PRODUCTION OF A SYNTHETIC REPEAT ARRAY SYSTEM FOR ARTIFICIAL CENTROMERE FORMATION IN MAIZE &

EPIGENETIC REGULATION OF KINETOCHORE SIZE IN PLANTS

by

HAN ZHANG

(Under the Direction of R. Kelly Dawe)

ABSTRACT

Centromeres are specified by a histone H3 variant CENH3 and a DNA-binding protein CENPC, both of which are conserved components of the inner kinetochores. In animals, ectopically localized CENH3 or CENPC is sufficient to assemble de novo kinetochores. In this study, we tested whether similar observations can be achieved in maize to create artificial centromeres. To this end, we have developed a satellite repeat array, which is composed of five different binding modules and was designed to be used as a tethering site. We showed that it is possible to synthetically engineer and introduce megabase repeat arrays into the maize genome by biolistic transformation, and that the array was sufficient to efficiently tether fluorescent proteins to specific loci.

We utilized the array system to specifically tether kinetochore protein CENH3 and CENPC to the chromosome arms. We showed that although these proteins may be transiently targeted to ectopic loci, they tended to be unstable and insufficient to induce neocentromere

formation. Rather, transgenic CENH3 or CENPC fused with different tethering proteins were localized at endogenous centromeres and caused strong dominant-negative affects on native kinetochores. Our observations suggest that plant kinetochores are highly stable and rarely move.

We also investigated centromere size variations in the grass species. We showed that the size of the CENH3 domains is correlated with genome size divided by chromosome number. We also observed that CENH3 domains are flexible and can change rapidly in new environments where two species are crossed. These observations lead to the speculation that the size of the CENH3 domain is a reflection of how much kinetochore area is required to stabilize the spindle microtubules and that cell size has a strong influence on the kinetochore stable state.

INDEX WORDS: centromeres, kinetochores, CENH3, CENPC, neocentromeres, artificial chromosomes, ectopic tethering, epigenetic regulation, centromere size, maize

PRODUCTION OF A SYNTHETIC REPEAT ARRAY SYSTEM FOR ARTIFICIAL CENTROMERE FORMATION IN MAIZE & EPIGENETIC REGULATION OF KINETOCHORE SIZE IN PLANTS

by

HAN ZHANG

Bachelor of Science, Fudan University, P.R.China, 2005

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2011

PRODUCTION OF A SYNTHETIC REPEAT ARRAY SYSTEM FOR ARTIFICIAL CENTROMERE FORMATION IN MAIZE & EPIGENETIC REGULATION OF KINETOCHORE SIZE IN PLANTS

by

HAN ZHANG

Major Professor: R. Kelly Dawe

Committee: Richard Meagher

Wayne Parrott

Sarah Covert

Xiaoyu Zhang

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
December 2011

DEDICATION

To the ones I love and the ones who love me.

ACKNOWLEDGEMENTS

I am so grateful to my major advisor Dr. R. Kelly Dawe for his guidance and support. To me, Kelly is a mentor who helped me with my research and kept me motivated, a role model who set examples for me in my career, a friend who cheered me up when I'm frustrated, and a parent whom I can turn to when I need help. I can't imagine what it would be like without him.

I would also like to acknowledge my committee members Dr. Richard Meagher, Dr. Wayne Parrott, Dr. Sarah Covert and Dr. Zhang Xiaoyu for their critical comments and invaluable suggestions on my research projects. I spent quite some time in each lab conducting experiments and felt honored and fortunate to have them on my committee. I also thank members of their labs for help with my project, especially Barbara Artelt and Bao Phan, who did all the biolistic transformations in maize. I also thank Dr. Michael Bender who was on my committee for the first three years of my PHD.

I would like to thank all current and past colleagues in the Dawe lab for their help and friendship. They made the Dawe lab a fantastic place to stay and work. I will miss you all after I leave. Special thanks to my undergraduate worker Harold Garner and Margaret Glover for all the tedious work of genotyping and FISH.

I thank all members of the Plant Functional Genomics Center for their scientific contributions to my research project over the past years. Particularly, I thank Yujun Han for his help with bioinformatic studies, Michael R. McKain for help with phylogenic comparative

analysis, and Lin Guo for all the technical help and her continuous encouragement throughout my graduate life.

I also appreciate the resources of the UGA Genetics department and Plant Biology Department in every aspect of my research and training. Particularly, I would like to thank the plant biology greenhouse staff Michael Boyd and Kevin Tarner for taking care of my plants.

Finally, I would like to thank my friends and family for their endless and unconditional love. Although they are not anywhere nearby, they are always there for help when I need them. My gratitude to them is beyond words.

To all the ones who helped me through this journey, I would not have achieved this goal alone. Thank you!

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

Plant Artificial Chromosomes

Artificial chromosomes are high capacity DNA vectors that have the potential of improving the accuracy and efficiency of gene delivery, as well as opening possibilities for large scale genome engineering (Brown *et al.* 2000). Artificial chromosomes can be generated by two approaches: the "top-down" approach, which is based on existing chromosomes engineered to reduce their DNA content to a minimum size, and the "bottom-up" approach, which relies on de novo assembly of defined sequences that enable replication and segregation of the mini-chromosomes.

The "top-down" approach involves fragmentation or truncation of existing chromosomes and can be achieved by irradiation or integration of telomere sequences. Fragmentation of chromosomes by irradiation has been reported in both animals and plants to produce mini-chromosomes but this method is not widely used (Carine *et al.* 1986, Riera-Lizarazu *et al.* 2000). More commonly, chromosomes are truncated by integration of telomeric repeats, which can provide chromosomes with new telomeres (Farr *et al.* 1991). The resulting broken chromosomes are transmissible in both somatic and germ cells. Therefore, they are adopted to produce mini-chromosomes and used as platforms for artificial chromosome formation (Farr *et al.* 1995, Heller *et al.* 1996). Plant mini-chromosomes can also be achieved by telomere-mediated chromosomal

truncation, but the frequency of such events is low compared with that in mammalian cells (Yu *et al.* 2006, Yu *et al.* 2007).

