Cellular Dynamics and Genomic Identity of Centromeres in the Cereal Blast Fungus

Vikas Yadav[1¶,#a], Fan Yang[2¶], Md. Hashim Reza[1], Sanzhen Liu[3], Barbara Valent[3], Kaustuv Sanyal[1*], 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, and Department of Biological Sciences, 1 Research Link, Singapore

[3]Department of Plant Pathology, Kansas State University, Manhattan, KS, USA

[#a]Current address: Department of Molecular Genetics and Microbiology, Duke University Medical Center, Durham, NC, USA

[*] Corresponding authors:

E-mail: naweed@tll.org.sg (NIN); sanyal@jncasr.ac.in (KS)

[¶]These authors contributed equally to this work.

Running title: Centromeres in *M. 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 rice blast fungus. PacBio genome assemblies and synteny analyses facilitated comparison of the centromere regions in distinct isolate(s) of rice blast, wheat blast, and in *M. poae*. Overall, this study identified unusual centromere dynamics and precisely mapped the centromere DNA sequences in the top model fungal pathogens that belong to the *Magnaporthales* and cause severe losses to global production of food crops and turf grasses.

Author summary

Magnaporthe oryzae is an important fungal pathogen that causes an annual loss of 10-30% of the 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 in the economically important cereal blast and summer patch pathogens. This study paves the way for understanding the contribution of heterochromatin in genome stability and virulence of the blast fungus.

Introduction

Faithful chromosome segregation is one of the essential processes 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 a few kilobases to several megabases. Unlike point centromeres, regional centromeres in an organism can span hundred of kbs and are often are defined by epigenetic factors. 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.

Magnaporthales is an order of ascomycete fungi comprising of many important plant pathogenic species including *Magnaporthe oryzae* (synonym of *Pyricularia oryzae*) and *Magnaporthe poae* (synonym of *Magnaporthiopsis poae*) [33]. *M. oryzae* includes host-adapted lineages (pathotypes) causing the devastating blast diseases in cereal crops including rice, wheat, barley and millets [34-36]. *M. poae* is responsible for summer-patch disease in turf grasses [37]. The *M. oryzae* lineage causes rice blast which remains a constant threat to agriculture-based economies due to significant damage to rice harvests. Recently, wheat blast disease, caused by the *M. oryzae Triticum* lineage has emerged as a major threat to global wheat production [38]. Rice blast has also become 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 [39, 40]. However, even with the availability of the genome sequence and annotated assembly, the centromere/kinetochore identity of the 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 additional conserved kinetochore proteins using *in silico* predictions (S1 Table). 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) (S1 Fig) 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 single dot-like and co-localized on chromatin, marked by mCherry-tagged histone H1 (Fig 1A and 1B). Further, co-localization of CenpA and CenpC signals confirmed their overlapping spatial positions in both mycelia and conidia (Fig 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) [41]. We localized SPBs by tagging Alp6 (MGG_01815, an ortholog of *S. cerevisiae* Spc98) with mCherry and observed that SPBs localize close to the clustered GFP-CenpA signals in *M. oryzae* (Fig 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 (S2A Fig). 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 (S2B Fig). Live-cell imaging during mitosis revealed dispersed GFP-CenpA signals localized along the mitotic spindle (Fig 2A and S1 Movie). 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 (Fig 2B, C and S2, S3 Movie). We observed that while the GFP-CenpA signals were spread out, they were localized in pairs, most likely representing the segregated kinetochore signals (Fig 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 (Fig 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 (S2C Fig and S4 Movie). 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 (S3 Fig, S5 and S6 Movie). Based on these observations, we propose a schematic model for the kinetochore and SPB dynamics during the mitotic cycle in rice blast where the dynamics of kinetochore clustering-declustering is most likely dependent on their direct link to the SPBs (Fig 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* [42]. This analysis revealed seven distinct CenpA-rich regions across the genome, one each on seven different contigs (Fig 3, Table 1 and S4 Fig). 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 independently by binding of another evolutionarily conserved kinetochore protein CenpC. ChIP-qPCR using the fungal strain expressing CenpC-GFP (Fig 3C) confirmed specific overlapping binding of CenpA and CenpC on each of these seven *CEN* regions. We also noticed

an extra albeit short region of 1200 bp on Contig 4 apart from the seven distinct peaks in CenpA ChIP-seq analysis. The enriched peak mapped to the gene encoding the vacuolar morphogenesis protein AvaB (MGG_01045). Using specific ChIP-qPCR primers for this region, the aforementioned CenpA enrichment on Contig 4 was confirmed to be an artifact (Fig 3D). 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%) (Fig 3B and Table 1). The centromeres in *M. oryzae* harbor a few repetitive elements (Fig 3B, S1 Dataset). 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 (S5 Fig). We then examined the transcriptional status and base modifications associated with centromeric chromatin using the published RNA-sequencing and bisulfite sequencing data [43, 44]. Centromeres in *M. oryzae* are found to be poorly transcribed and harbor 5mC DNA methylation (Fig 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 rice blast isolates

The MG8 genome assembly is based on the sequencing of the *M. oryzae* isolate 70-15, which represents a progeny of the Guy11 strain [39, 45, 46]. 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 (MG8) available for an *M. oryzae* rice pathogen (S6A Fig). 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 (Fig 4 and Table 1). We also observed two additional CenpA-enriched regions in the unassembled Supercontig8.8 of MG8 assembly for 70-15 (S6B Fig). Additionally, the identified centromere on chromosome 7 in this assembly matched the region previously predicted to harbor the centromere based on genetic analysis [47].

Next, we analyzed the recently published PacBio genome assembly of the *M. oryzae* field isolate FJ81278 [42] 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 (S7 Fig and Table 1). 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.

We further 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* (Fig

5A). However, a major part of the centromere sequences was absent in 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 certain 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 (Fig 5B and S2 Dataset). 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.

Intra- and Inter-species comparison of *CEN* sequences in *Magnaporthales*

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 decided to predict the centromeres in the wheat blast isolate B71 (Triticum pathotype of *M. oryzae;* MoT) as well as in *M. poae*, a root infecting pathogen that belongs to the *Magnaporthaceae* family [37]. Wheat blast isolate B71 genome was assembled to a chromosome level and exhibits a few chromosomal rearrangements as compared to the rice blast 70-15 genome assembly [48]. We identified seven putative centromeres, one in each chromosome, in the B71 genome. These centromeres in the wheat blast genome were long AT-rich regions (Fig 6A and Table 1). We then analyzed the centromere flanking regions between two genomes and

found that centromere locations between the rice blast and wheat blast strains are conserved (Fig 6B). Further analysis revealed that the centromeres in wheat blast B71 also harbor a few repeats, but are defined primarily by AT-richness, similar to those observed in the rice blast isolates. We analyzed Scaffold 1 representing the mini-chromosome in B71 [48] and identified a putative centromere based on a long AT-rich region (40,795 – 109,519 bp) (Fig 6C, Table 1).

Next, we extended our *in-silico* analysis to *M. poae* a distinct species within the *Magnaporthales* [33] and identified eight putative centromere regions across its genome (S8 Fig and Table 1). Three of these eight putative *CEN* regions are 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 in *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 sequences in *M. oryzae* and *M. poae* are highly divergent. Based on these results, we conclude that centromere DNA sequences in the *Magnaporthales* are rapidly evolving, whereas the properties of centromeric chromatin are likely conserved between the two species.

Discussion

Blast diseases caused by *M. oryzae* is exceedingly disastrous not only to rice production worldwide but also to wheat and other graminaceous crops [35, 38]. 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 in *M. oryzae* and studied their dynamics during different phases of the cell cycle at various developmental stages. We further identified the genomic loci that act as centromeres in this filamentous fungus. Based on the comparison of centromeric sequences among different host-adapted lineages of *M. oryzae*, and in the related species (*M. poae*), centromeres appear to be rapidly evolving in the *Magnaporthe* species complex, similar to those reported in several fungal genera [6, 8, 10, 12, 49, 50].

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 [51, 52]. 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, 53]. 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, 54-57]. 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 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*. Taken 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 [58]. 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 [59]. 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 [60]. 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 [61-63], proximity to DNA replication origins [64], DNA replication initiator proteins [65], homologous recombination-repair proteins [64, 66] and proteins that facilitate kinetochore clustering by tethering kinetochores to SPBs [22, 54]. Factors that favor local folding and looping of chromatin may also add to the process of centromere specification [4, 52, 67]. Repeats and transposons have been shown to play an essential role in centromere evolution [68-70]. Previous reports in *M. oryzae* suggested the presence of multiple clusters of repeat elements across the genome [43, 47]. 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 rice blast 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 predicted *CEN* regions in the latter. Overall, these results suggest that while the centromere DNA sequence properties, not the DNA sequence *per se,* remain conserved in the fungal order *Magnaporthales*, 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 this important family of cereal killers.

Materials and Methods

Fungal strains and culture conditions

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 S3 Fig, S5 and S6 Movie that were performed using B157 strain). All strains were propagated on prune-agar (PA) medium or complete medium (CM) as described previously [71].

