Cellular dynamics and genomic identity of centromeres in Rice Blast

Vikas Yadav[1,*,$], Fan Yang[2,*], Md. Hashim Reza[1], Kaustuv Sanyal[1,#] and Naweed I. Naqvi[2,#]

[1]Molecular Biology and Genetics Unit, Jawaharlal Nehru Centre for Advanced Scientific Research, Jakkur, Bangalore, India; [2]Temasek Life Sciences Laboratory, 1 Research Link, Singapore.

*These authors contributed equally to this work.

[$]Present address: Department of Molecular Genetics and Microbiology, Duke University Medical Center, Durham, NC, USA.

[#]Corresponding authors

Naweed I. Naqvi (naweed@tll.org.sg)

Kaustuv Sanyal (sanyal@jncasr.ac.in)

Running title: Centromeres in *M. oryzae*.

Keywords: CenpA, CenpC, Kinetochore, Spindle pole body, Mitosis, *Magnaporthe oryzae*.

Abstract

A series of well-synchronized events mediated by kinetochore-microtubule interactions ensure faithful chromosome segregation in eukaryotes. Centromeres scaffold kinetochore assembly and are among the fastest evolving chromosomal loci in terms of the DNA sequence, length, and organization of intrinsic elements. Neither the centromere structure nor the kinetochore dynamics is well studied in plant pathogenic fungi. Here, we sought to understand the process of chromosome segregation in the rice blast fungus, *Magnaporthe oryzae*. High-resolution confocal imaging of GFP-tagged inner kinetochore proteins, CenpA and CenpC, revealed an unusual albeit transient declustering of centromeres just before anaphase separation in *M. oryzae*. Strikingly, the declustered centromeres positioned randomly at the spindle midzone without an apparent metaphase plate *per se*. Using chromatin immunoprecipitation followed by deep sequencing, all seven centromeres were identified as CenpA-rich regions in the wild-type Guy11 strain of *M. oryzae*. The centromeres in *M. oryzae* are regional and span 57 to 109 kb transcriptionally poor regions. No centromere-specific DNA sequence motif or repetitive elements could be identified in these regions suggesting an epigenetic specification of centromere function in *M. oryzae*. Highly AT-rich and heavily methylated DNA sequences were the only common defining features of all the centromeres in the Rice Blast. Detailed gene synteny analyses helped identify and compare the centromere regions in distinct isolates of *M. oryzae* and its related species *Magnaporthe poae.* Overall, this study identifies unusual centromere dynamics and maps the centromere DNA sequences in the top model fungal pathogen *M. oryzae*, which causes severe losses in global rice production.

Significance

Magnaporthe oryzae is an important fungal pathogen that causes an annual loss of 10-30% rice crop due to the devastating blast disease. In most organisms, kinetochores are arranged either in the metaphase plate or are clustered together to facilitate synchronized anaphase separation of chromosomes. In this study, we show that the initially clustered kinetochores separate and position randomly prior to anaphase in *M. oryzae*. Centromeres, identified as the site of kinetochore assembly, are regional type without any shared sequence motifs in *M. oryzae*. Together, this study reveals atypical kinetochore dynamics and identifies functional centromeres in *M. oryzae*, thus paving the way to define heterochromatin boundaries and understand the process of kinetochore assembly on epigenetically specified centromere loci.

Introduction

Faithful chromosome segregation is one of the essential processes that is required for maintaining genome integrity in dividing cells. This process is successfully carried out by the attachment of microtubules, emanating from opposite spindle poles, to the proteinaceous multi-subunit structure, the kinetochore, that is pre-assembled onto centromeres (1, 2). The centromere forms a crucial part of this machinery and yet, it is one of the most rapidly evolving loci in eukaryotic genomes (3, 4). On the contrary, the proteins that bind to centromere DNA are evolutionary conserved (2). Centromere DNA shows a wide diversity in the length and composition of the underlying DNA sequence. A few fungal species, like *Saccharomyces cerevisiae*, harbor centromeres that are less than 400 bp comprising of conserved DNA sequence elements to form point centromeres (5). Most others possess regional centromeres that span from few kilobases to several megabases. Unlike point centromeres, regional centromeres in an organism often do not share any conserved DNA sequence features. For example, the regional centromeres in *Schizosaccharomyces pombe* and *Candida tropicalis* have a homogenized central core flanked by inverted repeats (6, 7). Likewise, the regional centromeres in *Cryptococcus neoformans* possess specific retrotransposons that are present randomly therein (8). In contrast, centromeres in *Candida albicans*, *Candida lusitaniae*, and *Candida dubliniensis* differ between all chromosomes and lack a conserved DNA sequence element (9-11). Centromeres in filamentous fungi like *Neurospora crassa*, on the other hand, span long stretches of repetitive DNA but lack a consensus sequence or pattern (12, 13). Metazoans and plants also have regional centromeres that are up to few megabases long, and mostly consist of repetitive DNA or transposons (14-16).

Despite this sequence divergence, centromeres in most studied organisms are bound by a centromere-specific histone H3 variant CENP-A /CenH3/Cse4, also known as the hallmark of centromere identity (4, 17). CENP-A forms the foundation of the kinetochore assembly and is essential for cell viability in all organisms studied until date. Evolutionary conservation of CENP-A along with other kinetochore proteins also provides an efficient tool to identify centromeres. Additionally, studies with fluorescently-labeled inner kinetochore proteins such as CENP-A or CENP-C/Cen-C/Mif2 has led to an understanding of spatial dynamics of the kinetochore within the nucleus (18-22). These studies established that kinetochores in most yeast species are clustered throughout the nuclear division, and unlike metazoan *CEN*, do not align on a metaphase plate. However, more recently, some variations to the metaphase plate or kinetochore clustering have been reported revealing the diversity in this phenomenon. Kinetochores remain clustered throughout the cell cycle in two well-studied ascomycetes, *S. cerevisiae,* and *C. albicans* (23, 24). In *S. pombe,* kinetochores undergo a brief declustering during mitosis but remain clustered otherwise (18, 25). Another ascomycete, *Zymoseptoria tritici*, shows multiple kinetochore foci instead of a single cluster during interphase although their localization dynamics during mitosis remains unexplored (26). On the other hand, the cells of a basidiomycete *C. neoformans* display multiple foci of kinetochores in interphase, but kinetochores gradually cluster during mitosis (19, 22). Even the phenomenon of centromere/kinetochore clustering is observed in *Drosophila* that depends on centric chromatin rather than specific DNA sequences (27).