The "bottom-up" approach requires in vitro assembly of three components for chromosome replication and maintenance: origins of replication, telomeres and centromeres. The first de novo assembly of artificial chromosomes was developed in *S.cerevisiae* by transforming plasmids containing all three DNA elements into yeast (Murray and Szostak 1983). Origins of replication are poorly defined in higher eukaryotes. It is generally presumed that large genomic DNA fragments contain origins of replication in some form. For this reason, only centromere-specific alpha satellites and telomeric DNAs were transfected into human cells and successfully produced human artificial chromosomes (Harrington *et al.* 1997, Ikeno *et al.* 1998, Henning *et al.* 1999).

Given the success of constructing human artificial chromosomes, several laboratories have carried out similar experiments in plants. In the first published experiments, rice was transformed with bacterial artificial chromosomes (BACs) containing maize centromeric repeats or rice centromeric satellites. Extensive cytological examination revealed only stable integration but not autonomous mini-chromosomes (Phan *et al.* 2007). In similar study, circular vectors containing satellite repeats, retrotransposon or other type of genomic DNA were delivered into maize by particle bombardment and were shown to be meiotically transmissible (Carlson *et al.* 2007). However, doubts remain about the size of the presumed mini-chromosomes and the mechanism of inheritance (Houben *et al.* 2008). In a third study, origins of replication and telomeric repeats were assembled into maize centromeric BACs and then transformed into maize (Ananiev *et al.* 2009). Mini-chromosomes were detected in the transformed plants but mainly resulted from genome rearrangement rather than de novo assembly of the input DNA (Ananiev *et*

al. 2009). Although these recent reports are encouraging, de novo assembly of plant artificial chromosomes similar to the ones described in humans has not yet been reported, which is likely due to the high complexity and limited understanding of plant centromere function.

Centromere Identity and Kinetochore Assembly

Centromeres are the primary constrictions on the metaphase chromosomes where kinetochores are assembled to interact with microtubules. In eukaryotes, centromeres are essential for proper chromosome segregation and inheritance of the genetic content.

Despite the conserved functions, centromeres in different organisms often vary in length, structure and sequence. For example, budding yeast *Saccharomyces cerevisiae* has a point centromere of 125 bp but in most higher eukaryotes, centromeres are composed of long tandem repeats spanning thousands to millions of base pairs (Black and Cleveland 2011). Compared to the size variation, sequence compositions of centromeres across species differ more dramatically. Until now no universal motif has been found among centromeric DNA. Occasionally the centromeric sequences evolve so rapidly that little homology can be detected even between closely related species. For instance, no homologous centromeric sequences are shared between sibling species of *Drosophila* (Lohe and Brutlag 1987).

Centromeres are best characterized in *Saccharomyces cerevisiae*, where a DNA segment as short as 125 bp is sufficient for centromere function. Centromeres of budding yeast are composed of three domains: highly conserved CDEI and CDEIII domains, and an AT-rich CDEII domain (Fitzgerald-Hayes *et al.* 1982). These DNA elements recruit a centromere binding complex CBF3 via sequence-specific interactions (Lechner and Carbon 1991).

However, in all other species studied, including fission yeast *Schizosaccharomyces pombe*, the presence of the centromeric DNA does not necessarily guarantee centromere activities. In human dicentric chromosomes, two centromeric satellite domains are present but only one is able to recruit kinetochore proteins (Earnshaw *et al.* 1989, Page *et al.* 1995, Sullivan and Schwartz 1995). Similar centromere inactivation has also been observed in plants to produce stable dicentric chromosomes (Han *et al.* 2006).

Centromeres can also form spontaneously in chromosomal regions lacking centromere-specific sequences. These ectopically established centromeres are referred to as "neocentromeres" (Voullaire *et al.* 1993, Karpen and Allshire 1997, Nasuda *et al.* 2005, Topp *et al.* 2009). Despite their DNA sequences, neocentromeres are capable of assembling kinetochore proteins for spindle attachment and chromosome movement. More recently, it has been reported that de novo centromeres can be assembled on synthetic sequences (Kiermaier *et al.* 2009, Lacefield *et al.* 2009, Barnhart *et al.* 2011, Gascoigne *et al.* 2011). The fact that ectopic centromeres can be formed independent of centromeric sequences suggests that centromeres are specified by epigenetic information.

Centromere Maintenance and Propagation

Functional centromeres are characterized by the underlying chromatin that contains a histone H3 variant called CENH3 (also known as Cse4 in yeast (Meluh *et al.* 1998), HTR12 in *Arabidopsis* (Talbert *et al.* 2002), CID in *Drosophila* (Henikoff *et al.* 2000), HCP-3 in *C.elegans* (Buchwitz *et al.* 1999) and CENPA in humans (Earnshaw and Rothfield 1985)). In all the eukaryotes studied, CENH3 replaces the canonical H3 only in the centromere cores. Particularly, only a portion of the centromeric satellite repeats are incorporated into the CENH3-

containing nucleosomes (Black and Cleveland 2011) and inactive centromeres of the dicentric chromosomes are excluded from the CENH3-specified chromatin (Earnshaw and Migeon 1985, Warburton *et al.* 1997, Han *et al.* 2006).

Like conventional histone H3, CENH3 is mainly composed of two domains (Figure 1.1): an amino-terminal tail that is dramatically divergent from H3, and a histone fold domain that is relatively well conserved (Henikoff *et al.* 2001). The N-terminal region of CENH3 is dispensable for centromere targeting (Wieland *et al.* 2004, Black *et al.* 2007) but recent discoveries suggest that it may be involved in CENH3 loading dynamics during meiosis (Ravi *et al.* 2011). In contrast, the histone fold domain of CENH3 is sufficient for centromere localization (Lermontova *et al.* 2006, Black *et al.* 2007). Specifically, the small loop1 and helix 2 within the histone core of CENH3, known as the CENPA centromere-targeting domain (CATD, Figure 1.1), confers centromere specification (Vermaak *et al.* 2002, Black *et al.* 2004).

