RNAi is a critical determinant of centromere evolution in closely related fungi

Vikas Yadav[a], Sheng Sun[b], R. Blake Billmyre[b], Bhagya C. Thimmappa[a], Terrance Shea[c], Robert Lintner[c], Guus Bakkeren[d], Christina A. Cuomo[c], Joseph Heitman[b], and Kaustuv Sanyal[a,1]

[a]Molecular Biology and Genetics Unit, Jawaharlal Nehru Centre for Advanced Scientific Research, Bangalore 560064, India; [b]Department of Molecular Genetics and Microbiology, Duke University Medical Center, Durham, NC 27710; [c]Infectious Disease and Microbiome Program, Broad Institute of Massachusetts Institute of Technology and Harvard, Cambridge, MA 02142; and [d]Agriculture and Agri-Food Canada, Summerland Research and Development Centre, Summerland, BC V0H1Z0, Canada

Edited by Steven Henikoff, Fred Hutchinson Cancer Research Center, Seattle, WA, and approved February 11, 2018 (received for review August 3, 2017)

The centromere DNA locus on a eukaryotic chromosome facilitates faithful chromosome segregation. Despite performing such a conserved function, centromere DNA sequence as well as the organization of sequence elements is rapidly evolving in all forms of eukaryotes. The driving force that facilitates centromere evolution remains an enigma. Here, we studied the evolution of centromeres in closely related species in the fungal phylum of Basidiomycota. Using ChIP-seq analysis of conserved inner kinetochore proteins, we identified centromeres in three closely related *Cryptococcus* species: two of which are RNAi-proficient, while the other lost functional RNAi. We find that the centromeres in the RNAi-deficient species are significantly shorter than those of the two RNAi-proficient species. While centromeres are LTR retrotransposon-rich in all cases, the RNAi-deficient species lost all full-length retroelements from its centromeres. In addition, centromeres in RNAi-proficient species are associated with a significantly higher level of cytosine DNA modifications compared with those of RNAi-deficient species. Furthermore, when an RNAi-proficient *Cryptococcus* species and its RNAi-deficient mutants were passaged under similar conditions, the centromere length was found to be occasionally shortened in RNAi mutants. In silico analysis of predicted centromeres in a group of closely related *Ustilago* species, also belonging to the Basidiomycota, were found to have undergone a similar transition in the centromere length in an RNAi-dependent fashion. Based on the correlation found in two independent basidiomycetous species complexes, we present evidence suggesting that the loss of RNAi and cytosine DNA methylation triggered transposon attrition, which resulted in shortening of centromere length during evolution.

retrotransposons | *Cryptococcus* | *Ustilago* | experimental evolution | DNA methylation

The centromere is a specialized DNA locus that is required for assembly of a multiprotein complex, the kinetochore, which drives faithful chromosome segregation in eukaryotes. Centromeres can be classified as point centromeres of short DNA sequences of <400 bp (e.g., *Saccharomyces cerevisiae*), and regional centromeres which are long, and can range between a few kilobases (kb) (e.g., *Schizosaccharomyces pombe*) to several megabases (Mb) (e.g., humans and plants) (1, 2). The repetitive DNA present in regional centromeres (core and pericentromeres) consists of either arrays of satellite DNA or transposons or both. Transposons, despite being present in lower copy numbers than the satellites, are proposed to play a major role in regional centromere evolution (1, 3, 4). Transposon domestication at the centromere can also give rise to functional domains or repeats, including satellite DNA repeats in a centromere. The *dh*/*dg* repeats in the fission yeast or α-satellite repeats in humans are proposed to be the result of such domestication events (4). The presence of CENP-B, a centromere DNA binding protein that shows a high level of similarity with DNA transposons in humans and its homologs in other organisms, provides another line of evidence toward the role of transposons in the evolution of centromere structure and function (5, 6). Notably, a large number of transposons present in centromeres are RNA transposons or retrotransposons, which propagate through RNA intermediates in the "copy–paste" mode. These elements are different from DNA transposons, which excise themselves from the original site and move to a new target site in the genome, and thus propagate in the "cut–paste" mode (7). Retrotransposons, owing to their mode of propagation, have been proposed as architects of the regional centromeres (4). These elements can produce multiple copies of themselves and integrate into the same centromere locus or centromeres of other chromosomes (3, 4, 8, 9).

A retroelement must be transcribed to generate its RNA intermediates required for transposition. A low level of transcription from the centromere in a range of organisms, including budding yeast, fission yeast, mouse, humans, and maize, has been reported (10, 11). Studies on centromere function in various organisms suggest that neither a complete absence of transcription nor a high level of transcription at the centromere supports assembly of the multisubunit kinetochore complex on the centromere DNA (10, 12). Long noncoding centromeric transcript RNAs are shown to be required for loading of critical kinetochore proteins such as CENP-C (13, 14). In fission yeast, transcripts from *dh*/*dg* repeat regions of

Significance

The "centromere paradox" refers to rapidly evolving and highly diverse centromere DNA sequences even in closely related eukaryotes. However, factors contributing to this rapid divergence are largely unknown. Here, we identified large regional, LTR retrotransposon-rich centromeres in a group of human fungal pathogens belonging to the *Cryptococcus* species complex. We provide evidence that loss-of-functional RNAi machinery and possibly cytosine DNA methylation trigger instability of the genome by activation of centromeric retrotransposons presumably suppressed by RNAi. We propose that RNAi, together with cytosine DNA methylation, serves as a critical determinant that maintains repetitive transposon-rich centromere structures. This study explores the direct link between RNAi and centromere structure evolution.

Author contributions: V.Y., S.S., C.A.C., J.H., and K.S. designed research; V.Y., S.S., R.B.B., T.S., R.L., and C.A.C. performed research; V.Y., S.S., R.B.B., B.C.T., G.B., C.A.C., J.H., and K.S. analyzed data; and V.Y., S.S., G.B., C.A.C., J.H., and K.S. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The data reported in this paper (including ChIP-seq and PacBio sequencing) have been deposited under NCBI BioProject [accession no. PRJNA395628 (sequences, including ChIP-seq and PacBio sequencing) and PRJNA13691 (Nanopore data)].