Besides CENP-A, several other chromatin features are known to be associated with centromeres. For example, centromeres are devoid of genes/ORFs and exhibit a significantly low level of polyA transcription as compared to the rest of the genome (8, 28). Furthermore,

centromeres in many organisms are heterochromatic in nature and harbor the heterochromatic marks like H3K9di/trimethylation and DNA methylation (8, 13, 29). A preference for AT-rich DNA sequence is evident for centromere formation in some organisms (13, 30-32). It is important to note that none of these features exclusively define centromeres and, in most cases, the importance of an individual factor in defining centromere loci is not well understood. However, the presence of such features on discrete chromosomal loci may pave the way for predicting centromeres in organisms in which genome tractability is difficult.

Magnaporthaceae is a family of ascomycete fungi comprising of many important plant pathogenic species including *Magnaporthe oryzae* and *Magnaporthe poae*. *M. oryzae* causes the devastating blast disease in cereal crops including rice, wheat, barley and millet (33, 34). *M. poae* is responsible for summer-patch disease in turf grasses (35). *M. oryzae*, also known as rice blast, is a constant threat to agriculture-based economies due to significant damage to rice harvests. In recent years, rice blast has also emerged as a model pathosystem for studying host-pathogen interactions due to the availability of the genome sequence, fully characterized infection cycle, genetic tractability and economic significance of the fungus (36, 37). However, even with the availability of the genome sequence and annotated assembly, the centromere/kinetochore identity of the rice blast fungus remains unexplored or poorly defined. Here, we first studied and characterized orthologs of CENP-A and CENP-C, two well-conserved kinetochore proteins, to understand the kinetochore dynamics in this organism and used these kinetochore proteins as tools to identify bona fide centromeres.

Results

Kinetochores are clustered during interphase in M. oryzae

A subset of putative kinetochore proteins was previously annotated in *M. oryzae* (12). We expanded the list further by identifying putative orthologs of the remaining conserved kinetochore proteins using *in silico* predictions (SI Appendix, Table S1). Multiple sequence alignment established the identity of at least two most conserved inner kinetochore proteins: CenpA (MGG_06445, a homolog of CENP-A) and CenpC (MGG_06960, a homolog of CENP-C) (SI Appendix, Figure S1) in *M. oryzae*. CenpA and CenpC of *M. oryzae* share 73% and 42% sequence identity with their *N. crassa* counterparts CenH3 and CEN-C, respectively. Next, we functionally expressed the GFP-tagged CenpA and CenpC from their native genomic loci in the wild-type Guy11 strain of *M. oryzae*. GFP-CenpA and CenpC-GFP signals appeared as a single dot-like and co-localized on chromatin, marked by mCherry-tagged histone H1 (Figure 1A, 1B). Further, co-localization of CenpA and CenpC signals confirmed their overlapping spatial positions in both mycelia and conidia (Figure 1C). Clustering of kinetochores is a hallmark feature of many yeast and fungal genera. Such clustered kinetochores are often found in close proximity to the spindle pole bodies (SPBs) (38). We localized SPBs by tagging Alp6 (ortholog of *S. cerevisiae* Spc98) with mCherry and observed that SPBs localize close to the clustered GFP-CenpA signals in *M. oryzae* (Figure 1D). These results indicate that kinetochore localization during interphase in *M. oryzae* is similar to that observed in other ascomycetes. Our attempts to delete *CENPA* or *CENPC* in *M. oryzae* failed indicating that both are essential for cell viability. This result was further corroborated by conditional repression of *CENPA* using the Tet-off system. The *Tet-GFP-CENPA* strain ceased to grow on culture media supplemented with doxycycline, the condition in which Tet-driven *CENPA* expression is shut down (SI Appendix,

Figure S2A). Overall, the conserved sequence features and the subcellular localization patterns confirmed that CenpA and CenpC are evolutionarily conserved kinetochore proteins in *M. oryzae*.

Kinetochores in M. oryzae undergo declustering-clustering dynamics during mitosis

To study the cellular dynamics of kinetochores in *M. oryzae*, we localized microtubules by expressing mCherry-TubA or GFP-TubA fusion protein and co-localized it with GFP-CenpA. During interphase, the microtubules are mostly localized throughout the cytoplasm (SI Appendix, Figure S2B). Live-cell imaging during mitosis revealed dispersed GFP-CenpA signals localized along the mitotic spindle (Figure 2A and Movie S1). Strikingly, the declustered dot-like signals of GFP-CenpA then segregated into two halves in a non-synchronous manner. Once segregated, the signals began to cluster again and localized as two bright foci close to poles of the mitotic spindle. To further probe the dynamics of kinetochore segregation, we performed high-resolution imaging in mitotic cells expressing GFP-CenpA (Figure 2B, C, Movie S2 and S3). We observed that while the GFP-CenpA signals were spread out, they were localized in pairs, most likely representing the segregated kinetochore signals (Figure 2B, time 00:32). We were able to count fourteen discrete spots of GFP-CenpA corresponding to 14 kinetochores of the seven duplicated chromosomes. These results suggest that kinetochores in *M. oryzae* remain largely unclustered during mitosis. It was further supported by co-localization of GFP-CenpA with a SPB markerAlp6-mCherry during the mitotic stages (Figure 2D). In pre-mitotic cells, we observed two duplicated spots of Alp6-mCherry that co-localized with replicated clustered GFP-CenpA signals. During mitosis, GFP-CenpA signal localized as multiple puncta scattered in between the two SPBs represented by Alp6-mCherry. After the division, the GFP-

CenpA/kinetochores clustered again and localized adjacent to the SPBs (SI Appendix, Figure S2C, and Movie S4). Taken together, we conclude that kinetochores decluster during mitosis in *M. oryzae*, and align along the mitotic spindle. Furthermore, we infer that an equatorial plate alignment of the kinetochores is not evident in *M. oryzae*, indicating a lack of a well-defined metaphase plate. Similar dynamics of the kinetochore and microtubules were observed in *M. oryzae* cells during pathogenic development and *in planta* conditions (SI Appendix, Figure S3, Movie S5, and S6). Based on these observations, we propose a schematic model for the kinetochore and SPB dynamics during the mitotic cycle in rice blast where kinetochore clustering-declustering dynamics is most likely dependent on their direct link to the SPBs (Figure 2E). During mitosis, this link is likely broken, and the clustering is thus perturbed. We infer that such timely and dynamic kinetochore clustering/declustering is crucial for proper chromosome segregation in *M. oryzae*.