For conidiation, fungal strains were grown on PA agar plates at 28 °C for two days in the dark followed by exposure to continuous light at room temperature for five days. Conidia were harvested using an inoculation loop by gently scraping the culture surface with sterile distilled water. The resulting conidial suspension was filtered through two layers of Miracloth (Calbiochem) to remove mycelial debris and was adjusted to obtain the required concentration after counting with a hemocytometer. *Agrobacterium tumefaciens*-mediated transformation (ATMT) of *M. oryzae* was carried out as described previously [71, 72]. Transformants were screened for antibiotic resistance using the respective selection media, i.e., CM with 250 µg/ml Hygromycin B, Basal media with 50 µg/ml ammonium glufosinate (selection for Basta) or Chlorimuron-ethyl (selection for Sulfonylurea). Transformants were verified for correct genomic integration by diagnostic PCR and sequencing. The strains thus validated and used in this study are listed in S2 Table. The plasmids and primers used for epifluorescence labeling in *M. oryzae* strains are listed in S3 and S4 Table, respectively. The detailed information about the plasmid construction is available in the S1 Text.

Microscopy and image processing

Unless otherwise stated, live cell microscopy imaging was performed on a motorized inverted Nikon Eclipse Ti-E microscope with perfect focus system equipped with a Yokogawa CUS-X1 spinning disk confocal system and a CFI Plan Apo VC 100x/1.4 NA oil objective lens. The images were captured using 16-bit digital Orca-Flash4.0 sCMOS camera (Hamamatsu Photonic K.K) and at 491 nm 100 mW (for green fluorescence) and at 561 nm 50 mW (for red fluorescence) laser illumination operated by MetaMorph Premier Software (Ver. 7.7.5, Universal Imaging). The maximum projection images were obtained from Z stacks of 0.5 µm-spaced

sections using the built-in MetaMorph module. Image processing was performed using Imaris (Bitplane) and Fiji (https://imagej.net/Fiji). A Live-SR module (Gataca Systems), which is based on optically demodulated structured illumination technique with online processing, was additionally mounted on the same spinning disk confocal system during image acquisitions for Figure 2A. The high-resolution images for Figure 2B and 2C were acquired using an Andor Dragonfly high-speed confocal microscope equipped with an iXon888 EMCCD camera and a 100x oil objective lens. The raw images were immediately processed using the integrated Fusion software and the in-built deconvolution feature.

Chromatin immunoprecipitation

The ChIP experiment was performed using the protocol described previously with a few modifications [73]. *M. oryzae* strain expressing GFP-CenpA or CenpC-GFP fusion protein was grown in 150 ml complete media for three days at 28 °C with continuous shaking at 150 rpm. Fungal mycelia were collected using two layers of Miracloth (Calbiochem), and the harvested mycelia were washed with sterile water. Mycelia were cross-linked by suspending them in 1% formaldehyde solution in 20 mM HEPES/pH7.4 for 20 min with continuous shaking at 100 rpm. Glycine was added to the suspension at a final concentration of 0.125M, and the mix was further incubated at room temperature for an additional 10 min. Cross-linked mycelia were harvested using Miracloth and rinsed with water. The excess water was removed by gently patting the mycelia mass with paper towels followed by snap-freezing in liquid nitrogen. The frozen mass was then stored at -80 °C until used. For each ChIP experiment, 80-100 mg of frozen mycelia were ground in liquid nitrogen using mortar-pestle, and powdered mycelia were resuspended in 1 ml of nuclei isolation buffer (10 mM MES-KOH/10 mM NaCl/10 mM KCl/2.5 mM EDTA

pH8.0/250 mM sucrose/0.1 mM spermine/0.5 mM spermidine-free base/1 mM DTT). Nuclei were separated from debris by filtering them through two layers of Miracloth and pelleted by centrifugation at 13000 rpm for 10 min at 4 °C. The pellet was resuspended in 1 ml of lysis buffer (50 mM HEPES pH 7.5/150 mM NaCl/1 mM EDTA/0.1% Na-deoxycholate/1% Triton-X and 0.1% SDS). The resuspended nuclei were subjected to sonication using a Bioruptor (Diagenode) for 60 cycles of the 30s on and 30s off bursts at the high level, and the fragmented chromatin (300 - 600 bp) was isolated by centrifugation at 13000 rpm, 10 mins at 4 °C. A part of the chromatin fraction (100 µl) was used for input DNA (I) preparation and the remaining chromatin solution was divided into two halves of 450 µl each. In one of the tubes, 20 µl of GFP-TRAP beads (ChromoTek) were added for immunoprecipitation (+) while the other tube was incubated with 20 µl of blocked agarose beads (ChromoTek) to use as the negative control (-). The tubes were incubated at 4 °C for 12 h on a rotator. The beads were then washed, and bound chromatin was eluted in 500 µl of elution buffer (1% SDS, 0.1M $NaHCO_3$). All three fractions (I, + and -), were de-crosslinked and DNA was isolated using phenol: chloroform extraction followed by ethanol precipitation. The precipitated DNA was air dried and dissolved in 25 µl of MilliQ water containing 25 µg/ml RNase (Sigma-Aldrich). I and + samples were subjected to ChIP-sequencing for GFP-CenpA ChIP in *M. oryzae*. CenpC-GFP ChIP samples (I, + and -) were subjected to qPCR with centromere-specific primers along with a non-centromeric primer set. The fold enrichment for CenpC at the centromere was plotted using GraphPad Prism software.

Analysis of sequencing data

GFP-CenpA ChIP sequencing was performed at the Clevergene Biocorp. Pvt. Ltd., Bengaluru, India. A total of 52,604,910 and 38,186,320 150 bp paired-end reads were obtained for Input and IP samples, respectively. The reads were mapped to Guy11 PacBio genome (GCA_002368485.1), MG8 genome assembly (GCA_000002495.2) and FJ81278 genome assembly (GCA_002368475.1) using Geneious 9.0 (http://www.geneious.com/) with default conditions. Each read was allowed to map only once randomly anywhere in the genome. The alignments were exported to bam files, sorted and visualized using Integrative Genomics Viewer (IGV, Broad Institute). The images from IGV were imported into Adobe Photoshop (Adobe systems) and scaled for representation purpose. The ChIP-sequencing reads have been deposited under NCBI BioProject Accession ID PRJNA504461.

RNA-sequencing data (SRR1568068) from a previous study was downloaded from the NCBI website and was aligned to genomes using RNA-seq alignment default parameters in Geneious 9.0. Similarly, bisulfite sequencing data (SRR653493) was obtained from NCBI, and the reads were aligned to genomes using Geneious default aligner. The aligned files were exported into bam files and visualized using IGV. The RNA-seq reads were plotted in logarithmic scale for the visualization purpose. GC-content was calculated using Geneious 9.0 with a sliding window size of 200 bp. The data was exported as wig files and further visualized using IGV.

For *M. poae*, the GC content and RNA-seq data were plotted as described for *M. oryzae*. The RNA-seq data was downloaded from NCBI (SRR057701) and was aligned to the reference genome (GCA_000193285.1).

The repeats annotations presented in this study were done based on repeat sequences described in a previous study [37]. BLASTn analysis was carried out using these repeat

sequences, results were sorted, and hits with 100% query coverage were extracted. The hits were mapped onto the respective genome assemblies of Guy11, FJ81278, B71 or *M. poae* and visualized using IGV.

Synteny analysis

The synteny analysis between centromere flanking regions was performed using Easyfig software [74] (http://mjsull.github.io/Easyfig/). The graphs were plotted with default condition (except the color settings), and GC percentage was plotted as 200 bp sliding window. The dot-plot analysis for centromere sequences was carried out using the gepard software with a 100-bp window [75].

Acknowledgments

We thank Clevergene Biocorp Pvt. Ltd., Bengaluru for generating the CenpA ChIP-sequencing reads. We thank the Fungal Pathobiology group at TLL for useful discussions and suggestions; and the TLL Bioimaging Facility for help with confocal microscopy and image analyses.

References

1. Musacchio A, Desai A. A Molecular View of Kinetochore Assembly and Function. Biology. 2017;6(1). doi: 10.3390/biology6010005. PubMed PMID: 28125021; PubMed Central PMCID: PMC5371998.

2. Cheeseman IM. The kinetochore. Cold Spring Harb Perspect Biol. 2014;6(7):a015826. doi: 10.1101/cshperspect.a015826. PubMed PMID: 24984773; PubMed Central PMCID: PMC4067989.