[1]To whom correspondence should be addressed. Email: sanyal@jncasr.ac.in.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1713725115/-/DCSupplemental.

Published online March 5, 2018.

the centromere are targets of the RNA interference (RNAi) machinery and are required for heterochromatin formation (15, 16). Lack of proper heterochromatinization at *dh*/*dg* repeats causes abnormal centromere function leading to chromosome missegregation, although RNAi does not directly affect CENP-A loading in this organism (16). These results highlight the importance of transcription of centromeres. However, the level of transcription at the centromere must be regulated at an optimal level for proper kinetochore assembly and genome stability.

By regulating transcription from retrotransposons, present at the centromere or elsewhere in the genome, RNAi contributes to genome defense (17, 18). The functional RNAi pathway involves key proteins, including Dicer (Dcr), Argonaute (Ago), and RNA-dependent RNA polymerase (Rdp). The RNAi machinery targets double-stranded RNA being generated from retroelements/repeats leading to its degradation and in some cases promoting heterochromatinization through repressive histone modifications (such as H3K9 dimethylation) and/or DNA methylation (15, 19). Thus, by controlling expression of transposons at centromeres, RNAi could play an important regulatory role in the structural evolution of centromeres. However, except in *S. pombe*, no specific study has been conducted to explore a possible link between RNAi and centromere structure–function evolution (9).

The presence/absence of retrotransposons or repetitive DNA sequences can also contribute to rapid divergence of centromeres in a species complex. For example, centromeres in three related species of the *Schizosaccharomyces* clade, *S. pombe*, *Schizosaccharomyces japonicus*, and *Schizosaccharomyces octosporus*, exhibit divergence in the centromeric architecture (9). *S. japonicus* centromeres contain mostly transposons, whereas *S. pombe* and *S. octosporus* contain repeat-rich centromeres, and have lost most of the active transposons (9). Another group of fungi, the *Candida* species complex in the CTG clade, have been studied extensively with respect to centromere structure and function. *Candida albicans*, *Candida dubliniensis*, and *Candida lusitaniae* harbor repeatless centromeres where centromere DNA sequences are unique and different in each species (20–22). On the other hand, *Candida tropicalis* contains highly homogenized repeat-associated fission yeast-like centromeres (23). Both of these genera, *Schizosaccharomyces* and *Candida*, belong to the phylum Ascomycota of the fungal kingdom.

In this study, we identified centromeres in three closely related *Cryptococcus* species in the fungal phylum Basidiomycota. The three species, *Cryptococcus neoformans* (type strain H99), *Cryptococcus deneoformans* (reference isolate JEC21), and *Cryptococcus deuterogattii* (outbreak isolate R265), are haploid and diverged from a common ancestor as recently as 34 Mya (24–26). All three species are human pathogens and two of these primarily infect immunocompromised individuals, including HIV/AIDS patients, to cause cryptococcal meningitis, a leading cause of death in these patients. While *C. neoformans* and *C. deneoformans* are the most studied species in this complex, a recent rise in cases of drug resistance and infections in otherwise healthy individuals have attracted attention to the lesser explored species *C. deuterogattii* (27, 28). Unlike *C. neoformans* and *C. deneoformans*, *C. deuterogattii* lost key components of the functional RNAi machinery (29). Here, we show that centromeres in all three *Cryptococcus* species are regional centromeres featuring retroelements or their remnants. Notably, the RNAi-deficient *C. deuterogattii* species harbors significantly shorter centromeres compared with the RNAi-proficient species *C. neoformans* or *C. deneoformans*. We also predicted centromere locations in the *Ustilago* species complex, a group of plant pathogens in the Basidiomycota, and observed a similar correlation between the loss of RNAi and shortening of putative centromeres. Overall, this study provides mechanistic insights into the structural evolution of centromeres in an RNAi-dependent manner.

Results and Discussion

***Cryptococcus* Species Harbor Large Regional Centromeres.** We previously reported the probable structure of centromeres in *C. neoformans* as long ORF-free regions that are rich in retrotransposable elements (25). In addition, centromeres in *C. deneoformans* were predicted as a single transposon-rich locus on each of the 14 chromosomes (26). In this study, we sought to identify the centromeres experimentally in three pathogenic species: *C. neoformans*, *C. deneoformans*, and *C. deuterogattii*, of the *Cryptococcus* species complex, including verifying centromere locations predicted earlier in *C. neoformans* and *C. deneoformans*.

To achieve this goal, two of the evolutionarily conserved inner kinetochore proteins, CENP-A and CENP-C, were identified in each of the three species (*SI Appendix*, Fig. S1 *A* and *B*). Both of these proteins were tagged with mCherry and showed localization patterns as reported previously (Fig. 1*A* and *SI Appendix*, Fig. S1*C*) (30). To identify functional centromeres in *C. neoformans*, we performed CENP-A and CENP-C chromatin immunoprecipitation (ChIP) followed by deep sequencing (ChIP-seq). ChIP-seq analysis revealed overlapping binding of both proteins at a single locus on each of the 14 chromosomes of *C. neoformans* (*SI Appendix*, Fig. S2 and Table S1). The binding regions of both proteins were largely overlapping and these regions spanned across the ORF-free, poorly transcribed regions on most chromosomes (Fig. 1*B* and *SI Appendix*, Fig. S3). The minor variations observed between the binding sites of these two proteins could be due to different sequencing approaches employed. While no specific patterns could be detected in CENP-A/CENP-C binding across these regions, occasional dips in the binding pattern of both proteins were observed (*SI Appendix*, *SI Materials and Methods*). The length of the CENP-A– and CENP-C–bound regions varied from 20 kb to 40 kb in *C. neoformans* unlike centromeres of many other fungi where the CENP-A–bound region remains nearly constant across chromosomes (2). A similar CENP-A binding pattern across the entire stretch of the centromere was found in *Neurospora crassa* (31). It is important to note here that centromeres in *C. neoformans* were not completely assembled, as they contain a few sequence gaps in the current assembly. We attempted to close these gaps by PacBio as well as Sanger sequencing (*SI Appendix*, *SI Materials and Methods*). This resulted in a significant improvement of the genome assembly as we could close the sequence gaps in 11 of 14 centromeres, leaving only *CEN3*, *CEN11*, and *CEN14* with some sequence gaps. While this is the best assembly obtained thus far for the *C. neoformans* genome, the presence of additional breaks/gaps cannot be excluded.

