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. To investigate impact of centromere loss on the *C. deuterogattii* genome, either centromere 9 or 10 was deleted. Deletion of either centromere resulted in neocentromere formation and interestingly, the genes covered by these neocentromeres maintained wild-type expression levels. In contrast to cen9Δ mutants, cen10Δ mutant strains exhibited growth defects and were aneuploid for chromosome 10. At an elevated growth temperature (37°C), the cen10Δ chromosome was found to have undergone fusion with another native chromosome in some isolates and this fusion restored wild-type growth. Following chromosomal fusion, the neocentromere was inactivated, and the native centromere of the fused chromosome served as the active centromere. The neocentromere formation and chromosomal fusion events observed in this study in *C. deuterogattii* may be similar to events that triggered genomic changes within the *Cryptococcus/Kwoniella* species complex and may contribute 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 (Cheeseman, 2014). Most centromeres are regional, sequence-independent, and defined by the replacement of the canonical histone H3 by the histone homolog CENP-A (CenH3 or Cse4) (Henikoff and Furuyama, 2010). In humans, centromeres contain higher-order α-satellite DNA arrays that span 0.1 to 4.8 Mb (McNulty and Sullivan, 2018), which is in contrast to most fungal centromeres, which contain transposable elements and repetitive sequences (Friedman and Freitag, 2017). Fungal regional centromeres range from the small centromeres of Candida albicans, (the CENP-A 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 regional centromeres described in Neurospora crassa, (which range from 174 to 287 kb and consist mainly of truncated transposable elements) (Sanyal et al., 2004; Smith et al., 2011). Similar to mice, some fungi have pericentric regions (Guenatri et al., 2004). The most prominent examples are the centromeres of Schizosaccharomyces pombe, which have a CENP-A-enriched region comprised of a central core flanked by heterochromatic pericentric regions divided into outer and inner repeats (Ishii et al., 2008; Rhind et al., 2011). Budding yeast have sequence-dependent centromeres, which are short and have a conserved organization with three centromere DNA elements (consensus DNA elements (CDEs) I-III) (Kobayashi et al., 2015). However, the budding yeast Naumovozyma castellii has unique consensus DNA elements that differ from those of other budding yeast species (Kobayashi et al., 2015).

Infrequently, centromeres can be spontaneously inactivated, resulting in neocentromere formation (i.e., evolutionarily new centromeres) (Ventura et al., 2007). 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 (Burrack and Berman, 2012; Tolomeo et al., 2017; Ventura et al., 2007)). In addition, several studies have
described neocentromere formation after deletion of native centromeres by molecular genetic engineering in fungi, chickens, and *Drosophila* (Alkan et al., 2007; Ishii et al., 2008; Ketel et al., 2009; Shang et al., 2013). In some organisms, the formation of neocentromeres can be deleterious, leading to disease, cancer, or infertility (Burrack and Berman, 2012; Garsed et al., 2014; Nergadze et al., 2018; Scott and Sullivan, 2013; Warburton, 2004). For example, human neocentromeres are often identified in liposarcomas (Garsed et al., 2014). However, neocentromere formation also can be beneficial, leading to speciation (Ventura et al., 2007).

Fungal neocentromeres are well described in the diploid yeast *C. albicans* and the haploid fission yeast *S. pombe* (Ishii et al., 2008; Ketel et al., 2009; Thakur and Sanyal, 2013). Deletion of *C. albicans* native centromere 5 or 7 has been shown to induce neocentromere formation and does not result in chromosome loss (Ketel et al., 2009; Thakur and Sanyal, 2013). In these cases, neocentromeres conferred chromosomal stability similar to the native centromere (Ketel et al., 2009; Mishra et al., 2007). Deletion of a native centromere in *S. pombe* led to either neocentromere formation or chromosome fusion (Ishii et al., 2008; Ohno et al., 2016). *S. pombe* neocentromeres formed in telomere-proximal regions near heterochromatin, and neocentromere organization featured a CENP-A-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 (Ishii et al., 2008). Deletion of kinetochore proteins (*mhf1Δ* and *mhf2Δ*) led to a shift of CENP-A binding, resulting in a CENP-A-enriched region directly adjacent to the native centromere (Lu and He, 2019).

In some cases, neocentromeres span genes that are silenced, such as the neocentromeres in *C. albicans*. However the mechanisms that mediate 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* (Ketel et al., 2009). Neocentromeres of *S. pombe* can also span genes. These genes are upregulated during nitrogen starvation and expressed at low levels during stationary growth in wild-type cells, but are silenced under all conditions tested when spanned by neocentromeres. In addition to neocentromeric genes, genes located within native centromeres have been identified in other fungi as well as rice and chicken (Nagaki et al., 2004; Schotanus et al., 2015; Shang et al., 2013).

Recently, the centromeres of the human pathogenic fungus *Cryptococcus deuterogattii* were characterized and compared to those of the closely related species *Cryptococcus neoformans* (centromeres ranging from 27 to 64 kb), revealing dramatically smaller centromeres in *C. deuterogattii* (ranging from 8.7 to 21 kb) (Janbon et al., 2014; Yadav et al., 2018). *C. deuterogattii* is responsible for an ongoing outbreak in the Pacific Northwest regions of Canada and the United States (Fraser et al., 2005). In contrast to the sister species *C. neoformans*, *C. deuterogattii* commonly infects immunocompetent patients (Fraser et al., 2005). *C. deuterogattii* is a haploid basidiomycetous fungus with 14 chromosomes (D’Souza et al., 2011; Farrer et al., 2015; Yadav et al., 2018). The dramatic reduction in centromere size in *C. deuterogattii* may be attributable to loss of the RNAi pathway (Farrer et al., 2015; Yadav et al., 2018). The centromeres of *C. deuterogattii* consist of truncated transposable elements, and active transposable elements are missing throughout the genome (Yadav et al., 2018). This is in stark contrast to *C. neoformans*, which has active transposable elements in centromeric regions (Dumesic et al., 2015; Janbon et al., 2014; Yadav et al., 2018).

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 centromeres of chromosomes 10 and 9 were deleted, leading to characterization of the first neocentromeres in the *Basidiomycota* phylum of the fungal kingdom.
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 YPD (yeast extract, peptone, and dextrose) media, and grown for two days at 30°C. Liquid YPD cultures were inoculated from single colonies of solid media and grown, while shaking, at 30°C 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 dominant 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 (Billmyre et al., 2017; Davidson et al., 2002). Deletion of CEN9 was performed by CRISPR-CAS9 mediated transformation with two guide RNAs flanking CEN9 and homologous recombination was mediated by the introduction of an overlap PCR product as previously described (Fan and Lin, 2018). Transformants were selected on YPD medium containing G418 (200 µg/mL) or nourseothricin (100 µg/mL).

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 CENP-A, was N-terminally fused to the gene encoding the fluorescent mCherry protein by overlap PCR, and C. deuterogattii strains were biolistically transformed as previously described (Billmyre et al., 2017). A subset of cen9Δ mutants were biolistically transformed with an overlap PCR product containing CENPC C-terminally fused with GFP. As three cen10Δ mutants have a neocentromere that spans the gene encoding CENPC, a subset of cen10Δ mutants were transformed instead with an overlap PCR product containing MIS12 C-terminally fused with GFP. Both PCR products encoding CENPC-GFP and MIS12-GFP were
randomly integrated in the genome and confirmed by a PCR spanning either CENPC-GFP or MIS12-GFP.

**Growth and competition assays**

Three replicate cultures for seven independent cen10Δ 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, three independent replicate cultures (cen9Δ, cen10Δ, control, and wild type) were grown overnight in 8 mL YPD. Subsequently, the cell density of the cultures was determined using a hemocytometer. For each independent cenΔ 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 (CFUs) and 2) a YPD plate containing G418 or nourseothricin to calculate the proportion of cen10Δ mutant CFUs compared to the wild-type CFUs. 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 cen10Δ 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 (Langmead,
Previously generated RNA sequencing reads (NCBI, SRA: SRR5209627) were remapped to the *C. deuterogattii* R265 reference genome by HISAT2 according to the default settings (de Oliveira Schneider et al., 2012; Pertea et al., 2016). Genomes were *de novo* assembled with Spades using the default conditions (Bankevich et al., 2012). Genome assemblies were confirmed with PCRs using primers flanking the chromosome fusions and the PCR products obtained span the chromosomal fusions (Figure S7). The read coverage at chromosome fusions was analyzed and compared to the average read coverage of the contig (Figure S7).

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

ChIP analyses were performed as previously described with minor modifications (Schotanus et al., 2015; Soyer et al., 2015). In short, 500 mL YPD cultures (1000 ml YPD for Mis12 ChIPs) were grown overnight at 30°C, after which 37% formaldehyde was added to a final concentration of 0.5% for crosslinking. 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. The crosslinking time of Mis12-GFP tagged isolates was extended to 45 minutes. 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 (cOmplett™ 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 (Soyer et al., 2015). 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 cen10Δ-A, cen10Δ-B and cen10Δ-C were sequenced with a HiSeq 2500 sequencer, and single reads of 50 bp were obtained. All other ChIP-seq samples were sequenced with a NovaSeq 600 sequencer and 50 bp PE reads were obtained. For each centromere mutant and the wild type, a ChIPed and input sample were sequenced. Reads were mapped to the reference genome, similar to the whole-genome sequencing reads. To analyze the ChIP-seq data the ChIPed sample was normalized with the input sample and visualized with the IGV viewer.

**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 (Findley et al., 2012).

**Deposited data**

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

Deletion of centromere 9 and 10 results in neocentromere formation

To determine if neocentromere formation occurs in the *C. deuterogattii* reference strain R265, either centromere 9 or 10 was deleted. Biolistic transformation was used to replace centromere 10 (*CEN10*) with either the *NAT* or *NEO* dominant drug-resistance gene via homologous recombination. Centromere 9 (*CEN9*) was deleted by CRISPR-Cas9-mediated transformation. Two guide RNAs flanking the centromere were used and *CEN9* was replaced with a *NAT* dominant drug-resistance gene by 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 for *cen10Δ* (Figure S1, S2). Multiple independent *cen9Δ* and *cen10Δ* deletion mutants (*cen9Δ*-A to -F and *cen10Δ*-A to -G) were obtained from independent transformations. Pulsed-field gel electrophoresis (PFGE) confirmed that *cenΔ* mutants had a wild-type karyotype and that chromosome 9 and 10 remained linear, because a circular chromosome would not have entered the gel (Figure S3).

The formation of neocentromeres in *C. deuterogattii* was infrequent. A total of 99 independent biolistic transformations resulted in only seven confirmed *cen10Δ* mutants (7/21 total candidate transformants, 33% homologous integration), 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% homologous integration). We estimate that the likelihood of deleting a centromere and recovering a viable colony is at least 1000-fold lower than would be expected from the deletion of a non-essential gene. The deletion of *CEN9* was more efficient as this was mediated by CRISPR-Cas9 cleavage with two guide RNAs and a repair allele.

Chromatin immunoprecipitation of mCherry-CENP-A followed by high-throughput sequencing (ChIP-seq) for six *cen9Δ* (-A to -F) and seven *cen10Δ* mutants (-A to -G) was performed. Prior to the ChIP-seq experiment, all of the centromere deletion mutants were streak purified from single colonies. The sequence reads were mapped to a complete
whole-genome assembly, followed by the normalization of the reads by subtraction of the input from the ChIPed sample (Yadav et al., 2018). To quantify the ChIP-seq data, the CENP-A-enriched regions were compared with the centromeres previously identified based on CENPC enrichment. Both the CENP-A- and CENPC-enriched peaks were congruent for all of the native centromeres (Yadav et al., 2018). This analysis identified 13 of the 14 native centromeres (CEN1-8, CEN11-14 and depending on the centromere mutant either CEN9 or CEN10), indicating that, as expected, the native centromere of chromosome 9 or 10 was missing in all of the cen9Δ and cen10Δ deletion mutants respectively (Figure 1). Instead, neocentromeres were observed.

