Crosstalk of DNA replication and chromosome segregation machinery ensures genome stability in *Candida albicans*

A thesis submitted for the degree of

Doctor of Philosophy

by

Lakshmi Sreekumar

Molecular Biology and Genetics Unit

Jawaharlal Nehru Centre for Advanced Scientific Research

Jakkur, Bangalore-560064

November 2018

Dedicated to my parents

DECLARATION

I do hereby declare that the work described here in this thesis titled '**Crosstalk of DNA replication and chromosome segregation machinery ensures genome stability in *Candida albicans'*** has originally been carried out by myself under the guidance and supervision of **Kaustuv Sanyal,** Professor, Molecular Biology and Genetics Unit, Jawaharlal Nehru Centre for Advanced Scientific Research, Bangalore-560064, India. In keeping with the norm of reporting the scientific observations, due acknowledgements have been made whenever the work described was carried out in collaboration with other researchers. Any omission, which might have occurred by oversight or misjudgement, is regretted.

Lakshmi Sreekumar

Place: Bangalore

Date:

Jawaharlal Nehru Center for Advanced Scientific Research

Kaustuv Sanyal PhD, FNA, FASc, FNASc
Professor & Tata Innovation Fellow

27 November 2018

Certificate

This is to certify that the work described in the thesis titled '**Crosstalk of DNA replication and chromosome segregation machinery ensures genome stability in *Candida albicans'*** is the result of investigations carried out by **Ms. Lakshmi Sreekumar** in the Molecular Biology and Genetics Unit, Jawaharlal Nehru Centre for Advanced Scientific Research, Bangalore, India, under my supervision and guidance. The results presented here have not previously formed the complete basis for the award of any other diploma or degree.

Kaustuv Sanyal

Molecular Mycology Laboratory, Molecular Biology & Genetics Unit, JNCASR, Jakkur Post, Bangalore 560064, India
Tel: +91 80 2208 2878; email: sanyal@jncasr.ac.in; web: http://www.jncasr.ac.in/sanyal

ACKNOWLEDGMENTS

TABLE OF CONTENTS

Results

LIST OF TABLES AND FIGURES

Introduction

Results

Discussion

Materials and methods

ABBREVIATIONS

ºC	Degree Celsius
µg	Microgram
µl	Microliter
µm	Micrometer
µM	Micromolar
ARS	Autonomously Replicating Sequence
ACS	ARS consensus sequence
BLAST	Basic Local Alignment Tool
bp	Base pair
CEN	centromere
CENPA	Centromere protein A
ChIP	Chromatin immunoprecipitation
ChIP-seq	ChIP-sequencing
Chr	Chromosome
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
EGTA	Ethyleneglycoltetraaceticacid
GFP	Green fluorescent protein
h	Hour
Hi-C	Chromosome conformation capture
kb	Kilobase
kDa	kilodalton
M	Molar
Mb	Megabase
MCM	Minichromosome maintenance complex
mg	Milligram
min	Minute
ml	Millilitre
mM	Millimolar
MNase	Micrococcal nuclease

NAT	Nourseothricin acetyl transferase
ORC	Origin recognition complex
ORF	Open reading frame
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
RNA	Ribonucleic acid
RNAi	RNA interference
rpm	Revolutions per minute
RT	Room temperature
s	Second
SDS	Sodium dodecyl sulphate
SEM	Standard error of mean
SIDD	Stress induced duplex destabilization
qPCR	Quantitative PCR

1. INTRODUCTION

Eukaryotic cell cycle

The eukaryotic cell cycle comprises of well-orchestrated cellular events that mediate accurate DNA duplication and high-fidelity chromosome segregation to yield two identical daughter cells during mitosis. Cell cycle completion occurs in distinct phases- the preparatory phase or the interphase and cell division or the mitotic phase. Interphase comprises of the two gap phases G_1 and G_2, each of which are intercepted by DNA replication occurring in S phase and the mitotic M phase, respectively, ensuring equal distribution of nuclear and cytoplasmic contents to yield progeny. The two concomitant DNA metabolic processes, namely DNA replication and chromosome segregation ensure genome stability in cellular systems. A majority of the molecular players and pathways operating in these processes are conserved from unicellular eukaryotes like yeast to the more complex developmentally regulated multicellular organisms including humans, even though species specific differences render properties to study genome biology of various organisms.

Chromosome and its inheritance

The genetic material is tightly packed in the nucleus of every eukaryotic cell as a complex of DNA and proteins bound to it. In the interphase cells, chromosomes appear as loose stands of DNA-protein complexes, which are later visualized as rod-shaped bodies during mitosis. There are multiple loci on a chromosome where replication of DNA initiates, called origins of replication. These DNA elements help to facilitate chromosome duplication by a semi-conservative manner of DNA replication. Since eukaryotic DNA replication is discontinuous, it creates the end-replication problem. This is solved by telomeres and telomere associated proteins. Post replication, chromosomes have to be equally partitioned to the daughter cells with the help of the centromere, a partitioning locus present on every eukaryotic chromosome. Centromeres facilitate accurate segregation of chromosomes with the help of spindle microtubules that pull chromosomes to opposite poles of a dividing cell. Hence, these distinct loci on every chromosome help in stable propagation of the genetic material. Although these processes occur in specific phases of the cell cycle, there is significant crosstalk between molecular players involved, that is pertinent to drive a stable genome. The alternate cycling of growth and division of cells is evolutionarily conserved from yeast to humans and a plethora of model organisms have been used to dissect molecular and genetic pathways concerning the same.
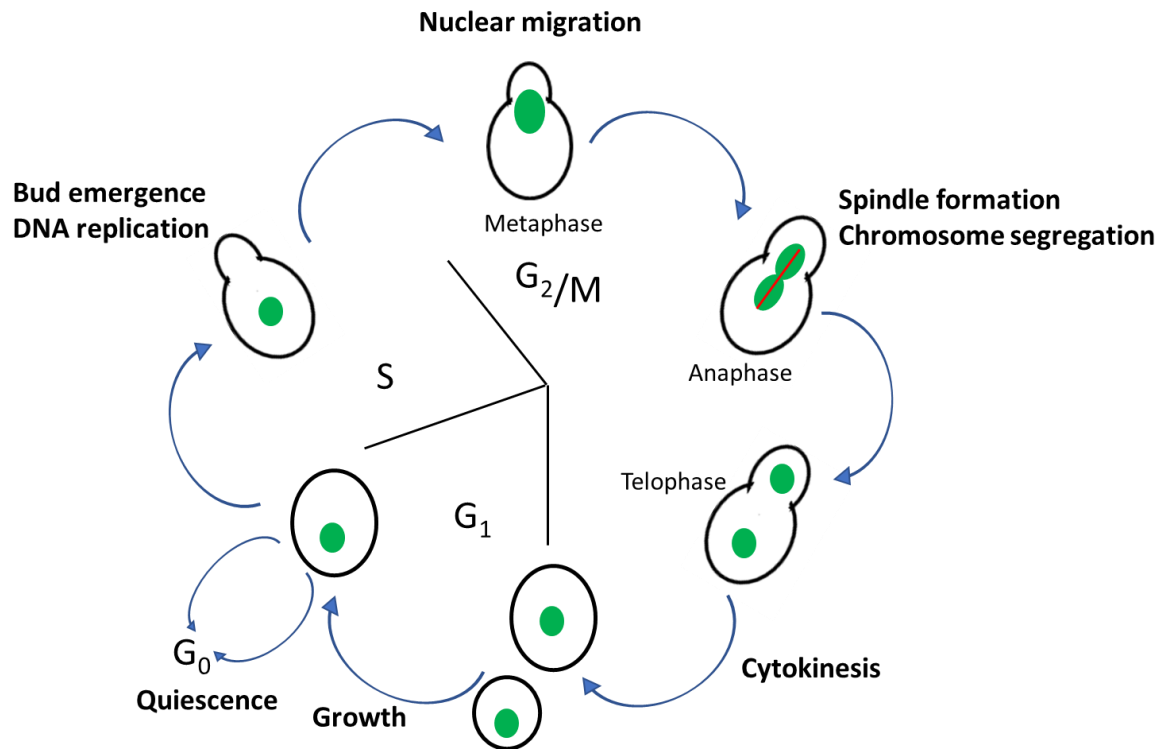
Figure 1.1 Different stages of the eukaryotic cell cycle. A model of budding yeast cell cycle represents various stages of an unbudded G_1 cell that undergoes budding accompanied by cell growth and duplication of chromosomes (green) at S phase. The budded cell post DNA replication eventually segregates its duplicated genetic content in mitosis with the help of spindle microtubules (red) to give rise to two identical daughter cells during cell division at cytokinesis. If the cell is depleted of nutrients, it can go into a resting of quiescent stage called G_o.

Regulation of cell cycle

Progression of cell cycle form one stage to another is biochemically regulated by activation and inactivation of cyclin dependent kinases (CDKs) and their association with an oscillatory protein called cyclin (Culotti and Hartwell 1971, Hartwell 1971, Hartwell 1971). A single CDK can interact with multiple cyclins, thereby rendering specificity to the CDK action. Once activated, CDKs phosphorylate downstream effector molecules that determine the individual stage of the cell cycle. This oscillatory, yet orderly pattern of cyclin-CDK action makes sure that the cell progresses forward through individual phases of the cell cycle well in time before cell division occurs. Apart from this regulation, there are surveillance mechanisms to make sure that DNA is not damaged during the process of replication and the replicated chromosomes are aligned and attached properly by the mitotic spindle apparatus.

These checkpoint mechanisms get activated as a cascade of biochemical reactions with downstream effectors. Control of cell size is monitored by G_1 checkpoint that regulates cell size (Figure 1.1). Transition from G_2 to M is monitored by kinases (Cdc25 and Wee1) which respond to cell size and nutritional status (Nurse, Thuriaux et al. 1976). Throughout interphase, the DNA undergoes several physical and chemical changes which make it prone to damage. Additionally, DNA can be subject to exogenous insults like chemical mutagens and ionizing radiations causing genomic instability. To circumvent this, the DNA damage response pathways at G_1/S boundary, S phase and G_2 phase operate to maintain CDKs in an inactive state till the damage/ lesion is repaired. These pathways are fairly conserved from yeast to humans, although many of the seminal studies have been done in fungal systems.

Yeast as a model to study cell cycle

Yeast is one of the simplest unicellular eukaryotes having many of the cellular components and pathways conserved in the developmentally complex humans. Moreover, its genetic amenability and the availability of a huge palette of genetic tools makes it an ideal system to study cell cycle related events. The bakers' yeast, *Saccharomyces cerevisiae* is the long-term popular model which was also one of the first eukaryotes to have its genome fully sequenced. The fission yeast, *Schizosaccharomyces pombe* is also a widely used model to study cell division and has been the subject of discovery of many "cell division cycle" (*cdc*) genes. With the advent of molecular biology and genomics, genomes of several pathogenic and non-pathogenic yeast species have been sequenced and homologs of essential cell cycle proteins have been characterized. One such yeast is the hemiascomycetous pathogenic budding yeast, *Candida albicans* which is the organism chosen for this study.

DNA replication

DNA is a self-replicating molecule. The faithful duplication of the genetic material is achieved at the synthetic or S-phase of the cell cycle. Replication origins are multiple distinct sites on every chromosome where DNA replication initiates and the replication forks proceed bidirectionally. These initiator sequences act as a platform for multi-protein sub-complexes to assemble and facilitate the opening up of the double-stranded DNA to form "replication bubbles". Following melting of the bubble, DNA polymerase synthesises nascent DNA using the parental template in a semi-conservative manner. The end of replication is marked by two

genetically identical DNA molecules which are ready to be segregated into the daughter cells. DNA replication comprises of three distinct steps: initiation, elongation and termination.

Initiation of DNA replication

Eukaryotic DNA replication is initiated by a series of biochemical events beginning with the binding of the origin recognition complex (ORC), which flag mark potential sites of origin assembly (Bell and Dutta 2002). In G_1, ORC, Cdc6, and Cdt1 together load or "license" the inactive minichromosome maintenance (MCM) helicase complexes to form the pre-replication complex, the pre-RC (Figure 1.2). The helicase MCM helps in template DNA unwinding and is activated at the G_1/ S transition with the help of phosphorylation events mediated by Dbf4-dependent kinase (DDK) and cyclin-dependent kinase (CDK). DDK acts first, phosphorylating MCM subunits to load Cdc45 and Sld3. In this pre-initiation complex, pre-IC, CDK phosphorylates Sld3 and Sld2, leading to the recruitment of GINS (which is the active functional helicase) and additional factors, including DNA polymerases, to complete the replisome assembly and initiate DNA synthesis. Although the requisite sequence of events is shared by all active origins, they differ in timing and efficiency of initiation during S phase (Aparicio 2013).

Autonomously Replication Sequences (ARS)

In yeast, chunks of DNA belonging to origins when cloned in episomes yield high transformation efficiency implying that they can replicate autonomously (Brewer and Fangman 1987). These autonomously replicating sequences (ARSs), first discovered in budding yeast, seemed to contain two features, a higher AT content and the presence of an 11 bp sequence, (A/T)TTTAT(A/G)TTT(A/T), the ARS Consensus Sequence (ACS) (Newlon and Theis 1993). ACS confers replicative functions to the origins in the sensu stricto *Saccharomyces* group (Nieduszynski, Knox et al. 2006). ARSs have a functional A element, which is the binding site of ORC and the ancillary B element that facilitates binding of transcription factors to create a more relaxed chromatin template for replication (Newlon and Theis 1993). The pathogenic yeast, *Candida glabrata* uses a 17 bp AT-rich region as an ACS (Descorps-Declere, Saguez et al. 2015), whereas the methylotropic yeast, *Pichia pastoris* utilizes a combination of AT-rich and GC-rich origins to fulfil its genome duplication (Liachko, Youngblood et al. 2014). With the exception of such budding yeast species studied so far, origins have been more difficult to identify in other eukaryotes since they do not strictly/solely rely on the underlying DNA sequence for ORC binding. Fission yeast origins
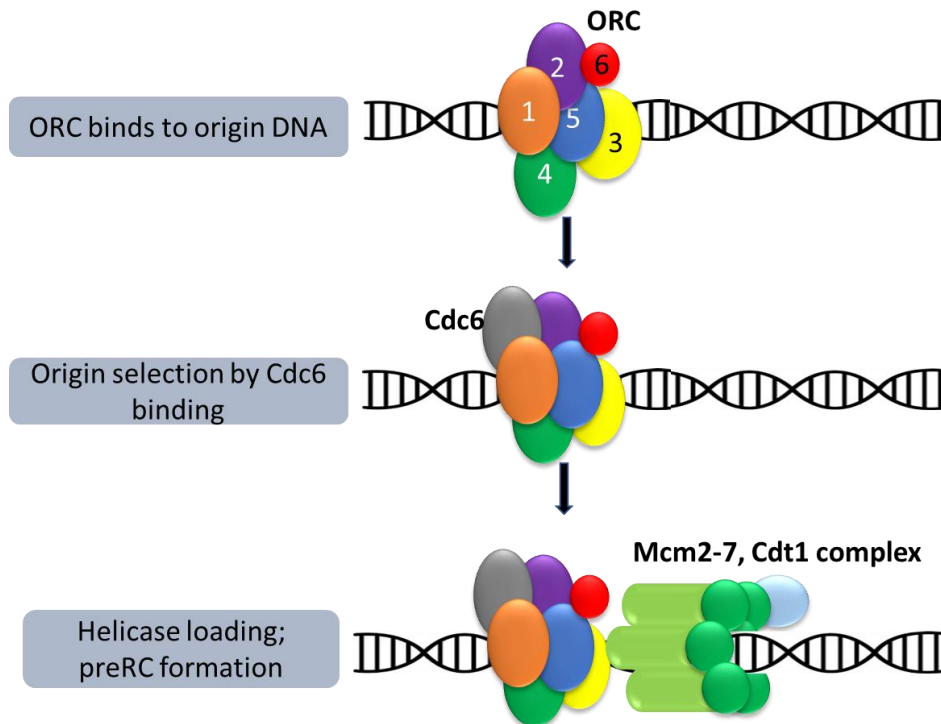
Figure 1.2 Assembly of the pre-replication complex. The hexameric ORC binds to the replication origin DNA through species specific mechanisms. The accessory protein, Cdc6 binds to the ORC-DNA complex and facilitates MCM loading. Mcm2-7 along with Cdt1 form a ring around the origin DNA that potentiates it for firing by forming the pre-replicative complex.

share features with metazoans with complex genomes. Metazoan origins are GC-rich as seen in case of *Drosophila* and mouse embryonic cells (Cayrou, Coulombe et al. 2011). The lack of DNA consensus sequences, and presence of inefficient origins firing only in a subset of cell cycles in fungal species like *S. pombe* (Patel, Arcangioli et al. 2006) strongly suggests the loose requirement of DNA sequence in imparting replicative properties to an origin.

Origin recognition complex (ORC)

All eukaryotic origins of DNA replication assemble a conserved protein complex assembly, the origin recognition complex (ORC), to recognize diverse DNA sequences. The hexameric ORC complex consists of the Orc1-6 polypeptides, which together act as a scaffold for binding and assembly of other key initiation factors. Upon recruitment to start sites, ORC binds Cdc6, a necessary prerequisite for helicase loading. (Leonard and Mechali 2013). First identified as an ARS binding protein in budding yeast, ORCs have been found to demonstrate significant conservation across eukaryotes (Bell and Stillman 1992). Structurally, the core components and subunit organisation show broad conservation even

though species specific differences exist that generate functional variations. Orc1-5 exhibit significant conservation across species, whereas Orc6 is the least conserved one. Out of the six proteins, Orc1, Orc4 and Orc5 contain the ATPase domain of the AAA+ family, which bind to ATP to cause a conformational change that results in binding of the interacting partner, Cdc6 (Lee and Bell 1997). Orc4 has retained the AAA+ signature Walker A and Walker B sequence motifs in all organisms with the exception of *S. cerevisiae*. While structural elements contributing to ATP binding is conserved in these subunits, only Orc1 possesses ATPase activity. The fission yeast Orc4 is known to bind to the minor groove of AT-rich sequences utilizing its AT-hook motifs, thereby facilitating origin usage (Chuang and Kelly 1999, Dai, Chuang et al. 2005). The binding of Cdc6 to ORC is ATP driven, and this complex is the functional initiator of origins. Structural studies of the *Drosophila* ORC revealed a subunit order of Orc1→ Orc4→ Orc5→ Orc3→ Orc2 around the ORC ring. Stable subcomplexes of ORC are seen to form across species. Unlike *D. melanogaster* and *S. cerevisiae* ORC, which form stable heterohexamers, human Orc1 and Orc6 loosely associate with an Orc2-5 core (Parker, Botchan et al. 2017). A recent cryoelectron microscopy structure of the budding yeast ORC revealed a conserved role of ORC in modulating DNA structure to facilitate origin selection and usage (Li, Lam et al. 2018)

The mode of ORC binding varies in different fungal and metazoan species studied so far (Figure 1.3). *S. cerevisiae* ORC is recruited via the ACS element (Marahrens and Stillman 1992). *S. pombe* relies on AT-rich asymmetric sequences that facilitate Orc4 binding using their AT-hook motifs (Chuang and Kelly 1999). The AT-hook motif is absent in *S. cerevisiae* and metazoan Orc4. Metazoan Orc6 is used to loosely tether the entire complex to DNA in a manner analogous to *S. pombe* Orc4. The initial recruitment is followed by subsequent stabilization and stable origin association. In metazoan origins, transcriptional regulators like heterochromatic protein 1 or HP1 help to tether ORC to chromosomes and likely affect its function. Another factor, ORCA (ORC Associated) has been shown to directly recruit ORC to chromosomal origins in humans and promote replication licensing (Prasanth, Shen et al. 2010).

Minichromosome Maintenance Complex (MCM)

Originally identified in genetic screens aimed to discover factors required for minichromosome maintenance in yeast (Maine, Sinha et al. 1984), MCMs are known to have poignant roles in ARS replication and maintenance, as *mcm* mutants display increased rate of
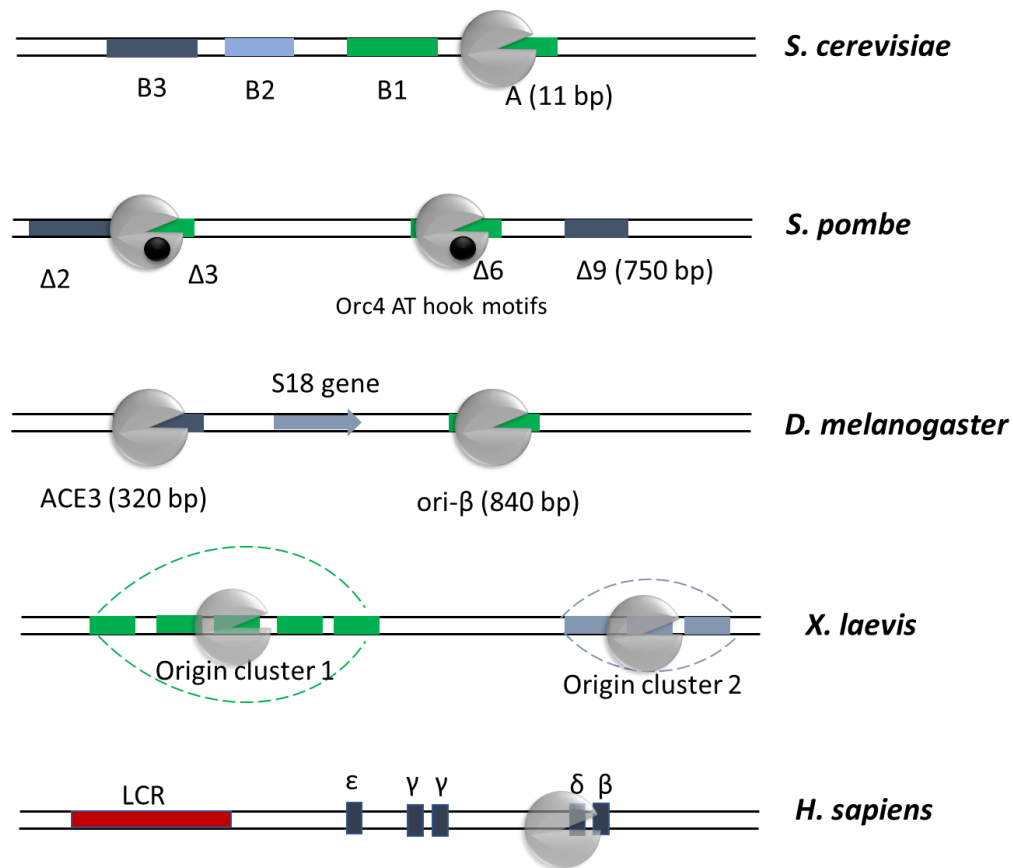
Figure 1.3 Varying modes of ORC binding to the origin DNA across species. The budding yeast *S. cerevisiae* has defined sequence signature i.e., the A element and the B element. The A element consists of an 11 bp ACS that binds ORC (grey sphere) in a sequence specific manner. The fission yeast *S. pombe* ORCs, on the other hand, bind to origins with the help of AT hook motifs (black circle) present in their Orc4 subunit. *D. melanogaster* ORCs show more promiscuous binding, but have few well defined origins placed in between gene bodies. *X. laevis* origins exist as replicon clusters that share similar activation timing, some strongly firing than the other. The replication program in *X. laevis* is stochastic and developmentally regulated. *H. sapiens* on other hand show more diverse origin features with larger gaps within developmentally regulated origin clusters located within transcriptionally start sites (TSSs).

mitotic recombination and chromosome loss (Sinha, Chang et al. 1986, Yan, Gibson et al. 1991). Mcm2-7 constitutes the helicase assembly, which have a conserved AAA+ ATPase domain and from heterohexamers. Unlike ORCs, all six subunits of MCM contain the complete set of catalytic residues that support ATP hydrolysis to enable opening of the duplex DNA. All MCMs contain a conserved MCM box that has the Walker A and the Walker B motifs (Forsburg 2004). MCM orthologues among species are more conserved than MCM paralogue in the same species. For example, Mcm2 from humans is more similar to Mcm2 in yeast (an orthologue) than to Mcm4 of humans (a paralogue). The DNA binding N-

terminal domain of the MCM are important for both origin DNA melting and unwinding. MCMs are nevertheless key players in DNA replication initiation and help in origin licensing. The active form of the helicase is generated by the CDK-DDK-mediated phosphorylation of the MCM subunits (Forsburg 2004).

In a majority of the species, MCMs are nuclear localised and their association with chromatin is cell cycle regulated. Mutational analysis of *S. pombe MCM2* revealed that the assembly of the MCM complex occurs in the cytoplasm and this is necessary for the nuclear import and retention (Forsburg, Sherman et al. 1997). Contrary to *S. pombe,* in *S. cerevisiae,* MCMs are mostly localised to the nucleus in the S phase (Forsburg, Sherman et al. 1997, Forsburg 2004). *In vitro* experiments suggest that Mcm4,6,7 form a trimer by themselves to form the MCM core complex and this complex is important for its helicase activity. This interaction is disrupted by addition of Mcm2. However, *in vivo* evidence suggests that bulk of the MCMs are assembled as hexamers (Forsburg 2004). These contrasting data could indicate that there are functionally distinct pools of MCMs in the cell which can operate in a complex-independent manner. Intriguingly, only a subset of MCMs are associated with origins, whereas the rest of the pool is liberally distributed throughout the nucleus.

Cardinal features of replication origins

The number and uniform distribution of origins across the length of a chromosome is an important criterion to fulfil the timely duplication of the genetic material of an organism (Newman, Mamun et al. 2013, Prioleau and MacAlpine 2016). In *S. cerevisiae*, approximately 400 ORC binding sites have been identified, but only a subset of them 'fire' at a given time (Wyrick, Aparicio et al. 2001, Nieduszynski, Knox et al. 2006). This implies that not all pre-RC binding sites act as functional DNA replication origins at every S phase of the cell cycle. In most eukaryotes, replication origins are defined more flexibly as they rely very little on a DNA sequence requirement for origin specification (Parker, Botchan et al. 2017). For a region to be defined as an origin of DNA replication, it has to be biochemically capacitated by the availability of initiation factors and reside in a chromatin environment that is favourable for initiating replication. Apart from the composition of the underlying DNA sequence, chromatin properties of regions surrounding replication origins augment its function. Nucleosome-free regions (NFRs) are permissible to replication and the replisome organisation imposed by nucleosome positioning is phylogenetically widespread in multiple

eukaryotic systems (Radman-Livaja and Rando 2010, Li, Zhong et al. 2014). NFRs provide easier access of ORCs to DNA and phased nucleosomes provide additionally favourable regions for ORC binding (Muller, Park et al. 2010, Hizume, Yagura et al. 2013) (Figure 1.3A). Stress induced duplex destabilization (SIDD) regions have been known to be associated with chromosomal attachment regions. Also, sites susceptible to duplex DNA destabilization occur at binding sites of transcription factors and regulatory proteins (Bi and Benham 2004, Ak and Benham 2005). SIDD sites can be located based on the energy value needed to open a nucleotide base at the corresponding location (destabilization energy). Hence, low energy states are associated with low SIDD values (SIDD valleys) that corelate with "easier-to-open" DNA. Evidently, genome-wide replication origins in several species so far have been successfully deciphered by mapping the binding sites of the pre-RC components using microarray and ChIP-sequencing.

Origin timing and efficiency

The differential timing of DNA replication ensures complete duplication of the genome along with regulated gene expression. Origin timing and efficiency are important parameters that are measured as population averages through distinct methods. Replication origins have characteristic time of firing, where the early origins often passively replicate the later or relatively less efficient origins. Some origins are active in a majority of cells within a population making them "efficient", whereas the others are rarely used and are dormant. Timing and efficiency of neighbouring regions are interrelated, where the early and more efficient origins replicate their late lesser efficient neighbours. Taken together, early firing, efficient origins have a higher probability of firing over the period of S phase.

Chromosomal environment greatly influences the origin firing timing because implications of chromatin structure, transcription or subnuclear localisation are known to dictate origin control. Sir3, a chromatin modifier required for subtelomeric heterochromatin assembly, is involved in delayed replication of subtelomeric regions (Stevenson and Gottschling 1999). Similarly, the deletion of *RPD3*, a deacetylase, advances the timing of most non-telomeric late firing origins. Tethering of the acetyl-transferase Gcn5 to a late firing origin advances its timing indicating that early firing is sensitive to histone acetylation levels (Vogelauer, Rubbi et al. 2002). Many genes possess a conserved replication timing owing to their physiological requirement to replicate at a certain point in the S phase. Histone genes,
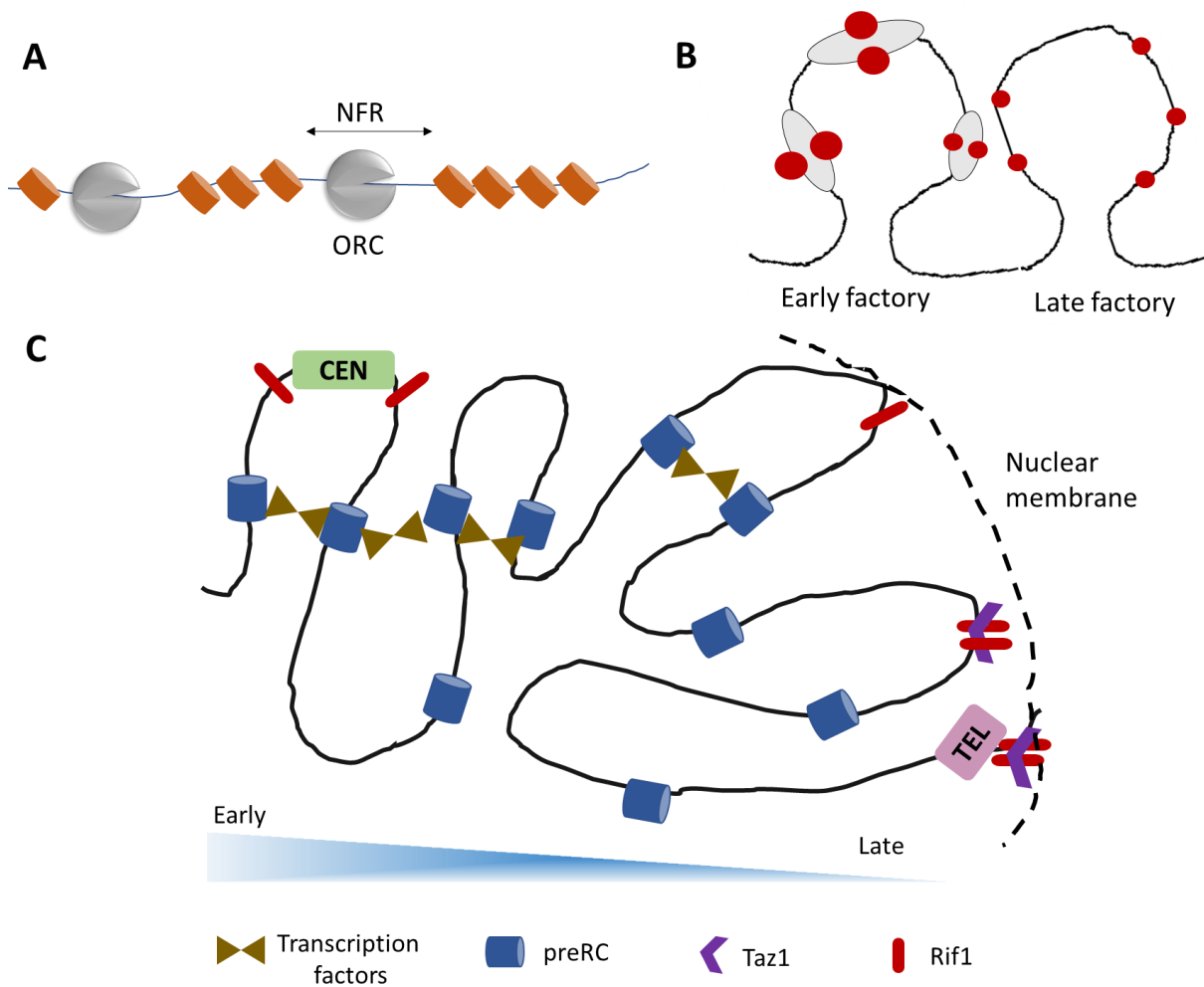
Figure 1.4 Factors determining origin location and firing in space and time. (**A**) Replication origins are preferably located in nucleosome-free regions (NFRs) because this facilitates ORC binding. (**B**) Early replicating domains are clustered together to facilitate easier access of core and accessory proteins required for initiation events. These replication "factories" are formed by chromatin looping and formation of topologically associated domains (TADs). (**C**) Similarly timed replication domains are brought close together by transcription factors like forkhead transcription factor Fkh1/2. Conserved replication timing of centromeres in yeast (early replicating) and telomeres (late replicating) are brought about by chromatin associated proteins such as Taz1 and Rif1.

for example are replicated early to ensure maximal expression in S phase (Muller and Nieduszynski 2017).

Transcription factors act as global determinants of replication timing (Figure 1.3C). The forkhead transcription factors, Fkh1 and Fkh2 are associated with early replicating regions, which harbour binding sites for these proteins (Knott, Peace et al. 2012). However, early firing of centromeres is not attributed to Fkh1/2 stating that centromeres in yeast species have an intrinsic early origin associated activity. In *S. pombe*, telomeres are late replicating

whereas centromeres are early replicating, even though they share similar DNA sequence elements and heterochromatin environment. Rif1 and Taz1, the telomere binding proteins delay the replication timing of sub-telomeric regions. Rif1, surprisingly interacts with pericentromeres advancing their replication timing (Hayano, Kanoh et al. 2012, Yamazaki, Hayano et al. 2013). Temporal regulation by Rif1 is similarly seen in human cell lines as well. Binding of Abf1 on the *S. cerevisiae* ARS elements enables them to provide a nucleosome free environment for MCM loading. Overall, the impairment of these timing factors causes a global loss is replication timing control.

Origin specification and location

S. cerevisiae has strict DNA sequence-specific discrete origins that are evenly placed across the genome and fire in about 90% of the cell cycle (Fangman and Brewer 1991). This does not seem to be the case in its distant cousin, fission yeast, which has relatively inefficient origins (~30%) showing stochastic firing properties (Patel, Arcangioli et al. 2006). Furthermore, origins in fly and frog embryos are regulated in a more non-random stochastic fashion, because some cells utilize a different combination of origins than others, giving rise to cell type specific replication programs (Rhind 2006). Hence, budding yeast seems to be the exception in eukaryotic replication rather than a general rule. But the occurrence of stochasticity in fission yeast and frog genomes gives rise to the "random gap problem" where, due to the non-random distribution of origins, large gaps would be created between the replication bubbles that would take abnormally longer times to complete replication. An elegant rescue to this problem stems from the proposition that origins that fire stochastically show increased efficiency as S-phase progresses (Patel, Arcangioli et al. 2006, Rhind 2006). Although this "increasing efficiency model" ensures that no gaps will persist beyond S-phase, it produces random patterns of replication that is inconsistent with the defined replication timing in metazoan genomes. The existence of late replicating origins itself indicates that there is a deterministic model for origin firing. The "relative efficiency model" reconciles the stochastic firing of individual origins with defined replication timing pattern of genomic regions (Rhind 2006). In short, regions that have efficient origins almost always replicate early in every S phase. These models provide tentative explanation for accommodating opposing features of stochastic firing and a deterministic model for origin firing.

Whole genome chromosome conformation capture (3C) analysis in budding yeast revealed that early origins form enriched interactions and cluster together (Duan, Andronescu

et al. 2010) consistent with another study where similarly timed origins tend to colocalise in the nucleus (Heun, Laroche et al. 2001). Fkh1/2-activated origins (early) are separated from the Fkh1/2- repressed origins (late), and each of these formed distinct interaction clusters indicating that these groups occupy designated areas within the nucleus (Knott, Peace et al. 2012). In *Candida glabrata,* 3C and deep sequencing revealed a total of 275 ARSs and 253 replication origins and also showed the clustering of early replicating origins using Hi-C interaction (Descorps-Declere, Saguez et al. 2015). Clustering of replication origins with similar replication timing (Figure 1.4B) leads to formation of replication zones that are defined by the availability of the limited initiator proteins. In mammals, it is the replication factors that are relocated to the nuclear periphery than the chromosomes themselves. Hence, clustering of chromosomal domain helps in preferential distribution of initiation and catalytic proteins that are limiting in concentration, for the timely completion of the replication program. Evidently, spatiotemporal regulation of initiation events controls replication timing of diverse genomes (Aparicio 2013).

Chromosome segregation

To maintain genetic integrity, eukaryotic chromosomes must be transmitted with utmost fidelity to the next cell cycle. This is carried out by specified loci on every chromosome, the centromere, that directs the assembly of multi-layered proteinaceous machinery, the kinetochore, through which sister chromatids are pulled towards opposite poles with the help of spindle microtubules during mitosis. Molecular mechanisms and genetic pathways of centromere-kinetochore and kinetochore-microtubule interactions have been explored in a wide variety of species.

Centromere

A centromere is classically defined as the primary constriction on a metaphase chromosome that holds the sister chromatids together, binds to spindle microtubules and brings about their separation during anaphase (WALTHER FLEMMING *et al*., 1882). Despite having a conserved and essential function, centromeres are among the fastest evolving DNA sequence loci in eukaryotic genomes. Centromeres are typically located in regions of a genome with less gene density and a reduced rate of recombination. While in most

organisms, they occupy discrete sites on monocentric chromosomes, they may extend up to the entire length of a holocentric chromosome.

Kinetochore

Kinetochores are large macromolecular complexes assembled on the centromere DNA that act as mechano-sensors for pulling forces generated by microtubules. Several centromeric proteins were first identified serendipitously using sera from patients suffering from the CREST disease. This led to the identification of three antigens, CENPA, CENPB and CENPC (Palmer, O'Day et al. 1987). Subsequently, other members of the CENP family were identified in several species. Localised to the most proximal part of the centromeric chromatin is the constitutive centromere associated network (CCAN) (reviewed in (Musacchio and Desai 2017). Centromeric chromatin possesses specialized nucleosomes which comprises of the centromere specific histone variant of H3, CENPA. CENPA dimerizes with H4 and gets incorporated into centromeric chromatin. Most of the CCAN orthologs have been identified in *S. cerevisiae* as the Ctf19 complex, with the notable exception of the CBF3 complex. The outer kinetochore transduces the force generated by the depolymerising microtubules to move chromosomes. It primarily comprises of the KMN complex containing Knl1, Mis12 and Ndc80 sub-complexes. These proteins are required for the end-on microtubule binding. Relative to CCAN, KMN is broadly conserved across species. Apart from KMN, the Dam1 and SKA complexes helping in kinetochore-microtubule coupling by tracking dynamic microtubules and binding with the cognate Ndc80 complex specifically (reviewed in (Musacchio and Desai 2017). A recent study shed light on the diversification of eukaryotic kinetochore and by examining 70 different kinetochore proteins revealed that the KMN complex proteins Spc24, Spc25, Dsn1 are absent in *D. melanogaster* and *C. elegans* even though their functional counterparts are present in these species (van Hooff, Tromer et al. 2017). The kinetochore of *Trypanosoma brucei* comprises of proteins that are not homologous to the "canonical" kinetochore proteins (Akiyoshi and Gull 2014). *D. melanogaster* and *C. elegans* lack most of the CCAN member proteins. C. elegans has a CENPA like histone hcp-3 (Buchwitz, Ahmad et al. 1999) and two homologues of CENP-F, hcp-1 and hcp-2 (Moore, Morrison et al. 1999).

CENPA and the associated chromatin

The centromere-specific variant H3, CENPA is the epigenetic determinant of centromeres in most eukaryotes. It is known by different names: Cse4 in budding yeast
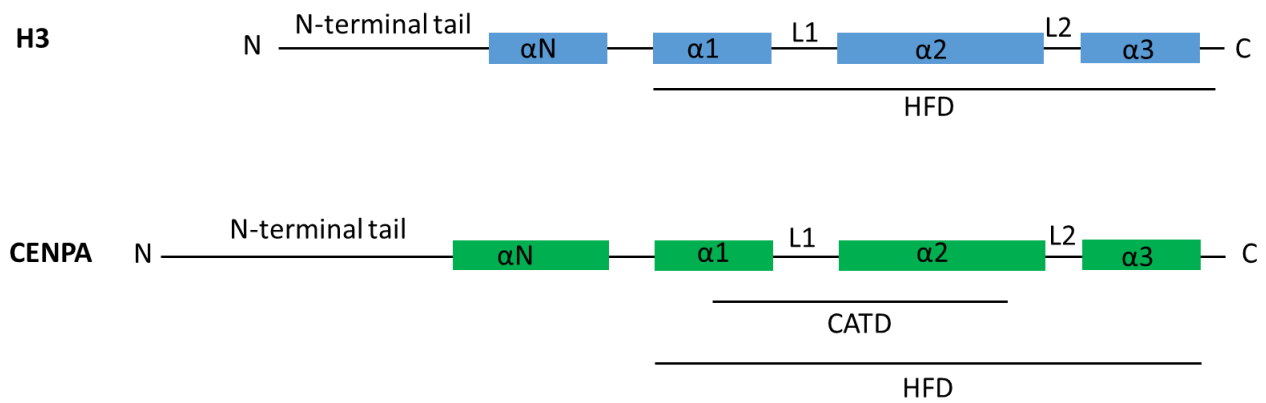
Figure 1.5 Structural features of histone H3 and CENPA. The histone fold domains (HFD) in both H3 and CENPA is comprised of the helices α1, α2, α3 and loops L1 and L2. The centromere targeting domain (CATD) specific to CENPA lies within the α1 and α2 region of the HFD.

(Stoler, Keith et al. 1995), Cnp1 in fission yeast (Takahashi, Chen et al. 2000), Cid in *Drosophila* (Henikoff, Ahmad et al. 2000) and CENPA in humans. Similar to H3, it interacts with H4. But CENPA nucleosomes protect shorter stretches of DNA (100-120 bp) that is enhanced by CENPC binding. CENPA is an evolutionary adaptor molecule that connects diverse centromere DNA sequences to well conserved kinetochore proteins. Being a histone, it possesses a histone fold domain (HFD) (Figure 1.5). Within the two helices of the HFD lies the CENPA targeting domain (CATD), which is required and is sufficient for CENPA targeting to centromere DNA (Black, Brock et al. 2007, Black and Cleveland 2011). Following targeting to CEN chromatin, CENPA is stabilized by the binding of kinetochore proteins, the assembly of which varies from yeast to humans. The nature of CENPA chromatin differs from the canonical H3 chromatin. Unlike H3 chromatin, CENPA chromatin is resistant to folding into higher order nucleosomes. This creates vast arrays of "open" chromatin structure. The nature of CEN chromatin has been modelled and experimentally deciphered in various organisms. In *S. cerevisiae*, centromeric nucleosome subunits are resolved into distinct ladder like pattern, indicating the highly ordered nature of CEN nucleosomes (Bloom and Carbon 1982). In *S. pombe,* the central core and associated repeats show little or no evidence of regularly spaced nucleosomes giving rise to an atypical chromatin structure (Polizzi and Clarke 1991). In *C. albicans,* periodic nucleosome arrays are disrupted at active centromeres giving rise to a smeary MNase digestion pattern (Baum, Sanyal et al. 2006). In *Drosophila*, a hemisome consisting of one molecule each of CENPA, H4, H2A and H2B have been purified. Alternatively, in *S. cerevisiae*, the non-histone

chaperone Scm3 has been proposed to be a part of the CENPA nucleosome (reviewed in (Malik and Henikoff 2009)) . Topologically, CENPA nucleosomes of *Drosophila* and budding yeast induce positive supercoils as opposed to H3 nucleosomes that induce negative supercoils (Furuyama and Henikoff 2009).

Factors that determine centromere specification and maintenance

DNA sequence

DNA sequence provides the underlying platform for the establishment of centromeric chromatin and facilitates loading of kinetochore proteins. The DNA sequence at the centromere varies across eukaryotes. Most organisms harbour a CENPA binding region and a surrounding heterochromatin (pericentromeric region) together constituting a functional centromere on a chromosome. The small point centromeres in *S. cerevisiae* harbour a short 125 bp sequence that serves as a sequence dependent functional centromere. Conserved sequence elements like CDEI, CDEII and CDEIII render these point centromeres genetic properties and also stabilize an ARS plasmid (Fitzgerald-Hayes, Clarke et al. 1982). Point centromeres are also found in *C. glabrata* (Kitada, Yamaguchi et al. 1997), *Ashbya gossypii* (Dietrich, Voegeli et al. 2004) and *Kluyveromyces lactis* (Heus, Zonneveld et al. 1993). The unconventional point centromeres of *Naumovozyma castellii* bears unique CDE elements that binds to the conserved CBF3 complex ortholog, indicating an independent evolutionary origin from *Saccharomyces* (Kobayashi, Suzuki et al. 2015) (Figure 1.6A).

C. albicans has unique and different centromere DNA sequence on each of its chromosome spanning 3-5 kb in length (Sanyal, Baum et al. 2004). A closely related species, *Candida dubliniensis* also harbours intermediate centromeres similar to *C. albicans* (Padmanabhan, Thakur et al. 2008). The centromere DNA sequence in *Candida lusitaniae* is unique for each chromosome and spans 4 - 4.5 kb (Kapoor, Zhu et al. 2015). The repeat associated centromeres of *Candida tropicalis* contain a 2-5 kb CENPA-rich region surrounded by 2-5 kb inverted repeats (Chatterjee, Sankaranarayanan et al. 2016). *S. pombe* has centromeres with a 10-15 kb CENPA binding region covering the central core (*cc*) and innermost repeats (*imr*), which is immediately flanked by outermost repeats (*otr*) consisting of *dh* and *dg* repeats (Clarke, Amstutz et al. 1986, Steiner, Hahnenberger et al. 1993). This pericentromeric heterochromatin spans 10-60 kb around the CENPA binding region.
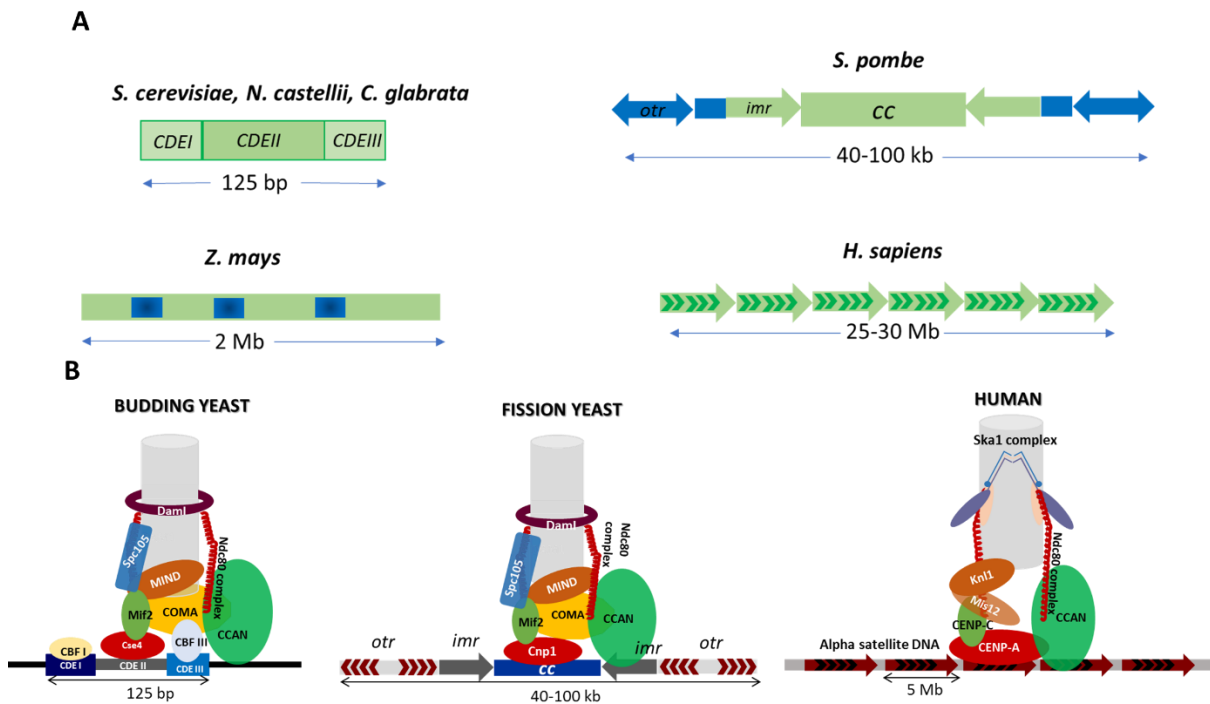
Figure 1.6 Diverse centromere DNA sequences bind to conserved kinetochore proteins. (A) Centromere structure in different species shows the simple point centromeres of *S. cerevisiae*, repetitive large regional centromeres of *S. pombe* with the central core and inner repeat *imr* being the CENPA binding region and the outer repeat (blue) acting as the pericentromeric chromatin. The large regional centromeres of maize are interspersed with retrotransposons (blue box), human centromeres are higher order structures composed of alpha satellite repeats (green arrows). **(B)** (Left) The point centromere in budding yeast containing CDE I, CDE II and CDE III elements holds point centromere specific kinetochore proteins (Cbf I, Cbf III). The CCAN (green) which consists of the CENP-T, S, X, W and other proteins is connected to the microtubule (grey cylinder) binding proteins Dam1 (purple), Ndc80 complex (red spiral) by MIND complex. (Right) Human centromeres are in the form of alpha satellite DNA repeats (maroon arrows), where each monomer (black arrow) spans 171 bp. The SkaI complex (pink-purple-blue trio) forms an inverted W structure along the microtubules. The KMN complex has Ndc80, Mis12 (light brown), Knl1 (dark brown).

Centromeres of the filamentous yeast, *Neurospora crassa* contain retrotransposon like elements in addition to the LINE elements (Centola and Carbon 1994). Recent studies in the basidiomycetous yeast, *Cryptococcus neoformans* revealed large regional centromeres of 20-65 kb rich in retrotransposons (Yadav, Sun et al. 2018). Hence, quite evidently, fungal centromeres provide a wide array of centromere structure and organization (Figure 1.6).

As one goes up the evolutionary ladder from unicellular yeast species to developmentally regulated, multicellular organisms, the criteria for a region to act as a centromere becomes hazier. Long stretches of complex AT rich repeats interspersed with

transposons make up the repetitive centromeres in *Drosophila* which has no DNA sequence specific to the centromere alone (Sun, Le et al. 2003). Using a functional minichromosome Dp1187 derived from the X-chromosome, it was revealed that satellite sequences at the centromere are interrupted by 5 retrotransposons and an island of 'complex' DNA (Sun, Wahlstrom et al. 1997). Chicken centromeres are associated with long tandem repeats and contain MHM repeats units, the copy number of which are chromosome specific (Shang, Hori et al. 2010). In humans, centromeres are associated with alpha satellite repeats, the monomeric form (170 bp) of which is arranged in a tandem head to tail fashion to form an array. These can form Higher Order Repeats (HOR) forming 3-5 Mb of large regional centromere interspersed with LINE, SINE and LTR elements (Schueler and Sullivan 2006). The human Y chromosome centromere was recently sequenced using Nanopore technology which revealed the complete genomic definition of a human centromere (Jain, Koren et al. 2018, Jain, Olsen et al. 2018). Within homogenous arrays, different alpha satellite dimeric units were observed on CENPA containing nucleosomes rendering structural and conformational anomaly to the centromeric chromatin in humans (Thakur and Henikoff 2018). Similar to humans, centromeres in the plant *Arabidopsis* are enriched in 178 bp satellite repeat organized in tandem arrays spanning 0.4- 1.4 Mb (Round, Flowers et al. 1997). Similarly, in *Zea mays*, there are mainly two kinds of centromeric repeat sequences, a 156 bp tandem CentC repeat and centromeric retrotransposon of maize (CRM). These sequences interact with CenH3 nucleosomes (Zhong, Marshall et al. 2002). The clusters of alpha-satellite DNA in humans and *Arabidopsis* are organized in a similar way although DNA sequences of the repeat subunits differ significantly. Such a plethora of diversity in centromere organization across species counts for the genome complexity observed from unicellular eukaryotes to multicellular organisms.

Most plant and animal species harbour monocentric chromosomes. In contrast, certain insect species (butterflies, moths, dragonflies, bugs and lice) have holocentric chromosomes where centromere activity is distributed along the length of the chromosome. It was revealed in a comparative study that CENPA is absent in species with holocentric chromosomes and present in insects with monocentric chromosomes (Drinnenberg, deYoung et al. 2014). In spite of this anomaly, many of the kinetochore proteins were still present in these species. Cytological observations in the worm *C. elegans* have shown the existence of holocentric chromosomes where spindle fibres attach along the length of a chromosome and pull them apart (Albertson and Thomson 1982). Experimental evidence in *C. elegans* suggest that

holocentromeres in this worm are composed of 700 individual point centromeres that are distributed along the length of the chromosome (Steiner and Henikoff 2014). This independent transition to holocentricity is a possible way to introduce changes in a centromere that does not require a CenH3 protein.

Heterochromatin and RNAi machinery

The mechanisms involved in heterochromatin formation (*de novo*) and maintenance (propagation) work hand in hand and sometimes in separation to ensure genomic stability (Hall, Shankaranarayana et al. 2002). A prominent and well worked out histone modification is the Clr4-HP1 mediated methylation at H3K9 which acts as a binding platform for Heterochromatin Protein 1, HP1 or Swi6 (fission yeast homolog) (Bernard, Maure et al. 2001, Hall, Shankaranarayana et al. 2002, Allshire and Madhani 2018) (Figure 1.7A). This is seen at centromeric and arm heterochromatin in fission yeast. Sir proteins that deacetylate, H3 promote H3K9 methylation by Clr4. RNAi is essentially used for heterochromatin establishment rather than maintenance (Zofall and Grewal 2006). The RNAi pathway involves RNA pol II transcribing heterochromatin repeats which are then processed by Dicer as single stranded siRNA to recruit silencing factors (Verdel, Jia et al. 2004) (Figure 1.7A). Ago1 is part of a complex that associates with an RNA dependent RNA polymerase (RdRP) and the methylase Clr4 (Sugiyama, Cam et al. 2005, Zhang, Mosch et al. 2008). Once nucleated, heterochromatin can spread or expand its domain irrespective of the underlying DNA sequence. The spreading need not occur in a linear fashion, but a heterochromatin domain and a nearby domain that provides a spatial environment to harbour key modifications can spread the seeds to newer locations. A visible decline in this frequency has been reported with an increasing distance from the nucleation site (Obersriebnig, Pallesen et al. 2016). Heterochromatin spreading is a double-edged sword. In order to avoid potentially deleterious gene silencing, there are mechanisms to ensure restriction of this spreading. *S. pombe* has physical boundary elements to prevent such a spreading. tRNA genes such as TFIIIC binding sites are present near centromeric and MTL boundaries (Noma, Cam et al. 2006). The histone variant H2A.Z is deposited at NFRs at the +1 nucleosomes of actively transcribing TSSs (Zofall, Fischer et al. 2009).

Heterochromatin has pivotal roles in centromere and kinetochore function. In *S. pombe,* flanking pericentromeric heterochromatin directs histone modification and mediates CENPA
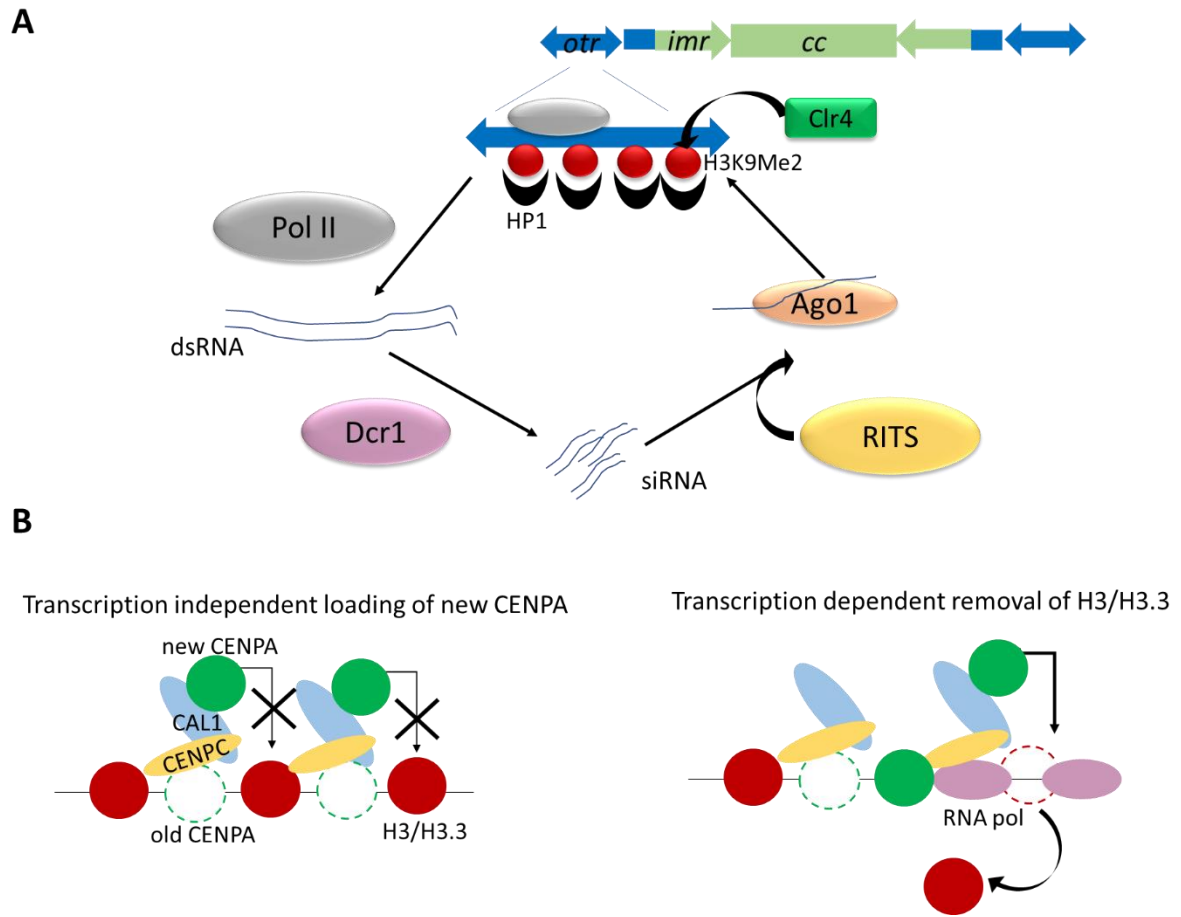
Figure 1.7. Heterochromatin and transcription machinery influence centromere identity. (A) The outer repeats of *S. pombe* centromeres produce transcripts that are processed by the RNAi machinery into single stranded siRNA to induce the formation of silent chromatin. This circuit also helps to recruit Swi6/ HP1 to dimethylated lysine nucleosomes at the pericentromeric repeats. This not only maintains centromeric silencing, but also restricts the localisation of CENPA to the central core. **(B)** Transcription-coupled eviction of the placeholder H3.3 nucleosomes has been observed in *Drosophila* centromeres. CAL1-mediated targeting of the CENPA/H4 complex to centromeres is restricted by H3.3, which are removed with the help RNA polymerase II, permitting incorporation of new CENPA molecules.

establishment, limiting its assembly to the central core. (Scott, Merrett et al. 2006). Ectopic heterochromatin formation prevents kinetochore assembly in this organism. A second, yet non-obvious role of centromeric heterochromatin involves sister-chromatid cohesion that is mediated by cohesins. During the metazoan metaphase, most sister chromatids are associated only at their centromeres, as centromeric cohesins are protected from degradation (Bernard, Maure et al. 2001). However, in *S. pombe*, centromeric cohesion is tightened by entrapment of higher levels of cohesion and this is facilitated by its physical association with the HP1 homolog, Swi6 (Nonaka, Kitajima et al. 2002). Cells with a defective heterochromatin

machinery fail to assemble functional kinetochores leading to severe chromosome segregation defects (Pidoux, Uzawa et al. 2000). Hence, a heterochromatic environment facilitates centromere establishment (CENPA loading) and also its stable propagation to the next cell cycle.

Transcription

A high level of transcription or a complete lack of it is incompatible with CEN function. Transcription plays an important role in specifying CEN identity and maintenance. However, pervasive level of transcription helps in CEN function as studied extensively in *S. pombe* centromeres. Stalling of RNA polymerase at CEN is a trigger for remodeling events that help in depositing CENPA at fission yeast centromeres (Catania, Pidoux et al. 2015). This stall also indicates that centromeres are difficult-to-transcribe regions. Due to redundant sequence features like presence of transcription start sites and promoter like elements, the central domain of *S. pombe* can create an unusual transcriptional environment that is permissive for CENPA establishment at the central core (Choi, Stralfors et al. 2011, Allshire and Ekwall 2015). Apart from fission yeast, transcripts from the centromere have been detected in species such as maize (Topp, Zhong et al. 2004), *Drosophila* (Rosic, Kohler et al. 2014), humans (Chan, Marshall et al. 2012) and budding yeast (Ohkuni and Kitagawa 2011). A pan-fungal analysis of RNAi proteins revealed that a few species, including *C. glabrata* and *Ustilago maydis*, have lost all of the proteins required for functional RNAi during the course of evolution, whereas species including *C. albicans* and *C. tropicalis* harbor a cryptic RNAi machinery (Nakayashiki, Kadotani et al. 2006). A recent study in the pathogenic *Cryptococcus* species complex correlated the loss of RNAi with the length of centromeres, thereby proposing that RNAi helps in maintaining long repetitive, transposon-rich centromeres (Yadav, Sun et al. 2018). Whether the RNAi machinery has a functional significance in the centromere biology of this species complex and other fungal pathogens remains unexplored. Apart from RNAi, centromere transcription can also play a functional role through long non-coding RNA (ncRNA).

Even in humans, the transcripts generated from the Higher Order Repeats (HOR) of the alpha-satellite DNA interact with CENPA, rendering structural stability to CEN chromatin (McNulty, Sullivan et al. 2017). A recent study is *Drosophila* elucidates that transcription coupled remodeling is required for CENPA incorporation (Bobkov, Gilbert et al. 2018) (Figure 1.7B). Transcription induced destabilization of nucleosomes helps in eviction of

nucleosomes by closely spaced Pol II complexes. RNA pol mediated transcription at the fly centromere helps CENPA to transit from an unstable chromatin-associated state to stably incorporated nucleosomes (Bobkov, Gilbert et al. 2018). Hence, the study of heterochromatin and transcription in a broad range of model organisms has answered many questions pertaining to their establishment and function of centromeres. However, aspects regarding the heritability of centromeric heterochromatin and what facilitates transcription at these regions are still unexplored.

Replication timing of the centromere

DNA replication at the centromeres is strictly a temporally regulated process. Centromeric regions are replicated in the earliest part of the S-phase in all the *Saccharomyces* species (Pohl, Brewer et al. 2012), *C. albicans* (Koren, Tsai et al. 2010), *S. pombe* (Rhind 2006) and the parasite *Trypanosoma brucei* (Calderano, Drosopoulos et al. 2015). The temporal effect on firing has been studied in yeast species including *C. albicans* where an early replicating origin flanking the centromere advances the replication timing of the centromere (Koren, Tsai et al. 2010). Strikingly, deletion of the native centromere, gives rise to a neocentromere with the activation of an early firing neo-origin. This clearly states that centromeric location establishes replication timing of the adjacent regions. Early replication of CENs due to early firing proximal origins can be attributed to their characteristic clustering and nuclear sub-positioning. Even in *S. cerevisiae*, relocation of the CEN to a late firing region advanced its replication timing further reinstating that the mere presence of a CEN sequence can modulate replication timing (Pohl, Brewer et al. 2012). A slight variant form of this timing influence exists in the fission yeast, where the heterochromatin environment at the centromeres and subtelomeric regions is similar. A Swi6 mediated recruitment of DDK ensures early replication of CEN, matK locus which could otherwise be late replicating owing to their heterochromatic environment (Bannister, Zegerman et al. 2001, Hayashi, Takahashi et al. 2009). Contrary to yeast and protozoan cells, metazoan centromeres replicate late during S phase since CENPA loading in these organisms is replication uncoupled.

Spatial cues within the nucleus

Most fungal centromeres are clustered near spindle pole bodies (SPBs). This association may result in folding back of chromosomes and positioning them such that telomeres are juxtaposed in the interphase nucleus, giving rise to the Rabl conformation (Rabl, 1885). This
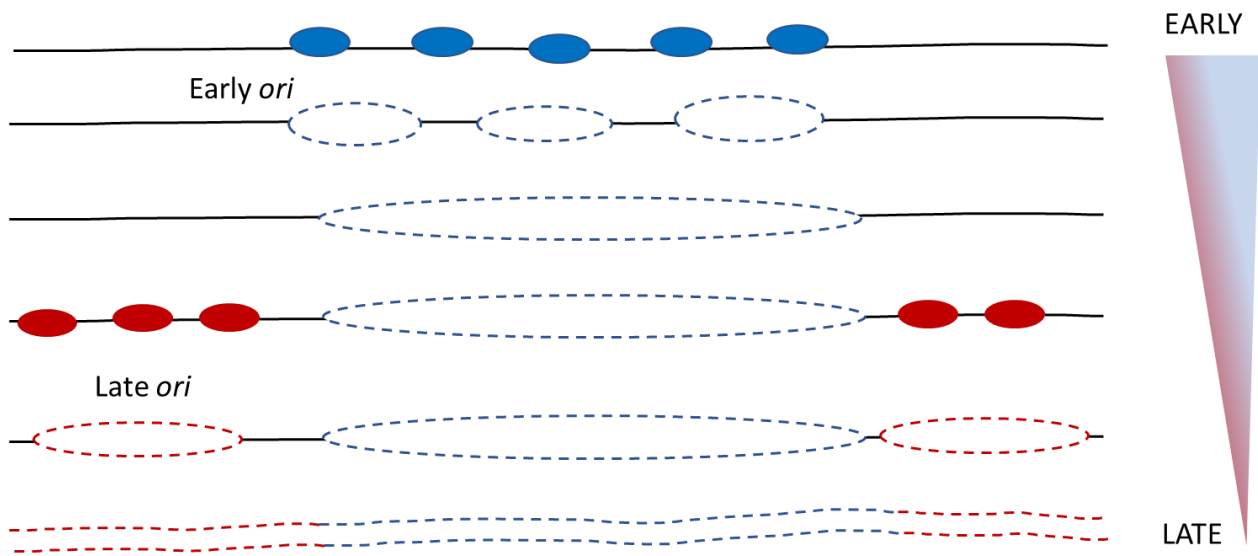
Figure 1.8. Temporal regulation of DNA replication. Fungal centromeres are early replicating. Early firing origins flanking centromere initiate DNA synthesis and replicate large parts of the genome. Subsequently, late replicating (including telomeric regions) regions are replicated later in the S phase.

phenomenon has been shown to occur in fungal species including *C. albicans, C. tropicalis,* and *Fusarium graminearum.*

Using 3C experiments, clustered centromere DNA regions were shown to be present in close spatial proximity, leading to physical interactions between different centromeres (Duan, Andronescu et al. 2010). It has been proposed that the clustering of centromeres aids in determining the site of centromere formation in these organisms. According to this hypothesis, a part of the nucleus is enriched with a pool of CENPA proteins to form a CENPA rich zone or CENPA cloud (Fukagawa and Earnshaw 2014). It was proposed that the region of a chromosome that is near this CENPA cloud would attract a higher level of CENPA and thus serve as a preferred site for centromere formation. In *S. cerevisiae*, for example, a locally enriched population of accessory CENPA molecules at pericentromeric chromatin has been shown to serve as a reservoir for rapid incorporation of CENPA in the event of premature eviction from centromeres (Haase, Mishra et al. 2013). Further evidence supporting the CENPA cloud hypothesis stems from studies in *C. albicans* in which neocentromeres were formed close to the native centromere in most cases (Thakur and Sanyal 2013, Burrack, Hutton et al. 2016). Interestingly, centromeres were found to be unclustered in premitotic *C. neoformans* cells that eventually cluster at the onset of mitosis

(Kozubowski, Yadav et al. 2013). Whether this centromere clustering also arises as a result of physical interactions among centromeres is not yet known. In the vertebrate interphase nuclei, studying 3D genome organization using 4C analysis revealed the clustered nature of centromeres forming compact chromatin (Nishimura, Komiya et al. 2018). The neocentromeres formed in these vertebrate cell lines are commonly associated with specific heterochromatin-rich regions in 3D, which was shown to suppress centromeric drift in non-repetitive centromeres. Hence, the 3D architecture of centromere and its associated chromatin within the nucleus provides ample insights into the spatial cues required to specify centromere location.

Plasticity of CEN chromatin

The fact that the evolutionarily conserved protein CENPA can bind to diverse DNA sequences across and even within species suggests the malleable nature of centromeric chromatin. This feature has been exemplified in many natural and artificially induced scenarios like gene silencing, centromere repositioning and neocentromere formation.

Transgene silencing at the centromere

Position effect variegation (PEV) refers to varying expression pattern of a gene due to its translocation to a specific position in the genome. Originally identified in fruit flies through the variegated expression states of the translocated white gene causing mosaic eye colouration patterns (Schotta, Ebert et al. 2003), in fission yeast, PEV has been known to cause heterochromatin spreading at the mating type locus (Ayoub, Goldshmidt et al. 1999). Telomeric silencing of a transgene in *S. pombe* does not alter the mRNA levels of the transgene, unlike in budding yeast. Within the central core of the centromere, however, PEV is observed due to the differential positioning of CENPA nucleosomes with respect to H3 nucleosomes rendering a variable expression of the underlying reporter gene (Allshire, Javerzat et al. 1994, Grewal and Klar 1996, Yao, Liu et al. 2013). Additionally, heterochromatin marks from pericentromeric regions can "spread" to the core, mitigating gene silencing. This involves breaching of the tRNA boundary element that separates it from the adjacent euchromatic regions. There exists an epigenetically heritable mechanism that acts as a spatial cue to create a memory of CENPA occupied chromatin. Over-expression of CENPA leads to an enhanced rate of silencing at the core. Regardless of the species, PEV

does influence variegated gene expression that is inherited in a clonal fashion. PEV has since then been widely used in *S. pombe* as a convenient readout to study the fidelity of nucleosome mediated inheritance (Li, Yi et al. 2017).

Neocentromeres and centromere repositioning

The episodic occurrence of centromere activity at non-centromere sequences, neocentromeres, strongly suggests the epigenetic nature of centromeres (Figure 1.9A). The first occurrence of a neocentromere in a human genome was reported to be formed on a marker chromosome, Mardel10 and was devoid of any alpha satellite repeat (Voullaire, Slater et al. 1993). Human neocentromeres have no alpha-satellite, but they can incorporate kinetochore proteins. Subsequently, CENPA enrichment patterns in neocentromeres identified have been found to be variable and often less abundant than the endogenous centromere (Ross, Woodlief et al. 2016, Sullivan, Maloney et al. 2016). Most of the identified human neocentromeres are seeded on acentric fragments consisting of inverted duplications of a distal portion of a chromosome arm (Rocchi, Archidiacono et al. 2012). Additionally, majority of the CENPA domain at neocentromeres harbour protein coding genes with similar transcription competence as the native context. In humans, neocentric inverted duplications generally occur at the distal end of chromosome ends. Neocentromere formation does not affect gene expression per se. However, rewiring of the pericentromeric regions might affect the gene structure and hence the consequent expression (Sullivan, Maloney et al. 2016).

Neocentromere features across species reveal certain commonalities as well as certain species-specific attributes. *S. pombe* forms sub-telomeric neocentromeres (Ishii, Ogiyama et al. 2008) whereas in humans they are more prevalent at sub-metacentric regions (Warburton 2004). In contrast, most neocentromeres have been detected at *CEN* proximal loci in *Drosophila* (Maggert and Karpen 2001), chicken (Shang, Hori et al. 2013), *C. albicans* (Thakur and Sanyal 2013), barley and maize. The assembly of ectopic CENPA as a "CENPA cloud" surrounding the endogenous CEN and proximity of neocentromere hotspots to native *CEN* in these organisms indicates that CENPA is peppered on CEN adjacent loci and can get rapidly incorporated to the CEN locus in cases of CENPA eviction (Haase, Mishra et al. 2013, Fukagawa and Earnshaw 2014).

A Neocentromere formation

non-centromeric DNA

kinetochore

B Centromere repositioning

centromeric DNA

Repositioned kinetochore

C Centromere inactivation

TSA treatment

non-functional kinetochore

Preventing dicentric formation

non-functional kinetochore
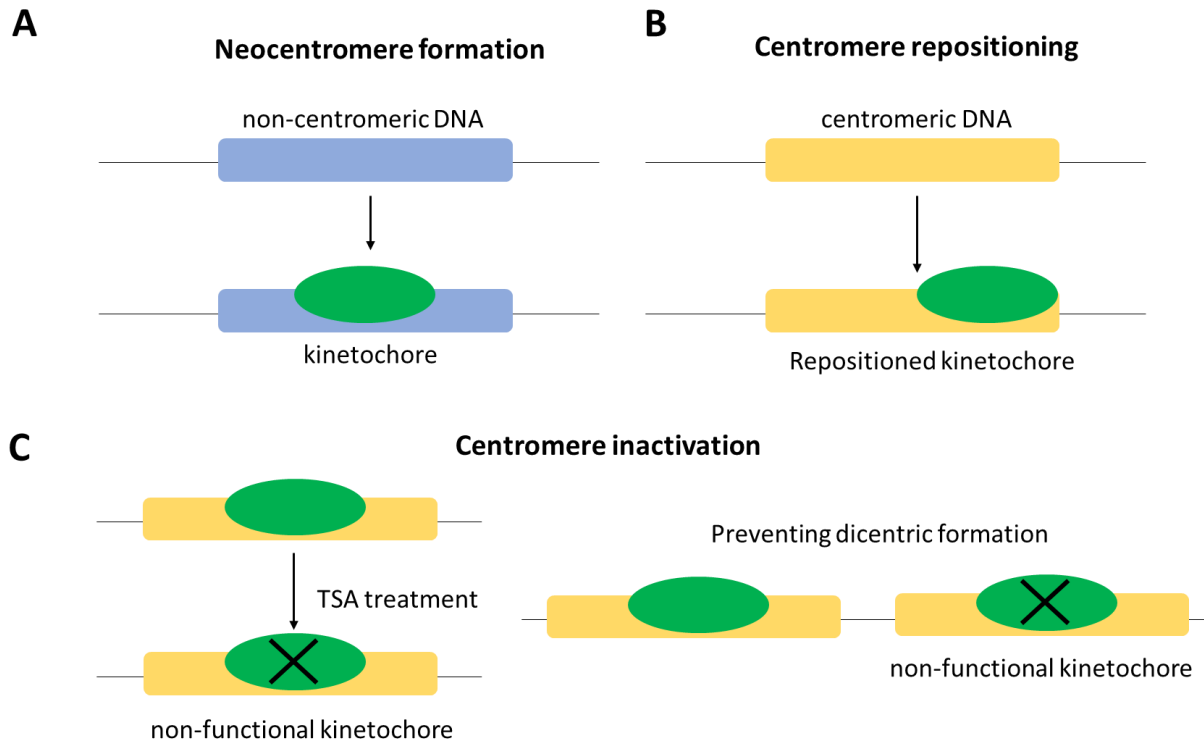
Figure 1.9. Epigenetic nature of centromeres. (A) Formation of a functional kinetochore complex on an ectopic locus has been shown to occur in cases when the native centromeres are deleted forming neocentromeres. **(B)** The position of KT can be drifted or shifted leading to a repositioned CEN. **(C)** Native centromeres can get inactivated with either by erasing the epigenetic marks using chemical inhibitors like Trichostatin A or by the formation of a conditional ectopic centromere

Centromere repositioning can be either a gradual or a one-step *de novo* event (Schubert 2018) (Figure 1.9B). Gradual loss of the old centromere and its emergence at a new locus could occur if one of the (repetitive) sequences of ancestral centromere is detectable at the new position. In rare cases, the old centromere remains unchanged in sequence, but gets inactivated by the absence of CENPA and kinetochore proteins (Rocchi, Archidiacono et al. 2012). However, it still remains elusive as to how this functional switch occurs fast enough that chromosome instability is avoided. Such unloading and mis-incorporation is fast enough to avoid creation of a transient acentric or a dicentric fragment that would render an unstable genome, making this a sudden rather than a gradual course of centromere repositioning. This repositioned centromere will be maintained as long as it does not alter the underlying gene expression. Hence, neocentromeres are preferably formed at gene poor, low GC content areas (Federico, Pappalardo et al. 2017). If two "unequal" centromeres form on the same chromosome, then the survival of the fittest is decided by the centromere that has a

favourable proportion of CENPA nucleosomes, in which case the "weaker" centromere gets inactivated (Figure 1.9C). Reactivation of the weaker centromere can occur if a rearrangement separates the active and inactive centromere, in which case the latter can seed CENPA on an otherwise acentric fragment (Han, Gao et al. 2009). Hotspots for neocentromeres represent areas that are pre-disposed to centromere formation. They are "latent" centromeres which represent position of inactivated ancestral centromeres (Marshall, Chueh et al. 2008).

Centromeres are known to "drift" in vertebrate chicken cells upon prolonged culturing (Hori, Kagawa et al. 2017). Similar to chicken cells, centromere drift has also been observed in fission yeast cells (Yao, Liu et al. 2013). This is only possible because centromeres in these organisms are specified and propagated by sequence independent mechanisms. Generally, centromere drift is considered deleterious to cells, as it can supress the expression of crucial genes (Fukagawa 2013). However, since the mechanism for centromere formation is epigenetically regulated, these events may be naturally occurring albeit at low frequencies.

Crosstalk between DNA replication and chromosome segregation machinery

Centromeric chromatin undergoes major changes in nucleosome composition, structure and architecture during successive cell cycles. Such alterations in specialized chromatin facilitate kinetochore formation in mitosis to ensure proper chromosome segregation. This calls for coordinated orchestration of centromeric chromatin dynamics during interphase, especially during S phase when the DNA is opened up and centromeric chromatin is disassembled and reassembled. There is an ever-increasing evidence of a crosstalk between replication machinery and centromeric chromatin to ensure its table propagation.

Chromatin replication

During S phase, nucleosomes dissociate from the nascent DNA strand and have to be redistributed to the replicated DNA. CENPA containing nucleosomes are diluted during replication. Two models that have been proposed to explain this patterned distribution (Sullivan, Blower et al. 2001, Dunleavy, Almouzni et al. 2011). The placeholder model suggests that post eviction of the parental CENPA molecules, the gaps created are occupied by H3 nucleosomes, suggesting a placeholder function for H3. On the contrary, the gap

filling model suggests that upon CENPA splitting, the gaps are maintained to incorporate new CENPA molecules. In *Drosophila*, where CENPA and H3 occupy distinct non-overlapping domains at the centromeres, H3.3 acts as placeholder for CENPA incorporation (Dunleavy, Almouzni et al. 2011). The H3 containing domains contain H3K4dimethylation marks that are important for CENPA loading by the CENPA specific chaperone, Holiday Junction Recognition Protein (HJURP) and also act as replication initiation zones to facilitate the discontinuous centromere replication in this organism. This trend is seen in case of human centromeres where nascent CENPA is loaded in an oscillatory pattern within the existing CENPA domain (Ross, Woodlief et al. 2016). Unlike in fission yeast, where the central core has maximum CENPA molecules as compared to the inner and outer repeats, vertebrate centromeres have abundant CENPA at the edges than the centre of the alpha satellite domain. This is an effective strategy to restrict the CEN chromatin to a region of repetitive DNA, as evidenced in fly and vertebrate centromeres (Ross, Woodlief et al. 2016). During the late S phase when human centromeres are replicated, nucleosome are disrupted and old histones are recycled with the help of anti-silencing function Asf1 and Mcm2 (Huang, Stromme et al. 2015, Richet, Liu et al. 2015).

The impact of chromatin replication on genome stability and maintenance has been profoundly exemplified across eukaryotes. Removal of "old" histones and replenishment with "new" ones have to be carried out in a timely manner to ensure no perturbation in cell cycle timing. Factors that promote nucleosome turnover through disassembly and/or reassembly of histone exchange help in inhibiting heterochromatin. One such recently identified factor is the yeast homolog of the human SMARCAD1, Fft3 which suppresses nucleosome turnover post replication (Taneja and Grewal 2017, Taneja, Zofall et al. 2017). The presence of silent chromatin in progeny cells after DNA replication suggests that heterochromatin can be self-propagated in a manner that is independent of the underlying DNA sequence. This is an example of *cis*-inheritance of a chromatin state. This kind of stable *cis* inheritance is seen in the H3K9me dependent mating type locus silencing in fission yeast (Scott, Merrett et al. 2006) and is reminiscent of Polycomb-associated sites in *Drosophila* (Berry, Hartley et al. 2015).

Centromere proximal replication origins

In most fungal species studied so far, replication origins have no known conserved DNA sequence. However, their chromatin status and location with respect to essential DNA

elements like centromeres, telomeres, mating type-loci etc. are known to define replication timing of these regions. Regions containing fork termination sites often comprise of genetic elements that are difficult to replicate. Centromeres fall in these stall sites (Greenfeder and Newlon 1992). Unlike metazoans, most fungal species have early replicating centromeres (Raghuraman, Winzeler et al. 2001, Kim, Dubey et al. 2003, Koren, Tsai et al. 2010, Pohl, Brewer et al. 2012). The functional consequence of this temporally distinct replication timing is to maintain viability of cells in the face of a replication stress as seen in case of the segregation apparatus in *S. cerevisiae* (Feng, Bachant et al. 2009). Additionally, early replication timing ensures proper kinetochore assembly at the centromeres (Kitamura, Tanaka et al. 2007). In *C. albicans*, a replication coupled repair mechanism mediated by stalling of converging forks flanking the centromere helps in kinetochore loading. Another study correlates the inheritance of centromere position with a constitutively active origin of replication that is early replicating in every S-phase. (Koren, Tsai et al. 2010, Mitra, Gomez-Raja et al. 2014). Hence, these *CEN*-proximal origins seem to have a role more than just acting as initiator sites for DNA replication.

The physical proximity of a centromere and a replication origin has been known to maintain plasmid stability in *Yarrowia lipolytica,* (Fournier, Abbas et al. 1993). The association of chromatin binding proteins with S-phase kinases helps to initiate DNA synthesis at heterochromatin. One such study identifies the HP1 homolog, Swi6 as a factor along with Dbf4-dependent kinase (DDK) to promote the activation of origins responsible for replicating pericentromeric regions and *mat* locus in *S. pombe* cells (Hayashi, Takahashi et al. 2009). DDK is also involved in the recruitment of initiator proteins, Sld3-Sld7, with the help to the Ctf19 complex to facilitate centromere replication in budding yeast. Furthermore, DDK independently recruits the cohesion loaders, Scc2-Scc4 to these origins to maintain pericentromeric cohesion (Natsume, Muller et al. 2013). Hence, this positive regulation helps to coordinate replication timing of different genomic locations, especially heterochromatin replication. At the molecular level, there are correlates of centromere replication timing with CENPA loading.