3. Henikoff S, Ahmad K, Malik HS. The centromere paradox: Stable inheritance with rapidly evolving DNA. Science. 2001;293(5532):1098-102. Epub 2001/08/11. doi: 10.1126/science.1062939
293/5532/1098 [pii]. PubMed PMID: 11498581.
4. Yadav V, Sreekumar L, Guin K, Sanyal K. Five pillars of centromeric chromatin in fungal pathogens. PLoS Pathog. 2018;14(8):e1007150. Epub 2018/08/24. doi: 10.1371/journal.ppat.1007150. PubMed PMID: 30138484.
5. Roy B, Sanyal K. Diversity in requirement of genetic and epigenetic factors for centromere function in fungi. Eukaryot Cell. 2011;10(11):1384-95. Epub 2011/09/13. doi: EC.05165-11 [pii]
10.1128/EC.05165-11. PubMed PMID: 21908596; PubMed Central PMCID: PMC3209047.
6. Chatterjee G, Sankaranarayanan SR, Guin K, Thattikota Y, Padmanabhan S, Siddharthan R, et al. Repeat-associated fission yeast-like regional centromeres in the ascomycetous budding yeast *Candida tropicalis*. PLoS Genet. 2016;12(2):e1005839. doi: 10.1371/journal.pgen.1005839. PubMed PMID: 26845548; PubMed Central PMCID: PMC4741521.
7. Pidoux AL, Allshire RC. The role of heterochromatin in centromere function. Philosophical transactions of the Royal Society of London Series B, Biological sciences. 2005;360(1455):569-79. doi: 10.1098/rstb.2004.1611. PubMed PMID: 15905142; PubMed Central PMCID: PMC1569473.
8. Yadav V, Sun S, Billmyre RB, Thimmappa BC, Shea T, Lintner R, et al. RNAi is a critical determinant of centromere evolution in closely related fungi. Proc Natl Acad Sci U S A. 2018;115(12):3108-13. doi: 10.1073/pnas.1713725115. PubMed PMID: 29507212; PubMed Central PMCID: PMC5866544.
9. Kapoor S, Zhu L, Froyd C, Liu T, Rusche LN. Regional centromeres in the yeast *Candida lusitaniae* lack pericentromeric heterochromatin. Proc Natl Acad Sci U S A. 2015;112(39):12139-44. doi: 10.1073/pnas.1508749112. PubMed PMID: 26371315; PubMed Central PMCID: PMC4593121.
10. Padmanabhan S, Thakur J, Siddharthan R, Sanyal K. 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. 2008;105(50):19797-802. Epub 2008/12/09. doi: 0809770105 [pii]
10.1073/pnas.0809770105. PubMed PMID: 19060206; PubMed Central PMCID: PMC2604992.

11. Sanyal K, Baum M, Carbon J. Centromeric DNA sequences in the pathogenic yeast *Candida albicans* are all different and unique. Proc Natl Acad Sci U S A. 2004;101(31):11374-9. doi: 10.1073/pnas.0404318101.

12. Smith KM, Galazka JM, Phatale PA, Connolly LR, Freitag M. Centromeres of filamentous fungi. Chromosome Res. 2012;20(5):635-56. PubMed PMID: 22752455.

13. Smith KM, Phatale PA, Sullivan CM, Pomraning KR, Freitag M. Heterochromatin is required for normal distribution of *Neurospora crassa* CenH3. Mol Cell Biol. 2011;31(12):2528-42. Epub 2011/04/21. doi: MCB.01285-10 [pii]
10.1128/MCB.01285-10. PubMed PMID: 21505064; PubMed Central PMCID: PMC3133421.

14. Comai L, Maheshwari S, Marimuthu MPA. Plant centromeres. Curr Opin Plant Biol. 2017;36:158-67. doi: 10.1016/j.pbi.2017.03.003. PubMed PMID: 28411416.

15. Westhorpe FG, Straight AF. The centromere: epigenetic control of chromosome segregation during mitosis. Cold Spring Harb Perspect Biol. 2014;7(1):a015818. doi: 10.1101/cshperspect.a015818. PubMed PMID: 25414369; PubMed Central PMCID: PMC4292168.

16. Aldrup-Macdonald ME, Sullivan BA. The past, present, and future of human centromere genomics. Genes (Basel). 2014;5(1):33-50. PubMed PMID: 24683489; PubMed Central PMCID: PMC3966626.

17. Stellfox ME, Bailey AO, Foltz DR. Putting CENP-A in its place. Cell Mol Life Sci. 2013;70(3):387-406. doi: 10.1007/s00018-012-1048-8. PubMed PMID: 22729156; PubMed Central PMCID: PMC4084702.

18. Funabiki H, Hagan I, Uzawa S, Yanagida M. Cell cycle-dependent specific positioning and clustering of centromeres and telomeres in fission yeast. J Cell Biol. 1993;121(5):961-76. PubMed PMID: 8388878.

19. Kozubowski L, Yadav V, Chatterjee G, Sridhar S, Yamaguchi M, Kawamoto S, et al. Ordered kinetochore assembly in the human-pathogenic basidiomycetous yeast *Cryptococcus neoformans*. mBio. 2013;4(5):e00614-13. Epub 2013/10/03. doi: mBio.00614-13 [pii]
10.1128/mBio.00614-13. PubMed PMID: 24085781; PubMed Central PMCID: PMC3791896.

20. Kursel LE, Malik HS. Centromeres. Curr Biol. 2016;26(12):R487-90. doi: 10.1016/j.cub.2016.05.031. PubMed PMID: 27326706.

21. Thakur J, Sanyal K. A coordinated interdependent protein circuitry stabilizes the kinetochore ensemble to protect CENP-A in the human pathogenic yeast *Candida albicans*. PLoS Genet. 2012;8(4):e1002661. Epub 2012/04/27. doi: 10.1371/journal.pgen.1002661 PGENETICS-D-11-00911 [pii]. PubMed PMID: 22536162; PubMed Central PMCID: PMC3334883.

22. Yadav V, Sanyal K. Sad1 Spatiotemporally Regulates Kinetochore Clustering To Ensure High-Fidelity Chromosome Segregation in the Human Fungal Pathogen *Cryptococcus neoformans*. mSphere. 2018;3(4). Epub 2018/07/07. doi: 10.1128/mSphere.00190-18. PubMed PMID: 29976642; PubMed Central PMCID: PMCPMC6034078.

23. Jin QW, Fuchs J, Loidl J. Centromere clustering is a major determinant of yeast interphase nuclear organization. J Cell Sci. 2000;113 (Pt 11):1903-12. Epub 2000/05/12. PubMed PMID: 10806101.

24. Sanyal K, Carbon J. 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 2002;99(20):12969-74. doi: 10.1073/pnas.162488299.

25. Hou H, Zhou Z, Wang Y, Wang J, Kallgren SP, Kurchuk T, et al. Csi1 links centromeres to the nuclear envelope for centromere clustering. J Cell Biol. 2012;199(5):735-44. Epub 2012/11/21. doi: jcb.201208001 [pii] 10.1083/jcb.201208001. PubMed PMID: 23166349; PubMed Central PMCID: PMC3514793.

26. Schotanus K, Soyer JL, Connolly LR, Grandaubert J, Happel P, Smith KM, et al. Histone modifications rather than the novel regional centromeres of *Zymoseptoria tritici* distinguish core and accessory chromosomes. Epigenetics Chromatin. 2015;8:41. Epub 2015/10/03. doi: 10.1186/s13072-015-0033-5. PubMed PMID: 26430472; PubMed Central PMCID: PMCPMC4589918.

27. Padeken J, Mendiburo MJ, Chlamydas S, Schwarz HJ, Kremmer E, Heun P. The nucleoplasmin homolog NLP mediates centromere clustering and anchoring to the nucleolus. Mol Cell. 2013;50(2):236-49. doi: 10.1016/j.molcel.2013.03.002. PubMed PMID: 23562326.

28. Sun S, Yadav V, Billmyre RB, Cuomo CA, Nowrousian M, Wang L, et al. Fungal genome and mating system transitions facilitated by chromosomal translocations involving

intercentromeric recombination. PLoS Biol. 2017;15(8):e2002527. doi: 10.1371/journal.pbio.2002527. PubMed PMID: 28800596.

29. Cam HP, Sugiyama T, Chen ES, Chen X, FitzGerald PC, Grewal SI. Comprehensive analysis of heterochromatin- and RNAi-mediated epigenetic control of the fission yeast genome. Nature genetics. 2005;37(8):809-19. doi: 10.1038/ng1602. PubMed PMID: 15976807.

30. Diner RE, Noddings CM, Lian NC, Kang AK, McQuaid JB, Jablanovic J, et al. Diatom centromeres suggest a mechanism for nuclear DNA acquisition. Proc Natl Acad Sci U S A. 2017;114(29):E6015-E24. doi: 10.1073/pnas.1700764114. PubMed PMID: 28673987; PubMed Central PMCID: PMC5530661.

31. Lynch DB, Logue ME, Butler G, Wolfe KH. Chromosomal G + C content evolution in yeasts: systematic interspecies differences, and GC-poor troughs at centromeres. Genome biology and evolution. 2010;2:572-83. doi: 10.1093/gbe/evq042. PubMed PMID: 20693156; PubMed Central PMCID: PMC2997560.

32. Zhu Y, Engstrom PG, Tellgren-Roth C, Baudo CD, Kennell JC, Sun S, et al. Proteogenomics produces comprehensive and highly accurate protein-coding gene annotation in a complete genome assembly of *Malassezia sympodialis*. Nucleic Acids Res. 2017;45(5):2629-43. doi: 10.1093/nar/gkx006. PubMed PMID: 28100699; PubMed Central PMCID: PMC5389616.

33. Zhang N, Cai G, Price DC, Crouch JA, Gladieux P, Hillman B, et al. Genome wide analysis of the transition to pathogenic lifestyles in Magnaporthales fungi. Scientific Reports. 2018;8(1):5862. doi: 10.1038/s41598-018-24301-6.

