted function, and one encodes the kinetochore component CENPC. Several independent biolistic transformations were performed to delete CENPC, but all attempts were unsuccessful. This suggests that CENPC is an essential gene and might explain why the gene is still expressed even when bound by CENPA. In addition, introducing mCherry-CENPC by homologous recombination in the *cen10*Δ mutants was not successful, but tagging CENPC in the wild-type strain was effective [24]. This suggests that the chromatin may be changed due to neocentromere formation.

In fission yeast, deletion of the gene encoding the CENPC homolog cnp3 was lethal at 36°C, but mutants were still viable at 30°C [35]. However, CENPA was mislocalized in the *cnp3*Δ mutants. Another gene partially located inside a *C. deuterogattii*

neocentromere encodes the serine/threonine-protein phosphatase 2A activator 2 (*RRD2*). The *RRD2* homolog is not essential in *S. cerevisiae* [36]. The other three neocentromeric genes encode hypothetical proteins and are available as deletion mutants in *C. neoformans* gene deletion mutant libraries [31,37]. Genes contained in regions in which *C. deuterogattii* neocentromeres in *cen10*Δ mutants formed were actively expressed in the wild-type strain, and this is similar to human neocentromeres that can form in regions with or without gene expression [38,39]. Compared with other haploid fungi, the neocentromeric genes of *C. deuterogattii* are similar to the native centromeric genes of the haploid plant pathogenic fungus *Zymoseptoria tritici*. *Z. tritici* has short regional centromeres with an average size of 10.3 kb, and 18 out of 21 native centromeres have a total of 39 expressed genes [22].

_cen10_Δ mutants with two CENPA-enriched regions

In our study, some of the initial *cen10*Δ mutants had two CENPA-enriched regions on chromosome 10, suggesting a putative dicentric chromosome. However, CENPA was not equally distributed between the two CENPA-enriched regions; one peak was more enriched for CENPA (primary neocentromere) compared to the other (secondary neocentromere). The appearance of two CENPA-enriched regions of *C. deuterogattii cen10*Δ mutants could be explained a few ways. First, neocentromere formation could lead to a dicentric chromosome 10 in which the centromeres may differ in functional capacity. Dicentric chromosomes are not per definition unstable, for example the dominant-negative mutation of the mammalian telomere protein TRF2 results in chromosome fusions, leading to the formation of dicentric chromosomes [40]. The formation of dicentric chromosomes occurs in 97% of the fused chromosomes, which are stable for at least 180 cell divisions [40]. Several microscopic studies showed that chromosomes with two regions of centromere protein enrichment are stable [40–43]. This suggests that a dineocentric chromosome 10 could be stable in the population. Second, the two CENPA-enriched peaks could be the result of a mixed population and either due to an unstable primary neocentromere and/or aneuploidy. The primary neocentromere

could be associated with the majority of the cells, whereas the secondary CENPA peak would be only found in a small number of cells (and the primary neocentromere is lost in these isolates). This is reflected by lower CENPA enrichment for the secondary peak, and the hypothesis of putative dicentrics is due to a mixture of alleles in the population. Third, the neocentromeres could be unstable, which could lead to the formation of two CENPA-enriched regions with centromere function switching between the regions. However, our data would argue against this latter model. Prior to the ChIP-seq analysis of the *cen10*Δ mutants, colonies were isolated by streak purification (eight times), suggesting that the presence of two distinct CENPA peaks occurs continuously. Experimental evolution experiments, followed by ChIP-qPCR, could be conducted to test if the primary neocentromere becomes more stable over time.

_cen10_Δ mutants are partially aneuploid

Neocentromere formation in chickens results in a low number of aneuploid cells [15]. Based on whole genome sequencing of a population of cells, the *C. deuterogattii cen10*Δ isolates are partially aneuploid for chromosome 10. For fully aneuploid isolates, the coverage of Illumina reads is expected to be 2-fold; the *cen10*Δ isolates with two CENPA peaks showed aneuploidy levels up to 1.28-fold or were even euploid. This suggests that, like the chicken neocentromeric isolates, only a small number of cells in a population of *C. deuterogattii cen10*Δ isolates are aneuploid.