Differences between CENH3 and H3 have been proposed to play important roles in specifying centromere identity. Sequence features that distinguish CENH3 from the canonical histone H3 include a more divergent core histone fold, a slightly longer loop 1 region and a highly varying N-terminal tail (Henikoff *et al.* 2001). The histone fold domains of CENH3 only share a 78% identity in closely related species like human and mouse, in contrast to human and mouse H3s, which are 96% conserved over their entire lengths (Baker and Rogers 2006). Human CENH3 contains two extra amino acids in the loop 1 region, which directly contacts nucleosomal DNA. Mutations of these two residues reduce CENH3 retention at centromeres (Tachiwana *et al.* 2011). The amino-terminal tail of CENH3 is highly divergent and shares no sequence homology with the tail of canonical H3. Besides, little conservation can be found in the amino-termini of the CENH3 proteins in different species (Malik and Henikoff 2003).

In addition to variations on the nucleotide level, CENH3 nucleosomes also wind DNA with a different geometry compared to that of usual histones. Such distinct physical property of centromeric chromatin has been proposed to provide the foundation for centromere identity (Dalai *et al.* 2007, Henikoff and Furuyama 2010). Models for CENH3 nucleosomes include a hexasome where H2A-H2B dimers are replaced with non-histone protein Scm3 (Mizuguchi *et al.* 2007); a hemisome containing only one copy of each histone component (H2A, H2B, CENH3 and H4) and inducing positive DNA supercoils (Dalal *et al.* 2007, Furuyama and Henikoff 2009); and a tetrasome which features the absence of H2A-H2B dimers (Williams *et al.* 2009). Crystal structures of human and yeast CENH3 suggest that the overall configuration of CENH3-containing nucleosomes resembles that of the canonical histone octamers (Sekulic *et al.* 2010, Cho and Harrison 2011, Tachiwana *et al.* 2011, Zhou *et al.* 2011). It remains undetermined whether bacterially expressed CENH3 used for crystallography resembles the atomic structure of CENH3 in vivo and whether CENH3 nucleosomes in other organisms possess similar chromatin architectures as observed in human and yeast.

Centromeric chromatin must be stably propagated so that the epigenetic information marked by the underlying nucleosomes can be maintained every time DNA replicates. In marked contrast to nucleosomes containing canonical H3, nucleosomes containing CENH3 are assembled independent of DNA replication (Shelby *et al.* 2000, Lermontova *et al.* 2006, Jansen *et al.* 2007). As a consequence, pre-existing CENH3s are diluted to half of their original concentration after replication. For part of the cell cycle, centromere chromatin must exit in an intermediate state. Newly synthesized CENH3 can be loaded to centromeres containing 'placeholding' H3 nucleosomes that are deposited during S phase, or nucleosome gaps that are left bare, or centromeric octamers composed of one molecule of CENH3 and one molecule of H3

(Black and Cleveland 2011). It has been revealed that in humans H3.3 is deposited at centromeres as a placeholder in S phase (Dunleavy *et al.* 2011). H3.3 is a histone H3 variant that usually replaces the canonical histone H3 (H3.1) at transcriptionally active regions (Ahmad and Henikoff 2002a, Ahmad and Henikoff 2002b). Further investigations need to determine whether the presence of H3.3 at centromeres links CENH3 deposition with transcription and how H3.3 nucleosomes are displaced when new CENH3 nucleosomes are assembled.

A key question about centromere propagation is how pre-nucleosomal CENH3 is loaded onto newly replicated DNA. CENH3 chaperones have been recently identified in yeast (Scm3), drosophila (CAL1) and human (HJURP) (Mizuguchi *et al.* 2007, Dunleavy *et al.* 2009, Foltz *et al.* 2009, Mellone *et al.* 2011). These proteins complex with soluble CENH3 and histone H4, transiently localize at centromeres in a cell-cycle-dependent manner (Mizuguchi *et al.* 2007, Dunleavy *et al.* 2009, Foltz *et al.* 2009, Cho and Harrison 2011, Mellone *et al.* 2011, Zhou *et al.* 2011). Although functionally conserved, little homology can be detected between known CENH3 chaperones. For example, CENH3 chaperones in human and yeast only share a small region of homology (Sanchez-Pulido *et al.* 2009) and drosophila CENH3 chaperone CAL1 is *Dipteran*-specific (Mellone *et al.* 2011). Little is known about how such non-homologous proteins perform analogous functions in these organisms.

It is generally agreed that the pre-existing centromeres direct the local assembly of new CENH3 nucleosomes. However, what marks the centromere position recognized by soluble CENH3 still awaits further investigation. It can be the unique structures of centromeric nucleosomes themselves (Zhou *et al.* 2011), or distinct histone modification patterns at the centromeres (Sullivan and Karpen 2004, Shi and Dawe 2006, Bergmann *et al.* 2011), or proteins that specifically target centromeres. In the model proposed by Harrison and colleague (Cho and

Harrison 2011), centromere localization of the yeast CENH3 chaperone Scm3 is determined by Ndc10, a kinetochore protein in budding yeast which binds to the centromeric DNA specifically. This hypothesis is further supported by recent studies showing that new CENH3 nucleosome assembly requires a key kinetochore protein CENPC (Erhardt *et al.* 2008, Carroll *et al.* 2010, Moree *et al.* 2011). CENPC is a structural protein that links the inner centromere to the outer kinetochore domain (see Kinetochore Assembly section for detailed descriptions on CENPC) (Przewloka *et al.* 2011, Screpanti *et al.* 2011). In human, CENPC recruits a CENH3 loading factor M18BP1, which interacts with CENH3 chaperone HJURP. Before CENH3 deposition occurs, M18BP1 is targeted to the centromeres through a direct association with CENPC. Depletion of CENPC prevents M18BP1 localization and inhibits CENH3 chromatin assembly (Moree *et al.* 2011). Since CENPC is a conserved protein that has been identified in all eukaryotes studied, including yeast, animal and plants (Saitoh *et al.* 1992, Brown *et al.* 1993, Dawe *et al.* 1999, Ogura *et al.* 2004), it is more likely to be a universal pathway for CENH3 deposition, although other mechanisms may also exist (Figure 1.2).