Except for the neocentromere of isolate cen10Δ-E, the neocentromeres formed in close proximity to the native centromere (CEN9 and CEN10). Almost all neocentromeres were shorter than the native centromere, with the exception of cen10Δ-G which was larger than native centromere 10.

In three of the independent cen9Δ mutants (cen9Δ-B, -C and -E), neocentromeres formed at the same chromosomal location. Interestingly, two independent cen10Δ mutants (cen10Δ-A and cen10Δ-C) contained two CENP-A-enriched regions on chromosome 10, with a primary peak and a smaller secondary peak with reduced levels of CENP-A (1.3- to 1.75-fold lower) compared to the primary CENP-A peak. The chromosomal location of the secondary peak was similar to the neocentromere of cen10Δ-B (which had only one neocentromere).

The two CENP-A-enriched regions suggest 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 CENP-A-enriched region functions as a centromere and the second CENP-A-enriched region is not bound by the kinetochore (Figure 1).

The neocentromeres were located in unique, nonrepetitive sequences and were not flanked by repetitive regions. The GC content of neocentromeres is similar to the overall GC content of chromosome 9 and 10, whereas the native centromere has a lower
Comparing the reference genome with de novo genome assemblies of cen10Δ-A, cen10Δ-B, and cen10Δ-E confirmed that transposable elements did not enter these genomic regions during neocentromere formation (Table S4). Instead of spanning repeats and transposable elements like the native centromeres, neocentromeres span genes. The majority of the genes (24/28) flanking native centromeres are transcribed in the direction towards the native centromere. All of the neocentromeres observed span one or more genes and most of the flanking genes are transcribed in the direction away from the neocentromere.

Several of the genes spanned by neocentromeres are hypothetical genes, and interestingly, the kinetochore protein CENPC was located inside the neocentromere of cen10Δ-B and in the secondary peak of cen10Δ-A and -C (Table 1). The neocentromere in cen10Δ-B spanned 4.46 kb, was located 242 kb away from the 3’ region of the native CEN10, and was located 115 kb from the telomere. In addition to the gene encoding CENPC, the CENP-A-enriched region spanned a hypothetical protein (Table 1). The primary CENP-A-enriched region of cen10Δ-A and cen10Δ-C spanned a gene encoding a serine/threonine-protein phosphatase 2A activator 2 (RRD2) and a hypothetical protein. This neocentromere spanned 2.85 kb and was located closer to the native CEN10 (21 kb from the native centromere) than the neocentromere of cen10Δ-B and the secondary CENP-A peak of cen10Δ-A and cen10Δ-C. The neocentromere of cen10Δ-D was the smallest neocentromere (2.5 kb) and partially (88.4%) spanned a gene encoding a Ser/Thr protein kinase and formed 7.4 kb from the location of the native CEN10. The neocentromere of cen10Δ-E spanned two hypothetical proteins, was 4.38 kb in length and was located directly adjacent to the right telomere. Mutant cen10Δ-F had a neocentromere of 2.64 kb, which spanned one hypothetical gene completely and two genes (hypothetical and a hexokinase (HXK1)) partially; the neocentromere formed at a chromosomal location 20 kb 5’ of the native centromere. The neocentromere of cen10Δ-G was the largest neocentromere with a CENP-A-enriched region of 7.97 kb, and was in fact larger than the native CEN10. This neocentromere spanned four genes, including a
gene coding for a high osmolarity protein (Sho1) and three genes coding for hypothetical proteins.

Like the neocentromeres of cen10Δ mutants, the neocentromeres of cen9Δ mutants also spanned genes. All of the neocentromeres formed in a region within 26 kb of the chromosomal location of the native centromere 9. Interestingly, the neocentromeres of cen9Δ-B, cen9Δ-C and cen9Δ-E all formed at the same chromosomal location and had the same length (4.41 kb); these neocentromeres spanned three genes. One gene was completely covered by CENP-A and this gene encoded a transglycosylase SLT domain-containing protein. The two other genes (a gene encoding a xylosylphosphotransferase and a gene encoding glutamate synthase (NADPH/NADH)) were partially covered with CENP-A. Mutant cen9Δ-A had a 3.87-kb long neocentromere located 26 kb 3’ to the native centromere and spanned two genes. The first gene was completely spanned by CENP-A and encoded an ESCRT-II complex subunit (Vps25) protein. The second gene was only partially covered and encoded an iron regulator 1 protein. The neocentromere of cen9Δ-D was located directly to the left of the native centromere and was 4.37 kb in length. This neocentromere spanned two genes, coding for a hypothetical protein (92% covered by CENP-A) and a gene encoding for Derlin-2/3 that was completely covered by CENP-A. Lastly, the neocentromere of mutant cen9Δ-F was 3.83 kb in length and spanned one gene (encoding a xylosylphosphotransferase, Xpt1), which was completely covered by CENP-A. This neocentromere was located 12 kb away (3’) from the native centromere 9.

To test if the kinetochore was binding to the CENP-A-enriched regions of chromosomes 9 and 10, and to validate if the neocentromeres were fully functional as centromeres, two additional kinetochore proteins were epitope-tagged with GFP. As the neocentromeres of three cen10Δ mutants spanned the gene encoding CENP-C, all cen10Δ mutants were transformed with an overlap PCR product, expressing Mis12-GFP. cen9Δ mutants were transformed with an overlap PCR product expressing CENPC-GFP. In addition to the cen9Δ and cen10Δ mutants, the wild type was transformed with constructs expressing Mis12-GFP and CENP-C-GFP, and these served as controls. ChIP-qPCRs for
cen9Δ mutants, cen10Δ mutants, and wild-type strains with Mis12-GFP or CENP-C-GFP were performed. Because Mis12 is an outer kinetochore protein, the formaldehyde cross-linking was extended to 45 minutes (15 minutes was used for CENP-A and CENP-C) for this protein. For all qPCR analyses, the native centromere 6 was used as an internal control and for each neocentromere specific primer pairs were designed. For cen9Δ and cen10Δ mutants a similar level of Mis12 or CENP-C enrichment at the neocentromeres and native centromere 6 was observed. This suggested that the CENP-A-enriched regions of chromosome 9 of the cen9Δ mutants and chromosome 10 of cen10Δ mutants identified by ChIP-seq were functional centromeres and indeed neocentromeres (Figure S4).

Previously generated RNA sequence data was remapped to the C. deuterogattii reference strain R265 and analyzed to determine if the regions where neocentromeres formed in the cenΔ mutants were transcribed in the wild type (Figure 1)(Table S3) (de Oliveira Schneider et al., 2012). In the wild-type strain, all genes spanned by neocentromeres in the cenΔ mutants were expressed. The expression levels of the neocentromeric genes in cen9Δ and cen10Δ mutants were assayed by qPCR. The neocentromeric genes of chromosome 9 were normalized to actin. To compensate for the ploidy levels of chromosome 10 in cen10Δ mutants, a housekeeping gene located on chromosome 10 was used to normalize the expression of genes spanned by neocentromeres located on chromosome 10. The expression levels of the CENP-A-associated neocentromeric genes were all found to be similar to the wild-type strain (Figure 2).

**Neocentromere formation can reduce fitness**

cen10Δ mutants were noted to grow more slowly than wild type. To investigate this, the growth of cen10Δ and wild-type strains was measured during the course of a 22-hour cell growth experiment. The majority of cen10Δ mutants exhibited slower growth rates compared to the wild-type parental strain R265. Six of seven cen10Δ 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 (Figure 3).
contrast, one mutant (*cen10Δ*-E) 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, *cen10Δ* 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 *cen9Δ* or *cen10Δ* mutants grown in liquid YPD medium. With no growth defect, the expectation was that the wild-type strain and centromere deletion mutants would grow at the same growth rate, resulting in a 1:1 ratio. In fact, fewer *cen10Δ* cells were found in the population after growth in competition with the wild-type strain, and this observation is consistent with the slower doubling time of *cen10Δ* mutants resulting in reduced fitness compared to wild type (Figure 3). Compared to the wild-type cells, there were fewer *cen9Δ* mutant cells in the population. However, the number was closer to a 1:1 ratio. The ratio of the *cen9Δ* mutants in the population was similar to the ratio of the *cen10Δ*-E mutant, which had a wild-type growth rate. Due to this observation, we hypothesize that the growth rate of the *cen9Δ* mutants is similar to wild type.

**cen10Δ isolates are aneuploid**

Because deletion of a centromere could lead to defects in chromosome segregation, *cenΔ* mutants were assessed for aneuploidy (Figure 4). Overall, *cen10Δ* 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 mixed colony size phenotype as that observed in the *cen10Δ* mutants (Sun et al., 2014). To exacerbate the aneuploidy-associated slow growth phenotype, four *cen10Δ* mutants were grown at elevated temperature (37°C), causing these isolates to produce smaller, growth-impaired and larger, growth-improved colonies (Figure S5). 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 *cen10Δ*-B and
*cen10Δ-C* 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 of the cells (25 to 36%), were aneuploid. The remainder of the genome was euploid. Chromosome 10 of the small colonies derived from isolate *cen10Δ-A* and *cen10Δ-E* exhibited ploidy levels ranging from 1.1- to 1.14-fold, reflecting less aneuploidy. Importantly, for all of the large colonies derived from isolates *cen10Δ-A*, *cen10Δ-B*, *cen10Δ-C*, and *cen10Δ-E* the fold coverage of chromosome 10 was restored to the wild-type euploid level (1.0 fold compared to wild type). The ploidy levels of chromosome 10 were 1-fold for all of the large colonies compared to wild type, indicating that the ploidy level of chromosome 10 of the large colonies was restored to euploid.

**cen10Δ chromosome is rescued by chromosome fusion**

Based on whole-genome sequencing and PFGE analysis, fusion of *cen10Δ* 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 S6). In addition, the large colonies of *cen10Δ-A* were missing sequences for two genes in the 5′ subtelomeric region of chromosome 4. Large colonies of *cen10Δ-B* were missing 18.5 kb at the 5′ subtelomere of chromosome 7 (including eight genes). The large colonies of *cen10Δ-E* 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 encoded hypothetical proteins and four encoded proteins with predicted functions. Seven genes have homologs in *C. neoformans* and are present in *C. neoformans* deletion libraries (Table S5). This observation suggested 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 S6).