The *C. neoformans* genome shares a high level of gene synteny with that of *C. deneoformans* (*SI Appendix*, Fig. S4*A* and Table S2) (25, 32). By performing synteny analysis across centromere flanking regions of *C. neoformans*, we were able to predict putative centromeres in *C. deneoformans*. The predicted regions were large, ORF-free, poorly transcribed, and map to loci previously predicted as centromeres (*SI Appendix*, Figs. S3 and S4*B* and Table S1) (26). CENP-C ChIP-qPCR confirmed the authenticity of each of these regions as functional centromeres in all 14 chromosomes of *C. deneoformans* (Fig. 1*C*). Compared with the noncentromeric locus, we obtained significant enrichment of centromeric CENP-C binding using two pairs of primers located distantly on each of the 14 centromeres.

We performed PacBio and Nanopore sequencing for the *C. deuterogattii* genome and generated a complete, chromosome-level de novo genome assembly, improving on the previous assembly of 26 scaffolds (Fig. 2*A* and *SI Appendix*, *SI Materials and Methods*). A comparison of the *C. deuterogattii* genome with that of *C. neoformans* revealed a number of rearrangements between the two species (*SI Appendix*, Fig. S4*C* and Table S2). It was previously predicted that the *C. neoformans* and *C. deuterogattii* genomes have undergone an arm exchange involving chromosomes 1 and 2 (25). The new long-read assembly provides compelling evidence supporting this rearrangement. Next, we performed CENP-C ChIP-seq in *C. deuterogattii* and analyzed the data using the chromosomal assembly and obtained 14 binding peaks, one on each chromosome (Fig. 2*B* and *SI Appendix*, Fig. S4*D* and Table S1). Similar to *C. neoformans* and *C. deneoformans*, all

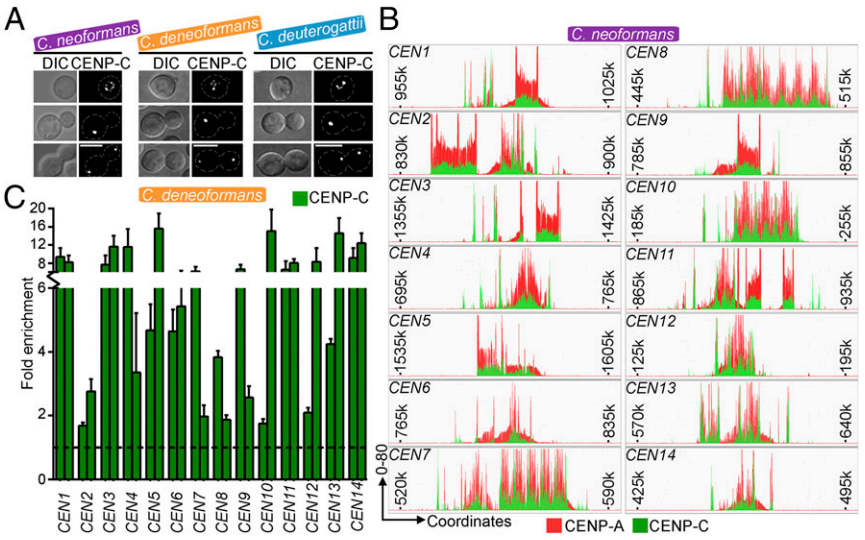

Fig. 1. The *Cryptococcus* species complex has large regional-type centromeres. (*A*) The subcellular localization patterns of a conserved kinetochore protein CENP-C at various cell cycle stages (interphase, prometaphase, and anaphase) in *C. neoformans*, *C. deneoformans*, and *C. deuterogattii*. (Scale bar, 5 μm.) (*B*) Overlapping binding of CENP-A (red) and CENP-C (green) on each of the 14 chromosomes of *C. neoformans* as revealed by ChIP-seq analysis of these proteins. An 80-kb DNA sequence harboring the centromere is shown for each of the 14 chromosomes. (*C*) CENP-C (mCherry)-ChIP-qPCR analysis confirmed enrichment of CENP-C on the predicted centromeres in *C. deneoformans*. Fold enrichment was normalized to a non-*CEN* region, the level of which is marked by the dotted line in the graph. Error bars represent SEM.

14 peaks mapped to ORF-free poorly transcribed regions in the *C. deuterogattii* genome (*SI Appendix*, Fig. S3).

Sequence analysis of the CENP-A/CENP-C–bound ORF-free centromere regions in all three species revealed the presence of retrotransposons Tcn1–Tcn6 (Fig. 3*A*) (33). While Tcn1–Tcn5 belong to the Ty3-*gypsy* family of retroelements, Tcn6 belongs to the Ty1-*copia* family of retroelements. The retroelements have a high degree of conservation (70% or more identity) across the three species (Dataset S1). However, they differ in their sequence from the centromeric retroelements of a closely related nonpathogenic *Cryptococcus* species, *Cryptococcus amylolentus* (*SI Appendix*, Fig. S5*A* and Dataset S1) (34). It is important to note that Tcn1–Tcn6 elements are not found in the *C. amylolentus* genome, indicating that these elements are specific to the pathogenic *Cryptococcus* species. Based on these results, we conclude that centromeres in pathogenic *Cryptococcus* species are large regional-type, and enriched in the same retroelements. These results suggest that neither the location nor the sequence elements of the centromeres have diverged significantly among these closely related fungal species. However, further analysis revealed that the centromeres in *C. deuterogattii* are significantly shorter, with an average length of around 14 kb compared with those of *C. neoformans* or *C. deneoformans*, which have an average length of 44 kb and 62 kb, respectively (Fig. 3*B*). Consequently, the CENP-C–bound regions were also found to be shortened in *C. deuterogattii* (5–15 kb) compared with *C. neoformans* (20–40 kb). Thus, we conclude that the *Cryptococcus* species have CENP-A/CENP-C–rich regional centromeres of varying lengths, but the centromere DNA sequence elements are highly conserved.