Centromere Establishment and Neocentromere Formation

Another central question in centromere biology is how centromeres are established. Studies to address this question mainly focus on neocentromeres. Several hypotheses have been proposed to explain the mechanism of neocentromere formation. It has been suggested that neocentromeres are formed because of the spreading of the original centromere activity. This idea is supported by the observations that experimentally occurring neocentromeres tend to form in proximity to endogenous centromeres (Maggert and Karpen 2001, Nasuda *et al.* 2005). However, human clinical data showed that the majority of human neocentromeres are formed at

the distal ends of chromosomes, suggesting that vicinity is not the sole reason for neocentromere formation (Marshall *et al.* 2008).

Histone modifications at centromeric chromatin are different from those at euchromatin or heterochromatin (Sullivan and Karpen 2004, Shi and Dawe 2006, Bergmann *et al.* 2011). For this reason, it has been proposed that the epigenetic state of the chromatin may influence neocentromere formation (Nakano *et al.* 2008). Nevertheless, no consistent conclusion has been drawn regarding what are the distinct epigenetic states that facilitate neocentromere formation. It is still unclear whether the boundaries between heterochromatin and euchromatin are hotspots for neocentromere formation (Olszak *et al.* 2011) or whether neocentromeres tend to form more frequently at subtelomeric regions (Ishii *et al.* 2008). Another discrepancy is whether the presence of transcription inhibits neocentromere formation (Lomiento *et al.* 2008, Ketel *et al.* 2009) or promotes centromere activity (Topp *et al.* 2004, Chueh *et al.* 2009, Choi *et al.* 2011).

A third hypothesis for neocentromere formation is the aberrant incorporation of CENH3 at ectopic sites. To test this idea several laboratories have conducted experiments to artificially target CENH3 to non-centromeric loci. It has been reported in multiple organisms that as the expression level of CENH3 increased, the proteins spread to chromosome arms (Van Hooser *et al.* 2001, Heun *et al.* 2006, Collins *et al.* 2007, Gascoigne *et al.* 2011). Ectopic CENH3 was able to recruit at least a subset of kinetochore components, including kinetochore protein CENPC (Van Hooser *et al.* 2001, Gascoigne *et al.* 2011). In some rare cases, CENH3 localized on the chromosome arms was sufficient to assemble nearly complete kinetochores and mediate poleward movement of otherwise acentric fragments (Heun *et al.* 2006, Olszak *et al.* 2011).

Recently, another study re-examined this issue by assembling CENH3 nucleosomes on magnetic beads. The authors showed that additional kinetochore components were recruited to

the beads and that de novo kinetochores capable of stabilizing microtubules were formed in vitro (Guse *et al.* 2011). It has also been reported that targeting HJURP, the human CENH3 chaperone to synthetic repeats on the chromosome arms is sufficient to induce neocentromere formation (Barnhart *et al.* 2011). These observations suggest that CENH3 specifies the epigenetic state of the chromatin and provides the foundation for neocentromere formation.

Neocentromeres can be assembled de novo if CENH3-containing nucleosomes are established artificially, but how naturally occurring neocentromeres are established is still largely unknown. Recent discoveries shed light on this question by showing that ubiquitin-mediated proteolysis is involved in neocentromere formation in nature (Collins *et al.* 2004, Hewawasam *et al.* 2010, Ranjitkar *et al.* 2010). These authors showed that in budding yeast, an E3 ubiquitin ligase called Psh1 regulates excess CENH3 and restricts CENH3 localization. Ectopically localized CENH3s can be the ubiquitylated by Psh1 and targeted for proteolysis while centromerically localized CENH3s are protected by their chaperones from degradation. In *psh1* mutants, however, CENH3 localizes to the euchromatic regions when overexpressed (Hewawasam *et al.* 2010, Ranjitkar *et al.* 2010). Even under regular circumstances, CENH3 can be targeted to non-centromeric regions. For example, CENH3s are recruited to double-strand DNA breaks in human and mouse cells (Zeitlin *et al.* 2009). These ectopically localized proteins are likely to result in stable neocentromeres when the proteolysis pathway is disrupted.

Kinetochore Assembly

Kinetochores are the proteinaceous structures assembled on the centromeres where spindle fibers attach to during cell divisions. In the kinetochores of yeast *S. cerevisiae* over 70 protein components have been identified, most of which are evolutionarily conserved (Meraldi *et*

al. 2006). The kinetochores are divided into two domains: outer kinetochore proteins that are assembled at the centromeres only during mitosis (Maiato *et al.* 2004) and inner kinetochore domains which bind to centromeric chromatin constitutively (Amor *et al.* 2004, Cheeseman and Desai 2008). Fifteen proteins have been found to be components of the constitutive centromere-associated network (CCAN) in human (Foltz *et al.* 2006, Hori *et al.* 2008), but among them only CENH3 and CENPC are highly conserved in all eukaryotes.

CENH3 replaces conventional H3 in the functional centromeres and provides the platform for kinetochore assembly. CENPC binds to the extreme C-terminus of CENH3 (consist of loop 2 and helix 3 of the histone fold, Figure 1.1) (Carroll *et al.* 2010) and serves as a principle linker between the inner centromere chromatin and outer kinetochores (Milks *et al.* 2009, Tanaka *et al.* 2009). In animals, CENPC is composed of three domains (Figure 1.3), a variant N-terminus, a poorly conserved central domain and a conserved C-terminal domain. The N-terminal domain of CENPC binds directly to the minichromosome instability 12 (Mis12) complex, a four-subunit complex at the outer kinetochores (Screpanti *et al.* 2011). It has also been revealed recently that this part of CENPC is sufficient for the recruitment of additional kinetochore components (Gascoigne *et al.* 2011, Przewloka *et al.* 2011). The central domain binds DNA in a sequence-independent manner and targets to the centromeres (Yang *et al.* 1996, Song *et al.* 2002, Trazzi *et al.* 2002, Milks *et al.* 2009). It has been suggested recently that localization of the central domain of CENPC at the centromeres is mediated via a direct and specific interaction with CENH3 nucleosomes rather than non-specific DNA binding (Carroll *et al.* 2010). The C-terminal domain can be further divided to two regions based on their sequence homology with yeast CENPC homolog Mif2 and are named domain II and III respectively (Brown 1995). Domain II contains a highly conserved 23 amino acid CENPC signature motif

(Brown 1995, Talbert *et al.* 2004) and domain III induces CENPC dimerization. Both domains are necessary for CENPC binding to the CENH3 loading factor M18BP1 and have a key role in CENH3 nucleosome assembly (Moree *et al.* 2011).