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
Based on *de novo* genome assemblies for the large colonies of *cen10Δ-A*, *cen10Δ-B*, and *cen10Δ-E*, chromosome 10 fused with chromosome 4, 7, or 1, respectively (Figure 5 and Table S5). In the large colony of *cen10Δ-A* (*cen10Δ-A-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 *cen10Δ-B-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. *cen10Δ-E-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 *cen10Δ* 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 CENP-A binding (Figure 6). For each neocentromere (either of the *cen10Δ-A* or *cen10Δ-B* mutant), CENP-A enrichment was tested with four primer sets located in the neocentromere. The CENP-A enrichment for these four locations was tested in 1) the initial *cen10Δ* mutant 2) a large colony derived from a specific *cen10Δ* mutant and 3) wild type. As expected, the ChIP-qPCR analysis showed CENP-A enrichment for the neocentromeres of the initial *cen10Δ-A* and *cen10Δ-B* mutants. The neocentromeric regions of *cen10Δ-A* and *cen10Δ-B* were not enriched with CENP-A in the wild-type strain, showing there was no occupancy by CENP-A prior to neocentromere formation. For all analyzed *cen10Δ* chromosome 10 fusion
isolates, the neocentromeres were not CENP-A-associated, and were similar to the wild-type background levels. Therefore, the neocentromeres were no longer active in the chromosome fusion strains(Figure 6). This suggests that the native centromere of the fusion partner of chromosome 10 (i.e., chromosome 1, 4, or 7) was the active centromere of the Chr10-Chr1, Chr10-Chr4, and Chr10-Chr7 fusions.

In addition to cen10Δ-A, -B, and -E mutants, whole-genome sequencing was performed for two large colonies of cen10Δ-C. Although it was not possible to identify the chromosome fusion based on whole-genome sequencing data for either of the large colonies of the cen10Δ-C mutant, PFGE analysis showed that cen10Δ-C-L1 had a fusion between chromosomes 10 and 13 (Figure 6b).

cen10Δ-C-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 was duplicated and was either a single chromosome or fused to another chromosome in this isolate. PFGE analysis suggested that this fragment was 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 the fact that the mutant still exhibited a fitness defect was consistent with this interpretation. The larger chromosome was euploid, suggesting that the unstable neocentromere(s), rather than causing aneuploidy, resulted in a fitness cost in this isolate.
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 (Yadav et al., 2018). By contrast, neocentromeres of *C. deuterogattii* span genes, lack repetitive elements, and like the native centromeres, are flanked by genes. In general (with one exception), the neocentromeres of *C. deuterogattii* are significantly shorter than the native centromeres, whereas most neocentromeres in other species have similar lengths as the native centromeres.

Native centromeres of *S. pombe* have a central core that is enriched with CENP-A and flanked by repetitive pericentric regions (Ishii et al., 2008). While neocentromere formation in *S. pombe* favors repeats in the pericentric regions, neocentromere formation is possible without the repetitive pericentric regions (Ishii et al., 2008). The majority of the neocentromeres in *C. albicans* and chickens are formed close to native centromeres due to seeding of CENP-A that is located near the native centromere (the so-called CENP-A cloud) (Ketel et al., 2009; Shang et al., 2013). The neocentromeres of *C. deuterogattii* follow the same trend and the majority of the neocentromeres formed close to the location of the native centromere. Our results and the earlier reports discussed, suggest that the chromosomal location of the native centromere is the main determinant of neocentromere formation. One exception was the neocentromere of *cen10*Δ-E, which directly flanked the left telomere. Interestingly, this was the only *cen10*Δ mutant that had a growth rate similar to wild type.

Several *C. deuterogattii* neocentromeres formed in the same location and a similar trend has been observed in neocentromere formation in *C. albicans* (Ketel et al., 2009). Evolutionary new centromeres (ECNs) in the largest crucifer tribe Arabideae originated several times independently and are located in the same chromosomal location (Mandáková et al., 2020). Although *C. deuterogattii* neocentromeres have the same chromosomal location, there is no apparent consensus between the different regions occupied by different neocentromeres. 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 can span genes, resulting in silencing or reduced gene expression. For example, all genes within five independent neocentromeres in *C. albicans* that spanned nine genes were suppressed (Burrack et al., 2016). 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 (Ishii et al., 2008). The native centromere 8 of rice contains an approximately 750-kb CENP-A-enriched region with four genes that are expressed in both leaf and root tissues of three closely related species (Fan et al., 2011; Nagaki et al., 2004). Neocentromeres of rice span genes that are expressed at similar levels as in the wild type (Zhang et al., 2013). Chicken neocentromeres have been induced on chromosome Z or 5 (Shang et al., 2013). 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 were only expressed during early embryonic stages (Shang et al., 2013). When a neocentromere formed, expression of the MAMDC2-encoding gene was reduced 20- to 100-fold. Chromosome 5 of chickens is diploid, and neocentromeres on this chromosome span genes that are expressed. The hypothesis behind this phenomenon 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 to three genes and that these genes are expressed at levels similar to the wild-type strain. Because the *cen10Δ* mutants of *C. deuterogattii* were aneuploid, the expression of genes spanned by chromosome 10 neocentromeres was normalized to expression levels of a housekeeping gene located on chromosome 10. Due to this internal chromosome 10 normalization, the hypothesis that the genes in the neocentromeres of *C. deuterogattii* are
still expressed because only one allele functions as a neocentromere can be rejected. The genes spanned by neocentromeres of cen9Δ mutants are also expressed at wild-type levels. As the cen9Δ mutants have uniform, wild-type colony sizes, the ploidy levels of these mutants were not tested and we hypothesize that these mutants are haploid/euploid. Based on this, and the modestly increased ploidy levels of cen10Δ mutants, we hypothesize the expression of neocentromeric genes is not due to the presence of an additional chromosome 10 allele. The expression of genes enriched for CENP-A chromatin is similar to that of wild type, and if the allelic hypothesis were valid, the expectation would be a 60% reduction in expression levels.

Genes contained in regions in which C. deuterogattii neocentromeres formed in cenΔ mutants 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 (Alonso et al., 2010; Marshall et al., 2008). However, the expression levels of the neocentromeric genes was lower than their neighboring genes. For example, the gene expression level of the gene (CNBG_5685) flanking native centromere 9 is three times higher than the genes spanned by neocentromeres in cen9Δ mutants. The same trend was observed in the cen10Δ mutants. Here, the expression level of the gene (CNBG_4365) 3’ flanking the neocentromere (primary CENP-A peak) of cen10Δ-A and cen10Δ-C was more than six times higher than the genes spanned by the neocentromere. Also, the neocentromere of cen10Δ-D is flanked by genes whose expression was either 16 or two times higher than the genes spanned by the neocentromere. This suggests that neocentromeres are formed in chromosomal regions with lower gene expression in C. deuterogattii. However, we have identified chromosomal regions that lack gene expression on chromosomes 9 and 10, although these regions were not close to the native centromere.

Of the C. deuterogattii genes spanned by the neocentromere region, one encodes the kinetochore component CENP-C. Several independent biolistic transformations were performed to delete the gene encoding CENP-C, 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 CENP-A. In addition, introducing mCherry-CENP-C by homologous
recombination in the cen10Δ mutants was not successful, but tagging CENP-C in the wild-type strain and in the cen9Δ mutants was effective (Yadav et al., 2018). This suggests that the chromatin may be changed due to neocentromere formation.

In fission yeast, deletion of the gene encoding the CENP-C homolog Cnp3 was lethal at 36°C, but mutants were still viable at 30°C (Suma et al., 2018). However, CENP-A was mislocalized in the cnp3Δ 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 (Higgs and Peterson, 2005). The other three neocentromeric genes encode hypothetical proteins and are available as deletion mutants in C. neoformans gene deletion mutant libraries (Liu et al., 2008; Sun et al., 2014).

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 (Schotanus et al., 2015).

cen10Δ mutants with two CENP-A-enriched regions

In our study, two of the initial cen10Δ mutants had two CENP-A-enriched regions on chromosome 10, suggesting a putative dicentric chromosome. However, CENP-A was not equally distributed between the two CENP-A-enriched regions; one peak was more enriched for CENP-A (primary neocentromere) compared to the other (secondary neocentromere). The appearance of two CENP-A-enriched regions of C. deuterogattii cen10Δ mutants could be explained in 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 by 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 (Stimpson et al., 2010). The formation of dicentric chromosomes occurred in 97% of the fused mammalian chromosomes, which were stable for at least 180 cell divisions (Stimpson et al., 2010).
Several microscopic studies showed that chromosomes with two regions of centromere-protein enrichment are stable (Higgins et al., 2005; Stimpson et al., 2012, 2010; Sullivan and Willard, 1998). This suggests that a dineocentric chromosome 10 could be stable in the population. Second, the two CENP-A-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 CENP-A 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 CENP-A 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 CENP-A-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 cen10Δ mutants, colonies were isolated by streak purification (eight times), suggesting that the presence of two distinct CENP-A peaks occurs continuously. Experimental evolution experiments, followed by ChIP-qPCR, could be conducted to test if the primary neocentromere becomes more stable over time. The independent cen10Δ-B mutant has only one neocentromere, and this neocentromere is located in the same region as the secondary CENP-A peak of cen10Δ-A and -C.

**cen10Δ mutants are partially aneuploid**

Neocentromere formation in chickens results in a low number of aneuploid cells (Shang et al., 2013). Based on whole-genome sequencing of a population of cells, the C. *deuterogattii cen10Δ* isolates are partially aneuploid for chromosome 10. For fully aneuploid isolates, the coverage of Illumina reads is expected to be 2-fold; the cen10Δ isolates with two CENP-A 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 cen10Δ* isolates are aneuploid.
cen10Δ 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 (Ketel et al., 2009). Similar results were reported for neocentromerices in chicken and S. pombe, in which strains with neocentromeres or chromosome fusion have a growth rate similar to the wild-type strain (Ishii et al., 2008; Shang et al., 2013). cen9Δ mutants have a growth rate and uniform colony sizes similar to wild type. In contrast, C. deuterogattii cen10Δ mutants have reduced fitness, and competition assays showed that cen10Δ mutants were less fit compared to the wild-type strain. There was no correlation between reduced fitness and abnormal cell morphology.

The competition assays showed that both cen9Δ and cen10Δ mutants grow slower than the wild type. However, the fitness defects of cen9Δ strains are not as severe as the fitness defects of cen10Δ mutants. If centromere deletions occurred in nature, we hypothesize that the wild type would outcompete all of the cenΔ isolates. The virulence of the cenΔ mutants was not assayed. Based on reduced fitness of the cenΔ mutants we hypothesize that pathogenicity of the cenΔ mutants would be lower than the wild type. However, when chromosome fusion occurs the growth rate is restored to a near wild-type level and we hypothesize that the isolates with 13 chromosomes could have virulence similar to the wild type. Several genes were lost due to the fusion events in the cenΔ mutants; to our knowledge these lost genes have not been associated with pathogenicity of C. deuterogattii.

Neocentromere stains exhibit impaired growth and chromosome fusion restores wild-type growth at elevated temperatures

Deletion of a centromere in S. pombe leads to either neocentromere formation or chromosome fusion due to a noncanonical homologous recombination pathway (Ishii et al., 2008; Ohno et al., 2016). This is in contrast to neocentromere formation in C. deuterogattii, which results in 100% neocentromere formation. Based on PFGE analysis,
the karyotype of the cenΔ isolates is wild type at 30°C, but chromosome fusion can occur at 37°C within the cen10Δ mutants and lead to improved growth at 30°C.

The location of the cen10Δ neocentromere had no influence on the ability to undergo chromosome fusion. Chromosome fusions in three cen10Δ mutants were analyzed in more detail. Of these three cen10Δ mutants one cen10Δ mutant has two enriched CENP-A regions, with the primary peak close to the native centromere 10 (within 21 kb). The second cen10Δ mutant has a neocentromere located 118 kb away from the telomere and the third cen10Δ mutant has a telocentric neocentromere.

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 CDC9 and DNL4 (Sinha et al., 2016).