_cen10_Δ mutants have reduced fitness

In *C. albicans*, deletion of centromere 5 results in neocentromere formation, and these isolates have fitness similar to the wild-type strain [13]. Similar results were reported for neocentromeres in chicken and *S. pombe*, in which strains with neocentromeres or chromosome fusion have a growth rate similar to the wild-type strain [9,15]. In contrast, *C. deuterogattii cen10*Δ mutants have reduced fitness, and competition assays show that *cen10*Δ mutants are less fit compared to the wild-type strain. There was no correlation between reduced fitness and abnormal cell morphology.

Neocentromere stains exhibit impaired growth and at elevated temperature restoration of wild-type growth occurs via chromosome fusion

Deletion of a centromere in *S. pombe* leads to either neocentromere formation or chromosome fusion due to a noncanonical homologous recombination pathway [9,44]. This is in contrast to neocentromere formation in *C. deuterogattii*, which results in 100% neocentromere formation. Based on PFGE analysis, the karyotype of the *cen10*Δ isolates is wild-type at 30°C, and chromosome fusion leads to improved growth at 37°C. The fused chromosomes have no or only short homology at the breakpoints that is insufficient for homologous recombination, suggesting that the chromosome fusions arise via MMEJ. Future experiments to test this hypothesis could involve deleting genes involved in the MMEJ pathway, such as *RAD51* or *RAD52* [45].

A prominent chromosome fusion occurred during the speciation of humans. Compared to other great apes, humans have a reduced karyotype, which is due to the fusion of two ancestral chromosomes that resulted in chromosome 2 in modern humans, Denisovans, and Neanderthals [46]. Human chromosome 2 still harbors signatures of telomeric repeats at the fusion point (interstitial telomeric sequences [ITS]), suggesting that this chromosome is derived from a telomere-telomere fusion. By synteny analysis, the inactive centromere of chimpanzee chromosome 2b can be identified on human chromosome 2, and there are relics of α satellite DNA at this now extinct centromere [46]. Moreover, a dominant negative mutation of the human telomeric protein TRF2 leads to telomere-telomere fusions, mainly between acrocentric chromosomes [40,47]. In the fungal species *Malassezia* chromosome breakage, followed by chromosome fusion has led to speciation[48]. The short regional centromeres (3-5 kb) are fragile and this led most likely to the chromosome reduction. By contrast in *C. deuterogattii,* the chromosomes involved in chromosomal fusion of the *cen10*Δ mutants were all metacentric, and fusion occurred in nontelomeric sequences.

Another example of telomeric fusions is the presence of ITS regions in several genomes. In budding yeast, the experimental introduction of an ITS into an intron of the *URA3* gene resulted in four classes of chromosome rearrangements, including: 1)

inversion, 2) gene conversion, 3) minichromosome formation due to deletion or duplication, and 4) minichromosome formation due to translocation [49]. Based on our *de novo* genome assemblies of the *C. deuterogattii* large-colony *cen10*Δ mutants, only chromosome fusions occurred with no signs of chromosome rearrangements. PFGE analysis of the initial *cen10*Δ mutants and the 37°C-derived large colonies showed that only chromosome 10 is fused to another chromosome and, except for the fused chromosome, all other chromosomes are wild-type. BlastN analysis in the *de novo* genome assemblies of the large colonies confirmed that the subtelomeric regions, which were lost due to chromosomal fusion, were not located on minichromosomes or inserted in other chromosomes. Thus, these chromosome fusions did not produce ITS regions, which would otherwise destabilize the genome.