Plant homologs of CENPC have been identified in maize and *Arabidopsis* based on sequence homology within the 23 amino acid region of the CENPC motif (Dawe *et al.* 1999, Ogura *et al.* 2004). Outside the defining motif, a small region of CENPC has been duplicated in the grass species (Talbert *et al.* 2004). Study in maize demonstrated that this region of CENPC binds both DNA and RNA in a sequence independent manner (Du *et al.* 2010). The authors also showed that removal or replacement of the DNA/RNA binding module with HIV integrase binding domain caused only a partial delocalization of CENPC *in vivo* (Du *et al.* 2010), indicating that another functional domain with centromere targeting capacity exists in plants. Considering that CENPC is a conserved structural protein in the inner kinetochores of all the eukaryotes, it is likely that plant CENPC performs similar functions as animals.

Given the importance of CENPC, several laboratories have tested the sufficiency of CENPC in kinetochore assembly. General overexpression of CENPC in chicken cell lines caused excess CENPC localization on chromosome arms but neocentromere formation was not observed (Fukagawa *et al.* 1999). In contrast, CENPC specifically targeted to non-centromeric loci is sufficient to assemble de novo kinetochores (Gascoigne *et al.* 2011, Przewloka *et al.* 2011). In one case, the N-terminal protein binding domain of CENPC was fused with a centrosome targeting protein (Przewloka *et al.* 2011). It was found that core components of the outer kinetochore but not CENH3 were recruited to the centrosomes (Przewloka *et al.* 2011). In another study, CENPC was targeted to synthetic binding motifs integrated on chromosome arms and was shown to interact with microtubules and function in chromosome segregation

(Gascoigne *et al.* 2011). These observations suggest that CENPC is a structural platform for kinetochore assembly and de novo kinetochores can be formed independent of the underlying CENH3 nucleosomes.

While CENH3 and CENPC provide the scaffold for kinetochore assembly, outer kinetochore components provide sites for microtubule attachment (Figure 1.4). Extensive genetic and structural analysis have revealed that the kinetochore-microtubule interface is mediated through a four-protein complex called nuclear division cycle 80 (Ndc80) (Cheeseman and Desai 2008). The Ndc80 complex is among the most conserved features of all kinetochores and forms a rod-like structure with globular regions at either end (Ciferri *et al.* 2005, Wei *et al.* 2005, Wei *et al.* 2006, Wei *et al.* 2007). One end of the rod is composed of Ndc80 and nuclear filament-containing protein 2 (Nuf2) (McCleland *et al.* 2003, DeLuca *et al.* 2005). The Ndc80/Nuf2 complex binds directly to tubulin monomers by electrostatic interactions (Ciferri *et al.* 2008). The other end of the rod is composed of spindle pole body component 24 (Spc24) and spindle pole body component 25 (Spc25), which face the inner kinetochore and interact with the MIS12 complex (Janke *et al.* 2001, Wigge and Kilmartin 2001, Bharadwaj *et al.* 2004). The Mis12 complex directly binds CENPC and serves as a protein interaction hub for outer kinetochore assembly (Maskell *et al.* 2010, Petrovic *et al.* 2010, Screpanti *et al.* 2011).

Homologs of Ndc80 and Mis12 have been identified in plants and localize outside the inner kinetochores (Sato *et al.* 2005, Du and Dawe 2007, Li and Dawe 2009). Other components of the Ndc80 complex and Mis12 complex can also be detected in plants by sequence similarity (Meraldi *et al.* 2006) but have not yet been shown to be kinetochore-localized or to interact directly with known kinetochore proteins.

Chromatin Protein Targeting by Tethering

Direct observations of chromosome dynamics are critical to our understanding of complicated cellular processes such as mitosis. However, research in this field is largely limited by our ability to visualize chromosome movement in vivo and the difficulties in knocking out the key components involved in these processes.

Utilization of the high affinity interactions between DNA binding proteins and their target sequences now made it possible to visualize and study specific regions of chromatin in living cells. Belmont and colleagues pioneered this approach by fusing lactose repressor (LacI) to a fluorescent protein (Robinett *et al.* 1996, Straight *et al.* 1996). The repressor protein can recognize and bind its corresponding lac operator (LacO), which is integrated as tandem repeat arrays into a specific locus of the chromatin. In this way, fluorescent proteins are tethered to the chromosomal regions and can be observed with fluorescence microscopes. First developed in mammalian cells (Robinett *et al.* 1996) and yeast (Straight *et al.* 1996) to visualize chromatin organization and dynamics, the system was rapidly adopted in other organisms such as bacteria (Webb *et al.* 1997), *C.elegans* (Carmi *et al.* 1998), *Drosophila* (Vazquez *et al.* 2001) and plants (Kato and Lam 2001) (reviewed in (Belmont 2001, Janicki and Spector 2003)).

The tethering strategy was then adapted to target chromatin-associated proteins to specific loci for functional analyses, which include but are not limited to: heterochromatin proteins to modify chromatin structure and mediate gene silencing (Danzer and Wallrath 2004); transcriptional activator/silencer to inactivate a synthetic human kinetochore (Nakano *et al.* 2008); and histone methyltransferase to induce heterochromatin assembly at euchromatic loci (Kagansky *et al.* 2009). More Recently, in budding yeast and animal cell culture, it has been

reported that kinetochore proteins fused with LacI can be tethered to the LacO arrays to initiate the assembly of nearly complete kinetochores (Kiermaier *et al.* 2009, Lacefield *et al.* 2009, Barnhart *et al.* 2011, Gascoigne *et al.* 2011). These studies demonstrated the feasibility to assemble kinetochores on specific sequences and avoided uncertainties associated with epigenetics.

Purpose of this Study

Centromeres are viewed as the key to producing artificial chromosomes, which enable large-scale genome engineering without multiple rounds of transformation. The ultimate goal of this research is to construct de novo kinetochores by the tethering approach to facilitate the formation of plant artificial chromosomes. Previously published works in the animal literature have already illustrated the potential of constructing artificial centromeres by tethering. In this study we intended to conduct similar tests in plants.