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 (Miga, 2017). 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 (Miga, 2017). Moreover, a dominant-negative mutation of the human telomeric protein TRF2 leads to telomere-telomere fusions, mainly between acrocentric chromosomes (Stimpson et al., 2010; Van Steensel et al., 1998). In the fungal species Malassezia, chromosome breakage followed by chromosome fusion has led to speciation (Sankaranarayanan et al., 2020). The short regional centromeres (3-5 kb) are fragile and this led most likely to chromosome reduction. By contrast in C. deuterogattii, the chromosomes involved in chromosomal fusion of the cen10Δ 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 (Aksenova et al., 2013). Based on our de novo genome assemblies of the C. deuterogattii large-colony cen10Δ mutants, chromosome fusions occurred with no signs of chromosome rearrangements. PFGE analysis of the initial cen10Δ 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.

Conclusions

Our work shows that, like in other model systems, neocentromeres can be induced in C. deuterogattii. However, C. deuterogattii neocentromeres have several unique characteristics, such as spanning genes whose expression is unaffected by centromere assembly. In some instances, deletion of CEN10 led to chromosome fusion, resulting in enhanced fitness and leading 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 (Luo et al., 2018; Shao et al., 2018).
Acknowledgements

We thank Tom Petes, Beth Sullivan, Kaustuv Sanyal, Sue Jinks-Robertson, Vikas Yadav, Shelby Priest, Shelly Clancey, and Inge van der Kloet for comments on the manuscript. We would like to thank all of the members of the Heitman and Sanyal labs who contribute to the bi-weekly Skype meeting. These studies were supported by NIH/NIAID grants R01 AI050113-15 and R37 MERIT award AI039115-22 to JH. JH is co-director and fellow of the CIFAR program Fungal Kingdom: Threats & Opportunities.

Author contributions

Conceptualization: KS and JH. Formal analysis: KS. Investigation: KS. 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

Aksenova AY, Greenwell PW, Dominska M, Shishkin AA, Kim JC, Petes TD, Mirkin SM. 2013. Genome rearrangements caused by interstitial telomeric sequences in yeast. Proc Natl Acad Sci U S A 1–6. doi:10.1073/pnas.1319313110

Alkan C, Ventura M, Archidiacono N, Rocchi M, Sahinalp SC, Eichler EE. 2007. Organization and evolution of primate centromeric DNA from whole-genome shotgun sequence data. PLoS Comput Biol 3:e181. doi:10.1371/journal.pcbi.0030181

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

Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin A V., Sirotkin A V., Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. 2012. SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477. doi:10.1089/cmb.2012.0021

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

Burrack LS, Berman J. 2012. Neocentromeres and epigenetically inherited features of centromeres. Chromosom Res 20:607–619. doi:10.1007/s10577-012-9296-x

Burrack LS, Hutton HF, Matter KJ, Clancey SA, Liachko I, Plemmons AE, Saha A, Power EA, Turman B, Thevandavakkam MA, Ay F, Dunham MJ, Berman J. 2016. Neocentromeres provide chromosome segregation accuracy and centromere clustering to multiple loci along a Candida albicans chromosome. PLoS Genet 12:e1006317. doi:10.1371/journal.pgen.1006317

Cheeseman IM. 2014. The kinetochore. Cold Spring Harb Perspect Biol 6:a015826. doi:10.1101/cshperspect.a015826

D’Souza CA, Kronstad JW, Taylor G, Warren R, Yuen M, Hu G, Jung WH, Sham A, Kidd SE, Tangen K, Lee N, Zeilmaker T, Sawkins J, McVicker G, Shah S, Gnerre S, Griggs A,
Zeng Q, Bartlett K, Li W, Wang X, Heitman J, Stajich JE, Fraser JA, Meyer W, Carter D, Schein J, Krzywinski M, Kwon-Chung KJ, Varma A, Wang J, Brunham R, Dyke M, Ouellette BFF, Siddiqui A, Marra M, Jones S, Holt R, Birren BW, Galagan JE, Cuomo CA. 2011. Genome variation in Cryptococcus gattii, an emerging pathogen of immunocompetent hosts. mBio 2:e00342-10. doi:10.1128/mBio.00342-10

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

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

Dumesic PA, Homer CM, Moresco JJ, Pack LR, Shanle EK, Coyle SM, Strahl BD, Fujimori DG, Yates JR, Madhani HD. 2015. Product binding enforces the genomic specificity of a yeast polycomb repressive complex. Cell 160:204–218. doi:10.1016/j.cell.2014.11.039

Fan C, Walling JG, Zhang J, Hirsch CD, Jiang J, Wing RA. 2011. Conservation and purifying selection of transcribed genes located in a rice centromere. Plant Cell 23:2821–30. doi:10.1105/tpc.111.085605

Fan Y, Lin Y. 2018. Multiple applications of a transient CRISPR-Cas9 coupled with electroporation (TRACE) system in the Cryptococcus neoformans species complex. Genetics 208:1357–1372. doi:10.1534/genetics.117.300656

Farrer RA, Desjardins CA, Sakthikumar S, Gujja S, Saif S, Zeng Q, Chen Y, Voelz K, Heitman J, May RC, Fisher MC, Cuomo CA. 2015. Genome evolution and innovation across the four major lineages of Cryptococcus gattii. mBio 6:e00868-15. doi:10.1128/mBio.00868-15

Findley K, Sun S, Fraser JA, Hsueh YP, Averette AF, Li W, Dietrich FS, 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. doi:10.1371/journal.pgen.1002528
Fraser JA, Giles SS, Wenink EC, Geunes-Boyer SG, Wright JR, Diezmann S, Allen A, Stajich JE, Dietrich FS, Perfect JR, Heitman J. 2005. Same-sex mating and the origin of the Vancouver Island Cryptococcus gattii outbreak. *Nature* **437**:1360–1364. doi:10.1038/nature04220

Friedman S, Freitag M. 2017. Centrochromatin of fungi In: Black BE, editor. Centromeres and Kinetochores. Cham, Switzerland: Springer International Publishing. pp. 85–109. doi:10.1016/S0092-8674(03)00115-6

Garsed DW, Marshall OJ, Corbin VDA, Hsu A, DiStefano L, Schröder J, Li J, Feng ZP, Kim BW, Kowarsky M, Landsell B, Brookwell R, Myklebost O, Meza-Zepeda L, Holloway AJ, Peđeutour F, Choo KHA, Damore MA, Deans AJ, Papenfuss AT, Thomas DM. 2014. The architecture and evolution of cancer neochromosomes. *Cancer Cell* **26**:653–667. doi:10.1016/j.ccell.2014.09.010

Guenatri M, Bailly D, Maison C, Almouzni G. 2004. Mouse centric and pericentric satellite repeats form distinct functional heterochromatin. *J Cell Biol* **166**:493–505. doi:10.1083/jcb.200403109

Henikoff S, Furuyama T. 2010. Epigenetic inheritance of centromeres. *Cold Spring Harb Symp Quant Biol* **75**:51–60. doi:10.1101/sqb.2010.75.001

Higgins AW, Gustashaw KM, Willard HF. 2005. Engineered human dicentric chromosomes show centromere plasticity. *Chromosom Res* **13**:745–762. doi:10.1007/s10577-005-1009-2

Higgs HN, Peterson KJ. 2005. Phylogenetic analysis of the formin homology 2 domain. *Mol Biol Cell* **16**:1–13. doi:10.1091/mbc.E04

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

Janbon G, Ormerod KL, Paulet D, Byrnes EJ, Yadav V, Chatterjee G, Mullapudi N, Hon C-C, Billmyre RB, Brunel F, Bahn Y-S, Chen W, Chen Y, Chow EWL, Coppée J-Y, Floyd-Averette A, Gaillardin C, Gerik KJ, Goldberg J, Gonzalez-Hilarion S, Gujja S, Hamlin JL, Hsueh Y-P, Ianiri G, Jones S, Kodira CD, Kozubowski L, Lam W, Marra M, Mesner LD,
Mieczkowski PA, Moyrand F, Nielsen K, Proux C, Rossignol T, Schein JE, Sun S, Wollschlaeger C, Wood IA, Zeng Q, Neuvéglise C, Newlon CS, Perfect JR, Lodge JK, Idnurm A, Stajich JE, Kronstad JW, Sanyal K, Heitman J, Fraser JA, Cuomo CA, Dietrich FS. 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. doi:10.1371/journal.pgen.1004261

Ketel C, Wang HSW, McClellan M, Bouchonville K, Selmecki A, Lahav T, Gerami-Nejad M, Berman J. 2009. Neocentromeres form efficiently at multiple possible loci in Candida albicans. PLoS Genet 5:e1000400. doi:10.1371/journal.pgen.1000400

Kobayashi N, Suzuki Y, Schoenfeld LW, Müller CA, Nieduszynski C, Wolfe KH, Tanaka TU. 2015. Discovery of an unconventional centromere in budding yeast redefines evolution of point centromeres. Curr Biol 25:2026–2033. doi:10.1016/j.cub.2015.06.023

Langmead B. 2010. Aligning short sequencing reads with Bowtie. Curr Protoc Bioinforma 11.7.1-11.7.14. doi:10.1002/0471250953.bi1107s32

Liu OW, Chun CD, Chow ED, Chen C, Madhani HD, Noble SM. 2008. Systematic genetic analysis of virulence in the human fungal pathogen Cryptococcus neoformans. Cell 135:174–188. doi:10.1016/j.cell.2008.07.046

Lu M, He X. 2019. Centromere repositioning causes inversion of meiosis and generates a reproductive barrier. Proc Natl Acad Sci U S A 116:21580–21591. doi:10.1073/pnas.1911745116

Luo J, Sun X, Cormack BP, Boeke JD. 2018. Karyotype engineering by chromosome fusion leads to reproductive isolation in yeast. Nature 560:392–396. doi:10.1038/s41586-018-0374-x

Mandáková T, Hloušková P, Koch MA, Lysak MA. 2020. Genome evolution in Arabideae was marked by frequent centromere repositioning. Plant Cell. doi:10.1105/tpc.19.00557

Marshall OJ, Chueh AC, Wong LH, Choo KHA. 2008. Neocentromeres: new insights into centromere structure, disease development, and karyotype evolution. Am J Hum
McNulty SM, Sullivan BA. 2018. Alpha satellite DNA biology: finding function in the recesses of the genome. *Chromosom Res* **26**:115–138. doi:10.1007/s10577-018-9582-3

Miga KH. 2017. Chromosome-specific centromere sequences provide an estimate of the ancestral chromosome 2 fusion event in hominin genomes. *J Hered* **108**:45–52. doi:10.1093/jhered/esw039

Mishra PK, Baum M, Carbon J. 2007. Centromere size and position in *Candida albicans* are evolutionarily conserved independent of DNA sequence heterogeneity. *Mol Genet Genomics* **278**:455–65. doi:10.1007/s00438-007-0263-8

Nagaki K, Cheng Z, Ouyang S, Talbert PB, Kim M, Jones KM, Henikoff S, Buell CR, Jiang J. 2004. Sequencing of a rice centromere uncovers active genes. *Nat Genet* **36**:138–45. doi:10.1038/ng1289

Nergadze SG, Piras FM, Gamba R, Corbo M, Cerutti F, McCarter JGW, Cappelletti E, Gozzo F, Harman RM, Antczak DF, Miller D, Scharfe M, Pavesi G, Raimondi E, Sullivan KF, Giulotto E. 2018. Birth, evolution, and transmission of satellite-free mammalian centromeric domains. *Genome Res* **28**:789–799. doi:10.1101/gr.231159.117