Loss of the RNAi Machinery and Shortening of Centromeres Are Correlated in *C. deuterogattii*. Ago, Dcr, and Rdp—the key proteins of the RNAi machinery—are all present in *C. neoformans* and *C. deneoformans* but all are absent in *C. deuterogattii* (29). We previously demonstrated that RNAi suppresses transcription and transposition of retrotransposons, including the Tcn elements in the RNAi-proficient species *C. neoformans* and *C. deneoformans* (35, 36). Real-time PCR analysis of two of the Tcn elements confirmed that the expression levels of these elements are higher in *C. deuterogattii* compared with *C. neoformans* (*SI Appendix*, Fig. S5*B*). To further investigate alterations in the centromere length observed, we performed a detailed analysis of the retrotransposons, Tcn1–Tcn6, in all three species. These retroelements are specifically enriched and mostly restricted (>95%) to centromeric regions in all three species (Fig. 3*C* and *SI Appendix*, Figs. S2 and S4). The most striking observation is that the centromeres in *C. neoformans* and *C. deneoformans* harbor a significant proportion of full-length retroelements (20–30%), whereas the *C. deuterogattii* centromeres are completely devoid of such full-length elements (Fig. 3*D*). Instead, retroelements present in *C. deuterogattii* centromeres are only remnants of transposons (Fig. 3*E*). They lack one or more of the essential domains (LTRs, reverse transcriptase, integrase) required for transposition activity, rendering them nonfunctional for further transposition. Thus, this analysis reveals that loss of RNAi in *C. deuterogattii* is correlated with the loss of full-length retrotransposons and an overall reduction in the length of the centromeres.

By comparison with *C. neoformans*, the genome of *C. deuterogattii* seems to have lost additional genes besides key enzymes in the RNAi pathway (24, 29, 37). These additional lost genes include genes involved in protein processing, protein degradation, and the mitochondrial oxidation pathway. In addition, we found

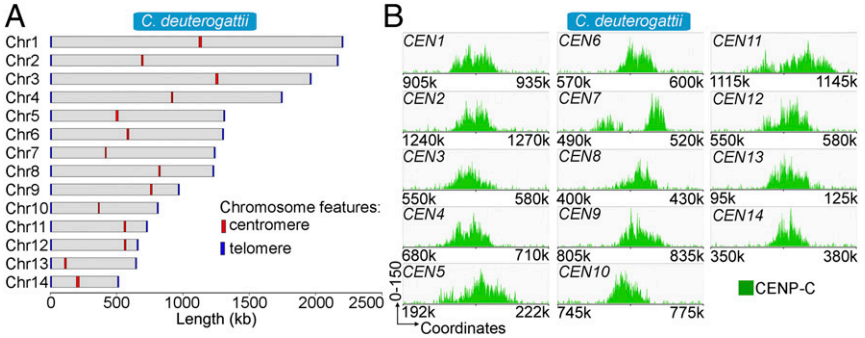

Fig. 2. ChIP sequencing and PacBio sequencing map centromeres on each of the 14 chromosomes in *C. deuterogattii*. (*A*) Map showing the 14 chromosomes in *C. deuterogattii* with the telomeres and centromeres marked. The centromeres are marked to scale, while the telomeres are marked as 10-kb regions for visualization purposes. (*B*) CENP-C (mCherry)-ChIP-seq identified 14 binding regions among 14 chromosomes in *C. deuterogattii*'s latest assembly. A 30-kb region spanning the CENP-C–bound region is shown for each chromosome.