Recently, two studies described *S. cerevisiae* strains with a dramatically reduced number of chromosomes: from 16 to as few as 1 or 2 chromosomes [50,51]. Chromosome fusions were induced via CRISPR-Cas9-mediated deletion of telomeres and centromeres, and except for a few single-nucleotide polymorphisms and the loss of the centromeres and telomere sequences, the genomic content of the resulting isolates is similar to the wild-type strain [51]. RNA-sequencing revealed that gene expression is similar to wild-type and only 0.5% of the genes were differentially expressed, despite that HiC contact maps showed that the isolate with 1 chromosome had a completely different structural organization in the nucleus.

Conclusions

Our work shows that, like in other model systems, ectopic centromeres can be induced in *C. deuterogattii*. However, *C. deuterogattii* neocentromeres have several unique characteristics, such as including genes whose expression is unaffected by centromere assembly. In some instances, deletion of *CEN10* led to chromosome fusion and this occurs in the *cen10*Δ mutants only at elevated temperature, results in enhanced fitness, and led to inactivation of the neocentromere. Presumably, deletion of other

centromeres could be carried out, leading to a *C. deuterogattii* strain with only one or a few chromosomes, as was recently reported in *S. cerevisiae* [50,51].

Acknowledgements

We thank Tom Petes, Beth Sullivan, Kaustuv Sanyal, Sue Jinks-Robertson, Vikas Yadav, and Shelby Priest for comments on the manuscript. We would like to thank all of the members of the Heitman and Sanyal labs who contributing to the bi-weekly Skype meeting. These studies were supported by NIH/NIAID grants R01 AI050113-15 and R37 Merit award AI039115-21 to JH.

Author contributions

Conceptualization: KS and JH. Formal analysis: KS. Investigation: KS and JH. Resources: JH. Data curation: KS and JH. Writing - original draft: KS. Writing – review & editing: KS and JH. Visualization: KS. Supervision: JH. Project administration: KS and JH. Funding acquisition: JH.

Declaration of Interests

The authors declare no competing interests.

References

1. Cheeseman, I.M. (2014). The kinetochore. Cold Spring Harb. Perspect. Biol. *6*, a015826.
2. Henikoff, S., and Furuyama, T. (2010). Epigenetic inheritance of centromeres. Cold Spring Harb. Symp. Quant. Biol. *75*, 51–60.
3. McNulty, S.M., and Sullivan, B.A. (2018). Alpha satellite DNA biology: finding function in the recesses of the genome. Chromosom. Res. *26*, 115–138.
4. Friedman, S., and Freitag, M. (2017). Centrochromatin of fungi. In Centromeres and Kinetochores, B. E. Black, ed. (Cham, Switzerland: Springer International Publishing), pp. 85–109.
5. Sanyal, K., Baum, M., and Carbon, J. (2004). Centromeric DNA sequences in the pathogenic yeast *Candida albicans* are all different and unique. Proc. Natl. Acad. Sci. U. S. A. *101*, 11374–9.
6. Smith, K.M., Phatale, P.A., Sullivan, C.M., Pomraning, K.R., and Freitag, M. (2011). Heterochromatin is required for normal distribution of *Neurospora crassa* CenH3. Mol. Cell. Biol. *31*, 2528–2542.
7. Guenatri, M., Bailly, D., Maison, C., and Almouzni, G. (2004). Mouse centric and pericentric satellite repeats form distinct functional heterochromatin. J. Cell Biol. *166*, 493–505.
8. Rhind, N., Chen, Z., Yassour, M., Thompson, D.A., Haas, B.J., Habib, N., Wapinski, I., Roy, S., Lin, M.F., Heiman, D.I., *et al.* (2011). Comparative functional genomics of the fission yeasts. Science *930*, 930–936.
9. Ishii, K., Ogiyama, Y., Chikashige, Y., Soejima, S., Masuda, F., Kakuma, T., Hiraoka, Y., and Takahashi, K. (2008). Heterochromatin integrity affects chromosome reorganization after centromere dysfunction. Science *321*, 1088–91.
10. Ventura, M., Antonacci, F., Cardone, M.F., Stanyon, R., D'Addabbo, P., Cellamare, A., Sprague, L.J., Eichler, E.E., Archidiacono, N., and Rocchi, M. (2007). Evolutionary formation of new centromeres in macaque. Science *316*, 243–246.
11. Tolomeo, D., Capozzi, O., Stanyon, R.R., Archidiacono, N., D'Addabbo, P.,

Catacchio, C.R., Purgato, S., Perini, G., Schempp, W., Huddleston, J., *et al.* (2017). Epigenetic origin of evolutionary novel centromeres. Sci. Rep. *7*, 41980.