CENPA loading

The replenishment of CENPA post replication is a cell cycle-mediated process. The mechanism of CENPA loading onto CEN chromatin involves three major steps: licensing of centromeres, loading of newly synthesised CENPA and maintenance of the newly
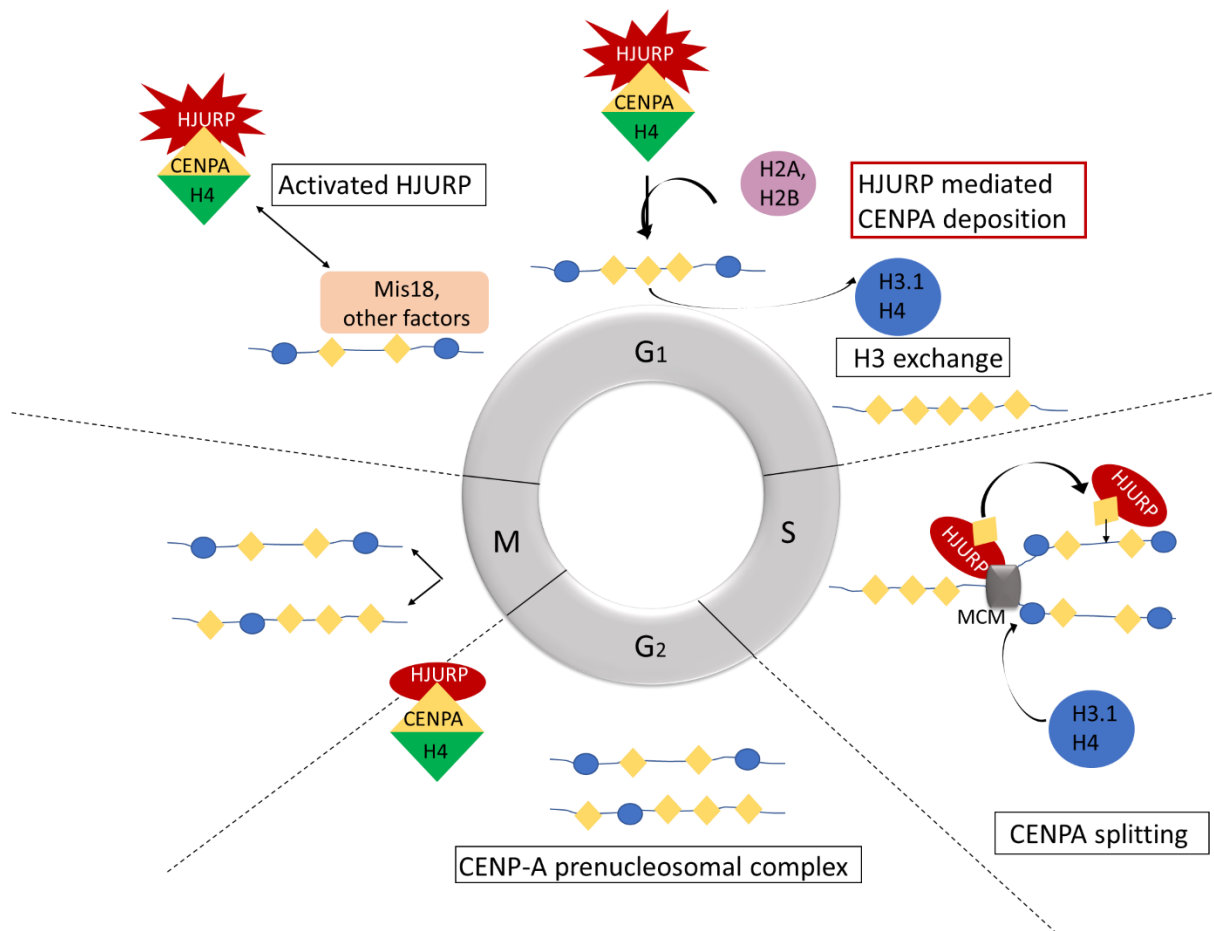
Figure 1.10. Centromeric assembly of CENPA mediated by HJURP in humans. Replication of sister chromatids requires new CENPA to be deposited specifically at centromeric loci in every cell cycle. Existing CENPA nucleosomes are distributed to sister chromosomes during DNA replication (CENPA splitting) with the help of HJURP that simultaneously interacts with old CENPA and Mcm2, following which CENPA levels are halved in G2 and M phases. Gaps are filled using H3 that acts as the placeholder. During G1 phase, new CENPA nucleosomes are incorporated using the chaperone, HJURP. H3 nucleosomes are evicted out to facilitate CENPA loading by specific chaperones. Restriction of CENPA assembly may be achieved by modification of the CENPA prenucleosomal complex or through the recruitment of other licensing factors to the centromere during G1, such as the Mis18 complex, that may serve to prime the centromere and restrict CENPA nucleosome assembly.

incorporated CENPA (De Rop, Padeganeh et al. 2012). The most upstream event is performed by the licensing complex, Mis18 which is required for CENPA localisation in many species (Fujita, Hayashi et al. 2007). The non-histone DNA binding chaperone, HJURP/Scm3 has been implicated in the loading of CENPA (Dunleavy, Roche et al. 2009). Scm3 is required for CENPA loading in both budding and fission yeast (Williams, Hayashi et al. 2009). In the budding yeast, loading of CENPA is replication coupled and occurs in S

Organism	CENPA loading time	CEN replication timing
S. cerevisiae	S- phase	Early S phase
S. pombe	G2 phase	Early S phase
C. albicans	Late anaphase	Early S phase
A. thaliana	G2 phase	Mid-late S phase
D. melanogaster	Anaphase/ metaphase	Mid-late S phase
G. gallus	Mitotic exit	Mid-late S phase
X. laevis	Mitotic exit- early G1	Mid-late S phase
H. sapiens	Late telophase-early G1	Mid-late S phase

Table 1.1. Stage specific deposition of CENPA and CEN chromatin replication timing across different eukaryotic species. The early replication of centromeres in yeast is associated with a replication coupled loading of CENPA. In higher eukaryotes like fruits flies and humans, CENPA loading is during or post mitosis as these organisms have late replicating centromeres.

phase, whereas higher eukaryotes display replication independent loading of CENPA. A mutant that is unable to distribute CENPA evenly would be sensitive to elevated levels of CENPA in the cell. One such mutant was located on a gene Rpt3, that was a member of the proteasome (responsible for degrading excess CENPA), which exhibited mis-segregation of mini-chromosomes and eventually was found to regulate CENPA distribution at fission yeast centromeres (Kitagawa, Ishii et al. 2014). MgcRacGAP was identified as a GTPase involved in centromeric identity, as depletion of this proteins leads to impairment in new CENPA nucleosome incorporation (Lagana, Dorn et al. 2010).

In most organisms, CENPA loading is uncoupled from replication (Figure 1.10, table 1.1). In *S. pombe*, new CENPA is deposited during G2 phase, while H3 acts as its placeholder during S phase (http://dx.doi.org/10.1101/215624). Fluorescence spectroscopic measurements in *C. albicans* revealed that there exists a cell cycle coupled oscillation of centromeric nucleosome in yeast (Shivaraju, Unruh et al. 2012). The intensity of CENPA clusters was found to double in anaphase suggesting that CENPA loading in this organism probably occurs outside the S-phase. In humans, the S phase retention of CENPA is mitigated by its simultaneous interaction with the specific chaperone HJURP and Mcm2 (Zasadzinska, Huang et al. 2018), which together transmit CENPA nucleosomes upon its disassembly ahead of the replication fork. HJURP is transiently known to associate with centromeres during S phase

and bind to pre-existing CENPA, and along with Mcm2 is required for centromeric nucleosome inheritance during S phase (Figure 1.10).

Non-replication functions associated with pre-RC components

Members of the pre-RC have established roles in cell cycle dependent dynamics at the centromere. Mcm2 is a known chaperone that hands over the old histones from the replication fork to Asf1 to recycle old histones and deposits them to the newly synthesised DNA (Hammond, Stromme et al. 2017). Apart from its definitive role in replication, increasing evidence suggests the role of MCMs in other DNA metabolic processes like transcription and chromatin remodelling. Mcm2 and Mcm5 are known to physically interact with the C terminal domain of RNA polymerase II facilitating transcription which may influence origin activity (Holland, Gauthier et al. 2002). MCMs are known to bind to nuclease hypersensitive regions. Recent studies highlight that Mcm2 and ASF1 cochaperone an H3-H4 dimer through histone-binding mode (Richet, Liu et al. 2015). This is true for both canonical H3 as well as H3 variants like H3.3 and CENPA (Huang, Stromme et al. 2015). A very recent study indicates the role played by Mcm2 in human embryonic stem cells to symmetrically partition modified histones to daughter cells using its histone-binding mode (Petryk, Dalby et al. 2018).

Apart from its established roles in replication initiation and origin usage, ORCs are significantly involved in transcriptional regulation with special emphasis on maintenance of heterochromatin loci. Pioneering studies in the budding yeast have yielded sufficient knowledge of ORCs as silencers of telomeres and mating type loci (Fox, Loo et al. 1995). For example, human and *Drosophila* ORC interact with HP1 to mitigate telomeric and heterochromatin silencing (Prasanth, Prasanth et al. 2004, Prasanth, Shen et al. 2010). The largest subunit of ORC, Orc1 contains the Bromo-Adjacent Homology domain (BAH) which bind to repressed chromatin states suggesting that Orc1 is a silencing protein. This domain is required for silencing but not for viability in *S. cerevisiae* (Hickman and Rusche 2010). Hence, an intimate crosstalk between the replication and centromere-kinetochore structure is evident in influencing genome stability.

Candida albicans

Life cycle

The genus *Candida* is comprised of a heterogenous group of organisms possessing both pathogenic and non-pathogenic species where several aspects of DNA metabolism have been studied. More than 90% of the infections are caused by *C. albicans, C. glabrata, C. parapsilopsis, C. tropicalis, C. krusei.* Additionally, the pathogenic members are known to be genetically more resistant to DNA damaging agents than the non-pathogenic ones (Rodrigues, Silva et al. 2014). *C. albicans* is opportunistic human pathogen. It is a diploid budding yeast belonging to the phylum Ascomycota (Fitzpatrick, Logue et al. 2006). It has a haploid genome size of 14324315 bp. It belongs to the CTG clade, where instead of leucine, the CTG codon shows a biased translation to serine. Despite having superficial similarities to *S. cerevisiae*, the life cycle and genome organization of *C. albicans* reflects several unique features such as regional epigenetically regulated centromeres with unique DNA sequence on each of its eight chromosomes (Sanyal, Baum et al. 2004, Baum, Sanyal et al. 2006), unusually long telomere repeats (McEachern and Hicks 1993) and evidence of a parasexual cycle with a program of concerted chromosome loss (Bennett and Johnson 2003).

C. albicans is a polymorphic fungus existing in yeast, hyphal, pseudo-hyphal and chlamydospore forms. Being an obligate diploid, it reproduces primarily through asexual means of budding. However, it possesses two mating type alleles in its genome, *MTL***a** and *MTL*α on chromosome 5 (Hull and Johnson 1999). Diploid cells of opposite mating types mate to give rise to tetrapoloids, which unlike conventional meiosis do not reduce the ploidy state, instead undergo a loss of multiple chromosomes or concerted chromosome loss (cCL). The parasexual cycle exists until the organism possesses a near diploid content of chromosomes.

Replication origins and ARSs

Identification of ARSs in *C. albicans* provided the first glimpse of a DNA replication origin in this organism several decades ago. A genomic DNA library of *C. albicans* was constructed in an *ADE2* plasmid of *S. cerevisiae* (Kurtz, Cortelyou et al. 1987). Upon transformation in an adenine auxotroph of *C. albicans*, a thousand transformants were obtained which included stable integrants as well. However, after eight generations in non-
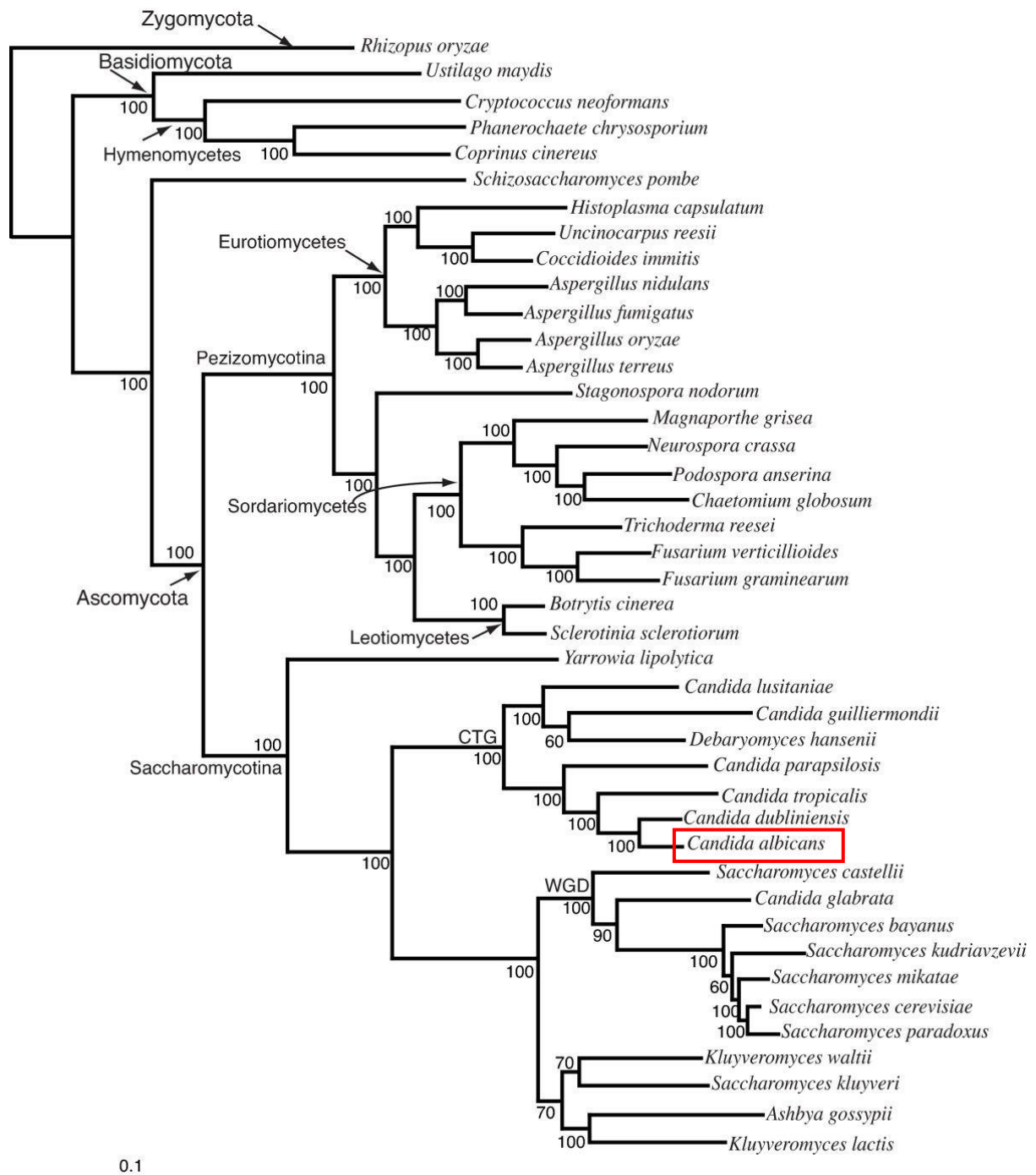
Figure 1.11 A phylogenetic tree showing the position of *C. albicans* (Fitzpatrick, Logue et al. 2006). The tree was constructed using the maximum likelihood method from the alignment of 153 universally distributed fungal genes.

selective media, only 1-2% of the progeny were still adenine prototroph (*ADE +*), indicating that ARS plasmids in this organism have very low mitotic stability. In one of the earlier

studies, an 8.6 kb fragment was isolated from *C. albicans* and a part of it was subcloned in a bacterial plasmid containing *CaLEU2* and *CaURA3* genes (Cannon, Jenkinson et al. 1990). These plasmids could yield a transformation efficiency of 2.15×10^3 CFU/μg and 1.91×10^3 CFU/μg in *S. cerevisiae* and *C. albicans,* respectively. Upon further sequence analysis, these ARS components were shown to be similar to *S. cerevisiae* ARS elements. In another report, a 15.3 kb fragment was cloned and it showed properties of an ARS plasmid. However, these ARS plasmids are subject to random integration and multimer formation in the genome (Herreros, Garcia-Saez et al. 1992).

Genomic origins were identified by the binding of ORC proteins and validated using the approach of 2-dimensional gel electrophoresis. A detailed analysis of the chromosome 5 and chromosome 7 in *C. albicans* revealed a strong correlation between the effects of the CEN proximal origin on CEN function (Mitra, Gomez-Raja et al. 2014). CEN proximal origins contributed to replication fork stalling at the CEN and enhanced their replication timing. This fork stalling or random termination was shown to be kinetochore- mediated. Upon deletion of these CEN proximal origins, the CEN function was compromised and occupancy of CENPA over the central core of CEN7 was seen to be ablated. In *C. albicans,* the replication timing of centromeres are advanced by the early firing of replication origins associated with it (Koren, Tsai et al. 2010). This holds true for non-native centromere or neocentromere formation as well, which replicate in concordance with the proximal early firing origins (Koren, Tsai et al. 2010, Mitra, Gomez-Raja et al. 2014). In another study, genome-wide ORC binding and nucleosome positioning profile of *C. albicans* revealed the presence of categorically distinct "epigenetic" *CEN* proximal origins and "hard-wired" sequence-dependent arm origins (Tsai, Baller et al. 2014).

Telomeres

C. albicans is an attractive model to study telomere biology because of the following reasons. Unlike other species, it contains unusually long (23 bp long) distinct and regular telomere repeat unit. Also, the overall length of telomeres can be varied depending on the growth conditions provided to the organism (McEachern and Hicks 1993). There exists a definite interplay between telomerase activity and telomere recombination for telomere maintenance. In wild- type *Candida* cells, it has been difficult to detect a senescent phenotype owing to heterogeneity in the sizes of Terminal Restriction Fragments (TRFs) of a single telomere studied over a period of time (Singh, Steinberg-Neifach et al. 2002). Telomeric

ORFs (TLOs) are a family of telomere associated ORFs in *C. albicans* and *C. dubliniensis* that encode a subunit of the Mediator complex which is used for the recruitment of RNA pol II during transcription initiation (Anderson, Baller et al. 2012, Sullivan, Berman et al. 2015). Telomere proximal genes exhibit higher noise levels largely due to intrinsic noise that is dependent on genome position, or telomere-adjacent gene expression noise (TAGEN) (Anderson, Gerstein et al. 2014). TAGEN generates expression variability due to local chromatin-mediated gene silencing.

Centromeres and neocentromeres

The small regional centromeres of *C. albicans* were identified by binding of the CENPA homolog, Cse4 that was associated with a 3-5 kb unique DNA sequence on every chromosome (Figure 1.12). Similarly, CEN sequences on all eight chromosomes were identified in the closely related yeast, *C. dubliniensis* (Padmanabhan, Thakur et al. 2008). Inability of a circular ARS plasmid carrying the CENPA-rich CEN region to produce a stable minichromosome in *Candida* cells reinstates that DNA sequence is not the only determinant of the CEN identity (Baum, Sanyal et al. 2006). Also, the centromere DNA sequence as such when integrated at a non-native locus failed to recruit CENPA indicating the little role played by DNA sequence to define centromere identity in this organism. Interestingly, all the eight CENs in these two organisms lack any conserved sequence including any motifs or repeats (Baum, Sanyal et al. 2006).

Upon deletion of a native centromere (*CEN5*) in *C. albicans*, neocentromeres were efficiently activated at various locations on the same chromosome. They could be proximal neocentromeres, formed close to the location of the native CEN or distal neocentromeres which are formed at other locations on the chromosomes (Ketel, Wang et al. 2009, Thakur and Sanyal 2013). A more comprehensive study in multiple chromosomes of *C. albicans* and additionally involving the closely related *C. dubliniensis* suggested that neocentromere formation is a conserved mechanism in these organisms and occurs in CEN-proximal regions (Thakur and Sanyal 2013). *C. albicans* has a combination of neocentromere properties observed in various plants, animals and fungal species asserting the conservation of such mechanisms across species (Marshall, Chueh et al. 2008). Additionally, the central core exhibits flexible positioning of CENPA nucleosomes that gets manifested when a transgene undergoes reversible transcriptional silencing (Thakur and Sanyal 2013), reminiscent of *S. pombe* centromeric chromatin (Allshire, Javerzat et al. 1994). Such peculiar features make
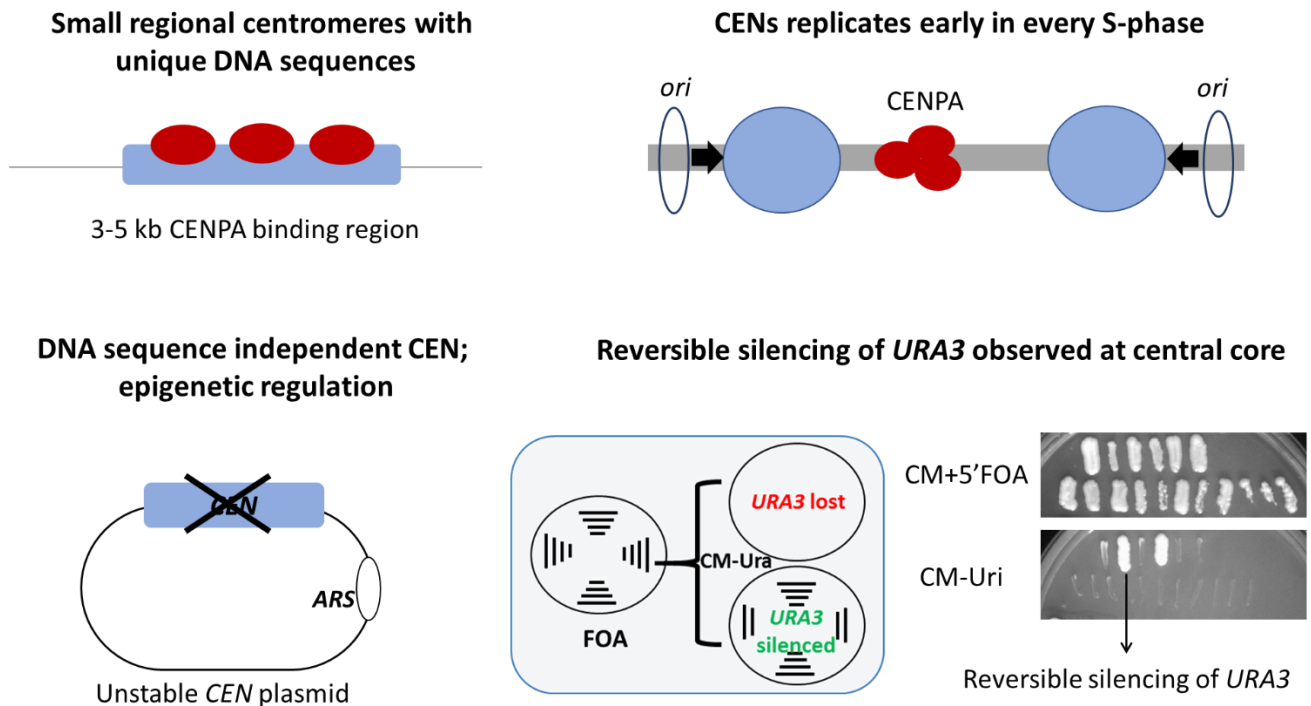
Figure 1.12. Defining characteristics of the centromeres in *C. albicans*. The small regional centromeres in *C. albicans* comprise of a 3-5 kb CENPA binding region. They do not yield a mitotically stable CEN-ARS plasmid stating that they are strictly epigenetic in regulation. They are flanked by early firing replication origins which make them the earliest replicating regions in every S phase. Centromeric chromatin exhibits reversible silencing of a transgene upon integration.

this organism an interesting candidate to study centromere biology (Figure 1.12, Figure 1.13).

Centromere associated proteins

Cse4, the yeast homolog of CENPA, was identified and localised as an intense dot-like signal, co-localizing with the nucleus in *C. albicans* (Sanyal and Carbon 2002). Similar to other organisms, Cse4 is an essential protein and is involved in kinetochore formation in *C. albicans*. Similarly, Mtw1, a homolog of human Mis12 in *C. albicans,* was characterized for its function in the process of kinetochore-microtubule-mediated chromosome segregation in *C. albicans* (Roy, Burrack et al. 2011). The essential DamI complex was shown to be indispensable for chromosome segregation in *C. albicans*, which operates via one kinetochore per microtubule interaction (Thakur and Sanyal 2011).

Recruitment of CENPA at the CEN in *C. albicans* is mediated by homologous recombination (HR) proteins, where the replication forks coming from CEN proximal origins stall at the
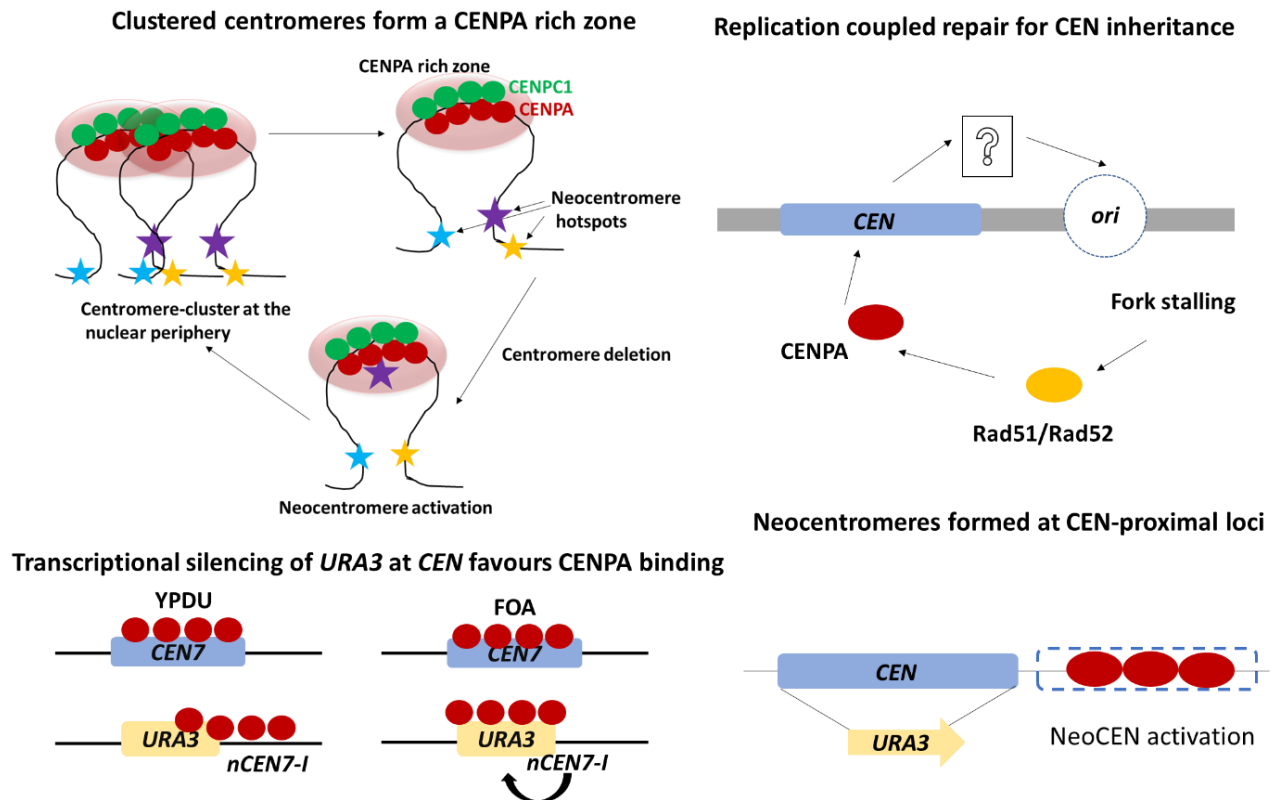
Figure 1.13 Mechanism of centromere establishment and maintenance in *C. albicans*. Unlike in metazoan centromeres, kinetochore complexes in *C. albicans* are clustered. This creates a CENPA rich zone at *CEN* proximal regions that facilitates the activation of neocentromeres at these hotspots. Depending on the transcriptional status of the transgene *URA3*, *CEN* chromatin can be assembled on it, wherein repressive chromatin environment favours CENPA binding. These epigenetic CENs once established are maintained by a replication coupled repair mediated CENPA deposition post replication of chromatin. This is regulated by HR proteins like Rad51/ Rad52 that physically interact with CENPA.

CEN in a kinetochore-dependent manner (Mitra, Gomez-Raja et al. 2014). Fork stalling at the CEN is reduced in the absence of HR proteins Rad51 or Rad52 (Figure 1.13). Null *rad51* or *rad52* mutant cells exhibit an increased kinetochore declustering and degradation of CENPA. The physical association of CENPA and Rad51/Rad52 in a complex is an indicator of a homologous recombination-mediated CENPA recruitment mechanism in this organism (Mitra, Gomez-Raja et al. 2014). This emerging role of HR proteins at the centromere raises an interesting possibility of the involvement of HJURP/Scm3 at the CEN. Additionally, outer kinetochore proteins influencing the localisation of CENPA suggests that the kinetochore sub-complex assembles in a unique interdependent concerted manner to form a stable kinetochore. Most strikingly, the kinetochore protects CENPA from proteasomal degradation in *C. albicans* (Thakur and Sanyal 2012). In *C. albicans,* where loading of CENPA occurs in

the late anaphase (Shivaraju, Unruh et al. 2012), it has been observed that non-chromatin associated CENPA or free CENPA becomes a target of the ubiquitin ligase pathway for proteasomal degradation.

Rationale of the present study

The establishment and maintenance of centromeric chromatin is largely regulated by epigenetic factors. However, these factors are not fully defined in several fungal species. The current study focuses on identifying factors acting in *cis* and *trans* with centromeric chromatin that not only favour establishment of centromeric chromatin to a particular locus on a chromosome, but also enable it to be faithfully maintained across generations. The organism chosen for the study is the budding pathogenic yeast, *C. albicans* with epigenetically regulated small regional centromeres that are early replicating. We attempted to decipher the "epigenetic" factors for centromere specification and maintenance using this elegant model system. Since *C. albicans* is known to have early replicating centromeres, we examined the crosstalk of replication machinery with centromeres and its associated components.

Summary of the present work

In this study, we aimed to examine various "epigenetic" factors that define centromere identity and maintenance in *C. albicans*. First, we identified genome wide replication origins in *C. albicans* in order to explore the replication architecture of the genome. Further, we examined the nature of centromere adjacent chromatin to gain novel insights into neocentromere activation in this pathogenic species. We then discovered a previously unknown role of replication initiation proteins in maintaining centromere function.

Replication origins in *C. albicans* have been poorly mapped with the exception of the discovery of a few ARS elements in the genome. In this study, we utilized the binding of a subunit of ORC, Orc4 to map replication origins by ChIP-sequencing. The binding sites identified, shared no common defining DNA sequence but were prevalent in the nucleosome-depleted regions of the genome. We did find tRNA gene (tDNA) motifs in a fraction of the identified origins. Analysis of the replicating timing of these origins revealed the tDNA

associated origins to be located at early replicating regions. Hi-C analysis revealed the clustering of early replicating regions with each other, showing minimal interactions with the later replicating regions.

Centromeres in *C. albicans* show authentic epigenetic regulation as there is no dependence on the underlying DNA sequence to dictate centromere formation. It is also known that neocentromere hotspots lie in the immediate vicinity of the native centromere location. So far, CENPA association is the only factor known to specify the centromere identity. Using Hi-C analysis and transgene silencing assay, we identify a pericentromeric heterochromatin spanning the CENPA binding region. The transient ectopic kinetochore formed is stabilized only upon the native centromere sequence deletion. The preferential formation of a neocentromere at a site that has already been primed with CENPA indicates that the number of molecules of CENPA at a CEN proximal region determines the site of kinetochore assembly. We allude that CENPA priming of a non-centromeric region can initiate centromere assembly in *C. albicans.*

Replication origins flanking the *C. albicans* centromeres exert a strong influence on replication timing of centromeres. Our ChIP sequencing analysis of Orc4 binding revealed the strong association of Orc4 at all *C. albicans* centromere, the occupancy of which was strikingly similar to CENPA. Orc4 was also bound to ectopic centromeres and neocentromeres, stating that Orc4 is an active component of *C. albicans* centromeres. However, the nature of interaction of CENPA and Orc4 remains elusive. Microscopic examination and standard ChIP- qPCR analysis revealed that Orc4 depletion causes kinetochore mis-segregation and eventually leads to declustered centromeres, a phenotype that we observed upon depletion of a helicase subunit, Mcm2 in *C. albicans.* These observations allude to a previously unknown role of ORC and MCM complexes (or distinct sub-complexes) in centromere function and CENPA stabilization. We discuss the implication of these results in the light of centromere establishment in *C. albicans.*

RESULTS

2. Identification of genome wide replication origins in *Candida albicans*

Identification of genome wide replication origins in *Candida albicans*

DNA replication origins are cardinal genomic loci that initiate and regulate the process of genome duplication once in every cell cycle. Hence, identification of the precise location of replication origins, which remains elusive in many fungal species, aids in understanding the genome organization of an organism. The pre-RC proteins identify replication origins and depending on their activation status, enable firing of origins. Hence, these proteins can be utilized to study the origin distribution in the genome, which comprises of both firing and non-firing origins.

ORCs are bound to potential origin loci throughout the *S. cerevisiae* cell cycle (Bell and Dutta 2002, Mendez and Stillman 2003). ORCs, initially identified as ARS binding proteins in *S. cerevisiae* (Marahrens and Stillman 1992), not only show broad conservation in core components and subunit organization across species but also display organism specific alterations that generate functional differences (Gavin, Hidaka et al. 1995, Gossen, Pak et al. 1995, Muzi-Falconi and Kelly 1995, Carpenter, Mueller et al. 1996, Grallert and Nurse 1996, Romanowski, Madine et al. 1996). Although ORC proteins have been studied well in organisms like *S. cerevisiae* (Lee and Bell 1997) and *S. pombe* (Chuang and Kelly 1999), its homolog in *C. albicans* has remained largely uncharacterized. To address this objective, we tried to identify and characterize the *C. albicans* homolog of one of the conserved pre-RC proteins, Orc4. Orc4 binds directly to the origin in *S. pombe* (Chuang and Kelly 1999) and similar to Orc1 and Orc2, binds to the major groove of the ACS DNA (Lee and Bell 1997). In this study, our broader objective to independently identify the precise location and distribution of replication origins in the genome of *C. albicans,* was initiated by examining the occupancy of CaOrc4, the Orc4 subunit of the conserved ORC in *C. albicans*. Distribution of this pre-RC protein gives a representative picture of genome-wide replication origin map which would comprise of both the active (firing) origins and passive (dormant) origins. We then examined the spatiotemporal features of Orc4 distribution across the *C. albicans* which revealed intriguing facets about the replication program of this pathogenic yeast.

Table 2.1. Features of replication origins in eukaryotic systems

Organism	Genome size (Median)	No. of replication origins/ARSs	Features of origins	Reference
S. cerevisiae	12.1227 Mb	~400 origins, 12000 ACS sites	Sequence dependent, AT rich, NFR-associated	(Nieduszynski, Knox et al. 2006, Xu, Aparicio et al. 2006)
S. pombe	12.6078 Mb	~400 origins	AT-rich, 0.5-1 kb long, NFR-associated	(Heichinger, Penkett et al. 2006)
K. lactis	10.7061 Mb	156 ARSs	50 bp ACS, found at intergenes	(Liachko, Bhaskar et al. 2010)
C. glabrata	12.6225 Mb	253 origins, 275 ARSs	17 bp AT-rich ACS; early Replicating clustered origins	(Descorps-Declere, Saguez et al. 2015)
P. pastoris	9.35815 Mb	311 ARSs	AT rich and GC rich ARSs; presence of a 20 bp GC-ACS, NFR-associated	(Liachko, Youngblood et al. 2014)
D. melanogaster	137.688 Mb	7,000-8,000 replication start sites (RSS)	Cell type specific firing, G4 proximal, RSS have high nucleosome occupancy	(Cayrou, Coulombe et al. 2011, Comoglio, Schlumpf et al. 2015)
X. laevis	2718.43 Mb	5-10 origins per cluster	AT rich asymmetric sequences	(Blow, Gillespie et al. 2001, Stanojcic, Lemaitre et al. 2008)
H. sapiens	2992.79 Mb	~13,000 Orc1 binding sites	Proximal to TSSs of coding and non-coding genes	(Dellino, Cittaro et al. 2013)

Orc4 is a conserved nuclear protein in *C. albicans*

Orc4 in *C. albicans* is a 564-aa long protein (https://doi.org/10.1101/430892) that contains the AAA+ domain which belongs to AAA+ family of ATPases (Walker, Saraste et al. 1982) associated with a variety of cellular activities (Figure 2.1A). This domain has been shown to be pivotal to the initiation of eukaryotic DNA replication and consists of the conserved Walker A, Walker B and the R (arginine) ring finger motifs (Walker, Saraste et al. 1982).
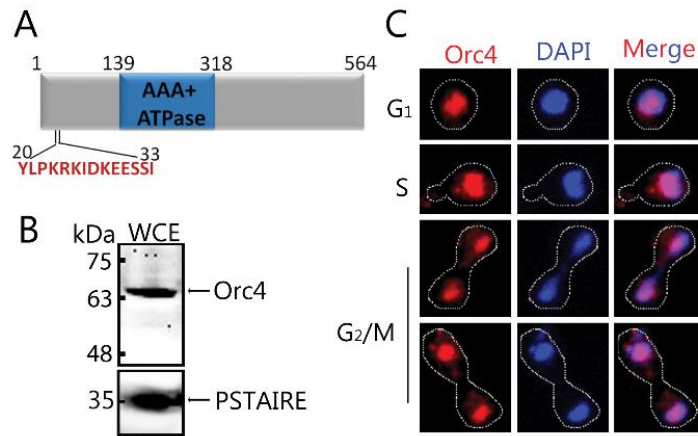
Figure 2.1. Expression and *in vivo* localisation of Orc4 in *C. albicans*. (**A**) Domain architecture of CaOrc4 reveals a 564 aa long polypeptide consisting of a central AAA+ ATPase domain. The peptide sequence chosen to raise the antibodies has been highlighted in red letters (residues 20-33). (**B**) Expression of Orc4 was verified by western blot with anti-CaOrc4 antibodies using whole cell extract (WCE) from *C. albicans* SC5314. Orc4 yielded a band at the expected molecular weight at ~64 kDa in a denaturing SDS PAGE. PSTAIRE was used as the loading control. (**C**) Wild-type SC5314 cells were fixed and stained with anti-Orc4 antibodies (red) and DAPI (blue) to study the intracellular localisation of Orc4 in *C. albicans*. Merged DAPI (blue) and CaOrc4 (red) images indicates CaOrc4 is localised to the nucleus throughout the cell cycle in *C. albicans*. Bar, 5 µm.

We raised polyclonal antibodies against a peptide sequence from the N-terminus of CaOrc4 (aa 20-33) (Figure 2.1A). Western blot with the whole cell extract of *C. albicans* SC5314 (*ORC4/ORC4*) yielded a strong specific band at the expected molecular weight of approximately 64 kDa when probed with anti-Orc4 antibodies (Figure 2.1B). PSTAIRE was used as the loading control. We performed indirect immuno-fluorescence microscopy using anti-CaOrc4 antibodies to examine cellular localisation of CaOrc4. CaOrc4 was found to be strictly localised to the nucleus at all stages of the *C. albicans* cell cycle (Figure 2.1C), a feature of the ORC proteins found to be conserved in *S. cerevisiae* as well (Dutta and Bell 1997).

Orc4 is essential for viability in *C. albicans*

Orc4 is an evolutionarily conserved essential subunit of the ORC complex across eukaryotes (Chuang and Kelly 1999, Dai, Chuang et al. 2005). In order to determine the essentiality of this gene in *C. albicans*, a diploid organism, a conditional mutant of *ORC4* was constructed by deleting one allele and replacing the endogenous promoter of the remaining *ORC4* allele with the repressive *MET3* promoter (Figure 2.2A) (Care, Trevethick et al. 1999).
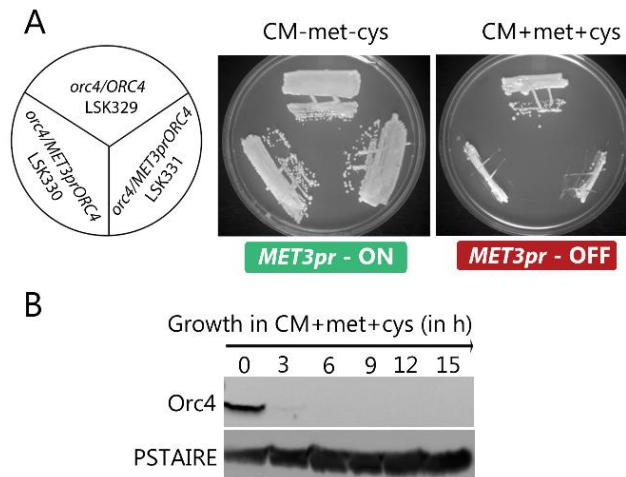
Figure 2.2. **Orc4 is essential for viability in** *C. albicans. (***A**) The promoter of *MET3*, is expressed in the absence of methionine and cysteine and repressed in the presence of both, was used for the controlled expression of *ORC4*. *C. albicans* cells where one deleted copy of *ORC4,* LSK329 (*ORC4/ORC4::FRT*), and two independent transformants of the conditional mutant (*ORC4::FRT/ MET3prORC4)*, LSK330 and LSK331 where the remaining wild-type copy was placed under the control of the *MET3* promoter were streaked on plates containing inducible (CM-met-cys) and repressible (CM, 5 mM cysteine and 5 mM methionine). Plate photographs were taken after 48 h of incubation at 30°C. (**B**) Western blot analysis using anti-Orc4 antibodies indicates time course depletion of Orc4 in the conditional mutant LSK330 when the strain was grown for the indicated time (0, 3h, 6h, 9h, 12h, 15h) in presence of 5mM methionine and 5mM cysteine. PSTAIRE was used as the loading control.

Inability of two independent transformants of the conditional mutant (*ORC4::FRT/MET3prORC4)*, LSK330 and LSK331 to grow in the presence of CM supplemented with 5mM methionine and 5mM cysteine (Figure 2.2A) indicates that Orc4 is essential for viability in *C. albicans*. The heterozygous null mutant of *ORC4*, LSK329 (*ORC4/ORC4::FRT*) grown under similar conditions did not show any growth defects. We confirmed the depletion of Orc4 protein levels from the cellular pool by performing a western blot analysis in the Orc4-depleted versus expressed conditions (Figure 2.2B). Following 3 h of repressive growth, we could not detect Orc4. Subsequently, we used antibodies raised against CaOrc4 as a tool to map its binding sites in the *C. albicans* genome.

Orc4 binds to discrete genomic loci in *C. albicans*

Binding regions of subunits of the pre-replication complex proteins has been widely used as a standard method of identifying genome-wide replication origins by ChIP-chip or ChIP-seq
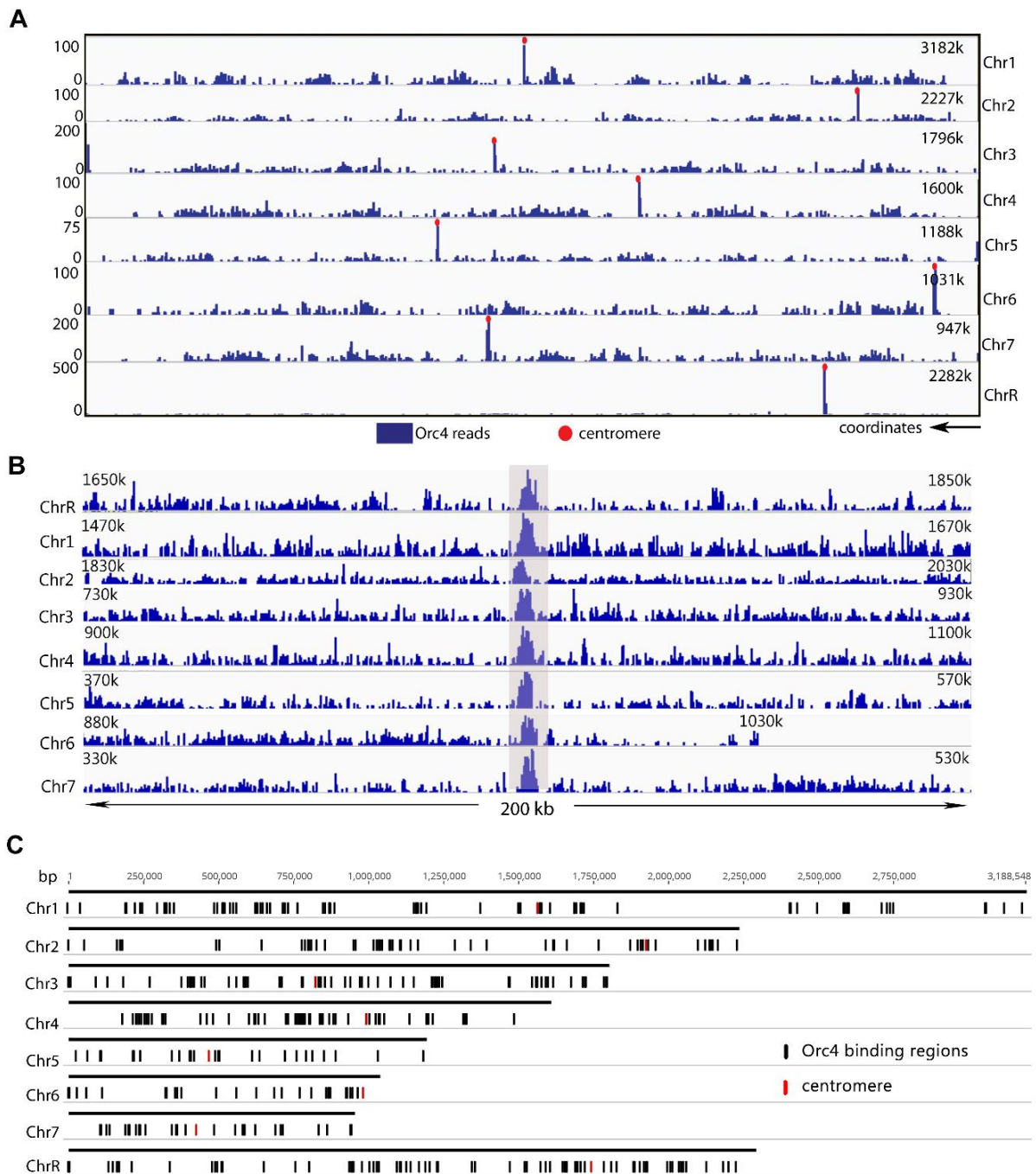
Figure 2.3. ChIP-sequencing analysis showing Orc4 binding across all chromosomes in *C. albicans*. (**A**) The relative number of Orc4 ChIP-sequencing reads obtained from the whole cell lysate from the Orc4 ChIP sequence reads aligned to the reference genome *C. albicans* SC5314 Assembly 21. (**B**) A 200 kb zoomed in region of all eight chromosomes in *C. albicans* with individual centromeres (grey shaded rectangle). Chromosome coordinates are shown on the *x*-axis and the track height for the aligned Orc4 (**C**) A genome-wide view of the Orc4 binding regions (black) along the length of all eight chromosomes in *C. albicans*.
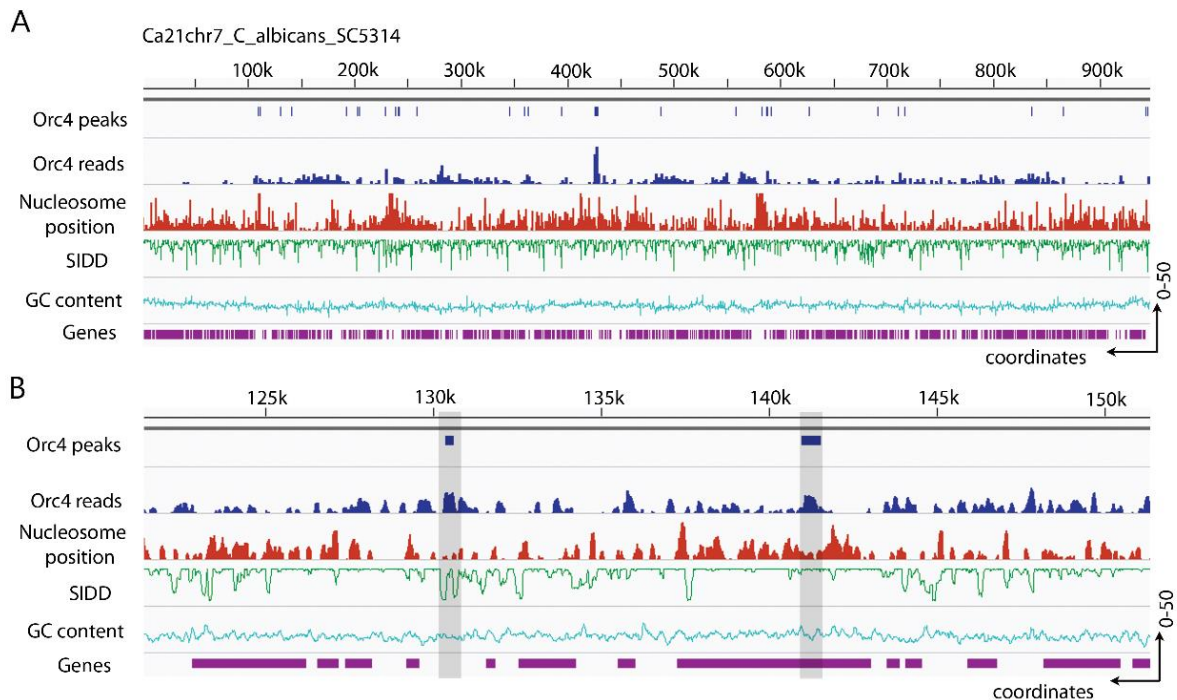
Figure 2.4. Orc4 binding regions are associated with nucleosome free regions (NFRs) and SIDD valleys in *C. albicans***.** (**A**) A representative snapshot of chromosome 7 displaying the whole chromosome distribution of Orc4 with the indicated peak positions (blue rectangle) and subtracted Orc4 ChIP-sequence reads (blue histogram). The nucleosome positioning profile (red histogram) indicates peaks and troughs (NFRs) corresponding to sequence reads obtained from the MNase-seq data (Tsankov, Thompson et al. 2010). Other genomic attributes such as the average GC content (blue) and gene density (purple) are shown below as visualized using IGV. (**B**) A 30 kb zoomed-in view of chromosome 7 highlights two Orc4 binding regions in grey with their corresponding Orc4 reads, nucleosome positioning, SIDD values, GC content, and gene density as visualized in IGV. The *x*-axis indicates the chromosomal coordinates and *y*-axis indicates the data range (see Materials and Methods for details).

---

sequencing assays (Wyrick, Aparicio et al. 2001). We wanted to examine Orc4 binding sites in the genome of *C. albicans* as a way to map potential replication origins in this organism. To do the same, we performed a ChIP sequencing experiment in asynchronously grown cells of *C. albicans* using anti-CaOrc4 antibodies. Our analysis yielded a total of 417 Orc4 binding sites with 414 of these belonging to various genomic loci (Figure 2.3 A,B,C) and 3 of these belonging to mitochondrial DNA. We found all centromeres to be highly enriched with Orc4. The length of CaOrc4 binding regions ranged from 200 bp to 3 kb. It is also to be noted here that ~61% of the Orc4 binding regions in our study were present in genic regions (252/414) deviating from the trend observed in *S. cerevisiae* where most of the chromosomal origins are located at intergenic regions (Xu, Aparicio et al. 2006).
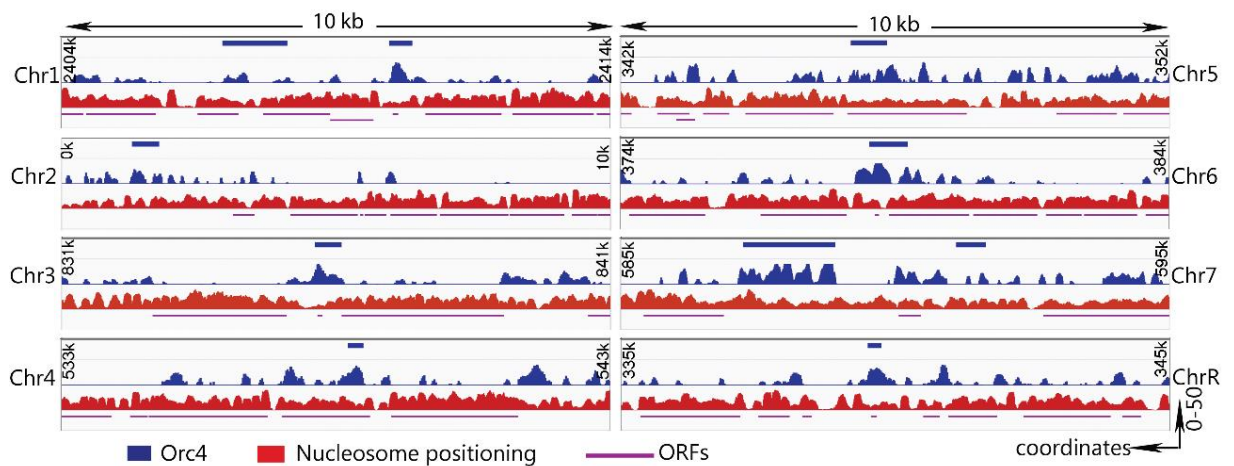
Figure 2.5. Orc4 binding regions lie on nucleosome-depleted areas of the *C. albicans* genome. A 10 kb zoomed-in region of each of the eight chromosomes in *C. albicans* shows the relative number of Orc4 ChIP sequence reads subtracted from the reads obtained from the whole cell lysate (blue histogram) aligned to the reference genome SC5314 Assembly 21. The nucleosome positioning profile (red histogram) indicates peaks and troughs corresponding to sequence reads obtained from the MNase-seq data from (Tsankov, Thompson et al. 2010). Troughs indicate nucleosome-free regions (NFRs). The position of Orc4 peaks (blue line) and genes (purple line) have been indicated in their respective colours. Chromosome coordinates are shown on the *x*-axis while the *y*-axis represents the individual track height for the aligned Orc4 sequence reads (top) and MNase-seq reads (bottom) on each chromosome as visualised using IGV.

Replication origins are found in nucleosome-depleted regions, as it provides easier access for DNA polymerase to open the double helix for the templated synthesis of nascent DNA (Lipford and Bell 2001). Hence, the nucleosome occupancies around replication origins are significantly lower than the flanking regions (Li, Zhong et al. 2014). We wanted to examine if the Orc4 binding sites occurred in nucleosome-free regions (NFRs) of the *C. albicans* genome. Hence, we utilized the nucleosome positioning data in *C. albicans* provided in a previous study (Eaton, Galani et al. 2010) and compared our Orc4 ChIP-sequencing peaks with the nucleosome positioning profile of corresponding regions (Figure 2.4 A, B, Figure 2.5). We also determined the SIDD (Bi and Benham 2004) profile for chromosome 7 and performed a comparison with the Orc4 binding regions on the same. Low SIDD values indicate that the region of a destabilized DNA duplex could be a putative replication origin. Combining *in silico* nucleosome positioning data and SIDD analysis data, we revealed that the Orc4 binding sites were correlated with low SIDD valleys which fall in nucleosome troughs (Figure 2.4A, B). These are typical features of DNA replication origins in most eukaryotic systems studied till date (Lipford and Bell 2001, Segurado, de Luis et al. 2003).
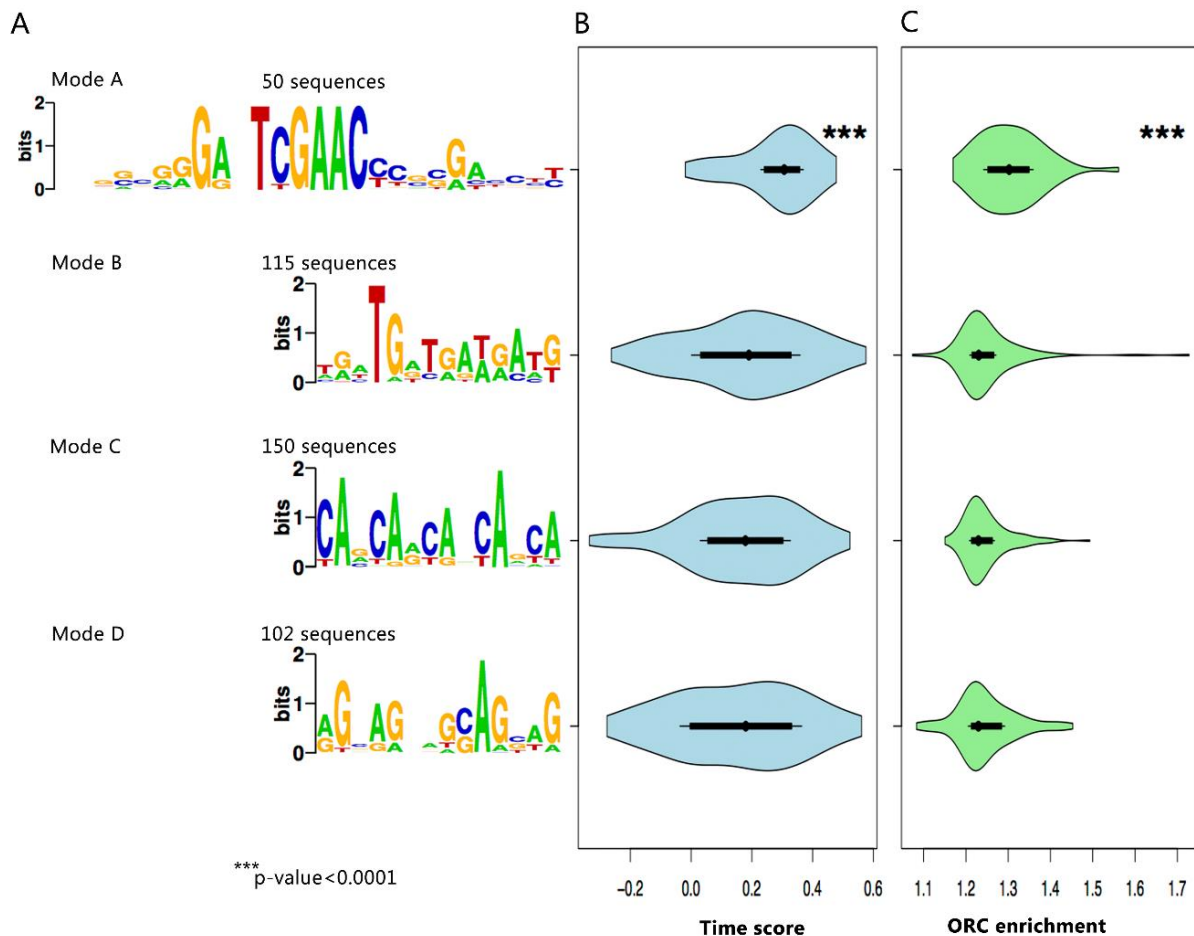
Figure 2.6. DNA motifs identified by DIVERSITY shows various distinct modes corresponding to their timing profile and ORC abundance. (A) The four different modes (motifs) identified by DIVERSITY (A, B, C, D) and their distribution across the 417 binding regions have been listed. Mode A corresponds to the tDNA motif. **(B)** Violin plots depicting the replication timing profile obtained from a previous report (Koren, Tsai et al. 2010) for all four modes shows higher time scores for Mode A associated origins. **(C)** Enrichment of Orc4 shows a significant increase for Mode A associated origins based on the ChIP-sequencing results. The other three modes depict enrichment values not significantly different from each other.

DNA sequence features of ORC binding regions in *C. albicans*

The G+C content of nuclear DNA for ascomycetous yeast species ranges from about 30–50% (Price, Fuson et al. 1978). The mean GC content of replication origins in *S. cerevisiae* is 31%, which is significantly lower than its genomic GC content of 37.9% making origins in this organism AT rich (Li, Zhong et al. 2014). The average GC content of all the DNA sequences covering the Orc4 binding regions in *C. albicans* in our study was calculated to be 34.3%, whereas its average genomic GC content is 33.4%. The number of origins on a

chromosome determines the spacing between initiation events and is a direct function of the chromosome length (Newman, Mamun et al. 2013). From our analysis, we find that the number of Orc4 binding regions per chromosome, or the ORC density is directly proportional to the length of the inter-origin distance (IOD) which is the ratio of chromosome length and number of Orc4 binding sites on the chromosome was calculated as 34.925 kb. A similar trend has been observed in well studied yeast genomes (Table 2.1) (Newman, Mamun et al. 2013). This can be extrapolated to estimate the theoretical inter-origin distance in *C. albicans*.

Conserved DNA sequence features (ACS motifs) are common in the *Saccharomyces* group (Nieduszynski, Knox et al. 2006). We used the *de novo* motif discovery tool DIVERSITY (Mitra, Biswas et al. 2018) on the *C. albicans* Orc4 binding regions. DIVERSITY allows for the fact that the profiled protein may have multiple modes of DNA binding. Here, DIVERSITY reports four binding modes (Figure 2.6A). The first mode, mode A is a strong motif GAnTCGAAC, present in 50 such regions, 49 of which were found to be located within tRNA gene bodies. The other three modes are low complexity motifs, TGATGA (mode B), CAnCAnCAn (mode C) and AGnAG (mode D). Strikingly, each of the 417 binding regions were associated with one of these motifs. Mode C has been identified before (Tsai, Baller et al. 2014). The association to tRNA genes has been demonstrated previously in a subset of *S. cerevisiae* replication origins as well (Wyrick, Aparicio et al. 2001). Taken together, this suggests that ORCs in *C. albicans* do not rely on a specific sequence feature for binding DNA. We sincerely acknowledge Dr. Leelavati Narlikar, NCL, Pune and Dr. Rahul Siddharthan IMSc, Chennai for performing the DIVERSITY analysis.

Replication origins in *C. albicans* are spatiotemporally regulated

Replication origins are spatially distributed and temporally regulated to ensure timely duplication of the genome as well as to avoid re-initiation events. Depending on the time of activation and efficiency, replication origins are classified as early and late domain/factories. To categorize the replication timing of Orc4 binding sites, we utilized the fully processed replication timing profile of *C. albicans* available from a previous study (Koren, Tsai et al. 2010). Upon categorizing the replication profile according to the different modes identified by DIVERSITY (Figure 2.7), we observe a significant advanced replication timing of the tRNA associated motifs (mode A) (Figure 2.6B). The other three modes (B, C, D) display no significant bias towards an early replication score. Moreover, we could correlate early
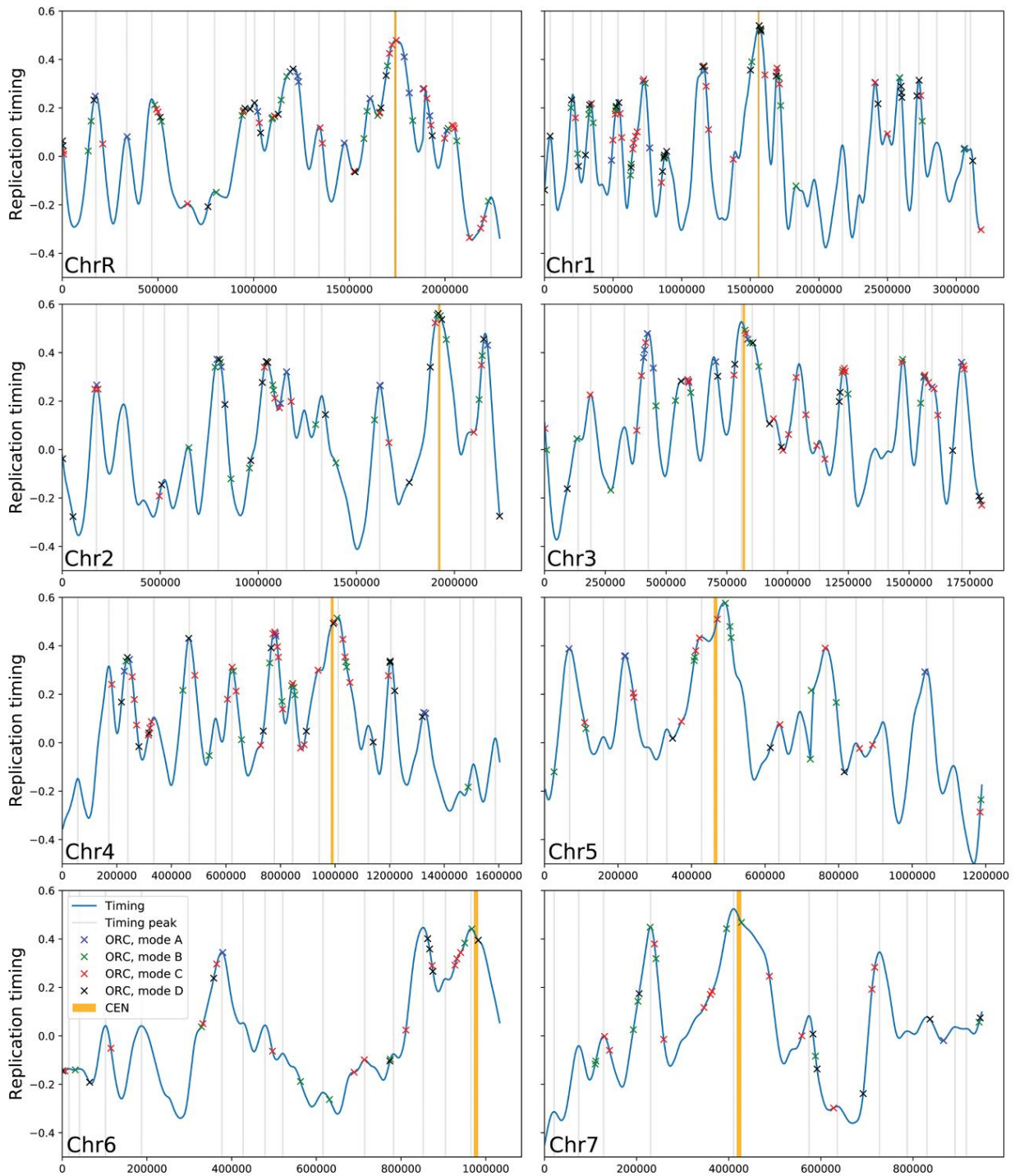
Figure 2.7. Replication timing profile of various modes associated with CaOrc4 binding.
CaOrc4 ChIP-seq peaks were aligned to the replication timing profile obtained from *C. albicans* from a previous report (Koren, Tsai et al. 2010). Color-coded stars indicate each of the four motifs identified by DIVERSITY which covers all the 414 chromosomal origins. Peaks represent early replicating regions, including the centromere (yellow lines). A significant fraction of the modes cluster towards the local maxima of the peaks. The x-axis represents chromosomal coordinates and y-axis shows replication timing scores.
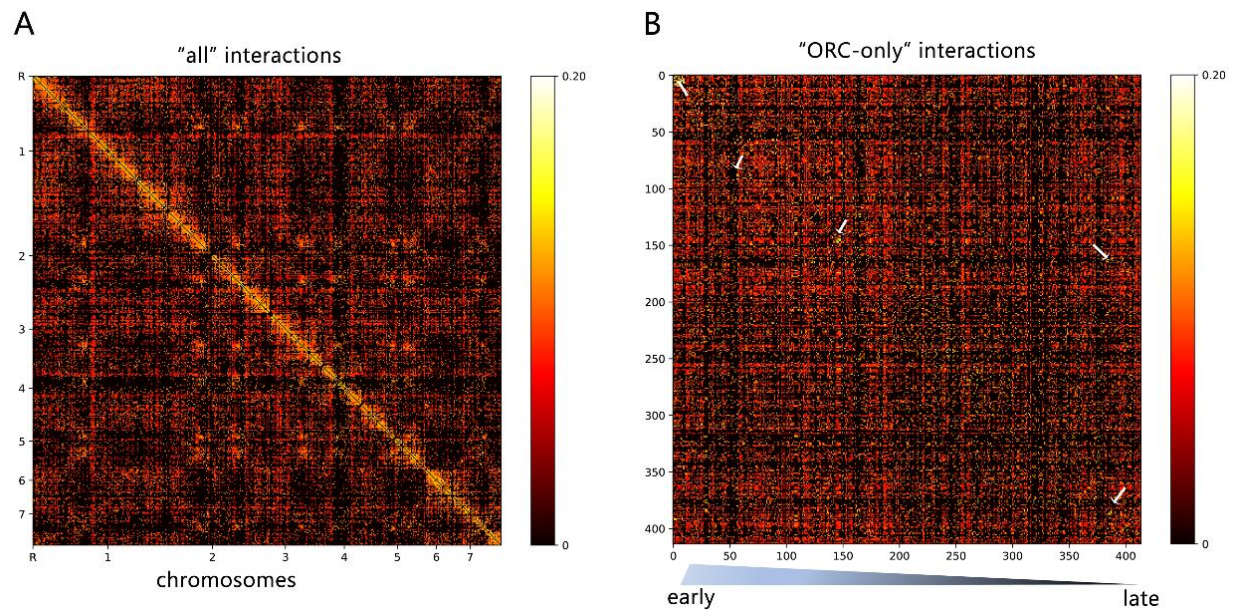
Figure 2.8. Early replicating regions interact among themselves to form clusters/ replication factories. (A) The Hi-C heatmap shows a whole-genome "all" heatmap representation of the Hi-C data (Burrack, Hutton et al. 2016) as a 7145x7145 matrix. The maximum value in the data was 0.2015 and the minimum was zero. For plotting, the values were log-transformed with a pseudocount of 0.0001. **(B)** The Hi-C "ORC-only" heatmap shows interactions between the 414 chromosomal ORC binding regions, ordered by timing (early to late), to the same colour scale as in (A). White arrows directing towards the yellow pixels indicate clustered/ strongly interacting origins. The analysis was performed at a resolution of 2 kb.

replication timing with an increased enrichment of Orc4 in these regions (Figure 2.6C). Additionally, all the motifs were located towards the local maxima of the timing peaks (Figure 2.7), suggestive that most of these regions are indeed chromosomal origins.

To locate these regions within the nuclear space, we mapped the interactions made by ORC binding regions with each other using the Hi-C data from a previous study in *C. albicans* (Burrack, Hutton et al. 2016). All the ORC binding regions were aligned with increasing order of their replication timing (early to late) and subsequent interactions were mapped (Figure 2.8B). Similar analysis was performed for the whole genome of *C. albicans* (Figure 2.8A). We observe that the overall "only-ORC" interactions are higher than the whole-genome "all" interactions, suggesting that ORC binding regions interact more than the average. Early replicating regions show a significantly higher interaction among themselves (Figure 2.9A), in agreement with previous observations in *C. glabrata* (Descorps-Declere, Saguez et al. 2015). Given that regions in this heatmap are ordered by timing and not genomic proximity, this suggests that regions with a similar timing in replication tend to
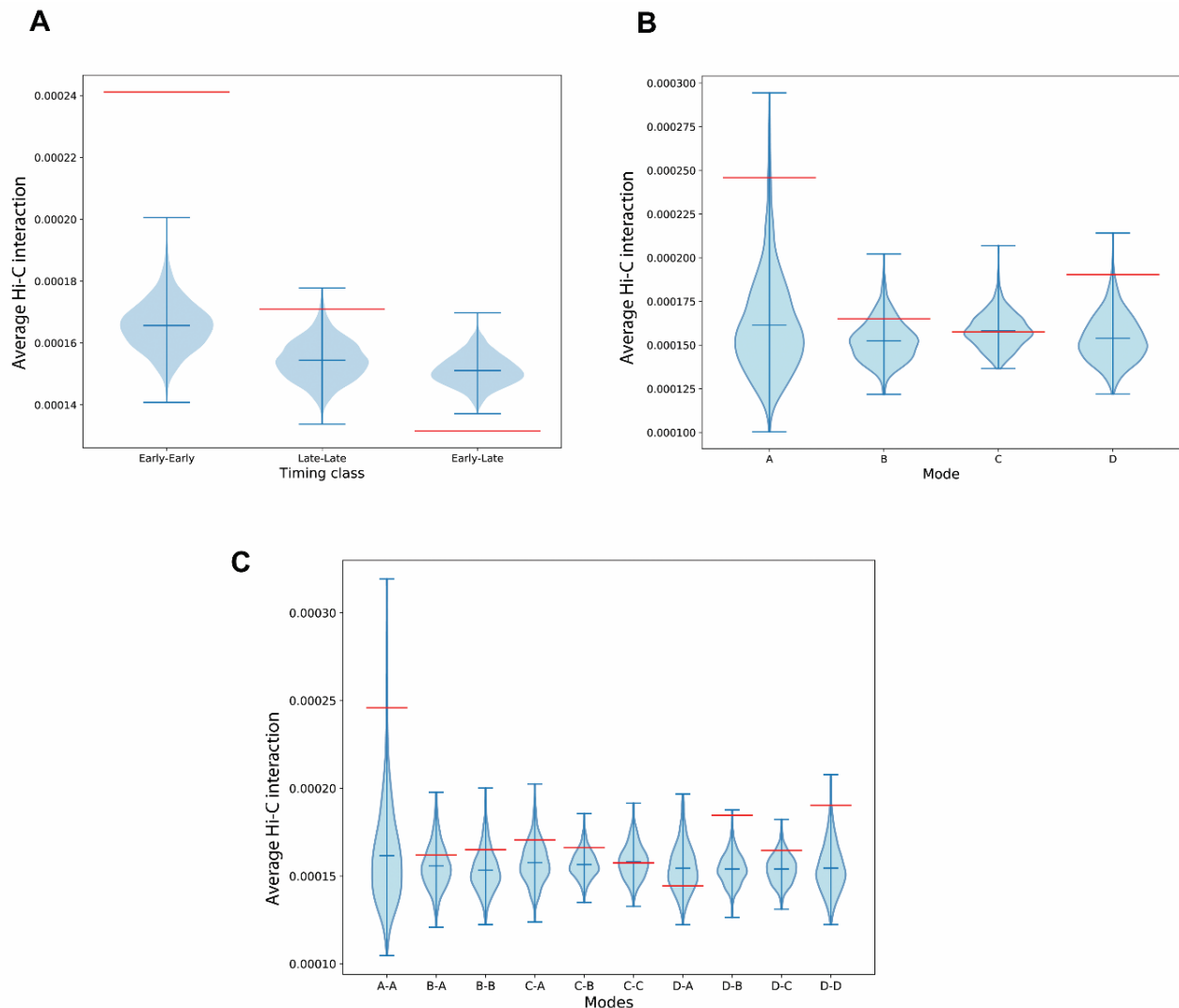
Figure 2.9. Spatiotemporal organization of *C. albicans* replication origins. (**A**) Violin plot depicts the average Hi-C interaction (red lines) of the early-early, late-late and early-late regions that have been classified based on their replication timing profile. Red horizontal lines indicate higher interaction values among early replicating domains, as obtained by the Hi-C experiment. The blue violins indicate mean interactions across 1000 randomizations. (**B**) Violin plot indicating average Hi-C interaction of all the four DIVERSITY modes (sequence motifs) to each other. Mode A (tRNA associated) shows the maximum interaction frequency among the 4 modes. (C) Hi-C mean interactions studied across all four modes has been depicted here. Red line in the violin plots indicates average Hi-C interactions (y-axis) across various combinations of modes (x-axis). Note that these are significantly higher for A-A interaction than other mode-mode combinations. The blue violins are mean interactions across 1000 randomizations.

associate together. Hi-C analysis also revealed that mode A containing origins, that show an early replication timing, form stronger interactions among themselves than all the other modes (Figure 2.9B, C). Hence, this alludes to the fact that the ORC binding regions identified in our study are the chromosomal origins in *C. albicans* as they associate with

categorically distinct domains separated in space and time of replication, facilitating origin function and usage. We sincerely acknowledge Dr. Rahul Siddharthan, IMSc, Chennai for replication timing analysis and Hi-C analysis.

3. Epigenetic factors of centromere formation of *C. albicans*

Epigenetic factors of centromere formation in *C. albicans*

Non-repetitive centromeres provide an excellent model to study characterization of centromeric chromatin. Moreover, the lack of a primary DNA sequence to specify centromeric location sets the premise to study epigenetics factors in controlling centromere formation and maintenance. In the ascomycetous budding yeast *Candida albicans,* the presence of unique CEN sequences on every chromosome (Sanyal, Baum et al. 2004) and the activation of neocentromeres at pre-determined hotspots proximal to the native centromere location (Thakur and Sanyal 2013) together reveal that the underlying DNA sequence is neither necessary nor sufficient for centromere formation (Sanyal, Baum et al. 2004, Baum, Sanyal et al. 2006, Thakur and Sanyal 2013). CENPA localisation on a transgene under selective conditions is known to correspond to its transcriptional status. Similar to *S. pombe* (Allshire, Javerzat et al. 1994), in *C. albicans* reversible silencing of the expression of a marker gene, *URA3,* captured by 5'FOA counter-selection, has been observed upon its integration at the CENPA binding region of the centromere endowing it a transcriptionally flexible status (Thakur and Sanyal 2013). Centromeres in *C. albicans* comprise regions of CENP-A occupied DNA that span 3-5 kb in length (Baum, Sanyal et al. 2006). The presence of replication origins and neocentromere hotspots within 30 kb of centromere 7, *CEN7* indicates that *CEN* proximal regions (pericentromeric regions) are important hubs that regulate centromere activity (Thakur and Sanyal 2013, Mitra, Gomez-Raja et al. 2014). Pericentromeric heterochromatin are the gatekeepers for CEN chromatin. Pericentromeres are devoid of kinetochore proteins but are enriched with heterochromatin marks. Territorial expansion of CENPA to euchromatic sites is known to have a notable repressive effect on gene expression of underlying regions. In corollary, the CENPA localisation status on a marker gene is reminiscent of its transcriptional status. We have limited knowledge regarding the boundaries confining CENPA chromatin to a 5 kb region. In this study, we wanted to define pericentromeric regions and examine the status of this pericentromeric heterochromatin. Using a combination of Hi-C analysis and transgene integration assay, we define the essential "epigenetic" elements that are required to form and maintain the centromeric chromatin in *C. albicans.* We also identify a novel factor that plays a determinant in the site of neocentromere formation in *C. albicans.*

Core CENPA-rich regions in *C. albicans* are flanked by a ~25 kb long unusual pericentromeric heterochromatin

The centromere DNA spans a region of 3-5 kb in *C. albicans* (Sanyal and Carbon 2002, Sanyal, Baum et al. 2004) bound by the CENPA homolog, Cse4 (Sanyal and Carbon 2002). In chromosome 7 (Chr7), chromosome 5 (Chr5) and chromosome 1 (Chr1) of *C. albicans,* neocentromere hotspots have been identified at centromere proximal regions (Thakur and Sanyal 2013). Additionally, replication origins are situated close to the CENPA binding regions in this organism making centromeres early replicating. (Thakur and Sanyal 2013, Mitra, Gomez-Raja et al. 2014). These features provide ample evidence that centromere proximal regions are important hubs that regulate centromere activity. To decipher the nature of interaction of centromeres with respect to the whole genome at the 3D level, we analyzed the Hi-C data of *C. albicans* from a previous report (Burrack, Hutton et al. 2016) and examined the inter-chromosomal (*trans*) and intra-chromosomal interactions (*cis*) interactions. The resolution of the Hi-C experiment was 2 kb. Our analysis revealed that each of the eight centromeres interacted with the adjacent "pericentromeric" regions at a contact probability higher than regions distal from the centromere (Figure 3.1A) making centromere-pericentromere *cis* interactions stronger than average chromosomal interactions. Also, the clustered centromeres of *C. albicans* interact in *trans* (with other centromeres) at a higher probability forming a compact chromatin environment than the average genome interaction found in bulk chromatin (Figure. 3.1B). Chr7 is the shortest and the most well studied chromosome in *C. albicans* making it the subject of our study. We examined all *cis* (centromeric, centromeric proximal and centromere distal) interactions on Chr7. The heatmap of Chr7 (Figure 3.1C left) indicates stronger interactions close to centromeres ("pericentromeric") (Figure 3.1D middle) than the ones extending towards the distal end of *CEN7* ("non-pericentromeric") (Figure 3.1C right). Further, we observed a 25 kb region centring on *CEN7* that closely interacts with the CENPA bound CEN mid-core (Figure 3.1D). This interaction was found to diminish with increasing distance from *CEN7* (Figure 3.1 E). These results state that centromeres in *C. albicans* form a compact chromatin that is closely interacting than the rest of the genome. This compaction enables us to identify a 25 kb pericentromeric chromatin flanking the CENPA bound central core.
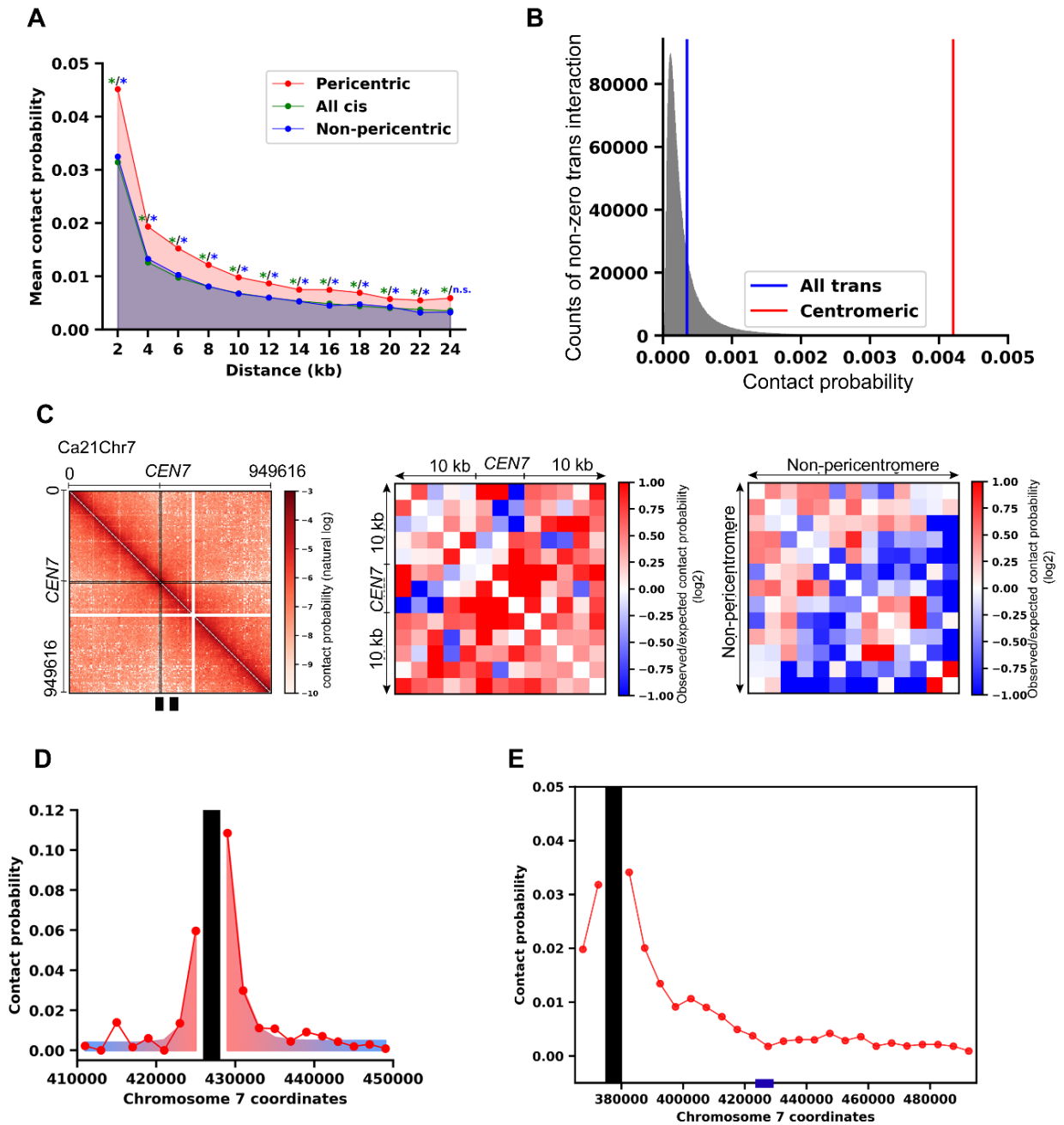
Figure 3.1. Centromeres in *C. albicans* are flanked by pericentromeric chromatin spanning ~25 kb centring the CENPA binding region. (A) The mean Hi-C contact probability (*y*-axis; bin size= 2 kb) separated at a given distance (*x*-axis) depicts significant increase in pericentromeric (centromeric region plus 10 kb upstream and downstream) interactions (red) over all *cis* interactions (green) and non-pericentromeric (a randomly selected region of same size as pericentromeric region and equidistant from the centromere of each chromosome) interactions (blue). (*: p-value<0.05; n.s.: not significant). **(B)** A histogram of non-zero *trans* contact probabilities (grey) from the genome-wide interaction matrix depicts that the mean contact probability of all *trans* (black line) is much lower than interactions among centromeric bins (red) (bin size=2kb). **(C)** Heatmaps of observed/expected contact probabilities (bin size=2kb) at Chr7 zoomed into a pericentromeric region (left) and a non-pericentromeric region (Chr7:440000-466000) with same size (right). The expected matrix was obtained from mean contact probabilities of all *cis* interactions at each distance. **(D)** The 3C profile (bin size=2kb) anchored on centromeric bin (Chr7:426000-428000) showing contact probabilities (red dots) between the anchor bin

58

(black) and its neighbouring bins on Chr7 indicates a strongly interacting 25 kb region. (**E**) The 3C profile (bin size=5kb) anchored on a bin 50kb downstream (Chr7:475000-480000) (black) (C) of *CEN7* (blue box). The red dots represent contact probabilities.

Transgene silencing frequency at pericentromeres decreases with increasing distance from centromere

To gain further insights into the pericentromeric regions, we sought to examine the transcriptional status of CEN-adjacent regions. We inserted the 1.4 kb *URA3* gene at ten independent *CEN7*-proximal loci in a strain that has two differentially marked arms of Chr7, J200 (Sanyal, Baum et al. 2004) (Figure 3.2A). We also performed integrations at a *CEN7*-distal locus (127 kb away from central core) and a *CEN5*-proximal locus (see Table 3.1 for loci of insertions). All insertions were performed at intergenic regions so as not to disrupt any ORFs. We plated approximately 1 million cells of each *URA3* integrant type on CM+5'FOA and replica plated 100 colonies resistant to 5'FOA on CM-Uri to obtain the rate of *URA3* silencing (Figure 3.2B). We also monitored the frequency of chromosome loss in these strains by examining the simultaneous loss of two markers, *ARG4* and *URA3* or *HIS1* and *URA3* (Table 3.2). We obtained reversibly silenced 5'FOA resistant colonies for all integrations except for 5L. We observed a steep decline in the percentage of reversible silencing of *URA3* (the ratio of the number of 5'FOA resistant colonies that grow on CM-Uri and the total number of 5'FOA resistant colonies analysed) from the *CEN7* core. *URA3* when inserted at *CEN7* core exhibited a significantly higher rate of silencing than the peripheral insertions (Figure 3.2C left). The clear trend of exponential decay in reversible silencing of *URA3*, correlated with contact probabilities made by the central core to the neighbouring regions (Figure 3.2C right), indicating that the clustered centromeres of *C. albicans* interact with pericentromeric regions to form a compact nuclear subdomain (up to 25 kb), the frequency of which is ablated at loci distal to the central core.