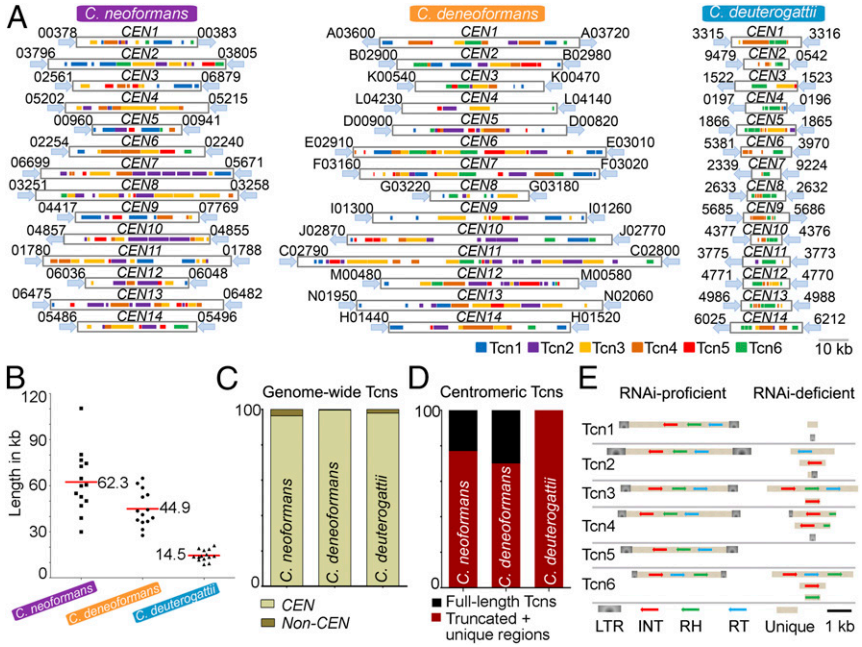

Fig. 3. Centromeres are enriched with retrotransposons in the *Cryptococcus* species complex. (*A*) The presence of various retrotransposons across the centromeres in *C. neoformans*, *C. deneoformans*, and *C. deuterogattii*. The numbers refer to centromere-flanking ORFs which are preceded by "CNAG_" for *C. neoformans*, "CN" for *C. deneoformans*, and "CNBG_" for *C. deuterogattii*. The diagrams are drawn to scale. (*B*) The length of each centromere of the respective species was plotted. Each dot represents one centromere, and the horizontal red line depicts the mean centromere length of the corresponding species. (*C*) Distribution of retrotransposons, Tcn1–Tcn6, in centromeres and across the genome in *C. neoformans*, *C. deneoformans,* and *C. deuterogattii*. (*D*) A bar diagram showing the distribution of full-length versus truncated Tcn elements at the centromeres in all three species. (*E*) Comparison of retrotransposon elements present in centromeres of RNAi-proficient (*C. neoformans* and *C. deneoformans*) and RNAi-deficient (*C. deuterogattii*) species. INT, integrase; LTR, long terminal repeat; RH, RNaseH; RT, reverse transcriptase; Unique, unique DNA sequence in each retroelement.

that the sole DNA methyltransferase encoding gene, *DNMT5*, is truncated in *C. deuterogattii* (*SI Appendix*, Fig. S6*A*). A recent study reported that putative centromeres of *C. neoformans* are methylated at the DNA level (38). Combining the available bisulfite sequencing results with our ChIP-seq results revealed that the centromere DNA sequence of *C. neoformans* is indeed extensively methylated (*SI Appendix*, Fig. S2). Incidentally, due to the presence of a truncated form of *DNMT5*, cytosine DNA methylation at centromeres is found to be absent in *C. deuterogattii* (*SI Appendix*, Fig. S6 *B* and *C*). The lack of centromere DNA methylation in *C. deuterogattii* was further supported by base modification analysis based on PacBio single-molecule real-time (SMRT) sequencing (Fig. 4*A* and *SI Appendix*, Fig. S6*D*). While *C. neoformans* centromeres harbor extensive base modifications, little or no methylation is associated with *C. deuterogattii* centromeres. Previous reports suggest that the DNA methylation suppresses recombination as well as regulates retrotransposition in various organisms (39, 40).

To strengthen support for our hypothesis on the roles of RNAi in centromere evolution, we also identified centromeres in another basidiomycete species complex, the *Ustilago* species complex. Similar to the *Cryptococcus* species complex, the *Ustilago* species complex harbors three species of which *Ustilago maydis* lost all three major proteins of the RNAi machinery along with the DNA methyltransferase while *Ustilago hordei* and *Ustilago bromivora* harbor the complete machinery (41, 42). In *Cryptococcus*, centromeres are present in large ORF-free and poorly transcribed regions of the genome. The centromeres in three *Cryptococcus* species are also syntenic with each other. Keeping in mind these centromeric features, we predicted centromeres in all three species of the *Ustilago* complex by in silico analysis and found that the RNAi-deficient species *U. maydis* possesses shorter centromeres (average length 12 kb) compared with both of the RNAi-proficient species *U. hordei* (average length 36 kb) and *U. bromivora* (average length 27 kb) (Fig. 4*B* and *SI Appendix*, Fig. S7 and Table S3). The centromere locations identified in *U. maydis* matched well with the ones predicted previously (43). Earlier studies have reported that *U. maydis* harbors fewer transposons than *U. hordei* and *U. bromivora* (41, 42). These observations further confirmed the correlation between loss of RNAi and a reduction in centromere length. We further extended our analysis to fungal species harboring regional centromeres, and in which RNAi status is known, and found that species with all components of the RNAi machinery harbor longer centromeres than the species that lack one or more of the RNAi components (Fig. 4*C* and *SI Appendix*, Table S4).

Rather than a single optimized DNA sequence, centromeres are among the most rapidly changing loci in the genome despite having a conserved and essential function, a phenomenon termed the "centromere paradox" (44). A number of studies verified that centromeres are evolving rapidly, even among very closely related species of plants and animals (3, 8, 45). This process of rapid evolution of centromeres is well studied in fungal species, especially among the members of the Ascomycota (9, 23, 46). The nature of centromeres in another significant fungal phylum, the Basidiomycota, was unknown. In this report, we identified centromeres in three pathogenic species of the *Cryptococcus* species complex. Using multiple sequencing and analysis methods, we significantly improved the genome assembly for *C. neoformans* and obtained a full genome-wide, chromosome-level assembly for *C. deuterogattii*. The three *Cryptococcus* species also provided an opportunity to test the role of RNAi in centromere evolution. We discovered that centromeres in the RNAi-deficient species are shorter compared with the RNAi-proficient species. Analysis of the centromere DNA sequence revealed that the RNAi-deficient species possesses fewer and truncated retrotransposons compared with the RNAi-proficient species. Taken together, our study reveals that centromeres evolve rapidly among closely related species in the Basidiomycota phylum of fungi.

