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

Sundar Ram Sankaranarayanan[1], Giuseppe Ianiri[2], Md. Hashim Reza[1], Bhagya C. Thimmappa[1,#], Promit Ganguly[1], Marco A. Coelho[2], Sheng Sun[2], Rahul Siddharthan[3], Christian Tellgren-Roth[4], Thomas L Dawson Jr.[5,6], Joseph Heitman[2,*] and Kaustuv Sanyal[1,*]

[1]Molecular Mycology Laboratory, Molecular Biology and Genetics Unit, Jawaharlal Nehru Center for Advanced Scientific Research, Jakkur P.O, Bengaluru- 560064.

[2]Department of Molecular Genetics and Microbiology, Duke University Medical Center, Durham, NC 27710, USA

[3]The Institute of Mathematical Sciences/HBNI, Taramani, Chennai 600 113, India

[4]National Genomics Infrastructure, Science for Life Laboratory, Department of Immunology, Genetics and Pathology, Uppsala University, 75108 Uppsala, Sweden

[5]Skin Research Institute Singapore, Agency for Science, Technology and Research (A*STAR), 138648, Singapore

[6]Medical University of South Carolina, School of Pharmacy, Department of Drug Discovery, 29425, USA

[#] Present address: Department of Biochemistry, Robert-Cedergren Centre for Bioinformatics and Genomics, University of Montreal, 2900 Edouard-Montpetit, Montreal, H3T1J4, QC, Canada

E-mail address of authors

SRS (sundar_ram@jncasr.ac.in), GI (giuseppe.ianiri@duke.edu), MHR (hashimreza@jncasr.ac.in), BCT (bhagyathimmappa@gmail.com), PG (promitganguly1@gmail.com), MAC (marco.dias.coelho@duke.edu), SS (sheng.sun@duke.edu), RS (rsidd@imsc.res.in), CTR (christian.tellgren@igp.uu.se), TLD (thomas_dawson@sris.a-star.edu.sg)

*Corresponding authors

Joseph Heitman (heitm001@duke.edu)

Kaustuv Sanyal (sanyal@jncasr.ac.in)

Classification

Biological sciences, Genetics

Keywords

Centromere loss, chromosome fusion, skin microbe, merotelic attachment, double strand breaks, kinetochore.

Abstract

Intra-chromosomal or inter-chromosomal genomic rearrangements often lead to speciation (1). Loss or gain of a centromere leads to alterations in chromosome number in closely related species. Thus, centromeres can enable tracing the path of evolution from the ancestral to a derived state (2). The *Malassezia* species complex of the phylum Basiodiomycota shows remarkable diversity in chromosome number ranging between six and nine chromosomes (3-5). To understand these transitions, we experimentally identified all eight centromeres as binding sites of an evolutionarily conserved outer kinetochore protein Mis12/Mtw1 in *M. sympodialis*. The 3 to 5 kb centromere regions share an AT-rich, poorly transcribed core region enriched with a 12 bp consensus motif. We also mapped nine such AT-rich centromeres in *M. globosa* and the related species *Malassezia restricta* and *Malassezia slooffiae*. While eight predicted centromeres were found within conserved synteny blocks between these species and *M. sympodialis*, the remaining centromere in *M. globosa* (*MgCEN2)* or its orthologous centromere in *M. slooffiae (MslCEN4*) and *M. restricta* (*MreCEN8*) mapped to a synteny breakpoint compared with *M. sympodialis*. Taken together, we provide evidence that breakage and loss of a centromere (*CEN2*) in an ancestral *Malassezia* species possessing nine chromosomes resulted in fewer chromosomes in *M. sympodialis*. Strikingly, the predicted centromeres of all closely related *Malassezia* species map to an AT-rich core on each chromosome that also shows enrichment of the 12 bp sequence motif. We propose that centromeres are fragile AT-rich sites driving karyotype diversity through breakage and inactivation in these and other species.

Significance statement

The number of chromosomes can vary between closely related species. Centromere loss destabilizes chromosomes and results in reduced number of chromosomes to drive speciation. A series of evidence from studies on various cancers suggest that an imbalance in kinetochore-microtubule attachments results in breaks at the centromeres. To understand if such events can cause chromosome number changes in nature, we studied six species of *Malassezia*, of which three possess eight chromosomes and others have nine chromosomes each. We find signatures of chromosome breakage at the centromeres in organisms having nine chromosomes. We propose that the break at the centromere followed by fusions of acentric chromosomes to other chromosomes could be a plausible mechanism shaping the karyotype of *Malassezia* and related organisms.

Introduction

Centromeres are the genomic loci on which the kinetochore, a multi-subunit complex, assembles to facilitate high fidelity chromosome segregation. The centromere-specific histone H3 variant CENP-A is the epigenetic hallmark of centromeres, and acts as the foundation to recruit other proteins of the kinetochore. A series of evidence suggest that centromeres are species specific and are among the most rapidly evolving regions in the genome even between closely related species (6-9). This is accompanied by concomitant evolution of CENP-A and the associated kinetochore proteins (10). Functional incompatibilities between centromeres result in uniparental genome elimination in interspecies hybrids (11, 12). The divergent nature of centromere is proposed to be a driving force for speciation (13, 14).

Recent studies show that asexual organisms, also by virtue of inter- and intra-chromosomal rearrangements, diversify into species clusters that are distinct in genotype and morphology (15). These genotypic differences include changes in both, the organization and the number of chromosomes. Centromere function is directly related to stabilization of the karyotype when a change in chromosome number occurs (2). Rearrangements in the form of telomere-telomere fusions and nested chromosome insertions (NCIs) are some of the major sources of chromosome number reduction (16). Such cases result in the formation of dicentric chromosomes that are subsequently stabilized by breakage-bridge-fusion cycles (17) or via inactivation of one of the centromeres through other mechanisms (18, 19). Well known examples of telomere-telomere fusion include, the formation of extant human chromosome 2 by fusion of two ancestral chromosomes (20), the reduction in karyotype seen within members of Saccharomycotina such as *Candida glabrata*, *Vanderwaltozyma polyspora*, *Kluyveromyces lactis*, and *Zygosaccharomyces rouxii* (2), and the exchange of chromosomal arms seen in plants (21). NCIs have predominantly shaped the karyotype evolution in grasses (22). Unlike the above cases, chromosome number reduction can also be driven by centromere loss.

The centromere-kinetochore complex is the chromosomal attachment site of spindle microtubules and experiences extensive physical forces through kinetochore-microtubule attachments during chromosome segregation. DNA double strand breaks (DSBs) have been mapped at the centromeres when improper kinetochore-microtubule attachments especially merotelic attachments remain unresolved (23-25). Such observations in various cancers are suggestive of centromere fission to be a potential source of chromosome breakage and aneuploidy (reviewed in 26). Merotelic attachments are common during normal cell division and occasionally such improper attachments remain undetected even in normal cells (27, 28).