Table 3.1. Coordinates for *URA3* insertion in *C. albicans*

Type of insertion	Coordinate of insertion	Distance from mid-CEN
5L	Ca21Chr7_417202-417203	10 kb (left of CEN7)
4L	Ca21Chr7_419529-419530	7.7 kb (left of CEN7)
3L	Ca21Chr7_422037-422038	5.2 kb (left of CEN7)
2L	Ca21Chr7_423682-423683	3.5 kb (left of CEN7)
1L	Ca21Chr7_425563-425564	1.7 kb (left of CEN7)
1R	Ca21Chr7_429198-429199	1.9 kb (right of CEN7)
2R	Ca21Chr7_432145-432146	4.9 kb (right of CEN7)
3R	Ca21Chr7_434069-434070	6.8 kb (right of CEN7)
4R	Ca21Chr7_437729-437730	10.4 kb (right of CEN7)
5R	Ca21Chr7_443546-443547	16.2 kb (right of CEN7)
Far-CEN	Ca21Chr7_299510-299511	127 kb (left of CEN7)
CEN5int	Ca21Chr5_477918-477919	7.5 kb (right of CEN5)

Table 3.2. Frequency of reversible silencing of *URA3* integration strains

Strain	Transformant no.	No. of 5'FOA^r colonies analyzed	% reversible silencing (%5'FOA^r Uri$^+$ His$^+$Arg$^+$)
5L	1	107	ND
	2	73	ND
	3	95	ND
4L	1	116	0.862
	2	103	ND
	3	117	0.854
3L	1	86	1.162
	2	96	2.083

	3	98	2.04
2L	1	97	1.03
	2	117	2.564
	3	100	ND
1L	1	110	10
	2	158	14.556
	3	160	3.125
CEN7::URA3/CEN7	J151	74	97.297
	J153	61	78.688
	J154	79	94.936
1R	1	101	98.019
	2	58	100
	3	78	91.025
2R	1	118	1.694
	2	111	0.9
	3	100	1
3R	1	111	0.9
	2	108	0.925
	3	88	1.136
4R	1	97	3.09
	2	96	1.04
	3	101	1.98
5R	1	138	0.724
	2	157	1.273
	3	100	ND
far-CEN	1	200	ND
	2	200	ND
	3	205	ND
CEN5 int	1	98	4.08
	2	114	1.75
	3	89	ND

ND= Not determined
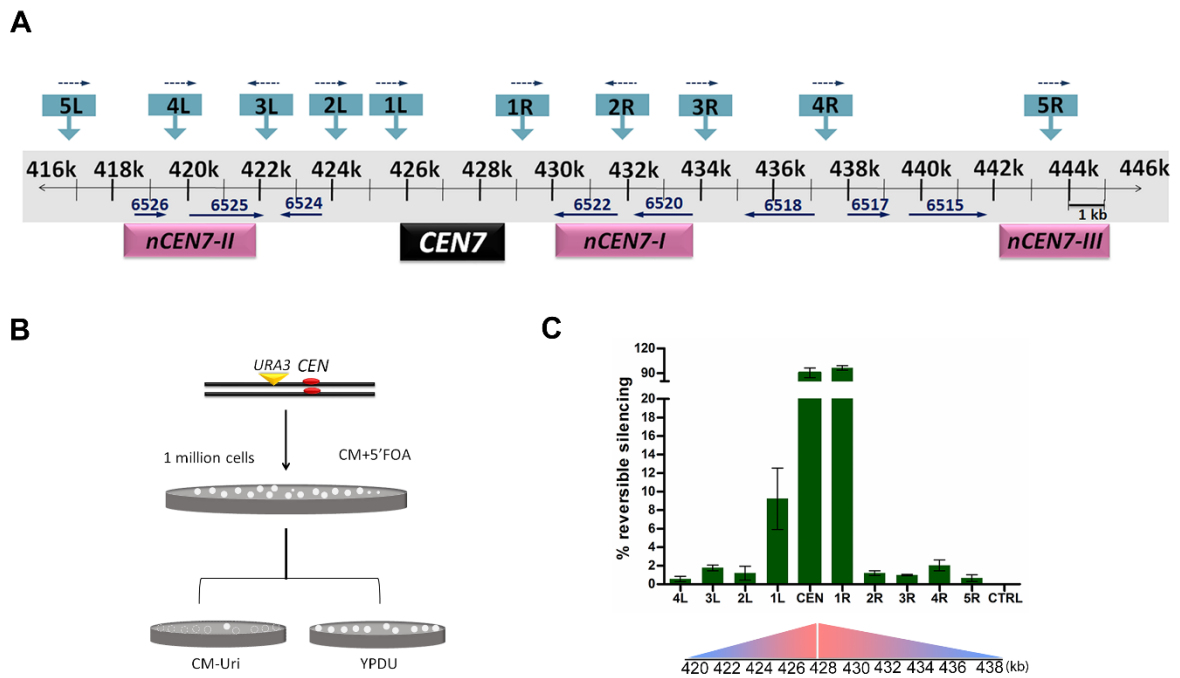
Figure 3.2 Transgene silencing frequency at pericentromeres decreases with increasing distance from centromere. (**A**) A line diagram of ~30 kb region on Chr7 surrounding *CEN7* shows individual *URA3* insertion locations (blue arrows) and previously mapped neocentromere hotspots (*nCEN7-I, nCEN7-II*). Arrowheads and numbers indicate positions and identities of the ORFs respectively. (**B**) The assay strategy used to screen reversibly silenced colonies derived from the *URA3* integrants using 5'FOA. (**C**) A decline in the percentage of reversible silencing of *URA3* from mid-*CEN7* (CEN) to the pericentromeric integrants (4L,3L,2L…5R, CTRL) was observed with increasing distance from native *CEN*. The phase exponential decay curve is colour coded with the graph in (3.1D). Coordinates for the respective insertions have been depicted below the black line.

Transgene silencing at the pericentromeres is associated with an ectopic kinetochore

Transcriptional silencing of *URA3* at the *C. albicans* CEN core is known to facilitate CENPA binding (Thakur and Sanyal 2013). We wanted to examine the consequence of *URA3* silencing in these pericentromeric insertions. ChIP experiments on the 5'FOA resistant colonies revealed that *URA3* is significantly enriched with CENPA when cells were grown in CM+5'FOA than CM-Uri indicating that transcriptional repression of *URA3* at pericentromeres favours CENPA binding in all the *URA3* insertions which yielded the 5'FOA resistant colonies (Figure.3.3 top panel, Figure 3.4A). We did not observe this phenomenon in the far-*CEN7* integrant (Figure 3.4B). We expressed Protein A-tagged Mtw1, the Mis12 homolog in *C. albicans,* (Roy, Burrack et al. 2011) and confirmed its expression by Western blot analysis (Figure 3.5 A). We detected its significant enrichment on *URA3* in the strains
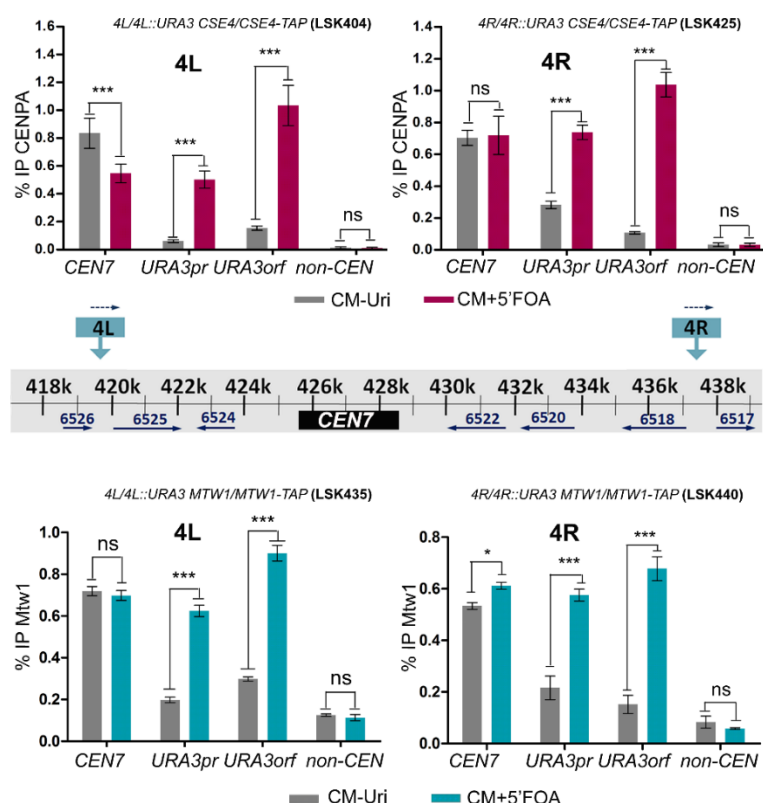
Figure 3.3. Transcriptional silencing of *URA3* at the pericentromeres favours formation of an ectopic kinetochore. A line diagram of the *URA3* insertions at locations 4L and 4R on Chr7 is shown (middle panel). Corresponding integrations are mentioned as graph titles. Anti-Protein A ChIP followed by qPCR analysis of the 5'FOA resistant colonies from the strains LSK404 (4L) and LSK425 (4R) was used to compare enrichment of CENPA on the indicated loci (*CEN7, URA3pr, URA3orf*, non-*CEN* region) in CM-Uri (grey) and CM+5'FOA (red) media (upper panel). Similar ChIP-qPCR assays (lower panels) performed on strains LSK435 (4L) and LSK440 (4R) showed significant enrichment of Mtw1 on the silent *URA3* locus (blue bar). Normalised CENPA and Mtw1 enrichment values indicate significant enrichment at *URA3* upon its transcriptional repression in 4L and 4R integrants (*** $p<0.001$, ** $p<0.01$, ns: $p>0.05$).

LSK437 (4L) and LSK440 (4R) indicating that *URA3* can form an ectopic centromere (ecCEN) when minimally transcribed (Figure 3.3 bottom panel). We performed similar ChIP-qPCR experiments for all other integrations on Chr7 (Figure 3.4 A) and one on Chr5 (Figure 3.4 C). We then examined whether the ecCEN formed was restricted to *URA3*. We performed qPCR analysis using primers corresponding to regions flanking the *URA3* locus in 4L (Figure 3.5 B) and 4R (Figure 3.5 C) integrations to assay for the level of enrichment of both CENPA and Mtw1 in these strains. Standard qPCR analysis revealed that apart from *CEN7*, CENPA and Mtw1 binding regions are limited to the repressed *URA3* locus.
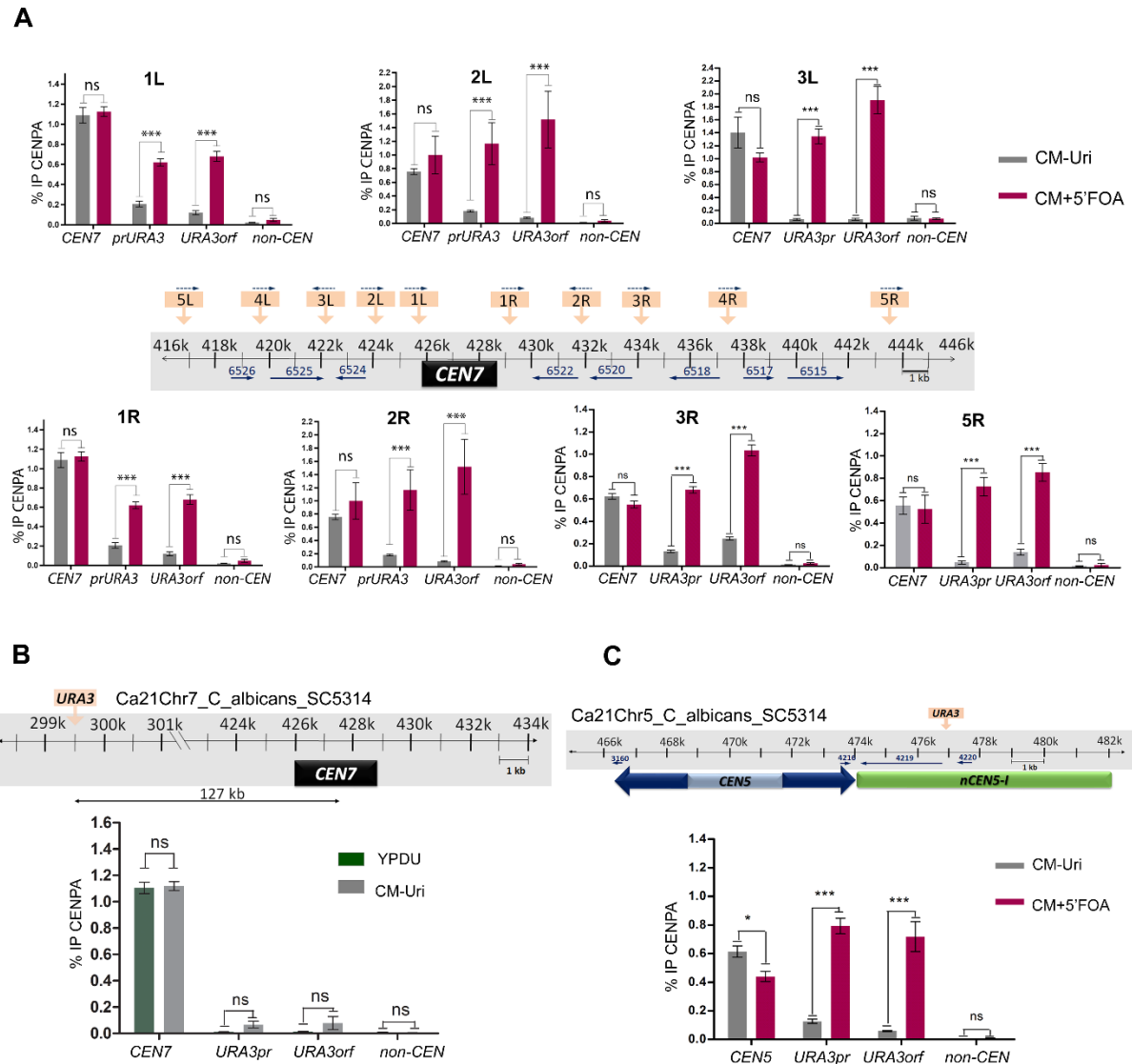
Figure 3.4. Ectopic centromeres are formed at pericentromeric regions of *C. albicans*. (A) Schematic of *URA3* (orange boxes) integrated at pericentromeres of Chr7 is shown (middle panel). Arrowheads and numbers indicate positions and identities of the ORFs. Corresponding sites (1L,2L…5R) are mentioned as graph titles (see Supplemental table S1 for integration coordinates). Standard ChIP-qPCR analysis (using anti-protein A antibodies) of the 5'FOA resistant colonies obtained from these strains was used to compare CENPA enrichment on the indicated loci (*CEN7*, *URA3pr*, *URA3orf*, non-*CEN* region) in CM-Uri (grey) and CM+5'FOA (red). **(B)** Schematic of *URA3* integration at a far-CEN locus, 127 kb away from *CEN7*. ChIP q-PCR results of this strain in YPDU (green bar) and CM-Uri (grey bar) show no significant enrichment of CENPA at the *URA3* locus. **(C)** *CEN5* of *C. albicans*, contains a mid-core (light blue) flanked by inverted repeats (dark blue arrows). *URA3* was integrated at the indicated location, at one of the neocentromere hotspots (*nCEN5-II*). ChIP qPCR results for the 5'FOA resistant colonies obtained from the integrant was grown in CM + 5'FOA (red bar) and CM- Uri (grey bar). Percent input values were normalised to corresponding values on *CEN1*. Statistical significance was determined by two-way ANOVA followed by Bonferroni post-tests (*** $p < 0.001$, ** $p < 0.01$, ns: $p > 0.05$).
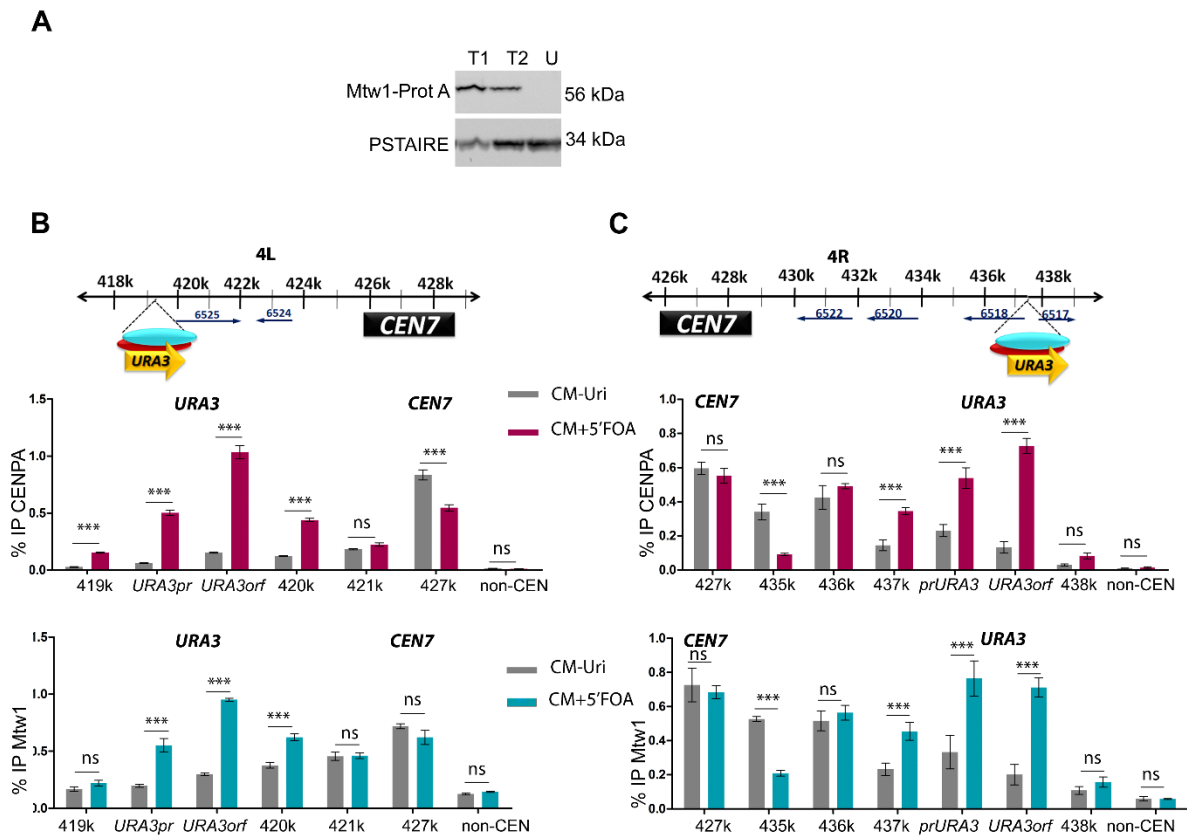
Figure 3.5. Kinetochore binding to ectopic centromeres is restricted to the silent *URA3* locus.
(A) Western blot analysis determines the expression level of the endogenous copy of *MTW1* tagged with Protein A (Prot A) in the strain RM1000AH. Mtw1-Prot A could be detected as a 56 kDa band (T_1, T_2) which was absent in the untagged control (U). PSTAIRE was used as the loading control. **(B)** Both CENPA and Mtw1 bind to the ectopic centromere at *URA3* when the 5'FOA resistant colonies from LSK404 (*4L/4L::URA3 CSE4/CSE4-TAP*) (top panels) and LSK437 (*4L/4L::URA3 CEN7 MTW1/MTW1-TAP*) (bottom panels), are grown in CM +5'FOA (red/ blue bar) or CM-Uri (grey bar). Primers flanking the *URA3* locus (Supplementary table S7) were used to check for the extended binding of CENPA and Mtw1 beyond *URA3*. **(C)** Similar ChIP-qPCR assays were done for 5'FOA resistant colonies from LSK425 (*4R/4R::URA3 CEN7 CSE4/CSE4-TAP*) and LSK440 (*4R/4R::URA3 MTW1/MTW1-TAP*). Percent input values were normalised to *CEN1*. ChIP q-PCR was performed in three independent transformants and technical triplicates for each transformant. Statistical significance was determined by two-way ANOVA followed by Bonferroni post-tests (*** $p<0.001$, ** $p<0.01$, ns: $p>0.05$).

Ectopic kinetochore formed at pericentromeres is transient

We further wanted to determine whether ecCEN can be stably propagated through mitosis by withdrawing the selection. We serially passaged the initial 5'FOA resistant colonies from LSK404 *(4L/4L::URA3)* and LSK425 *(4R/4R::URA3)* in non-selective media (YPDU) for up
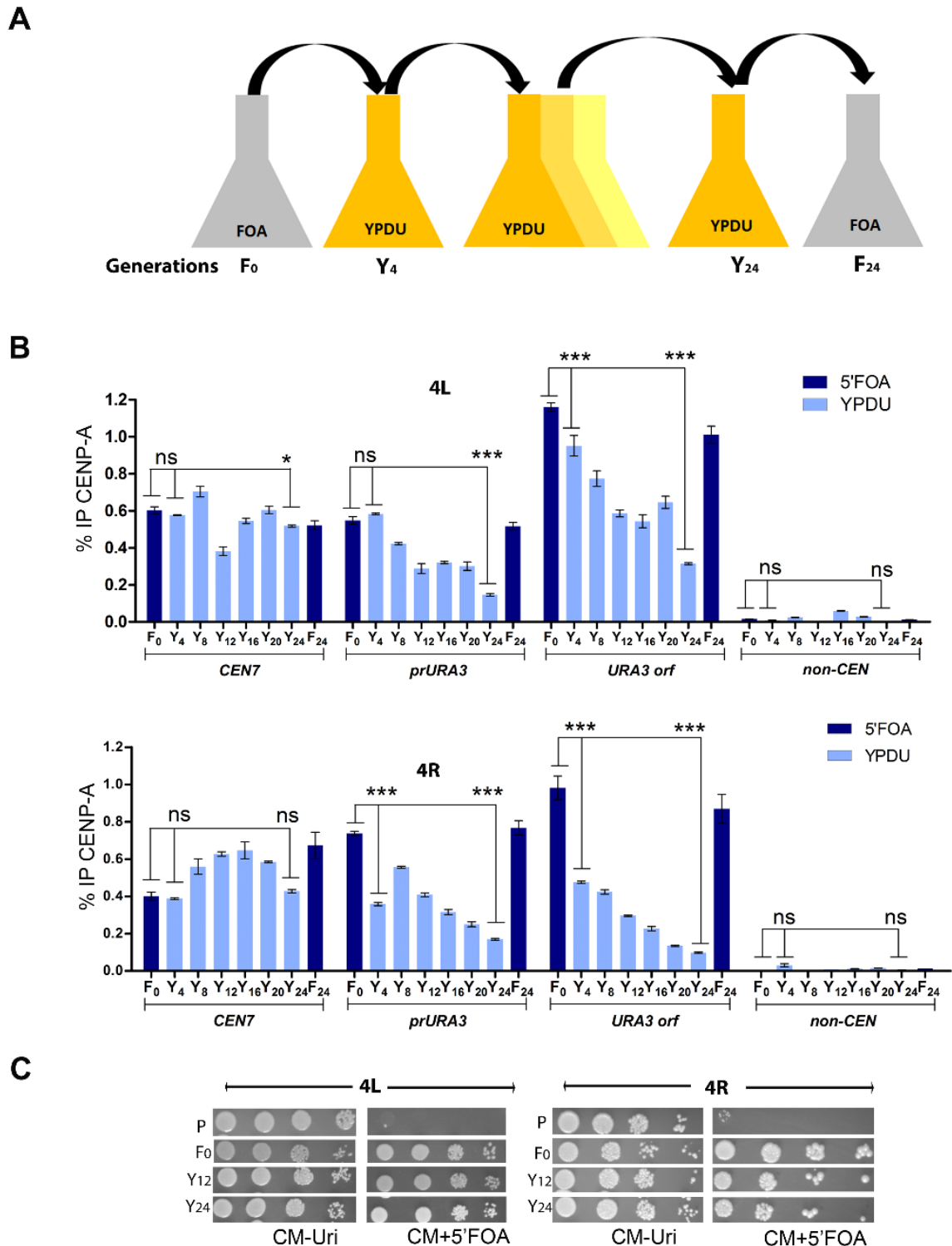
Figure 3.6. Ectopic kinetochore formed at *URA3* is transient and unstable. (A) Schematic of the experiment showing serial passaging of 5'FOA resistant colonies obtained from the strains LSK404 (*4L/4L::URA3 CSE4/CSE4-TAP*) and LSK425 (*4R/4R::URA3 CSE4/CSE4-TAP*) in non-selective media (YPDU). The primary 5'FOA resistant colony was grown in YPDU for the indicated number of generations and then regrown in CM+5'FOA. **(B)** ChIP using anti-Protein A antibodies followed by qPCR analysis reveals a steady decline in enrichment at *URA3* when cells were passaged in the non-selective media (light blue) Percent input values were normalised to *CEN1*. ChIP-qPCR was

performed in two independent transformants of 4L (top) and 4R (bottom) with technical triplicates for each transformant. Statistical significance was determined by one-way ANOVA followed by Bonferroni post-tests (*** p<0.001, ** p<0.01, ns: p>0.05). **(C)** A spotting assay showing the frequency of reversible silencing of the 5'FOA resistant colonies from strains 4L (left) and 4R (right) after they were grown in non-selective media for the indicated number of generations. Individual panels show serially diluted cultures of the 5'FOA sensitive strain (P), primary 5'FOA resistant colony (F_0) and 5'FOA resistant colony grown in YPD for 12 (Y_{12}) and 24 (Y_{24}) generations, spotted on CM+5'FOA and CM-Uri plates. Plates were incubated for 48 h at 30^0C and then photographed.

---

to 20 generations (Figure 3.6A). After every four doublings, cells were harvested and processed for ChIP using anti-Protein A antibodies. We observed a steady decline in the relative enrichment of CENPA at *URA3* with every doubling and after ~20 mitoses, the CENPA level was comparable to a state when cells were forced to express *URA3* (in CM-Uri) (Figure 3.6B). Additionally, we observed that if at any stage of passaging in non-selective media (YPDU), these cells were regrown in presence of selection (CM+5'FOA), they could reassemble the CENPA associated ecCEN on *URA3* (Figure 3.6C). Thus, transcriptional repression of a transgene within the 25 kb compact pericentromeric region favours the formation of a transient ectopic kinetochore.

Pre-existing CENPA molecules can prime a chromosomal location to form neocentromeres

Neocentromeres provide a way to study *de novo* centromere formation since they recapitulate all molecular events for centromere assembly under natural conditions on a non-native locus (Amor and Choo 2002, Craig, Wong et al. 2003, Marshall, Chueh et al. 2008). In *C. albicans*, neocentromeres are shown to get activated at CEN-proximal loci irrespective of the length of the centromere DNA deleted (Thakur and Sanyal 2013). There are four neocentromere hotspots that have been identified by ChIP-sequencing so far on Chr7-*nCEN7-I, nCEN7-II, nCEN7-III* and *nCEN7-IV*.

We wanted to examine that in the event of a centromere deletion, whether a cell would prefer to form a neocentromere on a pre-determined hotspot or on a CENPA-primed region located at the pericentromeric region. To address the same, we replaced the core 4.5 kb CENPA-rich *CEN7* region (Ca21Chr7 424475-428994) with the 1.2 kb *HIS1* sequence independently in two 5'FOA resistant strains, LSK443 (*4L/4L::URA3*) and LSK456 homolog (in *cis*) using
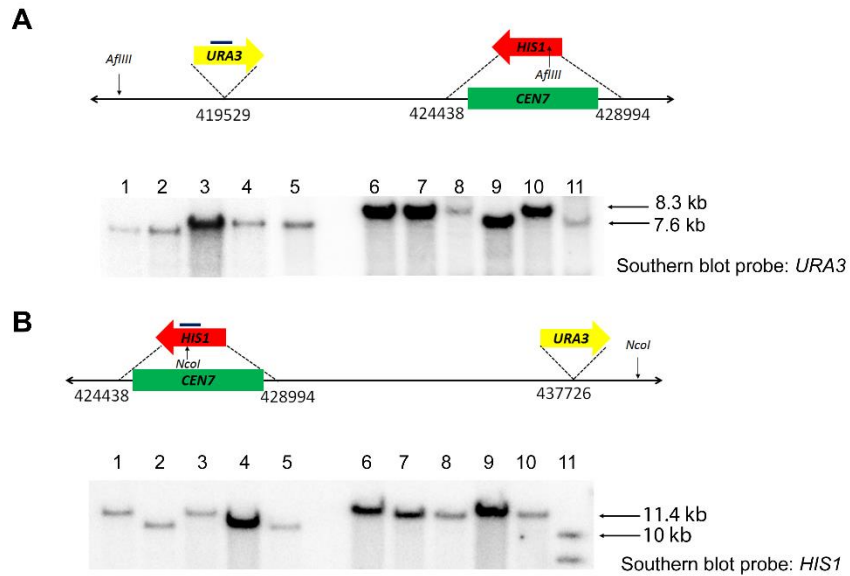
Figure 3.7. Southern analysis of *CEN7* deletion strains. (A) A line diagram showing restriction sites of pericentromeres of Chr7 when *URA3* (yellow arrow) is located at 4L (7.7 kb left of *CEN7).* *CEN7* (CaChr7 424475-428993) (green) has been replaced with *HIS1* (red). Genomic DNA from strains LSK445, LSK446, LSK 447, LSK 448, LSK449, LSK450, LSK451, LSK452, LSK453, LSK454 and LSK455 (lanes 1-11) were digested with *Afl*III, Southern hybridized and probed with a *URA3* fragment. The desired band of 8.3 kb suggests the presence of *URA3* and *HIS1* on the same homolog of Chr7. **(B)** A line diagram showing restriction digestion of pericentromeres of Chr7 when *URA3* is located at 4R (10.4 kb right of *CEN7). CEN7* (CaChr7 424475-428993) has been replaced with *HIS1*. Genomic DNA from strains LSK459, LSK460, LSK461, LSK462, LSK463, LSK464, LSK465, LSK466, LSK467 and LSK468 (lanes 1-10) were digested with *Nco*I, Southern hybridized and probed with a *HIS1* fragment. The desired band of 11.4 kb suggests the presence of *URA3* and *HIS1* on the same homolog of Chr7.

(*4R/4R::URA3*). We screened for colonies where *URA3* and *HIS1* were located on the same Southern hybridization (Figure 3.7) and obtained multiple transformants. We performed the same deletion in the corresponding 5'FOA sensitive *URA3* integrants and examined whether a CENPA primed region could assemble a functional kinetochore.

ChIP-qPCR analysis in the 5'FOA resistant strain LSK465 (*4R/4R::URA3 CEN7/CEN7::HIS1)* (Figure 3.8A) and LSK450 (*4L/4L::URA3 CEN7/CEN7::HIS1)* (Figure 3.8A) revealed that two independent kinetochore proteins, CENPA and Mtw1, assemble at *URA3* and neighboring regions, apart from the native centromere (Figure 3.8B, Figure 3.9B). We confirmed neocentromere formation on this altered chromosome by CENPA ChIP-sequencing in these strains, which revealed two new hotspots at CENPA-primed regions, *URA3nCEN7-I* and *URA3nCEN7-II* (Figure 3.8 C, D, Figure 3.9 C) (see table 3.3 for
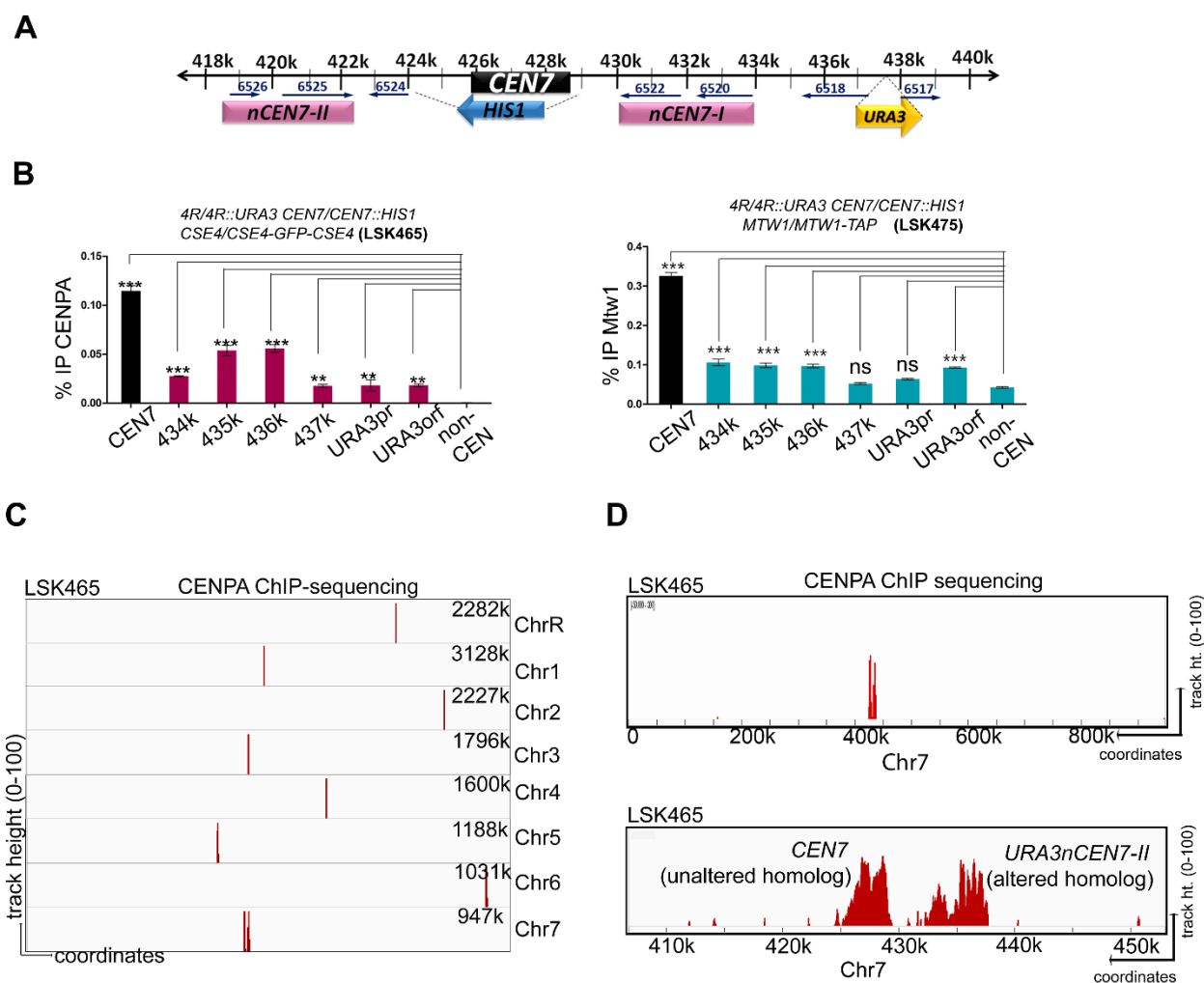
Figure 3.8. Pre-existing CENPA molecules can prime a chromosomal location for neocentromere formation. (A) In the diploid *C. albicans*, only one homolog of Chr7 where *CEN7* (CaChr7 424475-428994) has been replaced with *HIS1* in a *URA3* integrant strain (CaChr7 437729-437730) is shown. **(B)** Left: Relative enrichment of CENPA at native *CEN7* from the unaltered homolog (black) and at neocentromere *URA3nCEN7-II* (red) in the 5'FOA resistant strain LSK465. Right: Relative enrichment of Mtw1 at *CEN7* (black) and *URA3nCEN7-II* (blue) at the native centromere (427k) in the strain LSK465. Relative enrichment values of CENPA and Mtw1 indicate that the neocentromere formed on the altered homolog (*URA3nCEN7-II*) was mapped to a region surrounding the integration locus (CaChr7 435078-440387) error bars indicate SEM (***p<0.001, ** p<0.01, ns p>0.05). **(C)** ChIP-sequencing using anti-GFP antibodies in the strain LSK465 (*CSE4/CSE4-GFP-CSE4 4R/4R::URA3 CEN7/CEN7::HIS1)* reveals a single peak on all chromosomes, except Chr7 that shows two closely spaced CENPA peaks, centromere (*CEN7*) and neocentromere (*URA3nCEN7-II).* **(D)** CENPA ChIP-sequencing confirmed the presence of neocentromere in the strain LSK465, where the profile is a combination of two peaks, the one at *CEN7* is on the unaltered homolog the one at *URA3nCEN7-II)* is on the altered homolog. A 50 kb region harbouring *CEN7* depicts the track height (as on IGV) on *y*-axis and coordinates on the *x*-axis.
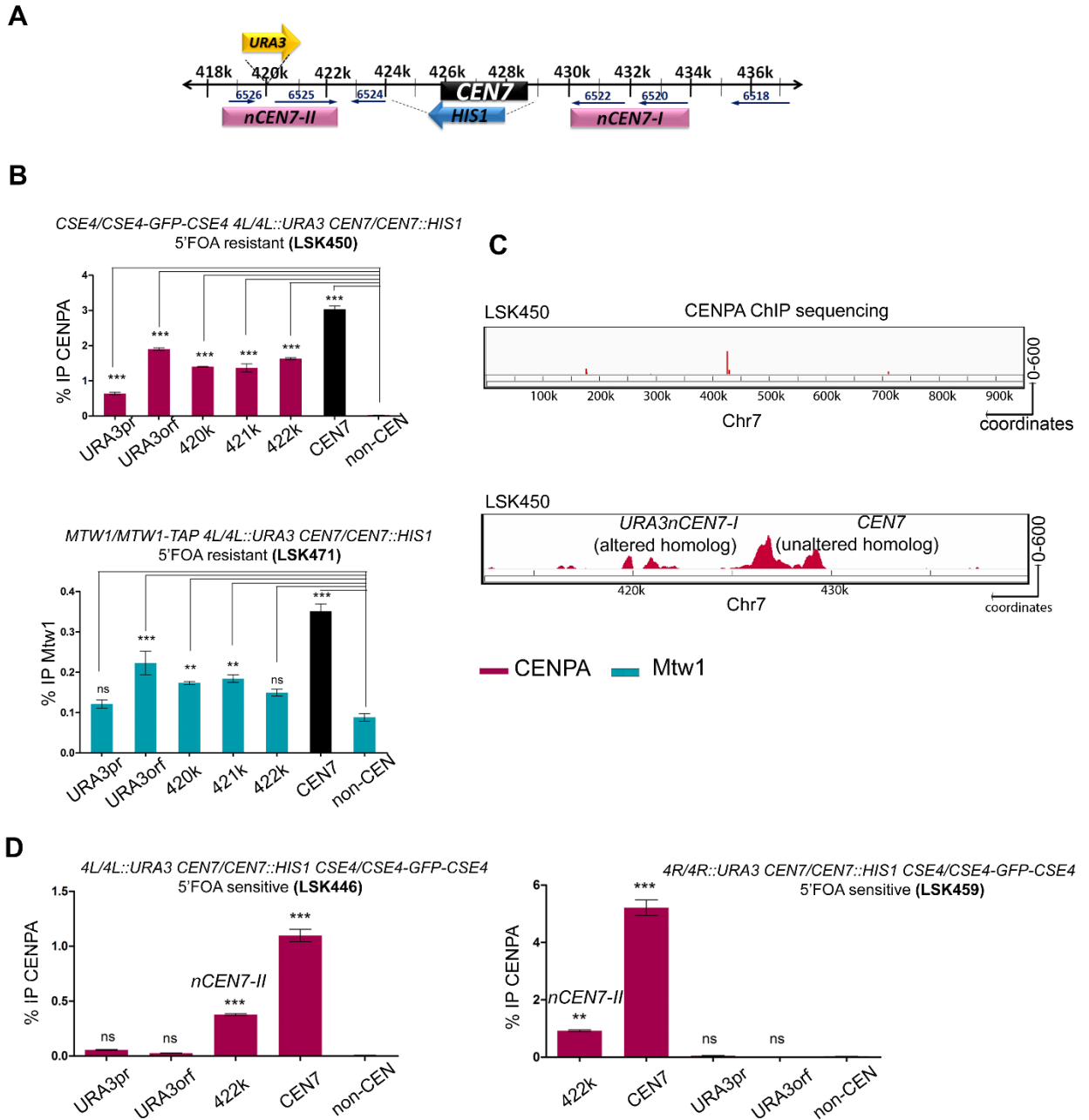
Figure 3.9. The number of CENPA molecules at a CEN proximal region determines the site of neocentromere formation. (A) In the diploid *C. albicans*, only one homolog of Chr7 where *CEN7* (Ca21Chr7 424475-428994) has been replaced with *HIS1* in a *URA3* integrant (Ca21Chr7 419529-419530) is shown. **(B)** Top: relative enrichment of CENPA at native *CEN7* from the unaltered homolog (black) and the neocentromere locus *URA3nCEN7-I* (red) in the 5'FOA resistant strain LSK450 (*CSE4/CSE4-GFP-CSE4 4L/4L::URA3 CEN7/CEN7::HIS1)*. Bottom panel indicates relative enrichment of Mtw1 at *CEN7* (black bar) and *URA3nCEN7-I* (blue) at the native centromere (427k) in LSK471 *(CSE4/CSE4-GFP-CSE4 4L/4L::URA3 CEN7/CEN7::HIS1 MTW1/MTW1-TAP)*. Relative enrichment of CENPA and Mtw1 indicate that neocentromere formed on the altered homolog (*URA3nCEN7-II*) was mapped to a region surrounding the integration locus (Ca21Chr7 435078-440387) (***p<0.001, ** p<0.01, ns p>0.05). **(C)** CENPA ChIP-sequencing confirmed the presence of neocentromere in the strain LSK450, where the profile is a combination of two peaks, the one at *CEN7* is on the unaltered homolog the one at *URA3nCEN7-I* is on the altered homolog. A 30 kb

70

region harbouring *CEN7* depicts the track height (as on IGV) on *y*-axis and coordinates on the *x*-axis. **(D)** CENPA ChIP followed by qPCR in the 5'FOA sensitive strains LSK446 (*CSE4/CSE4-GFP-CSE4 4L/4L::URA3 CEN7/CEN7::HIS1*) (left panel) and LSK459 (*CSE4/CSE4-GFP-CSE4 4R/4R::URA3 CEN7/CEN7::HIS1*) (right panel) indicates that neocentromeres are activated at the previously identified hotspot *nCEN7-II*. There was no CENPA enrichment seen on *URA3*. The experiment was performed in two independent transformants for each type of neocentric strain. Statistical significance was determined by one-way ANOVA followed by Bonferroni post-test (*** $p<0.001$, ** $p<0.01$, ns: $p>0.05$).

Table 3.3 Neocentromere coordinates of *CEN7* deletion strains (from CENPA ChIP-sequencing analysis)

Strain	Description	Coordinates for neocentromere
LSK450	4.5 kb *CEN7* deleted in 5'FOA resistant *URA3* integrant (4R)	Ca21Chr7 419629-422084
LSK465	4.5 kb *CEN7* deleted in 5'FOA resistant *URA3* integrant (4L)	Ca21Chr7 435078-440387

neocentromere coordinates). On the other hand, in the 5'FOA sensitive strains LSK465 and LSK450, neocentromeres formed at one of the pre-determined hotspots, *nCEN7-II* (Figure 3.9 D).

This alludes to the fact that an initial targeting of CENPA to a primed locus within the 25 kb compact region can render centromeric properties to that site if CENPA can enable its nucleation and the subsequent assembly of a functional kinetochore independent of selection or any other target mechanisms.

4. An implicit crosstalk between the pre-RC components and CENPA facilitates centromere establishment and activity

An implicit crosstalk between the pre-RC components and CENPA facilitates centromere establishment and activity

Centromeres and replication origins are often seen to be juxtaposed to each other from bacteria like *Bacillus subtilis* (Livny, Yamaichi et al. 2007) to yeast species like *Yarrowia lipolytica* (Vernis, Abbas et al. 1997). This physical proximity aids in centromere cohesion as well as ensures proper kinetochore assembly (Natsume, Muller et al. 2013). Additionally, *CEN* replication timing is pivotal in CENPA loading, where early replication of *CENs* ensures replication coupled loading of CENPA in *S. cerevisiae* (Pearson, Yeh et al. 2004). One of the mechanisms for the early replication of *CENs* in budding yeast is mediated by the timely recruitment of Dbf4 dependent kinase (DDK) at kinetochores with the help of the Ctf19 complex, which recruits replication initiator proteins to pericentromeric replication origins (Natsume, Muller et al. 2013). Hence, there is an intimate crosstalk of replication initiator proteins and kinetochore components in maintaining genome stability. Centromeres are the earliest to replicate in every S-phase of the *C. albicans* cell cycle by virtue of the early replicating origins flanking the centromere (Mitra, Gomez-Raja et al. 2014). Deletion of *CEN* proximal origins is also known to abrogate centromere function and debilitate kinetochore stability in this organism (Mitra, Gomez-Raja et al. 2014). Hence, the presence of early replication origins near the centromeres not only advances its replication timing, but also helps to seed CENPA to proximal regions in case of centromere inactivation i.e., neocentromere formation. We attempted to examine the relationship between this early replication program on centromere identity and function in *C. albicans.*

Orc4 binds to native and neocentromeres in *C. albicans*

Apart from the discrete genomic loci across all chromosome arms, the strong binding of Orc4 on all centromeres in *C. albicans* was an intriguing observation. This hints towards the possible role of replication initiator complexes in influencing centromere location and function. Upon comparison of the Orc4 enrichment with the CENPA occupancy in *C. albicans*, we observe that there is a significant overlap in the binding regions of both these proteins, indicating a strong physical association of ORCs at all centromeres (Figure 4.1A, table 3.1). We validated this binding by standard ChIP-qPCR assays and found Orc4 to be significantly enriched at all *C. albicans* centromeres (Figure 4.1B). We additionally examined for the presence of Orc4 in non-native centromeres. ChIP-qPCR analysis in the 5'FOA
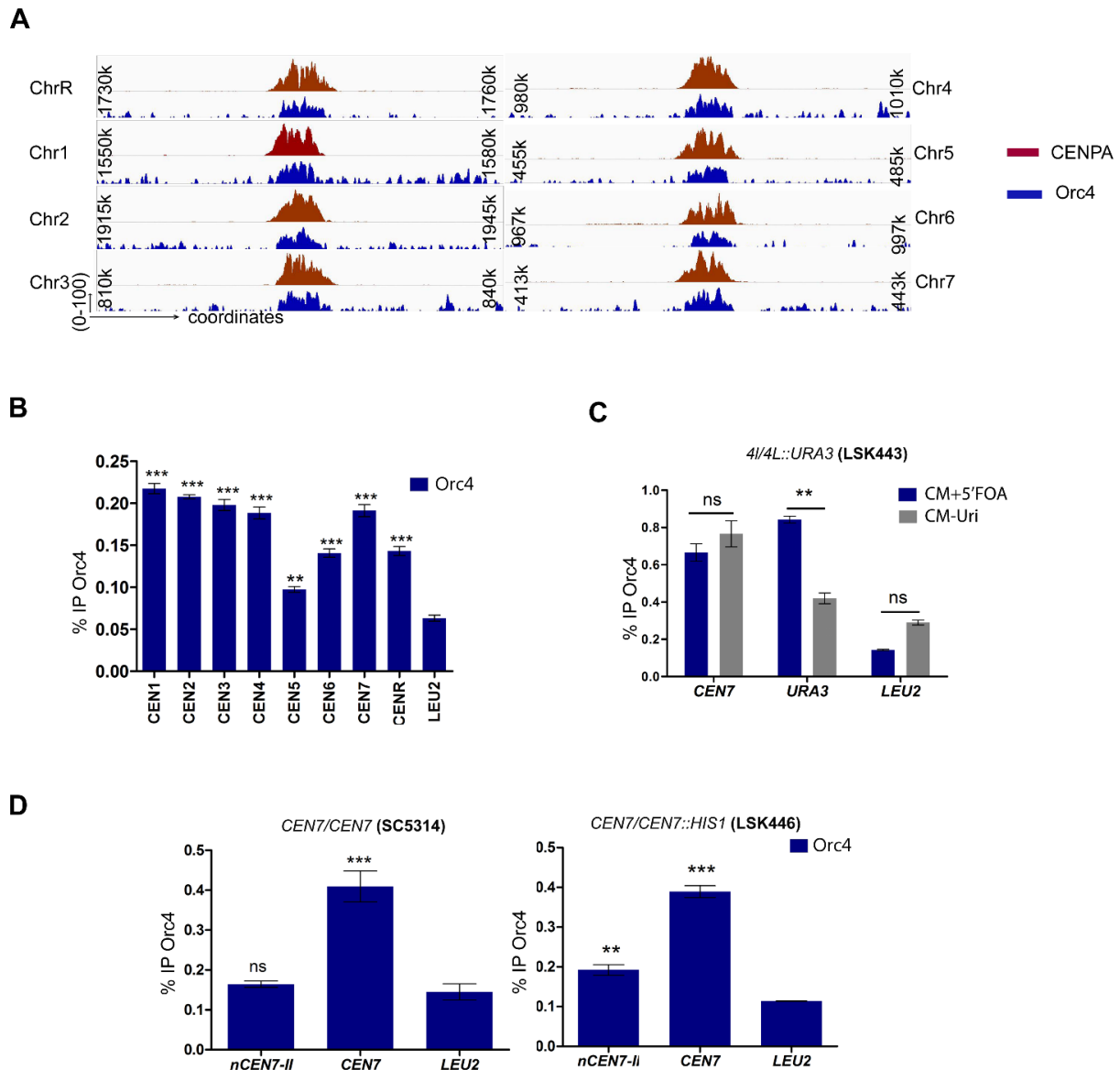
Figure 4.1 Orc4 is associated with native and ectopic centromeres in *C. albicans* (A) Orc4 ChIP sequencing revealed the highest binding of Orc4 at all eight *C. albicans* centromeres. The binding region of Orc4 (blue) showed a complete overlap with CENPA binding region (red). A 30 kb region harbouring each centromere (*x*-axis) was plotted against the subtracted ChIP sequencing reads (*y*-axis) for CENPA and Orc4. **(B)** Orc4 ChIP followed by standard qPCR assays was used to validate the enrichment of Orc4 at all *C. albicans* centromeres. *LEU2* was used as a control region. The experiment was performed in two replicates of SC5314. Statistical significance was determined by one-way ANOVA followed by Bonferroni post-tests (*** $p<0.001$, ** $p<0.01$, ns: $p>0.05$). **(C)** Orc4 ChIP followed by qPCR analysis in the strain LSK443 *(4L/4L::URA3)* revealed the significant enrichment of Orc4 on *URA3orf* in CM+5'FOA over CM-Uri. Statistical significance was determined by two-way ANOVA followed by Bonferroni post-tests (*** $p<0.001$, ** $p<0.01$, ns: $p>0.05$). **(D)** Orc4 ChIP qPCR in the wild type (*CEN7/CEN7*) (left side) and *CEN7* deletion strain LSK446 (*CEN7/CEN7::HIS1)* (right panel) indicates significant enrichment of Orc4 at a neocentromere hotspot over the control region (*LEU2*). Statistical significance was determined by one-way ANOVA followed by Bonferroni post-tests (*** $p<0.001$, ** $p<0.01$, ns: $p>0.05$).

Table 4.1. Chromosomal coordinates for Orc4 binding at centromeres based on *C. albicans* Assembly 21

Centromere	CENPA binding region Coordinates (length)	Orc4 binding region Coordinates (length)
1	Ca21Chr1 15662315-1566930 (4616 bp)	Ca21Chr1 1562748- 1566244 (3497 bp)
2	Ca21Chr2 1925206- 1929688 (4483 bp)	Ca21Chr2 1926183- 1929443 (3261 bp)
3	Ca21Chr3 822762-827727 (4966 bp)	Ca21Chr3 823057- 826863 (3807 bp)
4	Ca21Chr4 991382-996030 (4649 bp)	Ca21Chr4 992010- 995522 (3513 bp)
5	Ca21Chr5 4673814-472497 (5114 bp)	Ca21Chr5 468552- 471618 (3067 bp)
6	Ca21Chr6 979686-984007 (4322 bp)	Ca21Chr6 980541- 983910 (3370 bp)
7	Ca21Chr7 425129- 431652 (6524 bp)	Ca21Chr7 425910- 429297 (3388 bp)
R	Ca21ChrR 1742833- 1748598 (5766 bp)	Ca21ChrR 1743951- 1747274 (3324 bp)

resistant strain LSK443 revealed that similar to CENPA binding, the conditional ecCEN at *URA3* is enriched with Orc4 (Figure 4.1C). To validate the association of Orc4 at functional centromeres, we explored its binding to strains forming neocentromeres. Neocentromeres activated at *nCEN7-II* hotspot upon deletion of the 4.5 kb CENPA rich region on *CEN7* showed significant Orc4 enrichment on the altered homolog (Figure 4.1D). These observations strongly suggest that Orc4 is associated with all active centromeres in *C. albicans.*

The physical association of Orc4 to centromeres stabilizes CENPA

To examine it role in centromere function, we assayed for CENPA localisation in an *orc4* mutant, LSK331 (*ORC4::FRT/MET3prORC4 CSE4-GFP-CSE4/CSE4*). After 24 h of Orc4 depletion, significant reduction in chromatin bound CENPA by standard ChIP-qPCR analysis was observed (Figure 4.2A). We wanted to test the reciprocal relationship and hence employed the strain CAKS3b (Sanyal and Carbon 2002) where CENPA was placed under the
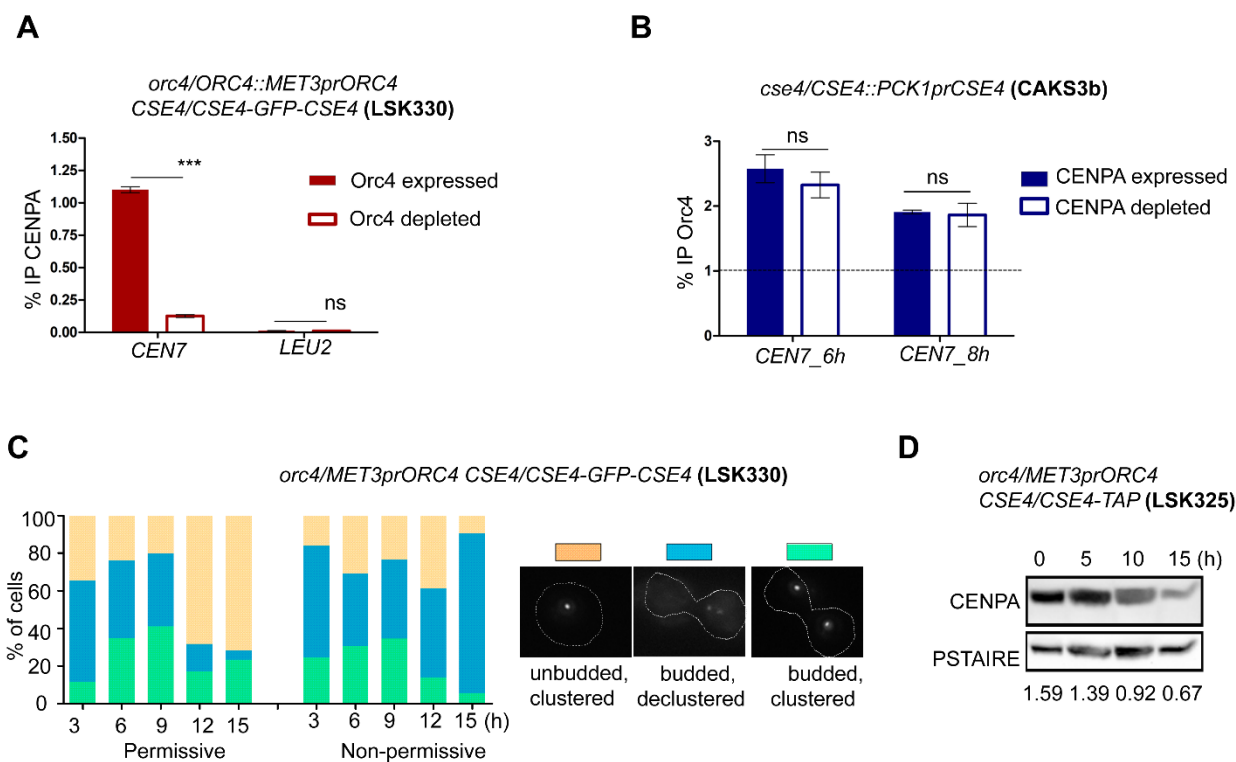
Figure 4.2 Orc4 stabilizes CENPA and preserves kinetochore integrity. (**A**) ChIP-qPCR using anti-GFP antibodies (CENPA) to reveal CENPA enrichment at the centromere upon Orc4 depletion in the strain LSK330 grown either in CM-met-cys or CM+5mM met + 5mM cys for 24h. (**B**) ChIP-qPCR using anti-Orc4 antibodies when CENPA is depleted for 6h and 8 h in YP with dextrose in the strain CAKS3b. Percent IP values are normalized to a non-centromeric control region, *LEU2*. Two-way ANOVA was used to determine statistical significance (***p<0.001, ns p>0.05). (**C**) CENPA segregation pattern was examined in an *orc4* conditional mutant LSK330. The strain was grown either in CM-met-cys or CM+5mM met +5mM cys for 3,6,9,12 and 15h to shut down the expression of *ORC4* and the percentage of cells showing a specific segregation phenotype of clustered kinetochores in small budded (yellow), unsegregated budded (blue) and segregated budded (green) was counted. (**D**) Western blot using anti-Protein A antibodies to estimate CENPA levels when Orc4 is depleted for 0,5,10,15 h, normalized with the loading control, PSTAIRE.

regulatable *PCK1* promoter. Depletion of CENPA for 6 h and 8h did not later the levels of Orc4 at the *CEN7* suggesting that Orc4 and CENPA are not reciprocally dependent on each other. Orc4 depletion caused severe kinetochore segregation defects. Time course depletion of Orc4 demonstrated around 90% cells to be have duplicated but unsegregated kinetochore (Figure 4.2C). These cells were predominantly large budded and upon extended incubation formed pseudo-hyphal structures indicating a cell cycle arrest. These phenotypes were corroborated by the fact that protein levels of CENPA measured by Western blot analysis were significantly reduced upon Orc4 depletion (Figure 4.2D). Hence, Orc4 has a direct role in stabilizing CENPA, thereby influencing centromere activity and kinetochore segregation.
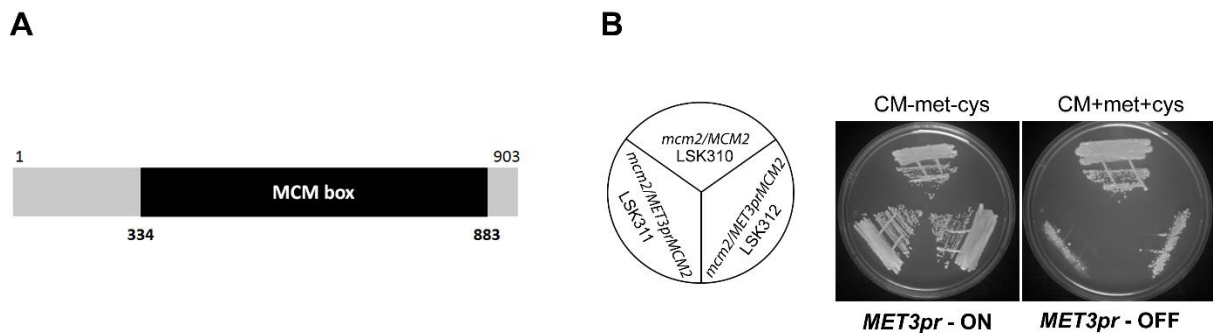
Figure 4.3. Mcm2 is a highly conserved essential protein in *C. albicans*. (**A**) Domain architecture of CaMcm2 reveals a 903 aa long polypeptide consisting of the conserved MCM box that houses the Walker A and Walker B motifs. (**B**) *C. albicans* cells where one deleted copy of *MCM2* has been deleted, LSK310 (*MCM2/MCM2::FRT*), and two independent transformants of the conditional mutant of *mcm2* (*MCM2::FRT/MET3prMCM2*), LSK311 and LSK312 where the remaining wild-type copy was placed under the control of the *MET3* promoter were streaked on plates containing inducible (CM-met-cys) and repressible (CM+5 mM cys + 5 mM met) media and photographed after 48 h of incubation at 30°C.

The helicase subunit, Mcm2 is a conserved protein and essential for viability in *C. albicans*

Even though ORCs flag mark replication origins in the genome, their subsequent activity is governed by assembly of the Mcm2-7 complex that primes the complex for replication initiation. Ca*MCM2* is annotated as an uncharacterized ORF (orf19.4354) in Candida Genome Database (candidagenome.org). BLAST analysis using *S. cerevisiae* Mcm2 as the query sequence revealed that this ORF translates to a 101.2 kDa protein that contains the conserved Walker A, Walker B and the R finger motif which together constitute the MCM box (Forsburg 2004) (Sreyoshi Mitra, PhD thesis) (Figure 4.3A). In order to determine the essentiality of this gene in *C. albicans* we constructed a conditional mutant of *MCM2*, LSK311(*mcm2/MET3prMCM2 CSE4-GFP-CSE4/CSE4*) by deleting one allele and replacing the endogenous promoter of the remaining *MCM2* allele with the repressive *MET3* promoter (Care, Trevethick et al. 1999). Inability of two independent transformants of the conditional mutant (*MCM2::FRT/ MET3prMCM2*), LSK311 and LSK312 to grow in the presence of CM supplemented with 5mM methionine and 5mM cysteine (Figure 4.3B) indicates that Mcm2 is essential for viability in *C. albicans*. The heterozygous null mutant of *MCM2*, LSK310 (*MCM2/MCM2::FRT*) grown under similar conditions did not show any growth defects.
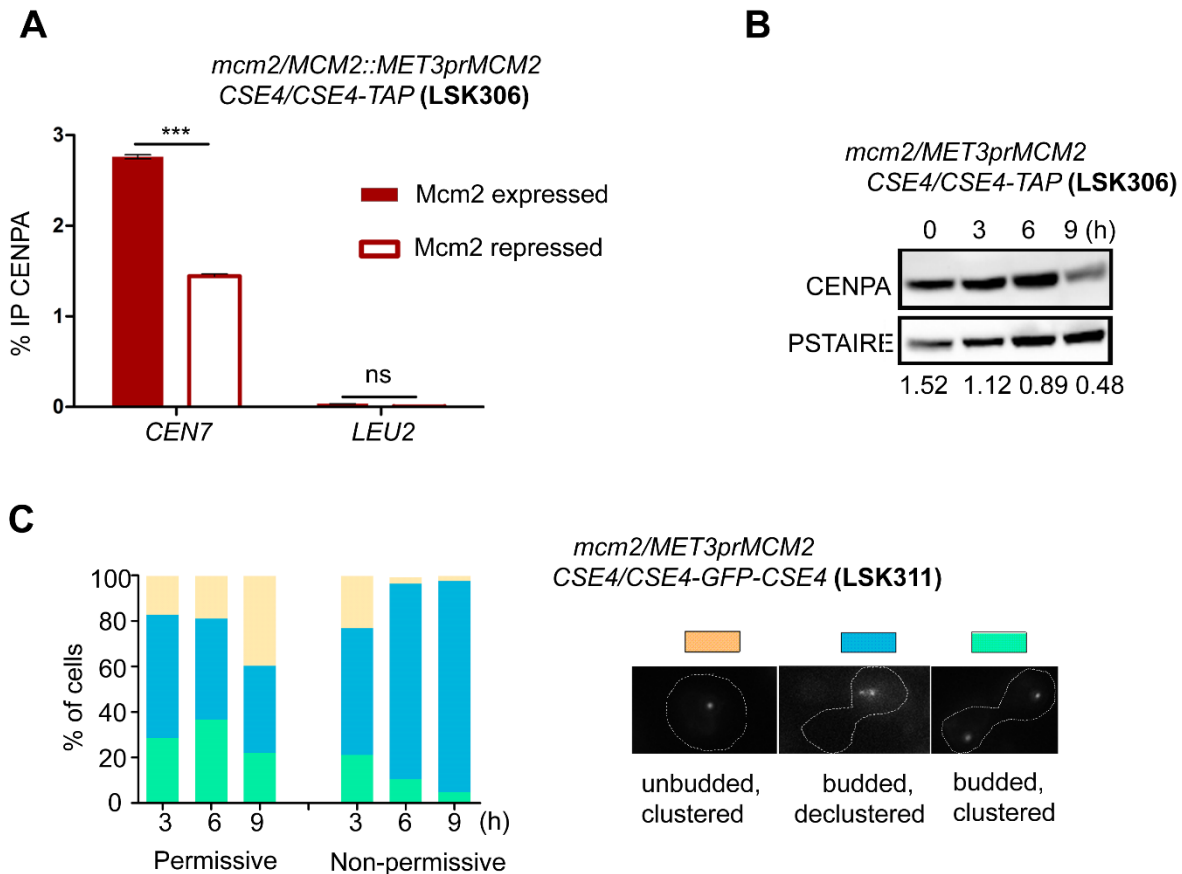
Figure 4.4 Mcm2 stabilizes CENPA. (**A**) ChIP-qPCR using anti-protein A antibodies revealed significant reduction in CENPA at *CEN7* in the strain LSK306 when grown either in CM-met-cys or CM+5mM met+5mM cys for 6 h. ChIP was performed in biological triplicates. Two-way ANOVA was used to determine statistical significance (***$p < 0.001$, ns $p > 0.05$). (**B**) Western blot showing protein levels of CENPA upon depletion of Mcm2 for 3,6,9 h. Normalization was performed using PSTAIRE. (**C**) CENPA (clustered kinetochore) segregation pattern was examined in an *mcm2* conditional mutant LSK311. The strain was grown either in CM-met-cys or CM+ 5mM met+ 5mM cys for 3,6,9 h to shut down the expression of *MCM2*. Approximately 100 cells from three independent transformants of *orc4* mutant were analyzed for this assay.

Mcm2 influences CENPA stability in *C. albicans*

Mcm2 is a known chaperone for H3-H4 incorporation at the replication fork. It is also known to bind to CENPA-H4 tetramer *in vitro*. We wanted to examine the impact of Mcm2 depletion on CENPA stability. We examined CENPA occupancy at *CEN7* in an *mcm2* conditional mutant LSK306. Post 6h of depletion of the protein, CENPA occupancy was reduced by 70% at the centromere (Figure 4.4A). Western blot analysis indicated a reduction of CENPA levels upon a time course Mcm2 depletion (Figure 4.4B). Similar to the

phenotypes observed upon Orc4 depletion, Mcm2 depletion showed impairment in kinetochore segregation (Figure 4.4C). Hence, we establish the role of two pre-RC components in CENPA stability in *C. albicans*.

5. DISCUSSION

Replication origins in *C. albicans* are not defined by a DNA sequence, but are spatiotemporally distributed

In the present study, we employed the binding of a pre-RC component, Orc4 to map its genome-wide binding sites in the pathogenic yeast *C. albicans*. We have characterized the Orc4 protein in *C. albicans* and examined its expression and localisation using antibodies against endogenous Orc4. The strong association of Orc4 at all eight centromeres is a notable observation in our ChIP-sequencing analysis. Centromeres in *C. albicans* do not contain an active firing origin since incoming replication forks flanking the centromere stall at the centromere in a kinetochore mediated manner (Mitra, Gomez-Raja et al. 2014). Since ORCs are associated with active and latent origins, it is possible that centromeres could act as a strong ORC binding, non-firing region similar to multiple other ORC binding regions of the genome. We could identify the genomic imprint of CaOrc4 binding as the putative replication origins in a medically relevant fungal pathogen that exhibits unusual genomic features (Lakshmi Sreekumar 2017).

We observed distinct features of replication origins in *C. albicans*. Majority of the them occupy genic locations, as compared to the expected location of active origins and functional ARSs in *S. cerevisiae* to intergenic regions (Xu, Aparicio et al. 2006) and hence are more reminiscent of the metazoan replication origins (Cayrou, Coulombe et al. 2011). Chromosomal regions flanking replication origins influence origin activity and ORC binding. Most of the Orc4 binding regions in our study were found to be associated with NFRs in congruence with the fact that replication origins reside in nucleosome-depleted regions of the genome. The previous genome-wide study on identification of ORC binding regions in *C. albicans* utilized antibodies against the *S. cerevisiae* ORC complex (Tsai, Baller et al. 2014). Moreover, all the origins identified in the previous study may not be an accurate depiction of the putative origin locations in *C. albicans* since the ORCs in *S. cerevisiae* recognize the AT-rich ACS associated origin sequences. The said study reported ~390 ORC binding sites but only 25% (106/414) of them are truly CaOrc4 binding sites that we identified. Since we used antibodies against an endogenous protein (CaOrc4) to map its binding sites in *C. albicans*, we possess a more reliable tool to identify replication origins in this organism.

Replication origins are spatially placed to complete the replication of the genetic material during S-phase. Yeast species that have similar DNA sequence elements in their origins show strong evolutionary pressure in maintaining regularly spaced replication origins. This is very

clear in case of the four yeast species, namely, *S. cerevisiae, Lachancea waltii, Lachancea kluyveri* and *Kluyveromyces lactis (Newman, Mamun et al. 2013)*. In spite of sharing the similar number of replication origins as *S. cerevisiae*, only 24% of the replication origins in *L. kluyveri* are conserved in the *S. cerevisiae* genome (Agier, Romano et al. 2013). In our study, we observe that the average separation of ORC binding sites across all chromosomes was ~35 kb. Previously calculated values for origin separation in various fungal species range from 26 kb in *S. cerevisiae* to 71 kb in *K. lactis* (Newman, Mamun et al. 2013). Unlike yeast species, replication origins in higher organisms like *Xenopus* are randomly placed and fire stochastically and hence, they are grouped as clusters that have similar replication timing (Romanowski, Madine et al. 1996, Blow, Gillespie et al. 2001).

S. cerevisiae has approximately 400 genomic sites that facilitate replication initiation, that comprises of only 5% of the total DNA sequences that harbour the ACS (Wyrick, Aparicio et al. 2001, Nieduszynski, Knox et al. 2006, Xu, Yanagisawa et al. 2012). This low origin usage suggests that active origins are selected by virtue of features other than the primary DNA sequence (local chromatin environment), even in an organism like *S. cerevisiae* that shows strong genetic requirements to specify replication origins. Species like *L. kluyveri* and *L. waltii* which diverged from *S. cerevisiae* before the whole genome duplication event, contain ACS-like elements, although active replication origins in these two species do not behave like evolutionary fragile sites, unlike in *S. cerevisiae* (Liachko, Tanaka et al. 2011, Agier, Romano et al. 2013). In contrast, *S. pombe* origins are 5-10 times larger sequences (0.5-1 kb) than *S. cerevisiae* and lack a consensus sequence analogous to the *S. cerevisiae* A element, relying on asymmetric AT sequences for ORC binding (Kong and DePamphilis 2002). At the other end of the spectrum are the metazoan ORCs that bind promiscuously to DNA and tend to rely on conserved structural features (like G-quadruplex) rather than specific consensus sequence for origin selection and specification (Cayrou, Coulombe et al. 2011). Unlike yeast, origins in metazoans are enriched with G-rich sequences and CpG islands, which have been proposed to serve two purposes- NFR maintenance and G-quadruplex formation (Parker, Botchan et al. 2017). Our analysis reveals the lack of a DNA sequence requirement for most of the ORC binding sites in *C. albicans*. However, we find a strong association of origins within many tRNA genes. Motif analysis using DIVERSITY revealed the presence of tRNA motifs in a subset (50/417) of the binding regions. tRNA genes along with histone genes and centromeres are known to exhibit conserved replication timing (Muller and Nieduszynski 2017). Moreover, in *S. cerevisiae,* there is a statistically significant bias for codirectional

transcription and replication of tRNA genes (Muller and Nieduszynski 2017). tRNA genes (tDNA) harbour sites for TFIIIC, TFIIIB and RNA polymerase III binding. These regions are actively transcribed and associated with nucleosome free regions. Apart from active transcription, tDNA also has established roles in genome architecture maintenance by acting as binding sites for SMC subunits (Glynn, Megee et al. 2004) and chromatin remodelers (Kogut, Wang et al. 2009). Additionally, tDNAs cluster near centromeres and recovers stalled forks (Thompson, Haeusler et al. 2003). Since centromeres are early replicating in fungal genomes, the presence of tDNAs in its vicinity might transduce an early replication program. Hence, the various motifs identified in our study hints towards differential usage and specification of origins facilitated by multiple modes of ORC binding in *C. albicans*.

The replication timing analysis performed in our study revealed that similar to *C. glabrata* origins, early origins in *C. albicans* are found to be clustered together. The stochastic activation of early origins, makes up for the uneven distribution of early origins in the genome. The 'replication wave' progresses with the sequential activation of pericentromeric origins to the chromosome arm origins. This considers the number of pre-RC factories that reside to facilitate the said initiation events. Initiation events convert early origin clusters to replication foci during S phase. As DNA replication progresses, more replisomes are formed, chromosomes are sufficiently mobilized which makes long range interactions more favourable with time. The limited initiation factors are better captured by the unfired/ dormant origins as they get slowly released from the early replicons. Hence, one can speculate the existence of topologically distinct domains that are separated in location and time as S-phase progresses. Considering the temporal compartmentalisation of S phase, if about a third of the origins are early replication, one would have ~140 origins in *C. albicans* that replicate a genome of 14 Mb during S phase. This implies that one early origin replicates a stretch of ~100 kb (50 kb on either side of the fork). Hence, the fork velocity (ratio of length of DNA replicated and S phase time) would be 2-3 kb per minute in *C. albicans*. This is very similar to the fork speed reported in *C. glabrata* (Descorps-Declere, Saguez et al. 2015).
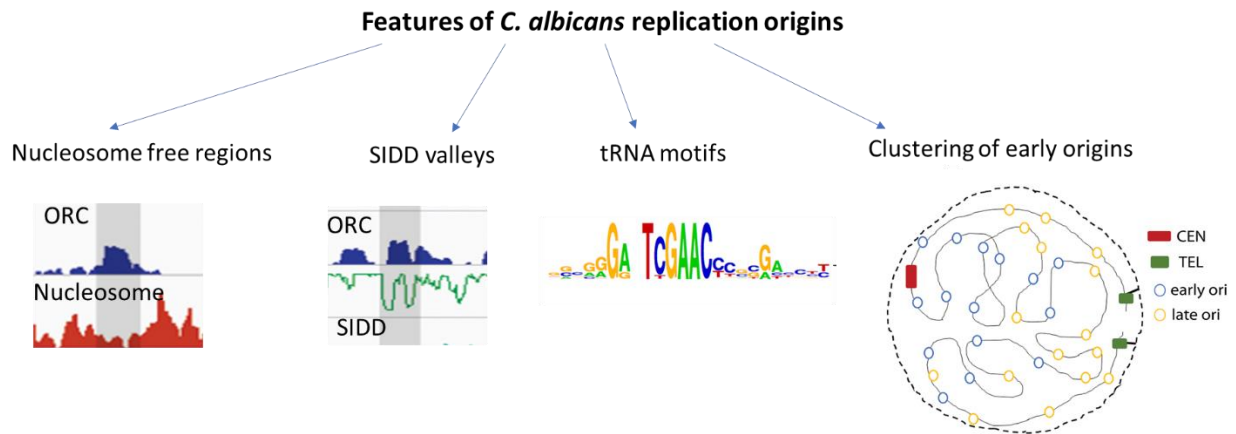
Figure 5.1 Characteristics of *C. albicans* replication origins. *C. albicans* origins are associated with nucleosome depleted regions and SIDD valleys. They do not carry a signature sequence but a fraction of them reside within tDNA. Organized clustering of similarly timed domains (early-early and late-late) helps in spatiotemporal regulation of *C. albicans* origins.

Most eukaryotes, unlike the *Saccharomyces* group of species, share similar attributes to specify replication origins. They do not have DNA consensus motifs and are rich in AT-rich sequence that are located in nucleosome depleted regions of the genome. *C. albicans* does not exhibit any dependence on DNA sequence features to specify essential chromosome elements. For example, centromeres in this organism show no requirement of a DNA sequence to bind to evolutionarily conserved kinetochore proteins (Sanyal, Baum et al. 2004) and are specified by epigenetic factors (Baum, Sanyal et al. 2006). Unlike centromeres, replication origins are present at multiple genomic locations. Hence, it is relevant to think that the lack of a consensus DNA sequence across all the replication origins in *C. albicans* can be considered as a breakpoint in chromosome evolution. This study helps in understanding the number and distribution of possible initiator sequences in a fungal genome and hence fills the missing gaps in deciphering fungal replication origins.

The number of CENPA molecules at the pericentromeric region determines the site of neocentromere formation.

Seeding of CENPA on DNA, the stability of centromeric chromatin during the cell cycle and its subsequent propagation involves a plethora of factors ranging from the primary DNA sequence and the chromatin context to crosstalk with DNA replication initiator and DNA damage repair proteins. Specific protein binding sites aid centromere formation in genetically

defined point centromeres (Lechner and Carbon 1991). Additional mechanisms must operate at multi-dimensional levels to spatiotemporally define centromere activity to a defined region in epigenetically regulated regional centromeres in most other organisms. Centromeric heterochromatin is distinct from arm heterochromatin, in terms of the degree of compaction and presence of topological adjusters like cohesin, condensin and topoisomerase II (Bloom 2014). The cruciform structure adopted by the centromeric chromatin in budding yeast facilitates cohesin maintenance on duplicated sister chromatids and orients the centromere towards the spindle pole (Stephens, Haase et al. 2011) (Lawrimore, Doshi et al. 2018). Centromere clustering gradually increases with cell cycle progression due to sister-chromatid cohesion during replication, and cohesion-mediated spindle-dependent clustering during anaphase (Lazar-Stefanita, Scolari et al. 2017). In this study, we determine the extent and functional consequence of centromeric chromatin compaction in *C. albicans.* The fact that inter-centromeric interactions are much stronger than the average genomic interactions, facilitates the formation of a CENPA cloud in a 3-dimensional milieu to enrich the local CENPA concentration in the clustered CENs of *C. albicans*. In the *S. cerevisiae* point centromeres, the presence of core and accessory CENPA molecules at the native centromeres and pericentromeres, respectively, helps in rapid incorporation of CENPA into the CEN chromatin during rogue loss events (Haase, Mishra et al. 2013), suggesting the dynamic nature of pericentromeric nucleosomes. Hence, the identification of a highly interacting 25 kb pericentromeric region in *C. albicans* enables us to dissect functional underpinnings of pericentromeres and spatial segregation of chromatin properties in and around active centromeres.

The strong reversible silencing of the transgene at the *C. albicans* central core, that is a readout of its flexible transcriptional status (Thakur and Sanyal 2013) is reminiscent of the repeat-associated centromere organization in *S. pombe,* where the central core shows variegated levels of marker gene expression whereas the outer repeats shut down the transgene expression due to heterochromatinization (Allshire, Javerzat et al. 1994, Karpen and Allshire 1997). Even though *S. pombe* outer repeats do not bind CENPA, they are considered an important component of a functional centromere (Clarke, Amstutz et al. 1986). Similarly, the pericentromeric regions identified in our study, probably possess pericentromeric properties in *C. albicans*. In *S. pombe*, CENPA can assemble on a non-centromeric region by competing out H3 (Castillo, Mellone et al. 2007), and the frequency of
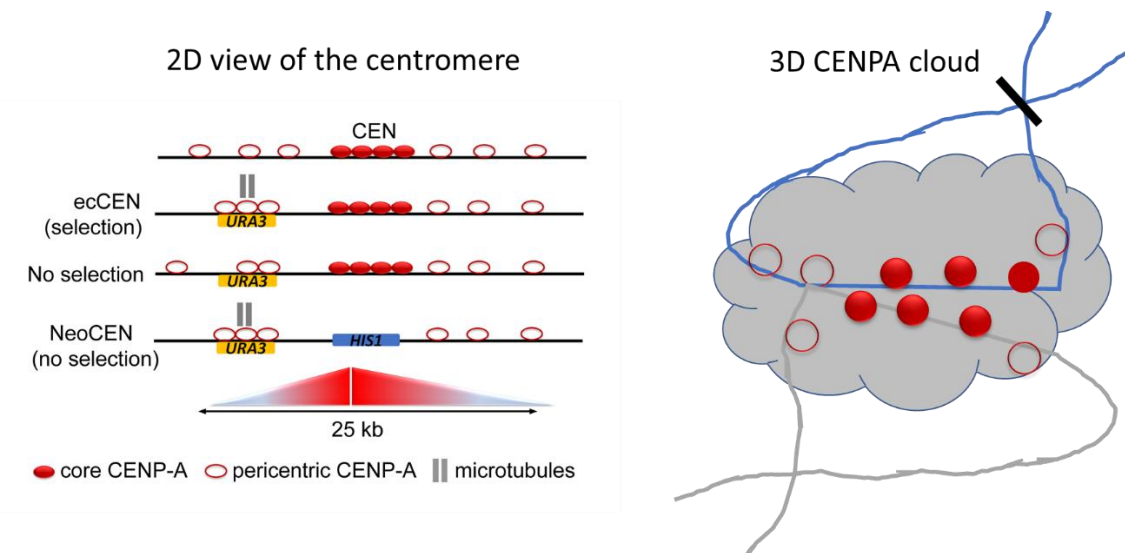
Figure 5.2. CENPA priming at pericentromeres is required for centromere formation. A 2D representation of the pericentromeric heterochromatin in *C. albicans* defined by a 25 kb unique sequence flanking the CENPA bound CEN core. The transient ectopic CEN created by transcriptional repression gets stabilized once the endogenous CEN DNA sequence is deleted. The preferential bias towards a CENPA primed region to assemble the kinetochore machinery, over a pre-determined hotspot indicates the requirement of a minimum number of CENPA molecules to initiate centromere/neocentromere formation. These CENPA molecules closely interact in 3D to form a CENPA rich zone (cloud) that facilitate neocentromere formation.

reversible silencing can be increased by overexpressing CENPA. In our study, the strong negative selection imposed on cells by 5'FOA, enables us to isolate rare individuals from a heterogenous population of cells that can transiently incorporate CENPA at an ectopic locus. We create this microenvironment by selecting cells with suppressed transcription of the transgene, which in this case favours centromere formation. Withdrawal of this selection dilutes those rare events in a population of cells which eventually form centromeres at the endogenous locus. However, even under selective growth conditions, only a few cells can tolerate the ecCEN because of the presence of the more dominating native centromere locus, eventually weeding out cells with ecCEN from the population. Formation of an ecCEN outside the native CEN strongly suggests the existence of non-centromeric CENPA molecules interspersed with the H3 nucleosome. Unlike *S. pombe,* CENPA overexpression in *C. albicans* does not lead to its extended occupancy beyond centromeric chromatin, it merely increases its occupancy at the native locus (Berman 2012). It is to be noted here that the ecCEN that was obtained by growth in selective media did not have an over-expression of CENPA and still harbored two intact copies of native *CEN7*.

CENPA associated chromatin, has self-propagating properties and hence relies on an epigenetic memory (Black, Brock et al. 2007). Our observation that any CENPA primed region within the identified pericentromeric boundaries (25 kb centring on the CEN core) can initiate neocentromere formation, emphasizes the importance of the number of CENPA molecules required to nucleate the kinetochore assembly. However, we have limited understanding regarding the determinants that act is favour of a particular locus on a chromosome to have a "centromere correctness". The interspecific differences in the site of neocentromere formation exist partly because of complexities in genome organization and in particular, the centromere organization. Relocation of a centromere to a more favourable site are powerful evolutionary forces for speciation in mitotically propagating organisms like *C. albicans* that lack a defined sexual cycle. The presence of unique, non-repetitive, epigenetic centromeres and the absence of well characterized heterochromatin and RNAi components makes its regulation non-obvious when compared to other species (Lakshmi Sreekumar 2017).

Orc4 and Mcm2 stabilize CENPA and influence centromere activity

C. albicans centromeres do not possess a firing origin (Mitra, Gomez-Raja et al. 2014). Replication forks originating from the centromere flanking origins stall at the centromere in a kinetochore dependent manner and facilitate new CENPA loading. Furthermore, CENPA loading is facilitated by the physical interaction of repair proteins like Rad51, Rad52 with CENPA, that are transiently localised to the kinetochore upon replication fork stalling (Mitra, Gomez-Raja et al. 2014). Hence, there is an interplay of the replication-repair machinery in maintaining centromere identity in *C. albicans*. We already know the existence of a CENPA rich zone facilitates incorporation of CENPA at the pericentromeres. Here, we postulate the existence of an "ORC-cloud" formed due to ORC abundance at all centromeres of *C. albicans*. Because centromeres have to be replicated in the earliest phase of the cell cycle, the origins flanking the centromere need to be highly efficient. This is also accounted for the biased distribution of the pre-RC subunits at or near the centromeres. Since CEN proximal origins are known to affect CENPA levels at the centromeres (Mitra, Gomez-Raja et al. 2014), the proximity of initiator proteins can be a primary facilitator of CENPA loading.