12. Burrack, L.S., and Berman, J. (2012). Neocentromeres and epigenetically inherited features of centromeres. Chromosom. Res. *20*, 607–619.
13. Ketel, C., Wang, H.S.W., McClellan, M., Bouchonville, K., Selmecki, A., Lahav, T., Gerami-Nejad, M., and Berman, J. (2009). Neocentromeres form efficiently at multiple possible loci in *Candida albicans*. PLoS Genet. *5*, e1000400.
14. Alkan, C., Ventura, M., Archidiacono, N., Rocchi, M., Sahinalp, S.C., and Eichler, E.E. (2007). Organization and evolution of primate centromeric DNA from whole-genome shotgun sequence data. PLoS Comput. Biol. *3*, e181.
15. Shang, W.-H.H., Hori, T., Martins, N.M.C.C., Toyoda, A., Misu, S., Monma, N., Hiratani, I., Maeshima, K., Ikeo, K., Fujiyama, A., *et al.* (2013). Chromosome engineering allows the efficient isolation of vertebrate neocentromeres. Dev. Cell *24*, 635–648.
16. Warburton, P.E. (2004). Chromosomal dynamics of human neocentromere formation. Chromosom. Res. *12*, 617–26.
17. Scott, K.C., and Sullivan, B.A. (2013). Neocentromeres: a place for everything and everything in its place. Trends Genet. *30*, 66–74.
18. Nergadze, S.G., Piras, F.M., Gamba, R., Corbo, M., Cerutti, F., McCarter, J.G.W., Cappelletti, E., Gozzo, F., Harman, R.M., Antczak, D.F., *et al.* (2018). Birth, evolution, and transmission of satellite-free mammalian centromeric domains. Genome Res. *28*, 789–799.
19. Garsed, D.W., Marshall, O.J., Corbin, V.D.A., Hsu, A., DiStefano, L., Schröder, J., Li, J., Feng, Z.P., Kim, B.W., Kowarsky, M., *et al.* (2014). The architecture and evolution of cancer neochromosomes. Cancer Cell *26*, 653–667.
20. Thakur, J., and Sanyal, K. (2013). Efficient neocentromere formation is suppressed by gene conversion to maintain centromere function at native physical chromosomal loci in *Candida albicans*. Genome Res., 638–652.
21. Mishra, P.K., Baum, M., and Carbon, J. (2007). Centromere size and position in

Candida albicans are evolutionarily conserved independent of DNA sequence heterogeneity. Mol. Genet. Genomics *278*, 455–65.