Ohno Y, Ogiyama Y, Kubota Y, Kubo T, 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. doi:10.1093/nar/gkv997

Pertea M, Kim D, Pertea GM, Leek JT, Salzberg SL. 2016. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat Protoc* **11**:1650–1667. doi:10.1038/nprot.2016.095

Quinlan AR, Hall IM. 2010. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**:841–842. doi:10.1093/bioinformatics/btq033

Rhind N, Chen Z, Yassour M, Thompson DA, Haas BJ, Habib N, Wapinski I, Roy S, Lin MF, Heiman DI, Young SK, Furuya K, Guo Y, Pidoux A, Chen HM, Robbertse B, Goldberg JM, Aoki K, Bayne EH, Berlin AM, Desjardins CA, Dobbs E, Dukaj L, Fan L, FitzGerald MG, French C, Gujja S, Hansen K, Keifenheim D, Levin JZ, Mosher RA, Muller CA,
Pfiffner J, Priest M, Russ C, Smialowska A, Swoboda P, Sykes SM, Vaughn M, Vengrova S, Yoder R, Zeng Q, Allshire RC, Baulcombe DC, Birren BW, Brown W, Ekwall K, Kellis M, Leatherwood J, Levin H, Margalit H, Martienssen RA, Nieduszynski CA, Spatafora JW, Friedman N, Dalgaard JZ, Baumann P, Niki H, Regev A, Nusbaum C. 2011. Comparative functional genomics of the fission yeasts. Science 930:930–936. doi:10.1126/science.1203357

Sankaranarayanan SR, Janiri G, Coelho MA, Reza MH, Thimmappa BC, Ganguly P, Vadnala RN, Sun S, Siddharthan R, Tellgren-Roth C, Dawson TL, Heitman J, Sanyal K. 2020. Loss of centromere function drives karyotype evolution in closely related Malassezia species. eLife 9. doi:10.7554/eLife.53944

Sanyal K, Baum M, 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. doi:10.1073/pnas.0404318101

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

Scott KC, Sullivan BA. 2013. Neocentromeres: a place for everything and everything in its place. Trends Genet 30:66–74. doi:10.1016/j.tig.2013.11.003

Shang W-HH, Hori T, Martins NMCC, Toyoda A, Misu S, Monma N, Hiratani I, Maeshima K, Ikeo K, Fujiyama A, Kimura H, Earnshaw WC, Fukagawa T. 2013. Chromosome engineering allows the efficient isolation of vertebrate neocentromeres. Dev Cell 24:635–648. doi:10.1016/j.devcel.2013.02.009

Shao Y, Lu N, Wu Z, Cai C, Wang S, Zhang LL, Zhou F, Xiao S, Liu L, Zeng X, Zheng H, Yang C, Zhao Z, Zhao G, Zhou JQ, Xue X, Qin Z. 2018. Creating a functional single-chromosome yeast. Nature 560:331–335. doi:10.1038/s41586-018-0382-x

Sinha S, Villarreal D, Shim EY, Lee SE. 2016. Risky business: Microhomology-mediated end joining. Mutat Res Mol Mech Mutagen 788:17–24. doi:10.1016/j.mrfmmm.2015.12.005
Smith KM, Phatale PA, Sullivan CM, Pomraning KR, Freitag M. 2011. Heterochromatin is required for normal distribution of *Neurospora crassa* CenH3. *Mol Cell Biol* **31**:2528–2542. doi:10.1128/MCB.01285-10

Soyer JL, Möller M, Schotanus K, Connolly LR, Galazka JM, Freitag M, Stukenbrock EH. 2015. Chromatin analyses of *Zymoseptoria tritici*: Methods for chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-seq). *Fungal Genet Biol* **79**:63–70. doi:10.1016/j.fgb.2015.03.006

Stimpson KM, Matheny JE, Sullivan BA. 2012. Dicentric chromosomes: unique models to study centromere function and inactivation. *Chromosome Res* **20**:595–605. doi:10.1007/s10577-012-9302-3

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

Sullivan BA, Willard HF. 1998. Stable dicentric X chromosomes with two functional centromeres. *Nat Genet* **20**:227–228. doi:10.1038/3024

Suma M, Kitagawa T, Nakase Y, Nakazawa N, Yanagida M, Matsumoto T. 2018. Fission yeast CENP-C (Cnp3) plays a role in restricting the site of CENP-A accumulation. *G3* **8**:2723–2733. doi:10.1534/g3.118.200486

Sun S, Billmyre RB, Mieczkowski PA, Heitman J. 2014. Unisexual reproduction drives meiotic recombination and phenotypic and karyotypic plasticity in *Cryptococcus neoformans*. *PLoS Genet* **10**:e1004849. doi:10.1371/journal.pgen.1004849

Thakur J, 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. doi:10.1101/gr.141614.112

Thorvaldsdóttir H, Robinson JT, Mesirov JP. 2013. Integrative genomics viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* **14**:178–192. doi:10.1093/bib/bbs017

Tolomeo D, Capozzi O, Stanyon RR, Archidiacono N, D’Addabbo P, Catacchio CR, Purgato
S, Perini G, Schempp W, Huddleston J, Malig M, Eichler EE, Rocchi M. 2017. Epigenetic origin of evolutionary novel centromeres. Sci Rep 7:41980. doi:10.1038/srep41980

Van Steensel B, Smogorzewska A, De Lange T. 1998. TRF2 protects human telomeres from end-to-end fusions. Cell 92:401–413. doi:10.1016/S0092-8674(00)80932-0

Ventura M, Antonacci F, Cardone MF, Stanyon R, D’Addabbo P, Cellamare A, Sprague LJ, Eichler EE, Archidiacono N, Rocchi M. 2007. Evolutionary formation of new centromeres in macaque. Science 316:243–246. doi:10.1126/science.1140615

Warburton PE. 2004. Chromosomal dynamics of human neocentromere formation. Chromosom Res 12:617–26. doi:10.1023/B:CHRO.0000036585.44138.4b

Yadav V, Sun S, Billmyre RB, Thimmappa BC, Shea T, Lintner R, Bakkeren G, Cuomo CA, Heitman J, Sanyal K. 2018. RNAi is a critical determinant of centromere evolution in closely related fungi. Proc Natl Acad Sci 201713725. doi:10.1073/pnas.1713725115

Zhang B, Lv Z, Pang J, Liu Y, Guo X, Fu S, Li J, Dong Q, Wu H-J, Gao Z, Wang X-J, Han F. 2013. Formation of a functional maize centromere after loss of centromeric sequences and gain of ectopic sequences. Plant Cell 25:1979–1989. doi:10.1105/tpc.113.110015
Tables

Table 1. Genes located inside neocentromeres.

The chromosomal locations, sizes, and GC content (%) for the native centromere and \( cen\Delta \) mutants are shown. For the neocentromeres, gene ID, predicted function, and the amount of CENP-A coverage are indicated.
| Native centromere | Chr coor (bp) | Size (kb) | Size compared to native centromere (%) | GC% | Genes spanned by neocentromere | Gene ID | % covered by Neocentromere | Exons inside neocentromere |
|------------------|---------------|-----------|----------------------------------------|------|-------------------------------|---------|---------------------------|--------------------------------|
| 9                | Chr9:755,771-762,621 | 6.84     | -                                      | 43.6 | -                             | -       | -                         | -                                      |
| cen9Δ-A          | Chr9:785,352-789,247 | 3.87     | 56.6                                   | 46.1 | Escr-II complex subunit (VPS25) | CNBG_5690 | 100                       | Last exon                              |
|                  |                |          |                                        |      | Iron regulator 1              | CNBG_9614 | 14.6          | Last exon                              |
| cen9Δ-B          | Chr9:775,164-780,756 | 4.41     | 64.5                                   | 46.6 | Xylosphosphotransferase       | CNBG_587 | 6.9                       | Last exon                              |
|                  |                |          |                                        |      | Tranglycosylase SLT domain-containing protein | CNBG_9613 | 100                       | Last exon                              |
| cen9Δ-C          | Chr9:775,164-780,756 | 4.41     | 64.5                                   | 46.6 | Xylosphosphotransferase       | CNBG_587 | 6.9                       | Last exon                              |
|                  |                |          |                                        |      | Glutamate synthase (NADPH/NADH) | CNBG_589 | 33.7          | Last exon                              |
| cen9Δ-D          | Chr9:750,902-755,294 | 4.37     | 63.9                                   | 41.9 | Hypothetical protein          | CNBG_584 | 92.8          | Last exon                              |
|                  |                |          |                                        |      | Derlin-3/3                   | CNBG_585 | 100                       | Last exon                              |
| cen9Δ-E          | Chr9:775,164-780,756 | 5.56     | 81.3                                   | 50   | Xylosphosphotransferase       | CNBG_587 | 6.9                       | Last exon                              |
|                  |                |          |                                        |      | Tranglycosylase SLT domain-containing protein | CNBG_9613 | 100                       | Last exon                              |
| cen9Δ-F          | Chr9:771,614-775,469 | 3.83     | 56.0                                   | 51.5 | Xylosphosphotransferase       | CNBG_587 | 100                       |                                    |
| Native centromere 10 | Chr10:362,876-369,657 | 6.77     | -                                      | 42.6 | -                             | -       | -                         | -                                      |
| cen10Δ-A         | Chr10:115,954-120,422 | 4.46     | 65.9                                   | 46.9 | CENPC/MIF2                    | CNBG_4461 | 88.3          | 1, 2, 3, 4 (only 5th is outside)          |
|                  |                |          |                                        |      | Hypothetical protein          | CNBG_4462 | 100                       |                                    |
| cen10Δ-B         | Chr10:391,090-393,946 | 2.85     | 42.1                                   | 48.9 | Serine/threonine-protein phosphatase 2A activator 2 (RRD2) | CNBG_945 | 10.6          | Last exon (5th)                        |
|                  |                |          |                                        |      | Hypothetical protein          | CNBG_446 | 100                       |                                    |
|                  |                |          |                                        |      | Hypothetical protein          | CNBG_4465 | 23.4          | Last exon (3rd)                        |
| cen10Δ-C         | Chr10:115,954-120,422 | 4.46     | 65.9                                   | 46.9 | CENPC/MIF2                    | CNBG_446 | 88.3          | 1, 2, 3, 4 (only 5th is outside)          |
|                  |                |          |                                        |      | Hypothetical protein          | CNBG_4462 | 100                       |                                    |
| cen10Δ-D         | Chr10:391,090-393,946 | 2.85     | 42.1                                   | 48.9 | Serine/threonine-protein phosphatase 2A activator 2 (RRD2) | CNBG_945 | 10.6          | Last exon (5th)                        |
|                  |                |          |                                        |      | Hypothetical protein          | CNBG_446 | 100                       |                                    |
|                  |                |          |                                        |      | Hypothetical protein          | CNBG_4465 | 23.4          | Last exon (3rd)                        |
| cen10Δ-E         | Chr10:101,4-135 | 4.38     | 64.7                                   | 53.2 | Hypothetical protein          | CNBG_1045 | 100                       |                                    |
|                  |                |          |                                        |      | Hypothetical protein          | CNBG_448 | 100                       |                                    |
| cen10Δ-F         | Chr10:342,517-345,159 | 2.64     | 39.0                                   | 45.5 | Hypothetical protein          | CNBG_438 | 18.6          | Last two exons                         |
|                  |                |          |                                        |      | Hexokinase (HXK1)            | CNBG_10075 | 100                       |                                    |
| cen10Δ-G         | Chr10:378,389-386,366 | 7.97     | 117.7                                  | 46.5 | Hypothetical protein          | CNBG_437 | 100                       |                                    |
|                  |                |          |                                        |      | Hypothetical protein          | CNBG_437 | 100                       |                                    |
|                  |                |          |                                        |      | Hypothetical protein          | CNBG_437 | 100                       |                                    |
|                  |                |          |                                        |      | Hypothetical protein          | CNBG_437 | 100                       |                                    |