Kinetochore protein binding identifies regional centromeres in M. oryzae

CenpA binding is a hallmark of functional centromeres in eukaryotes (4, 15). We used GFP-CenpA as a tool for molecular identification of centromeres in the *M. oryzae* genome. We utilized chromatin immunoprecipitation (ChIP) assays followed by deep sequencing (ChIP-seq) of GFP-CenpA-associated chromatin fragments and aligned the reads on the recently published PacBio genome assembly of the wild-type Guy11 strain of *M. oryzae* (39). This analysis revealed seven distinct CenpA-rich regions across the genome, one each on seven different contigs (Figure 3, Table 1 and SI Appendix, Figure S4). The CenpA binding spans a 57 to 109 kb region suggesting that *M. oryzae* possesses large regional centromeres. The centromere identity of these regions was further validated by binding of another evolutionarily conserved independent

kinetochore protein CenpC. ChIP-qPCR using the fungal strain expressing CenpC-GFP (SI Appendix, Figure S5A) confirmed specific overlapping binding of CenpA and CenpC on each of these seven *CEN* regions. We also observed an additional region of 1200 bp on Contig 4 apart from the seven distinct peaks in CenpA ChIP-seq analysis. The enriched peak was found to be present on the gene encoding the vacuolar morphogenesis protein AvaB (MGG_01045). Using specific ChIP-qPCR primers for this region, we confirmed that the aforementioned CenpA enrichment on Contig 4 was likely an artifact (SI Appendix, Figure S5B). Overall, the binding of two independent kinetochore proteins at seven long regions confirmed that these are indeed authentic centromeres of the corresponding chromosomes in *M. oryzae*.

A detailed analysis revealed that the seven centromeres in *M. oryzae* comprise of highly AT-rich sequences (≥67%) (Figure 3B and Table 1). The centromeres in *M. oryzae* harbor a few repetitive elements, mostly retrotransposons belonging to long terminal repeat (LTR) elements (Figure 3B, Dataset S1). However, these elements are neither exclusive to the centromeres nor common among the seven centromeres in *M. oryzae*. Further in-depth analysis of these regions did not reveal any common DNA sequence motif or repeats as supported by the dot-plot analysis of each centromere (SI Appendix, Figure S6). We then examined the transcriptional status and base modifications associated with centromeric chromatin using the published RNA-sequencing and bisulfite sequencing data (40, 41). Centromeres in *M. oryzae* are found to be poorly transcribed and harbor 5mC DNA methylation (Figure 3). Based on these results, we conclude that centromeres in *M. oryzae* do not share any common DNA sequence motif or repeat element and that AT-richness is likely the only defining sequence feature of all the centromeres in *M. oryzae*. Additionally, we also infer that centromeres in *M. oryzae* are large, regional and lie within transcriptionally-poor 5mC-rich DNA regions of the genome.

Centromere DNA sequences are rapidly evolving in Magnaporthe species

The MG8 genome assembly is based on the sequencing of the *M. oryzae* isolate 70-15, which represents a progeny of the Guy11 strain (36, 42, 43). This would mean that the 70-15 assembly must harbor at least 50% genome sequence of Guy11. The PacBio genome sequence of Guy11 provides a near complete end-to-end chromosome-wide coverage of the 70-15 genome, the only chromosome-level sequence assembly available for *M. oryzae* (SI Appendix, Figure S7A). Thus, we attempted to identify the centromere location in the 70-15 isolate by aligning CenpA ChIP-seq reads on to the MG8 assembly. This analysis revealed seven distinct peaks, one on each chromosome (Figure 4A, SI Appendix, Figure S7B, and Table S2). We also observed two additional CenpA-enriched regions in the unassembled Supercontig8.8 of MG8 assembly for 70-15 (SI Appendix, Figure S7C). Additionally, the identified centromere on chromosome 7 in this assembly mapped to the same region that was previously predicted to harbor the centromere based on genetic analysis (44).

Next, we analyzed the recently published PacBio genome sequence/assembly of the *M. oryzae* field isolate FJ81278 (39) to identify the centromere sequences and compare them with the 70-15 assembly. Mapping of CenpA ChIP-seq reads revealed nine distinct peaks in the FJ81278 genome assembly (SI Appendix, Figure S8, and Table S2). Three of these enriched regions were present at the end of three separate contigs (Contig 3, 14 and 16). By comparing genome assemblies of 70-15 and FJ81278, we concluded that contigs 3 and 14 are most likely parts of the same chromosome and the CenpA-enriched regions observed in these two contigs represent a single centromere (*CEN4*). Synteny analysis also revealed that the CenpA peaks in Contig 11 and 16 belong to the same chromosome. However, Contig11 of FJ81278 assembly

seems to be mis-assembled, since a part of this contig does not show synteny with any region of the 70-15 genome. Thus, we excluded this centromere (*CEN7*) region from further analysis.

Next, we compared the centromeres and the flanking regions from the genome assemblies of Guy11, 70-15, and FJ81278. Detailed synteny analyses revealed that the centromere flanking regions are conserved among these three isolates indicating that the overall position of centromeres is likely conserved in different strains/field isolates of *M. oryzae* (Figure 4B and SI Appendix, Figure S9). However, a major part of the centromere sequences was found to be missing from the 70-15 genome assembly as compared to Guy11 and FJ81278. It is important to note that the MG8 version of the 70-15 genome assembly is not complete and harbors a number of gaps. We believe that some of the centromere sequences are part of the unassembled Supercontig8.8 and are the CenpA-enriched regions observed in this fragment. The centromere sequences of Guy11 and FJ81278 isolates shared a high level of conservation with a few rearrangements. To explore this further, we performed a pair-wise comparison using sequences of respective centromeres from Guy11 and FJ81278 genomes. This analysis revealed that while most of the AT-rich sequence remains conserved between the two isolates, the repeat content varies significantly and accounts for almost all the observed rearrangements (Figure 4C, SI Appendix, Figure S10, and Dataset S2). These results suggest that repeat elements might shape the structure of centromeres in different isolates even though they may not be an integral part of centromeres.

Inter-species comparison of CEN sequences in Magnaporthaceae

The analysis in different isolates of *M. oryzae* further validated that centromeres in this species comprise of long AT-rich and transcription-poor regions. Using these parameters, we

attempted to predict centromeres in *M. poae*, the root infecting pathogen that belongs to the Magnaporthaceae family as well (35). We were able to identify eight putative centromere regions across the *M. poae* genome based on this *in-silico* analysis (SI Appendix, Figure S11, and Table S3). Three of these eight putative *CEN* regions were present at the end of different contigs. Since the chromosome number in *M. poae* is not established, it is uncertain whether all of these AT-rich regions represent bona fide centromeres in *M. poae*. We also found that these putative centromeres in *M. poae* harbor more repetitive DNA sequences than *M. oryzae* even though the genomic repeat content of *M. poae* is only 1.1% as compared to 10.1% in *M. oryzae*. Unlike, different isolates of *M. oryzae* that share a high level of centromere sequence conservation, the centromere sequence of *M. oryzae* and *M. poae* are highly diverged. Based on these results, we conclude that centromere DNA sequences in the *Magnaporthaceae* family are rapidly evolving, whereas the properties of centromeric chromatin are likely conserved between the two species.