We chose maize to study kinetochore organization and artificial centromere formation for several reasons. First, maize chromosomes are large and can be visualized easily under microscopes. A collection of antibodies against kinetochore proteins were made available in maize (Dawe *et al.* 1999, Yu *et al.* 1999, Zhong *et al.* 2002, Du and Dawe 2007, Li and Dawe 2009), which, combined with high resolution 3D light microscope, greatly facilitate the study of plant kinetochores. Besides, functional centromeres have been precisely mapped in at least two maize chromosomes (Wolfgruber *et al.* 2009), and the underlying nucleosomal structure helpful for centromere formation has been well characterized (Gent *et al.* 2011). Last but not the least, the purpose of this study is to construct artificial chromosomes, which will eventually be used in

crops for genome engineering. With its genome fully sequenced (Schnable *et al.* 2009) and mature techniques developed for transformation, maize, as a model organism, is the best choice for this kind of study.

Protein tethering has been successfully applied in *Arabidopsis* (Kato and Lam 2001, Matzke *et al.* 2003) but has not been achieved in maize. Four different DNA binding proteins, LacI (Kato and Lam 2001), tetracycline repressor (TetR) (Matzke *et al.* 2003), Gal4 (Aoyama and Chua 1997) and LexA (Moore *et al.* 2006) have been shown to be capable of binding their target DNA in a specific and high-affinity manner. In this study, we generated a very long synthetic tandem array containing all four DNA binding modules, and directly transformed it into maize by particle bombardment. We described three transformation events where tandem arrays up to 1.1 Mb were integrated on the chromosome arms as single intact loci. The utility of the arrays was demonstrated by showing that the arrays efficiently recruit a LacI-YFP fusion protein to single large fluorescent spots that are easily visible by standard microscopy. The details of this study are described in **Chapter 2**.

Kinetochore proteins CENH3 and CENPC have been identified and characterized in *Arabidopsis* and maize (Dawe *et al.* 1999, Talbert *et al.* 2002, Zhong *et al.* 2002, Ogura *et al.* 2004). In **Chapter 3** of the dissertation, we described the results of tethering CENH3 or CENPC with different binding domains in the array system and showed that plants kinetochore proteins are tightly regulated to prevent the formation of ectopic centromeres. Transgenic kinetochore proteins with fusion tags are more likely to destabilize the function of existing kinetochores rather than to induce the formation of neocentromeres. We also observed that CENH3 or CENPC may be transiently localized to non-centromeric loci, but they rarely resulted in neocentromeres, and tended to be smaller in size compared to endogenous kinetochores. This

observation led to our analysis on kinetochore size in different grass species, which is discussed in **Chapter 4**.

In recent months a number of exciting papers have demonstrated the advantage of using protein tethering to confer a genetic specification mechanism to centromeric regions (Barnhart *et al.* 2011, Gascoigne *et al.* 2011, Przewloka *et al.* 2011). Our study was designed using a similar approach, but we did not observe any neocentromere formation. The differences between our data and the published work in animal cell cultures can be a reflection of distinct mechanisms of kinetochore organization in varied species. They can also be attributed to our limited knowledge of plant kinetochores. Only a few kinetochore proteins have been identified in plants (Houben and Schubert 2003). How these proteins are recruited to the centromeres and how they interact with each other is still unclear. We proposed experiments to extend our understanding of plant kinetochore formation and neocentromere establishment in **Chapter 5**, with expectations that these experiments will improve the chance of artificial centromere formation in maize.

References

Ahmad, K. and Henikoff, S. (2002a) Histone H3 variants specify modes of chromatin assembly. *Proc Natl Acad Sci U S A*, **99 Suppl 4**, 16477-16484.

Ahmad, K. and Henikoff, S. (2002b) The histone variant H3.3 marks active chromatin by replication-independent nucleosome assembly. *Mol Cell*, **9**, 1191-1200.

Amor, D.J., Kalitsis, P., Sumer, H. and Choo, K.H. (2004) Building the centromere: from foundation proteins to 3D organization. *Trends Cell Biol*, **14**, 359-368.

Ananiev, E.V., Wu, C., Chamberlin, M.A., Svitashev, S., Schwartz, C., Gordon-Kamm, W. and Tingey, S. (2009) Artificial chromosome formation in maize (Zea mays L.). *Chromosoma*, **118**, 157-177.

Aoyama, T. and Chua, N.H. (1997) A glucocorticoid-mediated transcriptional induction system in transgenic plants. *Plant J*, **11**, 605-612.

Baker, R.E. and Rogers, K. (2006) Phylogenetic analysis of fungal centromere H3 proteins. *Genetics*, **174**, 1481-1492.

Barnhart, M.C., Kuich, P.H., Stellfox, M.E., Ward, J.A., Bassett, E.A., Black, B.E. and Foltz, D.R. (2011) HJURP is a CENP-A chromatin assembly factor sufficient to form a functional de novo kinetochore. *J Cell Biol*, **194**, 229-243.

Belmont, A.S. (2001) Visualizing chromosome dynamics with GFP. *Trends Cell Biol*, **11**, 250-257.

Bergmann, J.H., Rodriguez, M.G., Martins, N.M., Kimura, H., Kelly, D.A., Masumoto, H., . . . Earnshaw, W.C. (2011) Epigenetic engineering shows H3K4me2 is required for

HJURP targeting and CENP-A assembly on a synthetic human kinetochore. *EMBO J*, **30**, 328-340.

Bharadwaj, R., Qi, W. and Yu, H. (2004) Identification of two novel components of the human NDC80 kinetochore complex. *J Biol Chem*, **279**, 13076-13085.

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

Black, B.E., Foltz, D.R., Chakravarthy, S., Luger, K., Woods, V.L., Jr. and Cleveland, D.W. (2004) Structural determinants for generating centromeric chromatin. *Nature*, **430**, 578-582.

Black, B.E., Jansen, L.E., Maddox, P.S., Foltz, D.R., Desai, A.B., Shah, J.V. and Cleveland, D.W. (2007) Centromere identity maintained by nucleosomes assembled with histone H3 containing the CENP-A targeting domain. *Mol Cell*, **25**, 309-322.