22. Schotanus, K., Soyer, J.L., Connolly, L.R., Grandaubert, J., Happel, P., Smith, K.M., Freitag, M., and Stukenbrock, E.H. (2015). Histone modifications rather than the novel regional centromeres of *Zymoseptoria tritici* distinguish core and accessory chromosomes. Epigenetics Chromatin *8*, 41.
23. Nagaki, K., Cheng, Z., Ouyang, S., Talbert, P.B., Kim, M., Jones, K.M., Henikoff, S., Buell, C.R., and Jiang, J. (2004). Sequencing of a rice centromere uncovers active genes. Nat. Genet. *36*, 138–45.
24. Yadav, V., Sun, S., Billmyre, R.B., Thimmappa, B.C., Shea, T., Lintner, R., Bakkeren, G., Cuomo, C.A., Heitman, J., and Sanyal, K. (2018). RNAi is a critical determinant of centromere evolution in closely related fungi. Proc. Natl. Acad. Sci., 201713725.
25. Janbon, G., Ormerod, K.L., Paulet, D., Byrnes, E.J., Yadav, V., Chatterjee, G., Mullapudi, N., Hon, C.-C., Billmyre, R.B., Brunel, F., *et al.* (2014). Analysis of the genome and transcriptome of *Cryptococcus neoformans var. grubii* reveals complex RNA expression and microevolution leading to virulence attenuation. PLoS Genet. *10*, e1004261.
26. Fraser, J.A., Giles, S.S., Wenink, E.C., Geunes-Boyer, S.G., Wright, J.R., Diezmann, S., Allen, A., Stajich, J.E., Dietrich, F.S., Perfect, J.R., *et al.* (2005). Same-sex mating and the origin of the Vancouver Island *Cryptococcus gattii* outbreak. Nature *437*, 1360–1364.
27. D'Souza, C.A., Kronstad, J.W., Taylor, G., Warren, R., Yuen, M., Hu, G., Jung, W.H., Sham, A., Kidd, S.E., Tangen, K., *et al.* (2011). Genome variation in *Cryptococcus gattii*, an emerging pathogen of immunocompetent hosts. mBio *2*, e00342-10.
28. Farrer, R.A., Desjardins, C.A., Sakthikumar, S., Gujja, S., Saif, S., Zeng, Q., Chen, Y., Voelz, K., Heitman, J., May, R.C., *et al.* (2015). Genome evolution and innovation across the four major lineages of *Cryptococcus gattii*. mBio *6*, e00868-15.
29. Dumesic, P.A., Homer, C.M., Moresco, J.J., Pack, L.R., Shanle, E.K., Coyle, S.M., Strahl, B.D., Fujimori, D.G., Yates, J.R., and Madhani, H.D. (2015). Product binding

enforces the genomic specificity of a yeast polycomb repressive complex. Cell *160*, 204–218.

30. de Oliveira Schneider, R., de Souza Süffert Fogaça, N., Kmetzsch, L., Schrank, A., Vainstein, M.H., and Staats, C.C. (2012). Zap1 regulates zinc homeostasis and modulates virulence in *Cryptococcus gattii*. PLoS One *7*, e43773.
31. Sun, S., Billmyre, R.B., Mieczkowski, P.A., and Heitman, J. (2014). Unisexual reproduction drives meiotic recombination and phenotypic and karyotypic plasticity in *Cryptococcus neoformans*. PLoS Genet. *10*, e1004849.
32. Burrack, L.S., Hutton, H.F., Matter, K.J., Clancey, S.A., Liachko, I., Plemmons, A.E., Saha, A., Power, E.A., Turman, B., Thevandavakkam, M.A., *et al.* (2016). Neocentromeres provide chromosome segregation accuracy and centromere clustering to multiple loci along a *Candida albicans* chromosome. PLoS Genet. *12*, e1006317.
33. Fan, C., Walling, J.G., Zhang, J., Hirsch, C.D., Jiang, J., and Wing, R.A. (2011). Conservation and purifying selection of transcribed genes located in a rice centromere. Plant Cell *23*, 2821–30.
34. Zhang, B., Lv, Z., Pang, J., Liu, Y., Guo, X., Fu, S., Li, J., Dong, Q., Wu, H.-J., Gao, Z., *et al.* (2013). Formation of a functional maize centromere after loss of centromeric sequences and gain of ectopic sequences. Plant Cell *25*, 1979–1989.
35. Suma, M., Kitagawa, T., Nakase, Y., Nakazawa, N., Yanagida, M., and Matsumoto, T. (2018). Fission yeast CENP-C (Cnp3) plays a role in restricting the site of CENP-A accumulation. G3 *8*, 2723–2733.
36. Higgs, H.N., and Peterson, K.J. (2005). Phylogenetic analysis of the formin homology 2 domain. Mol. Biol. Cell *16*, 1–13.
37. Liu, O.W., Chun, C.D., Chow, E.D., Chen, C., Madhani, H.D., and Noble, S.M. (2008). Systematic genetic analysis of virulence in the human fungal pathogen *Cryptococcus neoformans*. Cell *135*, 174–188.
38. Alonso, A., Hasson, D., Cheung, F., and Warburton, P.E. (2010). A paucity of heterochromatin at functional human neocentromeres. Epigenetics and Chromatin

3, 6.