To investigate if centromere break can be a natural source of karyotype diversity in closely related species, we sought to identify centromeres in a group of *Malassezia* species that show variations in

chromosome number. *Malassezia* species are lipid-dependent basidiomycetous fungi that are a part of animal skin microbiome (29). The electrophoretic karyotype of some of these species is known and the chromosome number ranges between six and nine chromosomes (4, 5). This species complex includes 18 *Malassezia* species that are divided into three clades. Fungemia-associated species like *Malassezia furfur* belong to clade A, common inhabitants of skin such as *Malassezia sympodialis* form clade B, and clade C includes *Malassezia slooffiae* as the basal species that diverged from the common ancestor (30, 31). Strikingly, most of these species have a compact genome of less than nine megabases in size. In this study, we experimentally validated all the eight centromeres of *M. sympodialis.* We traced the transition between karyotype of 8 and 9 chromosomes within clade B by predicting centromeres in five other *Malassezia* species carrying either eight or nine chromosomes. Based on our results, we propose that the event of centromere break can act as a potential source of karyotype diversity and speciation in asexually propagating organisms.

Results

The predicted centromeres of *M. sympodialis* that maps to the GC troughs on each chromosome (32), resemble the AT-richness of CDEII of point centromeres in *Saccharomyces cerevisiae*. Organisms with point centromeres possess the CBF3 complex, a cognate protein complex specific to point centromeres (33). None of the point centromere-specific proteins could be detected in *M. sympodialis* (Table S1). We could however detect homologs of CENP-A, CENP-C, and most of the outer kinetochore proteins (Figure 1A and Table S1). Strikingly, none of the Ctf19 complex homologs that form the Constitutively Centromere Associated Network (CCAN) could be found (Table S1) suggesting at a loss of this protein complex in *M. sympodialis* as observed in other basidiomycetes such as *Cryptococcus neoformans* (34).

Kinetochores cluster and localize to the nuclear periphery in *M. sympodialis*

We functionally expressed an N-terminally GFP-tagged Mtw1 protein (Protein ID: SHO76526.1) from its native promoter and expression was confirmed by western blotting (Figure 1B). Upon staining with anti-GFP antibodies and DAPI (4′,6-diamidino-2-phenylindole), we could detect punctate localization of Mtw1 at the nuclear periphery (Figure 1C) consistent with the clustered kinetochore localization in other yeast species (35-37). Live-cell images of MSY001 (GFP-*MTW1*) cells show that the kinetochores (GFP-Mtw1) remain clustered throughout the cell cycle, starting from unbudded G1 cells in interphase to large budded cells in mitosis (Figure 1D).

Mtw1 localized to a single region at the GC minima of each *M. sympodialis* chromosome

To identify the centromeres, we performed ChIP-sequencing using the GFP-Mtw1 strain of *M. sympodialis* MSY001. Mapping the reads to the reference genome of *M. sympodialis* strain ATCC42132 (32) revealed one significantly enriched locus on each of the eight chromosomes (Figure 2A). The length of the Mtw1-enriched centromere regions identified from the ChIP-seq analysis range between 3167 bp and 5143 bp with an average length of 4165 bp (Table S2). However, the region of maximum enrichment (based on the number of sequenced reads aligned) mapped to the intergenic region harboring the GC trough (Figure S1A-B). The regions of enrichment we observed overlap and span the GC troughs that are predicted to be the centromeres in *M. sympodialis*, including the active genes located proximal to these troughs (Figure S1B). However, these ORFs do not show consensus in features such as direction of transcription or functional classification. We validated this enrichment by ChIP-qPCR analysis using primers homologous to the predicted centromeres and a control region distant from the centromere (Figure 2B).

Histone H3 is depleted at the core centromere with active genes at the pericentric regions in *M. sympodialis*

The presence of CENP-A nucleosomes should result in reduced histone H3 enrichment at the centromeres. To test this, we performed ChIP with anti-H3 antibodies and analyzed the immunoprecipitated (IP) DNA by qPCR. As compared to a control ORF region (190 kb away from *CEN1*), the pericentric regions flanking the core centromeres showed a marginal reduction in histone H3 enrichment which was further reduced at the core that maps to the GC minima with the highest enrichment of the kinetochore protein. That the core centromere region showing the maximum depletion of histone H3 coincided with the regions most enriched with Mtw1 further supports that histone H3, in these regions, is possibly replaced by its centromere specific variant CENP-A (Figure S1D).

Centromere sequences share a 12 bp long AT-rich consensus sequence motif

To understand the features of *M. sympodialis* centromeres, we analyzed the centromere DNA sequences for the presence of consensus motifs or structures such as inverted repeats. PhyloGibbs-MP (38, 39) predicted a 12 bp long AT-rich motif common to all of the centromere sequences of *M. sympodialis* (Figure 2C). We swept the Position Weight Matrix (PWM) across each chromosome of *M. sympodialis* and calculated the average log-likelihood ratio (LLR) in 500 bp sliding windows spaced at 100 bp, averaged over each 12 bp substring of each window. The LLR is the natural logarithm of the ratio of the likelihood of the 12 bp substring arising as a sample from the PWM, to the likelihood of it being generic "background". In each case, the global peak coincides with the centromere. This suggests that the AT-rich motif is more enriched at the centromeres than at any other region in the genome (Figure S3A). We

also searched for specific matches to the motif, genome-wide, by looking for site matches with a LLR greater than 7.5. There is one global peak per 500 bp window per chromosome, again matching the centromeres. A dot-plot aligning all the 10 kb bins containing centromere sequences against themselves generated using SyMap (40) further confirmed a lack of direct/inverted repeat structures (Figure S1E). The only unique feature of all eight centromere sequences is the presence of an AT-rich core (average AT content of 78% as compared to genome average of 41.5%) along with the 12 bp motif and a uniform kinetochore protein-bound region of 3 to 5 kb.

Chromosome number variation in *Malassezia* species Clade B

The basal species *M. slooffiae* (clade C) was reported to have 9 chromosomes (5). Within clade B, *M. globosa* and *M. restricta* were reported to have 9 chromosomes and *M. sympodialis* shown to have 8 chromosomes (4, 32). To validate the karyotype of *M. globosa* and *M. slooffiae*, we assembled their genomes *de novo* using long reads generated by PacBio sequencing technology. The *M. globosa* genome is a complete assembly in 9 chromosomes. We assigned each band on the pulsed field gel based on the sizes from our genome assembly and confirmed by Southern hybridization that chromosome 5 containing the rDNA locus (renamed as ChrR) migrates along with chromosome 3, higher than the expected size of 902 kb (Figure 3A, S2A). The assembled genome of *M. slooffiae* has 14 contigs of which 9 contigs have telomeres on both ends indicative of 9 chromosomes. The sizes from the genome assembly could be assigned to the bands observed in the pulsed field gel (Figure S2B). To further confirm the assigned karyotype and elucidate the basis of changes observed in the karyotype in more detail, we sought to map the centromeres of *M. globosa*, *M. slooffiae* and *M. restricta*.