The anaphase specific loading of new CENPA in *C. albicans* has been demonstrated earlier by fluorescence spectroscopic measurements (Shivaraju, Unruh et al. 2012). However,
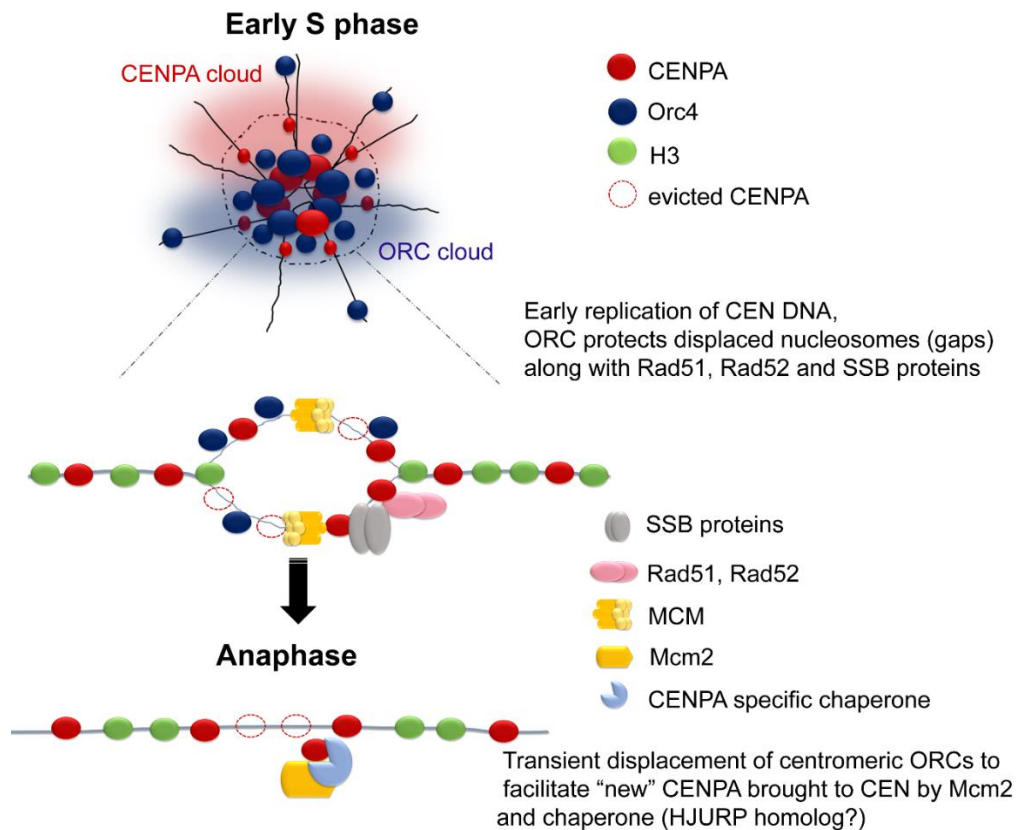
Figure 5.3. Propagation of CEN chromatin in *C. albicans*. CENPA rich centromeres are surrounded by the peripheral CENPA cloud in a compact chromatin environment. Along with the CENPA cloud, an ORC cloud exists at the centromeres to facilitate timely recruitment of initiator proteins for early CEN replication. The pericentromeres are interspersed with H3 and CENPA nucleosomes. In the early S phase, parental CENPA molecules are evicted from the central core and the gaps created are either occupied by a placeholder (unknown) or the CEN DNA has to be protected by the physical proximity and association of proteins like Orc4, along with Rad51, Rad52 and single stranded DNA binding proteins. Orc4 maintains centromeric heterochromatin and facilitates early CEN DNA replication. The pre-RC soon replicates other parts of the genome till the end of S phase. During anaphase, Orc4, though otherwise constitutively associated with the CEN chromatin, undergoes a transient dissociation to facilitate new CENPA loading by a specific chaperone (unknown) that is co-chaperoned by Mcm2. Hence, the epigenetic marks are propagated to the subsequent cell cycles.

specific chaperones and molecular pathways involved in the same are undeciphered. Our results reveal a reduction in CENPA levels and unsegregated kinetochore in the absence of Orc4. However, we do not see a reciprocal dependency, suggesting that ORC is an essential stabilizer of the kinetochore. We posit ORC to be an essential component for CENPA loading and kinetochore stability.

Members of the pre-RC have established roles in cell cycle dependent dynamics at the centromere. Mcm2 is a known chaperone that hands over the old histones from the

replication fork to anti-silencing function Asf1 to recycle old histones and deposits them to the newly synthesised DNA (Hammond, Stromme et al. 2017). Mcm2 and Asf1 cochaperone an H3-H4 dimer through histone-binding mode (Richet, Liu et al. 2015). This is true for both canonical H3 as well as H3 variants like H3.3 and CENPA (Huang, Stromme et al. 2015). A recent study indicates the role played by Mcm2 in mouse embryonic stem cells to symmetrically partition modified histones to daughter cells using its histone-binding mode (Petryk, Dalby et al. 2018). In humans, the S phase retention of CENPA is mitigated by its simultaneous interaction with the specific chaperone HJURP and Mcm2 (Zasadzinska, Huang et al. 2018), which together transmit CENPA nucleosomes upon its disassembly ahead of the replication fork. In the light of the existing evidence in metazoan systems and the results obtained in our study, Mcm2 emerges as an evolutionarily conserved factor required for eviction of old CENPA molecules and loading of newly synthesized ones by a possible interaction with a designated CENPA chaperone. Although our experiments demonstrate a strong genetic interaction between these two proteins, the physical interaction of Orc4-CENPA and Mcm2-CENPA is still speculative due to technical difficulties.

We hypothesize that during CEN chromatin replication at S phase, ORCs maintain the heterochromatin environment of CEN when "old" CENPA is evicted. These gaps or displaced nucleosomes could be filled with a placeholder like canonical H3 or a variant, or could be left unfilled, in which case the centromeric DNA has to be protected. Since ORCs have known heterochromatin functions in budding yeast, in *C. albicans* Orc4 (ORC cloud) might protect the "gapped" nucleosomes till the end of metaphase. During anaphase, only the centromeric ORCs are briefly displaced, to facilitate loading of "new" CENPA with the help of a specific chaperone such as Scm3/HJURP and Mcm2 which stabilizes the kinetochore complex. This pre-RC dependent S phase independent loading of CENPA is a novel aspect in the centromere biology of this organism.

ORCs in *S. cerevisiae* have established roles in heterochromatinization and MTL silencing (Foss, McNally et al. 1993, Hickman, Froyd et al. 2011). The centromere silencing mechanisms in *C. albicans* is relatively unknown as this organism lacks a functional RNAi machinery and H3K9me2 marks. In this regard, we envision the ORC family of proteins as a possible silencing factor for centromeres in this organism.

6. MATERIALS AND METHODS

Strains and primers

All the yeast strains used for the study have been listed in Table 6.1. The oligonucleotide primers used have been tabulated in Table 6.2.

Media, growth conditions and transformation

All strains of *C. albicans* where *URA3* was integrated on Chr7 and Chr5 were propagated in YPD (1% yeast extract, 2% peptone, 2% dextrose) with uridine, unless otherwise specified. All transformations were done in YPDU. The auxotrophs were selected on appropriate selection media, as mentioned previously. For the 5'FOA plating assays, complete media with 2 % agar were supplemented with 1 mg/ml 5'FOA (Sigma cat. no. F5013-1G). ChIP experiments for the silenced colonies were done in a) complete media supplemented with 10 mg/ml uridine and 1mg/ml 5'FOA (CM+5'FOA) and b) CM-Uri. Strains with neocentromeres were grown in YPDU. *ORC4* and *MCM2* mutants were grown either in CM-methionine-cysteine or in CM + 5mM methionine +5mM cysteine for the indicated number of hours. CAKS3b (Sanyal and Carbon 2002) was grown in YP with succinate (2%) for expressing CENPA and YP with dextrose (2%) for depleting CENPA for 6 and 8 h for the ChIP experiments. Transformation of *C. albicans* was performed using the lithium acetate mediated transformation technique, as described previously (Baum, Sanyal et al. 2006).

Strain construction

Construction of a conditional mutant for ORC4

In order to create a conditional null mutant of *ORC4* in *C. albicans*, a deletion cassette was constructed as follows: a 368 bp fragment (Ca21Ch5 coordinates 480170-479721) upstream of orf19.4221 was amplified using the primers ORC4_13/ORC4_14 from the genomic DNA of SC5314 and cloned as a *Kpn*I/*Xho*I fragment in pSFS2a (Reuss, Vik et al. 2004) to create pORC4US. A 490 bp fragment (Ca21Ch5 coordinates 478025-477535) downstream to orf19.4221 was amplified using ORC4_15/ORC4_16 and cloned as a *Sac*II/*Sac*I fragment in pORC4US to generate pORC4DEL (Table 6.2 for Primer list). The plasmid was linearized using *Kpn*I and *Sac*I, and transformed in *C. albicans* YJB8675 (Joglekar, Bouck et al. 2008) and SN148 to obtain the Nourseothricin resistant strains

LSK328 and LSK320, respectively. The marker was recycled to obtain the Nourseothricin sensitive strain LSK329 and LSK321. To inactivate the remaining allele, a conditional mutant was constructed by cloning the N-terminus of orf19.4221 (Ca21Ch5 coordinates 479720-479221) as a *BamH*I/*Pst*I fragment in pCaDIS (Care, Trevethick et al. 1999). The resulting plasmid (pMET3ORC4) was linearized using *Bgl*II and transformed in LSK329 and LSK321 to obtain independent transformants of the conditional mutant LSK330 and LSK325. Transformants were screened based on growth phenotype observed in complete media supplemented with 5mM methionine and 5 mM cysteine.

Construction of a conditional MCM2 mutant

In order to create a conditional null mutant of *MCM2* in *C. albicans*, a deletion cassette was constructed as follows: a 474 bp fragment (Ca21ChR coordinates 857151-856675) upstream of orf19.4354 was amplified using the primers MCM2_13/MCM2_14 from the genomic DNA of SC5314 and cloned as a *Kpn*I/*Xho*I fragment in pSFS2a to create pMCM2US. A 468 bp fragment (Ca21ChR coordinates 853962-853494) downstream to orf19.4354 was amplified using MCM2_15/MCM2_16 and cloned as *Sac*II/*Sac*I fragment in pMCM2US to generate pMCM2DEL (See Table 6.2 for Primer list). The plasmid was digested using *Kpn*I and *Sac*I, used to transform *C. albicans* YJB8675 and SN148 to obtain the Nourseothricin resistance strains LSK309 and LSK301, respectively. The marker was recycled to obtain the Nourseothricin sensitive strain LSK310 and LSKLSK302, respectively. To inactivate the remaining allele, a conditional mutant was constructed by cloning the N-terminus of orf19.4354 (Ca21ChR coordinates 856674-856164) as a *BamH*I/*Pst*I fragment in pCaDIS (47). The resulting plasmid (pMET3MCM2) was linearized using *Bgl*II and transformed in LSK310 to obtain independent transformants of the conditional mutant LSK311 and LSK306, respectively. Transformants were screened based on growth phenotype observed in complete media supplemented with 5mM methionine and 5 mM cysteine.

Construction of URA3 integration cassettes.

To construct the individual *URA3* integration cassettes, long primer pairs were designed. Briefly, 70 bp regions both upstream and downstream to the site of integration (Table 3.1) were incorporated in the primers as overhangs. The 1.4 kb *URA3* gene was amplified from the plasmid *pUC19-URA3* (Mitra, Gomez-Raja et al. 2014) or CIP10 (Murad, Lee et al. 2000) using the primers listed in table 6.2. The PCR products were independently transformed in the *C. albicans* J200 (Thakur and Sanyal 2013) and YJB8675 (Joglekar, Bouck et al. 2008).

The transformants were selected on complete media minus uridine (CM-Uri) and confirmed by PCR using specific primers (Table 6.2). Three independent transformants of each integration type was taken ahead for the assays. All the distances of individual *URA3* insertions are indicated with respect to the mid-point of *CEN7* which has been taken as Ca21Chr7_427262.

Construction of MTW1-Protein A tagged strains

To tag an endogenous copy of *MTW1* with Protein-A, an MTW1-TAP fragment was amplified from CAKS13 (Roy, Burrack et al. 2011) using primers PJ74 MTW1 TAP Not I and PJ 76 TAP rev SpeI. This sequence was then cloned as a *NotI/SpeI* fragment in pBS-NAT to obtain the plasmid *pMTW1-TAP(NAT).* This plasmid was linearized by using *PacI* and the resulting cassette was transformed in strains RM1000AH to obtain LSK436 *(MTW1/MTW1TAP(NAT).* Subsequently, the *URA3* integration cassettes for the 4L and 4R insertions were transformed in LSK436. For the neocentromere strains, LSK446 and LSK459 (5'FOA sensitive), LSK450 and LSK465 (5'FOA resistant), were transformed with *pMTW1-TAP(NAT)* fragment to obtain the strains LSK469/ LSK470/LSK473/LSK474 (5'FOA sensitive) and LSK471/ LSK472/LSK475/LSK476 (5'FOA resistant). All strains were confirmed by Western blot using anti-Protein A antibodies.

Construction of the CEN7 deletion strains (CaCEN7)

To delete one copy of *CEN7* (4.5 kb CENPA rich region), a deletion cassette was constructed as follows. A 1.4 kb fragment containing a 66 bp sequence of *CEN7* (Ca21Chr7424413-424472), a 70 bp sequence downstream of *CEN7* (Ca21Chr7 428994-429053) and a marker gene (*CaHIS1*) were amplified from pBS-HIS using the primers CEN7DHIS_FP and CEN7DHIS_RP. The PCR product was transformed in the *URA3* integrant strains LSK443 and LSK456. The transformants were selected on complete media lacking histidine. They were confirmed for *CEN7* deletion using the primers 2498-18 and HIS ORF_1. Transformants in *cis*-orientation (for *URA3* and *HIS1*) were screened on the basis of Southern Hybridisation (Southern 1975).

URA3 **Silencing assay**

Each of the *URA3* integrant was grown in YPDU overnight. Approximately, one million cells from three independent transformants of each kind of integration strain were plated on CM with 1 mg/ml 5'FOA. The plates were incubated at 30ºC up to 72 h. One hundred colonies from each plate were patched on complete media lacking uridine and on YPDU. These were simultaneously patched on CM-His and CM-Arg plates to detects events such as loss of the marker gene *URA3* or gene conversion. The colonies showing growth in CM-Uri were counted and the percentage of reversible silencing was determined. These colonies were then taken from the corresponding YPD patch and streaked on CM+5'FOA plates to obtain single colonies for the subsequent ChIP assays.

Serial passaging of 5'FOA resistant strains in YPDU

5'FOA resistant colonies obtained from two independent transformants of LSK404 and LSK425 were inoculated from their respective glycerol stocks in CM+5'FOA. These cells were harvested, washed and reinoculated into YPDU. Cultures were monitored for their growth and samples were withdrawn after every 4 doublings, for ChIP, with F_0 being the initial culture grown in 5'FOA. Y_4, Y_8, Y_{12}, Y_{20}, Y_{24} correspond to these reversibly silenced 5'FOA colonies grown on non-selective media for the indicated number of generations. Approximately 50 O.D. cells were harvested from each time point and ChIP was performed using anti-Protein A antibodies to examine the CENPA occupancy in these colonies. The cells from the last time point (Y_{24}) were washed and resuspended in fresh CM+ 5'FOA. Cells from the indicated time points (Y_{12}, Y_{24}) were washed and resuspend in sterile water. Serial dilutions of these along with the parental *URA3* insertion and 5'FOA resistant colony were made and spotted on CM-Uri and CM + 5'FOA. Plates were incubated for 72 h at 30ºC and photographed.

Generation of Orc4 antibodies

The peptide sequence from *C. albicans* Orc4 (YLPKRKIDKEESSI) was chemically synthesized and conjugated with Keyhole Limpet Hemocyanin (KLH) (GeneMed Synthesis, USA). The conjugated peptide (1 mg/ml) was mixed with equal volumes of Freund's complete adjuvant (Sigma, Cat no. F5881) and used as an antigen to inject non-immunized

rabbits as the priming dose. Three subsequent booster doses at an interval of two weeks (per immunization) were given using Freund's incomplete adjuvant (Sigma, Cat no. F5506). Following antibody detection using ELISA, major bleed was performed. The anti-serum was collected, IgG fractionated and affinity purified against the free peptide (AbGenex, India). The specificity of the purified antibody preparation was confirmed by western blot and immunolocalisation experiments.

Antibodies used

For Western blot analysis, we used rabbit anti-Protein A antibodies- 1:5000 (Sigma cat.no. P2921), mouse anti-PSTAIRE antibodies (Abcam cat no. 9866) in the dilution of 1: 5000 and rabbit anti-Orc4 antibodies in the dilution of 1:1000. Goat anti-rabbit IgG-HRP (dilution 1:10,000) (Bangalore Genei cat no. 105499) and goat anti-mouse IgG-HRP (dilution 1:10,000) (Bangalore Genei cat. no. HP06) were used as secondary antibodies. For indirect-immunofluorescence, we used anti-Orc4 antibodies in the dilution of 1:100 and Alexa Fluor goat anti-rabbit IgG 568 (Invitrogen Cat. No. 11011) in the dilution of 1:500. For ChIP, rabbit anti-Prot A antibodies (3ug/ ml IP), rabbit anti-Orc4 antibodies (10ug/ml IP) and mouse anti-GFP (Roche Cat# 11814460001) (4ug/ml) were used.

Western blotting

Approximately 3 O.D. equivalent cells were harvested and precipitated by 12.5% TCA overnight at -20°C. The pellet was spun down at 13000 rpm and washed with 80% acetone. Pellet was then dried and resuspended in lysis buffer (1% SDS, 1N NaOH) and SDS loading dye. Samples were boiled for 5 min and electrophoresed on a 10% polyacrylamide gel. Protein transfer was performed by semi-dry method for 40 min at 25V. Following protein transfer, the blot was blocked with 5% skimmed milk for an hour. The blot was incubated with anti-Orc4 antibodies or anti-PSTAIRE antibodies. The blot was washed thrice in PBST (1X PBS + 0.05% Tween) and incubated with goat anti-rabbit IgG-HRP. Following three PBST washes, the blot was developed using chemiluminescence method (SuperSignal West Pico Chemiluminescent substrate, Thermo scientific, cat no. 34080).

Fluorescence microscopy

For conditional expression of genes under the *MET3* promoter, cells were grown in permissive media (CM -met-cys) overnight. They were then grown in presence of CM + 5 mM met+5mM cys for the indicated time point, corresponding to the repressive phenotype. In each case, the cells were washed twice with water and resuspended in distilled water which was placed on a 2% agarose bed on a glass slide. Images were captured in 100x using Zeiss Axio Observer 7 and processed using ImageJ and Adobe photoshop.

Indirect immuno-fluorescence

Exponentially grown cultures of SC5314 was fixed with 37% formaldehyde. Spheroplasts were prepared using lysing enzyme (Sigma cat. no. L2265) and cells were fixed on poly-lysine coated slides using methanol and acetone. Cells were then incubated with 2% skimmed milk to block non-specific binding. Following ten PBS washes, cells were incubated with anti-Orc4 antibodies (1:100) for 1 h in a humid chamber. Post PBS washing, cells were incubated with the Alexa Fluor goat anti-rabbit IgG 568 (Invitrogen cat. no. 11011) in the dilution of 1:500 for one hour. The slide was mounted on a coverslip using DAPI (10ng/ul). Microscopic images were captured by a laser confocal microscope (Carl Zeiss, Germany) using LSM 510 META software with He/Ne laser (bandpass 565-615 nm) for Alexafluor 568 and a 2-photon laser near IR (bandpass~780 nm) for DAPI. Z-stacks were collected at 0.4-0.5 µm intervals and stacked projection images were processed using ImageJ and Adobe Photoshop.

Chromatin Immunoprecipitation (ChIP)

ChIP experiments for the reversibly silenced colonies were performed as follows. Each colony that was isolated from CM+5'FOA media was inoculated simultaneously in liquid media of CM + 5'FOA and CM-Uri and grown till log phase. Crosslinking was done for 15 min (for CENPA) or 30 min (for Mtw1) using formaldehyde to a final concentration of 1% and cells were quenched using 0.135 mM glycine for 5 min at room temperature. For Orc4 ChIP, cultures were grown in YPDU and crosslinked for 1 h, and processed similarly. Quenched cells were incubated in a reducing environment in presence of 9.5 ml distilled water and 0.5 ml of beta mercapto-ethanol (HiMedia cat. no. MB041). Spheroplasts were

made using lysing enzyme from *Trichoderma harzanium* (Sigma cat. no. L1412) and spheroplasting buffer (1M sorbitol, 0.1 M sodium citrate, 0.01 M EDTA) for 3 hrs at 37ºC. Cells were spun and washed, following which they were resuspended in 1 ml of lysis buffer (50 mM HEPES pH7.4, 1% Triton X-100, 140 mM NaCl, 0.1% Na-deoxycholate, 1mM EDTA). Chromatin was sheared using a Bioruptor (Diagenode) for 60 cycles of 30 s on and 30 s off. The sheared chromatin was checked on a 2% agarose gel for the appropriate size (300-500 bp). One tenth of the lysate was saved as the input (I) and the rest was split as two fractions- IP (+) and mock (-). To the IP fraction, required concentration of antibody was added and both tubes were incubated at 4ºC overnight. This was followed by adding Protein-A beads (Sigma cat. no. P3391) for 8 hrs. The beads were then washed in low and high salt conditions and finally eluted in elution buffer (1% SDS, 0.1 M sodium bicarbonate). The eluted samples were decrosslinked at 65ºC overnight and deproteinized. Following phenol-chloroform extraction, both I, + and - were ethanol precipitated. The DNA pellet was finally resuspended in 20 ul of MilliQ water. All three samples (I, +, -) were subjected to PCR reactions.

ChIP-qPCR analysis

The input and IP DNA were diluted appropriately and qPCR reactions were set up using primers listed in Table 6.2. The CENPA/ Mtw1/Orc4 enrichment was determined by the percentage input method. Two- way ANOVA and Bonferonni post tests were performed to determine statistical significance. All the percent IP values represented in the graphs comparing enrichment values in CM-Uri and CM+5'FOA are the ratio of percent IP of the regions indicated to the corresponding values of *CEN1*, which was used as an internal control to estimate the efficiency of the pulldown. For the ChIP experiments with the neocentromere strains, these values have not been normalised to *CEN1*.

ChIP-sequencing analysis

For the CENPA ChIP-seq, immunoprecipitated DNA and the corresponding DNA from whole cell extracts from strains LSK450 and LSK465 were quantified using Qubit before proceeding for library preparation. Around 5 ng ChIP and total DNA were used to prepare sequencing libraries using NEBNext Ultra DNA library preparation kit for Illumina (NEB,

USA). The library quality and quantity were checked using Qubit HS DNA (Thermo Fisher Scientific, USA) and Bioanalyzer DNA high sensitivity kits (Agilent Technologies, USA) respectively. The QC passed libraries were sequenced on Illumina HiSeq 2500 (Illumina Inc., USA). HiSeq rapid cluster and SBS kits v2 were to generate 50 bp single end reads. The reads were independently aligned onto the *C. albicans* SC5314 reference genome (v. 21) and a genome with an altered version of Chr7 using bowtie2 (v. 2.3.2) aligner. For the Orc4 ChIP-seq, subtracted reads were aligned onto the *C. albicans* SC5314 reference genome (v. 21) using bowtie2 (v. 2.3.2) aligner (Langmead, Trapnell et al. 2009). More than 95% of the reads mapped onto the reference genome (Control:97.74%; IP:96.13%). All the alignment files (BAM) were processed to remove PCR duplicate reads using Mark Duplicates module of Picard tools. These processed BAM files were further taken for identification of peaks by MACS2. These peaks were annotated with the *C. albicans* SC5314 reference and altered assembly annotation files. Visualisation of the aligned reads (BAM files) on the reference genome was performed using Integrative Genome Viewer (IGV) (https://software.broadinstitute.org/software/igv/).

Hi-C analysis

Wild-type *C. albicans* Hi-C data was downloaded from PRJNA308106 (Burrack, Hutton et al. 2016). FASTQ files containing 2 X 80bp paired-end (PE) reads were analyzed using hiclib package (http://mirnylab.bitbucket.org/hiclib/) (Imakaev, Fudenberg et al. 2012). First, each side of the PE reads was aligned separately to *C. albicans* reference genome (Ca21) using Bowtie 2 (Langmead and Salzberg 2012) with default parameters except for --very-sensitive option. This step was executed iteratively (iterative mapping) in which 3´ truncated reads was aligned to reference genome, starting from first 20 bases with increment of 5 bases in subsequent iteration till it reached to the end of read length. The reads which were uniquely mapped with MAPQ score ≥ 1 were saved at each iteration and rest were subsequently analyzed in next iteration. The alignment results from both sides were paired, keeping those reads which had at least one side aligned and were assigned to restriction fragments. The output read pairs and their alignment information as well as the assigned restriction fragments were saved in HDF5 file format. The fragment filter then removed reads those have: 1) only one side aligned; 2) both sides aligned to same restriction fragment; 3) two sides which were too close to each other. PCR redundancies (duplicates) were also removed and all the unique

valid pairs were binned into genomic intervals of 2 kb-5 kb (bin size). The resulted symmetric matrix was processed for bin filtering step, including removal of bins with <50% sequence information in reference genome and removal of 1% bins with low read coverage. Diagonal bins were excluded from further downstream analysis. The genome-wide interaction matrix was generated following bin bias correction as described (Imakaev, Fudenberg et al. 2012). The interaction matrix was then converted to a contact probability matrix where the sum of values in each row/column approached 1. The 3C profile anchored on a bin was generated using single row/column containing the anchor from the matrix. To plot distance-dependent contact probability curves, the mean *cis* contact probabilities (excluding the bins with 0 values) were calculated for each distance (bin size=2kb) for pericentromeric and non-pericentromeric regions. Mann-Whitney U test was performed for pericentromeric and all *cis* interactions, as well as for pericentromeric and non-pericentromeric interactions.

For the Orc4 binding regions, Hi-C interactions were analyzed according to the chromosome coordinates, different modes identified by DIVERSITY and also based on replication timing (early and late). The heatmap for the full genome was plotted using log-scaled values with a pseudocount of 0.000001 (10^{-6}). The heatmap for the "ORC-only" was plotted using values for the 2 kb windows overlapping with the midpoints of the origins, using the same scaling and colour scale as the full-genome heatmap. The violin plots were calculated for 1,000 randomizations of each dataset, where for each randomization, the chromosomal distribution and lengths of the regions were preserved.

Nucleosome positioning:

NucTools (https://homeveg.github.io/nuctools/) was used to analyse *C. albicans* MNase-seq data obtained from single end sequence data (Tsankov, Thompson et al. 2010). Bowtie2 was used to map single end reads onto *C. albicans* reference genome assembly version 21. Sorted .bam file was converted to .bed file using bowtie2bed.pl code provided in Nuctools. The single end reads were extended to an average of 150bp. All BED files were converted to occupancy files averaging nucleosomes occupancy values at each base, which were visualized in IGV. The data range cut-off was set to a Min 130 and Mid 130.

SIDD and GC content analysis:

We used WEBSIDD server to determine SIDD profile of the origins, using fixed window sizes. The online tool (http://orange.genomecenter.ucdavis.edu/benham/sidd/index.html) (Bi and Benham 2004) computes destabilization properties on the basis of denaturing behaviour of all the neighbouring base pairs experiencing superhelical stress as it occurs in vivo in a topologically constrained domain. Briefly, *C. albicans* Assembly 21 chromosome 7 was fragmented into 5 kb regions with 2.5 kb overlap. G(x) values corresponding to individual nucleotide were obtained from the WEBSIDD tool and merged to get a consolidated profile for Chr7 in .wig format, which was visualized using IGV. GC content for each chromosome was calculated with 250 bp sliding window, results were exported into .wig format for visualization using IGV.

Motif analysis for origins

For motif analysis, the *de novo* motif discovery tool DIVERSITY (Mitra, Biswas et al. 2018) was used with default web-server options on the 417 Orc4 ChIP-seq peaks. DIVERSITY is specially developed for ChIP-seq experiments profiling proteins that may bind DNA in more than one way.

Replication timing analysis

To analyze the replication timing of the ORC binding regions, fully processed timing data available in GSE17963_final_data.txt (Koren, Tsai et al. 2010) was used. A larger replication time value implies earlier replication. All the 414 genomic origins were aligned according to their timing scores, and categorized as early (first 207) and late (last 207) origins.

Table 6.1. *C. albicans* **strains used in the study.**

Name	Genotype	Description	Reference
J200	*Δura3::imm434/ Δura3::imm434 Δhis1::hisG/ Δhis1::hisG arg4::HIS1/ARG4 CSE4/CSE4-TAP(NAT)*	*CSE4-TAP(NAT)* in RM1000AH	(Thakur and Sanyal 2013)
8675	*Δura3::imm434/Δura3::imm434, Δhis1::hisG/Δhis1::hisG ,Δarg4::hisG/Δarg4::hisG, CSE4-GFP-CSE4/CSE4*	*CSE4-GFP-CSE4/CSE4*	(Joglekar, Bouck et al. 2008)
LSK401	*Δura3::imm434/ Δura3::imm434 Δhis1::hisG/ Δhis1::hisG arg4::HIS1/ARG4 5L/5L::URA3 CSE4/CSE4-TAP(NAT)*	5L_T1	This study
LSK402	*Δura3::imm434/ Δura3::imm434 Δhis1::hisG/ Δhis1::hisG arg4::HIS1/ARG4 5L/5L::URA3 CSE4/CSE4-TAP(NAT)*	5L_T2	This study
LSK403	*Δura3::imm434/ Δura3::imm434 Δhis1::hisG/ Δhis1::hisG arg4::HIS1/ARG4 5L/5L::URA3 CSE4/CSE4-TAP(NAT)*	5L_T3	This study
LSK404	*Δura3::imm434/ Δura3::imm434 Δhis1::hisG/ Δhis1::hisG arg4::HIS1/ARG4 4L/4L::URA3 CSE4/CSE4-TAP(NAT)*	4L_T1	This study
LSK405	*Δura3::imm434/ Δura3::imm434 Δhis1::hisG/ Δhis1::hisG arg4::HIS1/ARG4 4L/4L::URA3 CSE4/CSE4-TAP(NAT)*	4L_T2	This study
LSK406	*Δura3::imm434/ Δura3::imm434 Δhis1::hisG/ Δhis1::hisG arg4::HIS1/ARG4 4L/4L::URA3 CSE4/CSE4 -TAP(NAT)*	4L_T3	This study
LSK407	*Δura3::imm434/ Δura3::imm434 Δhis1::hisG/ Δhis1::hisG arg4::HIS1/ARG4 3L/3L::URA3 CSE4/CSE4-TAP(NAT)*	3L_T1	This study
LSK408	*Δura3::imm434/ Δura3::imm434 Δhis1::hisG/ Δhis1::hisG arg4::HIS1/ARG4 3L/3L::URA3 CSE4/CSE4-TAP(NAT)*	3L_T2	This study
LSK409	*Δura3::imm434/ Δura3::imm434 Δhis1::hisG/ Δhis1::hisG arg4::HIS1/ARG4 3L/3L::URA3 CSE4/CSE4-TAP(NAT)*	3L_T3	This study

LSK410	*Δura3::imm434/ Δura3::imm434 Δhis1::hisG/ Δhis1::hisG arg4::HIS1/ARG4 2L/2L::URA3 CSE4/CSE4-TAP(NAT)*	2L_T1	This study
LSK411	*Δura3::imm434/ Δura3::imm434 Δhis1::hisG/ Δhis1::hisG arg4::HIS1/ARG4 2L/2L::URA3 CSE4/CSE4-TAP(NAT)*	2L_T2	This study
LSK412	*Δura3::imm434/ Δura3::imm434 Δhis1::hisG/ Δhis1::hisG arg4::HIS1/ARG4 2L/2L::URA3 CSE4/CSE4-TAP(NAT)*	2L_T3	This study
LSK413	*Δura3::imm434/ Δura3::imm434 Δhis1::hisG/ Δhis1::hisG arg4::HIS1/ARG4 1L/1L::URA3 CSE4/CSE4-TAP(NAT)*	1L_T1	This study
LSK414	*Δura3::imm434/ Δura3::imm434 Δhis1::hisG/ Δhis1::hisG arg4::HIS1/ARG4 1L/1L::URA3 CSE4/CSE4-TAP(NAT)*	1L_T2	This study
LSK415	*Δura3::imm434/ Δura3::imm434 Δhis1::hisG/ Δhis1::hisG arg4::HIS1/ARG4 1L/1L::URA3 CSE4/CSE4TAP(NAT)*	1L_T3	This study
J151	*Δura3::imm434/ Δura3::imm434 Δhis1::hisG/ Δhis1::hisG arg4::HIS1/ARG4 CEN7/CEN7::URA3 CSE4/CSE4-TAP(NAT)*	CEN7::URA3_T1	(Thakur and Sanyal 2013)
J153	*Δura3::imm434/ Δura3::imm434 Δhis1::hisG/ Δhis1::hisG arg4::HIS1/ARG4 CEN7/CEN7::URA3 CSE4/CSE4-TAP(NAT)*	CEN7::URA3_T2	(Thakur and Sanyal 2013)
J154	*Δura3::imm434/ Δura3::imm434 Δhis1::hisG/ Δhis1::hisG arg4::HIS1/ARG4 CEN7/CEN7::URA3 CSE4/CSE4-TAP(NAT)*	CEN7::URA3_T3	(Thakur and Sanyal 2013)
LSK416	*Δura3::imm434/ Δura3::imm434 Δhis1::hisG/ Δhis1::hisG arg4::HIS1/ARG4 1R/1R::URA3 CSE4/CSE4-TAP(NAT)*	1R_T1	This study
LSK417	*Δura3::imm434/ Δura3::imm434 Δhis1::hisG/ Δhis1::hisG arg4::HIS1/ARG4 1R/1R::URA3 CSE4/CSE4-TAP(NAT)*	1R_T2	This study
LSK418	*Δura3::imm434/ Δura3::imm434 Δhis1::hisG/ Δhis1::hisG arg4::HIS1/ARG4 1R/1R::URA3 CSE4/CSE4-TAP(NAT)*	1R_T3	This study

LSK419	*Δura3::imm434/ Δura3::imm434 Δhis1::hisG/ Δhis1::hisG arg4::HIS1/ARG4 2R/2R::URA3 CSE4/CSE4-TAP(NAT)*	2R_T1	This study
LSK420	*Δura3::imm434/ Δura3::imm434 Δhis1::hisG/ Δhis1::hisG arg4::HIS1/ARG4 2R/2R::URA3 CSE4/CSE4-TAP(NAT)*	2R_T2	This study
LSK421	*Δura3::imm434/ Δura3::imm434 Δhis1::hisG/ Δhis1::hisG arg4::HIS1/ARG4 2R/2R::URA3 CSE4/CSE4-TAP(NAT)*	2R_T3	This study
LSK422	*Δura3::imm434/ Δura3::imm434 Δhis1::hisG/ Δhis1::hisG arg4::HIS1/ARG4 3R/3R::URA3 CSE4/CSE4-TAP(NAT)*	3R_T1	This study
LSK423	*Δura3::imm434/ Δura3::imm434 Δhis1::hisG/ Δhis1::hisG arg4::HIS1/ARG4 3R/3R::URA3 CSE4/CSE4-TAP(NAT)*	3R_T2	This study
LSK424	*Δura3::imm434/ Δura3::imm434 Δhis1::hisG/ Δhis1::hisG arg4::HIS1/ARG4 3R/3R::URA3 CSE4/CSE4-TAP(NAT)*	3R_T3	This study
LSK425	*Δura3::imm434/ Δura3::imm434 Δhis1::hisG/ Δhis1::hisG arg4::HIS1/ARG4 4R/4R::URA3 CSE4/CSE4-TAP(NAT)*	4R_T1	This study
LSK426	*Δura3::imm434/ Δura3::imm434 Δhis1::hisG/ Δhis1::hisG arg4::HIS1/ARG4 4R/4R::URA3 CSE4/CSE4-TAP(NAT)*	4R_T2	This study
LSK427	*Δura3::imm434/ Δura3::imm434 Δhis1::hisG/ Δhis1::hisG arg4::HIS1/ARG4 4R/4R::URA3 CSE4/CSE4-TAP(NAT)*	4R_T3	This study
LSK427	*Δura3::imm434/ Δura3::imm434 Δhis1::hisG/ Δhis1::hisG arg4::HIS1/ARG4 5R/5R::URA3 CSE4/CSE4-TAP(NAT)*	5R_T1	This study
LSK428	*Δura3::imm434/ Δura3::imm434 Δhis1::hisG/ Δhis1::hisG arg4::HIS1/ARG4 5R/5R::URA3 CSE4/CSE4-TAP(NAT)*	5R_T2	This study
LSK430	*Δura3::imm434/ Δura3::imm434 Δhis1::hisG/ Δhis1::hisG arg4::HIS1/ARG4 CEN5/CEN5-7 kb_right ::URA3 CSE4/CSE4TAP(NAT)*	CEN5_T1	This study
LSK431	*Δura3::imm434/ Δura3::imm434 Δhis1::hisG/ Δhis1::hisG arg4::HIS1/ARG4 CEN5/CEN5-7 kb_right ::URA3 CSE4/CSE4-TAP(NAT)*	CEN5_T2	This study

LSK432	*Δura3::imm434/ Δura3::imm434 Δhis1::hisG/ Δhis1::hisG arg4::HISl/ARG4 CEN5/CEN5-7 kb_right ::URA3 CSE4/CSE4-TAP(NAT)*	CEN5_T3	This study
LSK433	*Δura3::imm434/ Δura3::imm434 Δhis1::hisG/ Δhis1::hisG arg4::HISl/ARG4 CEN7::URA3_ 127 kb farCEN/ CEN7 CSE4/CSE4-TAP(NAT)*	FAR URA_T1	This study
LSK434	*Δura3::imm434/ Δura3::imm434 Δhis1::hisG/ Δhis1::hisG arg4::HISl/ARG4 CEN7::URA3_ 127 kb farCEN/ CEN7 CSE4/CSE4-TAP(NAT)*	FAR URA_T2	This study
LSK435	*Δura3::imm434/ Δura3::imm434 Δhis1::hisG/ Δhis1::hisG arg4::HISl/ARG4 CEN7::URA3_ 127 kb farCEN/ CEN7 CSE4/CSE4-TAP(NAT)*	FAR URA_T3	This study
LSK436	*Δura3::imm434/ Δura3::imm434 Δhis1::hisG/ Δhis1::hisG arg4::HISl/ARG4 MTW1/MTW1-TAP(NAT)*	MTW1-TAP IN RM1000AH	This study
LSK437	*Δura3::imm434/ Δura3::imm434 Δhis1::hisG/ Δhis1::hisG arg4::HISl/ARG4 4L/4L::URA3 MTW1/MTW1-TAP(NAT)*	MTW1-TAP IN 4L_T1	This study
LSK438	*Δura3::imm434/ Δura3::imm434 Δhis1::hisG/ Δhis1::hisG arg4::HISl/ARG4 4L/4L::URA3 MTW1/MTW1-TAP(NAT)*	MTW1-TAP IN 4L_T2	This study
LSK439	*Δura3::imm434/ Δura3::imm434 Δhis1::hisG/ Δhis1::hisG arg4::HISl/ARG4 4L/4L::URA3 MTW1/MTW1-TAP(NAT)*	MTW1-TAP IN 4L_T3	This study
LSK440	*Δura3::imm434/ Δura3::imm434 Δhis1::hisG/ Δhis1::hisG arg4::HISl/ARG4 4R/4R::URA3 MTW1/MTW1-TAP(NAT)*	MTW1-TAP IN 4R_T1	This study
LSK441	*Δura3::imm434/ Δura3::imm434 Δhis1::hisG/ Δhis1::hisG arg4::HISl/ARG4 4R/4R::URA3 MTW1/MTW1-TAP(NAT)*	MTW1-TAP IN 4R_T2	This study
LSK442	*Δura3::imm434/ Δura3::imm434 Δhis1::hisG/ Δhis1::hisG arg4::HISl/ARG4 4R/4R::URA3 MTW1/MTW- TAP(NAT)*	MTW1-TAP IN 4R_T3	This study
YJB8675	*Δura3::imm434/Δura3::imm434, Δhis1::hisG/Δhis1::hisG ,Δarg4::hisG/Δarg4::hisG, CSE4/CSE4-GFP-CSE4*	Cse4-GFP	This study
LSK443	*Δura3::imm434/Δura3::imm434, Δhis1::hisG/Δhis1::hisG ,Δarg4::hisG/Δarg4::hisG, 4L/4L::URA3 CSE4/CSE4-GFP-CSE4*	4L in Cse4-GFP_T1	This study

LSK444	*Δura3::imm434/Δura3::imm434, Δhis1::hisG/Δhis1::hisG ,Δarg4::hisG/Δarg4::hisG, 4L/4L::URA3 CSE4/CSE4-GFP-CSE4*	4L in Cse4-GFP_T2	This study
LSK445	*Δura3::imm434/Δura3::imm434, Δhis1::hisG/Δhis1::hisG ,Δarg4::hisG/Δarg4::hisG, 4L/4L::URA3 CSE4/CSE4-GFP-CSE4*	4L in Cse4-GFP_T3	This study
LSK446	*Δura3::imm434/Δura3::imm434, Δhis1::hisG/Δhis1::hisG ,Δarg4::hisG/Δarg4::hisG, 4L/4L::URA3 CSE4/CSE4-GFP-CSE4 CEN7/CEN7::HIS1*	CEN7 del in 4L (FOA[s], in *cis*)	This study
LSK447	*Δura3::imm434/Δura3::imm434, Δhis1::hisG/Δhis1::hisG ,Δarg4::hisG/Δarg4::hisG, 4L/4L::URA3 CSE4/CSE4-GFP-CSE4 CEN7/CEN7::HIS1*	CEN7 del in 4L (FOA[s], in *cis*)	This study
LSK448	*Δura3::imm434/Δura3::imm434, Δhis1::hisG/Δhis1::hisG ,Δarg4::hisG/Δarg4::hisG, 4L/4L::URA3 CSE4/CSE4-GFP-CSE4 CEN7/CEN7::HIS1*	CEN7 del in 4L (FOA[s], in *trans*)	This study
LSK449	*Δura3::imm434/Δura3::imm434, Δhis1::hisG/Δhis1::hisG ,Δarg4::hisG/Δarg4::hisG, 4L/4L::URA3 CSE4/CSE4-GFP-CSE4 CEN7/CEN7::HIS1*	CEN7 del in 4L (FOA[s], in *trans*)	This study
LSK450	*Δura3::imm434/Δura3::imm434, Δhis1::hisG/Δhis1::hisG ,Δarg4::hisG/Δarg4::hisG, 4L/4L::URA3 CSE4/CSE4-GFP-CSE4 CEN7/CEN7::HIS1*	CEN7 del in 4L (FOA[r], in *cis*)	This study
LSK451	*Δura3::imm434/Δura3::imm434, Δhis1::hisG/Δhis1::hisG ,Δarg4::hisG/Δarg4::hisG, 4L/4L::URA3 CSE4/CSE4-GFP-CSE4 CEN7/CEN7::HIS1*	CEN7 del in 4L (FOA[r], in *cis*)	This study
LSK452	*Δura3::imm434/Δura3::imm434, Δhis1::hisG/Δhis1::hisG ,Δarg4::hisG/Δarg4::hisG, 4L/4L::URA3 CSE4/CSE4-GFP-CSE4 CEN7/CEN7::HIS1*	CEN7 del in 4L (FOA[r], in *trans*)	This study
LSK453	*Δura3::imm434/Δura3::imm434, Δhis1::hisG/Δhis1::hisG ,Δarg4::hisG/Δarg4::hisG, CEN7/CEN7::HIS1::URA3_7.7kb left CSE4-GFP-CSE4/CSE4*	CEN7 del in 4L (FOA[r], in *trans*)	This study
LSK454	*Δura3::imm434/Δura3::imm434, Δhis1::hisG/Δhis1::hisG ,Δarg4::hisG/Δarg4::hisG,*	CEN7 del in 4L (FOA[r], in *trans*)	This study

	4L/4L::URA3 CSE4/CSE4-GFP-CSE4 CEN7/CEN7::HIS1		
LSK455	*Δura3::imm434/Δura3::imm434, Δhis1::hisG/Δhis1::hisG ,Δarg4::hisG/Δarg4::hisG, 4L/4L::URA3 CSE4/CSE4-GFP-CSE4 CEN7/CEN7::HIS1*	CEN7 del in 4L (FOA^r, in *trans*)	This study
LSK456	*Δura3::imm434/Δura3::imm434, Δhis1::hisG/Δhis1::hisG ,Δarg4::hisG/Δarg4::hisG, 4R/4R::URA3 CSE4/CSE4-GFP-CSE4*	4R in Cse4-GFP_T1	This study
LSK457	*Δura3::imm434/Δura3::imm434, Δhis1::hisG/Δhis1::hisG ,Δarg4::hisG/Δarg4::hisG, 4R/4R::URA3 CSE4/CSE4-GFP-CSE4*	4R in Cse4-GFP_T2	This study
LSK458	*Δura3::imm434/Δura3::imm434, Δhis1::hisG/Δhis1::hisG ,Δarg4::hisG/Δarg4::hisG, 4R/4R::URA3 CSE4/CSE4-GFP-CSE4*	4R in Cse4-GFP_T3	This study
LSK459	*Δura3::imm434/Δura3::imm434, Δhis1::hisG/Δhis1::hisG ,Δarg4::hisG/Δarg4::hisG, 4R/4R::URA3 CSE4/CSE4-GFP-CSE4 CEN7/CEN7::HIS1*	CEN7 del in 4R (FOA^s, in *cis*)	This study
LSK460	*Δura3::imm434/Δura3::imm434, Δhis1::hisG/Δhis1::hisG ,Δarg4::hisG/Δarg4::hisG, 4R/4R::URA3 CSE4/CSE4-GFP-CSE4 CEN7/CEN7::HIS1*	CEN7 del in 4R (FOA^s, in *cis*)	This study
LSK461	*Δura3::imm434/Δura3::imm434, Δhis1::hisG/Δhis1::hisG ,Δarg4::hisG/Δarg4::hisG, 4R/4R::URA3 CSE4/CSE4-GFP-CSE4 CEN7/CEN7::HIS1*	CEN7 del in 4R (FOA^s, in *cis*)	This study
LSK462	*Δura3::imm434/Δura3::imm434, Δhis1::hisG/Δhis1::hisG ,Δarg4::hisG/Δarg4::hisG, 4R/4R::URA3 CSE4/CSE4-GFP-CSE4 CEN7/CEN7::HIS1*	CEN7 del in 4R (FOA^s, in *trans*)	This study
LSK463	*Δura3::imm434/Δura3::imm434, Δhis1::hisG/Δhis1::hisG ,Δarg4::hisG/Δarg4::hisG, 4R/4R::URA3 CSE4/CSE4-GFP-CSE4 CEN7/CEN7::HIS1*	CEN7 del in 4R (FOA^s, in *trans*)	This study
LSK464	*Δura3::imm434/Δura3::imm434, Δhis1::hisG/Δhis1::hisG ,Δarg4::hisG/Δarg4::hisG, 4R/4R::URA3 CSE4/CSE4-GFP-CSE4 CEN7/CEN7::HIS1*	CEN7 del in 4R (FOA^r, in *cis*)	This study

LSK465	*Δura3::imm434/Δura3::imm434, Δhis1::hisG/Δhis1::hisG ,Δarg4::hisG/Δarg4::hisG, 4R/4R::URA3 CSE4/CSE4-GFP-CSE4 CEN7/CEN7::HIS1*	CEN7 del in 4R (FOA^r, in *cis*)	This study
LSK466	*Δura3::imm434/Δura3::imm434, Δhis1::hisG/Δhis1::hisG ,Δarg4::hisG/Δarg4::hisG, 4R/4R::URA3 CSE4/CSE4-GFP-CSE4 CEN7/CEN7::HIS1*	CEN7 del in 4R (FOA^r, in *cis*)	This study
LSK467	*Δura3::imm434/Δura3::imm434, Δhis1::hisG/Δhis1::hisG ,Δarg4::hisG/Δarg4::hisG, 4R/4R::URA3 CSE4/CSE4-GFP-CSE4 CEN7/CEN7::HIS1*	CEN7 del in 4R (FOA^r, in *trans*)	This study
LSK468	*Δura3::imm434/Δura3::imm434, Δhis1::hisG/Δhis1::hisG ,Δarg4::hisG/Δarg4::hisG, 4R/4R::URA3 CSE4/CSE4-GFP-CSE4 CEN7/CEN7::HIS1*	CEN7 del in 4R (FOA^r, in *trans*)	This study
LSK471	*Δura3::imm434/Δura3::imm434, Δhis1::hisG/Δhis1::hisG ,Δarg4::hisG/Δarg4::hisG, 4L/4L::URA3 CSE4/CSE4-GFP-CSE4 CEN7/CEN7::HIS1MTW1/MTW1-TAP(NAT)*	CEN7 del in 4L (FOA^r, in *cis*) MTW1-TAP	This study
LSK472	*Δura3::imm434/Δura3::imm434, Δhis1::hisG/Δhis1::hisG ,Δarg4::hisG/Δarg4::hisG, 4L/4L::URA3 CSE4/CSE4-GFP-CSE4 CEN7/CEN7::HIS1MTW1/MTW1-TAP(NAT)*	CEN7 del in 4L (FOA^r, in *cis*) MTW1-TAP	This study
LSK475	*Δura3::imm434/Δura3::imm434, Δhis1::hisG/Δhis1::hisG ,Δarg4::hisG/Δarg4::hisG, 4R/4R::URA3 CSE4/CSE4-GFP-CSE4 CEN7/CEN7::HIS1MTW1/MTW1-TAP(NAT)*	CEN7 del in 4R (FOA^r, in *cis*) MTW1-TAP	This study
LSK476	*Δura3::imm434/Δura3::imm434, Δhis1::hisG/Δhis1::hisG ,Δarg4::hisG/Δarg4::hisG, 4R/4R::URA3 CSE4/CSE4-GFP-CSE4 CEN7/CEN7::HIS1MTW1/MTW1-TAP(NAT)*	CEN7 del in 4R (FOA^r, in *cis*) MTW1-TAP	This study
LSK301	*Δura3::imm434/Δura3::imm434, Δhis1::hisG/Δhis1::hisG ,Δarg4::hisG/Δarg4::hisG, Δleu2::hisG/Δleu2::hisG, MCM2::NAT/MCM2*	*MCM2* heterozygous null (SN148)	This study
LSK302	*Δura3::imm434/Δura3::imm434, Δhis1::hisG/Δhis1::hisG ,Δarg4::hisG/Δarg4::hisG, Δleu2::hisG/Δleu2::hisG, MCM2::FRT/MCM2*	*MCM2* heterozygous null (SN148)	This study
LSK303	*Δura3::imm434/Δura3::imm434, Δhis1::hisG/Δhis1::hisG,Δarg4::hisG/Δarg4::hisG,*	*mcm2* conditional mutant (SN148)	This study

	Δleu2::hisG/Δleu2::hisG, MCM2:: FRT/MET3prMCM2		
LSK304	*Δura3::imm434/Δura3::imm434, Δhis1::hisG/Δhis1::hisG ,Δarg4::hisG/Δarg4::hisG, Δleu2::hisG/Δleu2::hisG, MCM2:: FRT/MET3prMCM2*	*mcm2* conditional mutant (SN148)	This study
LSK305	*Δura3::imm434/Δura3::imm434, Δhis1::hisG/Δhis1::hisG ,Δarg4::hisG/Δarg4::hisG, Δleu2::hisG/Δleu2::hisG, MCM2:: FRT/MET3prMCM2*	*mcm2* conditional mutant (SN148)	This study
LSK306	*Δura3::imm434/Δura3::imm434, Δhis1::hisG/Δhis1::hisG ,Δarg4::hisG/Δarg4::hisG, Δleu2::hisG/Δleu2::hisG, MCM2:: FRT/MET3prMCM2 CSE4 TAP(HIS)/CSE4*	*mcm2* conditional mutant (SN148) CENPA-Prot A	This study
LSK307	*Δura3::imm434/Δura3::imm434, Δhis1::hisG/Δhis1::hisG ,Δarg4::hisG/Δarg4::hisG, Δleu2::hisG/Δleu2::hisG, MCM2:: FRT/MET3prMCM2 CSE4 TAP(HIS)/CSE4*	*mcm2* conditional mutant (SN148) CENPA-Prot A	This study
LSK308	*Δura3::imm434/Δura3::imm434, Δhis1::hisG/Δhis1::hisG ,Δarg4::hisG/Δarg4::hisG, Δleu2::hisG/Δleu2::hisG, MCM2:: FRT/MET3prMCM2 CSE4 TAP(HIS)/CSE4*	*mcm2* conditional mutant (SN148) CENPA-Prot A	This study
LSK309	*Δura3::imm434/Δura3::imm434, Δhis1::hisG/Δhis1::hisG ,Δarg4::hisG/Δarg4::hisG, MCM2::NAT/MCM2 CSE4-GFP-CSE4/CSE4*	*MCM2* heterozygous null (8675)	This study
LSK310	*Δura3::imm434/Δura3::imm434, Δhis1::hisG/Δhis1::hisG ,Δarg4::hisG/Δarg4::hisG, MCM2::NAT/MCM2 CSE4-GFP-CSE4/CSE4*	*MCM2* heterozygous null (8675)	This study
LSK311	*Δura3::imm434/Δura3::imm434, Δhis1::hisG/Δhis1::hisG ,Δarg4::hisG/Δarg4::hisG, MCM2::FRT/MET3prMCM2 CSE4-GFP-CSE4/CSE4*	*mcm2* conditional mutant (8675)	This study
LSK312	*Δura3::imm434/Δura3::imm434, Δhis1::hisG/Δhis1::hisG ,Δarg4::hisG/Δarg4::hisG, MCM2::FRT/MET3prMCM2 CSE4-GFP-CSE4/CSE4*	*mcm2* conditional mutant (8675)	This study
LSK313	*Δura3::imm434/Δura3::imm434, Δhis1::hisG/Δhis1::hisG ,Δarg4::hisG/Δarg4::hisG, MCM2::FRT/MET3prMCM2 CSE4-GFP-CSE4/CSE4*	*mcm2* conditional mutant (8675)	This study

LSK320	*Δura3::imm434/Δura3::imm434, Δhis1::hisG/Δhis1::hisG ,Δarg4::hisG/Δarg4::hisG, Δleu2::hisG/Δleu2::hisG, ORC4::NAT/ORC4*	*ORC4* heterozygous null (SN148)	This study
LSK321	*Δura3::imm434/Δura3::imm434, Δhis1::hisG/Δhis1::hisG ,Δarg4::hisG/Δarg4::hisG, Δleu2::hisG/Δleu2::hisG, ORC4::NAT/ORC4*	*ORC4* heterozygous null (SN148)	This study
LSK322	*Δura3::imm434/Δura3::imm434, Δhis1::hisG/Δhis1::hisG ,Δarg4::hisG/Δarg4::hisG, Δleu2::hisG/Δleu2::hisG, ORC4::FRT/MET3prORC4*	*orc4* conditional mutant (SN148)	This study
LSK323	*Δura3::imm434/Δura3::imm434, Δhis1::hisG/Δhis1::hisG ,Δarg4::hisG/Δarg4::hisG, Δleu2::hisG/Δleu2::hisG, ORC4::FRT/MET3prORC4*	*orc4* conditional mutant (SN148)	This study
LSK324	*Δura3::imm434/Δura3::imm434, Δhis1::hisG/Δhis1::hisG ,Δarg4::hisG/Δarg4::hisG, Δleu2::hisG/Δleu2::hisG, ORC4::FRT/MET3prORC4*	*orc4* conditional mutant (SN148)	This study
LSK325	*Δura3::imm434/Δura3::imm434, Δhis1::hisG/Δhis1::hisG ,Δarg4::hisG/Δarg4::hisG, Δleu2::hisG/Δleu2::hisG, ORC4:: FRT/MET3prORC4 CSE4 TAP(HIS)/CSE4*	*orc4* conditional mutant (SN148) CENPA-Prot A	This study
LSK326	*Δura3::imm434/Δura3::imm434, Δhis1::hisG/Δhis1::hisG ,Δarg4::hisG/Δarg4::hisG, Δleu2::hisG/Δleu2::hisG, ORC4:: FRT/MET3prORC4 CSE4 TAP(HIS)/CSE4*	*orc4* conditional mutant (SN148) CENPA-Prot A	This study
LSK327	*Δura3::imm434/Δura3::imm434, Δhis1::hisG/Δhis1::hisG ,Δarg4::hisG/Δarg4::hisG, Δleu2::hisG/Δleu2::hisG, ORC4:: FRT/MET3prORC4 CSE4 TAP(HIS)/CSE4*	*orc4* conditional mutant (SN148) CENPA-Prot A	This study
LSK328	*Δura3::imm434/Δura3::imm434, Δhis1::hisG/Δhis1::hisG ,Δarg4::hisG/Δarg4::hisG, ORC4:NAT/ORC4 CSE4-GFP-CSE4/CSE4*	*ORC4* heterozygous null (8675)	This study
LSK329	*Δura3::imm434/Δura3::imm434, Δhis1::hisG/Δhis1::hisG ,Δarg4::hisG/Δarg4::hisG, ORC4:NAT/ORC4 CSE4-GFP-CSE4/CSE4*	*ORC4* heterozygous null (8675)	This study
LSK330	*Δura3::imm434/Δura3::imm434, Δhis1::hisG/Δhis1::hisG ,Δarg4::hisG/Δarg4::hisG, ORC4::FRT/MET3prORC4 CSE4-GFP-CSE4/CSE4*	*orc4* conditional mutant (8675)	This study

LSK331	*Δura3::imm434/Δura3::imm434,* *Δhis1::hisG/Δhis1::hisG ,Δarg4::hisG/Δarg4::hisG,* *ORC4::FRT/MET3prORC4 CSE4-GFP-CSE4/CSE4*	*orc4* conditional mutant (8675)	This study
LSK332	*Δura3::imm434/Δura3::imm434,* *Δhis1::hisG/Δhis1::hisG ,Δarg4::hisG/Δarg4::hisG,* *ORC4::FRT/MET3prORC4 CSE4-GFP-CSE4/CSE4*	*orc4* conditional mutant (8675)	This study
CAKS3b	*Δura3::imm434/ Δura3::imm434 Δhis1::hisG/* *Δhis1::hisG Δarg4::hisG/ Δarg4::hisG* *CSE4::PCK1prCSE4/ cse4::hisG:URA:hisG*	CENPA depletion	(Sanyal and Carbon 2002)

Table 6.2 Oligonucleotide primers used in the study.

Name	Sequence	Description
URA3 EXT HSP2_FP	GTTTCAGAATCCGAAAAAGTGACGAAACTTATCAT AATTGTACGAATATTCTTATCAAACACACCCTGAG CTTCCGGATAATAGGAATTG	Cassette primers for *URA3* integrated 10 kb left of CEN7
URA3 EXT HSP2_RP	GTTGCTCGAGGTTAGAGTCTATCTTGAAAAATTTT GTACATACAAACTGATATAACTCGACAATGGTCTT AGAAGGACCACCTTTGATTG	
URA3 AT HSP2_FP	CTCAAAAATACTTTAACAAACGGGTATATTGCTGA TATTCTGATTAAAACATTTGATCGTTTTATGTGAGC TTCCGGATAATAGGAATTG	Cassette primers for *URA3* integrated 7.7 kb left of CEN7
URA3 AT HSP2_RP	CTTAACCCCAGACAGTTTTAACAATTTAGACACTA CTACTAATTGCAACGTACTAACTAGTGAAACCCTT AGAAGGACCACCTTTGATTG	
19.6520_AvrII _F	AAACCCCCTAGGTTGCGAATATCTATTG	
19.6520_HindI II_MluI_R	AAACCCAAGCTTACGCGTAATGGTCCCATCAGCAG TGCA	Construction of pFA-URA3-I-SceI-TS-Orf 19.6520/65
19.6522_HpaI _MluI_F	CCCAAAGTTAACACGCGTCTGCCAACAAGAATGC AACT	
19.6522_SacII _R	CCCAAACCGCGGTATATTTTTGTTGTATCAGAATC CTACGCC	
L1_URA INT_FP	CACATATTTTTACTTTCTGTATTATTCAGATCTTTA CTCGTTGAAAAAAAATTTTTTTTTTCAAAAGCTTCC GGATAATAGGAATTG	Cassette primers for *URA3* integrated 3.5 kb left of CEN7

L1_URA INT_RP	GATGTAGTTGTATCTTTAATATCACAGTTATGATA AGGGTCGTGTATATGTGAACATGGATTTGCTTAGA AGGACCACCTTTGATTG	
PJ71	TGCTTACCATAATAGATGCTTAAAGCAACTAAAAT TAAGCTACTGGAAAGCTCCAGTGGTCCTAGATCCC GACTAATAGG	Cassette primers for *URA3* integrated 1.7 kb left of CEN7
PJ72	ATTCGGGCAATTGTGTTCGTTATTGGTGGTAAATA ATGGTAAGACTACTTGGCACATGTATAGAAGGAC CACCTTTGATTG	
PJ67	ATTGATTGAATTTATAGCGGAAAATGGATGACAAT TAAAGGTTACGTGACGCTTTTTGCTCCTAGATCCC GACTAATAGG	Cassette primers for *URA3* integrated 1.9 kb right of CEN7
PJ68	CTACATTTTCATGGACCAAACCCACTACAACACAT GCACCACACTGCACCTCCCCTAAAATAGAAGGAC CACCTTTGATTG	
19.6525_HpaI _MluI_F	CCCAAAGTTAACACGCGTGTCAATGCAGTCGTTGA ATAC	
19.6525_SacII _R	CCAAACCGCGGTTTCAATATCGCAGAGATGGGAT	Construction of pFA-URA3-I-SceI-TS-Orf 19.6524/65
19.6524_AvrII _F	AAACCCCCTAGGGAGTGATGATGAGATTAACCAG	
19.6524_HindI II_MluI_R	AAACCCAAGCTTACGCGTGCCTTATATGCCACCGA TGA	
R1_URA INT_FP	GTCAGAAATTGATTTATGGACGAGATAAGACTAA AATATGATTCTTCTAAAATCACATAATTAATTAGA GCTTCCGGATAATAGGAATTG	Cassette primers for *URA3* integrated 6.5 kb right of CEN7
R1_URA INT_RP	GTGTAACAAAAATTTGCAATCACATCATTGACAGC CACCACAGTTTTTTTATAATAAGTGATATTGTTAG AAGGACCACCTTTGATTG	
PJ70	TTGCTTTAAATGTTTCAAACCATAGGTATGAGTTT GGGTAGTATTTGGCGGAATTAATGTCCTAGATCCC GACTAATAGG	Cassette primers for *URA3* integrated 10.4 kb right of CEN7
PJ71	ATCACTCTTGTCGTTTATTGTAGATCACTAAAAGT AATGGTTGTGTGAATAACTCCTGCTTAGAAGGACC ACCTTTGATTG	
URA3 AT HSP3_FP	CAGTTTTAAGAAGGTTTACATTATTAGCCTACGAA CAAAGACAGGTTATGATAGGAAACAGAGCTCCTG TTTTTATTCAGCTTCCGGATAATAGGAATTG	Cassette primers for *URA3* integrated 16.3 kb right of CEN7
URA3 AT HSP3_RP	GCAATCGATCGTAAACGCCACTCAAGCTAAACTG AAAACTACTACGCCTAGAAGGCTAATCGGTACCA ATTAGAAGGACCACCTTTGATTG	

URA3 AT CRTL7_FP	GATCACATATGATTCTAGTACCACTAAACATTATCAACAACTATCATCAATTAGTAGAATTACTCTGAGAGCTTCCGGATAATAGGAATTG	Cassette primers for *URA3* integrated 100 kb left of CEN7
URA3 AT CRTL7_RP	CCACGTGGATTTTTAAAATCTCAATAGTTTCTATAGTGGTGGTATACCACTACTACGACTGTGGATTCATTAGAAGGACCACCTTTGATTG	
URA3 at nCEN5-II_FP	CAATTCCTATTATCCGGAAGCTGTCGTGTAAGGCGGTAAATGGTTTTGGTGGGTTTATTTTTCTTTAAAAATCCAGACATGTCTTGC	Cassette primers for *URA3* integrated 7.5 kb left of CEN5
URA3 at nCEN5-II_RP	CAATCAAAGGTGGTCCTTCTAACACACTATTTACTTGTGGTAAACATACTATTGGTTGATAATGATGTTAGCAATGGGTTTATGCTTATTTAC	
nCEN7-3	GCATACCTGACACTGTCGTT	q-PCR of *CEN7*
nCEN7-4	AACGGTGCTACGTTTTTTTA	
URA3 RT1	TGTTGAAAGTTGCTGTAGTG	q-PCR for *URA3* promoter
URA3 RT2	TGCAGGAAATAAGATTGC	
URA3 RT3	TCATCAGTGGGATCATTAGCA	q-PCR for *URA3* ORF
URA3 RT4	CACGTTGGGCAATAAATCCA	
CEN1 core RT1	CAATCTAGCATTTCCTTCACACA	q-PCR for *CEN1*
CEN1 core RT2	TGACGCAATGAAGTAGGTGAT	
CACH5F1	CCCGCAAATAAGCAAACACT	qPCR for CEN5
CACH5R1	TTCATGGAAGAGGGGTTTCA	
7S10 RTF	CTTGTAAATTTAATTGTCGCTGAGG	qPCR primers for neocentromere mapping
7S10 RTR	CGGATAATCGTCCAACATATGAC	
7S11 RTF	GTCTTCTGACCTACCCATCAC	
7S11 RTR	GAGGCGGAAGTTGGACC	
7S12 RTF	CGTTGTGGCAATTGTATTTATG	
7S12 RTR	GCCATAGCTTAGCAAATAACC	
7S13 RTF	CATGGCTAATCCAACAACACATG	
7S13 RTR	GCTGGCTCTTGTTCTTGTATC	
421K RT1	CCTATCGCCACAAGGGAGA	
421K RT2	CAACGACTGCATTGACTCTTT	
7S14 RTF	GGATGTTGAGTTCAAAGCCTG	
7S14 RTR	CCAGCCAAATAATCTAGCTGC	

4RTF	ATTTGTCCCATCCGTAATTGATTC	
3RTR	ACGTTTTACCAGCCTATGC	
18RTF	AATAGCTATATCAGTTGTCAGCTTAC	
17RTF	AATGCTTGGCCCTCAGTATAAC	
20RTF	ACTGAAGTCGGCTGTGATC	
19RTR	GATAACTGGACTCATTAGGCGAA	
16RTF	ACCAGGATAATCTAACTGGCAAC	
15RTR	CTATTGCCCCAATCAATAACCTT	
7F1RT	CAGTAAACGTCATCTCTTTTATACCT	
7R1RT	GGAAGTGTAACTATTGAGCTCC	
7F2RT	ATTAAATAGAATGCGGCAATACC	
7R2RT	ATTTTAAGGATGAGAGGTGTGG	
7F3RT	CTGGTATTCACAATGGAACGGT	
7R3RT	GTCACCCCAATTCAAATCACGT	
7F4RT	GGAGCTGGCGATCAATTTGT	
7R4RT	TCACACATGAGAGGACCGTT	
7F4A RTF	CGGATAATTGAAAGCAGCAATG	
7R4A RTR	CCACAACCTGTTGACGAG	
7F4B RT	GTAGGCGCGGATTTAATGTG	
7R4B RT	CCAACTTGTTTAGTTGTTGGATCTG	
7F4C RT	GACAAACACTCAAGGAGCAG	
7R4C RT	CTGCAAATCTATTGGAGGTGG	
7F5 RT	GGACAAAATCAGATACCAAGCC	
7R5 RT	GCTTTGGTCATACCAATACCAG	
7F6 RT	CTCCAAGAACATCAAATTGGG	
7R6 RT	CAAGGAAGTCATTTCTTCAGAAG	
CEN7DHIS_F P	GTAAACTTTTTCGATTCTCAATTTACTTTGAGGGCA TTGTCGCAAATGGAGATTCCTTACGATGGGAATTC CGGAATATTTATGAGAAAC	CEN7 DELETION WITH HIS1
CEN7DHIS_R P	CACAAAAATGCCCGCTAACAATACCATTAATTCCT ACTCCATGTACAGAATACCCAACATGCTTTGTATC GAATTCCGGGGATCCTGGAG	
HIS ORF_2	GGAGTAATGGTTAAACATTTTGC	
HIS ORF_1	CAAAGAAGCTGAACAATTCGAC	
ORC4 13	CGGGGTACCTTGGTTTGTAAAAATGTTGTTTC	Deletion cassette for ORC4
ORC4 14	CCGCTCGAGAAATAGTTTTACTCTTGAGTTAGC	
ORC4 15	TCCCCGCGGGTTATAGGTTGCTTTTAGTGC	
ORC4 16	TCCCCGCGGGTTATAGGTTGCTTTTAGTGC	
ORC4 11	CGCGGATCCATGAATTCACAGGACC	MET3 cassette cloning for ORC4
ORC4 12	AACTGCAGTGCCATTTAACTCTTTTAAGGCG	
MCM2_13	CGGGGTACCCTAATCCCATTTTGTTATGAATAT	

MCM2_14	CCGCTCGAGGGGTTGATTAAATAGTAATGTAATTAA TAAAG	Deletion cassette for MCM2
MCM2_15	TCCCCGCGGGTGATTAGTGGGTTATGG	
MCM2_16	CGGAGCTCTGCATTCCAGATTATTTTCTG	
MCM2_11	CGCGGATCCATGTCAAGTCCACCAGCTG	N term of Mcm2 (For *MET3pr* cloning)
MCM2_12	AACTGCAGGCGTCTTCATCTTCATCATCGTC	

Table 6.3. Southern blot strategy for *CEN7* deletion strains

Strain	Restriction enzyme	Primers used to amplify probe (length of probe)	Size of the expected band/wild type
4.5 kb *CEN7* deletion in URA3 at 4L locus	*Afl*III	URA3RT1, URA3 RT4 (870 bp)	8.3/7.6 kb
4.5 kb *CEN7* deletion in *URA3* at 4R locus	*Nco*I	HIS ORF_2, HIS ORF_1 (480 bp)	11.4/10 kb

Table 6.4 Softwares and online tools used

Name	Website/ Source
Candida genome database	http://www.candidagenome.org/
Integrative Genomics Viewer	http://software.broadinstitute.org/software/igv/
ESPript 3.0	http://espript.ibcp.fr/ESPript/ESPript/
WebSIDD	http://orange.genomecenter.ucdavis.edu/benham/sidd/index.html

7. REFERENCES

Agier, N., O. M. Romano, F. Touzain, M. Cosentino Lagomarsino and G. Fischer (2013). "The spatiotemporal program of replication in the genome of Lachancea kluyveri." Genome Biol Evol **5**(2): 370-388.

Ak, P. and C. J. Benham (2005). "Susceptibility to superhelically driven DNA duplex destabilization: a highly conserved property of yeast replication origins." PLoS Comput Biol **1**(1): e7.

Akiyoshi, B. and K. Gull (2014). "Discovery of unconventional kinetochores in kinetoplastids." Cell **156**(6): 1247-1258.

Albertson, D. G. and J. N. Thomson (1982). "The kinetochores of Caenorhabditis elegans." Chromosoma **86**(3): 409-428.

Allshire, R. C. and K. Ekwall (2015). "Epigenetic Regulation of Chromatin States in Schizosaccharomyces pombe." Cold Spring Harb Perspect Biol **7**(7): a018770.

Allshire, R. C., J. P. Javerzat, N. J. Redhead and G. Cranston (1994). "Position effect variegation at fission yeast centromeres." Cell **76**(1): 157-169.

Allshire, R. C. and H. D. Madhani (2018). "Ten principles of heterochromatin formation and function." Nat Rev Mol Cell Biol **19**(4): 229-244.

Amor, D. J. and K. H. Choo (2002). "Neocentromeres: role in human disease, evolution, and centromere study." Am J Hum Genet **71**(4): 695-714.

Anderson, M. Z., J. A. Baller, K. Dulmage, L. Wigen and J. Berman (2012). "The three clades of the telomere-associated TLO gene family of Candida albicans have different splicing, localization, and expression features." Eukaryot Cell **11**(10): 1268-1275.

Anderson, M. Z., A. C. Gerstein, L. Wigen, J. A. Baller and J. Berman (2014). "Silencing is noisy: population and cell level noise in telomere-adjacent genes is dependent on telomere position and sir2." PLoS Genet **10**(7): e1004436.

Aparicio, O. M. (2013). "Location, location, location: it's all in the timing for replication origins." Genes Dev **27**(2): 117-128.

Ayoub, N., I. Goldshmidt and A. Cohen (1999). "Position effect variegation at the mating-type locus of fission yeast: a cis-acting element inhibits covariegated expression of genes in the silent and expressed domains." Genetics **152**(2): 495-508.

Bannister, A. J., P. Zegerman, J. F. Partridge, E. A. Miska, J. O. Thomas, R. C. Allshire and T. Kouzarides (2001). "Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain." Nature **410**(6824): 120-124.

Baum, M., K. Sanyal, P. K. Mishra, N. Thaler and J. Carbon (2006). "Formation of functional centromeric chromatin is specified epigenetically in Candida albicans." Proc Natl Acad Sci U S A **103**(40): 14877-14882.

Bell, S. P. and A. Dutta (2002). "DNA replication in eukaryotic cells." Annu Rev Biochem **71**: 333-374.

Bell, S. P. and B. Stillman (1992). "ATP-dependent recognition of eukaryotic origins of DNA replication by a multiprotein complex." Nature **357**(6374): 128-134.

Bennett, R. J. and A. D. Johnson (2003). "Completion of a parasexual cycle in Candida albicans by induced chromosome loss in tetraploid strains." EMBO J **22**(10): 2505-2515.

Berman, J. (2012). "Candida albicans." Curr Biol **22**(16): R620-622.

Bernard, P., J. F. Maure, J. F. Partridge, S. Genier, J. P. Javerzat and R. C. Allshire (2001). "Requirement of heterochromatin for cohesion at centromeres." Science **294**(5551): 2539-2542.

Berry, S., M. Hartley, T. S. Olsson, C. Dean and M. Howard (2015). "Local chromatin environment of a Polycomb target gene instructs its own epigenetic inheritance." Elife **4**.

Bi, C. and C. J. Benham (2004). "WebSIDD: server for predicting stress-induced duplex destabilized (SIDD) sites in superhelical DNA." Bioinformatics **20**(9): 1477-1479.

Black, B. E., M. A. Brock, S. Bedard, V. L. Woods, Jr. and D. W. Cleveland (2007). "An epigenetic mark generated by the incorporation of CENP-A into centromeric nucleosomes." Proc Natl Acad Sci U S A **104**(12): 5008-5013.

Black, B. E. and D. W. Cleveland (2011). "Epigenetic centromere propagation and the nature of CENP-a nucleosomes." Cell **144**(4): 471-479.

Bloom, K. S. (2014). "Centromeric heterochromatin: the primordial segregation machine." Annu Rev Genet **48**: 457-484.

Bloom, K. S. and J. Carbon (1982). "Yeast centromere DNA is in a unique and highly ordered structure in chromosomes and small circular minichromosomes." Cell **29**(2): 305-317.

Blow, J. J., P. J. Gillespie, D. Francis and D. A. Jackson (2001). "Replication origins in Xenopus egg extract Are 5-15 kilobases apart and are activated in clusters that fire at different times." J Cell Biol **152**(1): 15-25.

Bobkov, G. O. M., N. Gilbert and P. Heun (2018). "Centromere transcription allows CENP-A to transit from chromatin association to stable incorporation." J Cell Biol.

Brewer, B. J. and W. L. Fangman (1987). "The localization of replication origins on ARS plasmids in S. cerevisiae." Cell **51**(3): 463-471.

Buchwitz, B. J., K. Ahmad, L. L. Moore, M. B. Roth and S. Henikoff (1999). "A histone-H3-like protein in C. elegans." Nature **401**(6753): 547-548.

Burrack, L. S., H. F. Hutton, K. J. Matter, S. A. Clancey, I. Liachko, A. E. Plemmons, A. Saha, E. A. Power, B. Turman, M. A. Thevandavakkam, F. Ay, M. J. Dunham and J. Berman (2016). "Neocentromeres Provide Chromosome Segregation Accuracy and Centromere Clustering to Multiple Loci along a Candida albicans Chromosome." PLoS Genet **12**(9): e1006317.

Calderano, S. G., W. C. Drosopoulos, M. M. Quaresma, C. A. Marques, S. Kosiyatrakul, R. McCulloch, C. L. Schildkraut and M. C. Elias (2015). "Single molecule analysis of Trypanosoma brucei DNA replication dynamics." Nucleic Acids Res **43**(5): 2655-2665.

Cannon, R. D., H. F. Jenkinson and M. G. Shepherd (1990). "Isolation and nucleotide sequence of an autonomously replicating sequence (ARS) element functional in Candida albicans and Saccharomyces cerevisiae." Mol Gen Genet **221**(2): 210-218.

Care, R. S., J. Trevethick, K. M. Binley and P. E. Sudbery (1999). "The MET3 promoter: a new tool for Candida albicans molecular genetics." Mol Microbiol **34**(4): 792-798.

Carpenter, P. B., P. R. Mueller and W. G. Dunphy (1996). "Role for a Xenopus Orc2-related protein in controlling DNA replication." Nature **379**(6563): 357-360.

Castillo, A. G., B. G. Mellone, J. F. Partridge, W. Richardson, G. L. Hamilton, R. C. Allshire and A. L. Pidoux (2007). "Plasticity of fission yeast CENP-A chromatin driven by relative levels of histone H3 and H4." PLoS Genet **3**(7): e121.

Catania, S., A. L. Pidoux and R. C. Allshire (2015). "Sequence features and transcriptional stalling within centromere DNA promote establishment of CENP-A chromatin." PLoS Genet **11**(3): e1004986.

Cayrou, C., P. Coulombe, A. Vigneron, S. Stanojcic, O. Ganier, I. Peiffer, E. Rivals, A. Puy, S. Laurent-Chabalier, R. Desprat and M. Mechali (2011). "Genome-scale analysis of metazoan replication origins reveals their organization in specific but flexible sites defined by conserved features." Genome Res **21**(9): 1438-1449.

Centola, M. and J. Carbon (1994). "Cloning and characterization of centromeric DNA from Neurospora crassa." Mol Cell Biol **14**(2): 1510-1519.

Chan, F. L., O. J. Marshall, R. Saffery, B. W. Kim, E. Earle, K. H. Choo and L. H. Wong (2012). "Active transcription and essential role of RNA polymerase II at the centromere during mitosis." Proc Natl Acad Sci U S A **109**(6): 1979-1984.

Chatterjee, G., S. R. Sankaranarayanan, K. Guin, Y. Thattikota, S. Padmanabhan, R. Siddharthan and K. Sanyal (2016). "Repeat-Associated Fission Yeast-Like Regional Centromeres in the Ascomycetous Budding Yeast Candida tropicalis." PLoS Genet **12**(2): e1005839.

Choi, E. S., A. Stralfors, A. G. Castillo, M. Durand-Dubief, K. Ekwall and R. C. Allshire (2011). "Identification of noncoding transcripts from within CENP-A chromatin at fission yeast centromeres." J Biol Chem **286**(26): 23600-23607.

Chuang, R. Y. and T. J. Kelly (1999). "The fission yeast homologue of Orc4p binds to replication origin DNA via multiple AT-hooks." Proc Natl Acad Sci U S A **96**(6): 2656-2661.

Clarke, L., H. Amstutz, B. Fishel and J. Carbon (1986). "Analysis of centromeric DNA in the fission yeast Schizosaccharomyces pombe." Proc Natl Acad Sci U S A **83**(21): 8253-8257.

Comoglio, F., T. Schlumpf, V. Schmid, R. Rohs, C. Beisel and R. Paro (2015). "High-resolution profiling of Drosophila replication start sites reveals a DNA shape and chromatin signature of metazoan origins." Cell Rep **11**(5): 821-834.

Craig, J. M., L. H. Wong, A. W. Lo, E. Earle and K. H. Choo (2003). "Centromeric chromatin pliability and memory at a human neocentromere." EMBO J **22**(10): 2495-2504.

Culotti, J. and L. H. Hartwell (1971). "Genetic control of the cell division cycle in yeast. 3. Seven genes controlling nuclear division." Exp Cell Res **67**(2): 389-401.

Dai, J., R. Y. Chuang and T. J. Kelly (2005). "DNA replication origins in the Schizosaccharomyces pombe genome." Proc Natl Acad Sci U S A **102**(2): 337-342.

De Rop, V., A. Padeganeh and P. S. Maddox (2012). "CENP-A: the key player behind centromere identity, propagation, and kinetochore assembly." Chromosoma **121**(6): 527-538.

Dellino, G. I., D. Cittaro, R. Piccioni, L. Luzi, S. Banfi, S. Segalla, M. Cesaroni, R. Mendoza-Maldonado, M. Giacca and P. G. Pelicci (2013). "Genome-wide mapping of human DNA-replication origins: levels of transcription at ORC1 sites regulate origin selection and replication timing." Genome Res **23**(1): 1-11.

Descorps-Declere, S., C. Saguez, A. Cournac, M. Marbouty, T. Rolland, L. Ma, C. Bouchier, I. Moszer, B. Dujon, R. Koszul and G. F. Richard (2015). "Genome-wide replication landscape of Candida glabrata." BMC Biol **13**: 69.

Dietrich, F. S., S. Voegeli, S. Brachat, A. Lerch, K. Gates, S. Steiner, C. Mohr, R. Pohlmann, P. Luedi, S. Choi, R. A. Wing, A. Flavier, T. D. Gaffney and P. Philippsen (2004). "The Ashbya gossypii genome as a tool for mapping the ancient Saccharomyces cerevisiae genome." Science **304**.

Drinnenberg, I. A., D. deYoung, S. Henikoff and H. S. Malik (2014). "Recurrent loss of CenH3 is associated with independent transitions to holocentricity in insects." Elife **3**.

Duan, Z., M. Andronescu, K. Schutz, S. McIlwain, Y. J. Kim, C. Lee, J. Shendure, S. Fields, C. A. Blau and W. S. Noble (2010). "A three-dimensional model of the yeast genome." Nature **465**(7296): 363-367.

Dunleavy, E. M., G. Almouzni and G. H. Karpen (2011). "H3.3 is deposited at centromeres in S phase as a placeholder for newly assembled CENP-A in G(1) phase." Nucleus **2**(2): 146-157.

Dunleavy, E. M., D. Roche, H. Tagami, N. Lacoste, D. Ray-Gallet, Y. Nakamura, Y. Daigo, Y. Nakatani and G. Almouzni-Pettinotti (2009). "HJURP is a cell-cycle-dependent maintenance and deposition factor of CENP-A at centromeres." Cell **137**(3): 485-497.

Dutta, A. and S. P. Bell (1997). "Initiation of DNA replication in eukaryotic cells." Annu Rev Cell Dev Biol **13**: 293-332.

Eaton, M. L., K. Galani, S. Kang, S. P. Bell and D. M. MacAlpine (2010). "Conserved nucleosome positioning defines replication origins." Genes Dev **24**(8): 748-753.

Fangman, W. L. and B. J. Brewer (1991). "Activation of replication origins within yeast chromosomes." Annu Rev Cell Biol **7**: 375-402.

Federico, C., A. M. Pappalardo, V. Ferrito, S. Tosi and S. Saccone (2017). "Genomic properties of chromosomal bands are linked to evolutionary rearrangements and new centromere formation in primates." Chromosome Res **25**(3-4): 261-276.

Feng, W., J. Bachant, D. Collingwood, M. K. Raghuraman and B. J. Brewer (2009). "Centromere replication timing determines different forms of genomic instability in Saccharomyces cerevisiae checkpoint mutants during replication stress." Genetics **183**(4): 1249-1260.

Fitzgerald-Hayes, M., L. Clarke and J. Carbon (1982). "Nucleotide sequence comparisons and functional analysis of yeast centromere DNAs." Cell **29**(1): 235-244.

Fitzpatrick, D. A., M. E. Logue, J. E. Stajich and G. Butler (2006). "A fungal phylogeny based on 42 complete genomes derived from supertree and combined gene analysis." BMC Evol Biol **6**: 99.

Forsburg, S. L. (2004). "Eukaryotic MCM proteins: beyond replication initiation." Microbiol Mol Biol Rev **68**(1): 109-131.

Forsburg, S. L., D. A. Sherman, S. Ottilie, J. R. Yasuda and J. A. Hodson (1997). "Mutational analysis of Cdc19p, a Schizosaccharomyces pombe MCM protein." Genetics **147**(3): 1025-1041.

Foss, M., F. J. McNally, P. Laurenson and J. Rine (1993). "Origin recognition complex (ORC) in transcriptional silencing and DNA replication in S. cerevisiae." Science **262**(5141): 1838-1844.

Fournier, P., A. Abbas, M. Chasles, B. Kudla, D. M. Ogrydziak, D. Yaver, J. W. Xuan, A. Peito, A. M. Ribet, C. Feynerol and et al. (1993). "Colocalization of centromeric and replicative functions on autonomously replicating sequences isolated from the yeast Yarrowia lipolytica." Proc Natl Acad Sci U S A **90**(11): 4912-4916.

Fox, C. A., S. Loo, A. Dillin and J. Rine (1995). "The origin recognition complex has essential functions in transcriptional silencing and chromosomal replication." Genes Dev **9**(8): 911-924.

Fujita, Y., T. Hayashi, T. Kiyomitsu, Y. Toyoda, A. Kokubu, C. Obuse and M. Yanagida (2007). "Priming of centromere for CENP-A recruitment by human hMis18alpha, hMis18beta, and M18BP1." Dev Cell **12**(1): 17-30.

Fukagawa, T. (2013). "Speciation mediated by centromeres." Dev Cell **27**(4): 367-368.