Discussion

Blast disease caused by *M. oryzae* is exceedingly disastrous not only to rice production worldwide but also to several graminaceous crops (34). Despite being such a vital plant pathogen, the fundamental cellular process of chromosome segregation is not well understood in this organism. In this work, we attempted to study the chromosome segregation machinery in *M. oryzae* at the molecular level. We tagged two evolutionarily conserved key kinetochore proteins and studied their dynamics during different phases of the cell cycle at various developmental stages in *M. oryzae*. We further identified the genomic loci that act as centromeres in this filamentous fungus. Based on a comparison of centromere sequences among different isolates of

M. oryzae and a related species, *M. poae*, centromeres appear to be rapidly evolving in *Magnaporthe*, as reported in several fungal species complex before (6, 8, 10, 12, 45, 46).

Kinetochores cluster together in a single locus at the nuclear periphery in many fungi. This locus is often referred to as the CENP-A-rich zone or CENP-A cloud (47, 48). It has been proposed that such a nuclear subdomain with a high concentration of CENP-A favor centromere seeding on the chromosomal regions in close proximity to it, in the absence of a centromere-specific DNA sequence. In most budding yeasts, kinetochores are clustered throughout the cell cycle except in *C. neoformans,* which shows clustered kinetochores only during mitosis (19). The kinetochore dynamics in *M. oryzae* is found to be similar to the "fission" yeast rather than that of the budding yeast species. It is possible that mitotic declustering of kinetochores is a feature of all yeasts/fungi that divide by septum formation. However, a more detailed analysis of kinetochore behavior in filamentous fungi like *N. crassa* and *Z. tritici* will be useful to establish this link. It is noteworthy that *Z. tritici* does not have a single centromere cluster, but kinetochores are arranged in multiple chromocenters, a process observed in some plant species (26, 49). We also observed that kinetochores align along the mitotic spindle in *M. oryzae*, though a proper metaphase plate formation was not evident. A similar kinetochore arrangement was also observed in a basidiomycete *C. neoformans* (19). The presence of such a structure in two evolutionarily distant fungal species suggests the existence of a transient formation of a structure, an arrangement alternative to the metaphase plate, across the fungal kingdom. In addition, co-localization of kinetochore proteins and SPBs revealed a close association between the two as observed in *S. pombe* (25). Our results also suggest that a direct interaction between the SPBs and kinetochores may facilitate kinetochore clustering. The SPB-kinetochore interaction has

been explored in other fungi and led to the identification of several uncharacterized proteins (22, 50-53). It remains to be seen whether or not such interactions occur in *M. oryzae* as well.

Centromere DNA sequences, despite being associated with a conserved and essential function, are highly divergent across species (3). The centromeres identified in *M. oryzae* further add to this diversity of centromere sequences. Our results show that centromeres in *M. oryzae* are long and AT-rich similar to those reported in *N. crassa* except that the centromeres are significantly shorter in *M. oryzae* (57-109 kb) compared to *N. crassa* (150-300 kb) (12, 13). The DNA methylation pattern observed in *M. oryzae* is similar to that of *N. crassa* as it is present at multiple loci in both the organisms and thus differs from that of *C. neoformans* where DNA methylation is restricted to only centromeres and telomeres (8, 13). Additionally, a specific pattern of centromeric histone binding has been reported in *N. crassa*, but no such pattern exists in *M. oryzae*. Since centromere DNA sequences are generally repeat-rich, they are poorly assembled which restricts finer analysis of *CEN* DNA sequence. For example, centromeres in *Fusarium graminearum* are proposed to be AT-rich, similar to that of *M. oryzae* and *N. crassa* (12). However, the exact nature of the centromere sequence of these regions remains unknown in *F. graminearum* due to sequence gaps in the genome assembly. Similarly, most of the centromere sequences are absent in the currently available 70-15 genome assembly of *M. oryzae*. Taking together, an improved genome assembly with complete chromosome-level sequence information is required for a better understanding of a complex genomic locus like the centromere.

Apart from filamentous fungi, AT-rich centromeres are present in other fungal species like *Malassezia sympodialis*, albeit the length of these regions is significantly smaller as compared to *M. oryzae* centromeres (32). The CDEII element of point centromeres present in the

budding yeast, *S. cerevisiae*, is also highly AT-rich (54). A recent study reports the presence of AT-rich centromeres of varying lengths in diatoms (30). Furthermore, the 171-bp alpha satellite repeat DNA present in human centromeres is also AT-rich in nature (55). Overall, these results suggest that AT-richness favors centromere function in many organisms. Intriguingly, *in vitro* experiments suggest that CENP-A binds with a lower affinity to an AT-rich DNA sequence (56). In contrast, the same study also revealed that the CENP-A chaperone Scm3 has a higher affinity towards AT-rich sequences. With more AT-rich centromeres being characterized, identifying the exact role of AT-rich sequences in centromere function is critical.

Regional centromeres of many organisms, including *M. oryzae*, do not share any common DNA sequence motifs. Rather, non-DNA sequence determinants mark centromeres in an epigenetic manner in many organisms. Some epigenetic determinants of centromere identity in fungi include early replicating regions of the genome (57-59), proximity to DNA replication origins (60), DNA replication initiator proteins (61), homologous recombination-repair proteins (60, 62) and proteins that facilitate kinetochore clustering by tethering kinetochores to SPBs (22, 50). Factors that favor local folding and looping of chromatin may also add to the process of centromere specification (4, 48, 63). Repeats and transposons have been shown to play an essential role in centromere evolution (64-66). Previous reports in *M. oryzae* suggested the presence of multiple clusters of repeat elements across the genome (40, 44). These studies also proposed that repeats play an important role in *M. oryzae* genome evolution and its association with the host. In this study, we find that the centromere location is close to these repeat clusters in some but not all chromosomes. Our results raise the possibility that centromere sequences in *M. oryzae* are prone to repeat-mediated evolution.

A comparison between two *M. oryzae* isolates, Guy11 and FJ81278, revealed that while the overall *CEN* DNA sequence between the two isolates is very similar, the repeat content at the centromeres of orthologous chromosomes varies widely. It is known that centromere DNA sequence among isolates of *N. crassa* can be different (12). The *CEN* sequences identified here would pave the way for a more detailed comparative analysis of centromeres in diverse isolates of *M. oryzae.* Such analyses will provide valuable insights into centromere evolution in this species and the potential impact of host factors on this process. A comparative genome analysis between *M. oryzae* and *M. poae* revealed the presence of a higher density of repeats in the latter. Overall, these results suggest that while the centromere DNA sequence properties, not the DNA sequence *per se*, remain conserved in this species complex, the centromere architecture is divergent and might have been shaped by the repeat elements. Further studies will provide more insights into the evolution of centromere DNA sequences and its possible link to host adaptation and variability in virulence within the *Magnaporthe* species complex.