CEN2 maps to a synteny breakpoint in *M. globosa*

Similar to *M. sympodialis* centromeres, we predicted the distinct GC minimum in each of the nine chromosomes of *M. globosa, M. slooffiae,* and *M. restricta* as the centromeres. The 12 bp AT-rich motif identified in *M. sympodialis* centromeres were enriched in these predicted centromeres as well (Figure 2C and S3B-D). The gene order and synteny across centromeres is largely conserved within closely related species (8, 41, 42). We mapped these regions in the context of gene synteny between each of these species and *M. sympodialis*. From this analysis, we find 8 GC troughs in *M. globosa* and *M. slooffiae* that share synteny with centromere regions of *M. sympodialis* (Figure 3B-C), supporting the fact that these regions are candidate centromeres in the corresponding species. In *M. restricta*, we find 7 predicted centromeres that are completely syntenous with *M. sympodialis* centromeres and one other centromere where partial synteny is maintained (Figure S4). However, a lack of synteny conservation was observed

in the centromeres of Chr2 in *M. globosa*, scaffold 4 in *M. slooffiae*, and scaffold 8 in *M. restricta* (Figure S4).

The GC trough corresponding to *MgCEN2* is flanked by genes that map to MsChr2 on one arm and MsChr4 on the other (Figure 3D). This putative centromere region marks a synteny breakpoint showing no homology in the *M. sympodialis* genome indicative of a loss of this centromere. We also observed that the genes flanking the breakpoint are conserved in *M. sympodialis* suggesting that the specific intergenic region was involved (Figure S2B). Evidence for internalization of telomere adjacent ORFs or presence of interstitial telomere repeats indicative of telomere-telomere fusions were not detected in the *M. sympodialis* genome. These observations strongly support our hypothesis that breakage of *CEN2* (or the orthologous ancestral *CEN*) and fusion of the two acentric arms to other chromosomes resulted in the chromosome number reduction observed between these species.

To map the first common ancestor to have experienced this break, we analyzed the regions flanking *CEN2* of *M. globosa*, and the centromeres of scaffold 4 of *M. slooffiae* and scaffold 8 of *M. restricta*. Conservation of synteny between *M. globosa* and *M. slooffiae* chromosomes at this locus suggests that the last common ancestor species contained 9 chromosomes (Figure 4A, yellow and blue circles). In scaffold 8 of *M. restricta*, the gene order is maintained with the centromere on one side while the genes on the other side were rearranged to a region 220 kb away from the centromere on the same scaffold consistent with further recombination at the locus. Besides *M. sympodialis*, *Malassezia nana*, *Malassezia caprae*, *Malassezia equina*, and *Malassezia dermatis* are some other species that form the clade B (Figure 4A). We identified putative centromeres using gene synteny and GC troughs in *M. nana* and *M. dermatis* because of their relatively better assembled genomes distributed in 13 and 18 scaffolds respectively. In both of these species, we detected 8 centromeres that share complete synteny conservation with *M. sympodialis* centromeres indicative of 8 chromosomes in these species (Figure S4 and Table S3). This was further supported by the enrichment of the 12 bp AT-rich motif at these predicted centromeres (Figure S3E-F). Based on this synteny analysis, we propose that centromere breakage would have occurred in the common ancestor with 9 chromosomes giving rise to an 8-chromosome karyotype that was inherited by *M. sympodialis* and related species.

Sequence conservation at centromeres

The orthologous centromeres in *M. sympodialis* and *M. globosa* were aligned using FSA (43), and 391 orthologous intergenic regions, identified with neighboring gene orthology and synteny, were similarly aligned. We found no significant difference in conservation rate between centromeric sequence and other intergenic sequence.

Discussion

A distinguishing feature of fungal centromeres/kinetochores is their clustered localization in the vicinity of SPBs at the nuclear periphery (44). Kinetochores in ascomycetes are clustered throughout the cell cycle with the exception of *Schizosaccharomyces pombe* and *Zymoseptoria tritici* (45, 46). The kinetochore proteins in the basidiomycetous yeast *C. neoformans* assemble in a stepwise manner and transit between unclustered and clustered states during interphase and mitosis respectively (47). The dynamics of the outer kinetochore in *M. sympodialis* reveals two features that distinguish the kinetochore assembly from another basidiomycete yeast *C. neoformans*: (a) the kinetochores are clustered across different stages of budding/cell cycle, and (b) the presence of clustered kinetochores across all cell cycle stages is suggestive of a constitutively assembled kinetochore. At the composition level, we could not detect the homologs of Constitutive Centromere Associated Network (CCAN) proteins similar to the observations in *C. neoformans*. With the advent of genetic manipulations (48-50), *Malassezia* species serve as an unconventional model system to understand structural and function diversity of kinetochores.

Most yeast species having single nucleosome-length long point centromeres have smaller genomes (< 14 Mb). While most *Malassezia* species including *M. sympodialis* and *M. globosa* have genomes that are highly compact (< 9 Mb), the centromeres are 3 to 5 kb in length. These small regional centromeres are distinct from the large, transposon-associated repetitive centromeres of the *Cryptococcus* species complex, the only other known basidiomycete centromere (42, 51). However, centromeres of similar lengths are reported in ascomycetous yeasts that include various *Candida* species (8, 52-54). Inverted repeats flanking the CENP-A enriched core have been reported to promote *de novo* centromere formation and subsequent stabilization of a replicative plasmid in *S. pombe, C. tropicalis,* and *Komagataella phaffii* (52, 55-58). No such structures were detected in the *M. sympodialis* and *M. globosa* centromeres (Figure S1E, S1F). Within the 3 to 5 kb long centromere, the region showing maximum Mtw1 enrichment mapped to the intergenic region containing the AT-rich centromere core. This is suggestive of a bipartite structure comprised of (a) a CDEII -like AT-rich core that shows maximum kinetochore enrichment and (b) the flanking pericentric regions that show basal levels of enrichment. Centromeres are known to be among the most rapidly evolving genomic loci (7-9). Strikingly, the centromeres identified in *M. sympodialis* show no signs of enhanced sequence divergence when compared to other orthologous intergenic regions in *M. globosa*. Adding to this, the 12 bp AT-rich consensus motif is enriched at the centromeres across different *Malassezia* species. While the functional significance of this conservation is unknown, it will be intriguing to test the role of the 12 bp centromere-specific AT-rich motif, the core, and the flanking sequence in centromere function. Testing these domains for centromere function in vivo by making centromeric plasmids in various *Malassezia* species is challenging due to technical limitations.

Besides *M. sympodialis*, *M. pachydermatis* and *M. furfur*, no other *Malassezia* species have been successfully transformed (48-50). Even in all the above cases, the genetic manipulations are performed by *Agrobacterium*-mediated transconjugation, which is not suitable for inserts to be maintained as circular plasmids. However, based on the conservation at the centromeres across species in this study (representative of Clade B and C), the 12 bp AT-rich motif can be a potential signature of centromeres in the *Malassezia* species complex.

The transcriptional state of chromatin has been reported to be a key determinant of centromere identity (59-64). Apart from heterochromatic histone marks, DNA marks such as cytosine DNA methylation are also enriched in *N. crassa* and *C. neoformans* centromeres (42, 65). The centromeres in *M. sympodialis* contain many ORFs that are transcribed (Figure S1B). The presence of transcribing ORFs has been documented earlier in centromeres of rice, maize and *Z. tritici* (46, 66, 67). Unlike these cases, our read count analysis did not reveal any significant difference in the transcription (RPKM values) of centromeric ORFs to that elsewhere in the genome (Figure S1C). In line with these results, homologs of the proteins commonly involved in heterochromatin formation such as Clr4, members of the RNAi complex, Swi6 and Dnmt5 could not be detected in *M. sympodialis*.