Fukagawa, T. and W. C. Earnshaw (2014). "The centromere: chromatin foundation for the kinetochore machinery." Dev Cell **30**(5): 496-508.

Furuyama, T. and S. Henikoff (2009). "Centromeric nucleosomes induce positive DNA supercoils." Cell **138**(1): 104-113.

Gavin, K. A., M. Hidaka and B. Stillman (1995). "Conserved initiator proteins in eukaryotes." Science **270**(5242): 1667-1671.

Glynn, E. F., P. C. Megee, H. G. Yu, C. Mistrot, E. Unal, D. E. Koshland, J. L. DeRisi and J. L. Gerton (2004). "Genome-wide mapping of the cohesin complex in the yeast Saccharomyces cerevisiae." PLoS Biol **2**(9): E259.

Gossen, M., D. T. Pak, S. K. Hansen, J. K. Acharya and M. R. Botchan (1995). "A Drosophila homolog of the yeast origin recognition complex." Science **270**(5242): 1674-1677.

Grallert, B. and P. Nurse (1996). "The ORC1 homolog orp1 in fission yeast plays a key role in regulating onset of S phase." Genes Dev **10**(20): 2644-2654.

Greenfeder, S. A. and C. S. Newlon (1992). "Replication forks pause at yeast centromeres." Mol Cell Biol **12**(9): 4056-4066.

Grewal, S. I. and A. J. Klar (1996). "Chromosomal inheritance of epigenetic states in fission yeast during mitosis and meiosis." Cell **86**(1): 95-101.

Haase, J., P. K. Mishra, A. Stephens, R. Haggerty, C. Quammen, R. M. Taylor, 2nd, E. Yeh, M. A. Basrai and K. Bloom (2013). "A 3D map of the yeast kinetochore reveals the presence of core and accessory centromere-specific histone." Curr Biol **23**(19): 1939-1944.

Hall, I. M., G. D. Shankaranarayana, K. Noma, N. Ayoub, A. Cohen and S. I. Grewal (2002). "Establishment and maintenance of a heterochromatin domain." Science **297**(5590): 2232-2237.

Hammond, C. M., C. B. Stromme, H. Huang, D. J. Patel and A. Groth (2017). "Histone chaperone networks shaping chromatin function." Nat Rev Mol Cell Biol **18**(3): 141-158.

Han, F., Z. Gao and J. A. Birchler (2009). "Reactivation of an inactive centromere reveals epigenetic and structural components for centromere specification in maize." Plant Cell **21**(7): 1929-1939.

Hartwell, L. H. (1971). "Genetic control of the cell division cycle in yeast. II. Genes controlling DNA replication and its initiation." J Mol Biol **59**(1): 183-194.

Hartwell, L. H. (1971). "Genetic control of the cell division cycle in yeast. IV. Genes controlling bud emergence and cytokinesis." Exp Cell Res **69**(2): 265-276.

Hayano, M., Y. Kanoh, S. Matsumoto, C. Renard-Guillet, K. Shirahige and H. Masai (2012). "Rif1 is a global regulator of timing of replication origin firing in fission yeast." Genes Dev **26**(2): 137-150.

Hayashi, M. T., T. S. Takahashi, T. Nakagawa, J. Nakayama and H. Masukata (2009). "The heterochromatin protein Swi6/HP1 activates replication origins at the pericentromeric region and silent mating-type locus." Nat Cell Biol **11**(3): 357-362.

Heichinger, C., C. J. Penkett, J. Bahler and P. Nurse (2006). "Genome-wide characterization of fission yeast DNA replication origins." EMBO J **25**(21): 5171-5179.

Henikoff, S., K. Ahmad, J. S. Platero and B. van Steensel (2000). "Heterochromatic deposition of centromeric histone H3-like proteins." Proc Natl Acad Sci U S A **97**(2): 716-721.

Herreros, E., M. I. Garcia-Saez, C. Nombela and M. Sanchez (1992). "A reorganized Candida albicans DNA sequence promoting homologous non-integrative genetic transformation." Mol Microbiol **6**(23): 3567-3574.

Heun, P., T. Laroche, M. K. Raghuraman and S. M. Gasser (2001). "The positioning and dynamics of origins of replication in the budding yeast nucleus." J Cell Biol **152**(2): 385-400.

Heus, J. J., B. J. Zonneveld, H. Y. de Steensma and J. A. van den Berg (1993). "The consensus sequence of Kluyveromyces lactis centromeres shows homology to functional centromeric DNA from Saccharomyces cerevisiae." Mol Gen Genet **236**(2-3): 355-362.

Hickman, M. A., C. A. Froyd and L. N. Rusche (2011). "Reinventing heterochromatin in budding yeasts: Sir2 and the origin recognition complex take center stage." Eukaryot Cell **10**(9): 1183-1192.

Hickman, M. A. and L. N. Rusche (2010). "Transcriptional silencing functions of the yeast protein Orc1/Sir3 subfunctionalized after gene duplication." Proc Natl Acad Sci U S A **107**(45): 19384-19389.

Hizume, K., M. Yagura and H. Araki (2013). "Concerted interaction between origin recognition complex (ORC), nucleosomes and replication origin DNA ensures stable ORC-origin binding." Genes Cells **18**(9): 764-779.

Holland, L., L. Gauthier, P. Bell-Rogers and K. Yankulov (2002). "Distinct parts of minichromosome maintenance protein 2 associate with histone H3/H4 and RNA polymerase II holoenzyme." Eur J Biochem **269**(21): 5192-5202.

Hori, T., N. Kagawa, A. Toyoda, A. Fujiyama, S. Misu, N. Monma, F. Makino, K. Ikeo and T. Fukagawa (2017). "Constitutive centromere-associated network controls centromere drift in vertebrate cells." J Cell Biol **216**(1): 101-113.

Huang, H., C. B. Stromme, G. Saredi, M. Hodl, A. Strandsby, C. Gonzalez-Aguilera, S. Chen, A. Groth and D. J. Patel (2015). "A unique binding mode enables MCM2 to chaperone histones H3-H4 at replication forks." Nat Struct Mol Biol **22**(8): 618-626.

Hull, C. M. and A. D. Johnson (1999). "Identification of a mating type-like locus in the asexual pathogenic yeast Candida albicans." Science **285**.

Imakaev, M., G. Fudenberg, R. P. McCord, N. Naumova, A. Goloborodko, B. R. Lajoie, J. Dekker and L. A. Mirny (2012). "Iterative correction of Hi-C data reveals hallmarks of chromosome organization." Nat Methods **9**(10): 999-1003.

Ishii, K., Y. Ogiyama, Y. Chikashige, S. Soejima, F. Masuda, T. Kakuma, Y. Hiraoka and K. Takahashi (2008). "Heterochromatin integrity affects chromosome reorganization after centromere dysfunction." Science **321**(5892): 1088-1091.

Jain, M., S. Koren, K. H. Miga, J. Quick, A. C. Rand, T. A. Sasani, J. R. Tyson, A. D. Beggs, A. T. Dilthey, I. T. Fiddes, S. Malla, H. Marriott, T. Nieto, J. O'Grady, H. E. Olsen, B. S. Pedersen, A. Rhie, H. Richardson, A. R. Quinlan, T. P. Snutch, L. Tee, B. Paten, A. M. Phillippy, J. T. Simpson, N. J. Loman and M. Loose (2018). "Nanopore sequencing and assembly of a human genome with ultra-long reads." Nat Biotechnol **36**(4): 338-345.

Jain, M., H. E. Olsen, D. J. Turner, D. Stoddart, K. V. Bulazel, B. Paten, D. Haussler, H. F. Willard, M. Akeson and K. H. Miga (2018). "Linear assembly of a human centromere on the Y chromosome." Nat Biotechnol **36**(4): 321-323.

Joglekar, A. P., D. Bouck, K. Finley, X. Liu, Y. Wan, J. Berman, X. He, E. D. Salmon and K. S. Bloom (2008). "Molecular architecture of the kinetochore-microtubule attachment site is conserved between point and regional centromeres." J Cell Biol **181**(4): 587-594.

Kapoor, S., L. Zhu, C. Froyd, T. Liu and L. N. Rusche (2015). "Regional centromeres in the yeast Candida lusitaniae lack pericentromeric heterochromatin." Proc Natl Acad Sci U S A **112**(39): 12139-12144.

Karpen, G. H. and R. C. Allshire (1997). "The case for epigenetic effects on centromere identity and function." Trends Genet **13**(12): 489-496.

Ketel, C., H. S. Wang, M. McClellan, K. Bouchonville, A. Selmecki, T. Lahav, M. Gerami-Nejad and J. Berman (2009). "Neocentromeres form efficiently at multiple possible loci in Candida albicans." PLoS Genet **5**(3): e1000400.

Kim, S. M., D. D. Dubey and J. A. Huberman (2003). "Early-replicating heterochromatin." Genes Dev **17**(3): 330-335.

Kitada, K., E. Yamaguchi, K. Hamada and M. Arisawa (1997). "Structural analysis of a Candida glabrata centromere and its functional homology to the Saccharomyces cerevisiae centromere." Curr Genet **31**(2): 122-127.

Kitagawa, T., K. Ishii, K. Takeda and T. Matsumoto (2014). "The 19S proteasome subunit Rpt3 regulates distribution of CENP-A by associating with centromeric chromatin." Nat Commun **5**: 3597.

Kitamura, E., K. Tanaka, Y. Kitamura and T. U. Tanaka (2007). "Kinetochore microtubule interaction during S phase in Saccharomyces cerevisiae." Genes Dev **21**(24): 3319-3330.

Knott, S. R., J. M. Peace, A. Z. Ostrow, Y. Gan, A. E. Rex, C. J. Viggiani, S. Tavare and O. M. Aparicio (2012). "Forkhead transcription factors establish origin timing and long-range clustering in S. cerevisiae." Cell **148**(1-2): 99-111.

Kobayashi, N., Y. Suzuki, L. W. Schoenfeld, C. A. Muller, C. Nieduszynski, K. H. Wolfe and T. U. Tanaka (2015). "Discovery of an unconventional centromere in budding yeast redefines evolution of point centromeres." Curr Biol **25**(15): 2026-2033.

Kogut, I., J. Wang, V. Guacci, R. K. Mistry and P. C. Megee (2009). "The Scc2/Scc4 cohesin loader determines the distribution of cohesin on budding yeast chromosomes." Genes Dev **23**(19): 2345-2357.

Kong, D. and M. L. DePamphilis (2002). "Site-specific ORC binding, pre-replication complex assembly and DNA synthesis at Schizosaccharomyces pombe replication origins." EMBO J **21**(20): 5567-5576.

Koren, A., H. J. Tsai, I. Tirosh, L. S. Burrack, N. Barkai and J. Berman (2010). "Epigenetically-inherited centromere and neocentromere DNA replicates earliest in S-phase." PLoS Genet **6**(8): e1001068.

Kozubowski, L., V. Yadav, G. Chatterjee, S. Sridhar, M. Yamaguchi, S. Kawamoto, I. Bose, J. Heitman and K. Sanyal (2013). "Ordered kinetochore assembly in the human-pathogenic basidiomycetous yeast Cryptococcus neoformans." MBio **4**(5): e00614-00613.

Kurtz, M. B., M. W. Cortelyou, S. M. Miller, M. Lai and D. R. Kirsch (1987). "Development of autonomously replicating plasmids for Candida albicans." Mol Cell Biol **7**(1): 209-217.

Lagana, A., J. F. Dorn, V. De Rop, A. M. Ladouceur, A. S. Maddox and P. S. Maddox (2010). "A small GTPase molecular switch regulates epigenetic centromere maintenance by stabilizing newly incorporated CENP-A." Nat Cell Biol **12**(12): 1186-1193.

Lakshmi Sreekumar, N. V., Kaustuv Sanyal (2017). Chromosome components important for genome stability in Candida albicans and related species. Candida albicans: Cellular and Molecular Biology. R. Prasad, Sprimger.

Langmead, B. and S. L. Salzberg (2012). "Fast gapped-read alignment with Bowtie 2." Nat Methods **9**(4): 357-359.

Langmead, B., C. Trapnell, M. Pop and S. L. Salzberg (2009). "Ultrafast and memory-efficient alignment of short DNA sequences to the human genome." Genome Biol **10**(3): R25.

Lawrimore, J., A. Doshi, B. Friedman, E. Yeh and K. Bloom (2018). "Geometric partitioning of cohesin and condensin is a consequence of chromatin loops." Mol Biol Cell **29**(22): 2737-2750.

Lazar-Stefanita, L., V. F. Scolari, G. Mercy, H. Muller, T. M. Guerin, A. Thierry, J. Mozziconacci and R. Koszul (2017). "Cohesins and condensins orchestrate the 4D dynamics of yeast chromosomes during the cell cycle." EMBO J **36**(18): 2684-2697.

Lechner, J. and J. Carbon (1991). "A 240 kd multisubunit protein complex, CBF3, is a major component of the budding yeast centromere." Cell **64**(4): 717-725.

Lee, D. G. and S. P. Bell (1997). "Architecture of the yeast origin recognition complex bound to origins of DNA replication." Mol Cell Biol **17**(12): 7159-7168.

Leonard, A. C. and M. Mechali (2013). "DNA replication origins." Cold Spring Harb Perspect Biol **5**(10): a010116.

Li, N., W. H. Lam, Y. Zhai, J. Cheng, E. Cheng, Y. Zhao, N. Gao and B. K. Tye (2018). "Structure of the origin recognition complex bound to DNA replication origin." Nature **559**(7713): 217-222.

Li, W., J. Yi, P. Agbu, Z. Zhou, R. L. Kelley, S. Kallgren, S. Jia and X. He (2017). "Replication stress affects the fidelity of nucleosome-mediated epigenetic inheritance." PLoS Genet **13**(7): e1006900.

Li, W. C., Z. J. Zhong, P. P. Zhu, E. Z. Deng, H. Ding, W. Chen and H. Lin (2014). "Sequence analysis of origins of replication in the Saccharomyces cerevisiae genomes." Front Microbiol **5**: 574.

Liachko, I., A. Bhaskar, C. Lee, S. C. Chung, B. K. Tye and U. Keich (2010). "A comprehensive genome-wide map of autonomously replicating sequences in a naive genome." PLoS Genet **6**(5): e1000946.

Liachko, I., E. Tanaka, K. Cox, S. C. Chung, L. Yang, A. Seher, L. Hallas, E. Cha, G. Kang, H. Pace, J. Barrow, M. Inada, B. K. Tye and U. Keich (2011). "Novel features of ARS selection in budding yeast Lachancea kluyveri." BMC Genomics **12**: 633.

Liachko, I., R. A. Youngblood, K. Tsui, K. L. Bubb, C. Queitsch, M. K. Raghuraman, C. Nislow, B. J. Brewer and M. J. Dunham (2014). "GC-rich DNA elements enable replication origin activity in the methylotrophic yeast Pichia pastoris." PLoS Genet **10**(3): e1004169.

Lipford, J. R. and S. P. Bell (2001). "Nucleosomes positioned by ORC facilitate the initiation of DNA replication." Mol Cell **7**(1): 21-30.

Livny, J., Y. Yamaichi and M. K. Waldor (2007). "Distribution of centromere-like parS sites in bacteria: insights from comparative genomics." J Bacteriol **189**(23): 8693-8703.

Maggert, K. A. and G. H. Karpen (2001). "The activation of a neocentromere in Drosophila requires proximity to an endogenous centromere." Genetics **158**(4): 1615-1628.

Maine, G. T., P. Sinha and B. K. Tye (1984). "Mutants of S. cerevisiae defective in the maintenance of minichromosomes." Genetics **106**(3): 365-385.

Malik, H. S. and S. Henikoff (2009). "Major evolutionary transitions in centromere complexity." Cell **138**(6): 1067-1082.

Marahrens, Y. and B. Stillman (1992). "A yeast chromosomal origin of DNA replication defined by multiple functional elements." Science **255**(5046): 817-823.

Marshall, O. J., A. C. Chueh, L. H. Wong and K. H. Choo (2008). "Neocentromeres: new insights into centromere structure, disease development, and karyotype evolution." Am J Hum Genet **82**(2): 261-282.

McEachern, M. J. and J. B. Hicks (1993). "Unusually large telomeric repeats in the yeast Candida albicans." Mol Cell Biol **13**(1): 551-560.

McNulty, S. M., L. L. Sullivan and B. A. Sullivan (2017). "Human Centromeres Produce Chromosome-Specific and Array-Specific Alpha Satellite Transcripts that Are Complexed with CENP-A and CENP-C." Dev Cell **42**(3): 226-240 e226.

Mendez, J. and B. Stillman (2003). "Perpetuating the double helix: molecular machines at eukaryotic DNA replication origins." Bioessays **25**(12): 1158-1167.

Mitra, S., A. Biswas and L. Narlikar (2018). "DIVERSITY in binding, regulation, and evolution revealed from high-throughput ChIP." PLoS Comput Biol **14**(4): e1006090.

Mitra, S., J. Gomez-Raja, G. Larriba, D. D. Dubey and K. Sanyal (2014). "Rad51-Rad52 mediated maintenance of centromeric chromatin in Candida albicans." PLoS Genet **10**(4): e1004344.

Moore, L. L., M. Morrison and M. B. Roth (1999). "HCP-1, a protein involved in chromosome segregation, is localized to the centromere of mitotic chromosomes in Caenorhabditis elegans." J Cell Biol **147**(3): 471-480.

Muller, C. A. and C. A. Nieduszynski (2017). "DNA replication timing influences gene expression level." J Cell Biol **216**(7): 1907-1914.

Muller, P., S. Park, E. Shor, D. J. Huebert, C. L. Warren, A. Z. Ansari, M. Weinreich, M. L. Eaton, D. M. MacAlpine and C. A. Fox (2010). "The conserved bromo-adjacent homology domain of yeast Orc1 functions in the selection of DNA replication origins within chromatin." Genes Dev **24**(13): 1418-1433.

Murad, A. M., P. R. Lee, I. D. Broadbent, C. J. Barelle and A. J. Brown (2000). "CIp10, an efficient and convenient integrating vector for Candida albicans." Yeast **16**(4): 325-327.

Musacchio, A. and A. Desai (2017). "A Molecular View of Kinetochore Assembly and Function." Biology (Basel) **6**(1).

Muzi-Falconi, M. and T. J. Kelly (1995). "Orp1, a member of the Cdc18/Cdc6 family of S-phase regulators, is homologous to a component of the origin recognition complex." Proc Natl Acad Sci U S A **92**(26): 12475-12479.

Nakayashiki, H., N. Kadotani and S. Mayama (2006). "Evolution and diversification of RNA silencing proteins in fungi." J Mol Evol **63**(1): 127-135.

Natsume, T., C. A. Muller, Y. Katou, R. Retkute, M. Gierlinski, H. Araki, J. J. Blow, K. Shirahige, C. A. Nieduszynski and T. U. Tanaka (2013). "Kinetochores coordinate pericentromeric cohesion and early DNA replication by Cdc7-Dbf4 kinase recruitment." Mol Cell **50**(5): 661-674.

Newlon, C. S. and J. F. Theis (1993). "The structure and function of yeast ARS elements." Curr Opin Genet Dev **3**(5): 752-758.

Newman, T. J., M. A. Mamun, C. A. Nieduszynski and J. J. Blow (2013). "Replisome stall events have shaped the distribution of replication origins in the genomes of yeasts." Nucleic Acids Res **41**(21): 9705-9718.

Nieduszynski, C. A., Y. Knox and A. D. Donaldson (2006). "Genome-wide identification of replication origins in yeast by comparative genomics." Genes Dev **20**(14): 1874-1879.

Nishimura, K., M. Komiya, T. Hori, T. Itoh and T. Fukagawa (2018). "3D genomic architecture reveals that neocentromeres associate with heterochromatin regions." J Cell Biol.

Noma, K., H. P. Cam, R. J. Maraia and S. I. Grewal (2006). "A role for TFIIIC transcription factor complex in genome organization." Cell **125**(5): 859-872.

Nonaka, N., T. Kitajima, S. Yokobayashi, G. Xiao, M. Yamamoto, S. I. Grewal and Y. Watanabe (2002). "Recruitment of cohesin to heterochromatic regions by Swi6/HP1 in fission yeast." Nat Cell Biol **4**(1): 89-93.

Nurse, P., P. Thuriaux and K. Nasmyth (1976). "Genetic control of the cell division cycle in the fission yeast Schizosaccharomyces pombe." Mol Gen Genet **146**(2): 167-178.

Obersriebnig, M. J., E. M. Pallesen, K. Sneppen, A. Trusina and G. Thon (2016). "Nucleation and spreading of a heterochromatic domain in fission yeast." Nat Commun **7**: 11518.

Ohkuni, K. and K. Kitagawa (2011). "Endogenous transcription at the centromere facilitates centromere activity in budding yeast." Curr Biol **21**(20): 1695-1703.

Padmanabhan, S., J. Thakur, R. Siddharthan and K. Sanyal (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.

Palmer, D. K., K. O'Day, M. H. Wener, B. S. Andrews and R. L. Margolis (1987). "A 17-kD centromere protein (CENP-A) copurifies with nucleosome core particles and with histones." J Cell Biol **104**(4): 805-815.

Parker, M. W., M. R. Botchan and J. M. Berger (2017). "Mechanisms and regulation of DNA replication initiation in eukaryotes." Crit Rev Biochem Mol Biol **52**(2): 107-144.

Patel, P. K., B. Arcangioli, S. P. Baker, A. Bensimon and N. Rhind (2006). "DNA replication origins fire stochastically in fission yeast." Mol Biol Cell **17**(1): 308-316.

Pearson, C. G., E. Yeh, M. Gardner, D. Odde, E. D. Salmon and K. Bloom (2004). "Stable kinetochore-microtubule attachment constrains centromere positioning in metaphase." Curr Biol **14**(21): 1962-1967.

Petryk, N., M. Dalby, A. Wenger, C. B. Stromme, A. Strandsby, R. Andersson and A. Groth (2018). "MCM2 promotes symmetric inheritance of modified histones during DNA replication." Science.

Pidoux, A. L., S. Uzawa, P. E. Perry, W. Z. Cande and R. C. Allshire (2000). "Live analysis of lagging chromosomes during anaphase and their effect on spindle elongation rate in fission yeast." J Cell Sci **113 Pt 23**: 4177-4191.

Pohl, T. J., B. J. Brewer and M. K. Raghuraman (2012). "Functional centromeres determine the activation time of pericentric origins of DNA replication in Saccharomyces cerevisiae." PLoS Genet **8**(5): e1002677.

Polizzi, C. and L. Clarke (1991). "The chromatin structure of centromeres from fission yeast: differentiation of the central core that correlates with function." J Cell Biol **112**(2): 191-201.

Prasanth, S. G., K. V. Prasanth, K. Siddiqui, D. L. Spector and B. Stillman (2004). "Human Orc2 localizes to centrosomes, centromeres and heterochromatin during chromosome inheritance." EMBO J **23**(13): 2651-2663.

Prasanth, S. G., Z. Shen, K. V. Prasanth and B. Stillman (2010). "Human origin recognition complex is essential for HP1 binding to chromatin and heterochromatin organization." Proc Natl Acad Sci U S A **107**(34): 15093-15098.

Price, C. W., G. B. Fuson and H. J. Phaff (1978). "Genome comparison in yeast systematics: delimitation of species within the genera Schwanniomyces, Saccharomyces, Debaryomyces, and Pichia." Microbiol Rev **42**(1): 161-193.

Prioleau, M. N. and D. M. MacAlpine (2016). "DNA replication origins-where do we begin?" Genes Dev **30**(15): 1683-1697.

Radman-Livaja, M. and O. J. Rando (2010). "Nucleosome positioning: how is it established, and why does it matter?" Dev Biol **339**(2): 258-266.

Raghuraman, M. K., E. A. Winzeler, D. Collingwood, S. Hunt, L. Wodicka, A. Conway, D. J. Lockhart, R. W. Davis, B. J. Brewer and W. L. Fangman (2001). "Replication dynamics of the yeast genome." <u>Science</u> **294**(5540): 115-121.

Reuss, O., A. Vik, R. Kolter and J. Morschhauser (2004). "The SAT1 flipper, an optimized tool for gene disruption in Candida albicans." <u>Gene</u> **341**: 119-127.

Rhind, N. (2006). "DNA replication timing: random thoughts about origin firing." <u>Nat Cell Biol</u> **8**(12): 1313-1316.

Richet, N., D. Liu, P. Legrand, C. Velours, A. Corpet, A. Gaubert, M. Bakail, G. Moal-Raisin, R. Guerois, C. Compper, A. Besle, B. Guichard, G. Almouzni and F. Ochsenbein (2015). "Structural insight into how the human helicase subunit MCM2 may act as a histone chaperone together with ASF1 at the replication fork." <u>Nucleic Acids Res</u> **43**(3): 1905-1917.

Rocchi, M., N. Archidiacono, W. Schempp, O. Capozzi and R. Stanyon (2012). "Centromere repositioning in mammals." <u>Heredity (Edinb)</u> **108**(1): 59-67.

Rodrigues, C. F., S. Silva and M. Henriques (2014). "Candida glabrata: a review of its features and resistance." <u>Eur J Clin Microbiol Infect Dis</u> **33**(5): 673-688.

Romanowski, P., M. A. Madine, A. Rowles, J. J. Blow and R. A. Laskey (1996). "The Xenopus origin recognition complex is essential for DNA replication and MCM binding to chromatin." <u>Curr Biol</u> **6**(11): 1416-1425.

Rosic, S., F. Kohler and S. Erhardt (2014). "Repetitive centromeric satellite RNA is essential for kinetochore formation and cell division." <u>J Cell Biol</u> **207**(3): 335-349.

Ross, J. E., K. S. Woodlief and B. A. Sullivan (2016). "Inheritance of the CENP-A chromatin domain is spatially and temporally constrained at human centromeres." <u>Epigenetics Chromatin</u> **9**: 20.

Round, E. K., S. K. Flowers and E. J. Richards (1997). "Arabidopsis thaliana centromere regions: genetic map positions and repetitive DNA structure." <u>Genome Res</u> **7**(11): 1045-1053.

Roy, B., L. S. Burrack, M. A. Lone, J. Berman and K. Sanyal (2011). "CaMtw1, a member of the evolutionarily conserved Mis12 kinetochore protein family, is required for efficient inner kinetochore assembly in the pathogenic yeast Candida albicans." <u>Mol Microbiol</u> **80**(1): 14-32.

Sanyal, K., M. Baum and J. Carbon (2004). "Centromeric DNA sequences in the pathogenic yeast Candida albicans are all different and unique." <u>Proc Natl Acad Sci U S A</u> **101**(31): 11374-11379.

Sanyal, K. and J. Carbon (2002). "The CENP-A homolog CaCse4p in the pathogenic yeast Candida albicans is a centromere protein essential for chromosome transmission." <u>Proc Natl Acad Sci U S A</u> **99**(20): 12969-12974.

Schotta, G., A. Ebert, R. Dorn and G. Reuter (2003). "Position-effect variegation and the genetic dissection of chromatin regulation in Drosophila." <u>Semin Cell Dev Biol</u> **14**(1): 67-75.

Schubert, I. (2018). "What is behind "centromere repositioning"?" <u>Chromosoma</u> **127**(2): 229-234.

Schueler, M. G. and B. A. Sullivan (2006). "Structural and functional dynamics of human centromeric chromatin." <u>Annu Rev Genomics Hum Genet</u> **7**: 301-313.

Scott, K. C., S. L. Merrett and H. F. Willard (2006). "A heterochromatin barrier partitions the fission yeast centromere into discrete chromatin domains." <u>Curr Biol</u> **16**(2): 119-129.

Segurado, M., A. de Luis and F. Antequera (2003). "Genome-wide distribution of DNA replication origins at A+T-rich islands in Schizosaccharomyces pombe." <u>EMBO Rep</u> **4**(11): 1048-1053.

Shang, W. H., T. Hori, N. M. Martins, A. Toyoda, S. Misu, N. Monma, I. Hiratani, K. Maeshima, K. Ikeo, A. Fujiyama, H. Kimura, W. C. Earnshaw and T. Fukagawa (2013). "Chromosome engineering allows the efficient isolation of vertebrate neocentromeres." Dev Cell **24**(6): 635-648.

Shang, W. H., T. Hori, A. Toyoda, J. Kato, K. Popendorf, Y. Sakakibara, A. Fujiyama and T. Fukagawa (2010). "Chickens possess centromeres with both extended tandem repeats and short non-tandem-repetitive sequences." Genome Res **20**(9): 1219-1228.

Shivaraju, M., J. R. Unruh, B. D. Slaughter, M. Mattingly, J. Berman and J. L. Gerton (2012). "Cell-cycle-coupled structural oscillation of centromeric nucleosomes in yeast." Cell **150**(2): 304-316.

Singh, S. M., O. Steinberg-Neifach, I. S. Mian and N. F. Lue (2002). "Analysis of telomerase in Candida albicans: potential role in telomere end protection." Eukaryot Cell **1**(6): 967-977.

Sinha, P., V. Chang and B. K. Tye (1986). "A mutant that affects the function of autonomously replicating sequences in yeast." J Mol Biol **192**(4): 805-814.

Southern, E. M. (1975). "Detection of specific sequences among DNA fragments separated by gel electrophoresis." J Mol Biol **98**(3): 503-517.

Stanojcic, S., J. M. Lemaitre, K. Brodolin, E. Danis and M. Mechali (2008). "In Xenopus egg extracts, DNA replication initiates preferentially at or near asymmetric AT sequences." Mol Cell Biol **28**(17): 5265-5274.

Steiner, F. A. and S. Henikoff (2014). "Holocentromeres are dispersed point centromeres localized at transcription factor hotspots." Elife **3**: e02025.

Steiner, N. C., K. M. Hahnenberger and L. Clarke (1993). "Centromeres of the fission yeast Schizosaccharomyces pombe are highly variable genetic loci." Mol Cell Biol **13**(8): 4578-4587.

Stephens, A. D., J. Haase, L. Vicci, R. M. Taylor, 2nd and K. Bloom (2011). "Cohesin, condensin, and the intramolecular centromere loop together generate the mitotic chromatin spring." J Cell Biol **193**(7): 1167-1180.

Stevenson, J. B. and D. E. Gottschling (1999). "Telomeric chromatin modulates replication timing near chromosome ends." Genes Dev **13**(2): 146-151.

Stoler, S., K. C. Keith, K. E. Curnick and M. Fitzgerald-Hayes (1995). "A mutation in CSE4, an essential gene encoding a novel chromatin-associated protein in yeast, causes chromosome nondisjunction and cell cycle arrest at mitosis." Genes Dev **9**(5): 573-586.

Sugiyama, T., H. Cam, A. Verdel, D. Moazed and S. I. Grewal (2005). "RNA-dependent RNA polymerase is an essential component of a self-enforcing loop coupling heterochromatin assembly to siRNA production." Proc Natl Acad Sci U S A **102**(1): 152-157.

Sullivan, B. A., M. D. Blower and G. H. Karpen (2001). "Determining centromere identity: cyclical stories and forking paths." Nature Reviews Genetics **2**: 584.

Sullivan, D. J., J. Berman, L. C. Myers and G. P. Moran (2015). "Telomeric ORFS in Candida albicans: does mediator tail wag the yeast?" PLoS Pathog **11**(2): e1004614.

Sullivan, L. L., K. A. Maloney, A. J. Towers, S. G. Gregory and B. A. Sullivan (2016). "Human centromere repositioning within euchromatin after partial chromosome deletion." Chromosome Res **24**(4): 451-466.

Sun, X., H. D. Le, J. M. Wahlstrom and G. H. Karpen (2003). "Sequence analysis of a functional Drosophila centromere." Genome Res **13**(2): 182-194.

Sun, X., J. Wahlstrom and G. Karpen (1997). "Molecular structure of a functional Drosophila centromere." Cell **91**(7): 1007-1019.

Takahashi, K., E. S. Chen and M. Yanagida (2000). "Requirement of Mis6 centromere connector for localizing a CENP-A-like protein in fission yeast." <u>Science</u> **288**(5474): 2215-2219.

Taneja, N. and S. I. S. Grewal (2017). "Shushing histone turnover: It's FUN protecting epigenome-genome." <u>Cell Cycle</u> **16**(19): 1731-1732.

Taneja, N., M. Zofall, V. Balachandran, G. Thillainadesan, T. Sugiyama, D. Wheeler, M. Zhou and S. I. Grewal (2017). "SNF2 Family Protein Fft3 Suppresses Nucleosome Turnover to Promote Epigenetic Inheritance and Proper Replication." <u>Mol Cell</u> **66**(1): 50-62 e56.

Thakur, J. and S. Henikoff (2018). "Unexpected conformational variations of the human centromeric chromatin complex." <u>Genes Dev</u> **32**(1): 20-25.

Thakur, J. and K. Sanyal (2011). "The essentiality of the fungus-specific Dam1 complex is correlated with a one-kinetochore-one-microtubule interaction present throughout the cell cycle, independent of the nature of a centromere." <u>Eukaryot Cell</u> **10**(10): 1295-1305.

Thakur, J. and K. Sanyal (2012). "A coordinated interdependent protein circuitry stabilizes the kinetochore ensemble to protect CENP-A in the human pathogenic yeast Candida albicans." <u>PLoS Genet</u> **8**(4): e1002661.

Thakur, J. and K. Sanyal (2013). "Efficient neocentromere formation is suppressed by gene conversion to maintain centromere function at native physical chromosomal loci in Candida albicans." <u>Genome Res</u> **23**(4): 638-652.

Thompson, M., R. A. Haeusler, P. D. Good and D. R. Engelke (2003). "Nucleolar clustering of dispersed tRNA genes." <u>Science</u> **302**(5649): 1399-1401.

Topp, C. N., C. X. Zhong and R. K. Dawe (2004). "Centromere-encoded RNAs are integral components of the maize kinetochore." <u>Proc Natl Acad Sci U S A</u> **101**(45): 15986-15991.

Tsai, H. J., J. A. Baller, I. Liachko, A. Koren, L. S. Burrack, M. A. Hickman, M. A. Thevandavakkam, L. N. Rusche and J. Berman (2014). "Origin replication complex binding, nucleosome depletion patterns, and a primary sequence motif can predict origins of replication in a genome with epigenetic centromeres." <u>MBio</u> **5**(5): e01703-01714.

Tsankov, A. M., D. A. Thompson, A. Socha, A. Regev and O. J. Rando (2010). "The role of nucleosome positioning in the evolution of gene regulation." <u>PLoS Biol</u> **8**(7): e1000414.

van Hooff, J. J., E. Tromer, L. M. van Wijk, B. Snel and G. J. Kops (2017). "Evolutionary dynamics of the kinetochore network in eukaryotes as revealed by comparative genomics." <u>EMBO Rep</u> **18**(9): 1559-1571.

Verdel, A., S. Jia, S. Gerber, T. Sugiyama, S. Gygi, S. I. Grewal and D. Moazed (2004). "RNAi-mediated targeting of heterochromatin by the RITS complex." <u>Science</u> **303**(5658): 672-676.

Vernis, L., A. Abbas, M. Chasles, C. M. Gaillardin, C. Brun, J. A. Huberman and P. Fournier (1997). "An origin of replication and a centromere are both needed to establish a replicative plasmid in the yeast Yarrowia lipolytica." <u>Mol Cell Biol</u> **17**(4): 1995-2004.

Vogelauer, M., L. Rubbi, I. Lucas, B. J. Brewer and M. Grunstein (2002). "Histone acetylation regulates the time of replication origin firing." <u>Mol Cell</u> **10**(5): 1223-1233.

Voullaire, L. E., H. R. Slater, V. Petrovic and K. H. Choo (1993). "A functional marker centromere with no detectable alpha-satellite, satellite III, or CENP-B protein: activation of a latent centromere?" <u>Am J Hum Genet</u> **52**(6): 1153-1163.

Walker, J. E., M. Saraste, M. J. Runswick and N. J. Gay (1982). "Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold." EMBO J **1**(8): 945-951.

Warburton, P. E. (2004). "Chromosomal dynamics of human neocentromere formation." Chromosome Res **12**(6): 617-626.

Williams, J. S., T. Hayashi, M. Yanagida and P. Russell (2009). "Fission yeast Scm3 mediates stable assembly of Cnp1/CENP-A into centromeric chromatin." Mol Cell **33**(3): 287-298.

Wyrick, J. J., J. G. Aparicio, T. Chen, J. D. Barnett, E. G. Jennings, R. A. Young, S. P. Bell and O. M. Aparicio (2001). "Genome-wide distribution of ORC and MCM proteins in S. cerevisiae: high-resolution mapping of replication origins." Science **294**(5550): 2357-2360.

Xu, J., Y. Yanagisawa, A. M. Tsankov, C. Hart, K. Aoki, N. Kommajosyula, K. E. Steinmann, J. Bochicchio, C. Russ, A. Regev, O. J. Rando, C. Nusbaum, H. Niki, P. Milos, Z. Weng and N. Rhind (2012). "Genome-wide identification and characterization of replication origins by deep sequencing." Genome Biol **13**(4): R27.

Xu, W., J. G. Aparicio, O. M. Aparicio and S. Tavare (2006). "Genome-wide mapping of ORC and Mcm2p binding sites on tiling arrays and identification of essential ARS consensus sequences in S. cerevisiae." BMC Genomics **7**: 276.

Yadav, V., S. Sun, R. B. Billmyre, B. C. Thimmappa, T. Shea, R. Lintner, G. Bakkeren, C. A. Cuomo, J. Heitman and K. Sanyal (2018). "RNAi is a critical determinant of centromere evolution in closely related fungi." Proc Natl Acad Sci U S A **115**(12): 3108-3113.

Yamazaki, S., M. Hayano and H. Masai (2013). "Replication timing regulation of eukaryotic replicons: Rif1 as a global regulator of replication timing." Trends Genet **29**(8): 449-460.

Yan, H., S. Gibson and B. K. Tye (1991). "Mcm2 and Mcm3, two proteins important for ARS activity, are related in structure and function." Genes Dev **5**(6): 944-957.

Yao, J., X. Liu, T. Sakuno, W. Li, Y. Xi, P. Aravamudhan, A. Joglekar, Y. Watanabe and X. He (2013). "Plasticity and epigenetic inheritance of centromere-specific histone H3 (CENP-A)-containing nucleosome positioning in the fission yeast." J Biol Chem **288**(26): 19184-19196.

Zasadzinska, E., J. Huang, A. O. Bailey, L. Y. Guo, N. S. Lee, S. Srivastava, K. A. Wong, B. T. French, B. E. Black and D. R. Foltz (2018). "Inheritance of CENP-A Nucleosomes during DNA Replication Requires HJURP." Dev Cell.

Zhang, K., K. Mosch, W. Fischle and S. I. Grewal (2008). "Roles of the Clr4 methyltransferase complex in nucleation, spreading and maintenance of heterochromatin." Nat Struct Mol Biol **15**(4): 381-388.

Zhong, C. X., J. B. Marshall, C. Topp, R. Mroczek, A. Kato, K. Nagaki, J. A. Birchler, J. Jiang and R. K. Dawe (2002). "Centromeric retroelements and satellites interact with maize kinetochore protein CENH3." Plant Cell **14**(11): 2825-2836.

Zofall, M., T. Fischer, K. Zhang, M. Zhou, B. Cui, T. D. Veenstra and S. I. Grewal (2009). "Histone H2A.Z cooperates with RNAi and heterochromatin factors to suppress antisense RNAs." Nature **461**(7262): 419-422.

Zofall, M. and S. I. Grewal (2006). "RNAi-mediated heterochromatin assembly in fission yeast." Cold Spring Harb Symp Quant Biol **71**: 487-496.

8. LIST OF PUBLICATIONS

1. A comprehensive model to predict mitotic division in budding yeasts
   Sutradhar S., Yadav V., Sridhar S., **Sreekumar L.,** Bhattacharyya D., Ghosh S. K., …
   Sanyal K. (2015). A comprehensive model to predict mitotic division in budding
   yeasts. *Molecular Biology of the Cell*, *26*(22), 3954–3965.
   http://doi.org/10.1091/mbc.E15-04-0236

2. Centromere and Kinetochore: Essential Components for Chromosome Segregation
   (Book Chapter) Sridhar S., Dumbrepatil A., **Sreekumar L.,** Sankaranarayanan S.,
   Guin K., Sanyal K. (2017). Gene Regulation, Epigenetics and Hormone Signaling,
   FirstEdition. Edited by Subhrangsu S. Mandal.2017 Wiley-VCH VerlagGmbH& Co.
   KGaA.

3. Chromosome Components Important for Genome Stability in Candida albicans and
   Related Species (Book Chapter) **Sreekumar L.**, Varshney V. and Sanyal K (2017)
   Candida albicans: Cellular and Molecular Biology, Edited by R. Prasad, DOI
   10.1007/978-3-319-50409-4_12

4. Five pillars of centromeric chromatin in fungal pathogens (PEARLS Review).
    Yadav V, **Sreekumar L,** Guin K, Sanyal K (2018) *PLoS Pathog* 14(8): e1007150.
   https://doi.org/10.1371/journal.ppat.1007150

5. DNA replication initiator proteins facilitate CENPA loading on early replicating
   compact chromatin; *BioRxiv* (https://doi.org/10.1101/465880) **Lakshmi Sreekumar,**
   Priya Jaitly, Yao Chen, Bhagya C. Thimmappa, Amartya Sanyal, Leelavati Narlikar,
   Rahul Siddharthan and Kaustuv Sanyal

6. DNA replication initiator proteins facilitate CENPA loading on early replicating
   compact chromatin **Lakshmi Sreekumar,** Priya Jaitly, Yao Chen, Bhagya C.
   Thimmappa, Amartya Sanyal, Leelavati Narlikar, Rahul Siddharthan and Kaustuv
   Sanyal; *Genome Research* (under review)

A comprehensive model to predict mitotic division in budding yeasts

Sabyasachi Sutradhar[a],*, Vikas Yadav[b],*, Shreyas Sridhar[b],*, Lakshmi Sreekumar[b], Dibyendu Bhattacharyya[c], Santanu Kumar Ghosh[d], Raja Paul[a], and Kaustuv Sanyal[b]

[a]Department of Solid State Physics, Indian Association for the Cultivation of Science, Kolkata 700032, India; [b]Molecular Mycology Laboratory, Molecular Biology and Genetics Unit, Jawaharlal Nehru Centre for Advanced Scientific Research, Jakkur, Bangalore 560064, India; [c]Tata Memorial Centre, Advanced Centre for Treatment Research and Education in Cancer, Kharghar, Navi Mumbai 410210, India; [d]Department of Biosciences and Bioengineering, Indian Institute of Technology, Bombay, Powai, Mumbai 400076, India

ABSTRACT High-fidelity chromosome segregation during cell division depends on a series of concerted interdependent interactions. Using a systems biology approach, we built a robust minimal computational model to comprehend mitotic events in dividing budding yeasts of two major phyla: Ascomycota and Basidiomycota. This model accurately reproduces experimental observations related to spindle alignment, nuclear migration, and microtubule (MT) dynamics during cell division in these yeasts. The model converges to the conclusion that biased nucleation of cytoplasmic microtubules (cMTs) is essential for directional nuclear migration. Two distinct pathways, based on the population of cMTs and cortical dyneins, differentiate nuclear migration and spindle orientation in these two phyla. In addition, the model accurately predicts the contribution of specific classes of MTs in chromosome segregation. Thus we present a model that offers a wider applicability to simulate the effects of perturbation of an event on the concerted process of the mitotic cell division.

INTRODUCTION

Mitosis is a fundamental cellular process that enables faithful transmission of genetic material to the subsequent generation in eukaryotes. This process is well coordinated and requires the cumulative effort of several macromolecular machineries, including the centromere–kinetochore complex, the mitotic spindle, microtubule organizing centers (MTOCs), molecular motors, and microtubule-associated proteins (MAPs). The foundation for this process of

chromosome segregation is provided by a specialized chromatin structure, the centromere, upon which 60–80 proteins assemble to form the kinetochore (KT). The KT connects centromeric chromatin to the mitotic spindle. The mitotic spindle, nucleated by MTOCs, is a bipolar array of microtubules (MTs) that provides the force required to segregate chromosomes. This mitotic spindle is synergistically modulated by motor proteins (Mallik and Gross, 2004), the plus end–directed kinesins and the minus end–directed dyneins, and MAPs, which dynamically alter the rate of MT stability. The unequal rate of MT polymerization and depolymerization provides the push–pull forces that mediate poleward movement of segregated chromosomes into two daughter cells. Apart from requiring the assembly of the segregation machinery on the centromere and push–pull forces to enable chromosomes to segregate, proper spindle positioning and orientation is crucial for carrying out faithful segregation of chromosomes (Segal and Bloom, 2001; Kusch et al., 2002).

In most organisms, MTs are largely localized to the cytoplasm until spindle formation begins during mitosis. These cytoplasmic MTs (cMTs) emanate from either multiple cytoplasmic MTOCs, as in metazoans, or from a single nuclear envelope (NE)-embedded MTOC, as in the budding yeast *Saccharomyces cerevisiae*. The cMTs, along with motor proteins, influence nuclear positioning and

movement (Lee *et al.*, 2000; Fink *et al.*, 2006; Ten Hoopen *et al.*, 2012). On the onset of mitosis, cMTs reorganize themselves to form the mitotic spindle between the two poles (spindle pole bodies [SPBs] in yeast or centrosomes in metazoans). The less dynamic minus ends of MTs are anchored to the SPBs, while the more dynamic plus ends radiate outward to facilitate interactions with other cellular components. Some of these MTs interact with KTs to become KT MTs (kMTs) and provide the pulling force on chromosomes during anaphase. Cytoplasmic (astral) MTs make contact with the cell cortex, aiding in spindle positioning, while interpolar MTs (ipMTs) are formed when the plus ends of MTs originating from opposite poles interact via sliding, resulting in an antiparallel array at the midzone. The combination of pushing force provided by ipMTs on SPBs along with the pulling force from kMTs and cMTs aids in segregation during anaphase.

The spindle positioning is not only crucial for proper chromosome segregation but also defines the site of division. In general, the spindle is positioned centrally in the dividing cell, and thus a mother cell gives rise to equal-sized daughter cells by the fission mode of division. While most organisms undergo this type of division, a few show variations in spindle positioning and hence give rise to cell polarity (Horvitz and Herskowitz, 1992; Neumuller and Knoblich, 2009). This type of division is mostly observed during developmental stages of multicellular organisms and in stem cells (Knoblich, 2008; Neumuller and Knoblich, 2009). Budding yeasts also undergo a similar unequal cell division, in which the site of division is defined before spindle positioning (Fraschini *et al.*, 2008). A number of studies have been carried out to identify the factors that affect the dynamics of spindle positioning. Some of these regulatory factors are shown to be different between budding yeasts and multicellular organisms (Fraschini *et al.*, 2008; Neumuller and Knoblich, 2009).

The process of chromosome segregation during budding has been well studied in *S. cerevisiae* and *Candida albicans*, both belonging to the fungal phylum Ascomycota. Recently these processes were studied in another phylum of fungi, Basidiomycota, represented by the yeasts *Cryptococcus neoformans* and *Ustilago maydis.* Although these organisms belong to two major fungal phyla and divide by budding, a striking variation is observed regarding the site of nuclear division that takes place in the mother cell in ascomycetes but in the newly budded daughter cell in basidiomycetes (Heath, 1980; Straube *et al.*, 2005; Gladfelter and Berman, 2009; Kozubowski, Yadav, *et al.*, 2013). In ascomycetes, the nucleus moves close to mother–daughter cell junction (neck) and divides into two equal halves. One half then moves to the daughter cell, and the other half is retained in the mother cell. In contrast, the nucleus moves completely to the daughter cell before division in basidiomycetes. Nuclear division takes place in the daughter cell, after which a divided nuclear mass moves back to the mother cell, while the other half is retained in the daughter cell. To address the molecular basis for this observed variability in mitosis between these yeast species, we first developed a common computational model that was subsequently modified to simulate the fungal phylum–specific nuclear dynamics during mitosis in ascomycetes or basidiomycetes. Mitosis has been studied extensively in several ascomycetous yeasts. To begin with, we established a computational model with available parameters that are well characterized for ascomycetes and then introduced varying parameters measured in vivo for both ascomycetes and basidiomycetes to develop two independent models. These models predict that cMT bias is required for directional nuclear movement in both ascomycetes and basidiomycetes. Both the models also accurately simulate the altered conditions that prevail during mitosis upon treatment of cells with various MT-specific depolymerizing drugs. We conclude that the models developed in this study offer a wider application toward understanding the consequences of not only short-lived mitotic events but also the consequences of small perturbations in the entirety of the mitotic cycle.

RESULTS

Generation of a computational model that replicates in vivo parameters of mitotic events

To model mitosis in budding yeast, we considered simplified versions of several cellular components known to play a role in chromosome segregation, including 1) SPB/MTOCs; 2) the centromere–KT complex; 3) cohesin complexes connecting sister chromatids before anaphase; 4) the MT network consisting of kMT, ipMT, and cMT; 5) cell cortex and cortical dyneins modulating cMT dynamics; and 6) kinesins involved in sliding overlapping ipMTs (Figure 1, A and B, and Table 1). The mother cell was considered an ellipsoid, while the nucleus was considered a spherical object placed randomly within the mother cell at the onset of simulations. To mimic the experimental scenario, we resorted to the same geometrical parameters of the mother cell as observed in our experiments (Table 1). Budding was initiated at a random location on the surface of the mother cell growing at the experimentally observed growth speed. MTs were modeled as straight filaments, and MT dynamics was replicated by incorporating stochastic switching between growing (lengthening) and shrinking (shortening) states by using standard computational techniques (see *Materials and Methods*). The cell cortex was taken as a rigid wall that resists free polymerization of the cMTs by applying a resistive force at the cMT tip. During mitosis, cortically anchored dynein motors that walk toward the minus end of the cMT generate a pulling force on the SPB and provide directional movement of the nucleus/SPB toward the cMT tip (Figure 1, A and B; see *Materials and Methods* for details). It is widely believed that cMT–cortex interactions play a vital role in nuclear migration in yeasts during mitosis (Carminati and Stearns, 1997; Adames and Cooper, 2000; Baumgartner and Tolic, 2014). We incorporated the idea of asymmetric loading of different protein molecules between the SPBs to ensure a biased nucleation of the cMTs in the model (Markus *et al.*, 2012). Because we sought to understand the difference between the steady-state positioning of the spindles in ascomycetes and basidiomycetes, we ignored their instantaneous dynamics in this particular study.

It is observed that KTs remain clustered in yeasts during mitosis (Jin *et al.*, 2000; Anderson *et al.*, 2009; Kozubowski, Yadav, *et al.*, 2013; Varoquaux *et al.*, 2015). Hence, to avoid overlapping of individual KTs, each KT was modeled to have a hard-core excluded volume. Assembly of sister KTs occurred immediately after centromere replication. To simplify the model, we assumed newly assembled sister KTs were captured instantaneously. It is observed that, in yeasts, the KTs always remain attached to the SPB during mitosis, except for a few minutes during chromosome duplication (Tanaka *et al.*, 2005, 2010; Tanaka and Tanaka, 2009; Gandhi *et al.*, 2011). Compared with mammalian cells, the KT capture process in ascomycetes occurs faster. Considering that nuclear migration per se is a much slower process than KT capture, instantaneous capturing of KTs is not expected to change our model prediction. At the ipMT overlap region, plus end–directed molecular motors slide the MTs apart, generating a pushing force on the SPBs (Kapoor and Mitchison, 2001; Marco, Dorn, *et al.*, 2013). The kMT–KT interaction is mediated by spring-like KT fibrils (McIntosh *et al.*, 2008). Before anaphase, two opposing forces on the KTs, an outward pulling force
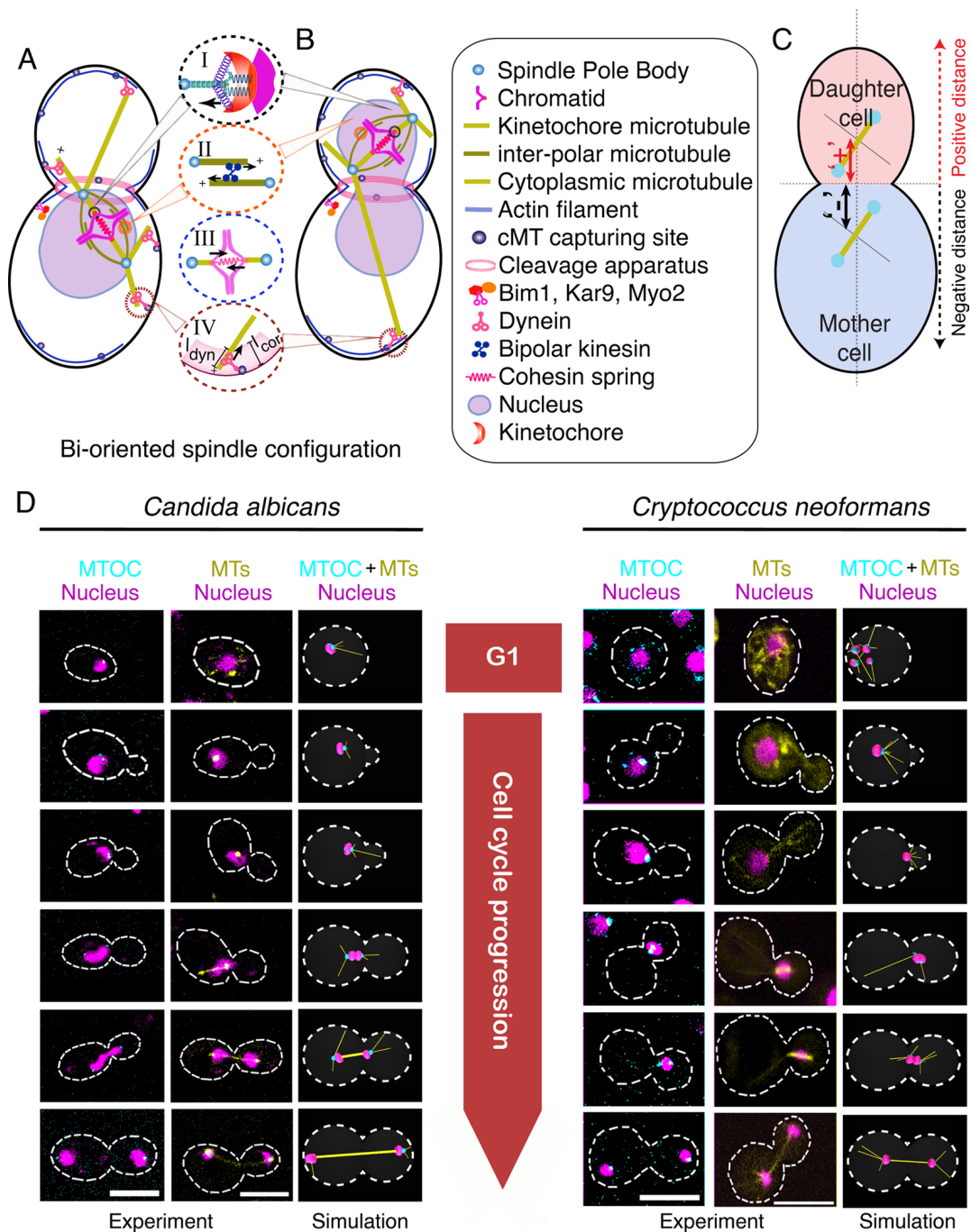
FIGURE 1: Model development to study mitotic progression in ascomycetes and basidiomycetes. (A and B) Schematic of biorientation of chromosomes that occurs within (A) the mother cell in *C. albicans* and (B) the daughter cell in *C. neoformans*. Various forces are responsible for proper biorientation in both these organisms and are depicted in detail as follows: Depolymerization of kMTs at the KT enables poleward movement of chromosome (circle I). KTs interact with kMTs through spring-like attachments that regulate kMT dynamics. MT depolymerization at the KT pulls chromosomes toward the SPB. Before anaphase, this poleward force is countered by the sliding force generated by the plus end–directed kinesins acting along the ipMTs (circle II) and the cohesive force between the sister chromatids, which is considered as a spring between sister chromatids (circle III). cMTs interact with the cell cortex, where dynein pulls SPBs toward the cortex (circle IV). (C) The representative sign convention for labeling the spindle distance from the neck. (D) Cell cycle phase–specific dynamics of nucleus, MTOCs, and MTs in ascomycetes (*C. albicans*) and basidiomycetes (*C. neoformans*) was monitored by imaging a GFP-tagged component of MTOC or MTs along with nuclear dynamics, represented by DAPI-stained nuclei in *C. albicans* and mCherry-tagged histone H4 in *C. neoformans*. In *C. albicans*, a single MTOC, visible in unbudded cells, forms two active SPBs during S phase (small budded cells). The duplicated SPBs then migrate away from each other to establish a bipolar spindle (~1.2 μm) in the mother cell during metaphase (large budded cells). In *C. neoformans*, multiple foci of MTOCs are observed at the beginning of the cell cycle. Observed MTOC foci merge together toward the onset of mitosis, forming an active SPB. After duplication, the SPBs migrate into the daughter bud and then establish a bipolar spindle evidenced by an increase in the distance (~1.6 μm) between the SPBs. The nucleus, MTOCs, and MTs are false colored as magenta, cyan, and yellow, respectively. Scale bars: 5 μm.

Abbreviation	Meaning	Value for ascomycetes	Value for basidiomycetes	Reference
N_{KT}	Number of KTs in haploid cell	16 (*S. cerevisiae*)	14 (*C. neoformans*)	Our experiment
N_{cMT}	Number of cMTs	4	$\propto 4\pi r_{MTOC}^2$	Kosco *et al.*, 2001
a_{cell}, b_{cell}, c_{cell}	Dimension of the cell	2.5–3.0 μm	2.50–3.0 μm,	Our experiment
l_{cor}	Width of cortex	0.2 μm	0.2 μm	Rodal *et al.*, 2005
K_{cor}	Spring constant of the cortex	5.0 pN/μm	5.0 pN/μm	This study
r_{nu}	Initial radius of the nucleus	1.0 μm	1.0 μm	Our experiment
r_{SPB}	Radius of single SPB	0.125 μm	0.125 μm	Seybold and Schiebel, 2013; Lee *et al.*, 2014
r_{KT}	Radius of single KT	0.05 μm	0.05 μm	Haase, Mishra, *et al.*, 2013
v_g, v_s	Growth, shrinkage velocity of MT	6.4 μm min^{-1}, 26.6 μm min^{-1}	10.4 μm min^{-1}, 28.6 μm min^{-1}	Fink *et al.*, 2006; Finley *et al.*, 2008
f_c, f_r	Catastrophe, rescue frequency of MT	0.34 min^{-1}, 0.02 min^{-1}	1.0 min^{-1}, 0.02 min^{-1}	Fink *et al.*, 2006; Finley *et al.*, 2008
f_s	Stall force of MT	1.7 pN	1.7 pN	Dogterom and Yurke, 1997
f^s_{dyn}	Force produced by single dynein	1.0 pN	1.0 pN	Muller *et al.*, 2008; Soppina *et al.*, 2009
λ_{dyn}	Density of dynein per unit length per MT	6.0 /μm	6.0/μm	Civelekoglu-Scholey *et al.*, 2006
λ_{ipMT}	Density of ipMT motor per unit length	1.0 /μm	1.0/μm	Civelekoglu-Scholey *et al.*, 2006
f_{ipMT}	Force produced by single ipMT motors	1.0 pN	1.0 pN	This study
η_{cell}	Viscosity of cytoplasm	5.0 pN s/μm^2	5.0 pN s/μm^2	Civelekoglu-Scholey *et al.*, 2006
η_{nu}	Viscosity of nucleoplasm	10.0 pN s/μm^2	10.0 pN s/μm^2	Civelekoglu-Scholey *et al.*, 2006
η_{NE}	Viscosity of NE	10.0 pN s/μm^2	10.0 pN s/μm^2	Civelekoglu-Scholey *et al.*, 2006
$K_{cohesion}$	Spring constant of the cohesion springs	0.1 pN/μm	0.1 pN/μm	Joglekar and Hunt, 2002
K_C	Spring constant of the KT–kMT attached springs	10.0 pN/μm	10.0 pN/μm	Civelekoglu-Scholey *et al.*, 2006; Sau *et al.*, 2014
$K_{fibrils}$	Spring constant of the KT fibrils	5.0 pN/μm	5.0 pN/μm	Civelekoglu-Scholey *et al.*, 2006; Sau *et al.*, 2014
C	Repulsion strength of KTs	1.0 pN/μm	1.0 pN/μm	This study

TABLE 1: Various parameters used to develop the model.

toward SPBs, driven by motor proteins and depolymerizing kMTs, and an inward cohesive force between the sister chromatids due to cohesin proteins must be balanced to satisfy the spindle-assembly checkpoint and subsequent entry to anaphase. For maintaining the experimentally observed separation between the KT cluster and SPBs, a length-dependent catastrophe of kMTs was incorporated (Foethke *et al.*, 2009; Sau *et al.*, 2014).

The nuclear mass always divides close to the mother bud junction in budding yeasts

The nuclear division or mitotic spindle formation in ascomycetes takes place in the mother cell, whereas it occurs in the daughter cell

in basidiomycetes (Heath, 1980; Straube *et al.*, 2005; Gladfelter and Berman, 2009; Kozubowski, Yadav, *et al.*, 2013; Figure 1D, fourth and fifth row). Although the spindle positioning in ascomycetes is a relatively well-studied process (Piatti *et al.*, 2006; Merlini and Piatti, 2011), very little is known about the same in basidiomycetes. We used fluorescence microscopy to understand the spindle and nuclear dynamics simultaneously in these two classes of yeasts. In ascomycetes, represented by *S. cerevisiae* and *C. albicans* henceforth, only one visible interphase MTOC serves as the SPB during mitosis (Figure 1D; Segal and Bloom, 2001). In contrast, among basidiomycetes, henceforth represented by *C. neoformans* and *U. maydis*, several MTOCs were seen spread throughout the cytoplasm during

interphase (Figure 1D; Straube *et al.*, 2003). These MTOCs subsequently coalesced to form an active SPB during mitosis.

In premitotic cells, SPBs are localized at a constant distance from each other after duplication, which segregates rapidly during the onset of mitosis (Figure 1D). The distance from the center of the mitotic spindle to the neck was measured in a number of cells (*n* = 30) during metaphase and early anaphase. The neck was taken as the origin, and the distance was marked as (+) or (–) for the presence of the spindle in the daughter cell or mother cell, respectively, during mitosis (Figure 1C). The net average neck–spindle distance for ascomycetes (–1 ± 0.22 μm) was found to be similar in basidiomycetes (+0.84 ± 0.23 μm). Thus the nucleus was found to be positioned close to the neck during mitosis, irrespective of the dynamics of nuclear movement in premitotic stages. In other words, the cellular machinery divides the nuclear mass into two equal halves across the neck in a well-conserved manner, irrespective of its earlier dynamics. The data obtained from these experiments and previously published results (Table 1) were incorporated into the universal model for mitosis described above to yield two working models, one each for ascomycetes and basidiomycetes (Figure 1, A and B, and Supplemental Videos 1 and 2).

Nuclear/spindle dynamics depends on the number of cMTs and dynein activity

Having developed these models, we probed for the underlying variation in nuclear migration observed between ascomycetes and basidiomycetes. Differential migration patterns and a large deformation of the nucleus during migration suggested that the magnitude of force pulling SPBs toward the bud is greater in basidiomycetes compared with ascomycetes (Straube *et al.*, 2005; Fink *et al.*, 2006; Kozubowski, Yadav, *et al.*, 2013). The larger force generated could either be due to an increased population of cMTs and/or a higher dynein activity at the cortical region. It is widely believed that ascomycetes nucleate ~4 cMTs (Kosco *et al.*, 2001), whereas the number of cMTs in basidiomycetes is unknown. A previous study using *U. maydis* showed that the number of MTs in this organism is 10–15, indicating a higher number of cMTs in basidiomycetes (Straube *et al.*, 2003).

Considering a conserved cMT–cortex interaction, our model revealed that the size of the cMT population must be more than eight for producing sufficient force to pull the nucleus into the daughter cell (Figure 2A). Assigning the number of cMTs as four for ascomycetes and eight for basidiomycetes, simulations predicted the mean distances between the neck and the spindle as –0.90 and +0.83 μm, respectively. These values are close to the experimental measurements (Figure 2B). Further, an increase in the density of cortical dyneins engaged in pulling the cMTs also provided enough pulling force for the migration of the nucleus into the daughter cell in the basidiomycetes model when other parameters were kept constant (Figure 2C). To test the model's prediction of requiring a greater number of cMTs in basidiomycetes for migration of the nucleus into the daughter cell, we counted the cMTs in *C. albicans* (Figure 2D) and *C. neoformans* (Figure 2E; see *Materials and Methods*). Our experiments revealed that *C. neoformans* has an approximately at least two times higher number of cMTs than *C. albicans* (Figure 2, D–F). It was observed that approximately six to 15 cMTs formed a dense mesh-like network in *C. neoformans*, with an average number of cMTs per cell being approximately nine (Figure 2F), while each *C. albicans* cell has three to five cMTs with an average of approximately four cMTs per cell (Figure 2F). The results presented above confirm the importance of cMT and dynein in positioning the spindle. Disruption of any of these components

leads to severe mitotic defects (Markus and Lee, 2011; Laan *et al.*, 2012; Xiang, 2012; Best *et al.*, 2013). Thus our model prediction, supported by experimental validation, confirms that an increased number of cMTs is required for migration of the nucleus/SPB into the daughter cell.

A biased "search and capture" by the cMTs is required for proper nuclear migration and spindle alignment

Next we analyzed whether the nuclear migration was governed by a random or polarized nucleation of cMTs. To test this, we performed simulations on both the models, keeping either an unbiased or a biased nucleation toward the daughter cell (Figure 3). The directional movement of the nucleus was impaired for an unbiased nucleation of cMTs, but the metaphase spindle length was found to be independent of cMT bias (Figure 3, A–D, and Supplemental Figure S1 and Videos 3 and 4). In both the cases, however, a biased dynamics of cMTs was crucial for proper nuclear migration (Figure 3, C and D, and Supplemental Figure S1). The resulting spindle–neck distances were found to be similar to the experimental values. Basidiomycetes showed some directional movement of the nucleus even in the absence of biased nucleation, which we attribute to a higher number of cMTs when compared with ascomycetes (Figures 2F and 3D). An unbiased cMT dynamics also failed to align the spindle with the mother–daughter cell axis, adding to the severity of the defect in these cases (Figure 3, E and F).

Clearly, biased cMTs produce a directed force on the nucleus/SPB, whereas uniformly nucleated cMTs generate force without any preferred directionality that therefore often fails to move the nucleus/SPB to the predefined position. Many studies revealed that a cortical actin-dependent mechanism, known as the "Kar9 pathway," utilizes a myosin-V, Myo2-based machinery to guide the plus ends of cMTs along the cortex toward the neck at the early stage of the cell cycle (Beach *et al.*, 2000; Yin *et al.*, 2000). Reports also suggest that asymmetric loading of Kar9 at the SPB can produce a chemical cue that leads to a biased nucleation of cMTs toward the neck (Liakopoulos, Kusch, *et al.*, 2003; Cepeda-Garcia *et al.*, 2010). The current model exploits these results, providing an additional line of evidence for the same.

The model accurately reproduces experimental outcomes of various drugs affecting MT dynamics

Next we tested the model by simulating the effects of two drugs that are known to affect the dynamics of specific classes of MTs in vivo. To depolymerize all MTs present in the cell, the MT-depolymerizing drug nocodazole was used, whereas methyl benzimidazole carbamate (MBC; Akera *et al.*, 2012) was used to disrupt ipMTs specifically. The depolymerizing kinetics for nocodazole treatment was simulated by increasing the catastrophe frequency of all the MTs in the model. Similarly, to simulate the effect of MBC, ipMT catastrophe was increased without altering the dynamics of cMTs or kMTs.

Our model accurately simulates the effect of nocodazole treatment, resulting in shorter spindles that failed to move to the bud (Figure 4 and Supplemental Videos 5 and 6). For both ascomycetes and basidiomycetes, the spindle length was drastically shorter than untreated cells (Figure 4, A and B). We also observed that the spindle is mispositioned and misaligned, as depicted by the higher mean values for the neck to spindle distance and a higher spindle orientation angle (Figure 4, C–F). Misaligned spindles are identified as those making angles greater than 30° with the mother–daughter axis, while mispositioned spindles position themselves farther than 1 μm away from the neck. Simulations for higher nocodazole concentration were achieved by increasing the catastrophe frequency
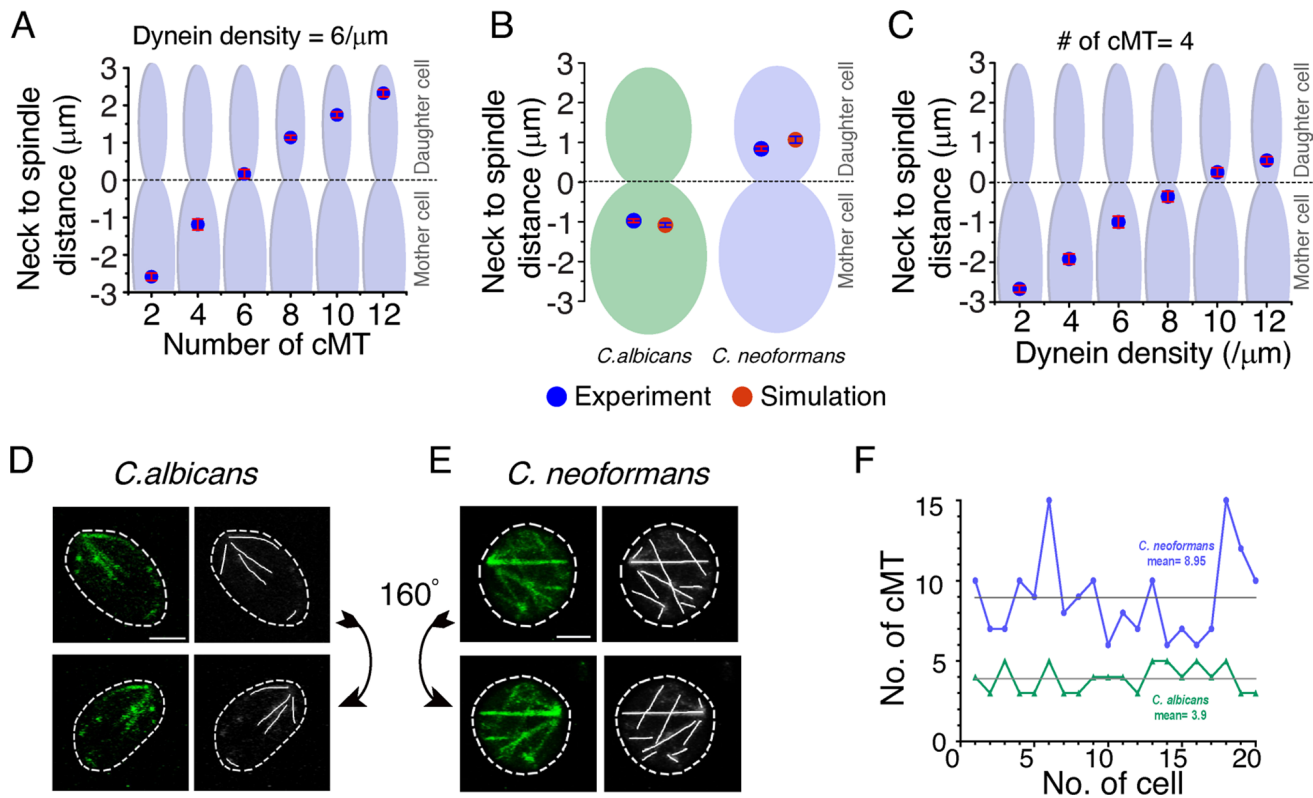
FIGURE 2: Dependence of nuclear migration on the number of cMTs and dynein activity in ascomycetes and basidiomycetes. (A) In silico measurements of the neck to spindle distance upon altering cMT numbers per cell during mitosis. We observed that, for a fixed density of cortical dynein, a higher number of cMTs leads to a deeper penetration of the spindle into the daughter cell (bud). With four cMTs, the observed spindle distance from the neck is close to −1.0 μm, which is similar to that observed in ascomycetes. As we increase the number of cMTs, the spindle moves closer to the neck, and when there are six cMTs, the spindle just crosses over into the daughter cell. The spindle is strongly pulled and moved deep into the daughter cell when the number of cMTs is eight or more, resembling what is observed in experiments. (B) Mean distance of the spindle from neck is plotted as observed in simulation (n = 100) and experiments (n = 30) for both *C. albicans* and *C. neoformans*. Experimental measurements were carried out in a strain that had MTOCs tagged with GFP. Differential interference contrast was used as a reference point for calculating spindle mid to neck distance. The mean distance of the spindle from the neck in *C. albicans* is estimated as −1.0 ± 0.22 μm in experiments, while our in silico model prediction with four cMTs turns out to be −1.0 ± 0.02 μm. On the other hand, the spindle to neck distance in *C. neoformans* with eight or more cMTs is found to be +0.84 ± 0.23 μm and +1.0 ± 0.05 μm from experiments and in silico measurements, respectively. (C) The spindle migration can also be affected by an alternative pathway involving cortical dyneins. An increase in the cortical dynein density for a fixed number of cMTs results in similar nuclear dynamics obtained previously by altering the cMT number. SEM is shown in red bars. (D and E) *C. albicans* (YJB12856) and *C. neoformans* (CNVY109) strains expressing Tub1-GFP were used to monitor and estimate cMTs. To rule out false positives in counting, we used high-resolution three-dimensionally rendered images to trace cMTs before estimation of their numbers. The cMTs in all stacks were taken into consideration. Two different views over the y-axis (0°, top panels; 160°, bottom panels) of the three-dimensionally rendered images are shown to improve the visibility of cMTs that may be masked by others in a given orientation. Scale bar: 2 μm. (F) The cMTs were counted in a large number of cells of *C. albicans* and *C. neoformans*. These values were plotted, and the calculated mean of cMTs per cell in each case is represented by a gray line. *C. neoformans* was found to contain six to 15 cMTs per cell, with an average of 8.95, while *C. albicans* was found to contain three to five cMTs per cell, with an average of 3.9.

of the MTs (8–12/min), resulting in shorter MTs. Under these circumstances, initial MTOC and KT clustering in basidiomycetes were highly affected due to altered cMT dynamics (Supplemental Figure S2). These simulated results correlated with the experimental results and accurately corroborated the in vivo observation of KTs failing to cluster upon nocodazole treatment, as reported previously (Kozubowski, Yadav, *et al.*, 2013).

After treatment with MBC, the spindle length was found to be shortened in basidiomycetes (Figure 4B and Supplemental Video 7). This effect was less drastic as compared with nocodazole, as the spindle length was longer for MBC-treated cells. Similar to the

nocodazole treatment, mispositioned and misaligned spindles were also observed (Figure 4, D and F) when treated with MBC. These results suggest that the role of ipMTs is not crucial for SPB separation, but they are required for spindle migration and orientation in basidiomycetes. The effect of MBC treatment could not be examined in ascomycetes, because MBC does not affect the cell cycle events of *C. albicans*, as reported earlier (Finley and Berman, 2005). Taken together, the consequences of experimental perturbation of various species of MTs on chromosome segregation are accurately simulated in the model. On the basis of these results, we conclude that the model developed is a robust one and
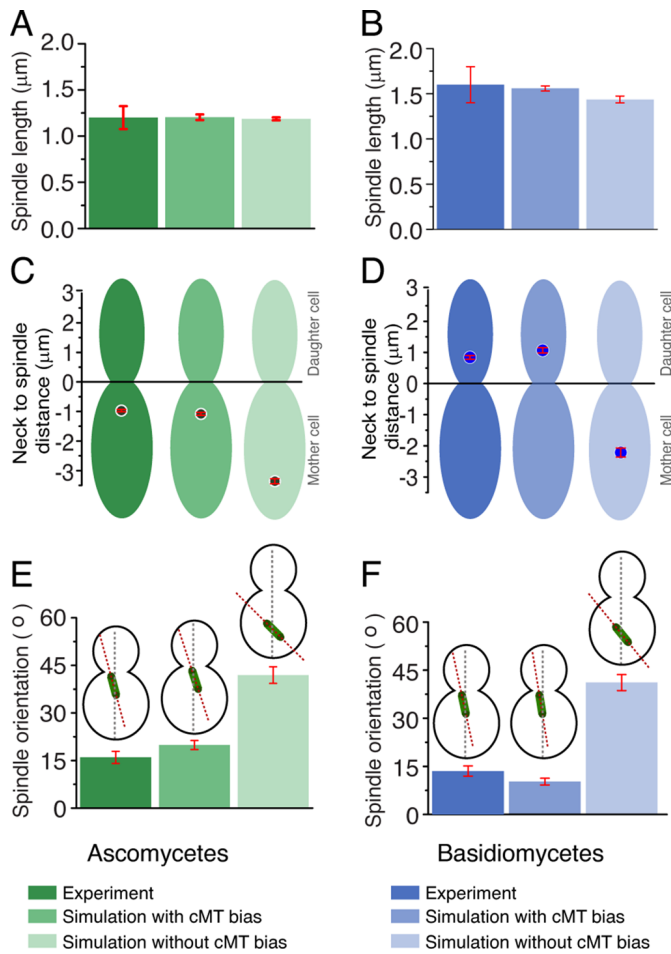
FIGURE 3: Biased vs. unbiased MT dynamics in maintaining spindle length, position, and orientation in ascomycetes and basidiomycetes. (A and B) The spindle length at metaphase was calculated in the model with biased or unbiased cMTs for both *C. albicans* and *C. neoformans*. Experimental measurements were carried out during mitosis to measure the spindle length. The model predicts that the spindle length at metaphase remains unaltered irrespective of cMT bias, as experimentally observed. (C and D) Neck to spindle distance is measured with or without cMT bias, while simultaneously comparing it with experimental data (wild type). With unbiased cMT, the spindle often failed to move to the predefined spatial positions in both ascomycetes and basidiomycetes. The spindle–neck mean distance changes from −1.0 to −3.35 μm during unbiased nucleation in ascomycetes. The spindle to neck mean distance in basidiomycetes, changes from +1.0 to −2.2 μm during unbiased nucleation. (E and F) Orientation of the spindle is calculated by measuring its tilt with respect to the mother–daughter cell axis. Unbiased cMT dynamics result in a larger angular orientation (~42° in ascomycetes and ~41° in basidiomycetes), with the mother–daughter cell axis reflecting the misaligned spindles for unbiased cMT dynamics. For biased cMT dynamics, the spindle aligned with the mother–daughter cell axis, and the angular orientation is measured as ~21° ± 2.1° for ascomycetes and ~10.25° for basidiomycetes. Red bars indicate SEM.

can replicate the events observed in the in vivo experiments with high precision.

DISCUSSION

In this study, we describe a model that accurately simulates the events of mitosis in distantly related budding yeasts belonging to the phyla Ascomycota and Basidiomycota. We also sought to
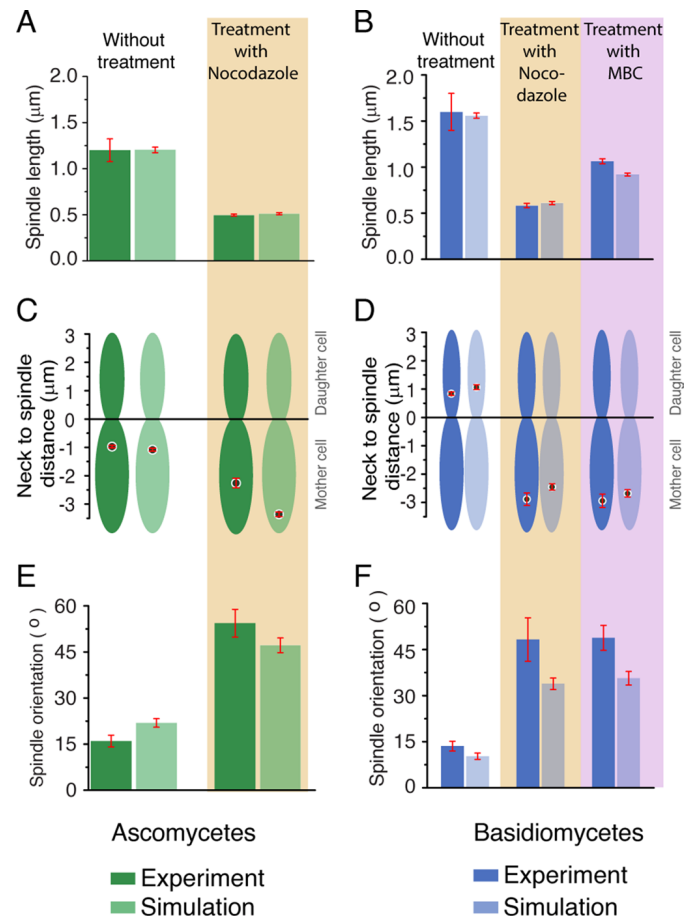
FIGURE 4: Comparison of in vivo and in silico results upon altering dynamics of MTs in ascomycetes and basidiomycetes. (A and B) Metaphase spindle lengths upon treatment with nocodazole or MBC are plotted for ascomycetes and basidiomycetes along with the unperturbed (dimethyl sulfoxide [DMSO] control) numerical and experimental data. For ascomycetes, we observed that, upon nocodazole treatment, the spindle length becomes ~0.50 ± 0.04 μm, which is in accordance with our model prediction ~0.52 ± 0.01 μm. Similarly, in basidiomycetes, the spindle length is reduced to ~0.61 μm from its wild-type spindle length of ~1.66 μm. This is in accordance with the experimental value 0.62 ± 0.04 μm. For MBC treatment, in basidiomycetes, the spindle length is shortened to 0.9 μm from its native value of 1.6 μm. This in silico result is in agreement with the experimental data for basidiomycetes as shown. (C and D) Measurements of spindle to neck distances for ascomycetes (with nocodazole) and basidiomycetes (either with nocodazole or MBC) revealed the inability of the spindles to move to their unperturbed spatial locations. The spindle always remained in the mother cell with an increased mean distance (~ −2.7 μm for nocodazole) from the neck compared with its wild-type value (~ −1.0 μm) in ascomycetes. In basidiomycetes, the spindle failed to move to the bud, and always remained in the mother cell during either nocodazole or MBC treatment. (E and F) The spindle orientation in the presence or absence of drugs was measured, and it was found to misalign with mother–daughter cell axis, as shown by the higher spindle orientation angle in both cases. Red bars indicate SEM.

understand the basis behind the differences in mitotic events observed in these two phyla. A universal model was developed for the budding mode of division, following which mitotic events were simulated and modeled both for ascomycetes and basidiomycetes by obtaining parameters either from the literature or through

experimental measurements. When this information was compared, we observed that variations in the MT organization, orientation, and dynamics account for most of the variations in mitotic events observed between these two classes of yeasts.

While making experimental measurements for basidiomycetes using *C. neoformans*, it was observed that, although the spindle migrates entirely to the daughter cell during mitosis, it is always positioned close to the neck at metaphase. Similarly in ascomycetes, the nucleus moves very close to the neck, where the division takes place during mitosis. This indicates that the site of nuclear division with respect to the site of cytoplasmic division remains conserved in these two classes of yeasts in spite of the other observed differences. In metazoans, fission yeast, and filamentous forms of many fungi (including the ones studied here), the site of nuclear division defines the site of cytoplasmic division (Balasubramanian *et al.*, 2000; Guertin *et al.*, 2002; Wang *et al.*, 2003; Gladfelter and Berman, 2009). Hence cells seem to have developed a mechanism, as nuclear division takes place close to the predefined cell cleavage site in budding yeasts. The interaction of MTs with septin proteins and other cleavage elements plays a determinant role in this process (Castillon *et al.*, 2003; Rodal *et al.*, 2005). Indeed, several reports in *S. cerevisiae* have shown that positioning of the spindle close to the neck is important for accurate chromosome segregation (Piatti *et al.*, 2006; Merlini and Piatti, 2011). In the absence of proper alignment and positioning of the mitotic spindle, the spindle positioning checkpoint (SPOC) is activated in these cells delaying chromosome segregation (Piatti *et al.*, 2006; Fraschini *et al.*, 2008). Hence this conserved distance observed between the neck to the spindle is possibly due to the SPOC activity that uses cMTs to monitor the location of the spindle in dividing cells (Moore *et al.*, 2009). Such a strict positioning of the spindle in basidiomycetes might reflect the conservation of the regulatory process.

The cytoskeletal elements, primarily MTs and their accessory network of proteins, have been shown to influence nuclear migration (Hwang *et al.*, 2003; Straube *et al.*, 2003; Martin *et al.*, 2004; Fink *et al.*, 2006; Gladfelter and Berman, 2009; Markus *et al.*, 2012; Kozubowski, Yadav, *et al.*, 2013). Our model revealed that nuclear migration toward the daughter cell would occur only if cMTs organized themselves in a biased manner in the direction of the newly emerging daughter cell. This was found to be applicable for both ascomycetes and basidiomycetes. Previously, several reports indicated that actin and other cytoskeleton elements reorganize during budding in yeasts, giving rise to cell polarity (Pruyne and Bretscher, 2000a,b). We show here that polarized MT nucleation plays an important role during the process of mitosis. This hypothesis is supported by experimental observations suggesting that an asymmetric recruitment of proteins (Myo2, Bim1, Kar9, etc.) may guide the cMTs to grow and stabilize toward the emerging daughter cell (bud; Miller *et al.*, 1998; Miller and Rose, 1998; Yin *et al.*, 2000; Huisman *et al.*, 2004; Markus *et al.*, 2012). Model simulations also predicted that at least eight cMTs are required to provide the necessary force to migrate the entire nucleus to the daughter cell in basidiomycetes. This was in agreement with our experimental observations that show each *C. neoformans* cell nucleates an average of approximately nine cMTs. Thus a greater number of cMTs in basidiomycetes (~9 cMTs/cell) as compared with ascomycetes (~4 cMTs/cell) provides a larger pulling force on the SPB toward the emerging daughter cell, resulting in a deeper penetration of the SPB in basidiomycetes. However, our model also predicts a redundant pathway in which an increased activity (population) of dynein motors present at the cortical region of the daughter cell could also provide sufficient force to pull the nucleus/SPB into the bud.