Brown, M.T. (1995) Sequence similarities between the yeast chromosome segregation protein Mif2 and the mammalian centromere protein CENP-C. *Gene*, **160**, 111-116.

Brown, M.T., Goetsch, L. and Hartwell, L.H. (1993) MIF2 is required for mitotic spindle integrity during anaphase spindle elongation in Saccharomyces cerevisiae. *J Cell Biol*, **123**, 387-403.

Brown, W.R., Mee, P.J. and Hong Shen, M. (2000) Artificial chromosomes: ideal vectors? *Trends Biotechnol*, **18**, 218-223.

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

Carine, K., Solus, J., Waltzer, E., Manch-Citron, J., Hamkalo, B.A. and Scheffler, I.E. (1986) Chinese hamster cells with a minichromosome containing the centromere region of human chromosome 1. *Somat Cell Mol Genet*, **12**, 479-491.

Carlson, S.R., Rudgers, G.W., Zieler, H., Mach, J.M., Luo, S., Grunden, E., . . . Preuss, D. (2007) Meiotic transmission of an in vitro-assembled autonomous maize minichromosome. *PLoS Genet*, **3**, 1965-1974.

Carmi, I., Kopczynski, J.B. and Meyer, B.J. (1998) The nuclear hormone receptor SEX-1 is an X-chromosome signal that determines nematode sex. *Nature*, **396**, 168-173.

Carroll, C.W., Milks, K.J. and Straight, A.F. (2010) Dual recognition of CENP-A nucleosomes is required for centromere assembly. *J Cell Biol*, **189**, 1143-1155.

Cheeseman, I.M. and Desai, A. (2008) Molecular architecture of the kinetochore-microtubule interface. *Nat Rev Mol Cell Biol*, **9**, 33-46.

Cho, U.S. and Harrison, S.C. (2011) Recognition of the centromere-specific histone Cse4 by the chaperone Scm3. *Proc Natl Acad Sci U S A*, **108**, 9367-9371.

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

Chueh, A.C., Northrop, E.L., Brettingham-Moore, K.H., Choo, K.H. and Wong, L.H. (2009) LINE retrotransposon RNA is an essential structural and functional epigenetic component of a core neocentromeric chromatin. *PLoS Genet*, **5**, e1000354.

Ciferri, C., De Luca, J., Monzani, S., Ferrari, K.J., Ristic, D., Wyman, C., . . . Musacchio, A. (2005) Architecture of the human ndc80-hec1 complex, a critical constituent of the outer kinetochore. *J Biol Chem*, **280**, 29088-29095.

Ciferri, C., Pasqualato, S., Screpanti, E., Varetti, G., Santaguida, S., Dos Reis, G., . . . Musacchio, A. (2008) Implications for kinetochore-microtubule attachment from the structure of an engineered Ndc80 complex. *Cell*, **133**, 427-439.

Collins, K.A., Camahort, R., Seidel, C., Gerton, J.L. and Biggins, S. (2007) The overexpression of a Saccharomyces cerevisiae centromeric histone H3 variant mutant protein leads to a defect in kinetochore biorientation. *Genetics*, **175**, 513-525.

Collins, K.A., Furuyama, S. and Biggins, S. (2004) Proteolysis contributes to the exclusive centromere localization of the yeast Cse4/CENP-A histone H3 variant. *Curr Biol*, **14**, 1968-1972.

Dalai, Y., Furuyama, T., Vermaak, D. and Henikoff, S. (2007) Structure, dynamics, and evolution of centromeric nucleosomes. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 15974-15981.

Dalal, Y., Wang, H., Lindsay, S. and Henikoff, S. (2007) Tetrameric structure of centromeric nucleosomes in interphase Drosophila cells. *PLoS Biol*, **5**, e218.

Danzer, J.R. and Wallrath, L.L. (2004) Mechanisms of HP1-mediated gene silencing in Drosophila. *Development*, **131**, 3571-3580.

Dawe, R.K., Reed, L.M., Yu, H.G., Muszynski, M.G. and Hiatt, E.N. (1999) A maize homolog of mammalian CENPC is a constitutive component of the inner kinetochore. *Plant Cell*, **11**, 1227-1238.

DeLuca, J.G., Dong, Y., Hergert, P., Strauss, J., Hickey, J.M., Salmon, E.D. and McEwen, B.F. (2005) Hec1 and nuf2 are core components of the kinetochore outer plate essential for organizing microtubule attachment sites. *Mol Biol Cell*, **16**, 519-531.

Du, Y. and Dawe, R.K. (2007) Maize NDC80 is a constitutive feature of the central kinetochore. *Chromosome Res*, **15**, 767-775.

Du, Y., Topp, C.N. and Dawe, R.K. (2010) DNA binding of centromere protein C (CENPC) is stabilized by single-stranded RNA. *PLoS Genet*, **6**, e1000835.

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

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

Earnshaw, W.C. and Migeon, B.R. (1985) Three related centromere proteins are absent from the inactive centromere of a stable isodicentric chromosome. *Chromosoma*, **92**, 290-296.

Earnshaw, W.C., Ratrie, H. and Stetten, G. (1989) Visualization of centromere proteins CENP-B and CENP-C on a stable dicentric chromosome in cytological spreads. *Chromosoma*, **98**, 1-12.

Earnshaw, W.C. and Rothfield, N. (1985) Identification of a family of human centromere proteins using autoimmune sera from patients with scleroderma. *Chromosoma*, **91**, 313-321.

Erhardt, S., Mellone, B.G., Betts, C.M., Zhang, W., Karpen, G.H. and Straight, A.F. (2008) Genome-wide analysis reveals a cell cycle-dependent mechanism controlling centromere propagation. *J Cell Biol*, **183**, 805-818.

Farr, C., Fantes, J., Goodfellow, P. and Cooke, H. (1991) Functional reintroduction of human telomeres into mammalian cells. *Proc Natl Acad Sci U S A*, **88**, 7006-7010.

Farr, C.J., Bayne, R.A., Kipling, D., Mills, W., Critcher, R. and Cooke, H.J. (1995) Generation of a human X-derived minichromosome using telomere-associated chromosome fragmentation. *EMBO J*, **14**, 5444-5454.