34. Talbot NJ. On the trail of a cereal killer: Exploring the biology of *Magnaporthe grisea*. Annu Rev Microbiol. 2003;57:177-202. Epub 2003/10/07. doi: 10.1146/annurev.micro.57.030502.090957. PubMed PMID: 14527276.

35. Fernandez J, Orth K. Rise of a Cereal Killer: The Biology of *Magnaporthe oryzae* Biotrophic Growth. Trends Microbiol. 2018;26(7):582-97. Epub 2018/02/06. doi: 10.1016/j.tim.2017.12.007. PubMed PMID: 29395728; PubMed Central PMCID: PMCPMC6003838.

36. Gladieux P, Condon B, Ravel S, Soanes D, Maciel JLN, Nhani A, Jr., et al. Gene Flow between Divergent Cereal- and Grass-Specific Lineages of the Rice Blast Fungus *Magnaporthe*

oryzae. MBio. 2018;9(1). Epub 2018/03/01. doi: 10.1128/mBio.01219-17. PubMed PMID: 29487238; PubMed Central PMCID: PMCPMC5829825.

37. Okagaki LH, Nunes CC, Sailsbery J, Clay B, Brown D, John T, et al. Genome Sequences of Three Phytopathogenic Species of the *Magnaporthaceae* Family of Fungi. G3 (Bethesda). 2015;5(12):2539-45. Epub 2015/09/30. doi: 10.1534/g3.115.020057. PubMed PMID: 26416668; PubMed Central PMCID: PMCPMC4683626.

38. Cruz CD, Valent B. Wheat blast disease: danger on the move. Tropical Plant Pathology. 2017;42(3):210-22. doi: 10.1007/s40858-017-0159-z.

39. Dean RA, Talbot NJ, Ebbole DJ, Farman ML, Mitchell TK, Orbach MJ, et al. The genome sequence of the rice blast fungus *Magnaporthe grisea*. Nature. 2005;434(7036):980-6. Epub 2005/04/23. doi: 10.1038/nature03449. PubMed PMID: 15846337.

40. Ebbole DJ. *Magnaporthe* as a model for understanding host-pathogen interactions. Annu Rev Phytopathol. 2007;45:437-56. Epub 2007/05/11. doi: 10.1146/annurev.phyto.45.062806.094346. PubMed PMID: 17489691.

41. McIntosh JR, O'Toole ET. Life cycles of yeast spindle pole bodies: getting microtubules into a closed nucleus. Biol Cell. 1999;91(4-5):305-12. PubMed PMID: 10518997.

42. Bao J, Chen M, Zhong Z, Tang W, Lin L, Zhang X, et al. PacBio Sequencing Reveals Transposable Elements as a Key Contributor to Genomic Plasticity and Virulence Variation in *Magnaporthe oryzae*. Mol Plant. 2017;10(11):1465-8. Epub 2017/08/26. doi: 10.1016/j.molp.2017.08.008. PubMed PMID: 28838703.

43. Jeon J, Choi J, Lee GW, Park SY, Huh A, Dean RA, et al. Genome-wide profiling of DNA methylation provides insights into epigenetic regulation of fungal development in a plant pathogenic fungus, *Magnaporthe oryzae*. Sci Rep. 2015;5:8567. Epub 2015/02/25. doi: 10.1038/srep08567. PubMed PMID: 25708804; PubMed Central PMCID: PMCPMC4338423.

44. Cao H, Huang P, Zhang L, Shi Y, Sun D, Yan Y, et al. Characterization of 47 Cys2 -His2 zinc finger proteins required for the development and pathogenicity of the rice blast fungus *Magnaporthe oryzae*. New Phytol. 2016;211(3):1035-51. Epub 2016/04/05. doi: 10.1111/nph.13948. PubMed PMID: 27041000.

45. Chao C-CT, Ellingboe AH. Selection for mating competence in *Magnaporthe grisea* pathogenic to rice. Canadian Journal of Botany. 1991;69(10):2130-4. doi: 10.1139/b91-267.

46. Leung H, Borromeo ES, Bernardo MA, Notteghem JL. Genetic analysis of virulence in the Rice Blast fungus *Magnaporthe grisea*. Phytopathology. 1988;78:1227-33.

47. Thon MR, Pan H, Diener S, Papalas J, Taro A, Mitchell TK, et al. The role of transposable element clusters in genome evolution and loss of synteny in the rice blast fungus *Magnaporthe oryzae*. Genome Biol. 2006;7(2):R16. Epub 2006/03/02. doi: 10.1186/gb-2006-7-2-r16. PubMed PMID: 16507177; PubMed Central PMCID: PMCPMC1431731.

48. Peng Z, Oliveira-Garcia E, Lin G, Hu Y, Dalby M, Migeon P, et al. Effector Gene Reshuffling Involves Dispensable Mini-chromosomes in the Wheat Blast Fungus. bioRxiv. 2018:359455. doi: 10.1101/359455.

49. Kobayashi N, Suzuki Y, Schoenfeld LW, Muller CA, Nieduszynski C, Wolfe KH, et al. Discovery of an unconventional centromere in budding yeast redefines evolution of point centromeres. Curr Biol. 2015;25(15):2026-33. doi: 10.1016/j.cub.2015.06.023. PubMed PMID: 26166782; PubMed Central PMCID: PMC4533239.

50. Rhind N, Chen Z, Yassour M, Thompson DA, Haas BJ, Habib N, et al. Comparative functional genomics of the fission yeasts. Science. 2011;332(6032):930-6. Epub 2011/04/23. doi: science.1203357 [pii]
10.1126/science.1203357. PubMed PMID: 21511999; PubMed Central PMCID: PMC3131103.

51. Shang WH, Hori T, Martins NM, Toyoda A, Misu S, Monma N, et al. Chromosome engineering allows the efficient isolation of vertebrate neocentromeres. Dev Cell. 2013;24(6):635-48. doi: 10.1016/j.devcel.2013.02.009. PubMed PMID: 23499358; PubMed Central PMCID: PMC3925796.

52. Thakur J, Sanyal K. Efficient neocentromere formation is suppressed by gene conversion to maintain centromere function at native physical chromosomal loci in *Candida albicans*. Genome Res. 2013;23(4):638-52. Epub 2013/02/27. doi: gr.141614.112 [pii]
10.1101/gr.141614.112. PubMed PMID: 23439889; PubMed Central PMCID: PMC3613581.

53. Simon L, Voisin M, Tatout C, Probst AV. Structure and function of centromeric and pericentromeric heterochromatin in *Arabidopsis thaliana*. Front Plant Sci. 2015;6:1049. doi: 10.3389/fpls.2015.01049. PubMed PMID: 26648952; PubMed Central PMCID: PMC4663263.

54. Fernandez-Alvarez A, Bez C, O'Toole ET, Morphew M, Cooper JP. Mitotic nuclear envelope breakdown and spindle nucleation are controlled by interphase contacts between centromeres and the nuclear envelope. Dev Cell. 2016;39(5):544-59. doi:

10.1016/j.devcel.2016.10.021. PubMed PMID: 27889481; PubMed Central PMCID: PMC5145006.

55. Hou H, Kallgren SP, Jia S. Csi1 illuminates the mechanism and function of Rabl configuration. Nucleus. 2013;4(3). Epub 2013/05/15. doi: 24876 [pii]. PubMed PMID: 23669133.

56. Jaspersen SL, Martin AE, Glazko G, Giddings TH, Jr., Morgan G, Mushegian A, et al. The Sad1-UNC-84 homology domain in Mps3 interacts with Mps2 to connect the spindle pole body with the nuclear envelope. J Cell Biol. 2006;174(5):665-75. doi: 10.1083/jcb.200601062. PubMed PMID: 16923827; PubMed Central PMCID: PMC2064310.

57. Richmond D, Rizkallah R, Liang F, Hurt MM, Wang Y. Slk19 clusters kinetochores and facilitates chromosome bipolar attachment. Mol Biol Cell. 2013;24(5):566-77. Epub 2013/01/04. doi: mbc.E12-07-0552 [pii]
10.1091/mbc.E12-07-0552. PubMed PMID: 23283988; PubMed Central PMCID: PMC3583661.

58. Baker RE, Rogers K. Genetic and genomic analysis of the AT-rich centromere DNA element II of *Saccharomyces cerevisiae*. Genetics. 2005;171(4):1463-75. Epub 2005/08/05. doi: 10.1534/genetics.105.046458. PubMed PMID: 16079225; PubMed Central PMCID: PMCPMC1350974.

59. Ohzeki J, Nakano M, Okada T, Masumoto H. CENP-B box is required for de novo centromere chromatin assembly on human alphoid DNA. J Cell Biol. 2002;159(5):765-75. Epub 2002/12/04. doi: 10.1083/jcb.200207112. PubMed PMID: 12460987; PubMed Central PMCID: PMCPMC2173396.

60. Xiao H, Mizuguchi G, Wisniewski J, Huang Y, Wei D, Wu C. Nonhistone Scm3 binds to AT-rich DNA to organize atypical centromeric nucleosome of budding yeast. Mol Cell. 2011;43(3):369-80. Epub 2011/08/06. doi: S1097-2765(11)00531-4 [pii]
10.1016/j.molcel.2011.07.009. PubMed PMID: 21816344.