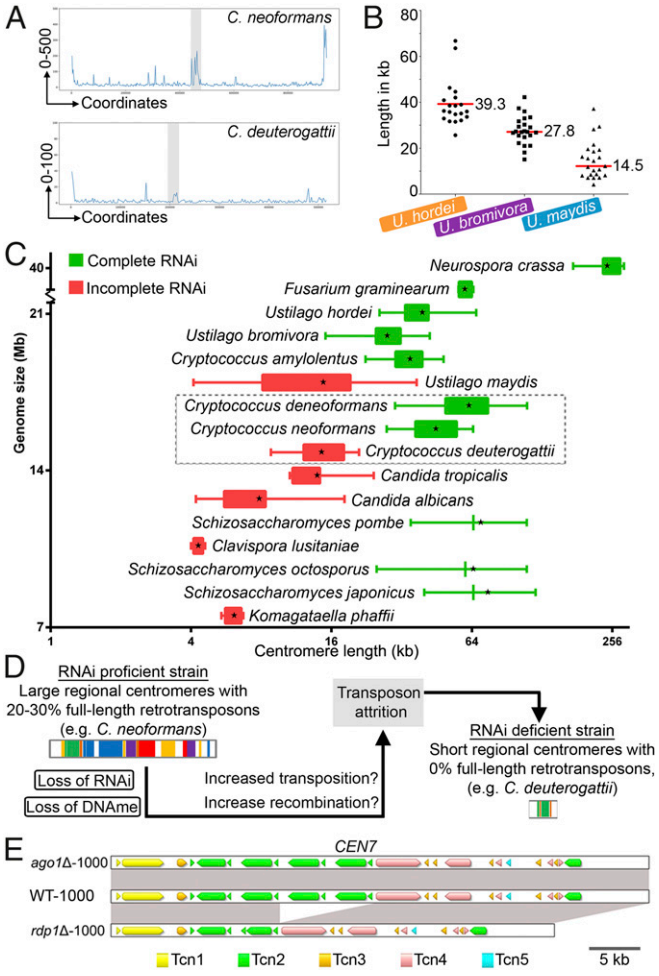

Fig. 4. RNAi as a key determinant of longer centromeres in closely related fungi. (*A*) Base modification analysis based on SMRT PacBio sequencing revealed a high level of methylation in *C. neoformans CEN14* (depicted as gray-shaded region) but not in *C. deuterogattii CEN14* (see *SI Appendix*, Fig. S6*D* for remaining centromeres). (*B*) Comparison of the predicted centromere length in *U. maydis*, *U. bromivora*, and *U. hordei*. Only 20 predicted centromeres are plotted for *U. hordei*, while all 23 are shown for both *U. maydis* and *U. bromivora*. Each dot represents one centromere, and the horizontal line depicts the mean value. (*C*) Graph showing the correlation between centromere (ORF-free region) lengths and status of RNAi among the fungal species. The star in each of the boxes represents the mean value, the boxes depict the range from the 25th percentile to 75th percentile values, and the terminal vertical lines mark the range of centromere lengths. Species lacking any one of the three key proteins (Ago, Dcr, or Rdp) of the RNAi machinery were considered to harbor incomplete RNAi machinery for this analysis (*SI Appendix*, Table S4). (*D*) Possible sequence of events that might have occurred due to loss of RNAi machinery and/or DNA methylation in an RNAi-proficient strain (*C. neoformans* or *C. deneoformans*) that led to a genome with truncated retrotransposons that are unable to transpose in an RNAi-deficient strain (*C. deuterogattii*). (*E*) PacBio sequencing results of strains passaged for 1,000 doublings revealed reduction in *CEN7* length in the *rdp1*Δ mutant but not in *C. neoformans* wild-type or the *ago1*Δ strains.

The majority of the siRNA in *C. neoformans* maps to centromeric retrotransposons, the level of which drops in RNAi-defective mutants of *C. neoformans* (36, 47). The key proteins required for a functional RNAi machinery were found to be absent in *C. deuterogattii* (29). We propose that loss of RNAi in this species might have led to amplification of retroelements, which in turn would have integrated into the centromere-proximal sites, leading to a transient elongation of the centromeres. Consequently, the presence of retroelements in close vicinity to each other might have enhanced the rate of recombination between these elements. Recombination would cause shortening of these regions and might render retrotransposons inactive. Further, the absence of RNAi along with cytosine DNA methylation can contribute to enhancing the rate of recombination between these elements (39, 40, 48). It is notable that both transposition and recombination can be damaging to the genome if they are associated with loss of essential genes. Thus, cells that have no essential genes inactivated after transposition and recombination would survive. Because the partial loss of a centromere may not affect its function (49, 50), the probability of such events being tolerated at the centromere locus may be higher. Thus, RNAi-deficient strains with a stabilized genome will be likely to have shorter centromeres with truncated retroelements, similar to those of *C. deuterogattii* or *U. maydis*, while the intermediate population during evolution might have possessed centromeres of varying length (Fig. 4*D*).

We attempted to test our hypothesis by performing experimental evolution experiments. *C. neoformans* and its derived RNAi mutants were grown for 1,000 doublings under standard laboratory conditions (*SI Appendix*, Fig. S8 and *SI Materials and Methods*). The passaged strains did not show any obvious growth defects as measured by their generation time (*SI Appendix*, Table S5). PacBio sequencing of wild-type and mutant strains (both passaged and unpassaged) revealed two rearrangements in the centromeric regions of RNAi mutants compared with the wild-type grown under similar conditions (*SI Appendix*, Table S6). *CEN7* of *rdp1*Δ mutant exhibited a reduction in length in the 1,000 doubling passaged strains compared with the 1,000 doubling wild-type and unpassaged strains (Fig. 4*E*). In addition, both unpassaged and passaged RNAi mutants carried a shorter *CEN2* compared with the wild type (*SI Appendix*, Fig. S8 and Table S6). These results suggest that centromeres are prone to structural alterations in the absence of RNAi. Overall, our study provides evidence of RNAi in maintaining the structure of retrotransposon-rich centromeres in fungi. Experimental evolution in RNAi as well as cytosine DNA methylation mutants under conditions that favor retrotransposon expression, such as during the sexual cycle, will be tested in the future to gain further insight into this process.

Transposons play a major role in shaping the evolution of genomes, including the centromere, in multiple ways (18, 51). It has been proposed that the centromeric repeats present in fission yeast, maize, and the α-satellite repeats in human centromeres evolved from transposable elements (4). Transposons also have been shown to play an active role in centromere evolution among closely related species in plants (3). Based on a study in the *Schizosaccharomyces* group, it was proposed that loss of transposons observed in *S. pombe* centromeres occurred as a result of recombination between LTR elements that remained present in the *S. japonicus* genome (9). The loss of retrotransposons was also correlated with a shift in transposon regulation from RNAi to the CENP-B homolog in *S. pombe*, Cbp1 (5). In this study, we show structural changes in centromeres mediated by retrotransposons. Centromeres in the RNAi-proficient species *C. neoformans* and *C. deneoformans* harbor full-length retrotransposable elements, whereas RNAi-deficient *C. deuterogattii* centromeres contain only footprints of these elements. We propose that the truncation of retrotransposons could have occurred due to recombination among retroelements rendering them inactive. One such example is reported in angiosperms where retrotransposon truncation via illegitimate recombination resulted in genome size reduction (52, 53). Loss of full-length retroelements can prove to be advantageous for *C. deuterogattii* as a pathogen. Host conditions might induce retrotransposition and damage a pathogen's genome despite suppression of active transposition by RNAi in an RNAi-proficient species (54, 55). While loss of RNAi in such situations can be lethal due to unregulated transposition, truncation and inactivation of full-length retroelements can provide selective advantages. The *C. deuterogattii* genome is 1.3 Mb smaller compared with *C. neoformans* and centromere shortening accounts for approximately one-third of this reduction (total centromere lengths being >630 kb in