The role of various MTs was determined by varying MT dynamics both by model simulations and by performing specific experiments, while scoring for the dynamics of the nuclear mass and SPBs through the progression of the cell cycle. The model predicted that disruption of all or only ipMTs results in the formation of short spindles without any nuclear movement toward the daughter cell. Experiments with depolymerizing drugs followed by measurement of the spindle length, neck to spindle distance, and the spindle axis revealed cells are arrested at the short-spindle stage. These experimental measurements, being in strong agreement with the model predictions, provide additional lines of validation for the developed models.

As our model can be used to simulate major transitions in mitosis, biological events spanning a short timescale can be incorporated to understand their global effects on the entire process of mitosis. Although we focused solely on the role of MTs in this study, the model can also be used to address other contributing factors and their roles among these systems. For example, using this model, we aim to further analyze the role of motor proteins during mitosis and to define their roles more specifically. However, this model also has certain limitations, which include consideration of only mechanical forces, absence of NE dynamics, and lack of regulation by the mitotic checkpoint. DNA replication was considered as an instantaneous process, and MT dynamics was taken as constant throughout the cell cycle, further adding to the model constraints. However, this model lays the foundation for follow-up work that will help create a more refined and comprehensive model. It is important to mention here that the above-mentioned limitations/assumptions do not affect the quantitative conclusions presented in this study. The predictive nature and robustness also remained unaltered when model parameters were varied within a permissible window. Nevertheless, a parallel set of pathways based on novel assumptions may exist that could produce results similar to the ones presented here.

In the present study, we developed a model to cover a large fraction of the mitotic cell cycle that is the first of its kind to our knowledge. We could successfully characterize different mitotic events, including the nuclear migration, spindle orientation, and spindle-length dynamics in a quantitative manner utilizing a holistic approach across two major fungal phyla. This type of systems biology approach to develop a predictive computational model may aid in identifying targets across human pathogenic yeasts for developing antifungal drugs.

MATERIALS AND METHODS
Model development
In this section, we describe the model variables and governing equations in a simplified configuration to explain the mitotic mechanics in ascomycetes and basidiomycetes (Figure 1, A and B).

Construction of mother and daughter cell
We consider the mother cell as an ellipsoid of dimensions a_{cell}, b_{cell}, and c_{cell} along the x, y, and z axes, respectively, whereas the nucleus is considered to be a sphere of radius $r_{nucleus}$ placed at a random location within the mother cell at the beginning of the simulation. In ascomycetes, the SPB is embedded into the NE (Kahana *et al.*, 1995; Jaspersen and Winey, 2004), whereas in basidiomycetes, no active SPB is reported to exist in the early stages of mitosis. However, multiple MTOCs are found that wade along the nuclear surface in basidiomycetes (Straube *et al.*, 2005; Fink *et al.*, 2006; Yamaguchi *et al.*, 2009). It has been observed that these MTOCs converge to a single mass during mitosis, leading to

SPB activation (Straube *et al.*, 2003). Formation of the bud can occur anywhere on the cell surface. We considered that the final volume of the bud is ~80–90% of the mother-cell volume, as observed in our experiments. We chose the growth rate of the daughter cell in our model such that it reflects the experimentally observed scenario. Our model predicts that the growth rate of the bud does not play any significant role in spindle positioning (Supplemental Figure S3); however, cell size variation affects the spindle positioning (Supplemental Figure S4).

Modeling of cMT and cMT and cortex–based interaction

The SPB nucleates cMTs that interact with the cell cortex via dynein motors (Carminati and Stearns, 1997; Adames and Cooper, 2000; Ten Hoopen *et al.*, 2012). MTs are modeled as straight filaments elongating with velocity v_g and shrinking with velocity v_s. Stochastic switching of MTs from a growing state to a shortening state and then the shortening state to the growing state occur with catastrophe frequency f_c and rescue frequency f_r, respectively (Mitchison and Kirschner, 1984). One can successfully simulate the dynamics of an MT using these four parameters. We assume that the cell cortex acts as a static wall that resists the growth of a cMT. The growth velocity of a cMT within the cortical region decreases as $v_g = v_g^0 \exp(-K_{cor} l_{dyn}/f_s)$ (Dogterom and Yurke, 1997; Janson *et al.*, 2003), where v_g^0 is the unconstrained growth velocity of an MT, K_{cor} is the stiffness of the cortex, l_{dyn} is the length of penetration of the cMT tip within the cortex, and f_s is the stall force per MT. Dyneins engage with the cMTs, growing within the cortical region of width l_{cor}, to pull the SPB and the nucleus toward the cortex (Carminati and Stearns, 1997; Adames and Cooper, 2000; Lee *et al.*, 2000; Ten Hoopen *et al.*, 2012). We assumed that several proteins like Tem1, Bub2, Bfa1, Elm1, and Num1 assemble close to the cleavage apparatus at the neck during mitosis (Huisman *et al.*, 2004; Rodal *et al.*, 2005; Cuschieri *et al.*, 2006; Baumgartner and Tolic, 2014). These protein molecules stabilize the cMTs and allow cMTs to interact with the cleavage apparatus (Castillon *et al.*, 2003). To achieve stable cMTs, rescue frequency of the cMTs is adjusted as a function of the distance of the cMT tip from neck as $f_r(x) \propto \exp(-x/l)$, where x is the distance between the tip of the cMT and neck, and l is a constant that is of the size of the cleavage apparatus (~0.2 µm). The pulling force due to the dyneins is calculated using the following expression:

$$f_{dyn} = l_{dyn} \lambda_{dyn} f_{dyn}^s \tag{1}$$

Here l_{dyn} is the penetration length of cMT within the cortical region, λ_{dyn} is the number of dynein motors engaged per unit length of the cMT, and f_{dyn}^s is the magnitude of the force exerted by a single dynein motor. Summing f_{dyn} over all the cMTs, we can estimate the total pulling force F_{dyn} on the SPB/nucleus. The cMTs also exert a net pushing force F_{push} (~1 pN) when the tip hits the cell periphery. Pushing force arises due to the polymerization of the cMT tip in contact with the cell cortex. To apply a bias to the cMTs, we explored several schemes, such as modulation of the dynamical parameters and differential cortical interaction between the mother and the daughter cells, independently.

Modeling KT and KT–MT interaction

During mitosis, KTs remain clustered and linked to SPBs through kMTs. To avoid any overlap among the KTs, we include inter-KT repulsion in a simplistic manner. Whenever two interacting KT spheres, each of radius r_{kt}, penetrate each other, a repulsive force keeps them separate. For the sake of simplicity, the repulsive force is considered as a linear function of the extent of overlap. The total repulsive force on the ith KT can be estimated as

$$F_{repul}^i = \sum_{j=1, j \neq i}^{N_{KT}} C d^j \tag{2}$$

Here C is a constant, N_{KT} is the number of KTs present in the nucleus and d^j is the maximum overlap length between the ith and the jth KT.

The dynamics of the kMT plus end plays a crucial role in positioning the KT in yeast (Gardner *et al.*, 2008). We assume that kMTs remain attached to the KTs throughout mitosis (Westermann *et al.*, 2007). The kMTs interact with the inner KT through spring-like KT fibrils (McIntosh *et al.*, 2008). As the polymerizing kMT tip penetrates the KT, it applies a pushing force on the KT, namely, $F_{poly} = l_{pen} K_{fibrils}$, where l_{pen} is the length of penetration of the kMT tip within the KT, and $K_{fibrils}$ is the effective spring constant of the KT fibrils. A depolymerizing kMT pulls the KT with a force $F_{depoly} = l_{out} K_c$ while trying to detach from the KT. Here l_{out} is the separation between the kMT tip and the KT, and K_c is the force constant of the kMT–KT connecting springs (Wei *et al.*, 2007; Powers *et al.*, 2009). We calculate the total force acting between the SPB and KT as $F_{SPB-KT} = \Sigma(F_{poly} + F_{depoly})$, where the sum over is the number of MTs interacting with a single KT. To maintain a constant distance between the SPB and the KT, we incorporated the notion of the length-dependent catastrophe of the kMT (Varga *et al.*, 2006; Varga, Leduc, *et al.*, 2009; Foethke *et al.*, 2009; Sau *et al.*, 2014), that is, catastrophe frequency of a kMT increases with its length l_{kMT} as $f_c = h l_{kMT}$.

Modeling ipMTs

After the duplication of the SPB, overlapping ipMTs facilitate mechanical interaction between the SPBs. Due to the presence of kinesin 5 motors along the overlap (Kapoor and Mitchison, 2001; Marco, Dorn, *et al.*, 2013), ipMTs tend to slide apart, essentially pushing the SPBs away from each other. The pushing force F_{ipMT} reads as

$$F_{ipMT} = l_{overlap} \lambda_{ipMT} f_{ipMT} \tag{3}$$

where $l_{overlap}$ is the total overlap length among all the ipMTs nucleated from the two SPBs, λ_{ipMT} is the linear density of kinesin motors engaged along the ipMTs, and f_{ipMT} is the force produced by a single ipMT motor.

During SPB duplication, KTs detach from the kMTs and reattach right after successful duplication. We assume all the KTs are captured instantly by the MTs nucleated from the mother and the daughter SPB, such that chromosomes are bioriented. The actual capture process and achievement of the biorientation, though, are far more complex and occur over a finite timescale (Marco, Dorn, *et al.*, 2013). In the present study, we ignore such details and focus on the spindle positioning and orientation during mitosis. It is noteworthy to mention that the KTs remain attached to the SPB throughout the cell cycle (except for 2–3 min during chromosome replication), and the average nuclear migration time is long (~1 h). Thus it is safe to assume the KT capturing process is "instantaneous." After chromosomal duplication, sister KTs remain attached to each other by cohesion springs. The cohesion springs, when stretched, generate tension between the sister KTs:

$$F_{cohesion} = K_{cohesion} x_{KT} \tag{4}$$

where $K_{cohesion}$ is the spring constant of the cohesion springs and x_{KT} is the separation between the sister KTs.

KTs always remain clustered in ascomycetes, whereas in basidiomycetes they are unclustered during interphase and each of them remains close to the NE in the beginning of mitosis. The KT clustering process ahead of mitosis was shown to be mediated by MTs (Kozubowski, Yadav, *et al.*, 2013). Further, the clustering of the MTOCs and KTs occurs at the same time before mitosis (V.Y. and K.S., unpublished data). Hence, in this model, we assume a direct interaction between the MTOCs on the outer surface of the NE with the adjacent KTs present on the inner surface. All MTOCs in *C. neoformans* nucleate MTs in random directions and interact with each other via these MTs. If, by chance, a searching MT from one MTOC captures another MTOC, they migrate along the connecting MT toward each other and coalesce to form a unified MTOC, conserving the total volume. The number of MTs nucleated from the merged MTOC is proportional to its surface area. The "search and capture" of MTOCs continues until all MTOCs merge together to form a single SPB. Because MTs can bend, the search extends along the nuclear periphery. In this way, two MTOCs situated on the diametrically opposite ends of the nucleus can interact with each other. Because self-assembly of MTOCs is an MT-driven phenomenon, the efficiency of this process depends on the selection of the dynamical parameters that determine MT life cycles. For instance, a very small catastrophe frequency leads to long MTs, which are efficient in capturing distant MTOCs; however, misdirected MTs waste valuable search time while completing their life cycles. Similarly, a very large catastrophe frequency leading to short MTs is also inefficient in locating distant MTOCs (Supplemental Figure S2). Thus MT dynamics are tuned and optimized to assemble MTOCs within experimental time frames.

The equation of motion for KT, SPB, and nucleus can now be written as

$$\frac{dx_{KT}^i}{dt} = \frac{F_{SPB-KT}^i + F_{cohesion}^i + F_{repul}^i}{\xi_{KT}^i} \tag{5}$$

$$\frac{dx_{SPB}}{dt} = \frac{\sum_{i=1}^{N_{KT}} F_{SPB-KT}^i + F_{dyn} + F_{push} + F_{ipMT}}{\xi_{SPB}} \tag{6}$$

$$\frac{dx_{Nu}}{dt} = \frac{F_{push} + F_{dyn}}{\xi_{Nu}} \tag{7}$$

Here the system of equations (Eqs. 5, 6, and 7) is derived in accordance with the well-known Stokes law, $v = F/\xi$; v, F, and ξ being the velocity, force, and viscous drag of a moving particle, respectively. The viscous drag obeys the formula $\xi = 6\pi\eta r$, where η is the coefficient of viscosity of the medium and r is the effective radius of the particle. Here ξ_{KT}, ξ_{SPB}, and ξ_{Nu} correspond to the effective drag on a KT, SPB, and nucleus, respectively. In our model, the medium in which the KTs, SPB, and the nucleus move are the nucleoplasm, NE, and cytoplasm of the cell, respectively. The superscript i in Eq. 1 stands for the ith KT and all the "x"s in Eqs. 5, 6, and 7 are the instantaneous coordinates of the objects considered here. After SPB duplication, another set of similar equations of motion for sister KTs and daughter SPB are incorporated. The constrained motion of the SPBs along the NE is achieved using a tangential coordinate system. The constraint is relaxed once the SPBs reach the diametrically opposite ends of the nucleus. At each time step, all the forces are calculated using Eqs. 1, 2, 3, and 4, and then Eqs. 5, 6, and 7 are solved numerically to update the positions of the KTs, SPBs, and nucleus, respectively. We explore a range of values for the model parameters (Table 1) to evaluate the model predictions.

Construction of fluorescently tagged strains

The MTOC markers, Tub4 in *C. albicans* and Spc98 in *C. neoformans*, were C-terminally tagged with green fluorescent protein (GFP) to visualize the dynamics of MTOC/SPB in live cells. In *C. albicans*, the 3′ part of the *TUB4* gene (Orf 19.1238) was amplified from the genome and cloned in a plasmid carrying the GFP-URA3, as a *SacII-SpeI* fragment. The plasmid was digested using *PacI* and transformed into *C. albicans*, SN148, to generate the Tub4-GFP–expressing strain. The resulting strain was used to visualize the MTOC with GFP and nuclear mass by 4′,6-diamidino-2-phenylindole (DAPI) staining during the imaging. For *C. neoformans*, histone H4 (ORF number CNAG_01648) was tagged with mCherry, and Spc98 (ORF number CNAG_01566) was tagged with GFP, using the overlap PCR strategy described earlier (Kozubowski, Yadav, *et al.*, 2013). For this purpose, 1 kb each of the 3′ part of the gene and 3′ untranslated region was amplified from the genome. A GFP-NAT (nourseothricin) or mCherry-NEO (neomycin) fragment (~3 kb) was amplified from pCN19 or pLK25 plasmids, respectively, and all three fragments were fused by overlap PCR, generating the cassettes. First, the Spc98-GFP-NAT cassette was transformed to get the Spc98-GFP strain, which was then transformed with H4-mCherry cassette to obtain a double-tagged strain. Similarly, the GFP-tubulin strain (Kozubowski, Yadav, *et al.*, 2013) was transformed with H4-mCherry cassette to obtain the strain. The list of strains used is given in Supplemental Table S1.

Microscopy and estimation of cMT number

The dynamics of fluorescently tagged proteins within cells across various cell cycle stages were captured using a Carl Zeiss confocal laser-scanning microscope LSM 510 META (Carl Zeiss, Jena, Germany). The images were then processed using the LSM 5 Image Examiner software (Carl Zeiss) and/or Adobe Photoshop (Adobe Systems, San Jose, CA).

The number of cMTs per cell was determined by the method described previously (Kosco *et al.*, 2001; Straube *et al.*, 2003). Briefly, the GFP-tagged Tub1 strain of *C. albicans* (YJB12856) or *C. neoformans* (CNVY109) was grown till log phase, harvested, and mounted on an 2% agarose pad containing synthetic complete media (2% dextrose, 0.67% YNB w/o amino acids, 0.2% amino acid mix, and 100 mg/l of uridine or uracil for *C. albicans* or *C. neoformans*, respectively). GFP-tagged tubulin images of *C. albicans* and *C. neoformans* cells were captured with identical settings using a Carl Zeiss LSM 510 META confocal microscope. Images were three-dimensionally rendered using ImageJ (National Institutes of Health, Bethesda, MD). The cMTs were tracked manually using three-dimensionally rendered images across all planes. Bright clustered signals of Tub1-GFP, which represented MTOCs, were excluded from counting. Subsequent processing was performed using ImageJ and Adobe Photoshop. Cell number versus cMTs/cell was plotted using Prism (GraphPad Software, La Jolla, CA), with the calculated mean drawn for both *C. albicans* and *C. neoformans*.

MT depolymerization experiments

We performed MT depolymerization experiments using nocodazole (Sigma-Aldrich, St. Louis, MO) and MBC (Sigma-Aldrich), MT-depolymerizing drugs, to disrupt either MTs or ipMTs, respectively. Both *C. albicans* (LSK111) and *C. neoformans* (CNVY197) GFP-tagged MTOC strains were grown overnight. The overnight culture was transferred to fresh media with an initial OD_{600} of 0.2. The culture was grown for

3 h to get the cells in log phase (OD$_{600}$ = 0.5–0.6). The cells were then treated with nocodazole (100 ng/ml for *C. neoformans* and 20 μg/ml for *C. albicans*) or MBC (1 μg/ml for *C. neoformans*) for 4 h. An aliquot of cells were treated with only dimethyl sulfoxide as a control. The cells were harvested after 4 h and washed with 1 ml of distilled water. Finally, cells were suspended in water, and images were captured using a microscope (DeltaVision, GE Healthcare Life Sciences) equipped with a CoolSnap HQ2 CCD. The images were then processed using ImageJ and Adobe Photoshop.

REFERENCES

Boldface names denote co–first authors.

Adames NR, Cooper JA (2000). Microtubule interactions with the cell cortex causing nuclear movements in *Saccharomyces cerevisiae*. J Cell Biol 149, 863–874.

Akera T, Sato M, Yamamoto M (2012). Interpolar microtubules are dispensable in fission yeast meiosis II. Nat Commun 3, 695.

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, 4131–4139.

Balasubramanian MK, McCollum D, Surana U (2000). Tying the knot: linking cytokinesis to the nuclear cycle. J Cell Sci 113, 1503–1513.

Baumgartner S, Tolic IM (2014). Astral microtubule pivoting promotes their search for cortical anchor sites during mitosis in budding yeast. PloS One 9, e93781.

Beach DL, Thibodeaux J, Maddox P, Yeh E, Bloom K (2000). The role of the proteins Kar9 and Myo2 in orienting the mitotic spindle of budding yeast. Curr Biol 10, 1497–1506.

Best HA, Matthews JH, Heathcott RW, Hanna R, Leahy DC, Coorey NV, Bellows DS, Atkinson PH, Miller JH (2013). Laulimalide and peloruside A inhibit mitosis of *Saccharomyces cerevisiae* by preventing microtubule depolymerisation-dependent steps in chromosome separation and nuclear positioning. Mol Biosyst 9, 2842–2852.

Carminati JL, Stearns T (1997). Microtubules orient the mitotic spindle in yeast through dynein-dependent interactions with the cell cortex. J Cell Biol 138, 629–641.

Castillon GA, Adames NR, Rosello CH, Seidel HS, Longtine MS, Cooper JA, Heil-Chapdelaine RA (2003). Septins have a dual role in controlling mitotic exit in budding yeast. Curr Biol 13, 654–658.

Cepeda-Garcia C, Delgehyr N, Juanes Ortiz MA, ten Hoopen R, Zhiteneva A, Segal M (2010). Actin-mediated delivery of astral microtubules instructs Kar9p asymmetric loading to the bud-ward spindle pole. Mol Biol Cell 21, 2685–2695.

Civelekoglu-Scholey G, Sharp DJ, Mogilner A, Scholey JM (2006). Model of chromosome motility in *Drosophila* embryos: adaptation of a general mechanism for rapid mitosis. Biophys J 90, 3966–3982.

Cuschieri L, Miller R, Vogel J (2006). Gamma-tubulin is required for proper recruitment and assembly of Kar9-Bim1 complexes in budding yeast. Mol Biol Cell 17, 4420–4434.

Dogterom M, Yurke B (1997). Measurement of the force-velocity relation for growing microtubules. Science 278, 856–860.

Fink G, Schuchardt I, Colombelli J, Stelzer E, Steinberg G (2006). Dynein-mediated pulling forces drive rapid mitotic spindle elongation in *Ustilago maydis*. EMBO J 25, 4897–4908.

Finley KR, Berman J (2005). Microtubules in *Candida albicans* hyphae drive nuclear dynamics and connect cell cycle progression to morphogenesis. Eukaryot Cell 4, 1697–1711.

Finley KR, Bouchonville KJ, Quick A, Berman J (2008). Dynein-dependent nuclear dynamics affect morphogenesis in *Candida albicans* by means of the Bub2p spindle checkpoint. J Cell Sci 121, 466–476.

Foethke D, Makushok T, Brunner D, Nedelec F (2009). Force- and length-dependent catastrophe activities explain interphase microtubule organization in fission yeast. Mol Systems Biol 5, 241.

Fraschini R, Venturetti M, Chiroli E, Piatti S (2008). The spindle position checkpoint: how to deal with spindle misalignment during asymmetric cell division in budding yeast. Biochem Soc Trans 36, 416–420.

Gandhi SR, Gierlinski M, Mino A, Tanaka K, Kitamura E, Clayton L, Tanaka TU (2011). Kinetochore-dependent microtubule rescue ensures their efficient and sustained interactions in early mitosis. Dev Cell 21, 920–933.

Gardner MK, Haase J, Mythreye K, Molk JN, Anderson M, Joglekar AP, O'Toole ET, Winey M, Salmon ED, Odde DJ, Bloom K (2008). The microtubule-based motor Kar3 and plus end-binding protein Bim1 provide structural support for the anaphase spindle. J Cell Biol 180, 91–100.

Gladfelter A, Berman J (2009). Dancing genomes: fungal nuclear positioning. Nat Rev Microbiol 7, 875–886.

Guertin DA, Trautmann S, McCollum D (2002). Cytokinesis in eukaryotes. Microbiol Mol Biol Rev 66, 155–178.

Haase J, Mishra PK, Stephens A, Haggerty R, Quammen C, Taylor RM II, Yeh E, Basrai MA, Bloom K (2013). A 3D map of the yeast kinetochore reveals the presence of core and accessory centromere-specific histone. Curr Biol 23, 1939–1944.

Heath IB (1980). Variant mitoses in lower eukaryotes: indicators of the evolution of mitosis. Int Rev Cytol 64, 1–80.

Horvitz HR, Herskowitz I (1992). Mechanisms of asymmetric cell division: two Bs or not two Bs, that is the question. Cell 68, 237–255.

Huisman SM, Bales OA, Bertrand M, Smeets MF, Reed SI, Segal M (2004). Differential contribution of Bud6p and Kar9p to microtubule capture and spindle orientation in *S. cerevisiae*. J Cell Biol 167, 231–244.

Hwang E, Kusch J, Barral Y, Huffaker TC (2003). Spindle orientation in *Saccharomyces cerevisiae* depends on the transport of microtubule ends along polarized actin cables. J Cell Biol 161, 483–488.

Janson ME, de Dood ME, Dogterom M (2003). Dynamic instability of microtubules is regulated by force. J Cell Biol 161, 1029–1034.

Jaspersen SL, Winey M (2004). The budding yeast spindle pole body: structure, duplication, and function. Annu Rev Cell Dev Biol 20, 1–28.

Jin QW, Fuchs J, Loidl J (2000). Centromere clustering is a major determinant of yeast interphase nuclear organization. J Cell Sci 113, 1903–1912.

Joglekar AP, Hunt AJ (2002). A simple, mechanistic model for directional instability during mitotic chromosome movements. Biophys J 83, 42–58.

Kahana JA, Schnapp BJ, Silver PA (1995). Kinetics of spindle pole body separation in budding yeast. Proc Natl Acad Sci USA 92, 9707–9711.

Kapoor TM, Mitchison TJ (2001). Eg5 is static in bipolar spindles relative to tubulin: evidence for a static spindle matrix. J Cell Biol 154, 1125–1133.

Knoblich JA (2008). Mechanisms of asymmetric stem cell division. Cell 132, 583–597.

Kosco KA, Pearson CG, Maddox PS, Wang PJ, Adams IR, Salmon ED, Bloom K, Huffaker TC (2001). Control of microtubule dynamics by Stu2p is essential for spindle orientation and metaphase chromosome alignment in yeast. Mol Biol Cell 12, 2870–2880.

Kozubowski L, Yadav V, Chatterjee G, Sridhar S, Yamaguchi M, Kawamoto S, Bose I, Heitman J, Sanyal K (2013). Ordered kinetochore assembly in the human-pathogenic basidiomycetous yeast *Cryptococcus neoformans*. mBio 4, e00614–00613.

Kusch J, Meyer A, Snyder MP, Barral Y (2002). Microtubule capture by the cleavage apparatus is required for proper spindle positioning in yeast. Genes Dev 16, 1627–1639.

Laan L, Pavin N, Husson J, Romet-Lemonne G, van Duijn M, Lopez MP, Vale RD, Julicher F, Reck-Peterson SL, Dogterom M (2012). Cortical dynein controls microtubule dynamics to generate pulling forces that position microtubule asters. Cell 148, 502–514.

Lee IJ, Wang N, Hu W, Schott K, Bahler J, Giddings TH Jr., Pringle JR, Du LL, Wu JQ (2014). Regulation of spindle pole body assembly and cytokinesis by the centrin-binding protein Sfi1 in fission yeast. Mol Biol Cell 25, 2735–2749.

Lee L, Tirnauer JS, Li J, Schuyler SC, Liu JY, Pellman D (2000). Positioning of the mitotic spindle by a cortical-microtubule capture mechanism. Science 287, 2260–2262.

Liakopoulos D, Kusch J, Grava S, Vogel J, Barral Y (2003). Asymmetric loading of Kar9 onto spindle poles and microtubules ensures proper spindle alignment. Cell 112, 561–574.

Mallik R, Gross SP (2004). Molecular motors: strategies to get along. Curr Biol 14, R971–R982.

Marco E, Dorn JF, Hsu PH, Jaqaman K, Sorger PK, Danuser G (2013). *S. cerevisiae* chromosomes biorient via gradual resolution of syntely between S phase and anaphase. Cell 154, 1127–1139.

Markus SM, Kalutkiewicz KA, Lee W-L (2012). Astral microtubule asymmetry provides directional cues for spindle positioning in budding yeast. Exp Cell Res 318, 1400–1406.

Markus SM, Lee WL (2011). Microtubule-dependent path to the cell cortex for cytoplasmic dynein in mitotic spindle orientation. Bioarchitecture 1, 209–215.

Martin R, Walther A, Wendland J (2004). Deletion of the dynein heavy-chain gene *DYN1* leads to aberrant nuclear positioning and defective hyphal development in *Candida albicans.* Eukaryot Cell 3, 1574–1588.

McIntosh JR, Grishchuk EL, Morphew MK, Efremov AK, Zhudenkov K, Volkov VA, Cheeseman IM, Desai A, Mastronarde DN, Ataullakhanov FI (2008). Fibrils connect microtubule tips with kinetochores: a mechanism to couple tubulin dynamics to chromosome motion. Cell 135, 322–333.

Merlini L, Piatti S (2011). The mother-bud neck as a signaling platform for the coordination between spindle position and cytokinesis in budding yeast. Biol Chem 392, 805–812.

Miller RK, Heller KK, Frisen L, Wallack DL, Loayza D, Gammie AE, Rose MD (1998). The kinesin-related proteins, Kip2p and Kip3p, function differently in nuclear migration in yeast. Mol Biol Cell 9, 2051–2068.

Miller RK, Rose MD (1998). Kar9p is a novel cortical protein required for cytoplasmic microtubule orientation in yeast. J Cell Biol 140, 377–390.

Mitchison T, Kirschner M (1984). Dynamic instability of microtubule growth. Nature 312, 237–242.

Moore JK, Magidson V, Khodjakov A, Cooper JA (2009). The spindle position checkpoint requires positional feedback from cytoplasmic microtubules. Curr Biol 19, 2026–2030.

Muller MJ, Klumpp S, Lipowsky R (2008). Tug-of-war as a cooperative mechanism for bidirectional cargo transport by molecular motors. Proc Natl Acad Sci USA 105, 4609–4614.

Neumuller RA, Knoblich JA (2009). Dividing cellular asymmetry: asymmetric cell division and its implications for stem cells and cancer. Genes Dev 23, 2675–2699.

Piatti S, Venturetti M, Chiroli E, Fraschini R (2006). The spindle position checkpoint in budding yeast: the motherly care of MEN. Cell Div 1, 2.

Powers AF, Franck AD, Gestaut DR, Cooper J, Gracyzk B, Wei RR, Wordeman L, Davis TN, Asbury CL (2009). The Ndc80 kinetochore complex forms load-bearing attachments to dynamic microtubule tips via biased diffusion. Cell 136, 865–875.

Pruyne D, Bretscher A (2000a). Polarization of cell growth in yeast. I. Establishment and maintenance of polarity states. J Cell Sci 113, 365–375.

Pruyne D, Bretscher A (2000b). Polarization of cell growth in yeast. II. The role of the cortical actin cytoskeleton. J Cell Sci 113, 571–585.

Rodal AA, Kozubowski L, Goode BL, Drubin DG, Hartwig JH (2005). Actin and septin ultrastructures at the budding yeast cell cortex. Mol Biol Cell 16, 372–384.

Sau S, Sutradhar S, Paul R, Sinha P (2014). Budding yeast kinetochore proteins, Chl4 and Ctf19, are required to maintain SPB-centromere proximity during G1 and late anaphase. PloS One 9, e101294.

Segal M, Bloom K (2001). Control of spindle polarity and orientation in *Saccharomyces cerevisiae*. Trends Cell Biol 11, 160–166.

Seybold C, Schiebel E (2013). Spindle pole bodies. Curr Biol 23, R858–R860.

Soppina V, Rai AK, Ramaiya AJ, Barak P, Mallik R (2009). Tug-of-war between dissimilar teams of microtubule motors regulates transport and fission of endosomes. Proc Natl Acad Sci USA 106, 19381–19386.

Straube A, Brill M, Oakley BR, Horio T, Steinberg G (2003). Microtubule organization requires cell cycle-dependent nucleation at dispersed cytoplasmic sites: polar and perinuclear microtubule organizing centers in the plant pathogen *Ustilago maydis.* Mol Biol Cell 14, 642–657.

Straube A, Weber I, Steinberg G (2005). A novel mechanism of nuclear envelope break-down in a fungus: nuclear migration strips off the envelope. EMBO J 24, 1674–1685.

Tanaka K, Kitamura E, Tanaka TU (2010). Live-cell analysis of kinetochore-microtubule interaction in budding yeast. Methods 51, 206–213.

Tanaka K, Tanaka TU (2009). Live cell imaging of kinetochore capture by microtubules in budding yeast. Methods Mol Biol 545, 233–242.

Tanaka TU, Stark MJ, Tanaka K (2005). Kinetochore capture and bi-orientation on the mitotic spindle. Nat Rev Mol Cell Biol 6, 929–942.

Ten Hoopen R, Cepeda-Garcia C, Fernandez-Arruti R, Juanes MA, Delgehyr N, Segal M (2012). Mechanism for astral microtubule capture by cortical Bud6p priming spindle polarity in *S. cerevisiae*. Curr Biol 22, 1075–1083.

Varga V, Helenius J, Tanaka K, Hyman AA, Tanaka TU, Howard J (2006). Yeast kinesin-8 depolymerizes microtubules in a length-dependent manner. Nat Cell Biol 8, 957–962.

Varga V, Leduc C, Bormuth V, Diez S, Howard J (2009). Kinesin-8 motors act cooperatively to mediate length-dependent microtubule depolymerization. Cell 138, 1174–1183.

Varoquaux N, Liachko I, Ay F, Burton JN, Shendure J, Dunham MJ, Vert JP, Noble WS (2015). Accurate identification of centromere locations in yeast genomes using Hi-C. Nucleic Acids Res 43, 5331–5339.

Wang H, Oliferenko S, Balasubramanian MK (2003). Cytokinesis: relative alignment of the cell division apparatus and the mitotic spindle. Curr Opin Cell Biol 15, 82–87.

Wei RR, Al-Bassam J, Harrison SC (2007). The Ndc80/HEC1 complex is a contact point for kinetochore-microtubule attachment. Nat Struct Mol Biol 14, 54–59.

Westermann S, Drubin DG, Barnes G (2007). Structures and functions of yeast kinetochore complexes. Annu Rev Biochem 76, 563–591.

Xiang X (2012). Nuclear positioning: dynein needed for microtubule shrinkage-coupled movement. Curr Biol 22, R496–R499.

Yamaguchi M, Biswas SK, Ohkusu M, Takeo K (2009). Dynamics of the spindle pole body of the pathogenic yeast *Cryptococcus neoformans* examined by freeze-substitution electron microscopy. FEMS Microbiol Lett 296, 257–265.

Yin H, Pruyne D, Huffaker TC, Bretscher A (2000). Myosin V orientates the mitotic spindle in yeast. Nature 406, 1013–1015.

9

Centromere and Kinetochore: Essential Components for Chromosome Segregation

Shreyas Sridhar, Arti Dumbrepatil, Lakshmi Sreekumar, Sundar Ram Sankaranarayanan, Krishnendu Guin, and Kaustuv Sanyal

Jawaharlal Nehru Centre for Advanced Scientific Research, Molecular Biology and Genetics Unit, Molecular Mycology Laboratory, Jakkur, Bangalore 560 064 India

9.1
Introduction

Perpetuation of life occurs by the fundamental property of cells to divide. A somatic cell undergoes a cell cycle that is comprised of essentially two periods: interphase and mitosis. Interphase can be further divided into G1, S, and G2. G1 and G2 constitute gap phases, involving cell growth that prepare cells for genome duplication in synthesis (S) phase and subsequent segregation in mitotic (M) phase, respectively. The mitotic cell cycle ensures equal division of the duplicated genetic content of the mother nucleus with the help of the kineto-chore and centromere. The kinetochore is a proteinaceous structure that assembles on centromere (*CEN*) DNA. The centromere/kinetochore generally appears as a constricted region of a metaphase chromosome (Figure 9.1a). The kineto-chore complex interacts with microtubules on one side and centromeric chro-matin on the other (Figure 9.1a). In most metazoans, multiple microtubules bind to each kinetochore, with an exception of certain budding yeasts where only a single microtubule appears to be associated with each kinetochore [1–4].

Apart from these general features of mitosis, organism-specific variations also exist. Mitosis is broadly classified in two types: closed mitosis and open mitosis (Figure 9.1b). This distinction primarily refers to the permeability of the nuclear envelope (NE), a bilayered membrane which along with the nuclear pore com-plexes (NPCs) regulate the entry and exit of molecules to and from the nucleus. Closed mitosis is considered to be the more primitive form of eukaryotic cell division, whereas open mitosis seems to have appeared several times during evo-lution. Plants and animals share open mitosis predominantly, while most fungi employ closed mitosis and variations of it. During closed mitosis, the NE

(a)

(b)
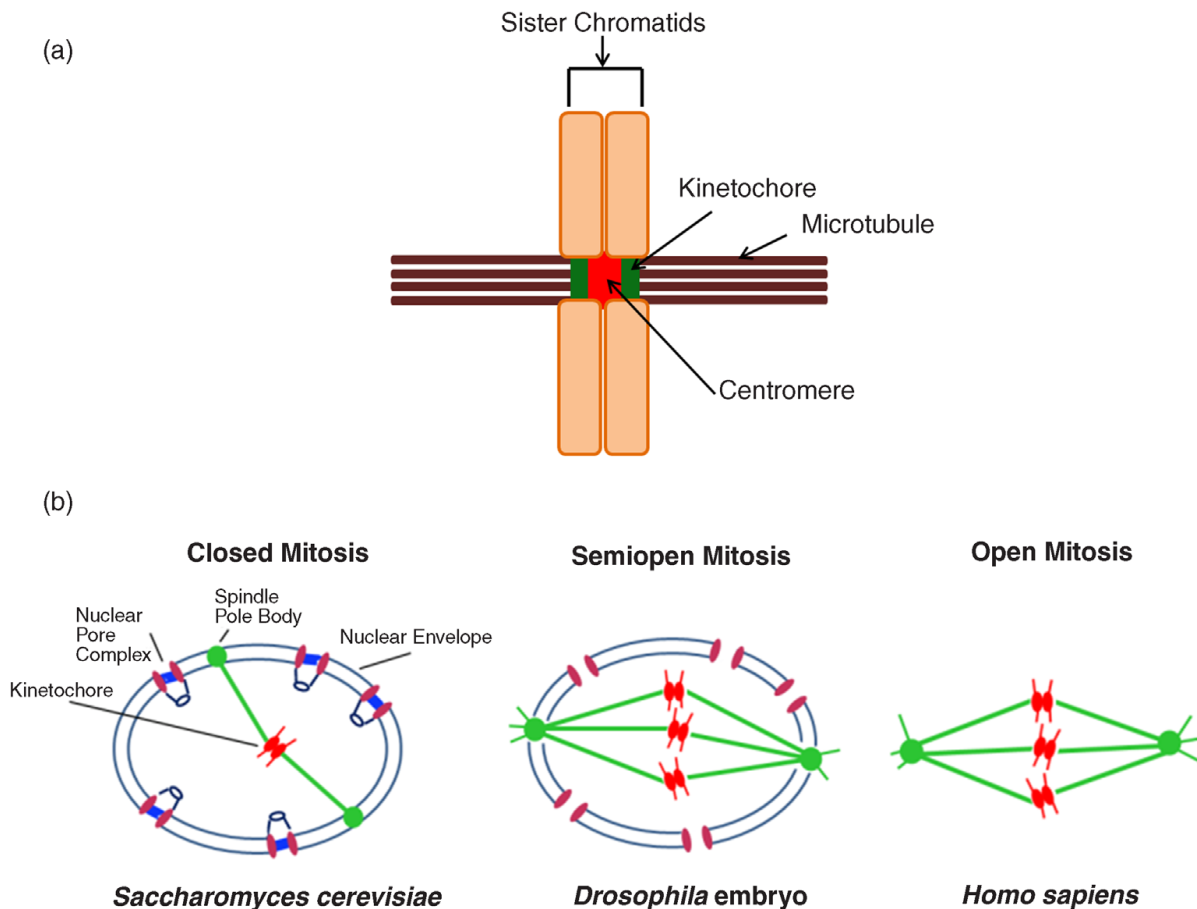
Figure 9.1 Modes of mitosis. (a) Schematic representation of a chromosome depicting microtubule attachment sites, the centromere/kinetochore complex. (b) Various forms of mitosis observed in the eukaryotic kingdom. Closed mitosis, in which the nuclear envelope never breaks down, is common in fungi (for example, *Saccharomyces cerevisiae*). The spindle pole bodies (SPBs) are embedded in the nuclear envelope. Chromosomes are attached to the spindle throughout cell cycle. Semiopen mitosis is exemplified by *Drosophila melanogaster* (early syncytial embryos), where disassembly of the nuclear pore complexes helps in partial opening of the nuclear envelope. Open mitosis, most commonly associated with metazoans (like humans), is characterized by the complete breakdown of the nuclear envelope during mitosis. The nuclear envelope reassembles after chromosome segregation.

remains intact throughout the cell cycle, the spindle forms within the nucleus followed by chromosomal segregation and subsequent nuclear fission. In contrast, the microtubule organizing centers (MTOCs) of many organisms are cytoplasmic and in order to facilitate the interaction of microtubules with kinetochores, the NE breaks down in what is termed as open mitosis [5]. However, the timing, duration, and extent of the NE breakdown vary considerably in different organisms and arrays of variant mitotic modes exist between the extremes of closed and open mitosis (Figure 9.1b). In this chapter, we discuss players and the process of chromosome segregation via mitotic cell cycle in fungal, animal, and plant kingdoms. We also discuss the growing knowledge of the same in protozoa as well.

9.1.1
Distinguishing Features of Mitosis

9.1.1.1 Fungi

Fungal systems undergo mitosis via a number of mechanisms [6]. In *Ustilago maydis*, with the aid of dynein motors, the microtubules pull the tip of the nucleus along with chromatin into the daughter cell. During this process, the NE breaks down in the vicinity of the spindle pole body (SPB) [7]. A similar observation was made in another basidiomycete *Cryptococcus neoformans* [8]. In a fission yeast *Schizosaccharomyces japonicus*, the NE breaks close to the central nuclear axis during anaphase B to prevent spindle collapse [9,10]. The NE was found to remain intact in many other fungi, including the popular and well-studied model organisms *Saccharomyces cerevisiae* [11] and *Schizosaccharomyces pombe* [12,13]. During closed mitosis, the spindle remains localized within the nucleus throughout the cell cycle. With the NE being closed, the only way molecules (e.g., tubulin) can pass in and out of the nucleus is via the NPCs. In addition, to facilitate spindle formation within the nucleus MTOCs are embedded in the NE. An exception to this mechanism was observed in *S. pombe*, where the SPB is in the cytoplasm during interphase but inserts itself into the NE during mitotic entry [14]. While the NE remains intact in *Aspergillus nidulans*, it was observed that the nucleus is permeable during mitosis [15]. Through subsequent studies, it was found that the NPCs but not the nuclear membranes enable this permeability [16,17]. A similar mechanism has also been shown to occur in *Magnaporthe grisea* [18], *Ceratocystis fagacearum, Fusarium oxysporum* [19], *and Fusarium verticillioides.*

9.1.1.2 Animals

Mitosis in higher eukaryotes, as in animals, involves disassembly of the NE and nuclear lamin that is absent in fungi. This process occurs between prophase and prometaphase stages to initiate open mitosis [5,20–23]. Examples of intermediate types of mitosis are observed in many invertebrates. During *Caenorhabditis elegans* early embryogenesis, lamins, and the NE completely disassemble only in late anaphase, leaving the NE intact during most cell cycle stages [24]. In embryonic and neuroblast development of *Drosophila melanogaster*, chromatin is surrounded by the NE till metaphase. The NE breakdown starts toward metaphase and ends postanaphase, while the spindle apparatus is enclosed in a fenestrated membrane.

9.1.1.3 Plants

Although the fundamental steps in cell division remain the same between animal and plant cells, a few variations are seen in the latter. The process of assembling the mitotic spindle apparatus is distinct in plant cells as they are devoid of centrosomes. While studies show the NE to harbor the MTOCs in most plants, it has also been proposed that microtubules emanate from the chromosomes themselves at random directions, and subsequently microtubules are aligned to

polarity with the aid of end-directed motor proteins. This is evidenced by the fact that gamma tubulin, a component of the animal centrosome, has been localized to *Vicia faba* kinetochores [25,26].

9.1.1.4 Protozoa

Closed mitosis, as observed in certain budding yeasts, is the characteristic of many protozoans [27,28]. Chromatin remains uncondensed throughout mitosis in the malaria parasite *Plasmodium*. In many protozoa including *Plasmodium*, MTOCs (kinetic centers, centriolar plaques, or centrosomes) nucleate microtubules from within the nucleus during prophase [6]. Among related coccidian parasites, *Toxoplasma* and *Eimeria*, a cone-shaped extension of nuclear membrane, centrocone aids in mitotic spindle nucleation [29]. While metaphase involves kinetochore capture by bipolar microtubules, the existence of a metaphase plate is unclear.

As it is evident, the critical process of cell division is accomplished with high efficiency via a wide variety of means to achieve equal distribution of genetic material between the daughters. The centromere DNA and the associated proteins, which form the kinetochore, play a major role in this process. While the kinetochore architecture is conserved, centromere DNA sequences are possibly the most rapidly evolving regions in the genome.

9.2
Centromeres

First identified as the primary constriction of the chromosome, the centromere is known to be a gene poor, transcriptionally silent, and recombination deficient chromosomal locus, as reviewed in Ref. [30]. While these features do not ensure the presence of a functional centromere on a chromosome, the unifying feature of all centromeres lies in their ability to provide a platform for the formation of a protein network, the kinetochore, which links chromosomes to the spindle apparatus. The structure of identified centromeres in various organisms differs significantly and can be classified into various types, however, they all are known to contain specialized chromatin. The unique chromatin state of the centromere is brought about by various factors including histone marks, RNAi, and specific protein binding [31]. The one common feature that all centromeres share is the presence of CENP-A, the centromeric histone H3 variant. CENP-A was originally identified from the immune sera of patients with limited systemic sclerosis (CREST) syndrome [32]. CENP-A comprises of a unique amino terminal domain and a C-terminal histone fold domain that shares significant homology with histone H3, yet is the most diverged of all histone H3 variants. The CENP-A targeting domain (CATD), a part of the histone fold domain, enables CENP-A to target to the centromere.

This quest to understand the centromere at the molecular level began more than 30 years ago with the identification and isolation of a functional centromere

from the unicellular budding yeast *S. cerevisiae* [33]. The identified centromere was DNA sequence-dependent as a single point mutation could inactive centromere function completely and thus genetically determined. Moreover, a plasmid molecule carrying a replication signal (<u>a</u>utonomously <u>r</u>eplicating <u>s</u>equence, *ARS*) and a 125 bp centromere DNA replicate and segregate as a minichromosome during mitosis and meiosis in this yeast. Subsequently, availability of the genome sequences, bioinformatics tools, and more sensitive biochemical and molecular biology techniques allowed scientists to identify centromeres from a large number of organisms.

The functional identification of centromeres relied on many of its unique properties such as, binding of an evolutionary conserved kinetochore protein, a minimal chromosomal locus that confers stability to a naked piece of DNA through mitosis and meiosis, or the presence of a region defined by tetrad analysis [34]. While certain organisms are known to have a holocentric chromosome with diffused centromeres along the length, chromosomes in most other organisms are monocentric. Centromeres of monocentric chromosomes can be classified into three broad types: (a) "point" centromeres where centromere function is primarily defined by conserved DNA motifs that are contained within short stretches (<400 bp) of DNA (Figure 9.2a), (b) "long regional" centromeres are defined by large stretches of CENP-A containing DNA interspersed with H3 containing nucleosomes (>40 kb) that are predominantly composed of repetitive and heterochromatic DNA (Figure 9.2b), and (c) "short regional" centromeres that are intermediate between point and large regional centromeres, with respect to CENP-A containing chromatin (<40 kb) that are often devoid of repetitive DNA and lack conserved protein binding motifs. While point centromeres employ specific DNA binding proteins to initiate kinetochore formation [31], it is becoming increasingly evident that in long and short regional centromeres epigenetic factors predominantly contribute to centromere establishment and maintenance [31,35].

9.2.1
Diversity in Organization of DNA Elements Across Centromeres

9.2.1.1 Fungi

Characterization of 125 bp centromeres in the *Hemiascomycetes* budding yeast *S. cerevisiae* was the first studied centromere. This magic sequence contains three <u>c</u>entromeric <u>D</u>NA <u>e</u>lements (CDEs). CDEI and CDEIII are conserved elements that are 8 bp and 25 bp, respectively [36,37]. CDEI and CDEIII flank CDEII that acts as a spacer (Figure 9.2a). Although CDEII is not conserved in its sequence, a high AT content (>86%) and an optimum length of CDEII are essential for centromere function. While deletion of CDEI, variations in AT-richness, and changes in the length of CDEII lead to only marginal segregation defects, deletion or mutations in central CCG element of CDEIII is not tolerated [38–40].

The point centromere of *S. cerevisiae* seems to have been inherited in only a few closely related species. Putative CDEI–CDEII–CDEIII elements are present

(a)

Budding yeast

(b)

Fission yeast

(c)

Human

(d)

Kinetoplastid
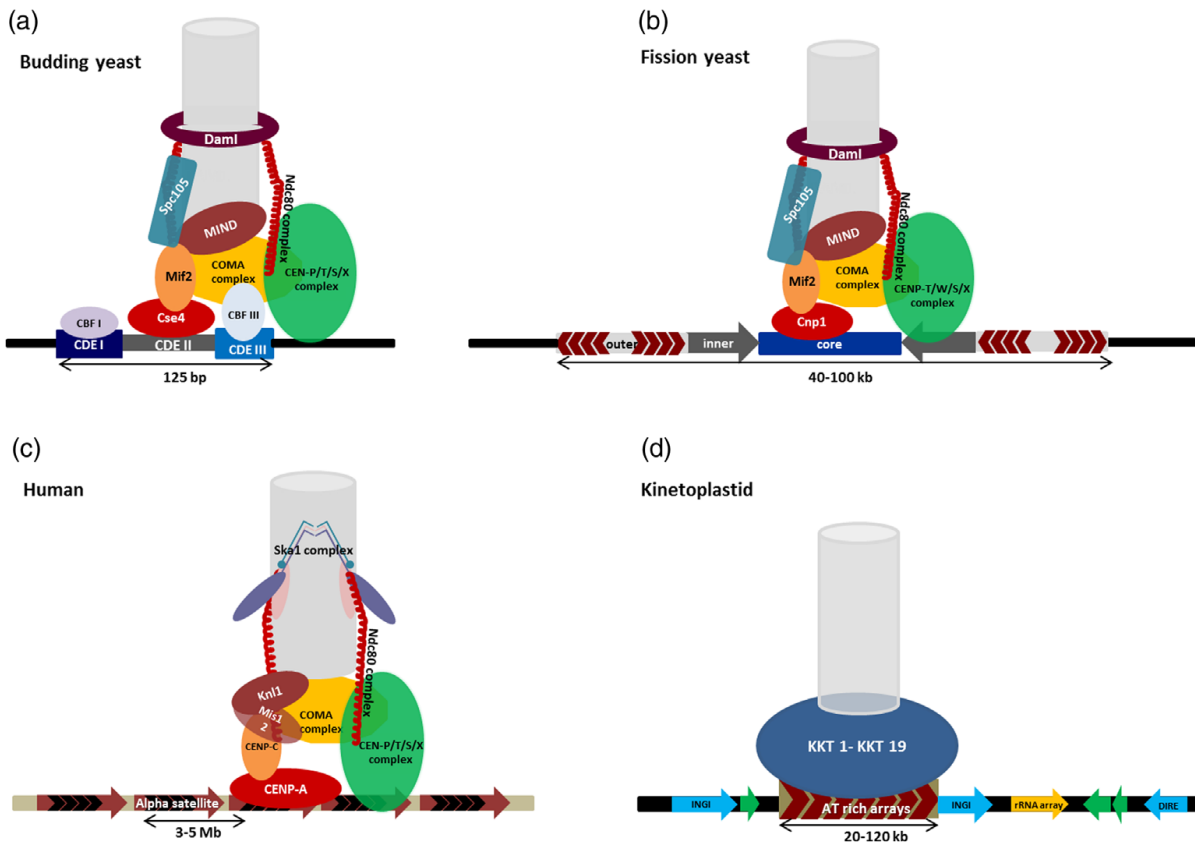
Figure 9.2 Schematic of centromeres and kinetochores across species. (a) In point centromere of the budding yeast, *S. cerevisiae*, the kinetochore complexes CBF1 and CBF3 bind CDEI and CDEIII elements in a sequence-specific manner and the Cse4 nucleosome wraps around the CDEII element. In addition, the inner kinetochore contains the COMA complex, Mif2 and Cnn1-contaning complex, which are homologs to the CCAN. This inner kinetochore network provides the platform for primary microtubule interacting complexes: the fungal specific Dam1 complex and the Ndc80 complex. (b) The fission yeast centromere contains a central core flanked by two sets of inverted repeats. While the kinetochore lacks CBF3 complex, only present in certain budding yeasts, other components including the CCAN and Dam1 complex are conserved. (c) Human centromeres consist of long repetitive DNA, called alpha satellite DNA. Apart from CENP-A, CENP-C and the CENP-T containing complex interact with the centromeric chromatin. Conserved CCAN proteins are present throughout the cell cycle. This network forms the base for the KMN network to form. While the Dam1 complex is absent, a functional homolog in Ska1 has been identified. (d) Unconventional kinetochores have been found to be associated with kinetoplastid centromeres that contain degenerate retrotransposons. Conserved kinetochore complexes/proteins have been depicted by similar topology and color. Microtubules are represented as gray cylinders.

in the chromosomes of *S. paradoxus, S. bayanus,* and *S. mikatae* as determined by sequence and synteny analyses [41,42]. Interestingly, in spite of a similar structure and function of centromeres, they can be species-specific as observed when comparing the centromere of *Candida glabrata* with *S. cerevisiae* by cross-species functional complementation. All *C. glabrata* centromeres contain CDEI–CDEII–CDEIII like elements, where CDEI and CDEIII were found to be required

for accurate chromosome segregation [43]. However, the *C. glabrata* centromeres could not function in *S. cerevisiae*. This species specificity of centromeres is further highlighted in the point centromeres of *Kluyveromyces lactis* and *Ashbya gossypii* in which the CDEII length is nearly twice as long (∼160 bp) as that of *S. cerevisiae* (78–86 bp). In addition, a 100 bp AT-rich element (CDE0) was found upstream of CDEI [42,44,45].

The ability of a centromere DNA to stabilize the mitotic inheritance of an *ARS* plasmid is a test of its function. In *Candida maltosa*, a 325 bp region was found to be sufficient to provide *ARS* plasmid stability [46]. Examination of this region revealed a conserved CDEI element followed by an AT-rich CDEII element but lacked any consensus CDEIII element. The centromeres of *Yarrowia lipolytica* lacked any consensus to CDEI or CDEIII elements although the AT-richness of the ∼200 bp centromere was similar to the CDEII element in *S. cerevisiae* [47].

Centromeres containing no obvious elements of DNA binding motifs came into light with the identification of the fission yeast centromeres. All the three centromeres in *S. pombe* contain a <u>c</u>entral <u>c</u>ore (cc) of 4–7 kb of unique non-homologous DNA, flanked by two sets of inverted repeats on either side [48–50]. The <u>innerm</u>ost <u>r</u>epeats (imr), closest to the cc, is composed of unique sequences on each centromere. <u>Out</u>e<u>r</u> (otr)most repeats are composed of two "dg" and "dh" repeats. Pericentric heterochromatin spans ∼10–60 kb on either side of the 10–15 kb of CENP-A chromatin (Figure 9.2b) [51–53]. Deletion of cc was found to abolish centromere function, while cc and dg repeats were found to be critical for maintaining an active centromere. In related fission yeast species, repeats were found to be present at all centromeres but appeared to be evolving fast. *S. pombe* and *Schizosaccharomyces octosporus* contain no transposable elements while *S. japonicus* contains clusters of inverted repeats in its centromeres [54].

Apart from repeat-rich long centromeres and DNA sequence motif-dependent short centromeres, the centromeres unearthed in *Candida albicans* and its close relative *Candida dubliniensis* neither contained repeats nor any detectable DNA motifs. The centromeres identified as Cse4 (CENP-A homolog in yeasts) binding sites, were found to be 3–5 kb in length, that are contained in 4–18 kb *ORF*-free regions [55,56]. All centromeres in these *Candida* species were found to contain unique DNA sequences, with the exception of *CEN5* that is made up of long inverted pericentric repeats. Surprisingly, the identified centromeres were not able to confer mitotic stability to an externally introduced naked *ARS* plasmid – hinting toward the possibility that centromeres may be epigenetically defined [57]. This was further supported by the formation of "neocentromeres" on unrelated sequences when a native centromere was replaced by a marker element [58–60]. Interestingly, the centromere DNA sequence of the orthologous chromosomes in *C. albicans* and *C. dubliniensis*, two closely related *Candida* species, was found to be the most rapidly changing locus of their genomes. Bioinfomatic analysis of *C. lusitania*, *Pichia stipites,* and *Debaryomyces hansenii* genomes have identified unique GC-poor regions as centromeres on each chromosome [61].

The study of ascomyceteous filamentous fungi *Neurospora crassa* identified centromeres that were found to be contained in a ~300 kb AT-region of DNA [62,63]. Centromeres were found to be repeat-rich and highly divergent. A mechanism of repeat induced point mutation (RIP) was found to be the major mediator of centromeric diversity that is operated via an unknown mechanism to convert C to A [62,64]. It was also observed that the centromeres contain retrotransposon like elements (Tcen, TglI, and Tgl2) in addition to LINE retro-transposons. The process of RIP was also suggested to occur in another filamentous fungi *A. nidulans*, whose centromeres are predicted to contain two degenerate LTR retrotransposons Dane1 and Dane2 [65,66].

The centromeres of the basidiomycotus fungi are relatively less studied. In *C. neoformans*, a pathogenic basidiomycete, the largest intergenic regions (20–65 kb) in each of the 14 chromosomes contained the centromere [67]. These centromeres are enriched with transposable elements (Tcn1–Tcn6) or their remnants. These findings were compared and validated in two serotypes A and D [67,68].

9.2.1.2 Animals

With the increase in genome complexity from yeast to metazoans, the factors that determine centromere function, whether genetic or epigenetic, become hazier. Most metazoan, including vertebrates have repeat-associated centromeres. Repeats adopt a heterochromatic state with distinct components and features. The centromere identifier, CID/CENP-A binding region extends to over 200–500 kb and 500–1500 kb in fruit flies and humans, respectively [69]. Using three-dimensional deconvolution microscopy analyses, it has been shown that CID/CENP-A in both flies and humans exists as a cylinder-like structure extending through the depth of the metaphase chromosome [69]. Immunofluorescence studies of CID and H3 on chromatin have revealed that a significant amount of H3 is present in centromeric fibers. Interestingly, CENP-A occupies only one-half to two-thirds of the centromeric chromatin fiber. This similarity in the organization of CENP-A/CID chromatin in flies and humans suggests that interspersed blocks of H3 and CENP-A are evolutionarily conserved aspects of centromere structure [69].

In certain taxa, kinetochores form along the length of the chromosome. This "holocentricity" or presence of a diffused kinetochore has been seen in two phyla of the animal kingdom: nematoda and arthropoda. The most widely studied among them is the nematode *C. elegans*. Genome analysis shows that *C. elegans* lack tandem repeats and show CENP-A binding sites over a wide array of unique sequences. Single base pair resolution studies have shown that this "polycentric" set up consists of hundreds of budding yeast-like point centromeric sites on each chromosome [70,71].

Among chicken chromosomes, a few centromeres have been shown to contain a 42 bp tandem repeat (CNM sequence) while most other chromosomes have long tandem repeats. However, chromosome 5 and chromosome Z lack tandem repetitive sequences. Instead, they contain a long-range repetitive region named MHM repeats, spanning around a 2 Mb region. Hence, all chicken centromeres

contain repeat units, the copy number of which is chromosome specific. This feature is quite similar in human centromeres [72].

Mammalian centromeres possess a complex structure containing long ordered repeats on which kinetochore proteins assemble. In humans, centromeres are present as long repetitive stretches called alpha satellite DNA. The alpha satellite DNA, which was identified first in the African Green monkey genome, comprises of 170 bp monomers arranged in a tandem head to tail fashion. Centromeric locus has been shown to have a higher order array of this repeat, spanning over 3–5 Mb. These stretches are frequently interspersed with LINE, SINE elements, and LTRs. In humans, at every other 171 bp monomer there is a CENP-B box. CENP-B has a dimerization domain so that it can bring two proteins together. It is a combination of two proteins, namely, CENP-A and CENP-B, at the alpha satellite DNA that brings about the higher order structure [34,69,73–75]. In short, primates have undergone a series of additions in their centromeric regions leading to the inclusion of new sequences (Figure 9.2c).

9.2.1.3 Plants

Much like the centromeres of higher eukaryotes, the composition of the plant centromeres identified till date is highly repeat associated, spanning megabases in length. Centromeres of *Brassicasea*, for instance, which includes *Arabidopsis* sp., have an array of 180 bp satellites ranging from 0.4–1.4 Mb as the primary centromere constituent [76]. This cen180 class of repeats includes other variations like the 169 bp repeats (also called pAge1 harboring a 9 bp deletion). In addition to these, other classes of repeats like a 500 bp repeat associated with all CENs and a 160 bp repeat located adjacent to the centromeres also exist [77]. It has also been shown that the cen180 repeats share homology with those found in cucumber and maize indicating a common ancestry [78]. In *Oryza* sp. of which rice is a part, centromeres consist of arrays of 155 bp repeat called the CentO repeat [79,80]. Similar in size is the CentC repeat (154 bp) that constitutes the maize centromeres [81]. With the presence of cereal centromere sequence 1 (CCS1) and Sau3A9 repeats, the centromeres of *Brachypodium* and sorghum bear no exception to the repeat associated trend [82,83].

An interesting feature of all plant centromeres studied till date is the presence of transposable elements as an integral structural component. A few families of retrotransposons have been found enriched or exclusive to the centromeres. This includes the centromeric retrotransposons (CR) family found to be associated with the centromeres in grasses [84]. Studies on cereals later showed that the CCS1 and Sau3A9 are indeed derivatives of the Ty3/Gypsy class of retrotransposons. The centromeres of rice and maize also have their share of CR variants in the form of CRR and CRM (CR of rice and maize, respectively) that are interspersed between the CentO and CentC repeats. Further these elements have been shown to be enriched in CENP-A chromatin immunoprecipitation (ChIP), indicating that they can be a functional part of centromeres. In *Arabidopsis*, the pericentromeric region contains copies of Athila transposon along with the blocks of cen180 repeats [77]. The centromeres of barley were also found to

contain a GC-rich satellite sequence along with retroelement cerebra, spanning up to 1.4 Mb in length [85]. The Ty1/copia-like retroelements have been associated with the centromeres of rye.

Apart from the aforementioned structural features, other important attributes of plant centromeres are mentioned further. There exists variation in the size of repeats between different centromeres of orthologous chromosomes. This indicates that rapid evolution of these sequences can occur in short time windows. This also gains support from the fact that there are variations in the sequences of the satellite repeats even among closely related species. For instance, short blocks of homology between the CentO and CentC satellites could possibly indicate their divergence from a common ancestor. There is a strong epigenetic component to these centromeres as it is evident from the occurrence of neocentromeres at nonnative loci. Even native centromeric sequences fail to assemble functional kinetochore machinery when reintroduced. This raises a possibility of an epigenetic component for centromere function in plants [86,87].

9.2.1.4 Protozoa

Identification of centromeres in protozoans posed several challenges including cloning of large AT-rich DNA segments, lack of chromatin condensation during mitosis, and limited genetic manipulation techniques forcing researchers to search for alternative methods. Etoposide-mediated topoisomerase II (Topo-II) cleavage was found to be a reliable means of identifying centromeres in these organisms [88]. Topo-II is a well-conserved protein that accumulates on centromeres at metaphase, playing a critical role in chromosome segregation [89]. Topo-II decatenates DNA by passaging an uncut strand of DNA through DNA breaks followed by ligation to repair the lesion. Etoposide, a Topo-II poison, blocks the re-ligation step resulting in DNA breaks. These cleavage sites are then mapped to specific chromosomal locations. Initially used to identify the *Plasmodium falciparum* centromeres [90], this technique has subsequently been extended to identify and validate several other protozoan centromeres.

The *P. falciparum* centromere was initially mapped to 6–12 kb gene-free loci that were GC poor (>3%) and located once per chromosome. Within these gene-free loci, a highly AT-rich region of 2.3–2.5 kb that contained a repetitive and core region was found, with the exception of chromosome 10. Subsequent sequence and genetic analysis revealed that both the core and repetitive region are essential for centromere function. Apart from a tightly restricted size range, no other conserved features were detected across chromosomes [90]. CENP-A ChIP analysis was performed to validate these results [91]. It was found that *Pf* CENP-A occupies a region of 4–4.5 kb encompassing a highly AT-rich core of 2–2.5 kb, which is devoid of any heterochromatin, on each chromosome. Similar AT-rich blocks were observed in contigs from chromosomes 4, 6, and 14 in *P. vivax* [90]. Subsequent synteny analysis in *P. berghi* using *Pf* centromeres lead to the finding of a highly AT-rich (96%) sequence spanning ~1.2 kb that contained a nonrepetitive core and a repetitive region [92]. This AT-rich segment of

DNA was sufficient for centromere establishment and faithful segregation of artificial chromosomes [92].

Centromere identification and characterization was extended to other apicomplexans. The centromeres of *Toxoplasma gondii* were identified on 12 of the 14 chromosomes by ChIP–chip analysis of CENP-A [93]. It revealed that CENP-A bound to regions of 16 ± 3.5 kb that lie largely in gene-free regions of 17 ± 5.6 kb. To validate these results, etoposide-mediated Topo-II cleavage was employed. In addition, centromeres were found to maintain close proximity to centrocones throughout the cell cycle [93]. The centromeres identified in apicomplexans yielded no sequence bias or conserved sequence elements similar to *C. albicans* and *C. dubliniensis. Trypanozoma cruzii* and *Trypanozoma brucii* are kinetoplastids, a group of protozoans, whose centromeres have been identified by telomere-associated centromere fragmentation [94,95]. A GC-rich transcriptional strand-switch domain of 16 kb in length, primarily constitutes the centromere in *T. cruzi*. This domain was predominantly composed of degenerate retrotransposons, which separated two large directional gene clusters that are transcribed toward the telomeres [94]. The identified chromosomal loci also coincide with etoposide-mediated Topo-II cleavage sites on chromosome 1. The 11 kb GC-rich domain is composed of degenerate retrotransposons (SIRE/VIPER). Etoposide-mediated Topo-II cleavage site mapping revealed the centromere location on chromosome 1 in *T. brucii* [96]. Although centromeres were initially indicated to include 2–8 kb array of ∼147 bp AT-rich tandem repeats, recent studies using long-range restriction mapping has fine mapped the size of the centromere on 8 chromosomes, revealing that the size of arrays varies from 20 to 120 kb [97]. In addition to varying sizes of centromeres, large heterogeneity between chromosome homologs was observed. This organization found in *T. brucii* is similar to centromeres of higher eukaryotes (Figure 9.2d).

9.3
Kinetochores

The kinetochore was first described as a specialized dish-shaped proteinaceous structure found at the periphery of the centromere DNA, as visualized from electron micrographs. The kinetochore is composed of more than 100 different proteins in vertebrate cells, each of which is present in multiple copies at each kinetochore. Based on the relative spatial localization of these proteins and their functions, they can be grouped into three main categories: (a) inner kinetochore proteins that are required to form the connection with the centromere DNA and provide a platform to assemble the kinetochore, (b) outer kinetochore proteins that form connections with microtubules, and (c) regulatory proteins that monitor or control the activities of the kinetochore. The first centromere-specific kinetochore proteins were discovered in clinical studies performed from patients with progressive systemic sclerosis (CREST syndrome) [32]. Advances in proteomics have enabled the identification of a large number of kinetochore

components [98–102]. However, functions of the many kinetochore components have been well worked out, little is known about how they are recruited to the centromere or how they assemble to form the complex kinetochore structure.

9.3.1
Kinetochore Architecture

From the centromere DNA on the chromosome to the dynamic plus end of the kinetochore microtubules, kinetochores measure ~75–100 nm, a region too small to resolve by conventional light microscopy. The current knowledge on the composition of this megastructure derives from the use of two elegant strategies to study protein complexes. While genetic screens for segregation defects and two hybrid screens to identify interacting proteins laid the platform, the development of high throughput techniques like affinity purification coupled with mass spectrometric analysis (LC-MS/MS) proved to be a powerful technique for identification of new kinetochore subunits in the later years. Although several limitations exist, careful analysis of results holds a promise for the possibility of systematically identifying all proteins within a purified complex [100]. For proteins that can be solubilized as components of native complexes, hydrodynamic analysis of the bait protein and those purifying with it can yield a clear picture of how coprecipitating components relate to each other. Most kinetochore complexes are highly elongated with frictional coefficients around 2.0 and axial ratios of 20:1. If such complexes are oriented vertically, they might contribute to bridging the distance between the centromere and microtubule tip(s). Each strata of protein complex, starting from the chromatin-associated layer is described further. These complexes together present an orchestra in harmony to carry out the primary functions of the kinetochore: (a) anchoring of the microtubules to the centromeres, (b) subsequent force and tension generation for the accurate segregation of the sister chromatids during cell division, and (c) activation of the spindle assembly checkpoint (SAC) to detect improper kinetochore attachments preventing cell cycle defects leading to aneuploidy.

9.3.2
Centromere DNA-Associated Layer

Proteins of this layer that form the inner kinetochore were the first to be identified and cloned starting with CENP-B that was shown to bind to a specific 17 bp sequence in alpha satellite DNA. Subsequent copurification of CENP-A isolated several other CENPs (M, N, T, H, I). Following studies led to identification of other factors that do not directly interact with CENP-A in humans [3,100,103] (Table 9.1).

9.3.2.1 Constitutive Centromere-Associated Network (CCAN)
The "inner kinetochore" consists of proteins that are most closely associated with centromeric chromatin. This included the isolation of CENP-A and other associated proteins in vertebrates, identified as the CCAN [103–106]. This

Table 9.1 Kinetochore subcomplexes across eukaryotic kingdoms.

		Fungi			Animalia			Plants	Protozoa	
		Ascomycota		Basidiomycota	Insecta	Nematoda	Chordata	Angiosperms	Apicoplexa	Euglenozoa
Complex name	Kinetochore protein	*Saccharomyces cerevisiae*	*Schizosaccharomyces pombe*	*Cryptococcus neoformans*	*Drosophila melanogaster*	*Caenorhabditis elegans*	*Homo sapiens*	*Arabidopsis thaliana*	*Plasmodium falciparum*	*Trypanosoma brucei*
CBF3	Ndc10	+	–	–	–	–	–	–	–	–
	Cep3	+	–	–	–	–	–	–	–	–
	Ctf13	+	–	–	–	–	–	–	–	–
	Skp1	+	+	+	+	+	+	+	+	–
CCAN	CENP-A	+	+	+	+	+	+	+	+	–
	CENP-B	–	+	•	•	•	+	•	•	–
	CENP-C	+	+	+	+	+	+	+	+	–
	CENP-F	•	•	•	•	+	+	+	•	–
	CENP-H	+	•	•	•	•	+	•	•	–
	CENP-I	+	+	•	•	•	+	•	•	–
	CENP-K	+	+	•	•	•	+	•	•	–
	CENP-L	+	+	•	•	•	+	•	•	–
	CENP-M	•	•	•	•	•	+	•	•	–
	CENP-N	+	+	•	•	•	+	•	•	–
	CENP-O	+	+	•	•	•	+	•	•	–
	CENP-P	+	+	•	•	•	+	•	•	–
	CENP-Q	+	+	•	•	•	+	•	•	–
	CENP-R	•	•	•	•	•	+	•	•	–
	CENP-S	+	+	•	•	+	+	•	•	–
	CENP-T	+	+	•	•	•	+	•	•	–
	CENP-U	+	+	•	•	•	+	•	•	–
	CENP-W	+	•	•	•	•	+	•	•	–
	CENP-X	+	+	•	•	+	+	•	•	–
	Nkp1	+	+	•	•	•	•	•	•	–
	Nkp2	+	+	•	•	•	•	•	•	–
Mis12	Mis12	+	+	+	+	+	+	•	•	
	Dsn1	+	+	+	+	+	+	•	•	
	Nnf1	+	+	+	+	+	+	•	•	
	Nsl1	+	+	+	+	+	+	•	•	
Ndc80	Ndc80	+	+	+	+	+	+	•	+	–
	Nuf2	+	+	+	+	+	+	•	+	–
	Spc24	+	+	+	•	+	+	•	+	–
	Spc25	+	+	+	+	+	+	•	+	–
Spc105	KNL-1	+	+	+	+	+	+	•	•	–
	Zwint	+	–	•	•	•	+	•	•	–
	Sos7	–	+	•	•	•	•	•	•	–
Dam1	Ask1	+	+	+	–	–	–	–	•	–
	Dad1	+	+	+	–	–	–	–	–	–
	Dad2	+	+	+	–	–	–	–	–	–
	Dad3	+	+	+	–	–	–	–	–	–
	Dad4	+	+	+	–	–	–	–	–	–
	Dam1	+	+	+	–	–	–	–	–	–
	Duo1	+	+	+	–	–	–	–	–	–
	Spc19	+	+	+	–	–	–	–	–	–
	Spc34	+	+	+	–	–	–	–	–	–
	Hsk1	+	+	+	–	–	–	–	–	–
Ska	Ska1	–	–	–	–	–	+	–	–	–
	Ska2	–	–	–	–	–	+	–	–	–
	Ska3	–	–	–	–	–	+	–	–	–

■ : Present
□ : Absent
▨ : Unidentified

network consists of subcomplexes that include CENP-C, CENP-H/I/K/M, CENP-L/N, CENP-O/P/Q/U, and the histone fold domain-containing proteins CENP-T/W and CENP-S/X [107–109]. Homologs of the CCAN proteins in the budding yeast, including the histone fold-containing proteins CENP-T/X/W/S have been identified. CENP-T is crucial for interacting and forming a scaffold for the Ndc80 complex, a major load bearing structure of the kinetochore. While CENP-T is essential for kinetochore function in humans, its nonessentiality in budding yeast might be due to the presence of a DNA sequence-dependent binding protein complexes, CBF3 [101,102]. The inner kinetochore of yeast contains orthologs of most human CCAN proteins. Mif2/CENP-C was one of the initially identified members of the CCAN and found to copurify with CENP-A nucleosomes.

9.3.3
Microtubule Interacting Layer

The critical kinetochore–microtubule interphase comprises of Mtw1/Mis12/MIND, Spc105/Knl1/Blinkin, Ndc80, Dam1/DASH, and a plethora of non-essential components such as motor proteins and checkpoint components [1,100,110,111] (Table 9.1).

9.3.3.1 KMN Network
The KMN network comprises of three major complexes: the KNL1/Spc107 complex, Mis12 complex, and the Ndc80 complex. This network of proteins form the major kinetochore microtubule (kMT) attachment sites and perform several key functions: modulates and anchors the dynamic kMTs, aids in force generation at the microtubule plus end, corrects errors in microtubule attachment, and plays role of a scaffold to recruit the spindle assembly checkpoint components. The KMN network also plays an important role of connecting the centromere binding proteins directly to the microtubules via the Ndc80 complex and KNL1 complex. The network links to the inner layer through attachments at both CENP-C and CENP-T through the Mis12 and Ndc80 complexes, respectively. These interactions orient the elongated Spc107 and Ndc80 along the axis at metaphase [112].

KNL1/Spc107 Complex. The KNL1 complex is composed of two proteins, Knl1 (Spc105 in yeast) and ZWINT (Ydc532, Kre28 in yeast). The stoichiometric ratio of Spc105: Kre28 has been estimated to be 1:2 [113–115]. This complex has been shown to play the critical role of a scaffold for the SAC recruitment onto the kinetochore, reviewed in Ref. [116].

Mis12 Complex MIS12 (Mtw1 or MIND in yeast), Nsl1/Dc31/Mis14, Nnf1/Pmf1, and Dsn1 constitute the four subunits of the Mis12 complex in a stoichiometric ratio of 1:1:1:1. Unlike the other two complexes of the KMN network, the Mis12 complex has not been implicated to bind microtubules directly, even though it is critical for the recruitment of the Ndc80 and KNL1/Spc107 complexes [111,117].

Ndc80 Complex: The Ndc80 complex is present as a 57 nm long heterotetrameric complex that is composed of the four subunits Ndc80/Hec1, Nuf2, Spc24, and Spc25 in a ratio of 1:1:1:1. The heterodimer is held together via α-helical coiled coil domains at the N-terminal of Spc24 and Spc25 and the C-terminal of Ndc80 and Nuf2 [118]. The coiled-coil domains at the C-terminal of Spc24–Spc25 have been shown to bind to CENP-T, Nsl1, and Dsn1 of the Mis12 complex in humans and fly. The N-terminal tail of both Ndc80 and Nuf2 contains a globular head with a positively charged Calponin homology (CH) domain that facilitates binding of this complex to the negatively charged microtubules [119,120].

Dam1 and Ska Complex. The Dam1 complex is a fungal exclusive kinetochore protein complex, whose components are essential in most known yeasts. It is a ~210 kDa heterodecamer whose components are Ask1, Dad1, Dad2, Dad3, Dad4, Dam1, Duo1, Hsk3, Spc19, and Spc34 [121–124]. The Ska complex in vertebrates has been recently implicated to be the functional counterpart of the fungal Dam1 complex. The Ska complex is essential for effective chromosome segregation. It consists of three subunits Ska1, Ska2, and Ska3 [125–127]. While the Dam1 complex forms rings around the microtubules, the Ska complex is known to form a wedge-like structure around the microtubules. These complexes do not share any sequence or structural homology but are known to play a vital role in the attachment of the kinetochore to the microtubules (Figure 9.2).

9.3.4
Kinetochore Assembly

9.3.4.1 Fungi

The fungal kinetochores are well studied and formed by more than 80 known proteins assembled on the centromere DNA [128]. As most budding yeasts undergo closed mitosis, it was observed that the components of the kinetochore are localized at the centromere all throughout the cell cycle, except for a few minutes in S-phase when the centromere DNA replicates. Unlike what is known from conventional mitosis, the kinetochores of this budding yeast were found to be attached to microtubules all throughout the cell cycle.

Initial identification of multiple complexes and proteins revealed the presence of kinetochore components that were unidentified in higher eukaryotes such as the CBF3 complex, Ctf19 complex, and components of the COMA complex [100]. Thus, in addition to the differences in centromere structure, subtle differences at kinetochore–centromeric chromatin interface exist between yeasts and other eukaryotes. In spite of such differences, recent studies have strongly hinted at more or less universal kinetochore architecture as many of the existing Ctf19 and COMA complex proteins have homologs in higher eukaryotes, constituting the CCAN (Table 9.1). Yeast homologs for the CENP-T/S/X/W DNA binding proteins, which form a nucleosome-like structure and scaffold for the Ndc80 complex in tandem with the Mis12 complex, were also identified [101,102].

Although the structure and function of the kinetochore is grossly conserved across eukaryotes, the process of assembly of this structure shows remarkable and interesting diversity. It is known that CENP-A and its chaperone Scm3 establish centromere identity [31], and the kinetochore formation is initiated by the CBF3 complex in *S. cerevisiae*. CENP-A together with the CBF3 complex recruits CENP-C and COMA complexes to the kinetochore [100,129]. This partially assembled kinetochore then brings Chl4, Iml3, and the Ctf3 complex. The presence of the COMA complex is also a prerequisite for the localization of the MIND complex, while the Spc107 and the Ndc80 complexes localize to the centromere independent of the CENP-A but dependent on the CBF3 complex [100]. Unlike the requirement of CENP-A for CENP-C recruitment in *S. cerevisiae*, CENP-C and CENP-A are mutually required for the localization of other KT proteins in *S. pombe* [119]. The KMN network complexes in *S. pombe* are interdependent on each other and Mis6 for their kinetochore localization. In contrast to these systems where a distinct hierarchal system of kinetochore assembly exists, it was observed in *C. albicans* that the entire kinetochore disassembled even when a single essential kinetochore protein was deficient [130]. In other words, a kinetochore protein required the presence of all other major kinetochore proteins for its centromeric localization.

The kinetochore proteins of all layers in the budding yeasts *S. cerevisiae* [117,131] and *C. albicans* [130,132] are constitutively present at the centromere, after their assembly soon after DNA replication, keeping the chromatids tethered throughout cell cycle. While in the fission yeast, *S. pombe*, all members of the Dam1 complex, except Dad1, arrive upon the onset of mitosis. Dad1 along with all other kinetochore proteins is constitutively present throughout cell cycle [133].

Analysis of the kinetochore of a basidiomycete, *C. neoformans*, revealed a pattern of kinetochore assembly that was more reminiscent to that of higher eukaryotes. Only the CCAN network proteins were found to be present throughout cell cycle while other outer kinetochore proteins are recruited upon the onset of mitosis in a stepwise manner in this organism [8].

While most conserved kinetochore proteins seem to have a common ancestry, the presence of CENP-B homologs in several higher eukaryotes and fungi point to convergent evolution of this kinetochore protein from pogo-like transposons. CENP-B homologs have been identified in *S. pombe, N. crassa, Penicillium sp*, and *Aspergillus sp*. [134]. While the human CENP-B has been shown to bind centromere DNA and EM has revealed its ability to contort centromere DNA in loops, its function toward centromere and kinetochore formation is still a mystery [135,136]. Although CENP-B homologs (Abp1, Cbh1, and Cbh2) in *S. pombe* have been implicated in centromere DNA binding, these proteins have also been shown to associate with the replication machinery, silencing of transposons, and genome maintenance during replication of repetitive elements [137].

9.3.4.2 Animals

The overall structure of kinetochore–microtubule interface is conserved across fungi but the attachment sites vary. CENP-A, -B, and -C are constitutively

present during the cell cycle, while the centromere-associated proteins CENP-E, CENP-F, and INCENP are present only on active centromeres and are conserved in chicken, mouse, and humans. Although CENP-A is the determinant of centromeric chromatin, presence of CENP-A alone is not sufficient to facilitate kinetochore formation in humans. The constitutive presence of a group of 15 CCAN proteins has been shown to be important for kinetochore assembly. Protein knockdown by RNAi in human cell lines revealed that CENP-A depletion led to mislocalization of most kinetochore proteins [3,109,138].

The interdependency of kinetochore proteins has been studied in various animal systems. In *C. elegans*, depletion of CENP-A prevents localization of CENP-C to chromosomes, while CENP-C depletion does not alter CENP-A localization indicating that CENP-A functions upstream to CENP-C. Similarly in *D. melanogaster*, CENP-C is dependent on CID (CENP-A homologue) for its centromere localization [3,69]. KNL1, another important kinetochore component, in association with the Mis12 complex plays a role in *Drosophila* kinetochore assembly. KNL1 is excluded from the interphase nucleus, on purpose to prevent precocious assembly of the Mis12 complex. It is imported to the mitotic nucleus during prophase where it interacts with Mis12, Nnf1a/b centromere interface, and recruits Mis12. This scaffold provides a platform to recruit Ndc80 and other outer kinetochore proteins to form a functional kinetochore layer [139].

9.3.4.3 Plants

Molecular details of the proteins and their organization in a plant kinetochore are less explored compared to other eukaryotes. Earlier studies identified plant kinetochore by the use of CREST sera that could specifically recognize kinetochores of plants such as maize and beans (*Vicia faba*). Similarly, monoclonal antibodies developed for certain mammalian kinetochore proteins also recognized proteins of the plant kinetochores. Such cross-reactivity strongly suggests conservation of some proteins in the kinetochore. CENP-A, a key component of the kinetochore machinery has been identified in many species including *A. thaliana* [140]. In addition, the protein KNL2 (Mis18 homolog) has been shown to play a role in CENP-A recruitment to the centromeres. Two other key kinetochore proteins, CENP-C and Mis12 have also been identified in many plants species [141].

9.3.4.4 Protozoa

Bioinformatic analysis suggested that many of the core machinery of the kinetochores that were found in eukaryotes are absent in kinetoplastids [142,143]. Possibly challenging well-established dogma in eukaryotic chromosome function, unconventional kinetochores were identified in *Trypanosoma bruci*, a kinetoplastid. Kinetoplastid kinetochore protein1–19 (KKT1 – KKT19) were identified by repetitive affinity purification and mass spectroscopy [144]. These proteins were found to localize onto possible centromeres that were identified earlier using etoposide-mediated Topo-II cleavage site mapping [96]. The localization patterns of these proteins and functions in chromosome segregation were

similar to other eukaryotic kinetochore proteins. Surprisingly, no homologs of these proteins were identified in other eukaryotes. Even the well-established epigenetic centromere marker CENP-A is absent in *T. bruci*. The identified KKT1–KKT19 proteins were found to be well conserved among other kinetoplastids (*T. cruzi, T. vivax, T. congolense, L. mexicana, Bodo saltans*, and *L. braziliensis*) [144]. These kinetochore proteins might have a similar structure to other known eukaryotes, as indicated by EM images that showed electron dense plaques that display end on attachments with microtubules [145].

Although a detailed architecture, assembly and composition of a kinetochore has not been worked out in other protozoans, CENP-A has been identified in *Plasmodium* species in addition to characterization of CENP-C [146]. The kinetochore in *Plasmodium* has been visualized by transmission electron microcopy (TEM) [147] and has been shown to localize on the periphery of the nucleus [91]. Other critical structural kinetochore components have been identified based upon sequence homology, including the Ndc80 complex proteins in several apicomplexans, archaeplastids, and amoebozoans [144].

9.4
Neocentromere

Functional centromere formation on an ectopic locus in the absence of a native centromere to maintain genomic stability has been demonstrated in several organisms ranging from yeast to humans. This newly formed centromere with comparable mitotic and/or meiotic stability to the native centromere is termed as the neocentromere. *De novo* neocentromere formation was first reported in 1993, when a rearrangement of chromosome 10 to an acentric chromosome mardel-10 was identified from cytogenetic screening of clinical samples from children with developmental abnormalities. Stable inheritance of the acentric chromosome mardel-10, lacking alpha satellite DNA, challenged the popular dogma of a DNA-sequence-based genetic identity for centromeres [148].

9.4.1
Naturally Occurring Neocentromeres

9.4.1.1 Humans
Since its discovery, approximately one hundred neocentromeres have been reported in humans. Except for two recently identified neocentromere on 18q22.1 and Xq27.1~27.2, all human neocentromeres formed on nonrepetitive DNA are devoid of alpha satellite repeats [149–152]. This evidence together with the observation that one of the centromeres underwent suppression in a dicentric chromosome formed through chromosomal recombination, strongly indicates the epigenetic nature of centromere identity [153].