In this study, we provide evidence for loss of a centromere by breakage resulting in a karyotype change between two closely placed group of species, the species with 9 chromosomes such as *M. globosa* and the ones with 8 chromosomes including *M. sympodialis*. Synteny breakpoints adjacent to the centromeres have been reported in *C. tropicalis* that has seven chromosomes - one less than that of *C. albicans* (52). Centromere loss by breakage in the pre-WGD ancestor was proposed to have reduced the *Ashbya gossypii* karyotype by one (2). Breakpoints of conserved synteny between mammalian and chicken chromosomes were also mapped to the centromeres (68). Similar consequences in the karyotype have been reported in cases where centromeres were experimentally excised. Besides neocentromere formation, survival by fusion of acentric chromosome arms has been shown in *S. pombe* (69). This strongly suggests that any compromise in centromere function has a direct role in shaping chromosome structure and karyotype of organisms.

What is the driving force for the karyotypic diversity observed within the *Malassezia* species? The centromere are the primary attachment sites for microtubules. Dysregulated mitotic spindles in the form of unchecked merotelic attachments has been implicated in the generation of intra-mitotic DSBs at centromeres, indicated by the accumulation of γH2AX in a majority of mammalian cells (23). This fragility is more pronounced in cases where AT-rich DNA is present. Studies of human fragile site FRA16D show that the AT-rich DNA (Flex1) results in fork stalling as a consequence of cruciform structure formation (70). The AT-rich core centromere sequence in *M. globosa* is also predicted to form

secondary structures. The replication fork stalling commonly observed in centromeres (71-73) can result in accumulation of single stranded DNA, providing impetus to form such secondary structures. In conjunction with unresolved merotelic attachments, these can result in DSBs at the centromere. Chromosomal breakage and aneuploidy results when cells fail to rectify these defects, as seen in cancers (74). In mammals, centromeric DSBs are repaired efficiently compared to regions elsewhere in the genome largely due to the presence of several homology tracts in the form of repetitive DNA sequences or the stiffness provided by the inherent heterochromatic state to facilitate ligation (75). In the absence of pericentric heterochromatin in *Malassezia* species, how efficiently NHEJ might repair the centromeric DSBs is not known. Merotelic attachments routinely occur in normal cells and are corrected early in mitosis prior to anaphase by various means (24, 27). When unchecked, this results in DSBs that elicit a stop anaphase signal by activation of the mitotic checkpoint (76). This delay facilitates the tension sensor protein Aurora B kinase-mediated detachment of microtubules from the kinetochore and establishment of proper attachments (77). Besides these, additional mechanisms such as the monopolin complex-mediated recruitment of condensins have been proposed to suppress merotely in organisms that lack pericentric heterochromatin (78-81). We could not detect homologs for any proteins of the monopolin complex in *Malassezia* except Csm1. Based on these lines of evidence, we propose that unresolved merotelic attachments could cause breaks at the centromere resulting in the diverse karyotypes seen within the *Malassezia* species complex (Figure 4B).

Materials and Methods

The strains, plasmids and primers used in this study are mentioned in the *SI appendix*. *Malassezia* strains were grown on modified-Dixon's media (Malt extract 36 g/L, Desiccated oxbile 20 g/L, Tween40 10 mL/L, Peptone 6g/L, Glycerol 2mL/L, Oleic acid 2.89 mL/L). *M. sympodialis* and *M. globosa* strains were grown at 30°C and 32°C respectively. *M. sympodialis* was transformed by *Agrobacterium* mediated transconjugation. All experimental procedures and sequence analysis are described in detail in *SI Materials and Methods*. The Mtw1 ChIP sequencing reads and the genome sequences assemblies of *M. globosa* and *M. slooffiae* reported in this paper have been deposited under NCBI BioProject (Accession number PRJNA509412).

Acknowledgments

We thank Clevergene Biocorp Pvt. Ltd., Bengaluru for generating the Mtw1 ChIP-sequencing data. S.R.S is a Senior Research Fellow supported by intramural funding from JNCASR. M.H.R is a National Postdoctoral Fellow (PDF/2016/002858), supported by the Science and Engineering Research Board (SERB), Department of Science and Technology (DST), Government of India. K.S is a Tata Innovation

Fellow (grant number BT/HRT/35/01/03/2017), is supported by a grant for Life Science Research, Education and Training (BT/INF/22/SP27679/2018) of Department of Biotechnology, Govt. of India and intramural funding from JNCASR. Studies are supported in part by NIH/NIAID R37 award AI39115-21 and R01 award AI50113-15 to J.H. T.L.D acknowledges the A* STAR Industry alignment fund H18/01a0/016, Asian Skin Microbiome Program. We thank the members of K.S lab and J.H lab for valuable discussions and comments during bi-weekly Skype meetings.

Author contributions

J.H, and K.S conceived and secured funding for the study. S.R.S, G.I, M.H.R, and S.S performed the experiments. R.S performed centromere sequence conservation analysis and identified the motif reported in this study. S.R.S, M.H.R, B.C.T, P.G, and M.D.C performed all the other bioinformatic analysis. T.L.D, and C.T.R sequenced and assembled the genomes of *M. globosa* and *M. slooffiae*. S.R.S and K.S wrote the manuscript with inputs from all the authors.

Declaration of Interests

The authors declare no competing interests.

References

1. Coghlan A, Eichler EE, Oliver SG, Paterson AH, & Stein L (2005) Chromosome evolution in eukaryotes: a multi-kingdom perspective. *Trends Genet* 21(12):673-682.
2. Gordon JL, Byrne KP, & Wolfe KH (2011) Mechanisms of chromosome number evolution in yeast. *PLoS Genet* 7(7):e1002190.
3. Theelen B*, et al.* (2018) *Malassezia* ecology, pathophysiology, and treatment. *Med Mycol* 56(suppl_1):S10-S25.
4. Boekhout T & Bosboom RW (1994) Karyotyping of *Malassezia* Yeasts - Taxonomic and Epidemiologic Implications. *Syst Appl Microbiol* 17(1):146-153.
5. Boekhout T, Kamp M, & Gueho E (1998) Molecular typing of *Malassezia* species with PFGE and RAPD. *Med Mycol* 36(6):365-372.
6. Roy B & Sanyal K (2011) Diversity in requirement of genetic and epigenetic factors for centromere function in fungi. *Eukaryot Cell* 10(11):1384-1395.
7. Bensasson D, Zarowiecki M, Burt A, & Koufopanou V (2008) Rapid evolution of yeast centromeres in the absence of drive. *Genetics* 178(4):2161-2167.
8. 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 U S A* 105(50):19797-19802.
9. Rhind N*, et al.* (2011) Comparative functional genomics of the fission yeasts. *Science* 332(6032):930-936.
10. Talbert PB, Bryson TD, & Henikoff S (2004) Adaptive evolution of centromere proteins in plants and animals. *J Biol* 3(4):18.
11. Ravi M & Chan SW (2010) Haploid plants produced by centromere-mediated genome elimination. *Nature* 464(7288):615-618.