Table 1. Genes located inside neocentromeres.
Figures

Figure 1. Centromere deletion leads to neocentromere formation.
For each panel, the chromosome coordinates are indicated. Genes (CDS) are shown in blue arrows and the truncated transposable elements, located in the native centromere (CEN9 or CEN10), are colored according to their class (Tcn4 in orange and Tcn6 in green). Previously generated RNA-sequencing obtained from wild-type cells was re-mapped and shown in green. In each panel, the wild-type CENP-A content is shown. In the wild type, CENP-A is only enriched at the native centromeres. For each cenΔ mutant, the neocentromeric region is shown by enrichment of CENP-A. (A) Schematic full overview of chromosome 9, the indentation represents the native centromere 9 position. The light grey area points to the zoomed-in chromosomal region shown with the detailed view of the native centromere (CEN9) and the location of the cen9Δ mutant neocentromeres. Neocentromeres of cen9Δ-B, cen9Δ-C and cen9Δ-E formed at the same chromosomal location. (B) Detailed view of the neocentromere of cen10Δ-B and the secondary CENP-A peak of cen10Δ-A and cen10Δ-C. (C) Overview of the chromosomal 10 region spanning 100 to 410 kb. cen10Δ-A and cen10Δ-C have two regions enriched with CENP-A (primary and secondary). (D) Schematic full overview of the full chromosome 10, the indentation represent the chromosomal location of the native centromere (CEN10). The light grey areas point to the zoomed-in chromosomal regions shown in panel C and below. The neocentromere of cen10Δ-E is indicated with an arrow. Lower panel, detailed view of the native centromere (CEN10) and the neocentromeres formed in cen10Δ-A, cen10Δ-C, cen10Δ-D, cen10Δ-F and cen10Δ-G mutants. (E) Detailed view of the telocentric neocentromere of cen10Δ-E.

Figure 2. Expression of neocentromeric genes.
Expression of the neocentromeric genes was assessed by qPCR for all cenΔ mutants and expression is shown as Log2^ΔΔCt. For cen10Δ-A, cen10Δ-B and cen10Δ-C, two genes were selected from each neocentromeric region, all other cenΔ mutants are represented by one gene spanned by CENP-A. cen10Δ-B has only one CENP-A-enriched region, and in this case,
the genes located within primary peak of cen10Δ-A and cen10Δ-C served as controls. The qPCRs of cen10Δ mutants are normalized with a housekeeping gene located on chromosome 10. The qPCRs of cen9Δ mutants are normalized with actin. Error bars show standard deviation.

**Figure 3. cen10Δ mutant strains have reduced fitness compared to the wild-type strain.**

(A) Six out of seven cen10Δ mutants had a longer doubling time and slower growth than the wild-type strain. In contrast cen10Δ-E grows similarly to the wild type. Error bars show standard deviation. (B) Doubling times and fold change compared to wild type are shown. (C) Competition assays with the wild type and cen9Δ and cen10Δ mutant strains. Mixed cultures (1:1) were grown overnight and plated with and without selection agents. After four days, colonies were counted and the percentage of cenΔ mutants (black) and wild type (grey) in each culture was plotted. As a control (C) a wild-type strain with a NAT marker was mixed with the wild type.

**Figure 4. cen10Δ mutants are aneuploid.**

The whole genomes of small and large colonies derived from four cen10Δ mutants were sequenced and read coverage (corresponding to ploidy levels) was plotted. Small colonies of cen10Δ 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 cen10Δ-B-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 cen10Δ-E was duplicated. In addition, cen10Δ-E-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. The native centromeric location is shown by a black square. Due to the deletion of
centromere 10, the location of the native centromere lacks sequence reads for each sample.

**Figure 5.** *cen10Δ* mutants undergo chromosome fusion leading to improved fitness at 37°C.

Chromosomal fusions were studied in detail for three *cen10Δ* mutants restored to wild-type growth levels at 37°C (large colonies). After chromosome fusion, the fused chromosomes of *cen10Δ-A-L* and *cen10Δ-B-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) *cen10Δ-A-L1*, the fusion occurred between chromosome 10 and chromosome 4. (B) In *cen10Δ-B-L1*, chromosomal fusion occurred between chromosomes 10 and 7. (C) *cen10Δ-E-L1* chromosomal fusion occurred between chromosomes 10 and 1.

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

(A) Neocentromeres are inactive after chromosomal fusion. For each neocentromere two qPCR primer pairs located in genes spanned by the neocentromere in *cen10Δ-A* and *cen10Δ-B* mutant were used in a ChIP-qPCR experiment. Analyzed is the CENP-A enrichment of 1) a *cen10Δ* mutant, 2) a large colony derived from the *cen10Δ* mutant, and 3) the wild-type strain. Centromere 6 (*CEN6*) was included as a positive control, and actin was included as a negative control. Data is shown for *cen10Δ-A*, *cen10Δ-A-L1*, *cen10Δ-B*, *cen10Δ-B-L2*, and wild type. For *cen10Δ-A* and *cen10Δ-A-L1* mutants, the chromosomal regions investigated are indicated according to the primary and secondary CENP-A peaks of the *cen10Δ-A* mutant. The *cen10Δ-B* mutant has only one CENP-A-enriched region which co-localized with the secondary CENP-A peak of *cen10Δ-A* and this region is labeled with neocen in *cen10Δ-B* and *cen10Δ-B-L1*. Error bars show standard
deviation. (B) PFGE analysis shows that the band corresponding to chromosome 10 was lost in the large colonies and instead larger bands appear due to the fusion of chromosome 10 with other chromosomes. *cen10Δ* 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.
Supplementary figure legends

Supplementary figure 1. Confirmation of centromere 10 deletion by PCR (together with Figure 1).

(A) The native centromeric regions of chromosome 9 in the wild type and cen9Δ mutants are shown. Genes flanking the centromeric region in the WT strain are indicated with an arrow and gene ID. The centromere was replaced by a nourseothricin (NAT) drug-resistance gene cassette, indicated with a rectangle (labelled NAT). Black lines indicate the length of the PCR product used to confirm the centromere 9 deletion. Black arrows indicate primers. (B) PCR confirmation of the centromere 9 deletion in cen9Δ mutants ran on an ethidium bromide-stained gel. The WT (W) and no-template control (B) are included as controls. For the spanning PCR, both primers are located outside of the transformed product and for the junction PCRs, one primer is outside of the transformed product and one primer is located inside the nourseothricin (NAT) drug-resistance gene cassette. (C) Centromere 10 region of the wild type (WT) and cen10Δ mutants are shown. Genes and nourseothricin (NAT) drug-resistance gene cassette, primers, and the length of the PCR product are labeled as in A. (D) PCR confirmation of the centromere 10 deletion in cen10Δ mutants ran on an ethidium bromide-stained gel. The WT (W) and no-template control (B) are included as controls.

Supplementary figure 2. Centromere 10 is deleted in cen10Δ isolates (together with Figure 1).

(A) The region corresponding to the native centromere 10 of the wild type and cen10Δ mutants is shown. Upstream and downstream flanking regions, used for homologous recombination, are indicated with rectangles (labeled “5’ region” or “3’ region”). The nourseothricin drug resistance gene cassette is indicated with a rectangle, (labeled NAT). Grey lines indicate XbaI restriction sites. PCR products of the 5’ and 3’ regions were used as Southern blot probes, and the expected restriction pattern is indicated above. (B) Southern blot analysis for the wild-type and cen10Δ mutant strains is presented. Left panel shows the Southern blot, right panel shows the ethidium bromide-stained gel prior to
Southern blotting. Both the wild type and \textit{cen10}Δ mutants have the expected restriction pattern which is shown in panel A.

**Supplementary figure 3. \textit{cen9}Δ and \textit{cen10}Δ mutants have a wild-type karyotype (together with Figure 1, 5 and 6).**

PFGE analysis with the wild type and \textit{cen9}Δ and \textit{cen10}Δ mutants. Cells were isolated from growing at 30°C. \textit{S. cerevisiae} chromosomes serve as size markers and are indicated with “M” on both sides of the ethidium bromide stained gel (Chromosome sizes shown on the left side). The chromosome sizes of the reference strain R265 are shown on the right side. In several cases, chromosomes co-migrate as indicated.

(A) PFGE analysis of \textit{cen10}Δ mutants. Order of samples from left to right are wild type and \textit{cen10}Δ mutants (-A to -G). (B) Southern blot analysis of the PFGE from panel A. A region of \textit{CEN10} served as the probe. On chromosome 10, the probe only has homology with the centromere. In the wild-type strain, the probe hybridized to chromosome 10 (arrow). \textit{cen10}Δ mutants lack centromere 10 and thus do not show hybridization of the probe. The probe has homology to centromeres of other chromosomes, resulting in cross-hybridization. (C) PFGE analysis of \textit{cen9}Δ mutants. Order of samples from left to right are \textit{cen9}Δ mutants (-A to -D) and wild type.

**Supplementary figure 4. ChIP-qPCR with additional kinetochore proteins**

To confirm the binding of the kinetochore to the CENP-A-enriched regions, two additional kinetochore proteins were tagged with GFP and ChIP-qPCR analysis was performed. \textit{cen9}Δ mutants were transformed with a construct expressing CENP-C-GFP and \textit{cen10}Δ mutants were transformed with a construct expressing Mis12-GFP. As a control, the wild type was transformed with a construct expressing CENP-C-GFP or Mis12-GFP. For each, ChIP-qPCR data is shown for 1) the internal positive control (\textit{CEN6}), 2) primer pair(s) specific for the neocentromere(s) and 3) enrichment compared to actin (set to 1). Error bars show standard deviation. (A) qPCR results of the ChIPs with \textit{cen10}Δ mutants transformed with Mis12-GFP. \textit{cen10}Δ-A and \textit{cen10}Δ-C have two CENP-A-enriched regions (primary peak and
secondary peak) and this is indicated in the figure. B) qPCR results of ChIPs with cen9Δ mutants transformed with CENP-C-GFP. C) qPCR results of the ChIP with the wild-type strain transformed with Mis12-GFP. This panel serves as a control for the ChIP-qPCRs performed in panel A. As the neocentromere of cen10Δ-B and the secondary peak of cen10Δ-A and cen10Δ-C are formed in the same chromosomal region, the qPCR reaction for this chromosomal region in the wild type is only shown once. Similarly, the primary peak of cen10Δ-A and cen10Δ-C formed in the same chromosomal location, and this region is only shown once in the wild type. D) qPCR results of the ChIP with the wild-type strain transformed with CENP-C-GFP. These qPCRs serve as a control for the ChIP-qPCRs performed in panel B. Three mutants have neocentromeres formed at the same chromosomal location (cen9Δ-B, cen9Δ-C and cen9Δ-E) and the CENP-C enrichment of this region in the wild type is only shown once.

**Supplementary figure 5. cen10Δ mutants with chromosomal fusion have a wild-type growth rate (together with Figure 3 and 6).**
The doubling times of large and small colonies derived from cen10Δ mutants at 30°C were determined. Large colonies had a growth rate similar to the wild-type strain, while small colonies and the initial cen10Δ mutants had a similar growth rate with one another but that was slower than the wild-type strain and the derived large colonies. (A) Growth curves for cen10Δ-A-derived isolates and the wild-type strain are shown. (B) Growth curves for cen10Δ-B-derived isolates and the wild-type strain are shown.