C. neoformans versus 203 kb in *C. deuterogattii*). A relatively faster generation time (*SI Appendix*, Table S5) of *C. deuterogattii* compared with *C. neoformans* or *C. deneoformans* may be due to the smaller genome of the former. Thus, we speculate that loss of full-length retroelements triggered by the loss of RNAi in *C. deuterogattii* could provide a pathogenic advantage over its related RNAi-proficient species in addition to other factors (27). There is as yet no established direct correlation between pathogenesis and the loss of RNAi along with full-length retrotransposon. The *Cryptococcus* species complex might prove to be a good model system with which to address such questions.

Materials and Methods

The strains and primers used in this study are listed in *SI Appendix*, Tables S7 and S8, respectively. *Cryptococcus* was grown in YPD (1% yeast extract, 2% peptone, and 2% dextrose) media at 30 °C. *Cryptococcus* cells were transformed using biolistics. Transformants were selected on YPD agar media containing 200 μg/mL of G418 (Sigma-Aldrich) or hygromycin (Invitrogen). Details of all of the experimental procedures and sequence analysis are given in *SI Appendix, SI Materials and Methods*. All of the sequencing data (including ChIP-seq and PacBio sequencing) have been deposited under NCBI BioProject accession no. PRJNA395628 and the Nanopore data under PRJNA13691. The reference number for each study is provided in *SI Appendix*, Table S9.

ACKNOWLEDGMENTS. We thank Genotypic Technology Private Limited (Bangalore, India) for *C. neoformans* CENP-A and CENP-C ChIP-seq library preparation and raw sequence data generation reported in this publication; the Duke University core sequencing facility; the University of North Carolina core sequencing facility; and the Broad Technology Labs. V.Y. is a Senior Research Fellow, supported by the Council of Scientific and Industrial Research, Government of India. K.S. is a Tata Innovation Fellow and is also supported by intramural funding from the Jawaharlal Nehru Centre for Advanced Scientific Research. S.S., R.B.B., and J.H. are supported by NIH/National Institute of Allergy and Infectious Diseases (NIAID) R37 Merit Award AI39115-20 and R01 award AI50113-13. C.A.C. is supported by NIH/NIAID Grant U19AI110818 to the Broad Institute.