61. Koren A, Tsai HJ, Tirosh I, Burrack LS, Barkai N, Berman J. Epigenetically-inherited centromere and neocentromere DNA replicates earliest in S-phase. PLoS Genet. 2010;6(8):e1001068. doi: 10.1371/journal.pgen.1001068. PubMed PMID: 20808889; PubMed Central PMCID: PMC2924309.

62. Kim SM, Dubey DD, Huberman JA. Early-replicating heterochromatin. Genes Dev. 2003;17(3):330-5. doi: 10.1101/gad.1046203. PubMed PMID: 12569122; PubMed Central PMCID: PMC195982.

63. Raghuraman MK, Winzeler EA, Collingwood D, Hunt S, Wodicka L, Conway A, et al. Replication dynamics of the yeast genome. Science. 2001;294(5540):115-21. doi: 10.1126/science.294.5540.115. PubMed PMID: 11588253.

64. Mitra S, Gomez-Raja J, Larriba G, Dubey DD, Sanyal K. Rad51-Rad52 mediated maintenance of centromeric chromatin in *Candida albicans*. PLoS Genet. 2014;10(4):e1004344. doi: 10.1371/journal.pgen.1004344. PubMed PMID: 24762765; PubMed Central PMCID: PMC3998917.

65. Sreekumar L, Jaitly P, Chen Y, Thimmappa BC, Sanyal A, Narlikar L, et al. DNA replication initiator proteins facilitate CENPA loading on early replicating compact chromatin. bioRxiv. 2018. doi: 10.1101/465880.

66. Cook DM, Bennett M, Friedman B, Lawrimore J, Yeh E, Bloom K. Fork pausing allows centromere DNA loop formation and kinetochore assembly. Proc Natl Acad Sci U S A. 2018;115(46):11784-9. Epub 2018/10/31. doi: 10.1073/pnas.1806791115. PubMed PMID: 30373818.

67. Anderson M, Haase J, Yeh E, Bloom K. Function and assembly of DNA looping, clustering, and microtubule attachment complexes within a eukaryotic kinetochore. Mol Biol Cell. 2009;20(19):4131-9. PubMed PMID: 19656849.

68. Gao D, Jiang N, Wing RA, Jiang J, Jackson SA. Transposons play an important role in the evolution and diversification of centromeres among closely related species. Front Plant Sci. 2015;6:216. Epub 2015/04/24. doi: 10.3389/fpls.2015.00216. PubMed PMID: 25904926; PubMed Central PMCID: PMC4387472.

69. Neumann P, Navratilova A, Koblizkova A, Kejnovsky E, Hribova E, Hobza R, et al. Plant centromeric retrotransposons: a structural and cytogenetic perspective. Mob DNA. 2011;2(1):4. Epub 2011/03/05. doi: 1759-8753-2-4 [pii] 10.1186/1759-8753-2-4. PubMed PMID: 21371312; PubMed Central PMCID: PMC3059260.

70. Wong LH, Choo KH. Evolutionary dynamics of transposable elements at the centromere. Trends Genet. 2004;20(12):611-6. Epub 2004/11/04. doi: S0168-9525(04)00283-5 [pii] 10.1016/j.tig.2004.09.011. PubMed PMID: 15522456.

71. Yang F, Naqvi NI. Sulfonylurea resistance reconstitution as a novel strategy for ILV2-specific integration in *Magnaporthe oryzae*. Fungal Genet Biol. 2014;68:71-6. Epub 2014/04/29. doi: 10.1016/j.fgb.2014.04.005. PubMed PMID: 24769367.

72. Rho HS, Kang S, Lee YH. Agrobacterium tumefaciens-mediated transformation of the plant pathogenic fungus, *Magnaporthe grisea*. Mol Cells. 2001;12(3):407-11. Epub 2002/01/24. PubMed PMID: 11804343.

73. Kim S, Mitchell TK. The application of ChIP-chip analysis in the rice blast pathogen. Methods Mol Biol. 2011;722:121-31. Epub 2011/05/19. doi: 10.1007/978-1-61779-040-9_8. PubMed PMID: 21590416.

74. Sullivan MJ, Petty NK, Beatson SA. Easyfig: a genome comparison visualizer. Bioinformatics. 2011;27(7):1009-10. Epub 2011/02/01. doi: 10.1093/bioinformatics/btr039. PubMed PMID: 21278367; PubMed Central PMCID: PMCPMC3065679.

75. Krumsiek J, Arnold R, Rattei T. Gepard: a rapid and sensitive tool for creating dotplots on genome scale. Bioinformatics. 2007;23(8):1026-8. Epub 2007/02/21. doi: 10.1093/bioinformatics/btm039. PubMed PMID: 17309896.

Figure legends

Fig 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.

Fig 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. t = minutes:seconds. 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. t = minutes:seconds. 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). t = minutes:seconds. 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.

Fig 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

Fig 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. **(C)** CenpC ChIP-qPCR analysis of cross-linked mycelia of the strain MGYF02 confirmed the centromere identity of each of the seven chromosomes of Guy11. Each bar represents the extent of enrichment obtained by one primer pair amplifying a unique sequence of each CenpA-bound region identified from the ChIP-seq analysis, and fold-enrichment values are normalized using a non-centromere region (ORF no. MGG_01917) as a control. Error bars represent the standard deviation calculated from three independent experiments. **(D)** ChIP-qPCR results showed that the gene MGG_01045 is not enriched with CenpC as observed in CenpA ChIP-seq analysis.

Fig 4. Identification of centromeres in *M. oryzae* strain 70-15 (MG8 assembly; Broad Institute). (A) Mapping of GFP-CenpA ChIP-seq reads to the reference MG8 genome assembly revealed the location of centromeres in the reference strain 70-15 of *M. oryzae*. Repeats, RNA-seq reads, and bisulfite sequencing reads were also mapped and are represented here for the comparative analyses. **(B)** 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.

Fig 5. Centromere DNA sequences in *M. oryzae* isolates are similar, but vary in repeat content. **(A)** Synteny analysis across centromeres and their flanking regions revealed the conservation of centromere flanking regions indicating that the centromere location is maintained in different isolates of *M. oryzae*. The gene annotations for the FJ81278 assembly are not available and hence are not represented in the maps. This analysis also revealed that centromere sequences are largely excluded from the current MG8 genome assembly as compared to that of Guy11 or FJ81278. A 200 kb region (with respect to the Guy11 genome assembly) for each centromere is represented in the maps. A few centromere flanking genes were also found missing from Chr7 in the MG8 assembly. BLAST analysis revealed the presence of these genes in the unassembled supercontigs8.8 of the genome assembly. **(B)** Dot-plot analysis of respective centromeres revealed that centromere sequences share considerable similarities in Guy11 and FJ81278. As shown in the graphs, the breaks observed in the dot-plot analysis overlapped with the presence/absence of repeat elements. The complete sequence of *CEN4* in FJ81278 was generated by fusing the two fragments, one each from contig 3 and 14. The individual fragments are shown using grey bars and are separated by a small thin black bar (equal to 100 bp). R denotes the repeat panels for both Guy11 and FJ81278.

Fig 6. Predicted centromeres in the wheat blast genome. (A) Chromosome maps showing the location of centromeres in the genome of the B71 isolate of wheat blast. The AT-rich, gene-free centromere regions are marked by a yellow star in each chromosome. **(B)** Synteny analysis for

centromeres and their flanking regions between the Guy11 and B71 genomes showed conservation of centromere locations therein. A 200-kb region is shown with respect to the Guy11 genome assembly. **(C)** Scaffolds representing the mini-chromosome in B71 were analyzed and are depicted with their gene and AT-richness graphs. A long AT-rich region (marked with dark green bar) in scaffold 1 (40,795 – 109,519) was identified and represents the putative centromere in the mini-chromosome in wheat blast.