9.4.1.2 Plants

While there are not many reports of naturally occurring neocentromeres in species other than humans, heterochromatic chromosomal elements in several plant species, including maize, wheat, and rye have some functional similarities to human neocentromeres. These subtelomeric heterochromatic chromosomal elements, known as knobs in maize, bind microtubules and contribute to the poleward movement of chromosomes in spite of several striking differences [154]. Most important of which is absence of centromeric cohesion between sister chromatids at the knob loci and stable coexistence of the knob with a functional centromere on the same chromosome [154]. In addition, unlike classical centromeres, knobs differ with respect to being dependent on other genomic loci, for example, Ab10 in maize [155,156], and lack of association with centromere defining kinetochore proteins such as CENP-C [157] and CENP-A [158].

9.4.2
Artificially Induced Neocentromeres

Although detailed studies on naturally occurring human neocentromeres and its inheritance have contributed critically to our understanding of a functional centromere, retrospective nature of the experiments has limited scope to study early events, because of the fact that they were identified long after establishment, as a part of stably inheritable karyotype [159]. To overcome this limitation, several methods of chromosome engineering have been developed that can induce neocentromere formation in various model systems to recreate the early events and observe the progression of neocentromere establishment in finer detail.

9.4.2.1 Fungi

Neocentromere formation in *S. pombe* was achieved by conditional deletion of native centromeres, utilizing controllable Cre recombinase expression and artificially placed target *loxP* sites, on either sides of the centromere. Appearance of transformants on selective media upon conditional deletion by inducing the Cre recombinase turned out to be a low chance event. While only a small fraction of survivors activated neocentromeres at subtelomeric heterochromatic regions, majority of the survivors were found to possess a chromosome that underwent telomeric fusion to retain the centromere excised acentric chromosome [160].

Neocentromere formation was induced when a native centromere was replaced by a selectable marker *URA3* in *C. albicans*. Among the two types of neocentromeres found based on their location, neocentromeres that positioned proximal to the native centromere were more frequent, while neocentromeres positioned away from the native centromere were less frequent [58]. A subsequent study observed that centromere deletion favors neocentromere formation only at a proximal region in *C. albicans*. In addition, the neocentric chromosome has been shown to lose the neocentromere when the native centromere is regained from the unaltered homolog by gene conversion to prevent centromere repositioning [60].

9.4.2.2 Animals

In an X-ray irradiation mutagenesis study in *Drosophila*, a 290 kb euchromatic fragment was found to exhibit centromere activity, when it was first translocated near a native centromere and subsequently native centromere was deleted by X-ray irradiation. On the other hand, the same fragment did not show centromeric activity in other euchromatic or heterochromatic context away from native centromere [161]. Interestingly, synthetic heterochromatin formation by tethering a heterochromatic protein HP1 to an ectopic site directly influences CID recruitment indicating a positive role of heterochromatin in neocentromere formation [162].

Similarly, chicken neocentromeres formed upon conditional deletion of native centromeres following the Cre-*loxP* strategy already described has demonstrated formation of neocentromeres on regions proximal to a preexisting native centromere [163].

9.4.3
Factors Relating to Neocentromere/Centromere Formation

9.4.3.1 Centromere-Associated Proteins

From studies in chicken and *C. albicans* it is evident that initiation of neocentromere formation is not genetically defined, rather epigenetically controlled, probably, by low levels of CENP-A present at the centromere proximal region [60,163]. Role of CENP-A as an epigenetic determinant has been further analyzed by tethering it to ectopic loci in *Drosophila*, where it seeds *de novo* centromere formation [164]. However, overexpression of CENP-A in *Drosophila* and humans showed ectopic localization on various parts of the genome without functional centromere formation [165,166]. While many of the kinetochore proteins are present at human and chicken neocentromeres, a complete absence of CENP-B [167] and reduced levels of Aurora B kinase [168] in human neocentromeres has been reported.

9.4.3.2 Chromatin

Nature of chromatin at the neocentromere loci is not uniform across studied organisms. While in *S. pombe* [160] and *Drosophila* [162], neocentromere formation is influenced by heterochromatin, chicken neocentromeres lack a well-characterized heterochromatic mark H3K9me3 [163,169]. In addition, neocentromeres of *C. albicans* present an interesting case where centromeric and neocentromeric chromatin can switch between a permissive euchromatic state to a nonpermissive heterochromatic state [60]. H3K4me2 and H3K36me3 were shown to be important components of centromeric chromatin in humans and *Drosophila* [170,171], but absent from maize, chicken, and *C. albicans* centromere and neocentromeres [163,172].

9.4.3.3 Replication Timing

Centromeres of yeasts replicate in early S-phase [173,174] that poses an intriguing hypothesis of replication-coupled loading of CENP-A at centromeric

chromatin [175,176], as early replicating DNA seems to be a preferred site for neocentromere formation in yeasts [159]. However, in *C. albicans,* neocentromeres can form on late replicating DNA but shifts its replication timing to early S-phase [177], ruling out an attractive notion that the replication timing is the only determinant of neocentromere initiation. On the contrary, chicken neocentromeres form on both early and late replicating DNA, which turned early/mid S-phase replicating DNA to late S-phase replicating, but did not change the replication timing of late replicating DNA [163]. These lines of evidence possibly indicate that neocentromere formation forces replication timing to match with that of the native centromeres, although it is not clear whether there is any preference of neocentromere formation on early or late replicating DNA.

9.5
Conclusions

Chromosome segregation, a highly orchestrated event, is essential for propagation of life. The accurate segregation of genetic and cellular assemblies to progeny requires tight regulation of multiple interdependent pathways and key molecular players. The underlying mechanisms for achieving high fidelity chromosome segregation are conserved across different kingdoms. This provides a unique platform to explore the mechanism, regulation, and maintenance of this fundamental process. The principal components, the centromere, and the kinetochore, which make up the segregation machinery, have been well studied in diverse organisms ranging from fungi to humans. However, we have seen in this chapter that subtle differences exist in the fundamental process of chromosome segregation that helps us to study the process of evolution. This evolutionary divergence is strongly evident in the diversity of the centromere structure and sequence.

Identification of the centromeres in various model organisms hinted at the possibility that there is no unifying *cis*-acting factor that specifies this unique chromosomal locus. More than thirty years of molecular dissections of this locus have identified an extensive array of cellular factors that point toward the requirement of "unique" chromatin rather than genomic sequence for the functioning of centromeres. Once a centromere is established how does it propagate after each round of DNA replication? While factors required for centromere establishment and its subsequent propagation through cell cycle could be different, making a distinction between these processes is imperative. Despite its rapid evolution, the centromere has maintained its conserved role of recruiting the kinetochore. The detailed knowledge of kinetochore architecture suggests that a high degree of structural conservation exists among eukaryotes. Better understanding of the kinetochore enabled researchers to functionally reconstitute this supramolecular complex [178]. Unfortunately, we still lack clear understanding of the dynamics of centromere/kinetochore formation and its regulation and maintenance. With the startling discovery of unconventional kinetochores in

certain protozoa, the general notion of the conserved kinetochore structure has taken a blow. On the other hand, centromeres, despite their conserved role in chromosome segregation, have rapidly diverged during evolution with some suggesting that they act as drivers of evolution themselves. Could these structural differences in centromeres/kinetochores between organisms, such as pathogens and humans, be a blessing in disguise: an inexhaustible source of drug targets? The roles of the kinetochore and centromere to aneuploidy, a common feature in many cancers, are being explored. Thus, the process of chromosome segregation continues to be an intriguing area of research for many years to come.

Acknowledgment

References

1 Westermann, S., Cheeseman, I.M. *et al.* (2003) Architecture of the budding yeast kinetochore reveals a conserved molecular core. *J. Cell Biol.*, **163** (2), 215–222.

2 Joglekar, A.P., Bouck, D. *et al.* (2008) Molecular architecture of the kinetochore-microtubule attachment site is conserved between point and regional centromeres. *J. Cell Biol.*, **181** (4), 587–594.

3 Gascoigne, K.E. and Cheeseman, I.M. (2011) Kinetochore assembly: if you build it, they will come. *Curr. Opin. Cell Biol.*, **23** (1), 102–108.

4 Thakur, J. and Sanyal, K. (2011) The essentiality of the fungus-specific Dam1 complex is correlated with a one-kinetochore-one-microtubule interaction present throughout the cell cycle, independent of the nature of a centromere. *Eukaryot. Cell*, **10** (10), 1295–1305.

5 Drechsler, H. and McAinsh, A.D. (2012) Exotic mitotic mechanisms. *Open. Biol.*, **2** (12), 120140.

6 Heath, I.B. (1980) Variant mitoses in lower eukaryotes: indicators of the evolution of mitosis. *Int. Rev. Cytol.*, **64**, 1–80.

7 Straube, A., Weber, I. *et al.* (2005) A novel mechanism of nuclear envelope break-down in a fungus: nuclear migration strips off the envelope. *EMBO J.*, **24** (9), 1674–1685.

8 Kozubowski, L., Yadav, V. *et al.* (2013) Ordered kinetochore assembly in the human-pathogenic basidiomycetous yeast *Cryptococcus neoformans. MBio*, **4** (5), e00614–e00613.

9 Aoki, K., Hayashi, H. *et al.* (2011) Breakage of the nuclear envelope by an extending mitotic nucleus occurs during anaphase in *Schizosaccharomyces japonicus. Genes Cells*, **16** (9), 911–926.

10 Yam, C., He, Y. *et al.* (2011) Divergent strategies for controlling the nuclear membrane satisfy geometric constraints during nuclear division. *Curr. Biol.*, **21** (15), 1314–1319.

11 Copeland, C.S. and Snyder, M. (1993) Nuclear pore complex antigens delineate

Chapter 12
Chromosome Components Important for Genome Stability in *Candida albicans* and Related Species

Lakshmi Sreekumar, Neha Varshney and Kaustuv Sanyal

Abstract Pathogenic microorganisms have been constantly evolving to battle the various responses elicited by the host upon their invasion. *Candida albicans* bears no exception to this trend. It is not only a fitter pathogen but also a successful gut commensal in humans. The extraordinary genome plasticity of *C. albicans* makes its survival possible in such widely diverse host niches. This chapter focuses on various chromosomal elements that are involved in maintaining the genome dynamics and stability in *C. albicans*. Here, we discuss molecular players of the basic cellular processes that lead to duplication of chromosomes, their faithful segregation in progeny, and chromosome maintenance in *C. albicans* and its related species.

12.1 Insights into DNA Replication

DNA is a self-replicating molecule. The faithful duplication of the genetic material is achieved at the synthetic or S-phase of the cell cycle. "Origins" are the sites on the chromosome where DNA replication initiates. These initiator sequences act as a platform for multi-protein sub-complexes to assemble and facilitate the opening up of the double-stranded DNA to form "replication bubbles". The rate limiting step in this pool is the assembly of the pre-replicative complex, which primarily comprises of evolutionarily conserved proteins like the Origin Recognition Complex (ORC), Cdc6, Cdt1, and the MiniChromosome Maintenance (MCM) helicase (see Table 12.1). Sequential activation of MCM helicase by loading of accessory factors like Cdc6 and Cdt1 makes it a tightly regulated and highly coordinated process. Upon loading of the helicase machinery, DNA polymerase is ready to start the addition of nucleotides to the template DNA (Masai et al. 2010). Termination of replication is governed by termination sites. Any damage to DNA during this

L. Sreekumar · N. Varshney · K. Sanyal (✉)
Molecular Mycology Laboratory, Molecular Biology and Genetics Unit, Jawaharlal Nehru Centre for Advanced Scientific Research, Jakkur, Bangalore 560064, India
e-mail: sanyal@jncasr.ac.in

233

process due to genomic insults like gamma rays or chemical mutagens or even hostile endogenous conditions in the cell is well taken care by the "sensors" or DNA damage check point proteins. All of these mechanisms are conserved across eukaryotes.

A replication origin or "*ori*" is a sequence where the replication bubble is first formed to initiate the templated synthesis of new DNA. These sequences may or may not have strict sequence dependence (Leonard and Mechali 2013; Rui 1999). Initially, when origins were examined in *Saccharomyces cerevisiae*, they were cloned in episomal vectors containing metabolic markers to yield high frequency transformants in auxotrophic strains. These Autonomously Replicating Sequences (*ARS*s) give rise up to a thousand colonies per microgram of the transforming DNA. *ARS* plasmids have proved to be efficient tools to clone yeast DNA sequences. Moreover, several genomic library preparations, in the past decades, have provided information about sequences that can act as putative origins in the genome.

Candida genus includes a diverse group of organisms possessing both pathogenic and non-pathogenic species where several aspects of DNA metabolism have been studied. Additionally, the pathogenic members are known to be genetically more resistant to DNA-damaging agents than the non-pathogenic ones (Rodrigues et al. 2014). Among *Candida* species, *ARS*s have been reported in *C. albicans, Candida guilliermondii, Candida utilis, and Candida glabrata*. In *C. albicans*, a genomic DNA library was constructed in an *ADE2* containing plasmid of *S. cerevisiae* (Kurtz et al. 1987). Upon transformation in an adenine auxotroph of *C. albicans*, a thousand transformants were obtained of which a minor fraction were genomic integrants as well. However, after eight generations of growth in non-selective media, only 1–2% of the progeny were still Ade+, indicating that *ARS* plasmids, if not integrated into the genome, have very low mitotic stability in this organism. In one of the earlier studies, an 8.6 kb fragment was isolated from *C. albicans* genome and a part of it was subcloned in a bacterial plasmid containing *CaLEU2* and *CaURA3* genes (Cannon et al. 1990). These plasmids could yield a transformation efficiency of 2.15×10^3 CFU/µg and 1.91×10^3 CFU/µg of plasmid DNA in *S. cerevisiae* and *C. albicans,* respectively. Upon further sequence analysis, these *ARS* components were shown to be similar to *S. cerevisiae ARS* elements. Subsequently, a 15.3 kb fragment was cloned in a bacterial plasmid and it showed properties of an *ARS* plasmid. However, these *ARS* plasmids are subject to random integration and multimer formation in the genome (Herreros et al. 1992).

Centromeres (*CENs*) are DNA loci that act as platforms for chromosome segregation. *CENs* in *C. albicans* were found to be the earliest replicating regions in the S-phase of every cell cycle (Koren et al. 2010). The replication profile of chromosome 1 showed that origins flanking the *CENs* are the first to fire during every S-phase. Genomic origins were identified by the binding of ORC proteins and validated using the approach of 2-dimensional gel electrophoresis. A more detailed analysis of the chromosome 5 and chromosome 7 in *C. albicans* revealed a strong correlation between the effects of the *CEN* proximal origin on *CEN* function (Mitra et al. 2014). *CEN* proximal origins contributed to replication fork stalling at the *CEN*. This fork stalling or random termination was shown to be a kinetochore-mediated phenomenon. Upon

Table 12.1 Protein components associated with major chromosomal elements in *C. albicans*

Name	Allele name	Feature type	Description	Orf no.
Replication origin-associated proteins				
MCM2	CR_03830C_B	Uncharacterized	Phosphorylated protein of unknown function; transcription is periodic with a peak at M/G1 phase	orf19.4354
MCM3	C2_07350W_B	Uncharacterized	Putative DNA replication protein; periodic mRNA expression that peaks at the M/G1 phase	orf19.1901
CDC54	C1_12550C_B	Uncharacterized	Putative pre-replication complex helicase subunit; periodic mRNA expression, peak at cell cycle M/G1 phase	orf19.3761
CDC46	C2_06250C_B	Uncharacterized	Putative hexameric MCM complex subunit; predicted role in control of cell division; periodic mRNA expression. peak at cell cycle M/G1 phase	orf19.5487
MCM6	CR_02110W_B	Uncharacterized	Putative MCM DNA replication initiation complex component; mRNA expression peaks at cell cycle M/G1 phase	orf19.2611
CDC47	C2_09020W_B	Uncharacterized	Phosphorylated protein described as having role in control of cell division; RNA abundance regulated by tyrosol and cell density	orf19.202
ORC1	C1_03070C_B	Uncharacterized	Putative origin recognition complex (ORC) large subunit; essential for viability; similar to *S. cerevisiae* Orc1p subunit	orf19.3000
ORC2	C2_10760C_B	Uncharacterized	Phosphorylated protein of unknown function	orf19.5358
ORC3	C3_03750C_B	Verified	Protein similar to *S. cerevisiae* Orc3p, ORC component	orf19.6942
ORC4	C5_02150C_B	Uncharacterized	Phosphorylated protein similar to *S. cerevisiae* Orc4	orf19.4221
ORC5	CR_06960W_B	Uncharacterized	Ortholog(s) have ATP binding, DNA replication origin binding activity	orf19.2369
ORC6	C1_01000C_B	Uncharacterized	Phosphorylated protein of unknown function	orf19.3289

(continued)

Table 12.1 (continued)

Name	Allele name	Feature type	Description	Orf no.
POL30	C4_01770W_B	Verified	Proliferating Cell Nuclear Antigen (PCNA), forms homotrimeric sliding clamp for DNA polymerases; cell density; stationary phase enriched protein	orf19.4616
CDC13	C1_00360W_B	Verified	Essential protein with similarity *to S. cerevisiae* Cdc13p, involved in telomere maintenance	orf19.6072
Centromere-associated proteins				
CSE4	C3_00860W_B	Verified	Centromeric histone H3 variant; role in structural changes of centromeric nucleosomes during cell cycle, centromeric DNA binding	orf19.6163
MIF2	C6_02780C_B	Verified	Centromere-associated protein; similar to CENP-C proteins; Cse4p and Mif2p colocalize at *C. albicans* centromeres	orf19.5551
MTW1	C2_09840W_B	Verified	Kinetochore component; the amount of Nuf2p and Mtw1p protein detected at each centromere is consistent with a single kinetochore-microtubule attachment site	orf19.1367
DAM1	C1_09730W_B	Verified	Component of DASH complex, microtubule plus end binding, chromosome segregation by coupling kinetochores to spindle microtubules	orf19.4837
DAD1	C1_13710C_B	Verified	Component of DASH complex, microtubule plus end binding, chromosome segregation by coupling kinetochores to spindle microtubules	orf19.5008.1
DAD2	C2_05210W_B	Verified	Component of DASH complex, microtubule plus end binding, chromosome segregation by coupling kinetochores to spindle microtubules	orf19.3551
ASK1	C4_01150W_B	Verified	Component of DASH complex, microtubule binding, chromosome segregation by coupling kinetochores to spindle microtubules	orf19.4675

(continued)

Table 12.1 (continued)

Name	Allele name	Feature type	Description	Orf no.
SPC19	C1_03980W_B	Verified	Component of DASH complex, microtubule plus end binding, chromosome segregation by coupling kinetochores to spindle microtubules	orf19.4473
RAD51	CR_02200C_B	Verified	Homologous recombination and DNA repair; slow growth and increased white-to-opaque switching frequency in null mutant	orf19.3752
RAD52	C6_00510C_B	Verified	Required for homologous DNA recombination, repair of UV- or MMS-damaged DNA, telomere length, UV-induced LOH; constitutive expression; slow growth, increased white-to-opaque switch	orf19.4208
MRE11	C7_01340W_B	Verified	Putative DNA double-strand break repair factor, $3'–5'$ exonuclease activity; involved in response to oxidative stress and drug resistance	orf19.6915
SCM3	C3_01770C_B	Not verified	Unknown	orf19.1668
Telomere-associated proteins				
STN1	C2_08470C_B	Verified	Protein involved in telomere maintenance; forms a complex with Ten1p	orf19.3631
TEN1	CR_01010W_B	Verified	Protein involved in telomere maintenance; forms a complex with Stn1; transcription is regulated upon yeast-hypha switch	orf19.3255
RAP1	C2_10080W_B	Verified	Transcription factor; binds telomeres and regulatory sequences in DNA; involved in telomere maintenance; represses hyphal growth under yeast-favoring conditions	orf19.1773
EST1	C5_05470W_B	Verified	Telomerase subunit; allosteric activator of catalytic activity, but not required for catalytic activity	orf19.4045

(continued)

Table 12.1 (continued)

Name	Allele name	Feature type	Description	Orf no.
TERT	C1_08130C_B	Verified	Telomerase reverse transcriptase; catalytic protein subunit of telomere synthesis; essential for telomerase activity; has telomerase-specific motif T and other conserved reverse transcriptase motifs	orf19.5089
EST3	C3_00430W_B	Verified		orf19.5423

Source: *Candida* genome database

deletion of these *CEN* proximal origins, the *CEN* function was compromised and occupancy of CENP-A over the central core of *CEN7* was seen to be ablated.

The dual nature of origins in the genome of *C. albicans* was shown in a study where the nature of origins was claimed to differ on the basis of the chromosomal location context (Tsai et al. 2014). A mini-*ARS* screen was used to identify bona fide *ARSs*. *CEN* proximal origins were termed "epigenetic," whereas arm origins were called "hard wired" sequence-dependent origins. Genome-wide origins were mapped based on ORC binding, nucleosome depletion patterns, and the afore-mentioned *ARS* screen. All the eight chromosomes of *C. albicans* showed the characteristic GC-skew pattern that is reminiscent of bacterial chromosomes.

Several *ARSs* have been reported in other *Candida* species. In the industrially important yeast *C. utilis,* a 6.6 kb *ARS* has been reported which was obtained after a library preparation and it had a mitotic stability of less than 1% (Iwakiri et al. 2005). Unlike other yeasts where the size of a functional *ARS* is limited to 100–200 bp, in *C. utilis* the smallest *ARS* is 2.8 kb, the rest being 5 kb and above. *C. guilliermondii* is an important yeast of biotechnological interest (Papon et al. 2013). In an attempt to standardize the electroporation method of transformation in this system, scientists stumbled upon an *ARS* located upstream of the *URA5* gene (Foureau et al. 2013). In *silico* analysis of the sequence predicted *ARS*-like elements (ALS) which apparently increased the transformation efficiency by a thousand fold. The most recent addition of a *Candida* species in the collectible of *ARS*s is in *C. glabrata* (Descorps-Declere et al. 2015). Though it is a member of the *Saccharomycota*, it underwent a whole-genome duplication (WGD) event followed by extensive loss of genes (Fitzpatrick et al. 2006). Its genome has several long tandem repeats or megasatellites. Using deep sequencing and chromosome conformation capture (3C) experiments, the replication landscape of *C. glabrata* was determined (Descorps-Declere et al. 2015). This study identified 253 replication origins and 275 *ARS*s in this pathogenic yeast. Using a time course S-phase study, centromeres, *MAT* loci, and most histone genes were found to be located in early replicating domains of the chromosome, whereas fragile sites and chromosome breakpoints were shown to be late-replicating. The chromosome conformation capture (3C) experiment showed clustering of early replicating origins which were not observed in non-subtelomeric megasatellites. In spite of their significant divergence, the

replication program of *C. glabrata* bore striking similarities to that of *S. cerevisiae*, once again proving that binding specificities of the MCM/ORC complexes are highly conserved on diverse DNA sequences.

Members of the pre-replicative complex (pre-RC) have not been characterized fully in the *Candida* species. Homologs of the MCM/ORC complex are yet to be characterized. The *C. albicans* Proliferating Cell Nuclear Antigen (PCNA) was shown to be an essential protein that functions as a sliding clamp for the DNA polymerase (Manohar and Acharya 2015). It is loaded by the clamp loader RFC. It is a homotrimeric ring encircling the double-stranded DNA. PCNA interacts physically with DNA polymerase ε. Cell cycle regulation is mediated tightly by the concerted activity of cyclin-CDK complexes. *C. albicans* has two B-type cyclins— Clb2 and Clb4. Clb2 is the functional homolog of Clb5/Clb6 (Ofir and Kornitzer 2010).

12.2   Segregating the Nuclear Material: When and How?

Aneuploidy is a major feature for a wide range of health disorders including cancers. Even in certain fungal species, like *C. albicans*, aneuploidy provides a unique advantage to the organism to maintain genome plasticity (reviewed in Sanyal 2012). The altered ploidy state seen in terms of genomic rearrangements and segmental aneuploidy of chromosomes stems from various exposures to stress conditions, one of them being a prolonged treatment of this fungal pathogen to antifungals thereby providing it a survival strategy in hostile environment niches.

Chromosomes are required to be segregated with utmost fidelity in a cell before it divides. Centromeres (*CEN*s) are DNA loci responsible for precise chromosome segregation. The proper attachment of the protein machinery residing on the *CEN* DNA, the kinetochore with the microtubules emanating from opposite spindle poles, is the determinant of high fidelity chromosome segregation. *CEN*s are typically located in regions of a genome with less gene density and a reduced rate of recombination. While in most organisms they occupy discrete sites on monocentric chromosomes, they may extend up to the entire length of a holocentric chromosome. A centromere can be physically located either at the middle of a chromosome (metacentric) or at the end of a chromosome (acrocentric or telocentric). Several centromeric proteins (CENPs) were first identified serendipitously by isolating anti-centromere antibodies from individuals with the CREST disease (Palmer et al. 1987). Subsequently, other members of CENP family were identified in many organisms. Centromere protein A (CENP-A)/Cse4 is a variant of histone H3 present exclusively at functional *CEN*s.

Centromeres are broadly classified as point, small regional, and large regional centromeres (reviewed in Roy and Sanyal 2011). The small regional centromeres of 3–5 kb non-repetitive, unique sequences were identified in three *Candida* species— *C. albicans*, *Candida dubliniensis,* and *Candida lusitaniae* (Sanyal et al. 2004; Padmanabhan et al. 2008; Kapoor et al. 2015). The *CEN* DNA of all eight

chromosomes of *C. albicans* were identified and characterized by chromatin immunoprecipitation using anti-CaCse4 antibodies (Sanyal et al. 2004). The Cse4-binding region was mapped to a 3–4.5 kb stretch, deletion of which resulted in a high frequency loss of only the altered chromosome. Similarly, *CEN* sequences on all eight chromosomes were identified in the closely related yeast, *C. dubliniensis* (Padmanabhan et al. 2008). Interestingly, all the eight *CEN*s in these two organisms lack any conserved sequence including any motifs or repeats (Mishra et al. 2007). However, *CEN1, CEN4, CEN5,* and *CENR* of both *C. albicans* and *C. dubliniensis* possess unique chromosome-specific pericentric-inverted repeats (IRs). In addition to IRs, LTRs were found within *CEN2, CEN3, CEN5,* and *CEN6* of *C. albicans.* The conservation of the relative position of the *CEN* in orthologous regions with respect to the adjacent *ORF*s (synteny) in these two *Candida* species hinted toward the possibility that the *CEN* DNA sequence alone may not determine its identity. This was further validated by an elegant experiment where an 85-kb chromosome fragment (CF) exhibiting *CEN* function in vivo failed to show de novo *CEN* activity when shuttled as a naked DNA from *C. albicans* back into *C. albicans* (Baum et al. 2006). Even the naked *CEN* DNA introduced 6.7 kb away from the native locus on chromosome 7 could not recruit CENP-A in *C. albicans.* Inability of a circular *ARS* plasmid carrying the CENP-A-rich *CEN* region to produce a stable minichromosome in *Candida* cells also bolstered the fact that DNA sequence is not the only determinant of the *CEN* identity. In budding yeast, the CENP-A occupying *CEN* chromatin has a distinct nucleosome occupancy pattern from the bulk chromatin in having a ladder of DNA fragments after partial digestion with micrococcal nuclease (MNase) which is a characteristic of regularly spaced nucleosomes. However, *C. albicans* gives a more distinct pattern in this assay, where mononucleosomes, dinucleosomes, and a smeared pattern were detected around *CEN7* with an absence of the nucleosome ladders. In the three species of *Candida* clade, *C. lusitaniae, Pichia stipitis,* and *Yarrowia lipolytica,* one obvious GC-poor region was found to be located on each chromosome speculated to be the *CEN,* even though the *CEN*s in *C. albicans* and *C. dubliniensis* are not seen to be GC-poor (Lynch et al. 2010). These *CEN* locations were experimentally validated in *C. lusitaniae* by ChIP-sequencing of two key centromeric proteins, Cse4 and Mif2 (Kapoor et al. 2015). The centromere DNA sequence in *C. lusitaniae* was unique for each chromosome and spanned 4–4.5 kb, similar to regional centromeres of *C. albicans.* A distinct pattern of histone modifications, methylated H3K79 and H3R2 but lack of methylation of H3K4, which is otherwise seen in regional centromeres, was enriched at centromeric chromatin in *C. lusitaniae.* However, unlike other regional centromeres, there was no evidence for the presence of pericentromeric heterochromatin in *C. lusitaniae.* The pericentromeric regions in *C. albicans* are assembled into an intermediate chromatin state harboring features of both heterochromatin and euchromatin (Freire-Beneitez et al. 2016). The regions flanking the central core are associated with nucleosomes that are hyperacetylated-like euchromatic regions and hypomethylated on H3K4-like heterochromatin. However, the pericentromeric regions in *C. lusitaniae* are not associated with hypoacetylated histones or sirtuin deacetylases that generates

heterochromatin in other yeasts. In other words, the lack of the *URA3* reporter gene repression when inserted adjacent to the centromere supports the fact that there is no pericentromeric heterochromatin in *C. lusitaniae* (Kapoor et al. 2015).

Strikingly, the structure of the *CEN* has been rewired in the closely related *Candida* species, *Candida tropicalis* where each of the seven chromosomes comprises a 2–5 kb non-repetitive mid-core region that forms the binding sites for CENP-A (Cse4) and CENP-C (Mif2) flanked by 2–5 kb inverted repeats (IRs) (Chatterjee et al. 2016). The repeat associated centromeres of *C. tropicalis* also share a high degree of sequence conservation with each other. In spite of the observed rapid change in the sequence and organization of *CEN*s in these closely related species, the AT-content of the CENP-A bound *CEN* DNA sequence was found to be similar in *C. albicans* and *C. tropicalis*. All *C. albicans CEN*s are free of transposons while one *CEN* in *C. tropicalis* is associated with retrotransposons (Mishra et al. 2007; Chatterjee et al. 2016). Interestingly, these small regional *CEN*s maintain their uniform *CEN* sizes ($\sim$3 kb) despite the absence of any obvious boundary elements such as the tRNA genes.

*C. glabrat*a exhibits similarities to *S. cerevisiae* in its *CEN* organization. A 451 bp fragment exhibiting *CEN* activity was isolated and sequenced (Kitada et al. 1997). It was shown to possess three elements similar to *S. cerevisiae*, Cg*CDEI*, Cg*CDEII*, and Cg*CDEIII*. Substitution mutation analysis revealed the requirement of at least Cg*CDEI* and Cg*CDEIII* for *CEN* function in this organism. Despite the presence of functional similarities between *CEN*s of *C. glabrata* and *S. cerevisiae*, there exists species specificity in *CEN* function. Similarly, in *C. maltosa* a 325 bp fragment was shown to harbor *CEN* activity with *CDEI* and *CDEII*-like regions of *S. cerevisiae* (Ohkuma et al. 1995).

One of the most astonishing examples of epigenetic changes within a genome is the formation of neocentromeres. Neocentromeres are *CEN*s arising at atypical chromosomal loci. *C. albicans* provides an excellent model system to study this process. Upon deletion of a native centromere (*CEN5*), neocentromeres could form efficiently in *C. albicans* in various locations on the same chromosome. They could be proximal neocentromeres, formed close to the location of the native *CEN* and distal neocentromeres which are formed at other locations on the chromosomes (Thakur and Sanyal 2013; Ketel et al. 2009). A more comprehensive study in multiple chromosomes (chromosome 1, 5 and 7) of *C. albicans* and additionally involving the closely related *C. dubliniensis* suggested that neocentromere formation is a conserved mechanism in these organisms and occurs in *CEN* proximal regions (Thakur and Sanyal 2013). The distal neocentromeres show low CENP-A enrichment as compared to the native *CEN*s leading to chromosome loss, asserting the fact that *CEN* proximal sites are the preferred sites for neocentromere formation (Thakur and Sanyal 2013). Since *CEN* chromatin can have a negative effect on the gene expression in *C. albicans*, all neocentromeres are formed in the intergenic regions of the chromosomes. Although deletion of the endogenous *CEN7* leads to neocentromere formation in *C. albicans*, in a fraction of strains the *CEN* was repositioned to the endogenous locus by gene conversion events through copying *CEN7* of the unaltered homolog (Thakur and Sanyal 2013). Additionally, there is

no correlation between the neocentromere formed and length of the deleted *CEN* region. Deletion of *CEN1* and *CEN5* led to neocentromere formation which was two to four times the size of native *CEN*s suggesting that length of the neocentromere is variable across chromosomes (Scott and Sullivan 2014). *C. albicans* has a combination of neocentromere properties observed in various plants, animals, and fungal species asserting the conservation of such mechanisms across species (Marshall and Choo 2009).

The early replicating timing of *CEN*s could be related to loading of CENP-A, that is also seen to take place in the early S-phase of other yeast species (Aravamudhan et al. 2013; Pearson et al. 2004; Takahashi et al. 2005). However, the formation of a neocentromere at a late-replicating domain in *C. albicans* has created a paradigm shift in the view that replication timing is not the sole determinant of de novo centromere assembly (Koren et al. 2010). The neocentromeres occurring in the late-replicating domain were accompanied by shifts in replication timing. These neocentromeres became the first to replicate and became associated with the origin recognition complex (Koren et al. 2010).

Kinetochores are large macromolecular complexes assembled on *CEN* DNA. The kinetochores are tri-layered structures revealed in the early microscopic images of mitotic chromosomes in human cells. It comprises an inner layer interacting directly with the *CEN* DNA, the outer layer forming the chromosomal attachment site for the microtubule plus end and middle layer, bridging the two layers (see Table 12.1). However, due to the small cellular size of unicellular organisms like yeasts, the ultra-structure of a kinetochore cannot be ascertained. Immuno-localization of kinetochore proteins in these organisms appears as puncta of clustered kinetochores at the nuclear peripheral regions located close to spindle pole bodies.

CaCse4, the yeast homolog of CENP-A, was identified and localized as an intense dot-like signal (a cluster of 16 kinetochores), co-localizing with the nucleus in *C. albicans* (Sanyal and Carbon 2002). Like in other organisms, CaCse4 is an essential protein and is involved in kinetochore formation in *C. albicans*. CaCse4 is required for proper chromosome segregation as depletion of this protein results in the accumulation of large buds in the population. Similarly, *CaMIF2* (CENP-C homolog) was shown to be an essential gene in *C. albicans*. CaMif2 colocalizes with CaCse4 and is enriched at all *CEN*s in *C. albicans* (Sanyal et al. 2004). Subsequently, CaMtw1, a homolog of human Mis12/Mtw1 in *C. albicans,* was characterized for its function in the process of kinetochore-microtubule mediated chromosome segregation in *C. albicans* (Roy et al. 2011). CaMtw1 is an essential protein required for G2/M progression and proper chromosome segregation during mitosis. It is required for spindle positioning and morphogenesis as well. A fungus-specific outer kinetochore protein complex, the Dam1 complex mediates attachment of the chromosomes to the mitotic spindle. It comprises of ten different subunits which oligomerize in various ways to form a ring that interacts with the microtubules. Dam1, Dad1, Dad2, Ask1, and Spc19 subunits of the Dam1 complex are shown to be essential for viability and are indispensable for proper chromosome segregation in *C. albicans* (Thakur and Sanyal 2011; Burrack et al. 2011). Dad2

shares functional similarity with Dam1 and was shown to be localized at the mid-zone in addition to its kinetochore localization. Although the recruitment of the Dam1 complex was shown to be independent of the kinetochore-microtubule interactions, the function of this complex was shown to be monitored by the spindle assembly checkpoint (SAC). Also, the Dam1 complex is required to prevent spindle elongation in early mitosis. The first biochemical study on a kinetochore protein in *C. albicans* was done on the Dam1 complex. Dad1 was shown to be an intrinsically disordered protein with a structure similar to its *S. cerevisiae* counterpart (Waldo et al. 2010).

The correlation of "one microtubule/kinetochore" was established in *S. cerevisiae* having a point *CEN* in contrast to multiple microtubules/kinetochore in *S. pombe* which contains a large regional *CEN* (Winey et al. 1995; Ding et al. 1993). Interestingly, the essentiality of the Dam1 complex can be correlated with one microtubule-one kinetochore type of interaction. Dam1 is essential in the budding yeasts, *S. cerevisiae* and *C. albicans,* while it is non-essential in the fission yeast, *S. pombe. C. albicans,* harboring a regional *CEN*, supports one microtubule/kinetochore-like interaction (Joglekar et al. 2008). Strikingly, CENP-A over-expression could rescue the depletion of Dam1 by increasing the level of other kinetochore proteins and hence microtubules to form a functional kinetochore (Burrack et al. 2011).

Recruitment of most of kinetochore proteins studied so far is regulated by CENP-A. However, a few master regulators such as Ndc10, Scm3 (*S. cerevisiae*), Mis6, the Mis16–Mis18 complex, and Ams2 (*S. pombe*) and Rad51–Rad52/CENP-C (*C. albicans*) were shown to influence CENP-A localization in various yeast species (Camahort et al. 2007; Hayashi et al. 2004; Takahashi et al. 2005; Mitra et al. 2014; Roy et al. 2011). Recruitment of CENP-A at the *CEN* in *C. albicans* is mediated by homologous recombination (HR) proteins, where the replication forks coming from *CEN* proximal origins stall at the *CEN* in a kinetochore-dependent manner (Mitra et al. 2014). Fork stalling at the *CEN* is reduced in the absence of HR proteins Rad51 or Rad52. Null *rad51* or *rad52* mutants exhibit an increased kinetochore declustering and degradation of CENP-A. The physical association of CENP-A and Rad51/Rad52 in a complex is an indicator of an HR-mediated CENP-A recruitment mechanism in this organism. CENP-A levels in HR mutants, such as *mre11, rad51,* and *rad52* null mutants, have been observed to be low in *C. albicans* (Mitra et al., 2014). This emerging role of HR proteins at the centromere raises an interesting possibility of the involvement of a Holliday Junction Recognition Protein, HJURP/Scm3 at the *CEN*. The timing of loading of CENP-A with respect to cell cycle varies in different species. Experiments suggest that CENP-A is deposited during S-phase in *S. cerevisiae* in contrast to its biphasic loading observed in *S. pombe*. Paradoxically, anaphase-specific loading of CENP-A was also indicated in contrast to the previous report of CENP-A deposition during S-phase in *S. cerevisiae*. Strikingly, *C. albicans* was also shown to have an anaphase coupled loading of CENP-A (Shivaraju et al. 2012).

While the kinetochore assembly occurs in a step-wise manner on a point *CEN* in *S. cerevisiae*, the kinetochore architecture is stabilized in a coordinated interdependent manner by its individual components in *C. albicans*. Intriguingly in *C. albicans*, the kinetochore proteins from the outer and middle layers influence the localization of the inner kinetochore protein, CENP-A. The *CEN* localization of CENP-A was shown to be remarkably reduced in the absence of the inner (Mif2/CENP-C), middle (Mis12/Mtw1), and outer (Dam1 complex and Nuf2) kinetochore proteins (Thakur and Sanyal 2012). Even Mif2 exhibited dependency on the Mtw1 for its recruitment (Roy et al. 2011). The unprecedented observation of outer kinetochore proteins influencing the localization of CENP-A in *C. albicans* suggests that the kinetochore sub-complex assembles in a unique interdependent concerted manner to form a stable kinetochore. The kinetochore collapses in the absence of its essential components indicating that the kinetochore may not be a layered structure in *C. albicans*. Most strikingly, the kinetochore protects CENP-A from proteasomal degradation in *C. albicans*. Even the newly synthesized CENP-A molecules fail to rescue the kinetochore integrity defects strengthening the fact that individual kinetochore components are absolutely essential for protecting the *CEN*-bound CENP-A molecules in *C. albicans*.

Several *Candida* species reside as a harmless commensal in the human gastrointestinal tract and genitourinary tract. However, they can be opportunistic human pathogens causing superficial to fatal systemic infections in immuno-compromised patients. Candidemia is the fourth most common cause of hospital-acquired infections with huge annual costs for medicare of patients. Hence, it is important to identify potential drug targets in these organisms. The outer kinetochore protein, the Dam1 complex, is essential and localized at the kinetochore throughout the cell cycle in *C. albicans*. Interestingly, no homologue of Dam1 has been found in metazoan system. The fungal specificity of the Dam1 complex and its crucial role in kinetochore-microtubule attachments makes it a potent drug target in this pathogenic fungus and its related species.

12.3   Cell Longevity Measures Taken by *Candida albicans*

To prevent the breakage and fusion of linear chromosomes in the nucleus, the cell employs a machinery to sequester the ends or "telomeres" such that it is masked from the exposure of DNA damage response. Telomeres act as caps at chromosomal ends and also maintain the replicative lifespan of chromosomes. Telomeres are G-rich repetitive sequences present at the end of every chromosome and are maintained by the reverse transcriptase enzyme, telomerase, that uses the 3'G-rich overhang as a template to synthesize these repetitive sequences (O'Sullivan and Karlseder 2010).

C. albicans is an interesting model to study telomere biology because of various reasons. Unlike other species, it contains unusually long (23 bp-long) distinct and regular telomere repeat units. Also, the overall length of telomeres can be varied

depending on the growth conditions provided to the organism, making this a novel yet interesting mechanism of telomere length regulation (McEachern and Hicks 1993). There exists a definite interplay between telomerase activity and telomere recombination for telomere maintenance. In wild-type *Candida* cells, it has been difficult to detect a senescent phenotype owing to heterogeneity in the sizes of Terminal Restriction Fragments (TRFs) of a single telomere studied over a period of time (Singh et al. 2002). Using computational and experimental data, homologs of telomerase proteins have been identified in *Candida* genomes. A rapid evolutionary divergence of telomere-associated proteins has been evident. Most of the studies on telomeres and telomerase components have been carried out in *C. albicans* so far with a very little documentation available in other *Candida* species. Telomere binding factors can be categorized as single-stranded and double-stranded nucleotide-binding proteins (see Table 12.1).

The Cdc13-Stn1-Ten1 (CST) complex forms a major heterotrimeric complex involved in telomere protection, via its oligosaccharide/oligonucleotide-binding (OB)-fold domain that recognizes single-stranded DNA (Lue et al. 2013). The *Candida* Cdc13 homologs are smaller than the *S. cerevisiae* counterpart, as they lack the OB1 domain at the N-terminus required for dimerization of Cdc13 (Lue and Chan 2013). However, the DNA-binding domain and the C-terminal OB4 domains are shared between *C. albicans, C. tropicalis,* and *S. cerevisiae.* Unlike Cdc13, the *Candida* Stn1 and Ten1 share considerable homology with their *S. cerevisiae* counterpart. Crystal structure analysis of the *C. tropicalis* Stn1–Ten1 complex reveals that these proteins do contain a single OB-fold domain. Additionally, the Stn1–Ten1 complex is structurally similar to the RPA complex (Rpa2–Rpa3) (Sun et al. 2009). Both these proteins and their interactions regulate telomere length as their null mutants exhibit long and heterogeneous telomeres. In the *C. albicans* strain BWP17, telomeres normally range from 1 to 5 kb in length. If the telomeres are "de-protected," it leads to the formation of extra-chromosomal telomeric circles resulting from recombination mediated telomere maintenance and Telomere Rapid Deletion (TRD), called t-loops (Tomaska et al. 2009).

Rap1 is a double-stranded DNA-binding protein which is required for mating type and transcriptional silencing. It is a multifunctional protein owing to its complex domain architecture. Rap1 homologs have three functional domains—the BRCT domain required for protein–protein interaction, the MYB domain for DNA binding, and the C-terminal RCT (Rap1 C-terminus) which is thought to interact with four other proteins, namely Rif1, Rif2, Sir3, and Sir4. The Rap1 homolog in *C. albicans* lacks the RCT domain (Uemura et al. 2004). Rap1 is not essential for cell viability, but null mutants show aberrant telomeric recombination leading to longer telomere lengths and an increased formation of t-circles (Yu et al. 2010; Biswas et al. 2003). Rap1 seems to have overlapping functions with Stn1–Ten1, but they are not redundant. Due to the lack of a C-terminal domain, the Sir-mediated subtelomeric silencing by Rap1 may not be mediated the same way in *C. albicans* as it occurs in other species studies so far (Feeser and Wolberger 2008). Apart from these proteins, Rif1/Rif2 and the Ku complex have been studied in other species, but their *Candida* homologs are yet to be characterized.

The telomerase complex is known to consist of a catalytic protein subunit (Est2p/TERT) and two non-catalytic subunits (Est1p, Est3p). The TERT protein is well conserved across species owing to its catalytic function. On the other hand, Est1 exhibits a low level of sequence conservation. The C-terminus of CaEst1, being involved in its binding to the telomerase RNA, Cdc13 and the telomerase DNA, is quite similar to ScEst1. CaEst3 shows the least level of conservation with its *S. cerevisiae* counterpart (Singh et al. 2002).

CaTER1 and CaEst2 (TERT) have been found to be essential for the catalytic activity of the telomerase complex (Singh et al. 2002). In the same report, the authors have proposed a potential telomere "capping" role of Est1p and TERT. In *C. albicans*, deletion of *EST2* results in progressive telomere attrition, thereby increasing the level of G-strand overhangs, whereas *EST1* or *EST3* deletion did not yield the same phenotype indicating that telomerase can physically protect telomeres in a catalytically independent manner. It has been speculated that the catalytic core complex physically blocks the access of degrading enzymes to telomere ends, thereby rendering its protection. Such an accumulation of G-strand overhangs was seen in null mutants of telomere protecting proteins like Cdc13 and the Ku complex. A similar phenotype was observed in a null mutant of the RNA component of telomerase, *TER1*. The notable feature of the *CaTER1* gene is that it possesses the largest *TER* gene identified so far (1,554 bp); however, the template for *CaTER1* gene is smaller (250 bp). The significance of this finding is still unknown (Hsu et al. 2007).

Telomeric *ORFs* (*TLOs*) are a family of telomere associated *ORFs* in *C. albicans* and *C. dubliniensis* that encode a subunit of the Mediator complex which is used for the recruitment of RNA pol II during transcription initiation. *C. albicans* has 15 TLOs, whereas *C. dubliniensis* has two (Haran et al. 2014; Anderson et al. 2012; Sullivan et al. 2015). This expansion of the *TLO* gene family could suggest a flexible transcriptional network in this organism that gives it the plasticity to thrive in varying environmental niches. These 15 *TLO* genes can be classified into three clades—*TLOα* (6 members), *TLOβ* (one member), and *TLOγ* (7 members) and a pseudogene. They primarily differ by the presence of a Long Terminal Repeat (LTR) that gives rise to splice variants. Members of *TLOγ* clade produce both spliced and unspliced transcripts. *TLOα* genes are expressed at the highest levels, whereas *TLOγ* genes are expressed at very low levels. A plethora of Mediator subunit composition, caused by a broad range of expression levels of *TLO* genes, provides a greater flexibility to *C. albicans* allowing it to acclimatize to a broad range of host niches.

Transient silencing by transcription is mediated by chromatin complexes. In several organisms, genes on telomeres are subject to Telomere Positioning Effect (TPE), where a biphasic open or closed chromatin state at a given telomere, switches the gene expression status as ON or OFF (Gottschling et al. 1990). TPE is dependent on Sir2p in *S. cerevisiae*. Cell to cell variation in transcription/translation levels, termed as "gene expression noise" gives rise to phenotypic variation in an isogenic population. Noise can be extrinsic from different environment cues or intrinsic, arising from allele to allele variation (Raser and O'Shea 2005). Telomere

Table 12.2 Unique genomic features of *C. albicans*

1. It shows enormous genome plasticity—segmental duplication, trisomy, monosomy for most chromosomes. These features provide *C. albicans* a survival advantage in the host
2. An epigenetic mechanism for centromere formation is evident in *C. albicans*. All eight chromosomes possess unique *CEN* DNA sequences
3. Efficient neocentromere formation has been observed when a native centromere is deleted in *C. albicans*. Genomic mechanisms such as gene conversion prevent centromere repositioning
4. A unique interdependent concerted manner of kinetochore formation takes place in *C. albicans*. The kinetochore integrity protects CENP-A from proteasomal-mediated degradation
5. *C. albicans* contains unusually long telomeric repeats, the length of which can be varied depending upon the environmental/culture conditions
6. *C. albicans* TLOs provide a wide range of transcriptional flexibility to adapt to diverse host niches

proximal genes exhibit higher noise levels largely due to intrinsic noise that is dependent on genome position, or Telomere-Adjacent Gene Expression Noise (TAGEN) (Anderson et al. 2014). TAGEN generates expression variability due to local chromatin-mediated gene silencing. Likewise, *TLO* genes, when placed in the internal locations of a chromosome, were shown to exhibit reduced noise levels. In *C. albicans,* TAGEN is regulated in a Sir2-dependent manner, largely promoter independent and is tightly associated with telomere position effect dynamics. A similar effect of colony to colony variation due to noise was seen for one gene, *EPA1* in *C. glabrata* (Juarez-Reyes et al. 2012).

Hence, *C. albicans* is an ideal system to study telomere biology, owing to its highly plastic genome and adaptability to diverse environment niches. Components of the telomere machinery that have diverged evolutionarily in terms of domain structure and loss or emergence of key factors suggests that this organism has a wealth of genomic tools that is different from other budding yeasts (see Table 12.2).

References

Anderson MZ, Baller JA, Dulmage K, Wigen L, Berman J (2012) The three clades of the telomere-associated TLO gene family of Candida albicans have different splicing, localization, and expression features. Eukaryot Cell 11:1268–1275

Anderson MZ, Gerstein AC, Wigen L, Baller JA, Berman J (2014) Silencing is noisy: population and cell level noise in telomere-adjacent genes is dependent on telomere position and sir2. PLoS Genet 10:e1004436

PEARLS

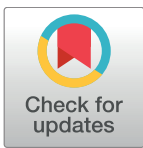# Five pillars of centromeric chromatin in fungal pathogens

Vikas Yadav, Lakshmi Sreekumar, Krishnendu Guin, Kaustuv Sanyal *

Molecular Mycology Laboratory, Molecular Biology and Genetics Unit, Jawaharlal Nehru Centre for Advanced Scientific Research, Jakkur, Bangalore, India

* sanyal@jncasr.ac.in

"*The greater the diversity, the greater the perfection.*"

—*Thomas Berry*

A centromere is classically defined as the primary constriction on a metaphase chromosome [1] that holds the sister chromatids together, binds to spindle microtubules, and brings about their separation during anaphase. Despite having a conserved and essential function, centromeres are among the fastest evolving DNA sequence loci in eukaryotic genomes [2]. With the advent of molecular biology techniques, centromeres could be mapped and sequenced in a large number of fungal species. The length of centromere DNA in fungi is found to be highly variable, classifying them as point ($<$400 bp), short regional ($>$400 bp, –20 kb), and large regional ($>$20 kb) [3]. Such diversity is achieved by different regulatory factors that have overlapping functions required for loading of the centromere-specific histone H3 variant centromere protein A/chromosome segregation 4(CENP-A/Cse4) to DNA to define centromere identity. Although genetic and epigenetic mechanisms of centromere formation across eukaryotes are largely conserved, there are examples of molecular innovation and genetic improvisation that help fungal species to maintain their ploidy across generations. In this review, we highlight five such genetic and epigenetic factors that define centromere identity in pathogenic fungi.

DNA sequence and organization of DNA sequence elements

DNA sequence features provide the necessary template to act as a binding platform for kinetochore proteins. The genus *Candida*, which harbors several pathogenic species, presents a diverse array of centromere types. *Candida glabrata* carries point genetic centromeres, much like the 125-bp DNA sequence that serves as a fully functional point centromere in the budding yeast *Saccharomyces cerevisiae* [3–5]. Typically, genetic centromeres have specific and conserved DNA sequence motifs and confer mitotic stability to otherwise unstable plasmids carrying an autonomous replicating sequence (ARS) during cell division. Despite high-structural homology in DNA sequence elements, the point centromeres of *C. glabrata* are not fully functional in *S. cerevisiae*, suggesting that centromere function is species-specific [5, 6]. Short regional genetic centromeres of *Candida tropicalis* comprise a central core flanked by inverted repeats, similar to those of the fission yeast *Schizosaccharomyces pombe* [3, 7]. The sequence and orientation of these repeats are important for centromere function. Due to the presence of inverted repeats, the centromeres in *C. tropicalis* can acquire a hairpin loop-like secondary structure that might be crucial for kinetochore assembly. *Candida albicans* and *Candida dubliniensis*, on the other hand, possess unique and different centromere DNA sequences on each of their chromosomes [8, 9]. While *C. tropicalis* centromeres can stabilize an ARS plasmid,

indicating a role of DNA sequence in centromere identity, the same does not hold true for
C. albicans. *Cryptococcus neoformans*, a basidiomycetous pathogen that diverged from *C. albicans* more than 900 million years ago, harbors large regional centromeres that are rich in centromere-specific retroelements [10–12]. While the presence of such retroelements hints
toward the functional dependence on DNA sequence in centromere function, more studies
are needed to explore such links. Some fungal centromeres possess specific DNA sequence features. For example, *Candida lusitaniae* (teleomorph *Clavispora lusitaniae*) and *Malassezia sympodialis* centromeres are present in highly AT-rich regions of the genome but lack any easily
detectable conserved sequence motifs or repeats [13, 14]. Whether any AT-rich DNA sequence
can act as a centromere in these organisms, similar to what is observed in diatoms [15],
remains an open question.

Centrochromatin—CENP-A and chromatin modifications

The conserved centromere-specific histone H3 variant CENP-A/Cse4 is specifically present at
all fungal centromeres identified to date but is largely excluded from other regions of the
genome. *Mucor circinelloides* and *Phycomyces blakesleeanus* are notable exceptions in this
regard, as they have no obvious CENP-A homologs, even though their centromeres are not yet
physically mapped [16]. CENP-A is considered as the epigenetic determinant of centromere
identity, as these molecules can seed the formation of a functional centromere in most organisms. This is supported by the fact that ectopic CENP-A incorporation can result in neocentromere formation, which is activated when an endogenous centromere becomes nonfunctional
[17]. Although it is not well understood how CENP-A acts as the epigenetic determinant of the
centromere, structural properties like a longer alpha N-terminal (αN) helix and the L1 loop
region and biophysical features of the CENP-A nucleosome array, such as higher condensation
properties, might be crucial for this role [18]. The process of CENP-A incorporation has been
studied in *C. albicans*. Like other species, the CENP-A chaperone Holliday junction recognition protein/suppressor of chromosome mis-segregation 3 (HJURP/Scm3) is found to be crucial for CENP-A loading in *C. albicans* (our unpublished results). Regional centromeres
harbor canonical histone H3 along with CENP-A. Post-translational modifications of histone
H3 are crucial in forming a functional kinetochore [19]. Biochemical studies revealed the
presence of heterochromatin histone marks such as demethylation of histone H3 lysine 9
(H3K9diMe) across the centromeres of *C. neoformans* [10]. Apart from histone marks, DNA
methylation has also been observed at the centromeres in *C. neoformans*, but its functional significance is unclear yet [10].

Transcription and RNAi

For a very long time, the centromere locus was considered heterochromatic and transcriptionally inert. While centromere regions are generally transcription poor, landmark studies in several yeast species revealed that small interfering RNAs (siRNAs) derived from pericentromeric
regions are necessary for centromere function [20, 21]. These studies indicated that centromere transcription is permissible and has a functional significance. A pan–fungal analysis
of RNA interference (RNAi) proteins revealed that a few species, including *C. glabrata* and
Ustilago maydis, have lost all of the proteins required for functional RNAi during the course of
evolution, whereas species including *C. albicans* and *C. tropicalis* harbor a cryptic RNAi
machinery [22]. A recent study in the pathogenic *Cryptococcus* species complex correlated the
loss of RNAi with the length of centromeres, thereby proposing that RNAi helps in maintaining long repetitive, transposon-rich centromeres [10]. Whether the RNAi machinery has a
functional significance in the centromere biology of this species complex and other fungal

pathogens remains unexplored. Apart from RNAi, centromere transcription can also play a functional role through long noncoding RNA (ncRNA). Indeed, pervasive levels of transcription have been documented in various fungal species [23, 24]. However, the transcripts generated from the centromeres are significantly low in number compared to the rest of the genome, as shown in the *Cryptococcus* and *Ustilago* species complex [10].

Replication and repair

Unlike metazoans, most fungal species have early replicating centromeres [25–27]. This temporally distinct replication timing not only allows better tolerance toward replication stress but also ensures proper kinetochore assembly at the centromeres. In *C. albicans*, centromeres replicate early in every synthesis phase (S-phase) and are associated with an early firing replicating origin [26]. Additionally, the formation of a neocentromere advances the replication time of the flanking region by activating an early replicating origin. This proximity effect was explained by a replication-coupled repair mechanism in a kinetochore-dependent manner [28]. Centromere-proximal origins stall randomly at the centromere, leading to accumulation of single-stranded DNA, which then recruits the homologous recombination proteins such as radiation sensitive (Rad)51 and Rad52. These proteins physically interact with CENP-A in *C. albicans* and load it at the site of stalled replication forks; that is, centromeres. How this process is regulated to occur only during S-phase remains unknown, with possible implications for the CENP-A chaperone Scm3. Based on studies in many other nonpathogenic fungal species, the physical proximity of a partitioning locus (centromere) and an initiator site (replication origin) is relevant when one dissects the functional aspects of genome maintenance. However, evidence toward this connection is just beginning to emerge.

Spatial location

Most fungal centromeres are clustered near spindle pole bodies (SPBs). This association may result in folding back of chromosomes and positioning them such that telomeres are juxtaposed in the interphase nucleus, giving rise to the Rabl conformation [29]. This phenomenon has been shown to occur in both animal and plant pathogens including *C. albicans*, *C. tropicalis*, and *Fusarium graminearum*. Using chromosome conformation capture (3C) experiments, clustered centromere DNA regions were shown to be present in close spatial proximity, leading to physical interactions between different centromeres [30–32]. It has been proposed that the clustering of centromeres aids in determining the site of centromere formation in these organisms. According to this hypothesis, a part of the nucleus is enriched with a pool of CENP-A proteins to form a CENP-A–rich zone or CENP-A cloud [33, 34]. It was proposed that the region of a chromosome that is near this CENP-A cloud would attract a higher level of CENP-A and thus serves as a preferred site for centromere formation. In *S. cerevisiae*, for example, a locally enriched population of accessory CENP-A molecules at pericentric chromatin has been shown to serve as a reservoir for rapid incorporation of CENP-A in the event of premature eviction from centromeres [35]. Further evidence supporting the CENP-A cloud hypothesis stems from studies in *C. albicans* in which neocentromeres were formed close to the native centromere in most cases [34]. In addition, neocentromeres in *C. albicans* change the spatial location to be a part of the centromere cluster by 3C experiments [31]. Interestingly, centromeres were found to be unclustered in premitotic *C. neoformans* cells that eventually cluster at the onset of mitosis [36]. Whether this centromere clustering also arises as a result of physical interactions among centromeres is not yet known.

Overall, here we summarize five key determinants among many that are needed for centromere identity in fungal pathogens (Fig 1). These factors may work sequentially or in parallel to
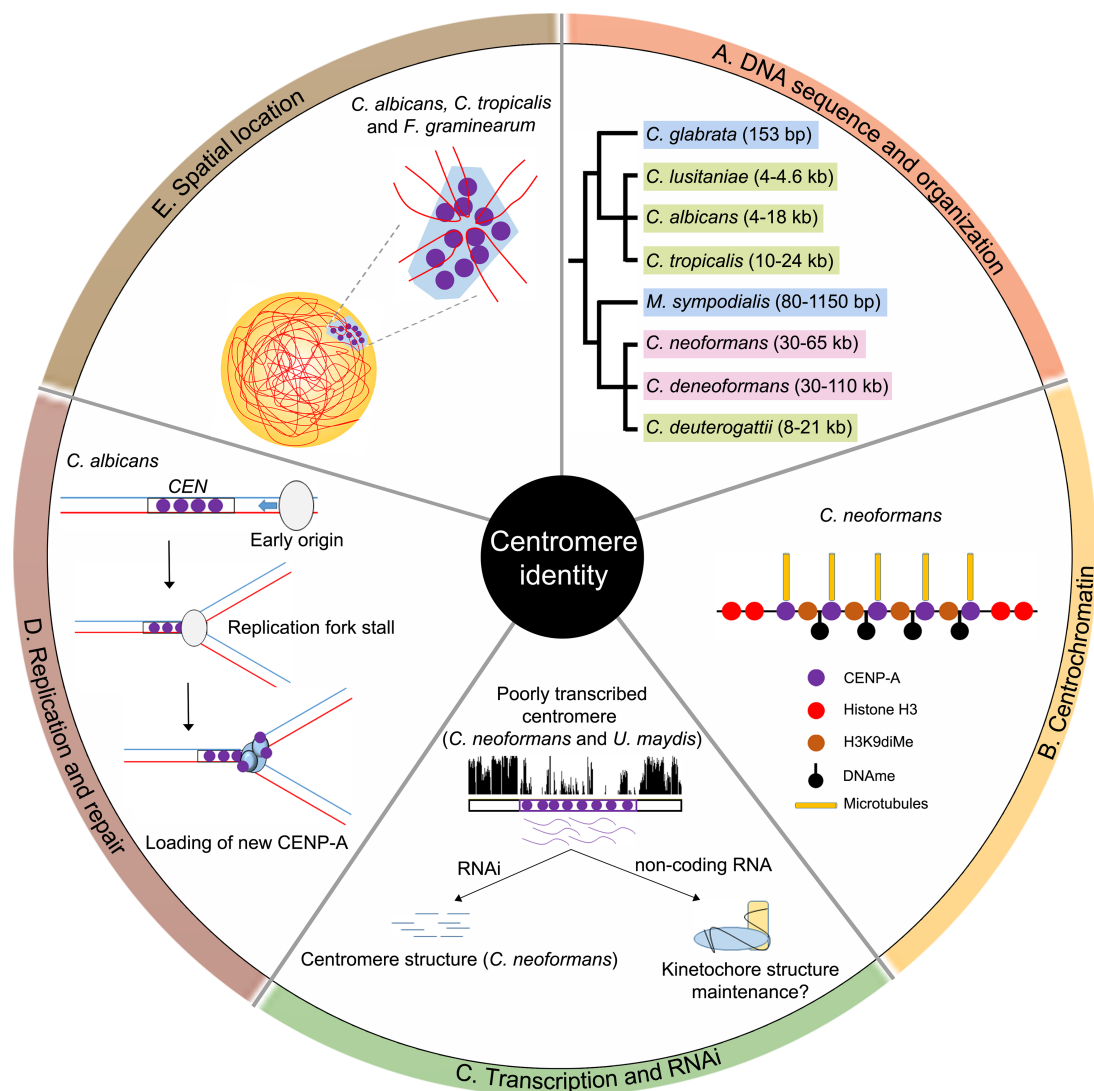
Fig 1. Five key determinants of centromere identity in pathogenic fungi. (A) The length and type of centromeres in various pathogenic fungi have been depicted in a cladogram. *Candida glabrata* and *Malassezia sympodialis* have point centromeres (blue); *C. albicans*, *C. tropicalis*, and *Cryptococcus deuterogattii* have small regional centromeres (green); and *C. neoformans* and *C. deneoformans* harbor large regional repetitive centromeres (pink). (B) Centromeric heterochromatin or centrochromatin refers to the various chemical modifications associated with the histones (both canonical and variant) as well as DNA at the centromeric locus. For example, centromere DNA is methylated at cytosines, and CENP-A nucleosomes are interspersed with H3K9-dimethylated nucleosomes in *C. neoformans*. (C) Transcription and a functional RNAi machinery at the centromere are required to preserve centromere identity in some organisms. The centromeres in the *Ustilago* and *Cryptococcus* species complexes are poorly transcribed. In *C. neoformans*, centromeric transcription is regulated by the RNAi machinery that silences centromeric retrotransposons to stabilize centromere structure. Unprocessed long noncoding RNA, whose function is unknown, is also produced from centromeres in this species. (D) DNA replication and repair proteins are known to play a key role in *C. albicans* centromere stability. The centromere proximal origins help in maintaining centromere function in this organism. Replication forks converging toward the centromere stall randomly, which leads to accumulation of single-stranded DNA (ssDNA) that recruits repair proteins like Rad51 and Rad52, along with new CENP-A molecules in *C. albicans*. (E) The spatial location of centromeres within the nucleus determines its activity and interaction with other nuclear subcompartments. All centromeres are clustered toward the nuclear periphery to form a CENP-A–rich zone in *C. albicans* and many other pathogenic fungi. This preferential spatial distribution helps to determine the site of centromere assembly in every cell cycle. CENP-A, centromere protein A; H3K9, histone H3 lysine 9; Rad, radiation sensitive; RNAi, RNA interference; ssDNA, single-stranded DNA.

https://doi.org/10.1371/journal.ppat.1007150.g001

ensure that centromere identity is maintained at all times in every cell cycle. For example, the repeats present at the centromere can form secondary structures, which may lead to double-stranded DNA breaks during replication and attract more CENP-A to these regions through DNA-repair pathways. The nuclear subdomain near SPBs is usually heterochromatic in nature, and the presence of the centromere cluster at a nuclear peripheral region contributes to poor transcription. Thus, a combination of factors can lead to more efficient CENP-A incorporation at the centromere. The CENP-A incorporation affects assembly of kinetochore proteins; for example, in *C. neoformans*, CENP-A assembly initiates the loading of other proteins, giving rise to a sequential kinetochore assembly in this organism [36]. On the contrary, *C. albicans* shows an interdependent kinetochore formation in which probably all kinetochore proteins assemble as a single complex [37]. However, our knowledge of the underlying mechanisms of CENP-A loading is still at its infancy, given the number of fungal species known to exist. The lack of knowledge is primarily due to difficulties associated with genetic manipulation of most pathogenic fungal species. Probing into the molecular mechanisms of centromere identity using diverse fungal species will yield significant insight into the structure-function-evolution of centromeres during speciation. Studies in *Candida* and *Cryptococcus* species complexes revealed that centromere-mediated recombination might have contributed to variations in the centromere structure and sequence [7, 10, 38]. Additionally, centromere sequences can also help in identification and classification of two or more closely related species. For example, *C. albicans* and *C. dubliniensis* centromere sequences show a high level of divergence while the rest of the genome sequences are highly homologous [9]. Discovery of highly efficient gene-targeting technologies such as the Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR-Cas9) system for various *Candida* and *Cryptococcus* species as well as the application of efficient and improved DNA sequencing technologies for obtaining better genome assemblies are some of the significant recent developments that will advance pathogenic fungal molecular genetic research.

Acknowledgments

References

1. Flemming W. Zellsubstanz, Kern und Zelltheilung: F. C. W. Vogel, Leipzig; 1882.

2. 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. https://doi.org/10.1126/science.1062939 PMID: 11498581.

3. 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. https://doi.org/10.1128/EC.05165-11 PMID: 21908596.

4. Clarke L, Carbon J. Isolation of a yeast centromere and construction of functional small circular chromosomes. Nature. 1980; 287(5782):504–9. Epub 1980/10/09. PMID: 6999364.

5. Kitada K, Yamaguchi E, Arisawa M. Isolation of a *Candida glabrata* centromere and its use in construction of plasmid vectors. Gene. 1996; 175(1–2):105–8. PMID: 8917084.

6. Kitada K, Yamaguchi E, Hamada K, Arisawa M. Structural analysis of a *Candida glabrata* centromere and its functional homology to the *Saccharomyces cerevisiae* centromere. Current genetics. 1997; 31 (2):122–7. PMID: 9021128.

7. 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. https://doi.org/10.1371/journal.pgen.1005839 PMID: 26845548.

8. 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. https://doi.org/10.1073/pnas.0404318101 PMID: 15272074

9. 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. https://doi.org/10.1073/pnas.0809770105 PMID: 19060206.

10. 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. https://doi.org/10.1073/pnas.1713725115 PMID: 29507212.

11. Janbon G, Ormerod KL, Paulet D, Byrnes EJ 3rd, Yadav V, Chatterjee G, et al. Analysis of the genome and transcriptome of *Cryptococcus neoformans* var. *grubii* reveals complex RNA expression and micro-evolution leading to virulence attenuation. PLoS Genet. 2014; 10(4):e1004261. Epub 2014/04/20. https://doi.org/10.1371/journal.pgen.1004261 PMID: 24743168.

12. Loftus BJ, Fung E, Roncaglia P, Rowley D, Amedeo P, Bruno D, et al. The genome of the basidiomycetous yeast and human pathogen *Cryptococcus neoformans*. Science. 2005; 307(5713):1321–4. https://doi.org/10.1126/science.1103773 PMID: 15653466.

13. 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. https://doi.org/10.1073/pnas.1508749112 PMID: 26371315.

14. 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. https://doi.org/10.1093/nar/gkx006 PMID: 28100699.

15. 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. https://doi.org/10.1073/pnas.1700764114 PMID: 28673987.

16. van Hooff JJ, Tromer E, van Wijk LM, Snel B, Kops GJ. Evolutionary dynamics of the kinetochore network in eukaryotes as revealed by comparative genomics. EMBO Rep. 2017; 18(9):1559–71. https://doi.org/10.15252/embr.201744102 PMID: 28642229.

17. Steiner FA, Henikoff S. Diversity in the organization of centromeric chromatin. Curr Opin Genet Dev. 2015; 31:28–35. https://doi.org/10.1016/j.gde.2015.03.010 PMID: 25956076.

18. Bui M, Walkiewicz MP, Dimitriadis EK, Dalal Y. The CENP-A nucleosome: a battle between Dr Jekyll and Mr Hyde. Nucleus. 2013; 4(1):37–42. https://doi.org/10.4161/nucl.23588 PMID: 23324462.

19. Fukagawa T. Critical histone post-translational modifications for centromere function and propagation. Cell Cycle. 2017:1–7. https://doi.org/10.1080/15384101.2017.1325044 PMID: 28598241.

20. 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. https://doi.org/10.1098/rstb.2004.1611 PMID: 15905142.

21. Volpe T, Schramke V, Hamilton GL, White SA, Teng G, Martienssen RA, et al. RNA interference is required for normal centromere function in fission yeast. Chromosome Res. 2003; 11(2):137–46. PMID: 12733640.

22. Nakayashiki H, Kadotani N, Mayama S. Evolution and diversification of RNA silencing proteins in fungi. J Mol Evol. 2006; 63(1):127–35. https://doi.org/10.1007/s00239-005-0257-2 PMID: 16786437.

23. Choi ES, Stralfors A, Castillo AG, Durand-Dubief M, Ekwall K, Allshire RC. Identification of noncoding transcripts from within CENP-A chromatin at fission yeast centromeres. J Biol Chem. 2011; 286 (26):23600–7. https://doi.org/10.1074/jbc.M111.228510 PMID: 21531710.

24. Ohkuni K, Kitagawa K. Endogenous transcription at the centromere facilitates centromere activity in budding yeast. Curr Biol. 2011; 21(20):1695–703. https://doi.org/10.1016/j.cub.2011.08.056 PMID: 22000103.

25. Kim SM, Dubey DD, Huberman JA. Early-replicating heterochromatin. Genes Dev. 2003; 17(3):330–5. https://doi.org/10.1101/gad.1046203 PMID: 12569122.

26. 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. https://doi.org/10.1371/journal.pgen.1001068 PMID: 20808889.

27. 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. https://doi.org/10.1126/science.294.5540.115 PMID: 11588253.

28. 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. https://doi.org/10.1371/journal.pgen.1004344 PMID: 24762765.

29. Rabl C. Ueber Zelltheilung. Morphologisches Jahrbuch. 1885; 10:214–330.

30. Duan Z, Andronescu M, Schutz K, McIlwain S, Kim YJ, Lee C, et al. A three-dimensional model of the yeast genome. Nature. 2010; 465(7296):363–7. Epub 2010/05/04. https://doi.org/10.1038/nature08973 PMID: 20436457.

31. Burrack LS, Hutton HF, Matter KJ, Clancey SA, Liachko I, Plemmons AE, et al. Neocentromeres provide chromosome segregation accuracy and centromere clustering to multiple loci along a *Candida albicans* chromosome. PLoS Genet. 2016; 12(9):e1006317. https://doi.org/10.1371/journal.pgen.1006317 PMID: 27662467.

32. 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. https://doi.org/10.1091/mbc.E09-05-0359 PMID: 19656849.

33. Fukagawa T, Earnshaw WC. The centromere: chromatin foundation for the kinetochore machinery. Dev Cell. 2014; 30(5):496–508. https://doi.org/10.1016/j.devcel.2014.08.016 PMID: 25203206.

34. 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. https://doi.org/10.1101/gr.141614.112 PMID: 23439889.

35. Haase J, Mishra PK, Stephens A, Haggerty R, Quammen C, Taylor RM 2nd, et al. A 3D map of the yeast kinetochore reveals the presence of core and accessory centromere-specific histone. Curr Biol. 2013; 23(19):1939–44. Epub 2013/10/01. https://doi.org/10.1016/j.cub.2013.07.083 PMID: 24076245.

36. 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. https://doi.org/10.1128/mBio.00614-13 PMID: 24085781.

37. 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. https://doi.org/10.1371/journal.pgen.1002661 PMID: 22536162.

38. 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. https://doi.org/10.1371/journal.pbio.2002527 PMID: 28800596.

1   **DNA replication initiator proteins facilitate CENPA loading on early replicating**

2   **compact chromatin**

3

4   Lakshmi Sreekumar[1], Priya Jaitly[1], Yao Chen[2], Bhagya C. Thimmappa[1], Amartya Sanyal[2], Leelavati

5   Narlikar[3], Rahul Siddharthan[4], Kaustuv Sanyal[1]

6

7   [1]Molecular Biology and Genetics Unit, Jawaharlal Nehru Centre for Advanced Scientific Research,

8   Bangalore, India; [2] School of Biological Sciences, Nanyang Technological University, 60 Nanyang

9   Drive, Singapore 637551, [3]Department of Chemical Engineering, CSIR-National Chemical

10   Laboratory, Pune 411008, India; [4]The Institute of Mathematical Sciences/HBNI, Taramani, Chennai

11   600113, India

12

13

14   *corresponding author

15   Kaustuv Sanyal

16   Molecular Biology & Genetics Unit

17   Jawaharlal Nehru Centre for Advanced Scientific Research

18   Jakkur, Bangalore - 560064

19   India

20   Email: sanyal@jncasr.ac.in

21   Telephone : +91-80-2208 2878

22   Fax : +91-80-2208 2766

23   Homepage: http://www.jncasr.ac.in/sanyal

24

25   Present address: Bhagya C. Thimmappa, Department of Biochemistry, Robert-Cedergren Centre for

26   Bioinformatics and Genomics, University of Montreal, 2900 Edouard-Montpetit, Montreal, H3T1J4,

27   QC, Canada

28

29

31   Running title: Inheritance of early replicating centromeric chromatin

32

33

1   **Abstract (250 words)**

2

3   The process of centromere formation enables the cell to conserve established genetic and epigenetic

4   information from the previous cell cycle and reuse it for future episodes of chromosome segregation.

5   CENPA asserts the role of an epigenetic requirement in maintaining active centromeres. Active

6   centromeres are subject to position effects which can cause its site of assembly to drift occasionally.

7   Determinants of neocentromere formation, when a native centromere is inactivated, remain elusive.

8   To dissect factors for centromere/neocentromere formation, here, we employed the budding yeast

9   *Candida albicans*, whose centromeres have unique and different DNA sequences, and exhibit

10   classical epigenetic regulation. We used CENPA-mediated reversible silencing of a marker gene,

11   *URA3,* as an assay to select cells with ectopic centromeres. We defined pericentric boundaries for *C.*

12   *albicans* centromeres by Hi-C analysis and these were located in early replicating domains. The

13   pericentric boundaries primed with CENPA served as sites of neocentromere formation in isolates

14   with ectopic centromeres, indicating that the number of non-centromeric CENPA molecules

15   determines neocentromere location. To understand the importance of early replication timing of

16   centromeres, we identified genome-wide binding sites of the Origin Recognition Complex subunit,

17   Orc4. A fraction of these Orc4 enriched regions located within tDNA, cluster towards early

18   replicating regions, and frequently interact among themselves than the late replicating regions,

19   demonstrating the spatiotemporal distribution of these regions. Strikingly, Orc4 is highly enriched at

20   centromeres of *C. albicans* and along with the helicase component Mcm2, stabilizes the kinetochore,

21   suggesting a role of pre-replication complex proteins as epigenetic determinants of centromere

22   identity.

23

24

25

26

1 **Introduction**

2 The centromeric histone H3 variant CENPA is known to assemble on various types of DNA

3 sequences although their position is predominantly confined to centromeres, chromosomal regions

4 responsible for faithful chromosome segregation. CENPA is an adaptor molecule between rapidly

5 diverged centromere (CEN) DNA sequences and the less diverged kinetochore machinery (Mellone

6 and Allshire 2003, Sullivan, Maloney et al. 2016). Centromere assembly requires the establishment of

7 centromeric "memory" by incorporation of CENPA into underlying chromatin, and its subsequent

8 stabilization by binding of kinetochore proteins (McNulty and Sullivan 2017). Extensive studies in

9 higher eukaryotes suggest that there are potential sites on a chromosome which are capable of

10 harboring a functional centromere, however they are kept dormant by the more *cis*-acting dominant

11 centromere (Amor and Choo 2002). This genetically well-defined locus is known to "drift" along the

12 length of a chromosome making them malleable structures. Vertebrate cell lines upon prolonged

13 culturing are subject to "centromere drift" (Hori, Kagawa et al. 2017) similar to the repeat-associated

14 regional centromeres of *Schizosaccharomyces pombe* that exhibit stochastic repositioning of CENPA

15 within CEN chromatin as a consequence of an oversized centromeric core (Yao, Liu et al. 2013). The

16 existence of such plasticity in CEN chromatin indicates that centromeres in vertebrates to the

17 unicellular fission yeast are specified and propagated by sequence independent mechanisms.

18 The episodic occurrence of centromere activity at non-centromere sequences, neocentromeres,

19 strongly suggests the epigenetic nature of centromeres. First observed in humans to rescue acentric

20 fragments (Voullaire, Slater et al. 1993), neocentromeres across species share common features as

21 well as certain species-specific attributes. *S. pombe* forms sub-telomeric neocentromeres (Ishii,

22 Ogiyama et al. 2008) whereas in humans they are more prevalent at sub-metacentric regions

23 (Warburton 2004). In contrast, most neocentromeres have been detected at CEN proximal loci in flies

24 (Maggert and Karpen 2001), and chicken (Shang, Hori et al. 2013). The assembly of ectopic CENPA

25 as a "CENPA-rich zone" surrounding the endogenous CEN and proximity of neocentromere hotspots

26 to native CEN in these organisms indicates that CENPA is peppered on CEN adjacent loci and can get

27 rapidly incorporated to the centromere in case of CENPA eviction (Haase, Mishra et al. 2013,

3

1   Fukagawa and Earnshaw 2014). Apart from location of CENPA chromatin, transcription plays an

2   important role in specifying CEN identity and maintenance. Centromeres are known to be "difficult to

3   transcribe" regions. However, pervasive level of transcription helps in CEN function as studied

4   extensively in *S. pombe* centromeres (Choi, Stralfors et al. 2011). Silencing of pericentromeric

5   heterochromatin by the RNAi-mediated pathway helps to create a microenvironment for CENPA

6   loading at the *S. pombe* central core (Allshire and Ekwall 2015, Catania, Pidoux et al. 2015). Even in

7   humans, the transcripts generated from the higher order repeats (HOR) of the alpha-satellite DNA

8   interact with CENPA, rendering structural stability to CEN chromatin (McNulty, Sullivan et al. 2017).

9   A recent study in *Drosophila* elucidates that a transcription-coupled remodeling is required for

10   CENPA incorporation (Bobkov, Gilbert et al. 2018). This reemphasizes the role of regulated

11   transcription to maintain centromere structure and function.

12   Non-repetitive centromeres provide an excellent model to study characterization of centromeric

13   chromatin. In the ascomycetous budding yeast *Candida albicans,* the presence of unique and different

14   CEN sequences on every chromosome (Sanyal, Baum et al. 2004) and the activation of

15   neocentromeres at pre-determined hotspots proximal to the native CEN location (Thakur and Sanyal

16   2013) together provide evidence that the underlying DNA sequence is neither necessary nor sufficient

17   for centromere formation (Sanyal, Baum et al. 2004, Baum, Sanyal et al. 2006, Thakur and Sanyal

18   2013). CENPA localization on a transgene under selective conditions is known to correspond to its

19   transcriptional status. Similar to *S. pombe* (Allshire, Javerzat et al. 1994), reversible silencing of the

20   expression of a marker gene, *URA3,* captured by 5'FOA counter-selection, has been observed upon its

21   integration at the CENPA binding region of the centromere in *C. albicans* endowing it a

22   transcriptionally flexible status (Thakur and Sanyal 2013).

23   For propagation of CEN chromatin, DNA replication ensures accurate assembly of centromeric

24   nucleosomes in a cell cycle specific manner. Replication origins are marked by the physical

25   association of the pre-replication complex (pre-RC) comprising of the hexameric origin recognition

26   complex (Orc1-6), minichromosome maintenance complex (Mcm2-7) and accessory proteins

27   (Leonard and Mechali 2013). These initiator complexes occupy discrete sites on a chromosome and

1   are temporally regulated to ensure complete genome duplication in the S phase. In the budding yeast

2   *Saccharomyces cerevisiae*, approximately 400 ORC binding sites have been identified, but only a

3   subset of them 'fire' at a given time (Wyrick, Aparicio et al. 2001, Nieduszynski, Knox et al. 2006).

4   This implies that not all ORC binding sites act as functional DNA replication origins in each cell

5   cycle. In most eukaryotes, replication origins are defined more flexibly as they rely very little on a

6   DNA sequence requirement for origin specification (Parker, Botchan et al. 2017). Based on the

7   presence of active firing and passive dormant origins, the genome can be classified into early, mid and

8   late replicating regions (Yamazaki, Hayano et al. 2013). Centromeres and replication origins are often

9   seen to be juxtaposed to each other from bacteria like *Bacillus subtilis* (Livny, Yamaichi et al. 2007)

10 to yeast species like *Yarrowia lipolytica* (Vernis, Abbas et al. 1997). This physical proximity aids in

11 centromere cohesion as well as ensures proper kinetochore assembly (Natsume, Muller et al. 2013).

12 Additionally, CEN replication timing is pivotal in CENPA loading, where early replication of CENs

13 ensures replication coupled loading of CENPA in *S. cerevisiae* (Pearson, Yeh et al. 2004). One of the

14 mechanisms for the early replication of CENs in budding yeast is mediated by the timely recruitment

15 of Dbf4-dependent kinase (DDK) at kinetochores with the help of the Ctf19 complex, which loads

16 replication initiator proteins to pericentromeric replication origins (Natsume, Muller et al. 2013).

17 Also, replication fork termination at the centromere promotes centromere DNA loop formation and

18 this is required for kinetochore assembly (Cook, Bennett et al. 2018). Centromeres are the earliest to

19 replicate in every S-phase of the *C. albicans* cell cycle (Koren, Tsai et al. 2010) by virtue of the early

20 replicating origins flanking the centromere (Mitra, Gomez-Raja et al. 2014). Deletion of CEN

21 proximal origins is also known to abrogate centromere function and debilitate kinetochore stability in

22 this organism (Mitra, Gomez-Raja et al. 2014). Hence, there is an intimate crosstalk of replication

23 origins, initiator proteins and kinetochore components in maintaining genome stability.

24 *C. albicans* centromeres are defined on the basis of a CENPA binding region which spans a 3-5 kb

25 unique DNA sequence on every chromosome (Sanyal, Baum et al. 2004). There is no functional

26 evidence for the existence of a pericentromeric boundary region to restrict CENPA spreading in *C.*

27 *albicans*, as seen in case of fission yeast centromeres (Karpen and Allshire 1997, Allshire and Ekwall

1   2015, Allshire and Madhani 2018). Unlike *S. pombe*, the genome of *C. albicans* does not encode an

2   HP1/Swi6-like protein, a methyl transferase like Clr4 required for H3K9me2 and components of a

3   fully functional RNAi machinery (Freire-Beneitez, Price et al. 2016). In this study, we defined the

4   pericentric boundaries of *C. albicans* by Hi-C analysis and a transgene silencing assay. A CENPA-

5   primed region within this pericentric boundary is found to serve as the neocentromere hotspot. By

6   identifying genome-wide binding sites of Orc4, we show that the pericentric regions lie in early

7   replicating highly interacting compact CEN-adjacent regions. We observe a strong physical

8   association of Orc4 to native and neocentromeres in *C. albicans.* The absence of Orc4 compromised

9   kinetochore integrity, a phenotype that we also observed upon depletion of another pre-RC

10 component, Mcm2. Thus, the genetic interaction between CENPA, Orc4 and Mcm2 revealed a

11 previously unidentified role of pre-RC components in maintaining active centromeres in this

12 pathogenic yeast.