39. Marshall, O.J., Chueh, A.C., Wong, L.H., and Choo, K.H.A. (2008). Neocentromeres: new insights into centromere structure, disease development, and karyotype evolution. Am. J. Hum. Genet. *82*, 261–282.
40. Stimpson, K.M., Song, I.Y., Jauch, A., Holtgreve-Grez, H., Hayden, K.E., Bridger, J.M., and Sullivan, B.A. (2010). Telomere disruption results in non-random formation of de novo dicentric chromosomes involving acrocentric human chromosomes. PLoS Genet. *6*, e1001061.
41. Stimpson, K.M., Matheny, J.E., and Sullivan, B.A. (2012). Dicentric chromosomes: unique models to study centromere function and inactivation. Chromosome Res. *20*, 595–605.
42. Sullivan, B.A., and Willard, H.F. (1998). Stable dicentric X chromosomes with two functional centromeres. Nat. Genet. *20*, 227–228.
43. Higgins, A.W., Gustashaw, K.M., and Willard, H.F. (2005). Engineered human dicentric chromosomes show centromere plasticity. Chromosom. Res. *13*, 745–762.
44. Ohno, Y., Ogiyama, Y., Kubota, Y., Kubo, T., and Ishii, K. (2016). Acentric chromosome ends are prone to fusion with functional chromosome ends through a homology-directed rearrangement. Nucleic Acids Res. *44*, 232–244.
45. Sinha, S., Villarreal, D., Shim, E.Y., and Lee, S.E. (2016). Risky business: Microhomology-mediated end joining. Mutat. Res. Mol. Mech. Mutagen. *788*, 17–24.
46. Miga, K.H. (2017). Chromosome-specific centromere sequences provide an estimate of the ancestral chromosome 2 fusion event in hominin genomes. J. Hered. *108*, 45–52.
47. Van Steensel, B., Smogorzewska, A., and De Lange, T. (1998). TRF2 protects human telomeres from end-to-end fusions. Cell *92*, 401–413.
48. Sankaranarayanan, S.R., Ianiri, G., Reza, H., Thimappa, B.C., Ganguly, P., Coelho, M.D., Sun, S., Siddharthan, R., Tellgren-Roth, C., Dawson Jr, T.L., *et al.* (2019).

Centromere-mediated chromosome break drives karyotype evolution in closely related Malassezia species. Submitted.

49. Aksenova, A.Y., Greenwell, P.W., Dominska, M., Shishkin, A.A., Kim, J.C., Petes, T.D., and Mirkin, S.M. (2013). Genome rearrangements caused by interstitial telomeric sequences in yeast. Proc. Natl. Acad. Sci. U. S. A., 1–6.

50. Shao, Y., Lu, N., Wu, Z., Cai, C., Wang, S., Zhang, L.L., Zhou, F., Xiao, S., Liu, L., Zeng, X., *et al.* (2018). Creating a functional single-chromosome yeast. Nature *560*, 331–335.

51. Luo, J., Sun, X., Cormack, B.P., and Boeke, J.D. (2018). Karyotype engineering by chromosome fusion leads to reproductive isolation in yeast. Nature *560*, 392–396.

52. Davidson, R.C., Blankenship, J.R., Kraus, P.R., De, M., Berrios, J., Hull, C.M., Souza, C.D., Wang, P., and Heitman, J. (2002). A PCR-based strategy to generate integrative targeting alleles with large regions of homology. Microbiology *2*, 2607–2615.

53. Billmyre, R.B., Clancey, S.A., and Heitman, J. (2017). Natural mismatch repair mutations mediate phenotypic diversity and drug resistance in *Cryptococcus deuterogattii*. eLife *6*, e28802.