Materials and Methods

Wild-type *M. oryzae* strain Guy11 (MAT1-2; a kind gift from Lebrun group, France) was used as the parent strain for all the experiments conducted in this study (except for the results shown in SI Appendix, Figure S3, Movie S5 and S6 that were performed using B157 strain). The strains thus validated and used in this study are listed in SI Appendix, Table S4. The plasmids and primers used for epifluorescence labeling in *M. oryzae* strains are listed in SI Appendix, Table S5, and S6, respectively. The detailed information about the plasmid construction is available in the SI Appendix, supplementary materials, and methods. Details of all the experimental procedures and sequence analysis are given in SI Appendix, Materials and

Methods. The ChIP-sequencing reads have been deposited under NCBI BioProject Accession ID PRJNA504461.

Acknowledgments

We thank Clevergene Biocorp Pvt. Ltd., Bengaluru for generating the CenpA ChIP-sequencing reads. VY was a Senior Research Fellow (SRF, grant number 09/733(0179)/2012/EMR-I) supported by Council of Scientific and Industrial Research (CSIR), Government of India. FY is supported by a postdoctoral fellowship from Temasek Life Sciences Laboratory (TLL) Singapore. MHR is a National Postdoctoral Fellow (N-PDF, fellowship reference number PDF/2016/002858), supported by the Science and Engineering Research Board (SERB), Department of Science and Technology (DST), Government of India. NIN acknowledges funding support from TLL, and the National Research Foundation (PMO, Singapore; NRF-CRP7-2010-02 and NRF-CRP16-2015-04). KS 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.

References

1. Musacchio A & Desai A (2017) A Molecular View of Kinetochore Assembly and Function. *Biology* 6(1).
2. Cheeseman IM (2014) The kinetochore. *Cold Spring Harb Perspect Biol* 6(7):a015826.
3. Henikoff S, Ahmad K, & Malik HS (2001) The centromere paradox: Stable inheritance with rapidly evolving DNA. *Science* 293(5532):1098-1102.
4. Yadav V, Sreekumar L, Guin K, & Sanyal K (2018) Five pillars of centromeric chromatin in fungal pathogens. *PLoS Pathog* 14(8):e1007150.
5. Roy B & Sanyal K (2011) Diversity in requirement of genetic and epigenetic factors for centromere function in fungi. *Eukaryot Cell* 10(11):1384-1395.
6. Chatterjee G, *et al.* (2016) Repeat-associated fission yeast-like regional centromeres in the ascomycetous budding yeast *Candida tropicalis*. *PLoS Genet* 12(2):e1005839.
7. Pidoux AL & Allshire RC (2005) The role of heterochromatin in centromere function. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 360(1455):569-579.
8. 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.
9. 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.
10. 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.
11. 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.
12. Smith KM, Galazka JM, Phatale PA, Connolly LR, & Freitag M (2012) Centromeres of filamentous fungi. *Chromosome Res* 20(5):635-656.
13. 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(12):2528-2542.
14. Comai L, Maheshwari S, & Marimuthu MPA (2017) Plant centromeres. *Curr Opin Plant Biol* 36:158-167.
15. Westhorpe FG & Straight AF (2014) The centromere: epigenetic control of chromosome segregation during mitosis. *Cold Spring Harb Perspect Biol* 7(1):a015818.
16. Aldrup-Macdonald ME & Sullivan BA (2014) The past, present, and future of human centromere genomics. *Genes (Basel)* 5(1):33-50.
17. Stellfox ME, Bailey AO, & Foltz DR (2013) Putting CENP-A in its place. *Cell Mol Life Sci* 70(3):387-406.
18. Funabiki H, Hagan I, Uzawa S, & Yanagida M (1993) Cell cycle-dependent specific positioning and clustering of centromeres and telomeres in fission yeast. *J Cell Biol* 121(5):961-976.
19. Kozubowski L, *et al.* (2013) Ordered kinetochore assembly in the human-pathogenic basidiomycetous yeast *Cryptococcus neoformans*. *mBio* 4(5):e00614-00613.
20. Kursel LE & Malik HS (2016) Centromeres. *Curr Biol* 26(12):R487-490.

21. Thakur J & Sanyal K (2012) A coordinated interdependent protein circuitry stabilizes the kinetochore ensemble to protect CENP-A in the human pathogenic yeast *Candida albicans*. *PLoS Genet* 8(4):e1002661.
22. Yadav V & Sanyal K (2018) Sad1 Spatiotemporally Regulates Kinetochore Clustering To Ensure High-Fidelity Chromosome Segregation in the Human Fungal Pathogen *Cryptococcus neoformans*. *mSphere* 3(4).
23. Jin QW, Fuchs J, & Loidl J (2000) Centromere clustering is a major determinant of yeast interphase nuclear organization. *J Cell Sci* 113 (Pt 11):1903-1912.
24. Sanyal K & Carbon J (2002) The CENP-A homolog CaCse4p in the pathogenic yeast *Candida albicans* is a centromere protein essential for chromosome transmission. *Proc Natl Acad Sci U S A.* 99(20):12969-12974.
25. Hou H, *et al.* (2012) Csi1 links centromeres to the nuclear envelope for centromere clustering. *J Cell Biol* 199(5):735-744.
26. 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.
27. Padeken J, *et al.* (2013) The nucleoplasmin homolog NLP mediates centromere clustering and anchoring to the nucleolus. *Mol Cell* 50(2):236-249.
28. Sun S, *et al.* (2017) Fungal genome and mating system transitions facilitated by chromosomal translocations involving intercentromeric recombination. *PLoS Biol* 15(8):e2002527.
29. Cam HP, *et al.* (2005) Comprehensive analysis of heterochromatin- and RNAi-mediated epigenetic control of the fission yeast genome. *Nature genetics* 37(8):809-819.
30. Diner RE, *et al.* (2017) Diatom centromeres suggest a mechanism for nuclear DNA acquisition. *Proc Natl Acad Sci U S A* 114(29):E6015-E6024.
31. Lynch DB, Logue ME, Butler G, & Wolfe KH (2010) Chromosomal G + C content evolution in yeasts: systematic interspecies differences, and GC-poor troughs at centromeres. *Genome biology and evolution* 2:572-583.
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. Talbot NJ (2003) On the trail of a cereal killer: Exploring the biology of *Magnaporthe grisea*. *Annu Rev Microbiol* 57:177-202.
34. Fernandez J & Orth K (2018) Rise of a Cereal Killer: The Biology of *Magnaporthe oryzae* Biotrophic Growth. *Trends Microbiol* 26(7):582-597.
35. Okagaki LH, *et al.* (2015) Genome Sequences of Three Phytopathogenic Species of the *Magnaporthaceae* Family of Fungi. *G3 (Bethesda)* 5(12):2539-2545.
36. Dean RA, *et al.* (2005) The genome sequence of the rice blast fungus *Magnaporthe grisea*. *Nature* 434(7036):980-986.
37. Ebbole DJ (2007) *Magnaporthe* as a model for understanding host-pathogen interactions. *Annu Rev Phytopathol* 45:437-456.
38. McIntosh JR & O'Toole ET (1999) Life cycles of yeast spindle pole bodies: getting microtubules into a closed nucleus. *Biol Cell* 91(4-5):305-312.
39. Bao J, *et al.* (2017) PacBio Sequencing Reveals Transposable Elements as a Key Contributor to Genomic Plasticity and Virulence Variation in *Magnaporthe oryzae*. *Mol Plant* 10(11):1465-1468.