12. Sanei M, Pickering R, Kumke K, Nasuda S, & Houben A (2011) Loss of centromeric histone H3 (CENH3) from centromeres precedes uniparental chromosome elimination in interspecific barley hybrids. *Proc Natl Acad Sci U S A* 108(33):E498-505.
13. Henikoff S, Ahmad K, & Malik HS (2001) The centromere paradox: stable inheritance with rapidly evolving DNA. *Science* 293(5532):1098-1102.
14. Malik HS & Henikoff S (2009) Major evolutionary transitions in centromere complexity. *Cell* 138(6):1067-1082.
15. Barraclough TG, Birky CW, Jr., & Burt A (2003) Diversification in sexual and asexual organisms. *Evolution* 57(9):2166-2172.
16. Lysak MA (2014) Live and let die: centromere loss during evolution of plant chromosomes. *New Phytol* 203(4):1082-1089.
17. McClintock B (1941) The Stability of Broken Ends of Chromosomes in Zea Mays. *Genetics* 26(2):234-282.
18. Han F, Gao Z, & Birchler JA (2009) Reactivation of an inactive centromere reveals epigenetic and structural components for centromere specification in maize. *Plant Cell* 21(7):1929-1939.
19. Sato H, Masuda F, Takayama Y, Takahashi K, & Saitoh S (2012) Epigenetic inactivation and subsequent heterochromatinization of a centromere stabilize dicentric chromosomes. *Curr Biol* 22(8):658-667.
20. JW IJ, Baldini A, Ward DC, Reeders ST, & Wells RA (1991) Origin of human chromosome 2: an ancestral telomere-telomere fusion. *Proc Natl Acad Sci U S A* 88(20):9051-9055.
21. Schubert I & Lysak MA (2011) Interpretation of karyotype evolution should consider chromosome structural constraints. *Trends Genet* 27(6):207-216.
22. Murat F*, et al.* (2010) Ancestral grass karyotype reconstruction unravels new mechanisms of genome shuffling as a source of plant evolution. *Genome Res* 20(11):1545-1557.
23. Guerrero AA*, et al.* (2010) Centromere-localized breaks indicate the generation of DNA damage by the mitotic spindle. *Proc Natl Acad Sci U S A* 107(9):4159-4164.
24. Guerrero AA, Martinez AC, & van Wely KH (2010) Merotelic attachments and non-homologous end joining are the basis of chromosomal instability. *Cell Div* 5:13.
25. Cimini D*, et al.* (2001) Merotelic kinetochore orientation is a major mechanism of aneuploidy in mitotic mammalian tissue cells. *J Cell Biol* 153(3):517-527.
26. Martinez AC & van Wely KH (2011) Centromere fission, not telomere erosion, triggers chromosomal instability in human carcinomas. *Carcinogenesis* 32(6):796-803.
27. Cimini D, Moree B, Canman JC, & Salmon ED (2003) Merotelic kinetochore orientation occurs frequently during early mitosis in mammalian tissue cells and error correction is achieved by two different mechanisms. *J Cell Sci* 116(Pt 20):4213-4225.
28. Rupa DS, Hasegawa L, & Eastmond DA (1995) Detection of chromosomal breakage in the 1cen-1q12 region of interphase human lymphocytes using multicolor fluorescence in situ hybridization with tandem DNA probes. *Cancer Res* 55(3):640-645.
29. Saunders CW, Scheynius A, & Heitman J (2012) Malassezia fungi are specialized to live on skin and associated with dandruff, eczema, and other skin diseases. *PLoS Pathog* 8(6):e1002701.
30. Wu G*, et al.* (2015) Genus-wide comparative genomics of *Malassezia* delineates its phylogeny, physiology, and niche Adaptation on human skin. *PLoS Genet* 11(11):e1005614.
31. Lorch JM*, et al.* (2018) *Malassezia vespertilionis* sp. nov.: a new cold-tolerant species of yeast isolated from bats. *Persoonia-Molecular Phylogeny and Evolution of Fungi* 41::56-70.
32. Zhu Y*, et al.* (2017) Proteogenomics produces comprehensive and highly accurate protein-coding gene annotation in a complete genome assembly of *Malassezia sympodialis*. *Nucleic Acids Res* 45(5):2629-2643.

33. Meraldi P, McAinsh AD, Rheinbay E, & Sorger PK (2006) Phylogenetic and structural analysis of centromeric DNA and kinetochore proteins. *Genome Biol* 7(3):R23.
34. van Hooff JJ, Tromer E, van Wijk LM, Snel B, & Kops GJ (2017) Evolutionary dynamics of the kinetochore network in eukaryotes as revealed by comparative genomics. *EMBO Rep* 18(9):1559-1571.
35. Euskirchen GM (2002) Nnf1p, Dsn1p, Mtw1p, and Nsl1p: a new group of proteins important for chromosome segregation in *Saccharomyces cerevisiae*. *Eukaryot Cell* 1(2):229-240.
36. Goshima G, Saitoh S, & Yanagida M (1999) Proper metaphase spindle length is determined by centromere proteins Mis12 and Mis6 required for faithful chromosome segregation. *Genes Dev* 13(13):1664-1677.
37. Roy B, Burrack LS, Lone MA, Berman J, & Sanyal K (2011) CaMtw1, a member of the evolutionarily conserved Mis12 kinetochore protein family, is required for efficient inner kinetochore assembly in the pathogenic yeast *Candida albicans*. *Mol Microbiol* 80(1):14-32.
38. Siddharthan R (2008) PhyloGibbs-MP: module prediction and discriminative motif-finding by Gibbs sampling. *PLoS Comput Biol* 4(8):e1000156.
39. Siddharthan R, Siggia ED, & van Nimwegen E (2005) PhyloGibbs: a Gibbs sampling motif finder that incorporates phylogeny. *PLoS Comput Biol* 1(7):e67.
40. Soderlund C, Bomhoff M, & Nelson WM (2011) SyMAP v3.4: a turnkey synteny system with application to plant genomes. *Nucleic Acids Res* 39(10):e68.
41. Byrne KP & Wolfe KH (2005) The Yeast Gene Order Browser: combining curated homology and syntenic context reveals gene fate in polyploid species. *Genome Res* 15(10):1456-1461.
42. Yadav V*, et al.* (2018) RNAi is a critical determinant of centromere evolution in closely related fungi. *Proc Natl Acad Sci U S A* 115(12):3108-3113.
43. Bradley RK*, et al.* (2009) Fast statistical alignment. *PLoS Comput Biol* 5(5):e1000392.
44. Duan Z*, et al.* (2010) A three-dimensional model of the yeast genome. *Nature* 465(7296):363-367.
45. Liu X, McLeod I, Anderson S, Yates JR, 3rd, & He X (2005) Molecular analysis of kinetochore architecture in fission yeast. *EMBO J* 24(16):2919-2930.
46. Schotanus K*, et al.* (2015) Histone modifications rather than the novel regional centromeres of *Zymoseptoria tritici* distinguish core and accessory chromosomes. *Epigenetics Chromatin* 8:41.
47. Kozubowski L*, et al.* (2013) Ordered kinetochore assembly in the human-pathogenic basidiomycetous yeast *Cryptococcus neoformans*. *mBio* 4(5):e00614-00613.
48. Ianiri G, Averette AF, Kingsbury JM, Heitman J, & Idnurm A (2016) Gene function analysis in the ubiquitous human commensal and pathogen *Malassezia* Genus. *mBio* 7(6).
49. Celis AM*, et al.* (2017) Highly efficient transformation system for *Malassezia furfur* and *Malassezia pachydermatis* using *Agrobacterium tumefaciens*-mediated transformation. *J Microbiol Methods* 134:1-6.
50. Ianiri G, Applen Clancey S, Lee SC, & Heitman J (2017) FKBP12-dependent inhibition of calcineurin mediates immunosuppressive antifungal drug action in *Malassezia*. *mBio* 8(5).
51. Sun S*, et al.* (2017) Fungal genome and mating system transitions facilitated by chromosomal translocations involving intercentromeric recombination. *PLoS Biol* 15(8):e2002527.
52. Chatterjee G*, et al.* (2016) Repeat-associated fission yeast-like regional centromeres in the ascomycetous budding yeast *Candida tropicalis*. *PLoS Genet* 12(2):e1005839.
53. 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 U S A* 112(39):12139-12144.
54. 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(31):11374-11379.