54. Thorvaldsdóttir, H., Robinson, J.T., and Mesirov, J.P. (2013). Integrative genomics viewer (IGV): high-performance genomics data visualization and exploration. Brief. Bioinform. *14*, 178–192.

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

56. Langmead, B. (2010). Aligning short sequencing reads with Bowtie. Curr Protoc Bioinforma., 11.7.1-11.7.14.

57. Pertea, M., Kim, D., Pertea, G.M., Leek, J.T., and Salzberg, S.L. (2016). Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. Nat. Protoc. *11*, 1650–1667.

58. Bankevich, A., Nurk, S., Antipov, D., Gurevich, A.A., Dvorkin, M., Kulikov, A.S., Lesin, V.M., Nikolenko, S.I., Pham, S., Prjibelski, A.D., *et al.* (2012). SPAdes: A new

genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. *19*, 455–477.

59. Soyer, J.L., Möller, M., Schotanus, K., Connolly, L.R., Galazka, J.M., Freitag, M., and Stukenbrock, E.H. (2015). Chromatin analyses of *Zymoseptoria tritici* : Methods for chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-seq). Fungal Genet. Biol. *79*, 63–70.

60. Findley, K., Sun, S., Fraser, J.A., Hsueh, Y.P., Averette, A.F., Li, W., Dietrich, F.S., and Heitman, J. (2012). Discovery of a modified tetrapolar sexual cycle in *Cryptococcus amylolentus* and the evolution of *MAT* in the *Cryptococcus* species complex. PLoS Genet. *8*, e1002528.

Materials and methods

Strains, primers, and culture conditions

Primers are listed in Supplementary Table S1. Strains used in this study are listed in Supplementary Table S2. All strains were stored in glycerol at -80°C, inoculated on solid media (YPD), and grown for two days at 30°C. Liquid cultures were inoculated from solid media and grown, while shaking, at 30°C in liquid YPD overnight.

Genetic manipulations

DNA sequences (1 to 1.5 kb) of the *CEN10*-flanking regions were PCR-amplified with Phusion High-Fidelity DNA Polymerase (NEB, Ipswich MA, USA). Flanking regions were fused on both sides of either the NEO or NAT selectable marker via overlap PCR, conferring G418 or nourseothricin resistance respectively. Deletion of *C. deuterogattii CEN10* was achieved through homologous recombination via biolistic introduction of an overlap-PCR product as previously described and the transformants were selected on YPD medium containing G418 (200 μg/mL) or nourseothricin (100 μg/mL) [52,53]. Subsequently, the 5' junction, 3' junction, and spanning PCR and Southern blot analyses were performed to confirm the correct replacement of *CEN10* by the appropriate drug resistance marker. To identify centromeres, the gene CNBG_0491, which encodes CENPA, was N-terminally fused to mCherry by overlap PCR, and *C. deuterogattii* strains were biolistically transformed as previously described [53].

Growth and competition assays

Three replicate cultures for seven independent *cen10*Δ deletion mutants and the wild-type strain were grown in liquid YPD at 30°C overnight. Cells were diluted to an OD_{600} of 0.01 and grown in 50 mL YPD at 30°C. The OD_{600} of the triplicate cultures was measured every two hours with a SmartSpec 3000 (BioRad) until stationary phase was reached (T = 22 h).

For competition assays, cultures were grown in 8 mL YPD. Subsequently, the cell density of the cultures was determined using a hemocytometer. For each independent *cen10*Δ deletion mutant, 500,000 cells were co-cultured in a 1:1 ratio with wild-type cells. After 24 hours, the cultures were inoculated on 1) a YPD plate to determine the total colony-forming units (cfu) and 2) a YPD plate containing G418 or nourseothricin to calculate the proportion of *cen10*Δ mutant compared to the wild-type colonies. Plates were incubated at 30°C and the colonies were counted after 4 days. The cell morphology of >1000 cells of the wild-type and of five *cen10*Δ mutant strains was analyzed, and the number of elongated cells was quantified (Figure S9).