Table 1. **Length and %GC content of centromeres in distinct *Magnaporthe* isolates.**

Isolate/ species	**Guy11**		**70-15***	**FJ81278****		**B71*****		***M. poae*******	
CEN#	**Coordinates**	**%GC**	**Coordinates**	**Coordinates**	**%GC**	**Coordinates**	**%GC**	**Coordinates**	**%GC**
CEN1	Contig15: 571735-678893 **(1,07,159)**	29.2	Chr1: 4669580-4690159 **(20,580)**	Contig6: 942880-1032698 **(89,819)**	26.8	Chr1: 5247711-5361150 **(113,440)**	29.4	GL876966: 1947276-2058771 **(111,496)**	25.5
CEN2	Contig7: 313767-411084 **(97,318)**	30.8	Chr2: 451419-471909 **(20,491)**	Contig1: 4838642-4937646 **(99,005)**	30.5	Chr2: 397914-483923 **(86,010)**	27.4	GL876967: 2010105-2095459 **(85,355)**	24.6
CEN3	Contig2: 3795849-3894639 **(98,791)**	33.0	Chr3: 5534011-5547746 **(13,736)**	Contig13: 469353-567903 **(98,551)**	34.6	Chr3: 6398799-6492490 **(93,692)**	25.8	GL876968: 291161-385721 **(94,561)**	23.8
CEN4	Contig10: 1000090-1063263 **(63,174)**	30.5	Chr4: 845436-854002 **(8,567)**	Contig14: 1-38471 + Contig3: 1-30023 **(68,694)**	30.0	Chr4: 815420-895650 **(80,231)**	24.8	GL876971: 2575155-2652612 **(77,458)**	24.6
CEN5	Contig4: 4014470-4071774 **(57,305)**	28.0	Chr5: 296487-302043 **(5,557)**	Contig4: 259241-318616 **(59,376)**	28.8	Chr5: 296404-367652 **(71,249)**	24.8	GL876972: 961971-1029283 **(67,313)**	23.3
CEN6	Contig5: 343714-452391 **(1,08,678)**	32.3	Chr6: 3852307-3861083 **(8,777)**	Contig12: 708384-794489 **(86,106)**	29.3	Chr6: 5687241-5758152 **(70,912)**	24.8	GL876975: 266207-359379 **(93,173)**	24.2
CEN7	Contig13: 573351-645424 **(72,074)**	30.5	Chr7: 2771164-2774628 **(3,465)**	-		Chr7: 3255069-3336736 **(81,668)**	26.2	GL876978: 1-33204 **(33,204)**	37.1
						Scaffold 1: 40795-109519 **(68,724)**		GL876979: 1-17025 **(17,025)**	24.6

Centromere numbers are according to the 70-15 genome assembly.

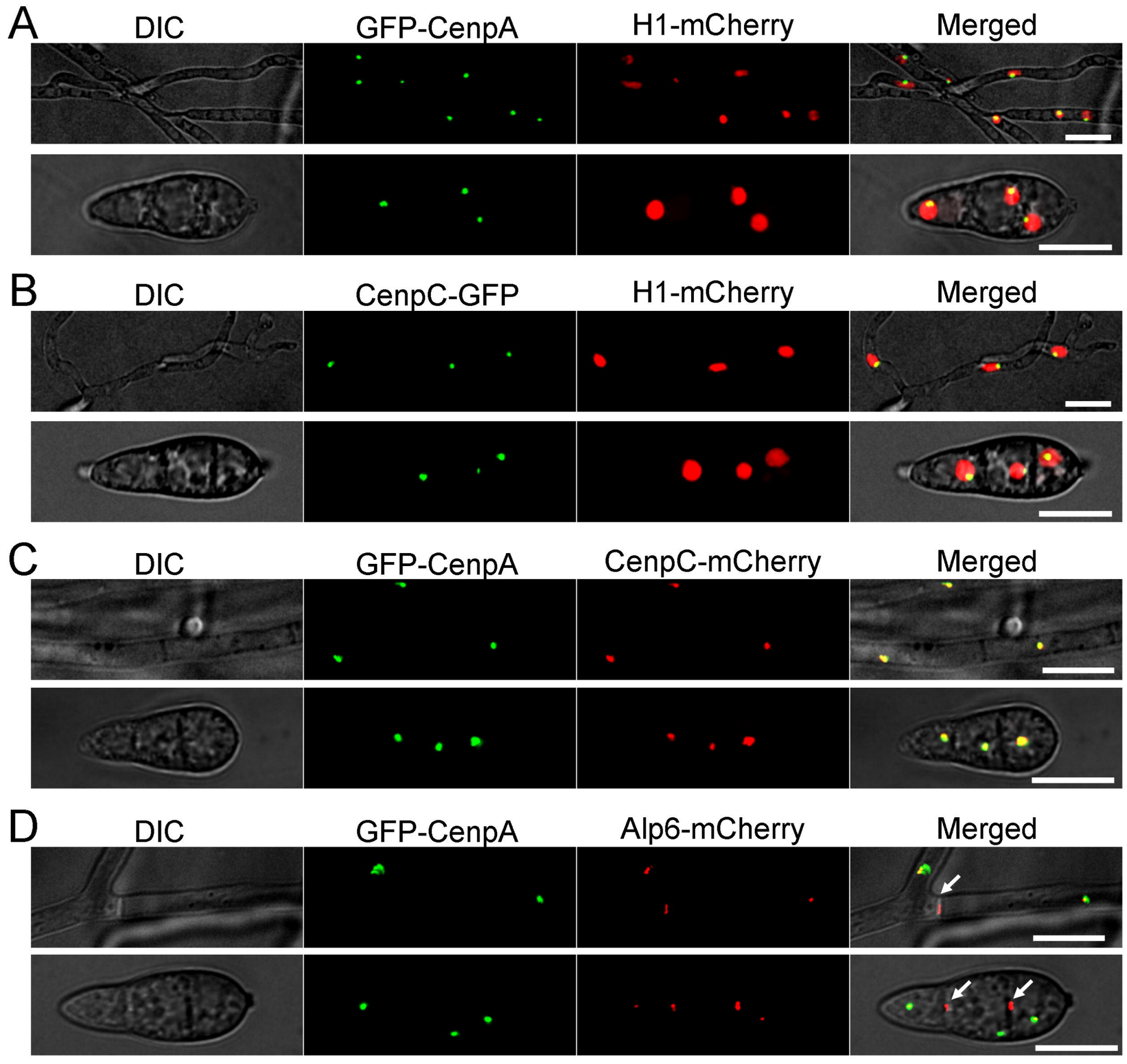
Numbers in the bracket represent centromere length in base pair.

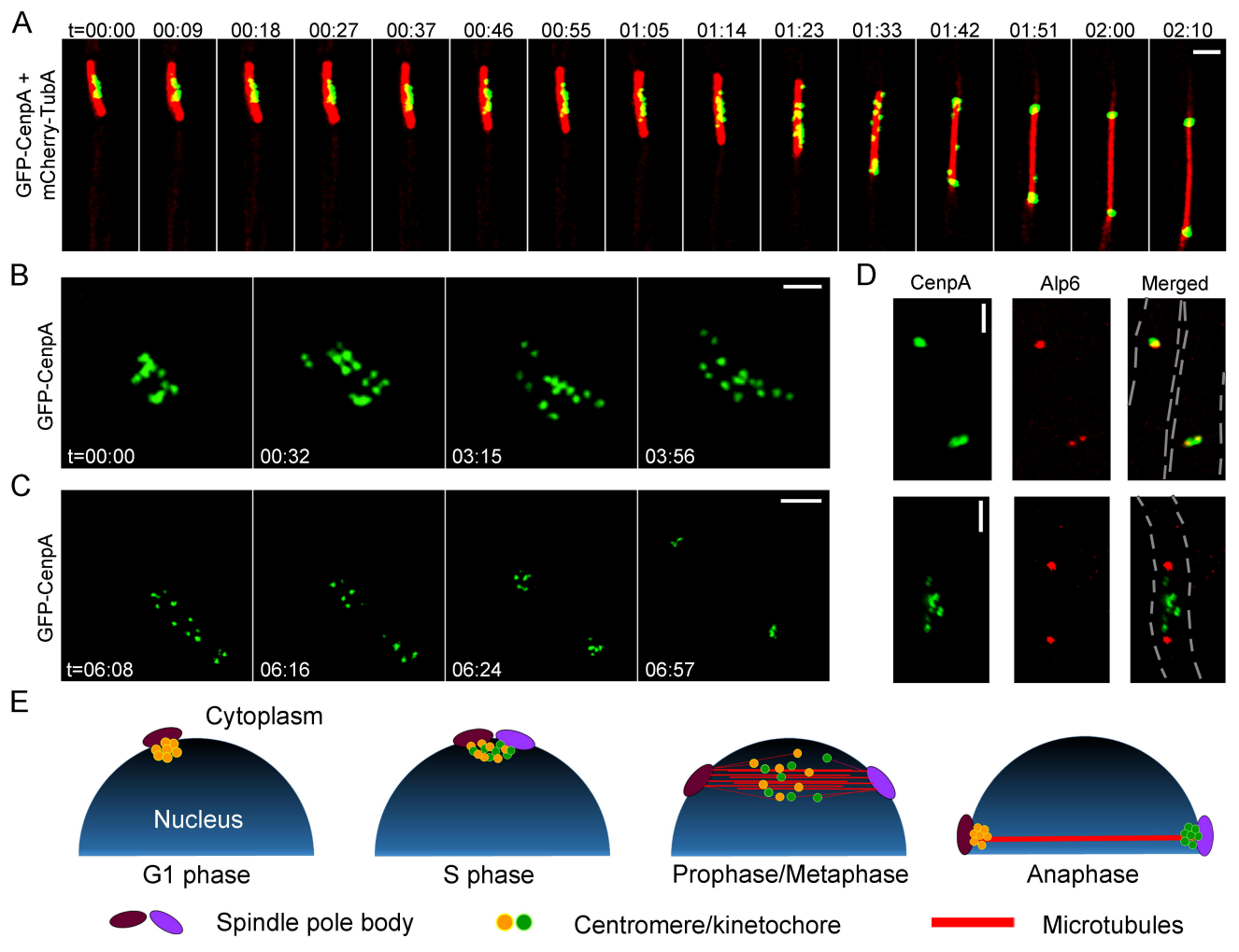
*Centromeres in 70-15 isolate contain breaks and hence %GC is not calculated for its centromeres.