55. Baum M, Ngan VK, & Clarke L (1994) The centromeric K-type repeat and the central core are together sufficient to establish a functional *Schizosaccharomyces pombe* centromere. *Mol Biol Cell* 5(7):747-761.
56. Coughlan AY, Hanson SJ, Byrne KP, & Wolfe KH (2016) Centromeres of the yeast *Komagataella phaffii (Pichia pastoris)* have a simple inverted-repeat structure. *Genome Biol Evol* 8(8):2482-2492.
57. Takahashi K*, et al.* (1992) A low copy number central sequence with strict symmetry and unusual chromatin structure in fission yeast centromere. *Mol Biol Cell* 3(7):819-835.
58. Luiza Cesca Piva JDM, Lidia Moraes, Viviane Reis, Fernando Araripe Goncalves Torres (2018) Genetic analysis of the *Komagataella phaffii* centromeres by a color-based plasmid stability assay. *bioRxiv 433417;* .
59. Chan FL & Wong LH (2012) Transcription in the maintenance of centromere chromatin identity. *Nucleic Acids Res* 40(22):11178-11188.
60. Volpe TA*, et al.* (2002) Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science* 297(5588):1833-1837.
61. Freire-Beneitez V, Price RJ, & Buscaino A (2016) The chromatin of *Candida albicans* pericentromeres bears features of both euchromatin and heterochromatin. *Front Microbiol* 7:759.
62. Bobkov GOM, Gilbert N, & Heun P (2018) Centromere transcription allows CENP-A to transit from chromatin association to stable incorporation. *J Cell Biol* 217(6):1957-1972.
63. Du Y, Topp CN, & Dawe RK (2010) DNA binding of centromere protein C (CENPC) is stabilized by single-stranded RNA. *PLoS Genet* 6(2):e1000835.
64. Talbert PB & Henikoff S (2018) Transcribing centromeres: noncoding RNAs and kinetochore assembly. *Trends Genet* 34(8):587-599.
65. Selker EU*, et al.* (2003) The methylated component of the *Neurospora crassa* genome. *Nature* 422(6934):893-897.
66. Nagaki K*, et al.* (2004) Sequencing of a rice centromere uncovers active genes. *Nat Genet* 36(2):138-145.
67. Wang K, Wu Y, Zhang W, Dawe RK, & Jiang J (2014) Maize centromeres expand and adopt a uniform size in the genetic background of oat. *Genome Res* 24(1):107-116.
68. International Chicken Genome Sequencing C (2004) Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* 432(7018):695-716.
69. Ishii K*, et al.* (2008) Heterochromatin integrity affects chromosome reorganization after centromere dysfunction. *Science* 321(5892):1088-1091.
70. Zhang H & Freudenreich CH (2007) An AT-rich sequence in human common fragile site FRA16D causes fork stalling and chromosome breakage in *S. cerevisiae*. *Mol Cell* 27(3):367-379.
71. Greenfeder SA & Newlon CS (1992) Replication forks pause at yeast centromeres. *Mol Cell Biol* 12(9):4056-4066.
72. Mitra S, Gomez-Raja J, Larriba G, Dubey DD, & Sanyal K (2014) Rad51-Rad52 mediated maintenance of centromeric chromatin in *Candida albicans*. *PLoS Genet* 10(4):e1004344.
73. Smith JG*, et al.* (1995) Replication of centromere II of *Schizosaccharomyces pombe*. *Mol Cell Biol* 15(9):5165-5172.
74. Kops GJ, Weaver BA, & Cleveland DW (2005) On the road to cancer: aneuploidy and the mitotic checkpoint. *Nat Rev Cancer* 5(10):773-785.
75. Rief N & Lobrich M (2002) Efficient rejoining of radiation-induced DNA double-strand breaks in centromeric DNA of human cells. *J Biol Chem* 277(23):20572-20582.

76. Dotiwala F, Harrison JC, Jain S, Sugawara N, & Haber JE (2010) Mad2 prolongs DNA damage checkpoint arrest caused by a double-strand break via a centromere-dependent mechanism. *Curr Biol* 20(4):328-332.
77. Strunnikov AV (2010) One-hit wonders of genomic instability. *Cell Div* 5(1):15.
78. Brito IL, Monje-Casas F, & Amon A (2010) The Lrs4-Csm1 monopolin complex associates with kinetochores during anaphase and is required for accurate chromosome segregation. *Cell Cycle* 9(17):3611-3618.
79. Burrack LS, Applen Clancey SE, Chacon JM, Gardner MK, & Berman J (2013) Monopolin recruits condensin to organize centromere DNA and repetitive DNA sequences. *Mol Biol Cell* 24(18):2807-2819.
80. Rumpf C*, et al.* (2010) Laser microsurgery provides evidence for merotelic kinetochore attachments in fission yeast cells lacking Pcs1 or Clr4. *Cell Cycle* 9(19):3997-4004.
81. Tada K, Susumu H, Sakuno T, & Watanabe Y (2011) Condensin association with histone H2A shapes mitotic chromosomes. *Nature* 474(7352):477-483.
82. Kim M*, et al.* (2018) Genomic tandem quadruplication is associated with ketoconazole resistance in *Malassezia pachydermatis*. *J Microbiol Biotechnol* 28(11):1937-1945.