**Supplementary figure 6. Deletion within subtelomeric regions in chromosome fusion isolates (together with Figure 4 and 6).**
Sequence reads were mapped to the reference R265 genome. Regions with sequence coverage are shown in blue, and those without sequence coverage are shown in white. Subtelomeric regions of the large cen10Δ colonies have lost sequence coverage, whereas cen10Δ small colonies have wild-type sequence coverage. Telomeric loss was not observed for the large colonies of cen10Δ-E. For all panels, genes (CDS) are shown on the top (light
blue). (A) Detailed view of the 3’ subtelomeric region of chromosome 10. Large colonies of cen10Δ-A and cen10Δ-B lost sequences corresponding to a region of 6.5 and ~8 kb. (B) Detailed view of the 3’ subtelomeric region of chromosome 4. Large colonies of cen10Δ-A have lost sequences corresponding to a 12-kb region. (C) Detailed view of the 3’ subtelomeric region of chromosome 7. Large colonies of cen10Δ-B lost sequences corresponding to an ~18.5-kb region.

Supplementary figure 7. Chromosome fusion in large cen10Δ colonies (together with Figure 5).

Sequence reads were mapped back to the de novo cen10Δ genome assemblies to verify the quality of the de novo cen10Δ mutant genome assemblies. For each panel the full length of the scaffold, which includes the fusion point, and a 100-bp detailed region is shown. Sequences homologous to chromosome 10 are depicted in blue and sequence reads mapped back to the de novo genome assemblies are shown in grey. (A) The de novo genome assembly of cen10Δ-A-L contains a scaffold (~211 kb) of a fused chromosome and this scaffold consists of an 81.97-kb region of chromosome 10 (blue) and a 129.38-kb region of chromosome 4 (green). The sequence coverage of this scaffold is 104 fold. The chromosome fusion occurred with a 2-bp overlap between chromosome 4 and 10. (B) The de novo genome assembly of cen10Δ-B-L contains a scaffold (~886 kb) and consists of an 80.46-kb region of chromosome 10 (blue) and an 806.09-kb region of chromosome 7 (red). The fusion point lacks overlap between chromosomes 7 and 10. The sequence coverage of this scaffold is 104 fold. (C) The de novo genome assembly of cen10Δ-E-L consists of a scaffold (~585 kb) of a fused chromosome and this scaffold consists of an 83.59-kb region of chromosome 10 (blue) and 501.37-kb of chromosome 1 (pink). The sequence coverage of this scaffold is 140 fold. The chromosome fusion occurred with a 6-bp overlap between chromosome 1 and 10. (D) PCR confirmation of the chromosome fusion occurring in the large colonies derived from cen10Δ-B, cen10Δ-C, and cen10Δ-A. For each chromosomal fusion, a PCR spanning the fusion was performed. For each chromosome fusion, PCRs were
performed for: 1) a large colony derived from the *cen10Δ* mutant, 2) the original *cen10Δ* mutant, 3) a small colony derived from growth at 37°C, and 4) the wild type.

**Supplementary figure 8. *cen10Δ* mutants have elongated cell morphology (together with Figure 1).**

Cell morphology of >1000 cells each for the wild-type strain and five *cen10Δ* mutant strains was analyzed, counted, and plotted as a percentage of total cell number. L = large, S = small, Scale bar = 10 µm. (A) Percentage of cells with elongated cell morphology. Formation of abnormal cell morphology is rare. The *cen10Δ*-C and *cen10Δ*-F mutants have an increased number of elongated cells. *cen10Δ*-D, *cen10Δ*-E, and *cen10Δ*-G mutants had <0.5% elongated cell morphology. (B) Representative view of wild-type and *cen10Δ*-F mutant cells. Several *cen10Δ*-F mutants had enlarged cell shapes and formed elongated cell clusters. (C) Representative view of wild-type colonies and a population of mixed colony sizes of *cen10Δ* mutants. Shown here is *cen10Δ*-A. Large colonies have a size to similar to wild type.
**Figure S2**

**A**
Probe: 5' region + 3' region

WT

| 0 | 2 | 4 | 6 | 8 | 10 | 12 | 14 | 16 | 18 kb |
|---|---|---|---|---|---|----|----|----|-------|
| 2.9 kb | | | | | | | | | |

CEN10

| 0 | 2 | 4 | 6 | 8 | 10 | 12 | 14 | 16 | 18 kb |
|---|---|---|---|---|---|----|----|----|-------|
| 2.1 kb | 6.8 kb | | | |

cen10Δ

| 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 9 | 10 kb |
|---|---|---|---|---|---|---|---|---|-------|
| 2.1 kb | | | | | | | | | |

B

| cen10Δ-A | cen10Δ-B | cen10Δ-C | cen10Δ-D | cen10Δ-E | cen10Δ-F | cen10Δ-G | False positive |
|---|---|---|---|---|---|---|----------------|

WT

- 10 kb
- 8 kb
- 6.8 kb
- 5.6 kb
- 5 kb
- 4 kb
- 3 kb
- 2 kb
- 1.5 kb
Figure S4

A  
**Mis12**

B  
**CENPC**

C  
**Mis12**

D  
**CENPC**

Fold enrichment compared to actin and input

| cen10Δ-A | cen10Δ-B | cen10Δ-C | cen10Δ-D | cen10Δ-E | cen9Δ-A | cen9Δ-B | cen9Δ-C | cen9Δ-D | cen9Δ-E |
|----------|----------|----------|----------|----------|---------|---------|---------|---------|---------|
| CEV6     | Primary  | ACT1     | CEV6     | Primary  | ACT1    | CEV6    | ACT1    | CEV6    | ACT1    |
| Neocentromere | ACT1 | NEocentromere | ACT1 | NEocentromere | ACT1 |

Cen10Δ-A, -B, & -C

Cen9Δ-A, -B, & -C

Fold enrichment compared to actin and input

| CEV6 | cen10Δ-A, -B, & -C | cen10Δ-D | cen10Δ-E | ACT1 |
|------|-------------------|----------|----------|------|
|     |                   |          |          |      |

| CEV6 | cen9Δ-A, -B, & -C | cen9Δ-D | cen9Δ-E | ACT1 |
|------|-------------------|----------|----------|------|
|     |                   |          |          |      |
Figure S7

1) Large cen10Δ colony (Chromosome fusion)
2) Original cen10Δ mutant
3) Small cen10Δ colony
4) Wild-type
5) No-DNA control

Expected size of PCR for chromosome fusion:
cen10Δ-B   4.1 kb
acen10Δ-E 3.1 kb
cen10Δ-A  2.4 kb
Figure S8

A

Percentage of cells with elongated morphology

WT  cen10Δ-C  cen10Δ-D  cen10Δ-E  cen10Δ-F  cen10Δ-G

B

WT  cen10Δ-F

C

WT  cen10Δ-A

Large colony

Small colony
| Primer Name | Purpose | Sequence |
|-------------|---------|----------|
| JOHE51249/KS1053 | lab identifier | △△ △△ △△ △△ △△ △△ |
| JOHE51096/KS964 | purpose | △△ △△ △△ △△ △△ △△ |
| JOHE51092/KS960 | sequence | △△ △△ △△ △△ △△ △△ |
| JOHE51089/KS957 | | △△ △△ △△ △△ △△ △△ |
| JOHE50698/KS811 | | △△ △△ △△ △△ △△ △△ |
| JOHE50697/KS810 | | △△ △△ △△ △△ △△ △△ |
| JOHE50418/KS726 | | △△ △△ △△ △△ △△ △△ |
| JOHE50412/KS720 | | △△ △△ △△ △△ △△ △△ |
| JOHE46287/KS586 | | △△ △△ △△ △△ △△ △△ |
| JOHE46286/KS585 | | △△ △△ △△ △△ △△ △△ |
| JOHE46271/KS570 | | △△ △△ △△ △△ △△ △△ |
| JOHE45788/KS515 | | △△ △△ △△ △△ △△ △△ |
| JOHE45785/KS512 | | △△ △△ △△ △△ △△ △△ |
| JOHE45637/KS502 | | △△ △△ △△ △△ △△ △△ |
| JOHE45410/KS475 | | △△ △△ △△ △△ △△ △△ |
| JOHE45397/KS470 | | △△ △△ △△ △△ △△ △△ |
| JOHE45393/KS466 | | △△ △△ △△ △△ △△ △△ |
| JOHE45381/KS454 | | △△ △△ △△ △△ △△ △△ |
| JOHE45317/KS434 | | △△ △△ △△ △△ △△ △△ |
| JOHE44970/KS407 | | △△ △△ △△ △△ △△ △△ |

For each primer, the lab identifier, purpose, and sequence are shown.
### Supplementary Table 2. Strains used in this study.

| Heitman lab ID | Name      | Strain          | Parental strain | Selection cassette | Reference          |
|---------------|-----------|-----------------|-----------------|--------------------|--------------------|
| 20096         | KS1       | R265 (Wild-type)| -               | -                  | Fraser et al (2005) |
| 20097         | KS2       | cen10 △-A       | R265            | NAT                | This study          |
| 20098         | KS3       | cen10 △-B       | R265            | NAT                | This study          |
| 20099         | KS4       | cen10 △-C       | R265            | NAT                | This study          |
| 20100         | KS5       | cen10 △-D       | R265            | NEO                | This study          |
| 20101         | KS6       | cen10 △-E       | R265            | NAT                | This study          |
| 20102         | KS7       | cen10 △-F       | R265            | NAT                | This study          |
| 20103         | KS8       | cen10 △-G       | R265            | NAT                | This study          |
| 20104         | KS9       | cen10 △-A-S1    | KS2             | NAT                | This study          |
| 20105         | KS10      | cen10 △-A-L1    | KS2             | NAT                | This study          |
| 20106         | KS11      | cen10 △-B-S1    | KS3             | NAT                | This study          |
| 20107         | KS12      | cen10 △-B-S2    | KS3             | NAT                | This study          |
| 20108         | KS13      | cen10 △-B-S3    | KS3             | NAT                | This study          |
| 20109         | KS14      | cen10 △-B-L1    | KS3             | NAT                | This study          |
| 20110         | KS15      | cen10 △-B-L2    | KS3             | NAT                | This study          |
| 20111         | KS16      | cen10 △-C-S1    | KS4             | NAT                | This study          |
| 20112         | KS17      | cen10 △-C-S2    | KS4             | NAT                | This study          |
| 20113         | KS18      | cen10 △-C-S3    | KS4             | NAT                | This study          |
| 20114         | KS19      | cen10 △-C-L1    | KS4             | NAT                | This study          |
| 20115         | KS20      | cen10 △-C-L2    | KS4             | NAT                | This study          |
| 20116         | KS21      | cen10 △-E-S1    | KS6             | NAT                | This study          |
| 20117         | KS22      | cen10 △-E-S2    | KS6             | NAT                | This study          |
| 20118         | KS23      | cen10 △-E-S3    | KS6             | NAT                | This study          |
| 20119         | KS24      | cen10 △-E-L1    | KS6             | NAT                | This study          |
| 20120         | KS25      | cen10 △-E-L2    | KS6             | NAT                | This study          |
| 20121         | KS26      | cen9 △-A        | R265            | NAT                | This study          |
| 20122         | KS27      | cen9 △-B        | R265            | NAT                | This study          |
| 20123         | KS28      | cen9 △-C        | R265            | NAT                | This study          |
| 20124         | KS29      | cen9 △-D        | R265            | NAT                | This study          |
| 20125         | KS30      | cen9 △-E        | R265            | NAT                | This study          |
| 20126         | KS31      | cen9 △-F        | R265            | NAT                | This study          |
| 20126         | KS32      | R265 MIS12      | R265            | NAT/NEO            | This study          |
| 20126         | KS33      | cen10 △-A MIS12 | KS2             | NAT/NEO            | This study          |
| 20126         | KS34      | cen10 △-B MIS12 | KS3             | NAT/NEO            | This study          |
| 20126         | KS35      | cen10 △-C MIS12 | KS4             | NAT/NEO            | This study          |
| 20126         | KS36      | cen10 △-D MIS12 | KS5             | NAT/NEO            | This study          |
| 20126         | KS37      | cen10 △-E MIS12 | KS6             | NAT/NEO            | This study          |
| 20126         | KS38      | R265 CENPC      | R265            | NAT/NEO            | This study          |
| 20126         | KS39      | cen9 △-A CENPC  | KS26            | NAT/NEO            | This study          |
| 20126         | KS40      | cen9 △-B CENPC  | KS27            | NAT/NEO            | This study          |
| 20126         | KS41      | cen9 △-C CENPC  | KS28            | NAT/NEO            | This study          |
| 20126         | KS42      | cen9 △-D CENPC  | KS29            | NAT/NEO            | This study          |
| 20126         | KS43      | cen9 △-E CENPC  | KS30            | NAT/NEO            | This study          |