40. Jeon J, *et al.* (2015) Genome-wide profiling of DNA methylation provides insights into epigenetic regulation of fungal development in a plant pathogenic fungus, *Magnaporthe oryzae*. *Sci Rep* 5:8567.
41. Cao H, *et al.* (2016) Characterization of 47 Cys2 -His2 zinc finger proteins required for the development and pathogenicity of the rice blast fungus *Magnaporthe oryzae*. *New Phytol* 211(3):1035-1051.
42. Chao C-CT & Ellingboe AH (1991) Selection for mating competence in *Magnaporthe grisea* pathogenic to rice. *Canadian Journal of Botany* 69(10):2130-2134.
43. Leung H, Borromeo ES, Bernardo MA, & Notteghem JL (1988) Genetic analysis of virulence in the Rice Blast fungus *Magnaporthe grisea*. *Phytopathology* 78:1227-1233.
44. Thon MR, *et al.* (2006) The role of transposable element clusters in genome evolution and loss of synteny in the rice blast fungus *Magnaporthe oryzae*. *Genome Biol* 7(2):R16.
45. Kobayashi N, *et al.* (2015) Discovery of an unconventional centromere in budding yeast redefines evolution of point centromeres. *Curr Biol* 25(15):2026-2033.
46. Rhind N, *et al.* (2011) Comparative functional genomics of the fission yeasts. *Science* 332(6032):930-936.
47. Shang WH, *et al.* (2013) Chromosome engineering allows the efficient isolation of vertebrate neocentromeres. *Dev Cell* 24(6):635-648.
48. Thakur J & Sanyal K (2013) Efficient neocentromere formation is suppressed by gene conversion to maintain centromere function at native physical chromosomal loci in *Candida albicans*. *Genome Res* 23(4):638-652.
49. Simon L, Voisin M, Tatout C, & Probst AV (2015) Structure and function of centromeric and pericentromeric heterochromatin in *Arabidopsis thaliana*. *Front Plant Sci* 6:1049.
50. Fernandez-Alvarez A, Bez C, O'Toole ET, Morphew M, & Cooper JP (2016) Mitotic nuclear envelope breakdown and spindle nucleation are controlled by interphase contacts between centromeres and the nuclear envelope. *Dev Cell* 39(5):544-559.
51. Hou H, Kallgren SP, & Jia S (2013) Csi1 illuminates the mechanism and function of Rabl configuration. *Nucleus* 4(3).
52. Jaspersen SL, *et al.* (2006) The Sad1-UNC-84 homology domain in Mps3 interacts with Mps2 to connect the spindle pole body with the nuclear envelope. *J Cell Biol* 174(5):665-675.
53. Richmond D, Rizkallah R, Liang F, Hurt MM, & Wang Y (2013) Slk19 clusters kinetochores and facilitates chromosome bipolar attachment. *Mol Biol Cell* 24(5):566-577.
54. Baker RE & Rogers K (2005) Genetic and genomic analysis of the AT-rich centromere DNA element II of *Saccharomyces cerevisiae*. *Genetics* 171(4):1463-1475.
55. Ohzeki J, Nakano M, Okada T, & Masumoto H (2002) CENP-B box is required for de novo centromere chromatin assembly on human alphoid DNA. *J Cell Biol* 159(5):765-775.
56. Xiao H, *et al.* (2011) Nonhistone Scm3 binds to AT-rich DNA to organize atypical centromeric nucleosome of budding yeast. *Mol Cell* 43(3):369-380.
57. Koren A, *et al.* (2010) Epigenetically-inherited centromere and neocentromere DNA replicates earliest in S-phase. *PLoS Genet* 6(8):e1001068.
58. Kim SM, Dubey DD, & Huberman JA (2003) Early-replicating heterochromatin. *Genes Dev* 17(3):330-335.

59. Raghuraman MK, *et al.* (2001) Replication dynamics of the yeast genome. *Science* 294(5540):115-121.
60. 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.
61. Sreekumar L, *et al.* (2018) DNA replication initiator proteins facilitate CENPA loading on early replicating compact chromatin. *bioRxiv*.
62. Cook DM, *et al.* (2018) Fork pausing allows centromere DNA loop formation and kinetochore assembly. *Proc Natl Acad Sci U S A* 115(46):11784-11789.
63. Anderson M, Haase J, Yeh E, & Bloom K (2009) Function and assembly of DNA looping, clustering, and microtubule attachment complexes within a eukaryotic kinetochore. *Mol Biol Cell* 20(19):4131-4139.
64. 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.
65. Neumann P, *et al.* (2011) Plant centromeric retrotransposons: a structural and cytogenetic perspective. *Mob DNA* 2(1):4.
66. Wong LH & Choo KH (2004) Evolutionary dynamics of transposable elements at the centromere. *Trends Genet* 20(12):611-616.

Figure legends

Figure 1. Localization patterns of CenpA and CenpC reveal kinetochores are closely associated with each other in *M. oryzae*. (A) The *M. oryzae* strain MGYF03 exhibits a single dot-like GFP-CenpA signal localized at the periphery of each nucleus marked by mCherry-histone H1 in both mycelia (upper panel) and conidia (lower panel). **(B)** Similarly, another inner kinetochore protein CenpC-GFP in the strain MGYF04 is found to be localized at the periphery of the mCherry-histone H1 marked nucleus in both mycelia (upper panels) and conidia (lower panels). **(C)** Co-localization of GFP-CenpA and CenpC-mCherry revealed complete overlapping signals in both mycelia and conidia in the MGYF05 strain. **(D)** In the strain MGYF08, the clusters of GFP-CenpA are closely associated with the spindle pole body (SPB) component Alp6-mCherry. In addition to SPBs, the Alp6 signals were also observed at the septa (white arrows). The fluorescence images shown here are maximum projections from Z stacks consisting of 0.5 µm-spaced planes. Scale bar = 10 µm.