Figures and figure legends

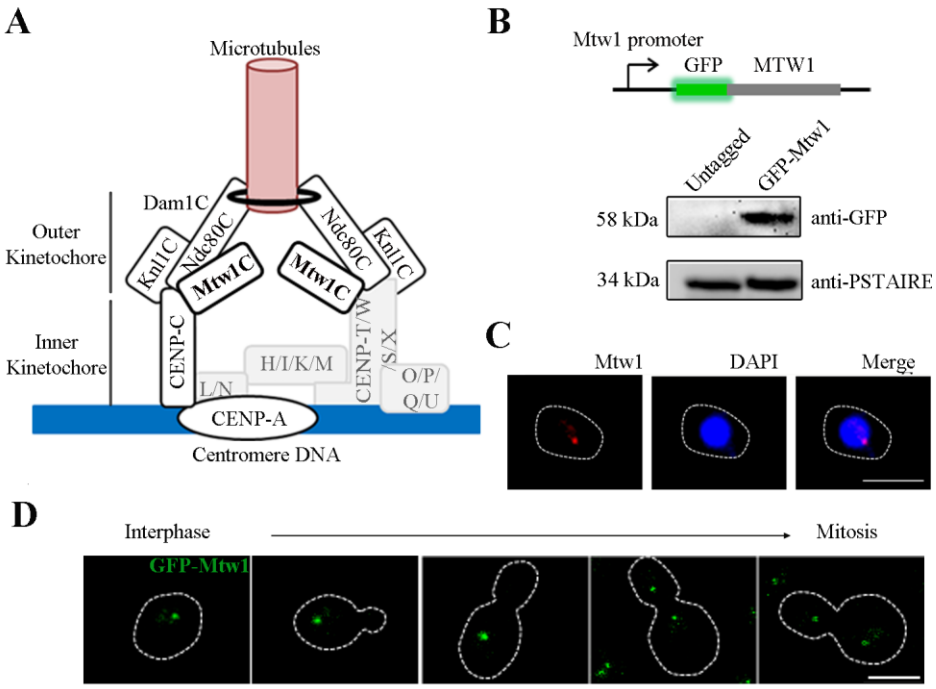

Figure 1. Kinetochores cluster and localized at the nuclear periphery in *M. sympodialis*. (A) Schematic of the organization of the kinetochore in *M. sympodialis*. Gray boxes indicate proteins absent in *M. sympodialis*. The outer kinetochore protein Mtw1 has been used as the kinetochore marker in the present study. (B) Line diagram representation of the tagging of GFP at the N-terminus of Mtw1. Immunoblot analysis using whole cell lysates prepared from the untagged strain (*M. sympodialis* ATCC42132) and GFP-Mtw1 expressing cells probed with anti-GFP antibodies and anti-PSTAIRE antibodies. PSTAIRE was used as a loading control. (C) Logarithmically grown cells expressing GFP-Mtw1 were fixed and stained with DAPI (blue) and anti-GFP antibodies (pseudo-colored as red). Bar, 2.5 µm. (D) Cell cycle stage-specific localization dynamics of GFP-Mtw1 from interphase (unbudded) through mitosis (large budded). Scale bar, 2.5 µm.

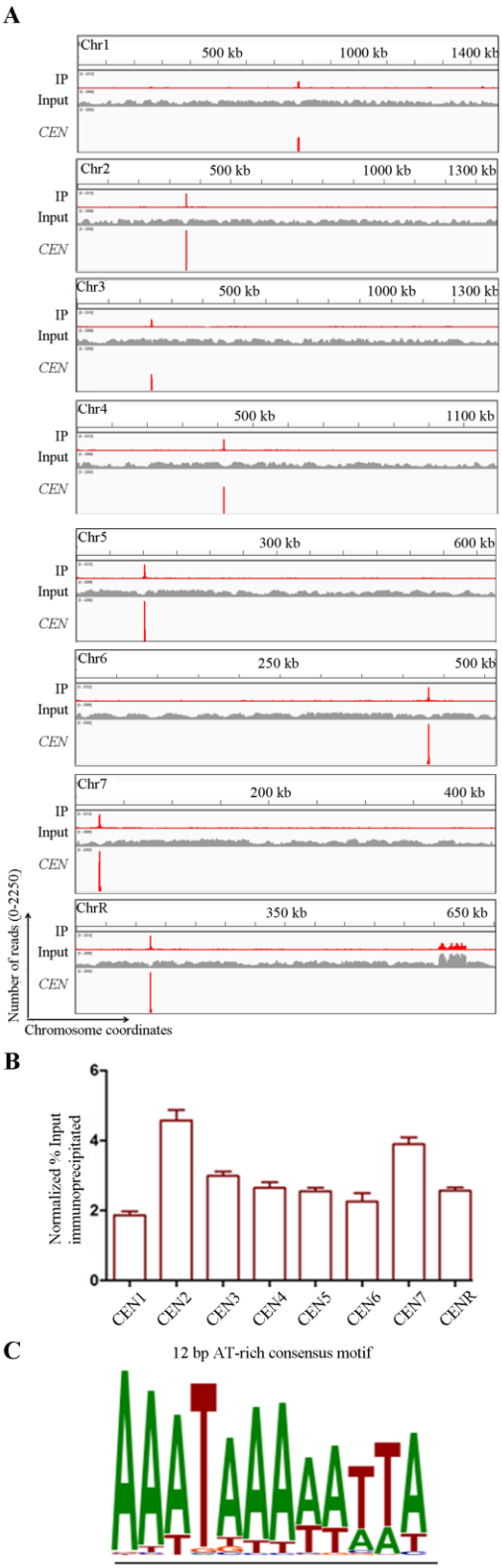

Figure 2. Single-peak localization of Mtw1 identifies centromeres on each of the eight chromosomes of *M. sympodialis*. (A) GFP-Mtw1 ChIP-seq reads mapped along each chromosome. The *x*-axis indicates

chromosome coordinates (in kb) and the *y*-axis indicates distribution of sequenced reads. “Input”, reads from total DNA; “IP,” reads from immunoprecipitated sample; *CEN*, Mtw1-enriched regions derived by subtracting input reads from those of the IP sample (peak values 0-2250). Additional peaks observed in both IP and input tracks on Chr5 are from the rDNA locus. (B) ChIP-qPCR assays validating the enrichment of Mtw1 at the centromeres. The *x*-axis indicates individual *CEN* regions and the *y*-axis indicates enrichment as normalized % input immunoprecipitated. Error bars indicate standard error mean (SEM). (C) Logo of consensus DNA sequence identified from *M. sympodialis* centromeres, graphically represented with the size of the base correlating to the frequency of occurrence.