1. Brown JD, O'Neill RJ (2014) *The Evolution of Centromeric DNA Sequences. eLS* (John Wiley & Sons, Hoboken, NJ).
2. Roy B, Sanyal K (2011) Diversity in requirement of genetic and epigenetic factors for centromere function in fungi. *Eukaryot Cell* 10:1384–1395.
3. Gao D, Jiang N, Wing RA, Jiang J, Jackson SA (2015) Transposons play an important role in the evolution and diversification of centromeres among closely related species. *Front Plant Sci* 6:216.
4. Wong LH, Choo KH (2004) Evolutionary dynamics of transposable elements at the centromere. *Trends Genet* 20:611–616.
5. Cam HP, Noma K, Ebina H, Levin HL, Grewal SI (2008) Host genome surveillance for retrotransposons by transposon-derived proteins. *Nature* 451:431–436.
6. Mateo L, González J (2014) *Pogo*-like transposases have been repeatedly domesticated into CENP-B-related proteins. *Genome Biol Evol* 6:2008–2016.
7. Goodier JL, Kazazian HH, Jr (2008) Retrotransposons revisited: The restraint and rehabilitation of parasites. *Cell* 135:23–35.
8. Comai L, Maheshwari S, Marimuthu MPA (2017) Plant centromeres. *Curr Opin Plant Biol* 36:158–167.
9. Rhind N, et al. (2011) Comparative functional genomics of the fission yeasts. *Science* 332:930–936.
10. Chan FL, Wong LH (2012) Transcription in the maintenance of centromere chromatin identity. *Nucleic Acids Res* 40:11178–11188.
11. Hall LE, Mitchell SE, O'Neill RJ (2012) Pericentric and centromeric transcription: A perfect balance required. *Chromosome Res* 20:535–546.
12. Pezer Z, Ugarković D (2008) Role of non-coding RNA and heterochromatin in aneuploidy and cancer. *Semin Cancer Biol* 18:123–130.
13. Du Y, Topp CN, Dawe RK (2010) DNA binding of centromere protein C (CENPC) is stabilized by single-stranded RNA. *PLoS Genet* 6:e1000835.
14. Scott KC (2013) Transcription and ncRNAs: At the cent(rome)re of kinetochore assembly and maintenance. *Chromosome Res* 21:643–651.
15. Volpe T, Martienssen RA (2011) RNA interference and heterochromatin assembly. *Cold Spring Harb Perspect Biol* 3:a003731.
16. Pidoux AL, Allshire RC (2005) The role of heterochromatin in centromere function. *Philos Trans R Soc Lond B Biol Sci* 360:569–579.
17. Dumesic PA, Madhani HD (2014) Recognizing the enemy within: Licensing RNA-guided genome defense. *Trends Biochem Sci* 39:25–34.
18. Fedoroff NV (2012) Presidential address. Transposable elements, epigenetics, and genome evolution. *Science* 338:758–767.
19. Matzke MA, Mosher RA (2014) RNA-directed DNA methylation: An epigenetic pathway of increasing complexity. *Nat Rev Genet* 15:394–408.
20. Padmanabhan S, Thakur J, Siddharthan R, Sanyal K (2008) Rapid evolution of Cse4p-rich centromeric DNA sequences in closely related pathogenic yeasts, *Candida albicans* and *Candida dubliniensis*. *Proc Natl Acad Sci USA* 105:19797–19802.
21. Kapoor S, Zhu L, Froyd C, Liu T, Rusche LN (2015) Regional centromeres in the yeast *Candida lusitaniae* lack pericentromeric heterochromatin. *Proc Natl Acad Sci USA* 112:12139–12144.
22. 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 USA* 101:11374–11379.
23. Chatterjee G, et al. (2016) Repeat-associated fission yeast-like regional centromeres in the ascomycetous budding yeast *Candida tropicalis*. *PLoS Genet* 12:e1005839.
24. D'Souza CA, et al. (2011) Genome variation in *Cryptococcus gattii*, an emerging pathogen of immunocompetent hosts. *mBio* 2:e00342-10.
25. Janbon G, et al. (2014) Analysis of the genome and transcriptome of *Cryptococcus neoformans* var. *grubii* reveals complex RNA expression and microevolution leading to virulence attenuation. *PLoS Genet* 10:e1004261.
26. Loftus BJ, et al. (2005) The genome of the basidiomycetous yeast and human pathogen *Cryptococcus neoformans*. *Science* 307:1321–1324.
27. Bielska E, May RC (2016) What makes *Cryptococcus gattii* a pathogen? *FEMS Yeast Res* 16:fov106.
28. Dixit A, Carroll SF, Qureshi ST (2009) *Cryptococcus gattii*: An emerging cause of fungal disease in North America. *Interdiscip Perspect Infect Dis* 2009:840452.
29. Feretzaki M, Billmyre RB, Clancey SA, Wang X, Heitman J (2016) Gene network polymorphism illuminates loss and retention of novel RNAi silencing components in the *Cryptococcus* pathogenic species complex. *PLoS Genet* 12:e1005868.
30. Kozubowski L, et al. (2013) Ordered kinetochore assembly in the human-pathogenic basidiomycetous yeast *Cryptococcus neoformans*. *mBio* 4:e00614-e13.
31. 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.
32. Sun S, Xu J (2009) Chromosomal rearrangements between serotype A and D strains in *Cryptococcus neoformans*. *PLoS One* 4:e5524.
33. Goodwin TJ, Poulter RT (2001) The diversity of retrotransposons in the yeast *Cryptococcus neoformans*. *Yeast* 18:865–880.
34. Sun S, et al. (2017) Fungal genome and mating system transitions facilitated by chromosomal translocations involving intercentromeric recombination. *PLoS Biol* 15:e2002527.
35. Janbon G, et al. (2010) Characterizing the role of RNA silencing components in *Cryptococcus neoformans*. *Fungal Genet Biol* 47:1070–1080.
36. Wang X, et al. (2010) Sex-induced silencing defends the genome of *Cryptococcus neoformans* via RNAi. *Genes Dev* 24:2566–2582.
37. Farrer RA, et al. (2015) Genome evolution and innovation across the four major lineages of *Cryptococcus gattii*. *mBio* 6:e00868-15.
38. Huff JT, Zilberman D (2014) Dnmt1-independent CG methylation contributes to nucleosome positioning in diverse eukaryotes. *Cell* 156:1286–1297.
39. Maloisel L, Rossignol JL (1998) Suppression of crossing-over by DNA methylation in *Ascobolus*. *Genes Dev* 12:1381–1389.
40. Mirouze M, et al. (2012) Loss of DNA methylation affects the recombination landscape in *Arabidopsis*. *Proc Natl Acad Sci USA* 109:5880–5885.
41. Laurie JD, et al. (2012) Genome comparison of barley and maize smut fungi reveals targeted loss of RNA silencing components and species-specific presence of transposable elements. *Plant Cell* 24:1733–1745.
42. Rabe F, et al. (2016) A complete toolset for the study of *Ustilago bromivora* and *Brachypodium* sp. as a fungal-temperate grass pathosystem. *eLife* 5:e20522.
43. Kämper J, et al. (2006) Insights from the genome of the biotrophic fungal plant pathogen *Ustilago maydis*. *Nature* 444:97–101.
44. Henikoff S, Ahmad K, Malik HS (2001) The centromere paradox: Stable inheritance with rapidly evolving DNA. *Science* 293:1098–1102.
45. Melters DP, et al. (2013) Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution. *Genome Biol* 14:R10.
46. Bensasson D, Zarowiecki M, Burt A, Koufopanou V (2008) Rapid evolution of yeast centromeres in the absence of drive. *Genetics* 178:2161–2167.
47. Dumesic PA, et al. (2013) Stalled spliceosomes are a signal for RNAi-mediated genome defense. *Cell* 152:957–968.
48. Ellermeier C, et al. (2010) RNAi and heterochromatin repress centromeric meiotic recombination. *Proc Natl Acad Sci USA* 107:8701–8705.
49. Mravinac B, et al. (2009) Histone modifications within the human X centromere region. *PLoS One* 4:e6602.
50. Steiner NC, Clarke L (1994) A novel epigenetic effect can alter centromere function in fission yeast. *Cell* 79:865–874.
51. Chuong EB, Elde NC, Feschotte C (2017) Regulatory activities of transposable elements: From conflicts to benefits. *Nat Rev Genet* 18:71–86.
52. Devos KM, Brown JK, Bennetzen JL (2002) Genome size reduction through illegitimate recombination counteracts genome expansion in *Arabidopsis*. *Genome Res* 12:1075–1079.
53. Vitte C, Panaud O (2005) LTR retrotransposons and flowering plant genome size: Emergence of the increase/decrease model. *Cytogenet Genome Res* 110:91–107.
54. Horváth V, Merenciano M, González J (2017) Revisiting the relationship between transposable elements and the eukaryotic stress response. *Trends Genet* 33:832–841.
55. Nicolás FE, Torres-Martínez S, Ruiz-Vázquez RM (2013) Loss and retention of RNA interference in fungi and parasites. *PLoS Pathog* 9:e1003089.