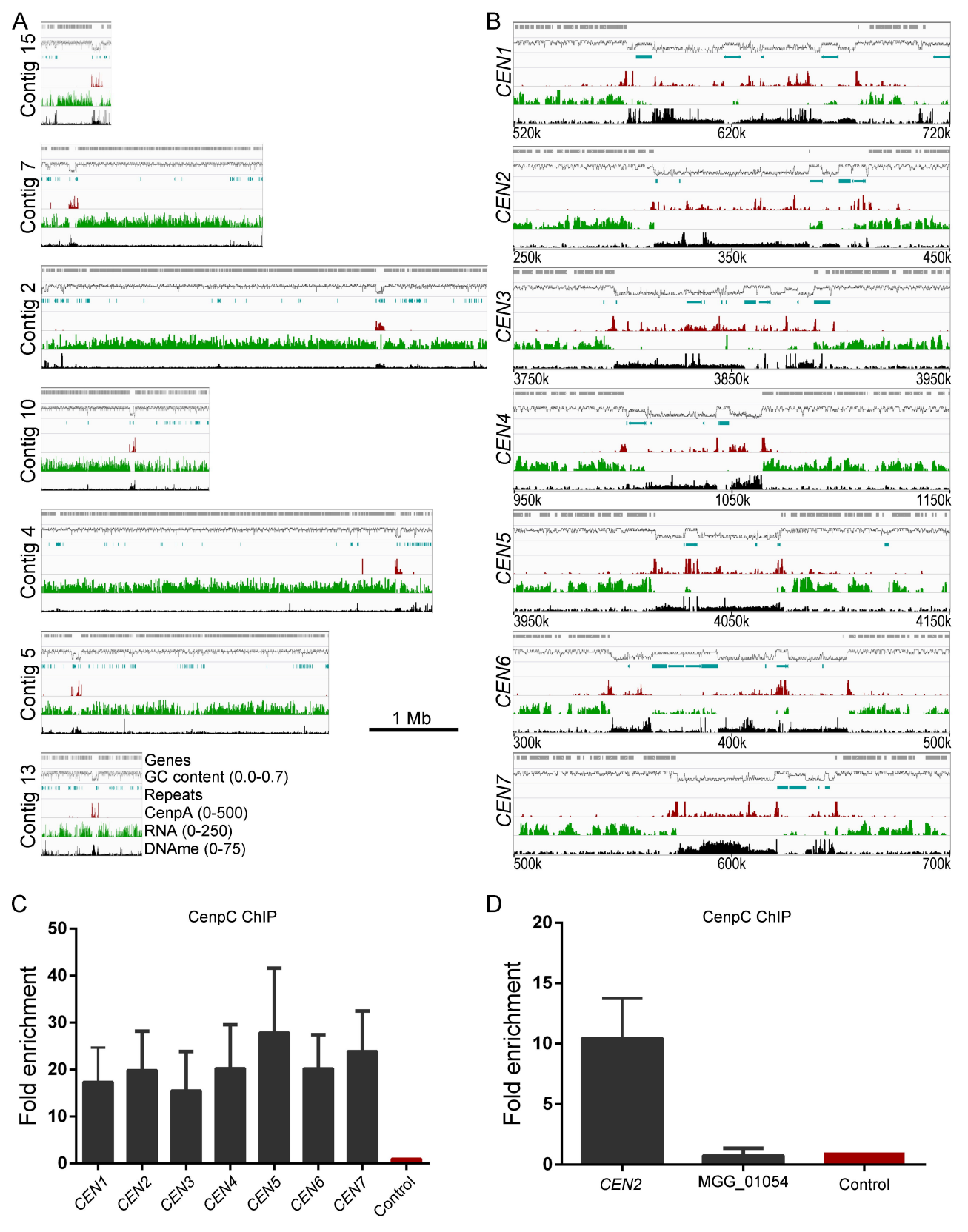
***CEN7* in FJ81278 could not be identified due to poor genome assembly.

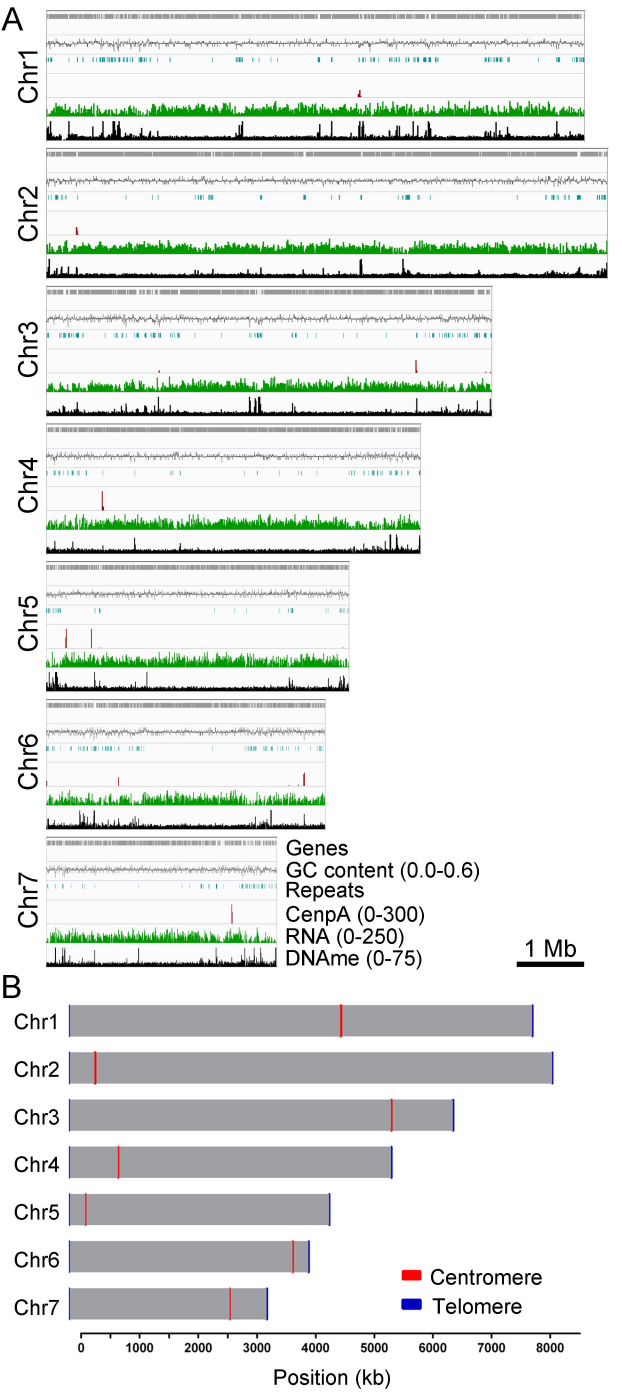
***Predicted centromere for mini-chromosome of B71 is also listed.

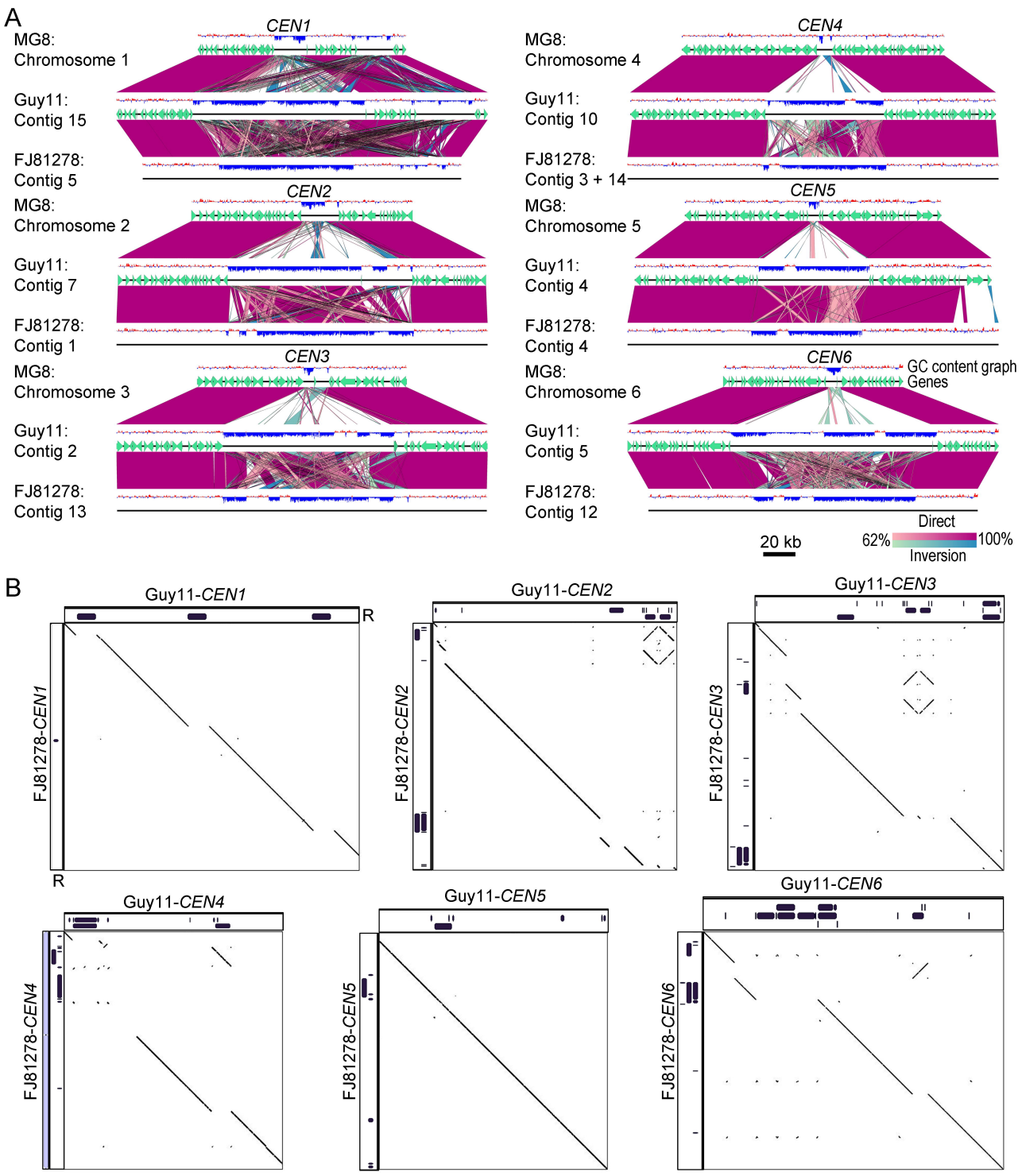
****Two of the putative centromeres in *M. poae* are present at the end of two separate contigs. Both are listed as different centromeres here.