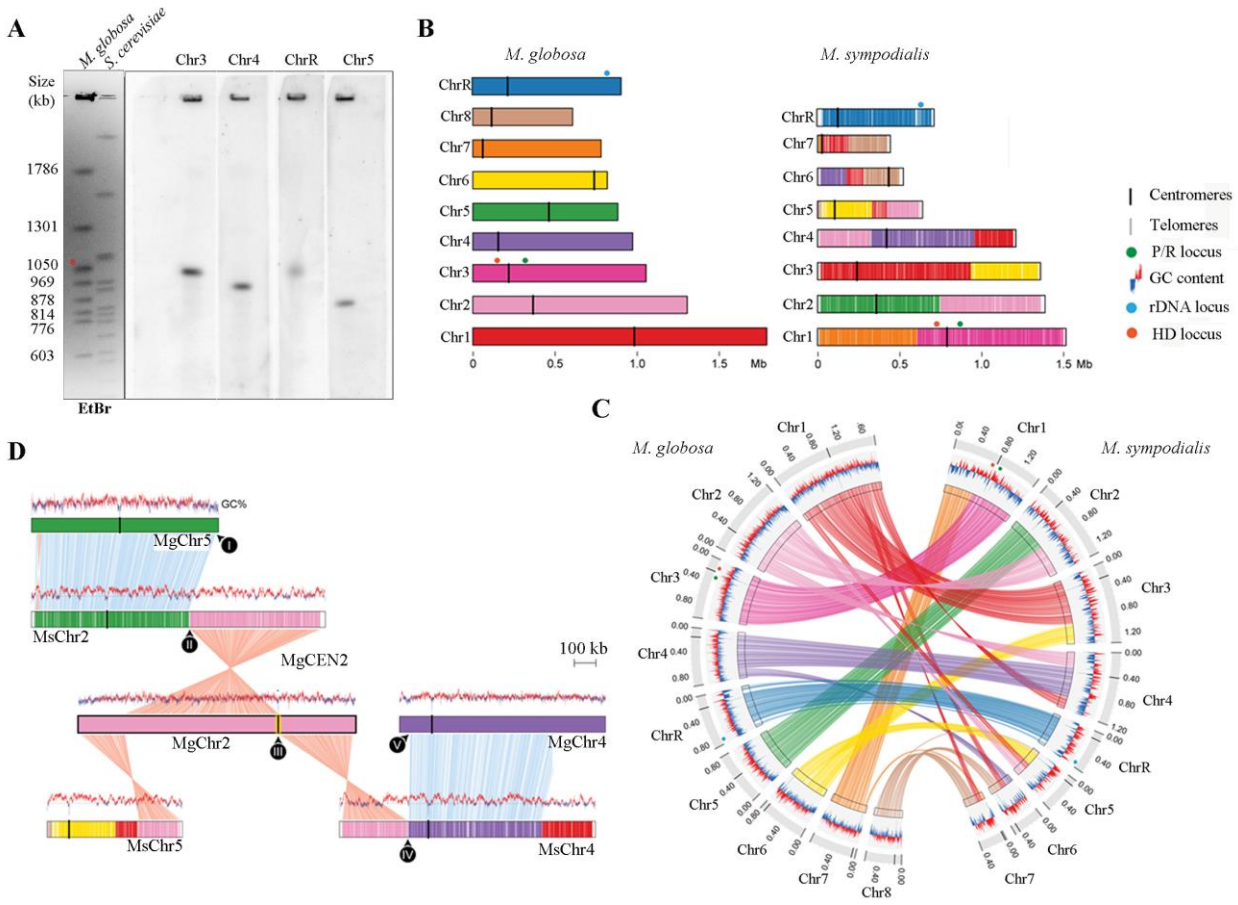

Figure 3. **Centromere of chromosome 2 of *M. globosa* (MgChr2) maps to a synteny breakpoint.**

(A) Chromosomes of *M. globosa* were resolved on CHEF gels and stained with ethidium bromide (EtBr) along with *S. cerevisiae* chromosomes used as size markers (also see Figure S2A). The gels were blotted and probed with unique sequences from Chr3, Chr4, Chr5, and ChrR (right panel). (B) The panel represents the karyotype and the position of centromeres, telomeres, rDNA loci, and the *HD* and *P/R MAT* loci of *M. globosa* and *M. sympodialis* as a linear bar diagram. *M. sympodialis* chromosomes are color coded based on their synteny with *M. globosa* chromosomes. (C) A circos plot depicting the conserved synteny blocks between the *M. globosa* (CBS7966 strain) and *M. sympodialis* chromosomes. Tracks from outside to inside represent- positions of centromeres and telomeres, GC content (red/blue line diagram) and colored connectors indicate regions of synteny between the two species. (D) Linear chromosome plot depicting the synteny between chromosome 2 of *M. globosa* and chromosomes 2, 4, and 5 of *M. sympodialis.* GC content (in %) is shown as red/blue line diagram above each chromosome. Pink connectors represent regions with synteny to MgChr2 and blue connectors represent those of MgChr4. Labels in black circles mark the synteny breakpoints. Synteny breakpoint of MgChr2 is marked as *MgCEN2*(III). The regions on MsChr2 and MsChr4 where the homologs of ORFs flanking the

breakpoint are located are marked II and IV. The synteny block start site between MgChr4 and MsChr4 of *M. globosa* is labeled V.

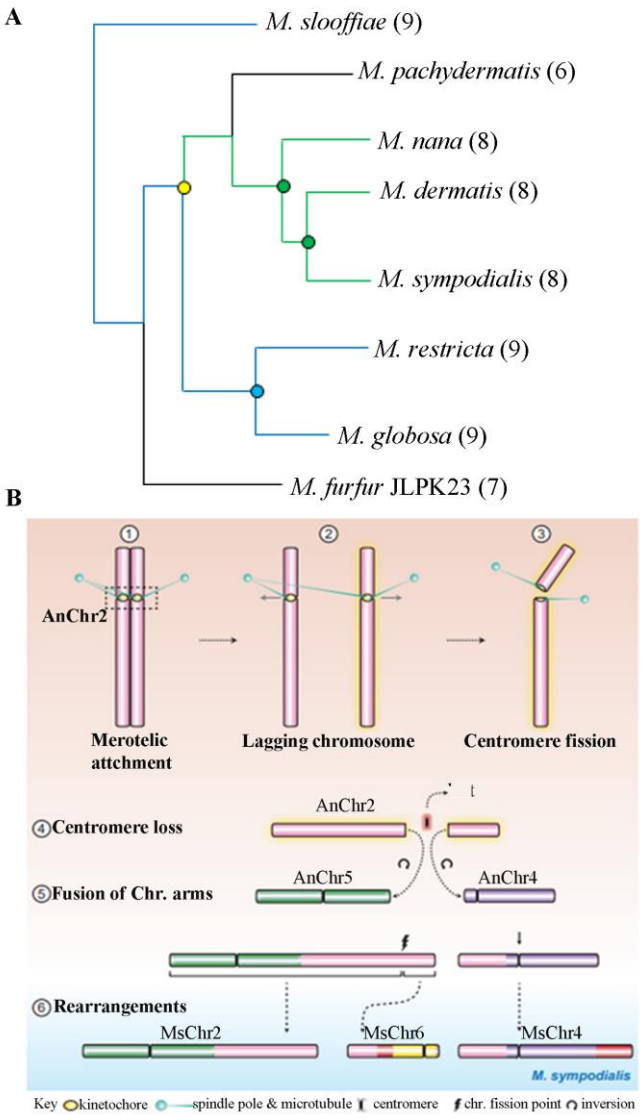

Figure 4. Karyotype evolution by centromere breakage and loss in *Malassezia* species

(A) Cladogram of closely related *Malassezia* species with their chromosome number in brackets (adapted from (31)). The chromosome numbers mentioned for *M. slooffiae* and *M. globosa* are based on results from this study. In case of *M. sympodialis*, *M. restricta*, *M. pachydermatis*, and *M. furfur*, the numbers are based on previous reports (4, 5, 32, 82). For *M. dermatis* and *M. nana*, the number of predicted centromeres, indicative of chromosome number, is mentioned in the brackets. Blue and green lines/circles indicate karyotype with 9 and 8 chromosomes respectively. The yellow circle marks the ancestral

karyotype with 9 chromosomes. (B) Schematic of the centromere break and the resulting reduction in chromosome number as a consequence of unresolved merotelic attachment and fusion of chromosome arms to other chromosomes. A karyotype with 9 chromosomes (as shown for *M. globosa*) is depicted as the ancestral state.