13

14 **Results**

15 **Core CENPA-rich regions in *C. albicans* are flanked by a ~25 kb long unusual pericentric**

16 **heterochromatin**

17 The centromere DNA spans a region of 3-5 kb in *C. albicans* (Sanyal and Carbon 2002, Sanyal, Baum

18 et al. 2004) bound by the CENPA homolog, Cse4 (Sanyal and Carbon 2002). The presence of

19 replication origins and neocentromere hotspots within 30 kb of centromere 7, *CEN7* indicates that

20 CEN proximal regions are important hubs that regulate centromere activity (Thakur and Sanyal 2013,

21 Mitra, Gomez-Raja et al. 2014). We analyzed the Hi-C data of *C. albicans* from a previous report

22 (Burrack, Hutton et al. 2016). Our analysis revealed that all the centromeres interacted with adjacent

23 "pericentric" regions at a higher probability than regions distal from the centromere (Fig 1A). Also,

24 the clustered centromeres of *C. albicans* interact both in *cis* (with the pericentric regions) (Fig. 1A)

25 and in *trans* (with other centromeres) (Supplemental fig. S1A) at a higher probability forming a

26 compact chromatin environment than the average genome interaction found in bulk chromatin

1   (Supplemental fig. 1B). Upon examining the intra-chromosomal interactions, we observed a 25 kb

2   region centring on *CEN7* that closely interacts with the CENPA bound CEN mid-core (Fig 1B). To

3   gain further insights into the pericentric regions, we sought to examine the transcriptional status of

4   CEN-adjacent regions. We inserted the 1.4 kb *URA3* gene at ten independent *CEN7*-proximal loci

5   (Fig. 1C) (see Supplemental table S1 for location of insertions) in a strain that has two differentially

6   marked arms of Chr7, J200 (Sanyal, Baum et al. 2004). We also performed integrations at a *CEN7*-

7   distal locus and a *CEN5*-proximal locus (Supplemental table S1). We plated approximately 1 million

8   cells of each *URA3* integrant type on CM+5'FOA and replica plated 100 colonies resistant to 5'FOA

9   on CM-Uri to obtain the rate of *URA3* silencing (Fig 1D). We also monitored the frequency of

10   chromosome loss in these strains by examining the simultaneous loss of two markers, *ARG4* and

11   *URA3* or *HIS1* and *URA3* (Supplemental table S2). We observed a steep decline in the percentage of

12   reversible silencing of *URA3* (the ratio of the number of 5'FOA resistant colonies that grew on CM-

13   Uri and the total number of 5'FOA resistant colonies analysed) from the *CEN7* core to the periphery.

14   *URA3* when inserted at *CEN7* core exhibited a significantly higher rate of silencing than the

15   peripheral insertions (Fig. 1E, see Supplemental table S2). The clear trend of exponential decay in

16   reversible silencing of *URA3*, correlated with contact probabilities made by the central core to the

17   neighbouring regions, indicating that the clustered centromeres of *C. albicans* interact with pericentric

18   regions to form a compact nuclear subdomain (up to 25 kb), the frequency of which is ablated at loci

19   distal to the central core (Supplemental figs. S1C, 1D).

20   **Transgene silencing at the pericentromeres is associated with a transient ectopic kinetochore**

21   Transcriptional silencing of *URA3* at the *C. albicans* CEN core is known to facilitate CENPA binding

22   (Thakur and Sanyal 2013). We wanted to examine the consequence of *URA3* silencing in these

23   pericentromeric insertions. ChIP experiments on the 5'FOA resistant colonies revealed that *URA3* is

24   significantly enriched with CENPA when cells were grown in CM+5'FOA than CM-Uri indicating

25   that transcriptional repression of *URA3* at pericentromeres favors CENPA binding in all the *URA3*

26   insertions which yielded the 5'FOA resistant colonies (Fig. 2 (top panel), Supplemental Figs. S2A,

27   S2C). We did not observe this phenomenon in the far-*CEN7* integrant (Supplemental Fig. S2B). We

7

1   expressed Protein A-tagged Mtw1 (Supplemental fig. S3A), the Mis12 homolog in *C. albicans,* (Roy,

2   Burrack et al. 2011) and detected its significant enrichment on *URA3* in LSK437 (4L) and LSK440

3   (4R) indicating that *URA3* can form an ectopic centromere (ecCEN) when minimally transcribed (Fig.

4   2, bottom panel). The overlapping CENPA and Mtw1 binding regions were limited to the repressed

5   *URA3* locus and did not extend to regions beyond it (Supplemental Figs. S3B, S3C).

6   We further wanted to determine whether ecCEN can be stably propagated through mitosis by

7   withdrawing the selection. We serially passaged the initial 5'FOA resistant colonies from LSK404

8   *(4L/4L::URA3)* and LSK425 *(4R/4R::URA3)* in non-selective media (YPDU) for up to 20 generations

9   (Supplemental Fig. S4A) (see Supplemental methods). We observed a gradual decline in the relative

10 enrichment of CENPA at *URA3* with every doubling and after ~20 mitoses, the CENPA level was

11 comparable to a state when cells were forced to express *URA3* (in CM-Uri) (Supplemental Fig. S4B).

12 Additionally, we observed that if at any stage of passaging in non-selective media (YPDU), these cells

13 were regrown in presence of selection (CM+5'FOA), they could reassemble the CENPA associated

14 ecCEN on *URA3* (Supplemental Fig. S4C). Thus, transcriptional repression of a transgene within the

15 25 kb compact pericentromeric region favors the formation of a transient ectopic kinetochore.

16 **Pre-existing CENPA molecules can prime a chromosomal location to form neocentromeres**

17 Neocentromeres provide a way to study *de novo* centromere formation since they recapitulate all

18 molecular events for centromere assembly under natural conditions on a non-native locus (Amor and

19 Choo 2002, Craig, Wong et al. 2003, Marshall, Chueh et al. 2008). In *C. albicans*, neocentromeres are

20 shown to get activated at CEN-proximal loci irrespective of the length of the centromere DNA deleted

21 (Thakur and Sanyal 2013). This prompted us to examine that in the event of a centromere deletion

22 whether a cell would prefer to form a neocentromere on a pre-determined hotspot or on a CENPA-

23 primed region located at the pericentric region. To address the same, we replaced the core 4.5 kb

24 CENPA-rich *CEN7* region (Ca21Chr7 424475-428994) with the 1.2 kb *HIS1* sequence independently

25 in two 5'FOA resistant strains, LSK443 (*4L/4L::URA3*) and LSK456 (*4R/4R::URA3*). We screened

26 for colonies where *URA3* and *HIS1* were located on the same homolog (in *cis*) using Southern

27 hybridization (Supplemental Figs. S5A,5B) (Supplemental table. S3) and obtained multiple

1   transformants. We performed the same deletion in the corresponding 5'FOA sensitive *URA3*

2   integrants and examined whether a CENPA primed region could assemble a functional kinetochore.

3   ChIP-qPCR analysis in the 5'FOA resistant strain LSK465 (*4R/4R::URA3 CEN7/CEN7::HIS1*) (Fig

4   3A, 3B) and LSK450 (*4L/4L::URA3 CEN7/CEN7::HIS1*) (Supplemental Figs. S7A, S7B) revealed

5   that two independent kinetochore proteins, CENPA and Mtw1, assemble at *URA3* and neighboring

6   regions, apart from the native centromere. We confirmed neocentromere formation on this altered

7   chromosome by CENPA ChIP-sequencing (Supplemental fig. S6), which revealed two new hotspots

8   at CENPA-primed regions, *URA3nCEN7-I* and *URA3nCEN7-II* (Fig 3C, Supplemental Fig. S6C) (see

9   Supplemental table S4 for neocentromere coordinates). On the other hand, in the 5'FOA sensitive

10 strains, neocentromeres formed at one of the pre-determined hotspots, *nCEN7-II* (Supplemental Fig

11 7D). This alludes to the fact that an initial targeting of CENPA to a primed locus within the 25 kb

12 compact region can render centromeric properties to that site if CENPA can enable its nucleation and

13 the subsequent assembly of a functional kinetochore independent of selection or any other target

14 mechanisms.

15 **Orc4 binds to discrete regions uniformly across the *C. albicans* genome to ensure efficient**

16 **completion of DNA replication in S phase**

17 Nuclear organization is important to study replication architecture. The mitotic propagation of

18 epigenetic marks is ensured by timely replication of the genome. However, to decipher the same, the

19 precise location and timing of replication origins is pivotal. We utilized the binding of an

20 evolutionarily conserved replication initiator protein, Orc4 to map putative replication origins in *C.*

21 *albicans.* Orc4 in *C. albicans* is a 564-aa long protein (https://doi.org/10.1101/430892) that contains

22 the AAA+ domain which belongs to the AAA+ family of ATPases (Walker, Saraste et al. 1982)

23 associated with a variety of cellular activities (Supplemental fig. 8A). We raised polyclonal antibodies

24 against a peptide sequence from the N-terminus of the native Orc4 (aa 20-33) (Supplemental fig. 8B)

25 of *C. albicans* (see Supplemental methods). Western blot with the whole cell extract of *C. albicans*

26 SC5314 *(ORC4/ORC4)* yielded a strong specific band at the expected molecular weight of

27 approximately 64 kDa when probed with purified anti-Orc4 antibodies (Supplemental fig. 8C).

1   Indirect immuno-fluorescence microscopy using anti-Orc4 antibodies revealed that Orc4 was found to

2   be strictly localized to the nucleus at all stages of the *C. albicans* cell cycle (Fig. 4A), a feature of the

3   ORC proteins found to be conserved in *S. cerevisiae* as well (Dutta and Bell 1997).

4   Orc4 is an evolutionarily conserved essential subunit of the origin recognition complex (ORC) across

5   eukaryotes (Chuang and Kelly 1999, Dai, Chuang et al. 2005). A conditional mutant of *orc4*

6   constructed by deleting one allele and replacing the endogenous promoter of the remaining *ORC4*

7   allele with the repressive *MET3* promoter (Care, Trevethick et al. 1999), showed growth impairment

8   of *C. albicans* cells (Fig. 4B). Hence, Orc4 is essential for viability in *C. albicans* as well. We

9   confirmed the depletion of Orc4 protein levels from the cellular pool by performing a western blot

10   analysis in the Orc4 repressed versus expressed conditions (Supplemental fig. 8D). Subsequently, we

11   used the purified anti-Orc4 antibodies as a tool to map its binding sites in the *C. albicans* genome.

12   ChIP sequencing in asynchronously grown cells of *C. albicans* using anti-Orc4 antibodies yielded a

13   total of 417 discrete Orc4 binding sites with 414 of these belonging to various genomic loci (Fig. 4C,

14   Supplemental fig. S8E) while the remaining three mapped to mitochondrial DNA. We validated one

15   region on each of the eight chromosomes by ChIP-qPCR (Supplemental fig. S8F). Strikingly, all

16   centromeres were found to be highly enriched with Orc4 (Supplemental fig. S8E). The length of Orc4

17   binding regions across the genome ranged from 200 bp to ~3 kb. Approximately 61% of the Orc4

18   binding regions in our study were present in genic regions (252/414) in *C. albicans* deviating from the

19   trend observed in *S. cerevisiae* where most of the chromosomal origins are located at intergenic

20   regions (Xu, Aparicio et al. 2006).

21   **Orc4-bound regions in *C. albicans* lack a common DNA sequence motif but are**

22   **spatiotemporally positioned across the genome**

23   Conserved DNA sequence features at replication origins are common in the *Saccharomyces* group

24   (Nieduszynski, Knox et al. 2006). We used the *de novo* motif discovery tool DIVERSITY (Mitra,

25   Biswas et al. 2018) on the *C. albicans* Orc4 binding regions. DIVERSITY allows for the fact that the

26   profiled protein may have multiple modes of DNA binding. Here, DIVERSITY reports four binding

27   modes (Fig. 5A left). The first mode, mode A is a strong motif GAnTCGAAC, present in 50 such

1   regions, 49 of which were found to be located within tRNA gene bodies. The other three modes were

2   low complexity motifs, TGATGA (mode B), CAnCAnCAn (mode C) and AGnAG (mode D).

3   Strikingly, each of the 417 binding regions were associated with one of these motifs. Mode C has

4   been identified before (Tsai, Baller et al. 2014). The association to tRNA genes has been

5   demonstrated previously in a subset of *S. cerevisiae* replication origins as well (Wyrick, Aparicio et

6   al. 2001). Taken together, this suggests that ORCs in *C. albicans* do not rely on a specific sequence

7   feature for binding DNA.

8   Replication origins are spatially distributed and temporally regulated to ensure timely duplication of

9   the genome as well as to avoid re-initiation events. Depending on the time of activation and

10   efficiency, replication origins are classified as early and late domain/factories. To categorize the

11   replication timing of Orc4 binding sites, we utilized the fully processed replication timing profile of

12   *C. albicans* available from a previous study (Koren, Tsai et al. 2010) and overlaid the DIVERSITY

13   motifs onto the timing profile (Supplemental fig. S9). We observed a significant advanced replication

14   timing of the tRNA associated motifs (mode A) (Fig. 5A middle). The other three modes (B, C, D)

15   display no significant bias towards an early replication score. Moreover, we could correlate early

16   replication timing with an increased enrichment of Orc4 in these regions (Fig. 5A right). Additionally,

17   all the motifs were located towards the local maxima of the timing peaks (Supplemental fig. S9).

18   To locate these regions within the nuclear space, we mapped the interactions made by ORC binding

19   regions with each other using the Hi-C data from a previous study in *C. albicans* (Burrack, Hutton et

20   al. 2016). All the ORC binding regions were aligned with an increasing order of their replication

21   timing (early to late) and subsequent interactions were mapped. Similar analysis was performed for

22   the whole genome of *C. albicans.* We observe that the overall "only-ORC" interactions are higher

23   than the whole-genome "all" interactions, suggesting that ORC binding regions interact more than the

24   average (Supplemental figs. S10A, S10B). Early replicating regions (Fig. 5B) show a significantly

25   higher interaction among themselves, in agreement with previous observations in *Candida glabrata*

26   (Descorps-Declere, Saguez et al. 2015). Given that regions in this heatmap are ordered by timing and

27   not genomic proximity, this suggests that regions with a similar timing in replication tend to associate

1   together. Hi-C analysis also revealed that mode A containing sites, that show an early replication

2   timing, form stronger interactions among themselves than all the other modes (Fig. 5B). Hence, it is

3   highly likely that a subset of ORC binding regions identified in our study are the chromosomal origins

4   in *C. albicans* as they associate with categorically distinct domains separated in space and time of

5   replication, facilitating origin function and usage.

6   **The strong physical association of Orc4 at *C. albicans* centromeres stabilizes CENPA**

7   Apart from the discrete genomic loci across all chromosome arms, the strong binding of Orc4 on all

8   centromeres in *C. albicans* was particularly striking. This hints towards the possible role of replication

9   initiator complexes in influencing centromere location and function. Upon comparison of the Orc4

10   enrichment with the CENPA occupancy in *C. albicans*, we observe that there is a significant overlap

11   in the binding regions of both these proteins, indicating a strong physical association of ORCs at all

12   centromeres (Fig. 6A, Supplemental Fig. S11A) (Supplemental table S5). We additionally examined

13   for the presence of Orc4 in non-native centromeres. ChIP-qPCR analysis in the 5'FOA resistant strain

14   LSK443 revealed that similar to CENPA binding, the conditional ecCEN at *URA3* is enriched with

15   Orc4 (Supplemental fig. S11B). To validate the association of Orc4 at functional centromeres, we

16   explored its binding to strains forming neocentromeres. Neocentromeres activated at *nCEN7-II*

17   hotspot upon deletion of the 4.5 kb CENPA rich region on *CEN7* showed a significant Orc4

18   enrichment on the altered homolog (Supplemental fig. S11C). These observations strongly suggest

19   that Orc4 is associated with all active centromeres in *C. albicans.* To examine its role in centromere

20   function, we assayed for CENPA localization in an *orc4* conditional mutant, LSK331

21   (*orc4/MET3prORC4 CSE4/CSE4-GFP-CSE4)*. Orc4 depletion caused severe chromosome mis-

22   segregation (Supplemental fig. S13A). ChIP-qPCR analysis revealed a significant reduction of

23   chromatin associated CENPA upon Orc4 depletion (Fig. 6B), which was corroborated by degradation

24   of CENPA protein levels (Supplemental fig. S13B). However, depletion of CENPA did not

25   significantly alter the levels of Orc4 bound to the centromere (Fig. 6C), indicating that Orc4 strictly

26   regulates CENPA localization at the centromere but not vice-versa. Hence, Orc4 has a direct role in

27   stabilizing CENPA, thereby influencing centromere activity and kinetochore segregation.

12

1   **The helicase subunit, Mcm2 influences CENPA stability and kinetochore segregation**

2   Even though ORCs flag mark replication origins in the genome, their subsequent activity is governed

3   by assembly of the Mcm2-7 helicase that primes the complex for replication initiation. There are

4   distinct subunits and subunits of the pre-replicative complex (pre-RC) which perform roles outside

5   replication initiation. *MCM2* is annotated as an uncharacterized ORF (Orf19.4354) in the *Candida*

6   *Genome Database* (candidagenome.org). BLAST analysis using *S. cerevisiae* Mcm2 as the query

7   sequence revealed that this Orf19.4354 translates to a 101.2 kDa protein that contains the conserved

8   Walker A, Walker B and the R finger motif which together constitute the MCM box (Forsburg 2004)

9   (Supplemental fig. 12). In order to determine the essentiality of this gene in *C. albicans*, we

10   constructed a conditional mutant of *mcm2,* LSK311 (*mcm2/MET3prMCM2 CSE4-GFP-CSE4/CSE4*)

11   by deleting one allele and replacing the endogenous promoter of the remaining *MCM2* allele with the

12   repressive *MET3* promoter (Care, Trevethick et al. 1999). Mcm2 was found to be essential for

13   viability (Fig 7B). We could detect severe kinetochore segregation defects in this mutant post 6 h of

14   depletion of the protein (Supplemental fig. 13C). Depletion of Mcm2 led to a reduction in CENPA at

15   the centromere (Fig. 7C, Supplemental fig. 13D). Hence, Mcm2 is possibly helping in loading the

16   CENPA-H4 dimer at the *C. albicans* centromere, similar to what has been reported in human cells

17   (Huang, Stromme et al. 2015). Taken together, we establish how an intricate crosstalk of DNA

18   replication initiator proteins and early replication program at the pericentric regions help load,

19   stabilize and propagate centromeric chromatin in absence of any obvious DNA sequence cues in *C.*

20   *albicans.*

21

22   **Discussion**

23   Seeding of CENPA on DNA, the stability of centromeric chromatin during the cell cycle and its

24   subsequent propagation involves a plethora of factors ranging from the primary DNA sequence and

25   the chromatin context to crosstalk with DNA replication initiator and DNA damage repair proteins.

26   Specific protein binding sites aid centromere formation in genetically defined point centromeres

13

1 (Lechner and Carbon 1991). Additional mechanisms must operate at multi-dimensional levels to

2 spatiotemporally define centromere activity to a defined region in epigenetically regulated regional

3 centromeres in most other organisms. Centromeric heterochromatin is distinct from arm

4 heterochromatin, in terms of the degree of compaction and presence of topological adjusters like

5 cohesin, condensin and topoisomerase II (Bloom 2014). The cruciform structure adopted by the

6 centromeric chromatin in budding yeast facilitates cohesin maintenance on duplicated sister

7 chromatids and orients the centromere towards the spindle pole (Stephens, Haase et al. 2011)

8 (Lawrimore, Doshi et al. 2018). Centromere clustering gradually increases with cell cycle progression

9 due to sister-chromatid cohesion during replication, and cohesin mediated spindle-dependent

10 clustering during anaphase (Lazar-Stefanita, Scolari et al. 2017). In this study, we determine the

11 extent and functional consequence of centromeric chromatin compaction in *C. albicans.* The fact that

12 inter-centromeric interactions are much stronger than the average genomic interactions facilitates the

13 formation of a CENPA cloud in a 3-dimensional milieu to enrich the local CENPA concentration in

14 the clustered CENs of *C. albicans*. In the *S. cerevisiae* point centromeres, the presence of core and

15 accessory CENPA molecules at the native centromeres and pericentromeres, respectively, helps in

16 rapid incorporation of CENPA into the CEN chromatin during rogue loss events (Haase, Mishra et al.

17 2013), suggesting the dynamic nature of pericentromeric nucleosomes. Hence, the identification of a

18 highly interacting 25 kb pericentric region in *C. albicans* enables us to dissect functional

19 underpinnings of pericentromeres and spatial segregation of chromatin properties, in this case, created

20 by the pericentric heterochromatin that acts as the reservoir of CENPA molecules.

21 The strong reversible silencing of the transgene at the *C. albicans* central core, that is a readout of its

22 flexible transcriptional status (Thakur and Sanyal 2013) is reminiscent of the repeat-associated

23 centromere organization in *S. pombe,* where the central core shows variegated levels of marker gene

24 expression whereas the outer repeats shut down the transgene expression due to

25 heterochromatinization (Allshire, Javerzat et al. 1994, Karpen and Allshire 1997). Even though *S.*

26 *pombe* outer repeats do not bind CENPA, they are considered an important component of a functional

27 centromere (Clarke, Amstutz et al. 1986). Similarly, the pericentric regions identified in our study,

1 probably possess pericentric properties in *C. albicans*. In *S. pombe*, CENPA can assemble on a non-

2 centromeric region by competing out H3 (Castillo, Mellone et al. 2007), and the frequency of

3 reversible silencing can be increased by overexpressing CENPA. In our study, the strong negative

4 selection imposed on cells by 5'FOA, enables us to isolate rare individuals from a heterogenous

5 population of cells that can transiently incorporate CENPA at an ectopic locus. However, even under

6 selective growth conditions, only a few cells can tolerate the ecCEN because of the presence of the

7 more dominating native centromere locus (*CEN7*), eventually weeding out cells with ecCEN from the

8 population. Formation of an ecCEN outside the native CEN strongly suggests the existence of that

9 non-centromeric CENPA molecules interspersed with the H3 nucleosome. Unlike *S. pombe,* CENPA

10 overexpression in *C. albicans* does not lead to its extended occupancy beyond centromeric chromatin,

11 it merely increases its occupancy at the native locus (Berman 2012). It is to be noted here that the

12 ecCEN that was obtained by growth in selective media did not have an over-expression of CENPA

13 and still harbored two intact copies of native *CEN7*. CENPA associated chromatin, has self-

14 propagating properties and hence relies on an epigenetic memory (Black, Brock et al. 2007). Our

15 observation that any CENPA-primed region within the identified pericentric boundaries (25 kb

16 centring on the CEN core) can initiate neocentromere formation, emphasizes the importance of the

17 number of CENPA molecules required to nucleate the kinetochore assembly. However, we have

18 limited understanding regarding the determinants that act is favor of a particular locus on a

19 chromosome to have a "centromere correctness".

20 Incorporation of CENPA into replicated chromosomes is uncoupled with DNA replication in most

21 organisms. CEN proximal replication origins facilitate early replication of CEN chromatin in *C.*

22 *albicans* (Koren, Tsai et al. 2010, Mitra, Gomez-Raja et al. 2014). Naturally, the identification of

23 genome-wide replication origins in *C. albicans* will reveal useful insights into the replication

24 architecture of this organism. Towards this objective, first, our analysis of the Orc4 binding regions

25 revealed the lack of a DNA sequence requirement for most of the ORC binding sites in *C. albicans*. A

26 previous genome-wide study on identification of ORC binding regions in *C. albicans* utilized

27 antibodies against the *S. cerevisiae* ORC complex (Tsai, Baller et al. 2014). The said study reported

1 ~390 ORC binding sites which exhibited a 25% overlap with the Orc4 binding regions identified in

2 our study. Since we used antibodies against an endogenous protein (CaOrc4) to map its binding sites

3 in *C. albicans,* we present an authentic depiction of Orc4 binding regions in the genome. We do find a

4 strong association of a fraction of these regions within many tRNA genes. tRNA genes along with

5 histone genes and centromeres are known to exhibit conserved replication timing (Muller and

6 Nieduszynski 2017). Moreover, in *S. cerevisiae* there is a statistically significant bias for codirectional

7 transcription and replication of tRNA genes (Muller and Nieduszynski 2017). tDNAs cluster near

8 centromeres and recovers stalled forks (Thompson, Haeusler et al. 2003). Since centromeres are early

9 replicating in fungal genomes, the presence of tDNAs in its vicinity might transduce an early

10 replication program. The various Orc4 binding DNA motifs identified in our study hints towards

11 differential usage and specification of origins facilitated by multiple modes of ORC binding in *C.*

12 *albicans.* Secondly, in spite of the sequence heterogeneity, these Orc4 binding regions could be

13 classified based on their replication timing, wherein the early replicating regions form closely

14 associated units and interact sparsely with the late replicating ones. This is reminiscent of the genome-

15 wide replication landscape of *C. glabrata* origins (Descorps-Declere, Saguez et al. 2015). The

16 stochastic activation of early origins, makes up for the uneven distribution of origins in the genome.

17 The 'replication wave' progresses with the sequential activation of pericentromeric origins to the

18 chromosome arm origins. Initiation events convert early origin clusters to replication foci during S

19 phase. As DNA replication progresses, more replisomes are formed, chromosomes are sufficiently

20 mobilized which makes long range interactions more favourable with time. Hence, one can speculate

21 the existence of topologically distinct domains that are separated in location and time as S-phase

22 progresses.

23 *C. albicans* centromeres do not possess a firing origin (Mitra, Gomez-Raja et al. 2014). Replication

24 forks originating from the centromere flanking origins stall at the centromere in a kinetochore

25 dependent manner and facilitate new CENPA loading. Furthermore, CENPA loading is facilitated by

26 the physical interaction of repair proteins like Rad51, Rad52 with CENPA, that are transiently

27 localized to the kinetochore upon replication fork stalling (Mitra, Gomez-Raja et al. 2014). Hence,

1   there is an interplay of replication-repair machinery in maintaining centromere identity in *C. albicans*.

2   We postulate that an ORC-cloud (Fig. 8) facilitates ORC abundance at all centromeres of *C. albicans*

3   and can be attributed to the early replication of centromeres in every S-phase which in turn influences

4   the loading of new CENPA. The anaphase specific loading of new CENPA in *C. albicans* has been

5   demonstrated earlier by fluorescence spectroscopic measurements (Shivaraju, Unruh et al. 2012).

6   However, specific chaperones and molecular pathways involved in the same are undeciphered. We

7   posit ORC to be an essential component for CENPA loading by maintaining a heterochromatin

8   environment at the centromeric locus. Members of the pre-RC have established roles in cell cycle

9   dependent dynamics at the centromere. Mcm2 is a known chaperone that hands over the old histones

10 from the replication fork to anti-silencing function Asf1 to recycle old histones and deposits them to

11 the newly synthesised DNA (Hammond, Stromme et al. 2017). Mcm2 and Asf1 cochaperone an H3-

12 H4 dimer through histone-binding mode (Richet, Liu et al. 2015). This is true for both canonical H3

13 as well as H3 variants like H3.3 and CENPA (Huang, Stromme et al. 2015). A recent study indicates

14 the role played by Mcm2 in mouse embryonic stem cells to symmetrically partition modified histones

15 to daughter cells using its histone-binding mode (Petryk, Dalby et al. 2018). In humans, the S phase

16 retention of CENPA is mitigated by its simultaneous interaction with the specific chaperone HJURP

17 and Mcm2 (Zasadzinska, Huang et al. 2018), which together transmit CENPA nucleosomes upon its

18 disassembly ahead of the replication fork. In the light of the existing evidence in metazoan systems

19 and the results obtained in our study, Mcm2 emerges as an evolutionarily conserved factor required

20 for eviction of old CENPA molecules and loading of newly synthesized ones (Fig. 8). Although our

21 experiments demonstrate a strong genetic interaction between these two proteins, the physical

22 interaction of Orc4-CENPA and Mcm2-CENPA is still speculative due to technical difficulties. We

23 hypothesize that during CEN chromatin replication at S phase, ORCs maintain the heterochromatin

24 environment of CEN when "old" CENPA is evicted (Fig. 8). During anaphase, centromeric ORCs are

25 briefly displaced, to facilitate loading of "new" CENPA with the help of a specific chaperone such as

26 Scm3/HJURP and Mcm2 which stabilizes the kinetochore complex. In the next cell cycle, Mcm2

27 associates with the MCM complex to license replication origins during G_1. ORCs in *S. cerevisiae*

28 have established roles in heterochromatinization and MTL silencing (Foss, McNally et al. 1993,

17

1 Hickman, Froyd et al. 2011). The centromere silencing mechanisms in *C. albicans* is relatively

2 unknown as this organism lacks a functional RNAi machinery and H3K9me2 marks. In this regard,

3 we envision the ORC family of proteins as a possible silencing factor for centromeres in this

4 organism.

5 **Methods**

6 All the strains and primers are listed in Supplemental Tables S6 and S7, respectively. Protocols and

7 experimental procedures have been mentioned in Supplementary information.

8 **Data access**

9 The sequencing data used in the study have been submitted to NBCI under the SRA accession number

10 PRJNA477284.

11 **Acknowledgements**

22 **References:**

23 Allshire, R. C. and K. Ekwall (2015). "Epigenetic Regulation of Chromatin States in
24 Schizosaccharomyces pombe." Cold Spring Harb Perspect Biol **7**(7): a018770.
25 Allshire, R. C., J. P. Javerzat, N. J. Redhead and G. Cranston (1994). "Position effect variegation at
26 fission yeast centromeres." Cell **76**(1): 157-169.
27 Allshire, R. C. and H. D. Madhani (2018). "Ten principles of heterochromatin formation and
28 function." Nat Rev Mol Cell Biol **19**(4): 229-244.
29 Amor, D. J. and K. H. Choo (2002). "Neocentromeres: role in human disease, evolution, and
30 centromere study." Am J Hum Genet **71**(4): 695-714.
31 Baum, M., K. Sanyal, P. K. Mishra, N. Thaler and J. Carbon (2006). "Formation of functional
32 centromeric chromatin is specified epigenetically in Candida albicans." Proc Natl Acad Sci U S A
33 **103**(40): 14877-14882.

1   Berman, J. (2012). "Candida albicans." Curr Biol **22**(16): R620-622.
2   Black, B. E., M. A. Brock, S. Bedard, V. L. Woods, Jr. and D. W. Cleveland (2007). "An epigenetic mark
3   generated by the incorporation of CENP-A into centromeric nucleosomes." Proc Natl Acad Sci U S A
4   **104**(12): 5008-5013.
5   Bloom, K. S. (2014). "Centromeric heterochromatin: the primordial segregation machine." Annu Rev
6   Genet **48**: 457-484.
7   Bobkov, G. O. M., N. Gilbert and P. Heun (2018). "Centromere transcription allows CENP-A to transit
8   from chromatin association to stable incorporation." J Cell Biol.
9   Burrack, L. S., H. F. Hutton, K. J. Matter, S. A. Clancey, I. Liachko, A. E. Plemmons, A. Saha, E. A.
10   Power, B. Turman, M. A. Thevandavakkam, F. Ay, M. J. Dunham and J. Berman (2016).
11   "Neocentromeres Provide Chromosome Segregation Accuracy and Centromere Clustering to
12   Multiple Loci along a Candida albicans Chromosome." PLoS Genet **12**(9): e1006317.
13   Care, R. S., J. Trevethick, K. M. Binley and P. E. Sudbery (1999). "The MET3 promoter: a new tool for
14   Candida albicans molecular genetics." Mol Microbiol **34**(4): 792-798.
15   Castillo, A. G., B. G. Mellone, J. F. Partridge, W. Richardson, G. L. Hamilton, R. C. Allshire and A. L.
16   Pidoux (2007). "Plasticity of fission yeast CENP-A chromatin driven by relative levels of histone H3
17   and H4." PLoS Genet **3**(7): e121.
18   Catania, S., A. L. Pidoux and R. C. Allshire (2015). "Sequence features and transcriptional stalling
19   within centromere DNA promote establishment of CENP-A chromatin." PLoS Genet **11**(3): e1004986.
20   Choi, E. S., A. Stralfors, A. G. Castillo, M. Durand-Dubief, K. Ekwall and R. C. Allshire (2011).
21   "Identification of noncoding transcripts from within CENP-A chromatin at fission yeast centromeres."
22   J Biol Chem **286**(26): 23600-23607.
23   Chuang, R. Y. and T. J. Kelly (1999). "The fission yeast homologue of Orc4p binds to replication origin
24   DNA via multiple AT-hooks." Proc Natl Acad Sci U S A **96**(6): 2656-2661.
25   Clarke, L., H. Amstutz, B. Fishel and J. Carbon (1986). "Analysis of centromeric DNA in the fission
26   yeast Schizosaccharomyces pombe." Proc Natl Acad Sci U S A **83**(21): 8253-8257.
27   Cook, D. M., M. Bennett, B. Friedman, J. Lawrimore, E. Yeh and K. Bloom (2018). "Fork pausing
28   allows centromere DNA loop formation and kinetochore assembly." Proc Natl Acad Sci U S A.
29   Craig, J. M., L. H. Wong, A. W. Lo, E. Earle and K. H. Choo (2003). "Centromeric chromatin pliability
30   and memory at a human neocentromere." EMBO J **22**(10): 2495-2504.
31   Dai, J., R. Y. Chuang and T. J. Kelly (2005). "DNA replication origins in the Schizosaccharomyces
32   pombe genome." Proc Natl Acad Sci U S A **102**(2): 337-342.
33   Descorps-Declere, S., C. Saguez, A. Cournac, M. Marbouty, T. Rolland, L. Ma, C. Bouchier, I. Moszer,
34   B. Dujon, R. Koszul and G. F. Richard (2015). "Genome-wide replication landscape of Candida
35   glabrata." BMC Biol **13**: 69.
36   Dutta, A. and S. P. Bell (1997). "Initiation of DNA replication in eukaryotic cells." Annu Rev Cell Dev
37   Biol **13**: 293-332.
38   Forsburg, S. L. (2004). "Eukaryotic MCM proteins: beyond replication initiation." Microbiol Mol Biol
39   Rev **68**(1): 109-131.
40   Foss, M., F. J. McNally, P. Laurenson and J. Rine (1993). "Origin recognition complex (ORC) in
41   transcriptional silencing and DNA replication in S. cerevisiae." Science **262**(5141): 1838-1844.
42   Freire-Beneitez, V., R. J. Price and A. Buscaino (2016). "The Chromatin of Candida albicans
43   Pericentromeres Bears Features of Both Euchromatin and Heterochromatin." Front Microbiol **7**: 759.
44   Fukagawa, T. and W. C. Earnshaw (2014). "The centromere: chromatin foundation for the
45   kinetochore machinery." Dev Cell **30**(5): 496-508.
46   Haase, J., P. K. Mishra, A. Stephens, R. Haggerty, C. Quammen, R. M. Taylor, 2nd, E. Yeh, M. A. Basrai
47   and K. Bloom (2013). "A 3D map of the yeast kinetochore reveals the presence of core and accessory
48   centromere-specific histone." Curr Biol **23**(19): 1939-1944.
49   Hammond, C. M., C. B. Stromme, H. Huang, D. J. Patel and A. Groth (2017). "Histone chaperone
50   networks shaping chromatin function." Nat Rev Mol Cell Biol **18**(3): 141-158.

Hickman, M. A., C. A. Froyd and L. N. Rusche (2011). "Reinventing heterochromatin in budding yeasts: Sir2 and the origin recognition complex take center stage." Eukaryot Cell **10**(9): 1183-1192.

Hori, T., N. Kagawa, A. Toyoda, A. Fujiyama, S. Misu, N. Monma, F. Makino, K. Ikeo and T. Fukagawa (2017). "Constitutive centromere-associated network controls centromere drift in vertebrate cells." J Cell Biol **216**(1): 101-113.

Huang, H., C. B. Stromme, G. Saredi, M. Hodl, A. Strandsby, C. Gonzalez-Aguilera, S. Chen, A. Groth and D. J. Patel (2015). "A unique binding mode enables MCM2 to chaperone histones H3-H4 at replication forks." Nat Struct Mol Biol **22**(8): 618-626.

Ishii, K., Y. Ogiyama, Y. Chikashige, S. Soejima, F. Masuda, T. Kakuma, Y. Hiraoka and K. Takahashi (2008). "Heterochromatin integrity affects chromosome reorganization after centromere dysfunction." Science **321**(5892): 1088-1091.

Karpen, G. H. and R. C. Allshire (1997). "The case for epigenetic effects on centromere identity and function." Trends Genet **13**(12): 489-496.

Koren, A., H. J. Tsai, I. Tirosh, L. S. Burrack, N. Barkai and J. Berman (2010). "Epigenetically-inherited centromere and neocentromere DNA replicates earliest in S-phase." PLoS Genet **6**(8): e1001068.

Lawrimore, J., A. Doshi, B. Friedman, E. Yeh and K. Bloom (2018). "Geometric partitioning of cohesin and condensin is a consequence of chromatin loops." Mol Biol Cell **29**(22): 2737-2750.

Lazar-Stefanita, L., V. F. Scolari, G. Mercy, H. Muller, T. M. Guerin, A. Thierry, J. Mozziconacci and R. Koszul (2017). "Cohesins and condensins orchestrate the 4D dynamics of yeast chromosomes during the cell cycle." EMBO J **36**(18): 2684-2697.

Lechner, J. and J. Carbon (1991). "A 240 kd multisubunit protein complex, CBF3, is a major component of the budding yeast centromere." Cell **64**(4): 717-725.

Leonard, A. C. and M. Mechali (2013). "DNA replication origins." Cold Spring Harb Perspect Biol **5**(10): a010116.

Livny, J., Y. Yamaichi and M. K. Waldor (2007). "Distribution of centromere-like parS sites in bacteria: insights from comparative genomics." J Bacteriol **189**(23): 8693-8703.

Maggert, K. A. and G. H. Karpen (2001). "The activation of a neocentromere in Drosophila requires proximity to an endogenous centromere." Genetics **158**(4): 1615-1628.

Marshall, O. J., A. C. Chueh, L. H. Wong and K. H. Choo (2008). "Neocentromeres: new insights into centromere structure, disease development, and karyotype evolution." Am J Hum Genet **82**(2): 261-282.

McNulty, S. M. and B. A. Sullivan (2017). "Centromere Silencing Mechanisms." Prog Mol Subcell Biol **56**: 233-255.

McNulty, S. M., L. L. Sullivan and B. A. Sullivan (2017). "Human Centromeres Produce Chromosome-Specific and Array-Specific Alpha Satellite Transcripts that Are Complexed with CENP-A and CENP-C." Dev Cell **42**(3): 226-240 e226.

Mellone, B. G. and R. C. Allshire (2003). "Stretching it: putting the CEN(P-A) in centromere." Curr Opin Genet Dev **13**(2): 191-198.

Mitra, S., A. Biswas and L. Narlikar (2018). "DIVERSITY in binding, regulation, and evolution revealed from high-throughput ChIP." PLoS Comput Biol **14**(4): e1006090.

Mitra, S., J. Gomez-Raja, G. Larriba, D. D. Dubey and K. Sanyal (2014). "Rad51-Rad52 mediated maintenance of centromeric chromatin in Candida albicans." PLoS Genet **10**(4): e1004344.

Muller, C. A. and C. A. Nieduszynski (2017). "DNA replication timing influences gene expression level." J Cell Biol **216**(7): 1907-1914.

Natsume, T., C. A. Muller, Y. Katou, R. Retkute, M. Gierlinski, H. Araki, J. J. Blow, K. Shirahige, C. A. Nieduszynski and T. U. Tanaka (2013). "Kinetochores coordinate pericentromeric cohesion and early DNA replication by Cdc7-Dbf4 kinase recruitment." Mol Cell **50**(5): 661-674.

Nieduszynski, C. A., Y. Knox and A. D. Donaldson (2006). "Genome-wide identification of replication origins in yeast by comparative genomics." Genes Dev **20**(14): 1874-1879.

Parker, M. W., M. R. Botchan and J. M. Berger (2017). "Mechanisms and regulation of DNA replication initiation in eukaryotes." Crit Rev Biochem Mol Biol **52**(2): 107-144.

Pearson, C. G., E. Yeh, M. Gardner, D. Odde, E. D. Salmon and K. Bloom (2004). "Stable kinetochore-microtubule attachment constrains centromere positioning in metaphase." Curr Biol **14**(21): 1962-1967.

Petryk, N., M. Dalby, A. Wenger, C. B. Stromme, A. Strandsby, R. Andersson and A. Groth (2018). "MCM2 promotes symmetric inheritance of modified histones during DNA replication." Science.

Richet, N., D. Liu, P. Legrand, C. Velours, A. Corpet, A. Gaubert, M. Bakail, G. Moal-Raisin, R. Guerois, C. Compper, A. Besle, B. Guichard, G. Almouzni and F. Ochsenbein (2015). "Structural insight into how the human helicase subunit MCM2 may act as a histone chaperone together with ASF1 at the replication fork." Nucleic Acids Res **43**(3): 1905-1917.

Roy, B., L. S. Burrack, M. A. Lone, J. Berman and K. Sanyal (2011). "CaMtw1, a member of the evolutionarily conserved Mis12 kinetochore protein family, is required for efficient inner kinetochore assembly in the pathogenic yeast Candida albicans." Mol Microbiol **80**(1): 14-32.

Sanyal, K., M. Baum and J. Carbon (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.

Sanyal, K. and J. Carbon (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.

Shang, W. H., T. Hori, N. M. Martins, A. Toyoda, S. Misu, N. Monma, I. Hiratani, K. Maeshima, K. Ikeo, A. Fujiyama, H. Kimura, W. C. Earnshaw and T. Fukagawa (2013). "Chromosome engineering allows the efficient isolation of vertebrate neocentromeres." Dev Cell **24**(6): 635-648.

Shivaraju, M., J. R. Unruh, B. D. Slaughter, M. Mattingly, J. Berman and J. L. Gerton (2012). "Cell-cycle-coupled structural oscillation of centromeric nucleosomes in yeast." Cell **150**(2): 304-316.

Stephens, A. D., J. Haase, L. Vicci, R. M. Taylor, 2nd and K. Bloom (2011). "Cohesin, condensin, and the intramolecular centromere loop together generate the mitotic chromatin spring." J Cell Biol **193**(7): 1167-1180.

Sullivan, L. L., K. A. Maloney, A. J. Towers, S. G. Gregory and B. A. Sullivan (2016). "Human centromere repositioning within euchromatin after partial chromosome deletion." Chromosome Res **24**(4): 451-466.

Thakur, J. and K. Sanyal (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.

Thompson, M., R. A. Haeusler, P. D. Good and D. R. Engelke (2003). "Nucleolar clustering of dispersed tRNA genes." Science **302**(5649): 1399-1401.

Tsai, H. J., J. A. Baller, I. Liachko, A. Koren, L. S. Burrack, M. A. Hickman, M. A. Thevandavakkam, L. N. Rusche and J. Berman (2014). "Origin replication complex binding, nucleosome depletion patterns, and a primary sequence motif can predict origins of replication in a genome with epigenetic centromeres." MBio **5**(5): e01703-01714.

Vernis, L., A. Abbas, M. Chasles, C. M. Gaillardin, C. Brun, J. A. Huberman and P. Fournier (1997). "An origin of replication and a centromere are both needed to establish a replicative plasmid in the yeast Yarrowia lipolytica." Mol Cell Biol **17**(4): 1995-2004.

Voullaire, L. E., H. R. Slater, V. Petrovic and K. H. Choo (1993). "A functional marker centromere with no detectable alpha-satellite, satellite III, or CENP-B protein: activation of a latent centromere?" Am J Hum Genet **52**(6): 1153-1163.

Walker, J. E., M. Saraste, M. J. Runswick and N. J. Gay (1982). "Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold." EMBO J **1**(8): 945-951.

Warburton, P. E. (2004). "Chromosomal dynamics of human neocentromere formation." Chromosome Res **12**(6): 617-626.

Wyrick, J. J., J. G. Aparicio, T. Chen, J. D. Barnett, E. G. Jennings, R. A. Young, S. P. Bell and O. M. Aparicio (2001). "Genome-wide distribution of ORC and MCM proteins in S. cerevisiae: high-resolution mapping of replication origins." Science **294**(5550): 2357-2360.

1   Xu, W., J. G. Aparicio, O. M. Aparicio and S. Tavare (2006). "Genome-wide mapping of ORC and
2   Mcm2p binding sites on tiling arrays and identification of essential ARS consensus sequences in S.
3   cerevisiae." BMC Genomics **7**: 276.
4   Yamazaki, S., M. Hayano and H. Masai (2013). "Replication timing regulation of eukaryotic replicons:
5   Rif1 as a global regulator of replication timing." Trends Genet **29**(8): 449-460.
6   Yao, J., X. Liu, T. Sakuno, W. Li, Y. Xi, P. Aravamudhan, A. Joglekar, Y. Watanabe and X. He (2013).
7   "Plasticity and epigenetic inheritance of centromere-specific histone H3 (CENP-A)-containing
8   nucleosome positioning in the fission yeast." J Biol Chem **288**(26): 19184-19196.
9   Zasadzinska, E., J. Huang, A. O. Bailey, L. Y. Guo, N. S. Lee, S. Srivastava, K. A. Wong, B. T. French, B. E.
10 Black and D. R. Foltz (2018). "Inheritance of CENP-A Nucleosomes during DNA Replication Requires
11 HJURP." Dev Cell.

12

13

14

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31
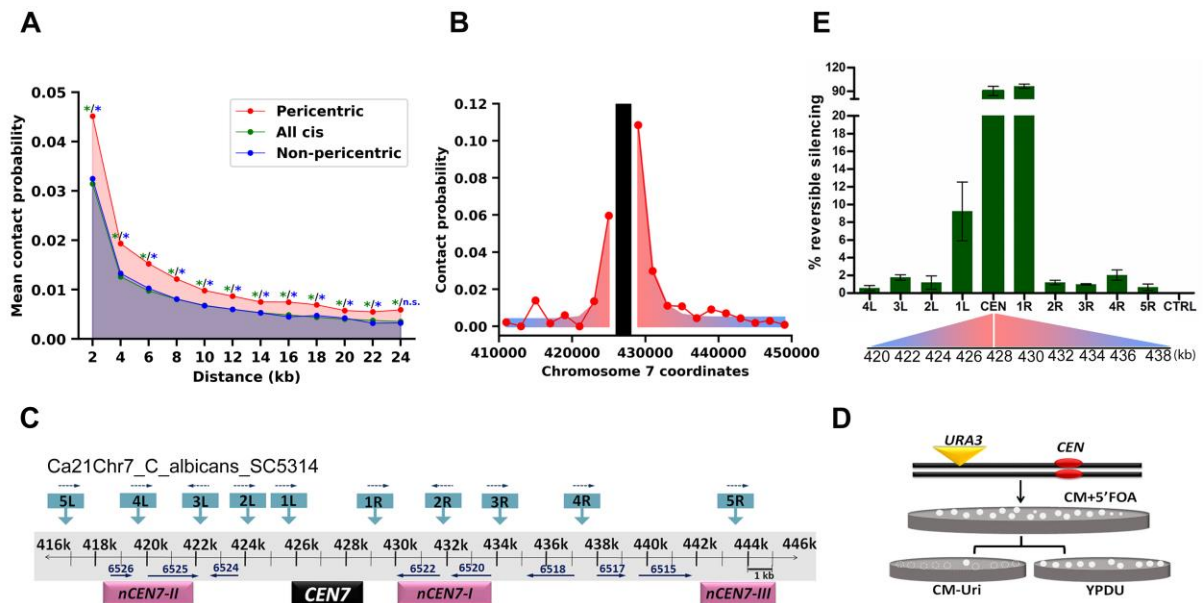
32

33

34

35

36

1   **Figures**

2   **Fig.1**

3

4

5   **Fig 1. Centromeres in *C. albicans* are flanked by pericentromeric chromatin spanning ~25 kb**
6   **centring the CENPA binding region. (A)** The mean Hi-C contact probability (*y*-axis; bin size= 2 kb)
7   separated at a given distance (*x*-axis) depicts significant increase in pericentric (centromeric region
8   plus 10 kb upstream and downstream) interactions (red) over all *cis* interactions (green) and non-
9   pericentric (a randomly selected region of same size as pericentric region and equidistant from the
10   centromere of each chromosome) interactions (blue). (*: p-value<0.05; n.s.: not significant). **(B)** The
11   3C profile (bin size=2kb) anchored on centromeric bin (Chr7:426000-428000) showing contact
12   probabilities (red dots) between the anchor bin (black) and its neighbouring bins on Chr7 indicates a
13   strongly interacting 25 kb region. **(C)** A line diagram of ~30 kb region on Chr7 surrounding *CEN7*
14   shows individual *URA3* insertion locations (blue arrows) and previously mapped neocentromere
15   hotspots (*nCEN7-I, nCEN7-II*). Arrowheads and numbers indicate positions and identities of the
16   ORFs respectively. **(D)** The assay strategy used to screen reversibly silenced colonies derived from
17   the *URA3* integrants using 5'FOA. **(E)** A decline in the percentage of reversible silencing of *URA3*
18   from mid-*CEN7* (CEN) to the pericentromeric integrants (4L,3L,2L…5R, CTRL) was observed with
19   increasing distance from native *CEN*. The phase exponential decay curve is colour coded with the
20   graph in (B). Coordinates for the respective insertions have been depicted below the black line.
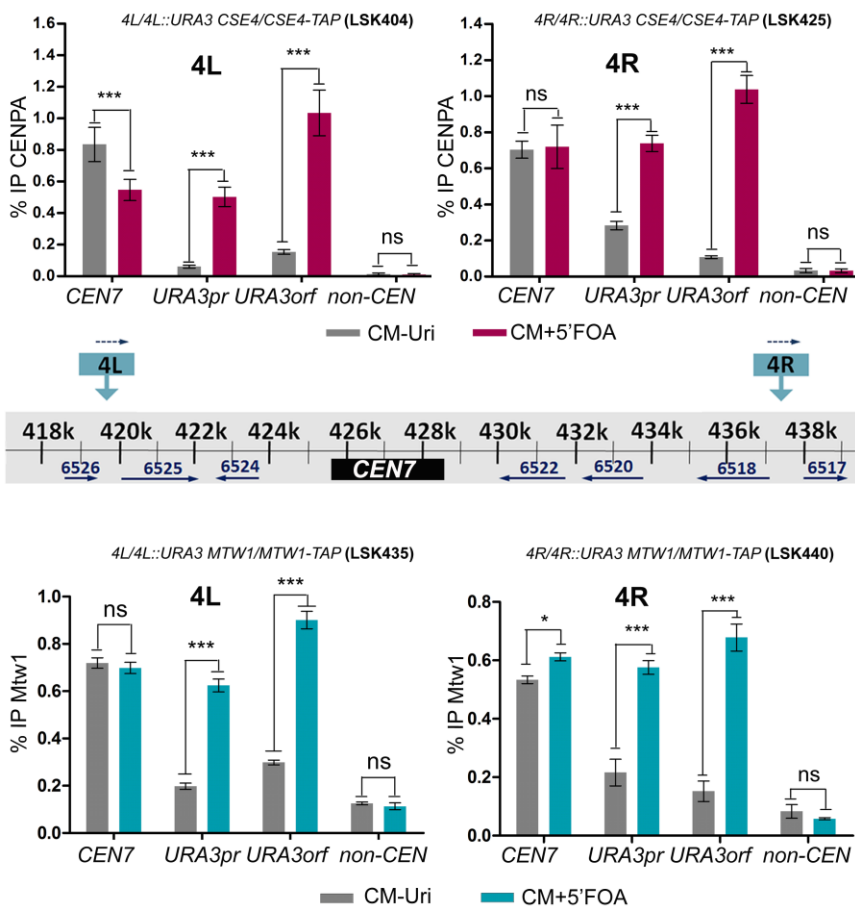
21

22

23

24

25

26

27

1 **Fig 2**

4 **Fig 2. Transcriptional silencing of *URA3* at the pericentromeres favours formation of an ectopic**
5 **kinetochore.** A line diagram of the *URA3* insertions at locations 4L and 4R on Chr7 is shown (middle
6 panel). Corresponding integrations are mentioned as graph titles. Anti-Protein A ChIP followed by
7 qPCR analysis of the 5'FOA resistant colonies from the strains LSK404 (4L) and LSK425 (4R) was
8 used to compare enrichment of CENPA on the indicated loci (*CEN7, URA3pr, URA3orf*, non-*CEN*
9 region) in CM-Uri (grey) and CM+5'FOA (red) media (upper panel). Similar ChIP-qPCR assays
10 (lower panels) performed on strains LSK435 (4L) and LSK440 (4R) showed significant enrichment of
11 Mtw1 on the silent *URA3* locus (blue bar). Normalised CENPA and Mtw1 enrichment values indicate
12 significant enrichment at *URA3* upon its transcriptional repression in 4L and 4R integrants (***
13 p<0.001, ** p<0.01, ns: p>0.05). (See Supplemental Figs S2 and S3 for ChIP-qPCR results of other
14 *URA3* integrants)

1 **Fig. 3**
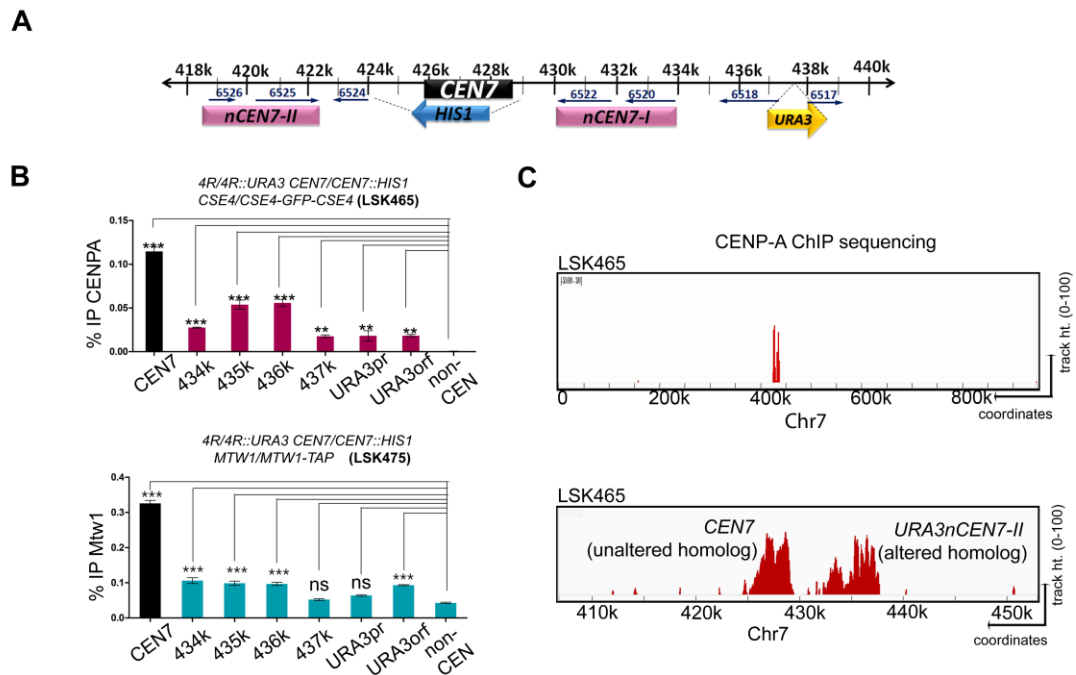
2

3 **Fig 3. Pre-existing CENPA molecules can prime a chromosomal location for neocentromere**
4 **formation. (A)** In the diploid *C. albicans*, only one homolog of Chr7 where *CEN7* (CaChr7 424475-
5 428994) has been replaced with *HIS1* in a *URA3* integrant strain (CaChr7 437729-437730) is shown.
6 **(B)** Top panel indicates relative enrichment of CENPA at native *CEN7* from the unaltered homolog
7 (black) and at neocentromere *URA3nCEN7-II* (red) in the 5'FOA resistant strain LSK465. Bottom
8 panel indicates relative enrichment of Mtw1 at *CEN7* (black) and *URA3nCEN7-II* (blue) at the native
9 centromere (427k) in the strain LSK465. Relative enrichment values of CENPA and Mtw1 indicate
10 that the neocentromere formed on the altered homolog (*URA3nCEN7-II*) was mapped to a region
11 surrounding the integration locus (CaChr7 435078-440387) error bars indicate SEM (***p<0.001, **
12 p<0.01, ns p>0.05). **(C)** CENPA ChIP-sequencing confirmed the presence of neocentromere in the
13 strain LSK465, where the profile is a combination of two peaks (top), the one at *CEN7* is on the
14 unaltered homolog the one at *URA3nCEN7-II)* is on the altered homolog. A 50 kb region (bottom)
15 harbouring *CEN7* depicts the track height (as on IGV) on *y*-axis and coordinates on the *x*-axis.
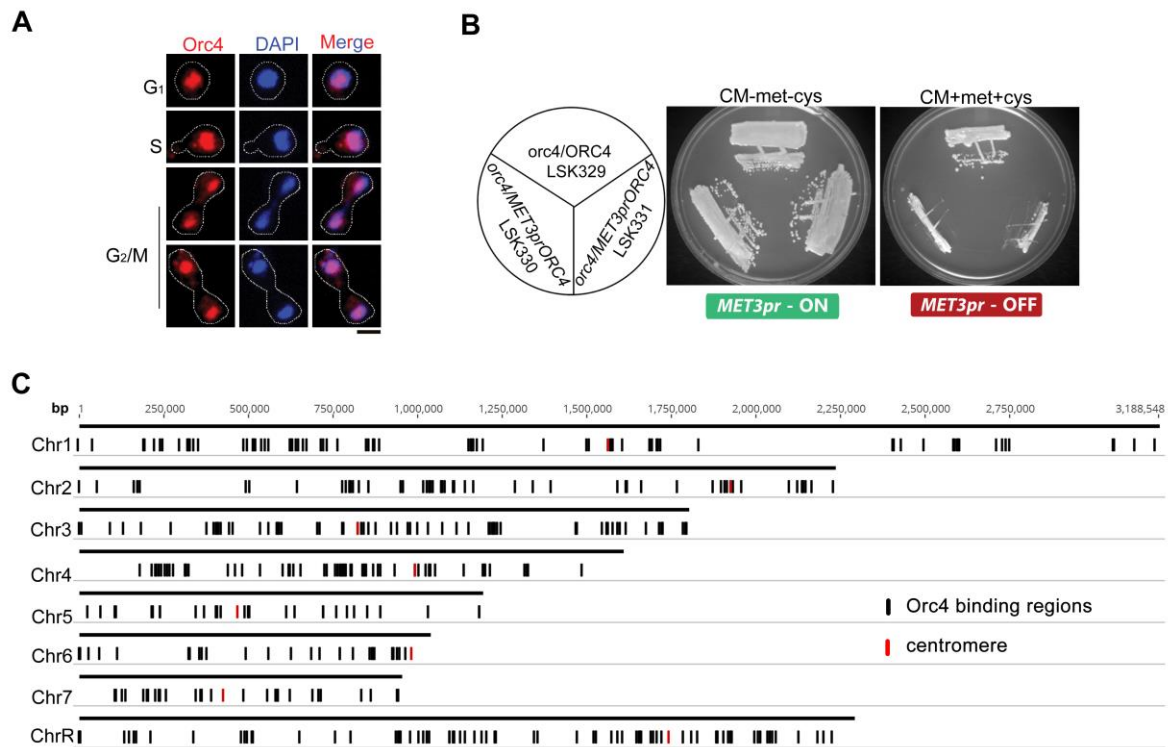
16

17

18

19

20

21

22

23

24

25

25

Fig 4.

Fig 4. The essential pre-RC component Orc4 binds to discrete loci in the *C. albicans* genome. (A) Intracellular localization of Orc4 in *C. albicans* SC5314 cells stained with anti-Orc4 antibodies (red) and DAPI (blue) indicates Orc4 is localized to the nucleus throughout the cell cycle in *C. albicans.* Bar, 5 μm. (**B**) The promoter of *MET3*, is expressed in the absence of methionine and cysteine and repressed in the presence of both, was used for the controlled expression of *ORC4*. *C. albicans* cells with one deleted copy of *ORC4*, LSK329 (*ORC4/ORC4::FRT*), and two independent transformants of the conditional mutant of *orc4* (*ORC4::FRT/MET3prORC4*), LSK330 and LSK331 where the remaining wild-type copy was placed under the control of the *MET3* promoter were streaked on plates containing inducible (CM-met-cys) and repressible (CM+ 5 mM cys + 5 mM met) media photographed after 48 h of incubation at 30°C. (**C**) A whole genome view of Orc4 binding regions (black bars) on each of the eight *C. albicans* chromosomes including all eight centromeres (red bar).

26

1 **Fig 5.**
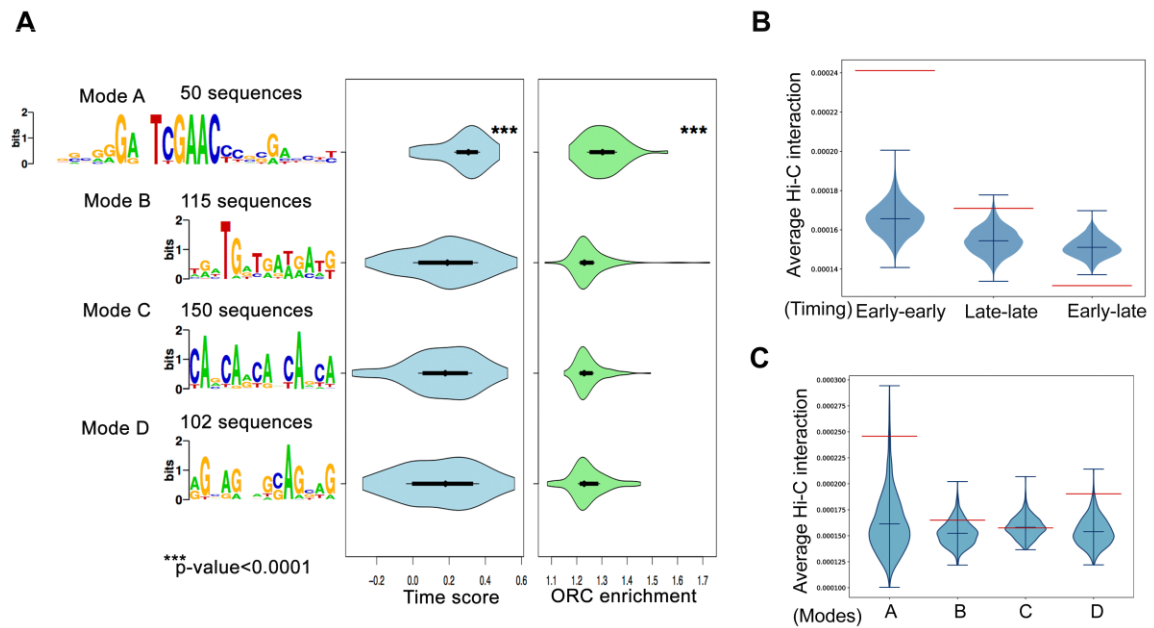
2

3 **Fig 5. Spatiotemporal distribution of DNA sequence independent replication origins in** *C.*
4 *albicans* (**A**) The four different modes (motifs) identified by DIVERSITY (A, B, C, D) and their
5 distribution across the 417 binding regions have been listed. Mode A corresponds to the tDNA motif.
6 Violin plots depicting the replication timing profile obtained from (Koren, Tsai et al. 2010) for all
7 four modes shows higher time scores for Mode A associated origins. Enrichment of Orc4 (based on
8 the ChIP-sequencing results) shows a significant increase for Mode A associated origins. The other
9 three modes depict enrichment values not significantly different from each other. (**B**) The average Hi-
10 C interaction (red lines) of the early-early, late-late and early-late regions that have been classified
11 based on their replication timing profile indicate higher interaction values among early replicating
12 domains. Blue violins indicate mean interactions across 1000 randomizations. (**C**) Mode A containing
13 regions show the highest average Hi-C interaction frequency (red lines) among themselves as
14 compared to the three other modes. Blue violins indicate mean interactions across 1000
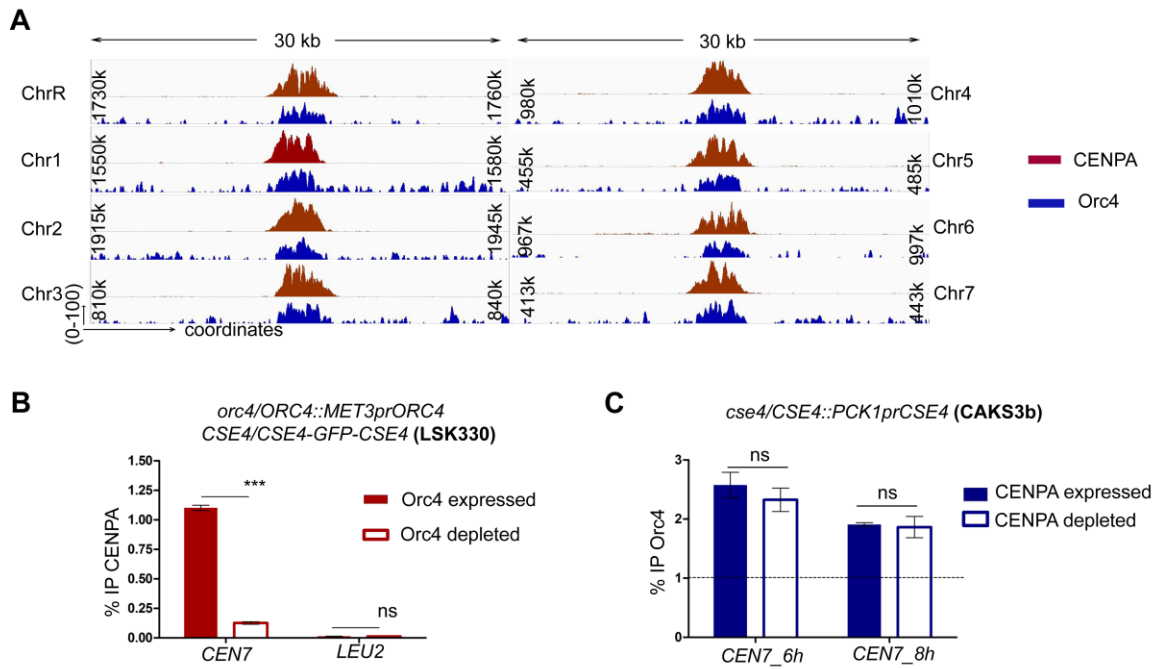15 randomizations.

16

17

18

19

20

21

22

23

1   **Fig 6.**

3   **Fig 6. Centromeric localization of Orc4 stabilizes CENPA. (A)** Orc4 ChIP sequencing revealed the
4   highest binding of Orc4 at all eight *C. albicans* centromeres. The binding region of Orc4 (blue)
5   showed a complete overlap with CENPA binding region (red). A 30 kb region harbouring each
6   centromere (*x*-axis) was plotted against the subtracted ChIP sequencing reads (*y*-axis) for CENPA and
7   Orc4. **(B)** ChIP-qPCR using anti-GFP antibodies (CENPA) revealed reduced CENPA enrichment at
8   the centromere upon Orc4 depletion in the strain LSK330 either in CM-met-cys or CM+5mM met +
9   5mM cys for 24h. (**C**) ChIP-q PCR using anti-Orc4 antibodies revealed no significant reduction in
10   centromeric Orc4 when CENPA was depleted for 6h and 8 h in YP with dextrose in the strain
11   CAKS3b. Percent IP values are normalized to a non-centromeric control region, *LEU2*. Two-way
12   ANOVA was used to determine statistical significance (***p<0.001, ns p>0.05).
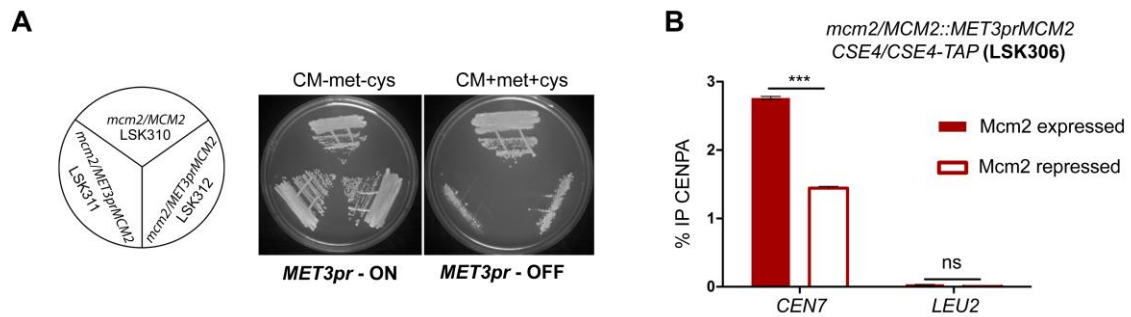
13

14

15

16

17

18

19

20

21

28

bioRxiv preprint first posted online Nov. 8, 2018; doi: http://dx.doi.org/10.1101/465880. The copyright holder for this preprint (which was not peer-reviewed) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license.

1 **Fig 7.**

2

3 **Fig 7. The helicase subunit Mcm2 is essential for viability and CENPA stability in *C. albicans*.**
4 (**A**) *C. albicans* cells where one deleted copy of *MCM2* has been deleted, LSK310
5 (*MCM2*/*MCM2*::*FRT*), and two independent transformants of the conditional mutant of *mcm2*
6 (*MCM2*::*FRT*/*MET3prMCM2*), LSK311 and LSK312 where the remaining wild-type copy was placed
7 under the control of the *MET3* promoter were streaked on plates containing inducible (CM-met-cys)
8 and repressible (CM+5 mM cys + 5 mM met) media and photographed after 48 h of incubation at
9 30°C. (**B**) ChIP-qPCR using anti-protein A antibodies revealed significant reduction in CENPA at
10 *CEN7* in the strain LSK306 when grown either in CM-met-cys or CM+5mM met+5mM cys for 6 h.
11 ChIP was performed in biological triplicates. Two-way ANOVA was used to determine statistical
12 significance (***p<0.001, ns p>0.05).

13
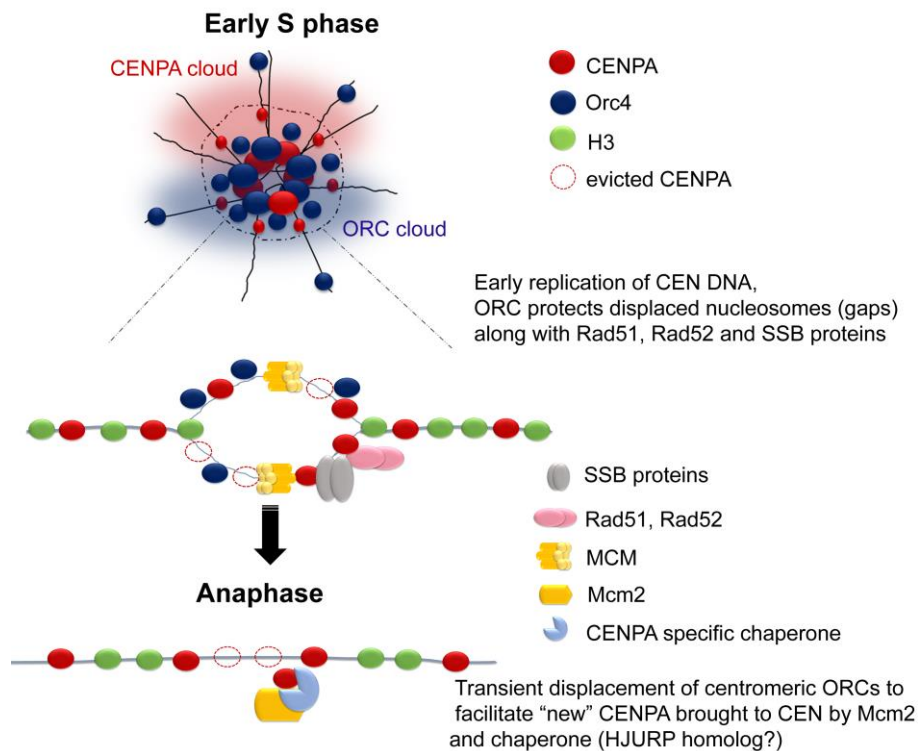
14

15

16

17

18

19

20

21

22

23

24

25

26

1 **Fig 8.**

2

3 **Fig 8. Propagation of CEN chromatin in *C. albicans*.** CENPA rich centromeres are surrounded by
4 the peripheral CENPA cloud in a compact chromatin environment. Along with the CENPA cloud, an
5 ORC cloud exists at the centromeres to facilitate timely recruitment of initiator proteins for early CEN
6 replication. The pericentromeres are interspersed with H3 and CENPA nucleosomes. In the early S
7 phase, parental CENPA molecules are evicted from the central core and the gaps created are either
8 occupied by a placeholder (unknown) or, the CEN DNA has to be protected by the physical proximity
9 and association of proteins like Orc4, along with Rad51, Rad52 and single stranded DNA binding
10 proteins. Orc4 maintains centromeric heterochromatin and facilitates early CEN DNA replication. The
11 pre-RC soon replicates other parts of the genome till the end of S phase. During anaphase, Orc4,
12 though otherwise constitutively associated with the CEN chromatin, undergoes a transient
13 dissociation to facilitate new CENPA loading by a specific chaperone (unknown) that is co-
14 chaperoned by Mcm2. Hence, the epigenetic marks are propagated to the subsequent cell cycles.

APPENDIX

A1: S-phase specific kinases (in continuation to Introduction page no. 4)**:**

Initiation of DNA replication requires initiation factors and S-phase kinases. There are two families of S-phase kinases in eukaryotes, Dbf4-dependant kinases (DDKs) and cyclin-dependant kinases (CDKs) (SANSAM *et al.* 2015). DDKs phosphorylate subunits of Mcm2-7 to activate the replicative helicase. DDK activity is conserved from yeast to humans. However, the mechanism of stimulation by CDKs in higher eukaryotes has diverged significantly. In *S. cerevisiae*, the S-CDKs phosphorylate Sld2 and Sld3 which is necessary for initiation (MASUMOTO *et al.* 2002). In metazoans, TRESLIN phosphorylation by S-CDK has been shown to regulate the length of S-phase (SANSAM *et al.* 2015). Re-initiation of DNA replication is prevented by a number of mechanisms (BLOW and DUTTA 2005). Once initiated, pre-RCs are disassembled from the origins as Mcm2-7 moves away from the origin (along the replication fork), releasing Cdt1 and Cdc6 from the origins. These proteins are degraded by proteolysis and exported from the nucleus. Such mechanisms prevent loading of Mcm2-7 onto the already fired origin (reviewed in (TRUONG and WU 2011). The metazoan specific protein Geminin inhibits pre-RC assembly by sequestering Cdt1 in an inactive complex such that it is unable to recruit Mcm2-7 (WOHLSCHLEGEL *et al.* 2000). At the end of mitosis, Geminin is degraded so that pre-RCs can be recruited at the G_1 phase (ARIAS and WALTER 2007).

A2: Methods to measure replication timing (in continuation to Introduction page no. 10)**:**

Replication timing has been measured across both yeast and mammalian genomes using various methods. Repli-seq is one such method that employs pulse labelling of asynchronous cells with a nucleotide analog such as BrdU, that labels replicating (nascent) DNA (RYBA *et al.* 2011; MARCHAL *et al.* 2018). This is followed by sorting of early and late S-phase cell fractions using flow cytometry. Immunoprecipitation using anti-BrdU antibodies yields nascent DNA that be sequenced. The ratio of the early and late nascent DNA fractions is used to calculate the replication timing. Another method employs sorting of G_1 and S-phase cells. The ratio of their DNA content using two dye microarray technology is a measure of the replication timing (KOREN *et al.* 2010; YAFFE *et al.* 2010).

A3: Replication timing ensures proper kinetochore assembly (in continuation to Introduction page no. 29)

In *S. cerevisiae*, CENs replicate during early S-phase. Kinetochores are tethered to the SPBs by microtubules during most part of the cell cycle. However, during a brief period in early S-

phase, centromeres are briefly detached from the microtubules and this time period coincides with centromere DNA replication. This brief disassembly of kinetochores is followed by reassembly and recapture by microtubules. These temporal mechanisms are especially useful to regulate kinetochore assembly and capture in organisms that undergo closed mitosis and have been reported in (KITAMURA *et al.* 2007).

A4: Sheared fragment sizes of chromatin obtained from *URA3* integrants grown in *URA3* expressed vs repressed conditions.
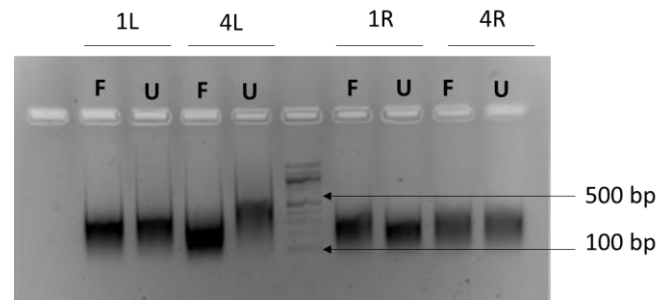

Figure A4: Agarose gel image depicting sheared fragment sizes of the chromatin obtained from the ChIP samples of the *URA3* integrants grown in *URA3* expressed (CM-Uri) vs repressed (CM+5-FOA) conditions. DNA from the starting material (SM) of the strains from the *URA3* integrants corresponding to the loci 1L, 4L, 1R, 4R were electrophoresed on a 2% agarose gel along with a 100 bp ladder. This was performed to detect any difference in the chromatin sizes between *URA3* expressed v/s repressed conditions. No observable difference between the chromatin sizes was detected between the 5-FOA and CM-Uri ChIPs for all the integrants tested.

A5: Melt curves of the primers used in ChIP-qPCR experiments:
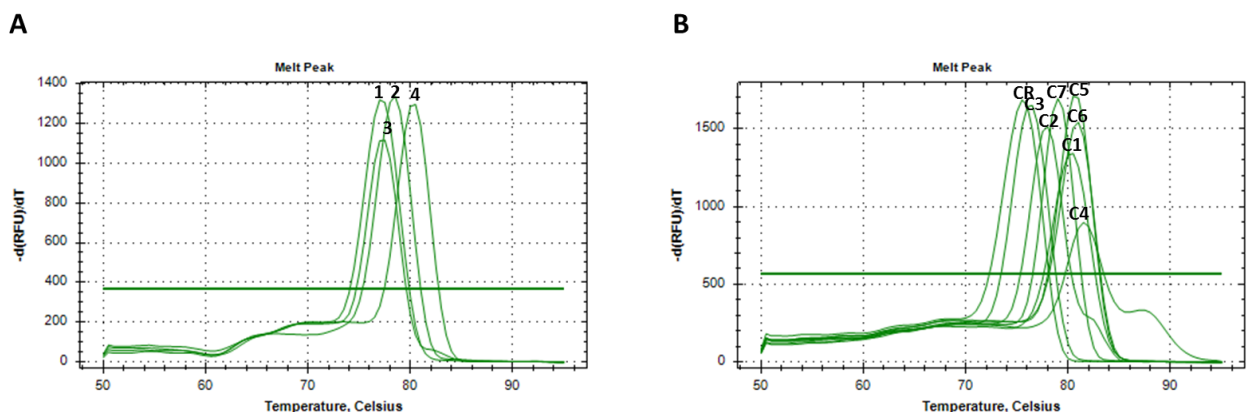

Figure A5: Melt curve analysis of the primer pairs used for CEN ChIP-qPCRs. (A) Sheared chromatin from wild type cells was amplified using the primers from *URA3* (1), *CEN7* (2), *nCEN7*-II (3), *Ctrl7* (4) and melt curve was plotted after 40 PCR cycles. **(B)** Melt curve analysis done for the primer sets used to amplify Orc4 ChIP DNA: C1-CR indicate melt curve peaks corresponding to the primers sets (C1_FP, RP/ C2_FP/RP… CR_FP/RP) from each of the eight chromosomes.

A6: Mcm2 protein levels are depleted in the conditional mutant LSK339

We have generated a conditional mutant of *MCM2* by replacing its endogenous promoter with the repressive *MET3* promoter in a heterozygous null background. We tagged the C-terminus of the only intact allele of *MCM2* with *TAP* to generate the strain LSK339 (*mcm2/MET3prMCM2-TAP(NAT)*). We measured the levels of Mcm2 upon growth of *C. albicans* cells in repressive media (in presence of 5mM methionine and 5mM cysteine). It was observed that after 6 h of growth in repressive media, the protein levels of Mcm2 were undetectable by western blot confirming the depletion.
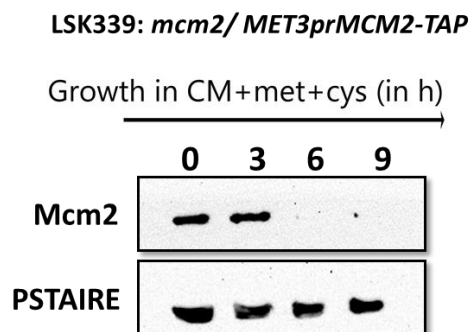

Figure A6: Time course depletion of Mcm2 in the conditional mutant LSK339. Western blot analysis using anti-Protein A antibodies indicates time course depletion of Mcm2 in the conditional mutant LSK339 when the strain was grown in the indicated time (0,3,6,9 h) in presence of 5mM methionine and 5mM cysteine. PSTAIRE was used as the loading control.

A7: A comparison of the ORC binding sites reported in Tsai *et al*., 2014 with the Orc4 binding sites reported in the present study

The previous study (Tsai *et al.* 2014) employed a microarray-based approach to identify genome-wide replication origins in *C. albicans*. The authors used antibodies against the whole-ORC complex of *S. cerevisiae* to pull down ORC proteins in *C. albicans*. Using this approach, they identified a total of 386 proposed origin of replication (proORIs), which were mostly located at intergenic regions and excluded from nucleosomes. They also observed a strong association between tRNA genes and some of the proORIs. Further, they classified some of the proORIs as bona fide ORIs, based on the detection of bubble arcs using 2D gel electrophoresis and by transformation of linear plasmids with cloned ARS fragments. They examined the presence of DNA consensus motifs associated with ORC binding sites using MEME SUITE and detected an AC-rich 15 bp within the proORIs. They also found a T-rich

motif within the centre of the proORI, overlapping with a nucleosome excluded region. They finally concluded that replication origins in *C. albicans* are defined by a proORI motif.

The putative replication origins identified in our study was performed using a ChIP-sequencing experiment with polyclonal antibodies against the Orc4 subunit of the *C. albicans* ORC. This was necessary since the mode of recognition of origin sequences in *C. albicans* might be different from that in *S. cerevisiae*. ChIP-sequencing revealed a total of 414 genomic sites and 3 mitochondrial binding sites of Orc4. Among these, 70% were located within gene bodies, contrary to the earlier report. We validated some of these by ChIP-qPCR. 2D gel electrophoresis will reveal if they are active origins, which is one of the future prospects of the study. We employed a *de novo* motif scan tool, DIVERSITY to examine the DNA consensus motifs associated with Orc4 binding sites. We could detect four DNA motifs, one of which was the tRNA motif reported earlier. Upon comparison with the replication timing data from an earlier report (KOREN *et al.* 2010), they were found to be localized within the early replicating regions of the genome. Overall, our study showed a 25% overlap in the ORC binding regions with the previous study (TSAI *et al.* 2014). The overlapping regions comprise of centromeres and tDNA associated origins, which are early replicating and highly enriched in Orc4.

Additional references

Arias, E. E., and J. C. Walter, 2007 Strength in numbers: preventing rereplication via multiple mechanisms in eukaryotic cells. Genes Dev 21**:** 497-518.

Blow, J. J., and A. Dutta, 2005 Preventing re-replication of chromosomal DNA. Nat Rev Mol Cell Biol 6**:** 476-486.

Kitamura, E., K. Tanaka, Y. Kitamura and T. U. Tanaka, 2007 Kinetochore microtubule interaction during S phase in Saccharomyces cerevisiae. Genes Dev 21**:** 3319-3330.

Koren, A., H. J. Tsai, I. Tirosh, L. S. Burrack, N. Barkai *et al.*, 2010 Epigenetically-inherited centromere and neocentromere DNA replicates earliest in S-phase. PLoS Genet 6**:** e1001068.

Marchal, C., T. Sasaki, D. Vera, K. Wilson, J. Sima *et al.*, 2018 Genome-wide analysis of replication timing by next-generation sequencing with E/L Repli-seq. Nat Protoc 13**:** 819-839.

Masumoto, H., S. Muramatsu, Y. Kamimura and H. Araki, 2002 S-Cdk-dependent phosphorylation of Sld2 essential for chromosomal DNA replication in budding yeast. Nature 415**:** 651-655.

Ryba, T., D. Battaglia, B. D. Pope, I. Hiratani and D. M. Gilbert, 2011 Genome-scale analysis of replication timing: from bench to bioinformatics. Nat Protoc 6**:** 870-895.

Sansam, C. G., D. Goins, J. C. Siefert, E. A. Clowdus and C. L. Sansam, 2015 Cyclin-dependent kinase regulates the length of S phase through TICRR/TRESLIN phosphorylation. Genes Dev 29**:** 555-566.

Truong, L. N., and X. Wu, 2011 Prevention of DNA re-replication in eukaryotic cells. J Mol Cell Biol 3**:** 13-22.

Tsai, H. J., J. A. Baller, I. Liachko, A. Koren, L. S. Burrack *et al.*, 2014 Origin replication complex binding, nucleosome depletion patterns, and a primary sequence motif can predict origins of replication in a genome with epigenetic centromeres. MBio 5**:** e01703-01714.

Wohlschlegel, J. A., B. T. Dwyer, S. K. Dhar, C. Cvetic, J. C. Walter *et al.*, 2000 Inhibition of eukaryotic DNA replication by geminin binding to Cdt1. Science 290**:** 2309-2312.

Yaffe, E., S. Farkash-Amar, A. Polten, Z. Yakhini, A. Tanay *et al.*, 2010 Comparative analysis of DNA replication timing reveals conserved large-scale chromosomal architecture. PLoS Genet 6**:** e1001011.