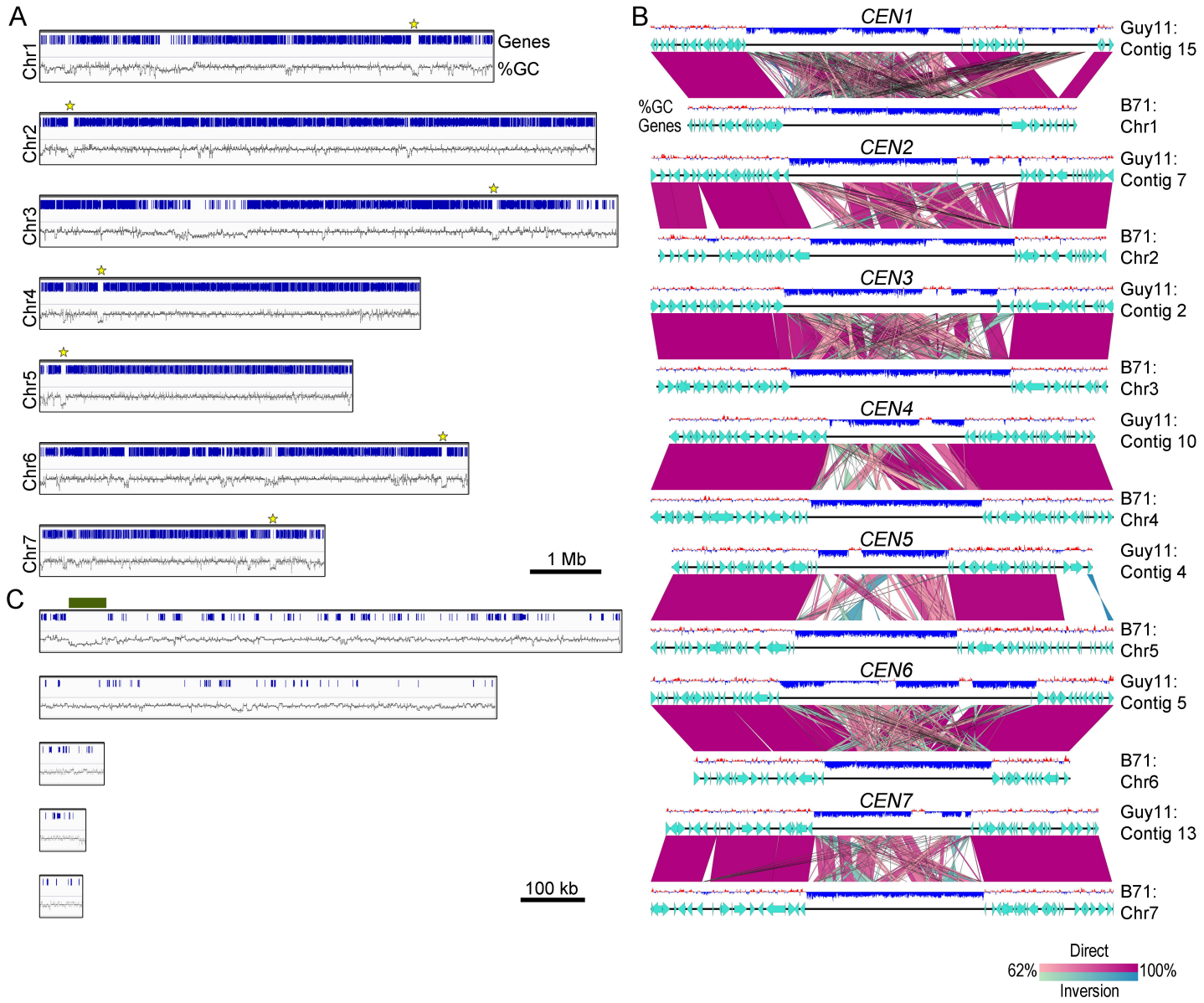

A
DIC
GFP-CenpA
H1-mCherry
Merged
B
DIC
CenpC-GFP
H1-mCherry
Merged
C
DIC
GFP-CenpA
CenpC-mCherry
Merged
D
DIC
GFP-CenpA
Alp6-mCherry
Merged


A
t=00:00
00:09
00:18
00:27
00:37
00:46
00:55
01:05
01:14
01:23
01:33
01:42
01:51
02:00
02:10
GFP-CenpA +
mCherry-TubA
B
GFP-CenpA
t=00:00
00:32
03:15
03:56
C
GFP-CenpA
t=06:08
06:16
06:24
06:57
D
CenpA
Alp6
Merged
E
Cytoplasm
Nucleus
G1 phase
S phase
Prophase/Metaphase
Anaphase
Spindle pole body
Centromere/kinetochore
Microtubules


A
Contig 15
Contig 7
Contig 2
Contig 10
Contig 4
Contig 5
Contig 13
1 Mb
Genes
GC content (0.0-0.7)
Repeats
CenpA (0-500)
RNA (0-250)
DNAme (0-75)
B
CEN1
520k
620k
720k
CEN2
250k
350k
450k
CEN3
3750k
3850k
3950k
CEN4
950k
1050k
1150k
CEN5
3950k
4050k
4150k
CEN6
300k
400k
500k
CEN7
500k
600k
700k
C
CenpC ChIP
Fold enrichment
0
10
20
30
40
50
CEN1
CEN2
CEN3
CEN4
CEN5
CEN6
CEN7
Control
D
CenpC ChIP
Fold enrichment
0
5
10
15
20
CEN2
MGG_01054
Control


A
Chr1
Chr2
Chr3
Chr4
Chr5
Chr6
Chr7
Genes
GC content (0.0-0.6)
Repeats
CenpA (0-300)
RNA (0-250)
DNAme (0-75)
1 Mb
B
Chr1
Chr2
Chr3
Chr4
Chr5
Chr6
Chr7
Centromere
Telomere
0
1000
2000
3000
4000
5000
6000
7000
8000
Position (kb)


A
CEN1
MG8:
Chromosome 1
Guy11:
Contig 15
FJ81278:
Contig 5
CEN2
MG8:
Chromosome 2
Guy11:
Contig 7
FJ81278:
Contig 1
CEN3
MG8:
Chromosome 3
Guy11:
Contig 2
FJ81278:
Contig 13
CEN4
MG8:
Chromosome 4
Guy11:
Contig 10
FJ81278:
Contig 3 + 14
CEN5
MG8:
Chromosome 5
Guy11:
Contig 4
FJ81278:
Contig 4
CEN6
MG8:
Chromosome 6
Guy11:
Contig 5
FJ81278:
Contig 12
GC content graph
Genes
20 kb
Direct
62%
100%
Inversion
B
Guy11-CEN1
FJ81278-CEN1
R
R
Guy11-CEN2
FJ81278-CEN2
Guy11-CEN3
FJ81278-CEN3
Guy11-CEN4
FJ81278-CEN4
Guy11-CEN5
FJ81278-CEN5
Guy11-CEN6
FJ81278-CEN6


A
Chr1
Genes
%GC
Chr2
Chr3
Chr4
Chr5
Chr6
Chr7
1 Mb
B
CEN1
Guy11:
Contig 15
%GC
Genes
B71:
Chr1
CEN2
Guy11:
Contig 7
B71:
Chr2
CEN3
Guy11:
Contig 2
B71:
Chr3
CEN4
Guy11:
Contig 10
B71:
Chr4
CEN5
Guy11:
Contig 4
B71:
Chr5
CEN6
Guy11:
Contig 5
B71:
Chr6
CEN7
Guy11:
Contig 13
B71:
Chr7
Direct
62%
100%
Inversion
C
100 kb