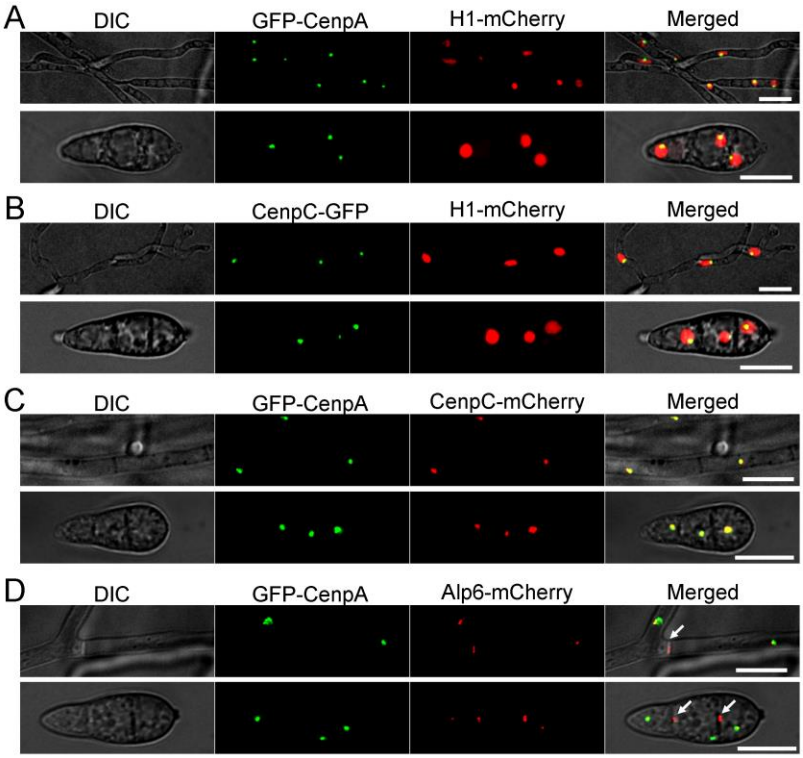
Figure 2. Kinetochores decluster momentarily but arrange randomly on the spindle axis before sister kinetochore separation during anaphase in *M. oryzae*. (A) Time-lapse imaging of strain MGYF07 cells exhibited that the GFP-CenpA signals separate from each other and move along the mitotic spindle (mCherry-TubA). Also see movie S1. The images shown are maximum projections of 0.3 µm-spaced *Z* stacks. Scale bar = 2 µm. **(B)** High-resolution time-lapse images showing the declustering of kinetochores (GFP-CenpA) during the process of mitosis in strain MGYF01 (Also, see Movie S2). The images were acquired with *Z* projections of 0.17-µm step size. Scale bar =1 µm. **(C)** High-resolution time-lapse images of MGYF01 cells showing the segregation dynamics of sister kinetochores in daughter cells during metaphase to anaphase transition and finally reclustering of kinetochores in post-anaphase cells (Also see movie S3). Scale bar = 2 µm. **(D)** Spatial organization of kinetochores (GFP-CenpA) and SPBs (Alp6-mCherry) in strain MGYF08 during pre-mitotic stage (upper panel) and early mitosis (lower panel). Scale bar = 2 µm. **(E)** A schematic depiction of centromere dynamics at specific stages of the cell cycle in *M. oryzae*. For simplification, chromosomes and astral microtubules are omitted in the schematic.

Figure 3. Identification of centromeres in *M. oryzae*. (A) Reads obtained from the GFP-CenpA ChIP-sequencing in the cross-linked mycelia of the strain MGYF01 identified one distinct enriched region on each of the seven contigs when aligned to the Guy11 genome assembly (see SI Appendix, Figure S4 for the remaining contigs). CenpA-bound regions overlap with AT-rich, poorly-transcribed regions on each contig, and harbor 5mC DNA methylation (see text for details). The numbers in the bracket with parameters represent the minimum and

maximum value along the *y*-axis. **(B)** The zoomed view of centromere regions in Guy11 depicting the presence of repeat elements, CenpA enrichment, polyA transcription and DNA methylation (5mC) status in these regions. A 200 kb region spanning the centromere is shown for each chromosome. The only common defining sequence feature of centromeres is AT-richness.

Figure 4. Centromere DNA sequences in *M. oryzae* isolates are similar but vary in repeat content. **(A)** A map showing seven chromosomes of *M. oryzae* with centromere locations marked on each chromosome. The chromosome length along with centromere length obtained from the ChIP-seq analysis is plotted to the scale on the available chromosome-wide 70-15 genome assembly. However, telomeres are shown as 10 kb regions on either side for each chromosome for visualization purpose. **(B)** Map showing synteny conservation across centromere flanking regions among three isolates of *M. oryzae*. A 200 kb region including the centromere (Guy11 as reference) is shown. Also, see SI Appendix, Figure S9. **(C)** Dot-plot analysis of centromere sequences revealed centromere sequences between Guy11 and FJ81278 are similar but differ in the repeat content. R denotes the repeat panels for both Guy11 and FJ81278. Also, see SI Appendix, Figure S10.