For each strain used in this study, the lab strain identifier, description, and parental strain are indicated.
Expression levels of genes where neocentromeres formed in the cen Δ mutants were analyzed in the R265 wild-type strain. Previously generated RNA sequencing data were remapped to the R265 reference genome, and the expression levels (FPKM) were analyzed for expression levels of genes located in the subtelomeric regions. Expression levels of genes where neocentromeres formed in the wild-type strain are presented in Supplementary Table 3. Neocentromeric regions are expressed in the wild-type strain.

### Table 3. Neocentromeric regions are expressed in the wild-type strain.

| Gene ID | Gene | FPKM value | Present after chromosome fusion |
|---------|------|------------|---------------------------------|
| CBGB_5690 | Escrt-II complex subunit (VPS25) | 36.986 |
| CBGB_9614 | Iron regulator 1 | 60.887352 |
| CBGB_5687 | Xylosphosphotransferase | 77.018387 |
| CBGB_9613 | Transglycosylase SL1 domain-containing protein | 68.94972 |
| CBGB_5689 | Glutamate synthase (NADPH/NADH) | 47.842888 |
| CBGB_5687 | Xylosphosphotransferase | 77.018387 |
| CBGB_9613 | Transglycosylase SL1 domain-containing protein | 68.94972 |
| CBGB_5689 | Glutamate synthase (NADPH/NADH) | 47.842888 |
| CBGB_5683 | Transitional endoplasmic reticulum ATPase | 261.714935 |
| CBGB_5685 | Dcm1 | 38.73462 |
| CBGB_5687 | Xylosphosphotransferase | 77.018387 |
| CBGB_9613 | Transglycosylase SL1 domain-containing protein | 68.94972 |
| CBGB_5689 | Glutamate synthase (NADPH/NADH) | 47.842888 |
| CBGB_5687 | Xylosphosphotransferase (SPT11) | 77.018387 |
| CBGB_4461 | CENP/MIF2 | 80.212288 |
| CBGB_4462 | Hypothetical protein | 76.691688 |
| CBGB_9459 | Serine/threonine-protein phosphatase 2A activator 2 (RRD2) | 61.127243 |
| CBGB_4366 | Hypothetical protein | 13.426547 |
| CBGB_4365 | Hypothetical protein | 279.7435 |
| CBGB_4461 | CENP/MIF2 | 80.212288 |
| CBGB_4462 | CENP/MIF2 | 67.691688 |
| CBGB_4379 | Serine/threonine-protein phosphatase 2A activator 2 (RRD2) | 61.127243 |
| CBGB_4366 | Hypothetical protein | 13.426547 |
| CBGB_4365 | Hypothetical protein | 279.7435 |
| CBGB_4383 | Hypothetical protein | 4.496051 |
| CBGB_4383 | Hypothetical protein | 54.773762 |
| CBGB_10450 | Set/Thr protein kinase | 23.199852 |
| CBGB_4495 | Hypothetical protein | 4.496051 |
| CBGB_4372 | Hypothetical protein | 30.983101 |
| CBGB_4371 | Hypothetical protein | 62.656879 |
| CBGB_4270 | CENPA | 187.139 |
| CBGB_5663 | Histone H3 | 952.68 |
| CBGB_0187 | Tubulin beta chain | 2070.41 |
| CBGB_0491 | CENPA | 1354.83 |
| CBGB_6141 | Median expression of all genes on chromosome 9 | 96.8559645 |
| CBGB_6142 | Median expression of all genes on chromosome 10 | 96.29 |

### Table 4. Expression levels of genes located in the subtelomeric regions.

| Gene ID | Gene | Present at other side of chromosome |
|---------|------|-------------------------------------|
| CBGB_10308 | 1R | X |
| CBGB_3720 | 1R | X |
| CBGB_10000 | 1L | Present at other side of chromosome |
| CBGB_2934 | 1L | Present at other side of chromosome |
| CBGB_10211 | 4R | X |
| CBGB_10105 | 4R | X |
| CBGB_6144 | 4L | Present at other side of chromosome |
| CBGB_10291 | 4L | Present at other side of chromosome |
| CBGB_10214 | 7R | X |
| CBGB_10183 | 7R | X |
| CBGB_2029 | 7R | X |
| CBGB_2030 | 7R | X |
| CBGB_2031 | 7R | X |
| CBGB_2032 | 7R | X |
| CBGB_2033 | 7R | X |
| CBGB_2034 | 7R | X |
| CBGB_2036 | 7R | X |
| CBGB_2485 | 7L | Present at other side of chromosome |
| CBGB_2484 | 7L | Present at other side of chromosome |

Expression levels of genes where neocentromeres formed in the cen Δ mutants were analyzed in the R265 wild-type strain. Previously generated RNA sequencing data were remapped to the R265 reference genome, and the expression levels (FPKM) were analyzed for the native genes in each region wherein a neocentromere was formed in the cen Δ mutants. Expression analysis of several housekeeping genes was included for control purposes, and the median RNA expression level of all genes located on chromosomes 9 and 10 are listed. Expression levels of genes located in the subtelomeric regions were also analyzed.
## Supplementary Table 4. Neocentromeres are not enriched with transposable elements.

| Isolate      | Chromosome/scaffold | % Identical matches | Alignment length | # Mismatches | # Gap openings | Alignment in query Start (bp) | End (bp) | Alignment in subject Start (bp) | End (bp) | Evalue | Bitscore |
|--------------|---------------------|---------------------|------------------|--------------|----------------|-------------------------------|---------|---------------------------------|---------|--------|----------|
| WT           | scaffold3.10        | 100                 | 4537             | 0            | 0              | 1                             | 4537    | 115939                          | 120475  | 0      | 8379     |
| WT           | scaffold3.10        | 100                 | 3179             | 0            | 0              | 1                             | 3179    | 390605                          | 393783  | 0      | 5871     |
| cen10Δ-A     | NODE_39             | 100                 | 4537             | 0            | 0              | 1                             | 4537    | 17760                           | 13224   | 0      | 8379     |
| cen10Δ-A     | NODE_11             | 100                 | 3179             | 0            | 0              | 1                             | 3179    | 242270                          | 245448  | 0      | 5871     |
| cen10Δ-B     | NODE_42             | 100                 | 4537             | 0            | 0              | 1                             | 4537    | 17765                           | 13229   | 0      | 8379     |
| cen10Δ-C-L1  | NODE_42             | 100                 | 3179             | 0            | 0              | 1                             | 3179    | 250468                          | 253646  | 0      | 5871     |
| cen10Δ-C-L1  | NODE_10             | 100                 | 4537             | 0            | 0              | 1                             | 4537    | 115954                          | 120490  | 0      | 8379     |
| cen10Δ-C-L1  | NODE_10             | 100                 | 3179             | 0            | 0              | 1                             | 3179    | 250468                          | 253646  | 0      | 5871     |
| cen10Δ-C-L2  | NODE_59             | 100                 | 4537             | 0            | 0              | 1                             | 4537    | 61353                           | 65889   | 0      | 8379     |
| cen10Δ-C-L2  | NODE_26             | 100                 | 3179             | 0            | 0              | 1                             | 3179    | 242325                          | 245503  | 0      | 5871     |
| cen10Δ-E     | NODE_41             | 100                 | 4537             | 0            | 0              | 1                             | 4537    | 25951                           | 21415   | 0      | 8379     |
| cen10Δ-E     | NODE_9              | 100                 | 3179             | 0            | 0              | 1                             | 3179    | 242271                          | 245449  | 0      | 5871     |

To exclude the possibility that transposable elements were deposited into the neocentromeres, BlastN searches in a database with de novo genome assemblies of cen10Δ-A, cen10Δ-B, cen10Δ-C, and cen10Δ-E were performed. As input the homologous wild-type sequence of the chromosomal location of the neocentromeres was used. All neocentromeres of the tested cen10Δ mutants have the same length as the homologous sequence in the wild-type.
Supplementary Table 5. Genes located in subtelomeric regions of chromosomes 1, 4, 7, and 10.

| Subtelomeric region | Gene ID | Position of gene based on telomere | Chromosome location | Putative function | Present after chromosome fusion | C. neoformans homolog | Putative function of C. neoformans homolog | Synthetic | C. neoformans homolog present in knockout library |
|---------------------|---------|------------------------------------|---------------------|--------------------|-------------------------------|----------------------|---------------------------------------------|----------|-----------------------------------------------|
| 10R                 | CNBG_2029 | 1                                 | 7:1,489-1,740,783   | Hypothetical protein | Present at other side of chromosome | Uniprot: D. eugeniostriatus | Hypothetical protein |          |                                 |
| 10L                 | CNBG_6141 | 1                                 | 10:4,399-5,971      | Hypothetical protein | Present at other side of chromosome | X                     | Hypothetical protein |          |                                 |
| 7L                  | CNBG_1030 | 1                                 | 7:1,233,786-1,234,252 | Hypothetical protein | Present at other side of chromosome | X                     | Hypothetical protein |          |                                 |
| 7R                  | CNBG_1031 | 1                                 | 7:1,234,189-1,236,927 | Hypothetical protein | Present at other side of chromosome | X                     | Hypothetical protein |          |                                 |
| 4R                  | CNBG_1032 | 1                                 | 7:1,240,863-1,240,989 | Hypothetical protein | Present at other side of chromosome | X                     | Hypothetical protein |          |                                 |
| 4L                  | CNBG_1033 | 1                                 | 7:1,241,786-1,241,985 | Hypothetical protein | Present at other side of chromosome | X                     | Hypothetical protein |          |                                 |

We have indicated whether each gene located in the subtelomeric regions was lost in the chromosome fusion of the large-colony knockout library. For each gene in the subtelomeric region of chromosome 1, 4, 7, and 10 several character traits as the chromosomal location, putative function, and the presence of putative C. neoformans homologs in existing mutant libraries are indicated. There is no correlation between the loss of genes with or without predicted function.