Whole genome sequencing, read mapping for aneuploidy/RNA-seq, and *de novo* genome assemblies

Genomic DNA was isolated using the CTAB protocol and sent to the Duke University Sequencing and Genomic Technologies Shared Resource facility for library preparation and Illumina sequencing. Sequencing was performed with a HiSeq 4000 sequencer, and 150 bp paired-end reads were generated. The resulting DNA sequence reads were trimmed, quality-filtered, and subsequently mapped with Bowtie2 to a complete PacBio, Nanopore-based, Illumina Pilon-corrected, whole genome assembly of the *C. deuterogattii* R265 reference genome (version R265_fin_nuclear). Reads were visualized with IGV [24,54–56]. Previously generated RNA sequencing reads (NCBI, SRA: SRR5209627) were remapped to the *C. deuterogattii* R265 reference genome by HISAT2 according to the default settings [30,57].

Genomes were *de novo* assembled with Spades using the default conditions [58]. Most of the large colonies of each independent *cen10*Δ mutant had the same chromosome fusion, suggesting that these colonies were not independent. The fusion regions of the large colonies of the *cen10*Δ-3 mutant were al unique. Genome assemblies were confirmed with PCRs using primers flanking the chromosome fusions and the PCR products obtained span the chromosomal fusions. The read coverage at chromosome fusions was analyzed and compared to the average read coverage of the contig (Figure S8).

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

ChIP analyses were performed as previously described with minor modifications [22,59]. In short, 50 mL YPD cultures were grown overnight at 30°C, after which 37% formaldehyde was added to a final concentration of 0.5%. The cultures were then incubated for 15 minutes, formaldehyde was quenched with 2.5 M glycine (1/20 volume), and cells were washed with cold PBS. Cells were resuspended in chromatin buffer (50 mM HEPES-NaOH, pH 7.5; 20 mM NaCl; 1 mM Na-EDTA, pH 8.0; 1% [v/v] Triton X-100; 0.1% [w/v] sodium deoxycholate [DOC]) containing protease inhibitors (cOmplet™ Tablets, mini EDTA-free EASYpack, Roche), followed by homogenization by bead beating with a miniBead beater (BioSpec products) using 18 cycles of 1.5 min on and 1.5 min off. The supernatant containing chromatin was sheared by sonication (24 cycles of 15 sec on, 15 sec off, burst at high level) (Bioruptor UCD-200, Diagenode). Chromatin was isolated by centrifugation, and the supernatant was divided into a sample fraction and a sonication control. The sample fraction was precleared with protein-A beads (1 to 3 hrs) and subsequently divided into two aliquots. One tube served as the input control, and a mCherry antibody (ab183628, Abcam) was added to the remaining half of the sample. The samples were incubated overnight at 4°C and then processed according to a previously published protocol [59]. After completing the ChIP experiment, the samples were analyzed by ChIP-qPCR or sent to the Duke University Sequencing and Genomic Technologies Shared Resource facility for library preparation and Illumina sequencing. Samples were sequenced with a HiSeq 2500 sequencer, and single reads of 50 bp were obtained. Reads were mapped to the reference genome, similar to the whole genome sequencing reads.

RNA isolation and qPCRs

Cells were grown in an overnight culture of 25 mL YPD at 30°C. RNA was isolated with TRIzol™ LS (Thermo Fisher Scientific) according to the manufacturer's instructions. Subsequently, cDNA was synthesized with the SuperScript™First-Strand Synthesis System (Thermo Fisher Scientific) according to the manufacturer's instructions. qPCRs were

performed in triplicate with Brilliant III Ultra-Fast SYBR® Green qPCR Master Mix (Agilent Technologies) on an ABI 1900HT qPCR machine.

Pulsed-field gel electrophoresis (PFGE)

Isolation of whole chromosomes and conditions for PFGE and chromoblot analysis were performed as previously described [60].

Deposited data

ChIP and whole genome sequencing reads and *de* novo genome assemblies were deposited under NCBI BioProject Accession ID: #####.