1 Figure 1



2

3

Figure 2

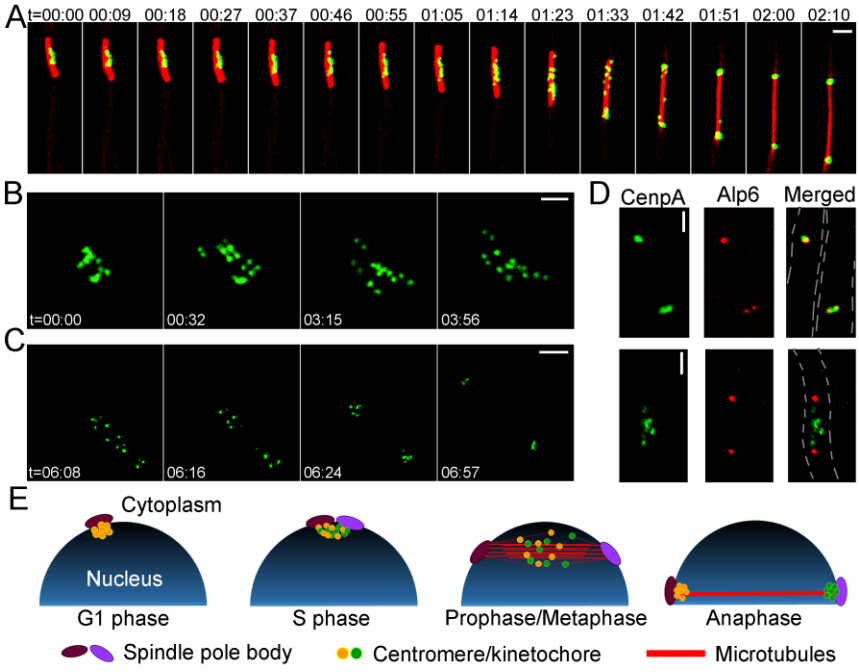
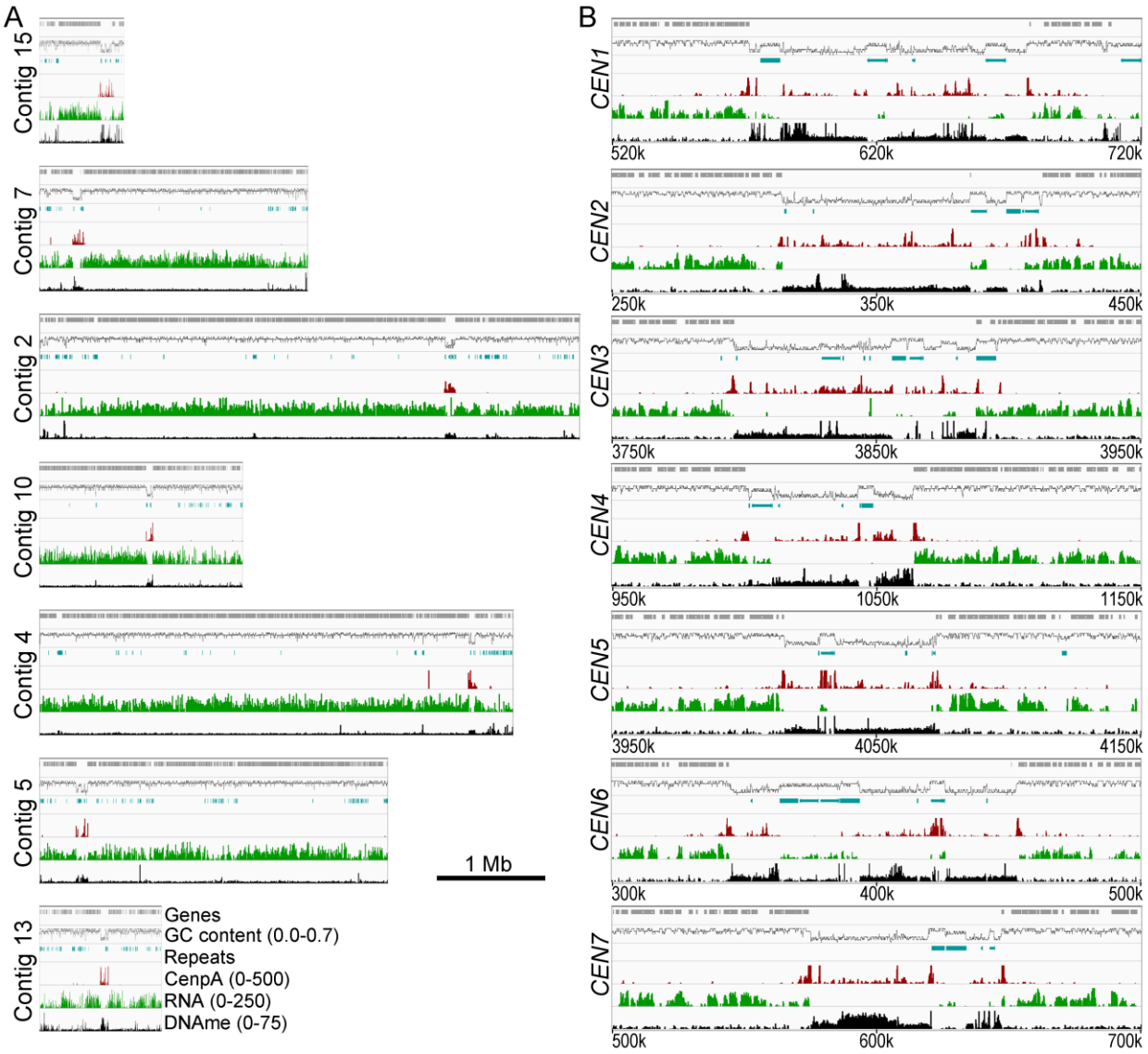
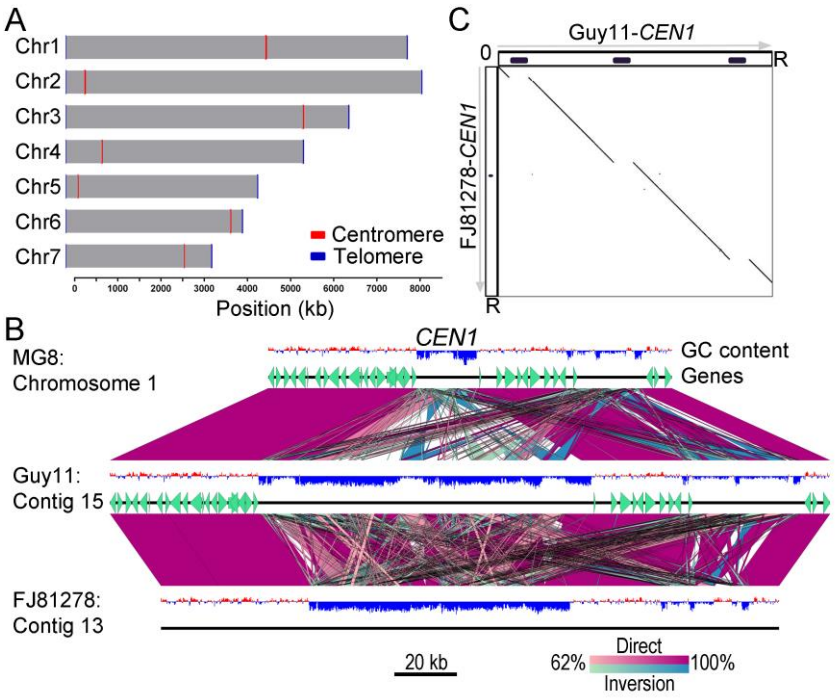


Figure 3



1 Figure 4



2

3

Table 1. **Length and GC% of centromeres in *Magnaporthe oryzae*.**

CEN #	**Centromere coordinates in Guy11**	**% GC content (Genomic GC content = 51.2%)**
CEN1	Contig15: 571735-678893 **(1,07,159)**	29.2
CEN2	Contig7: 313767-411084 **(97,318)**	30.8
CEN3	Contig2: 3795849-3894639 **(98,791)**	33.0
CEN4	Contig10: 1000090-1063263 **(63,174)**	30.5
CEN5	Contig4: 4014470-4071774 **(57,305)**	28.0
CEN6	Contig5: 343714-452391 **(1,08,678)**	32.3
CEN7	Contig13: 573351-645424 **(72,074)**	30.5

Centromere numbers are according to the 70-15 genome assembly.
Numbers in the bracket represent centromere length in base pair.