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

Foltz, D.R., Jansen, L.E., Bailey, A.O., Yates, J.R., 3rd, Bassett, E.A., Wood, S., . . . Cleveland, D.W. (2009) Centromere-specific assembly of CENP-a nucleosomes is mediated by HJURP. *Cell*, **137**, 472-484.

Foltz, D.R., Jansen, L.E., Black, B.E., Bailey, A.O., Yates, J.R., 3rd and Cleveland, D.W. (2006) The human CENP-A centromeric nucleosome-associated complex. *Nat Cell Biol*, **8**, 458-469.

Fukagawa, T., Pendon, C., Morris, J. and Brown, W. (1999) CENP-C is necessary but not sufficient to induce formation of a functional centromere. *EMBO J*, **18**, 4196-4209.

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

Gascoigne, K.E., Takeuchi, K., Suzuki, A., Hori, T., Fukagawa, T. and Cheeseman, I.M. (2011) Induced Ectopic Kinetochore Assembly Bypasses the Requirement for CENP-A Nucleosomes. *Cell*, **145**, 410-422.

Gent, J.I., Schneider, K.L., Topp, C.N., Rodriguez, C., Presting, G.G. and Dawe, R.K. (2011) Distinct influences of tandem repeats and retrotransposons on CENH3 nucleosome positioning. *Epigenetics Chromatin*, **4**, 3.

Guse, A., Carroll, C.W., Moree, B., Fuller, C.J. and Straight, A.F. (2011) In vitro centromere and kinetochore assembly on defined chromatin templates. *Nature*.

Han, F., Lamb, J.C. and Birchler, J.A. (2006) High frequency of centromere inactivation resulting in stable dicentric chromosomes of maize. *Proc Natl Acad Sci U S A*, **103**, 3238-3243.

Harrington, J.J., Bokkelen, G.V., Mays, R.W., Gustashaw, K. and Willard, H.F. (1997) Formation of de novo centromeres and construction of first-generation human artificial chromosomes. *Nature Genetics*, **15**, 345-355.

Heller, R., Brown, K.E., Burgtorf, C. and Brown, W.R. (1996) Mini-chromosomes derived from the human Y chromosome by telomere directed chromosome breakage. *Proc Natl Acad Sci U S A*, **93**, 7125-7130.

Henikoff, S., Ahmad, K. and Malik, H.S. (2001) The centromere paradox: stable inheritance with rapidly evolving DNA. *Science*, **293**, 1098-1102.

Henikoff, S., Ahmad, K., Platero, J.S. and Steensel, B.V. (2000) Heterochromatic deposition of centromeric histone H3-like proteins. *Proc. Natl. Acad. Sci. USA*, **97**, 716-721.

Henikoff, S. and Furuyama, T. (2010) Epigenetic Inheritance of Centromeres. *Cold Spring Harb Symp Quant Biol*.

Henning, K.A., Novotny, E.A., Compton, S.T., Guan, X.Y., Liu, P.P. and Ashlock, M.A. (1999) Human artificial chromosomes generated by modification of a yeast artificial chromosome containing both human alpha satellite and single-copy DNA sequences. *Proc Natl Acad Sci U S A*, **96**, 592-597.

Heun, P., Erhardt, S., Blower, M.D., Weiss, S., Skora, A.D. and Karpen, G.H. (2006) Mislocalization of the Drosophila centromere-specific histone CID promotes formation of functional ectopic kinetochores. *Dev Cell*, **10**, 303-315.

Hewawasam, G., Shivaraju, M., Mattingly, M., Venkatesh, S., Martin-Brown, S., Florens, L., . . . Gerton, J.L. (2010) Psh1 is an E3 ubiquitin ligase that targets the centromeric histone variant Cse4. *Mol Cell*, **40**, 444-454.

Hori, T., Amano, M., Suzuki, A., Backer, C.B., Welburn, J.P., Dong, Y., . . . Fukagawa, T. (2008) CCAN makes multiple contacts with centromeric DNA to provide distinct pathways to the outer kinetochore. *Cell*, **135**, 1039-1052.

Houben, A., Dawe, R.K., Jiang, J. and Schubert, I. (2008) Engineered plant minichromosomes: a bottom-up success? *Plant Cell*, **20**, 8-10.

Houben, A. and Schubert, I. (2003) DNA and proteins of plant centromeres. *Curr Opin Plant Biol*, **6**, 554-560.

Ikeno, M., Grimes, B., Okazaki, T., Nakano, M., Saitoh, K., Hoshino, H., . . . Masumoto, H. (1998) Construction of YAC-based mammalian artificial chromosomes. *Nat. Biotech.*, **16**, 431-439.

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

Janicki, S.M. and Spector, D.L. (2003) Nuclear choreography: interpretations from living cells. *Curr Opin Cell Biol*, **15**, 149-157.

Janke, C., Ortiz, J., Lechner, J., Shevchenko, A., Magiera, M.M., Schramm, C. and Schiebel, E. (2001) The budding yeast proteins Spc24p and Spc25p interact with Ndc80p and Nuf2p at the kinetochore and are important for kinetochore clustering and checkpoint control. *EMBO J*, **20**, 777-791.

Jansen, L.E., Black, B.E., Foltz, D.R. and Cleveland, D.W. (2007) Propagation of centromeric chromatin requires exit from mitosis. *J Cell Biol*, **176**, 795-805.

Kagansky, A., Folco, H.D., Almeida, R., Pidoux, A.L., Boukaba, A., Simmer, F., . . . Allshire, R.C. (2009) Synthetic heterochromatin bypasses RNAi and centromeric repeats to establish functional centromeres. *Science*, **324**, 1716-1719.

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

Kato, N. and Lam, E. (2001) Detection of chromosomes tagged with green fluorescent protein in live Arabidopsis thaliana plants. *Genome Biol*, **2**, RESEARCH0045.

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

Kiermaier, E., Woehrer, S., Peng, Y., Mechtler, K. and Westermann, S. (2009) A Dam1-based artificial kinetochore is sufficient to promote chromosome segregation in budding yeast. *Nat Cell Biol*, **11**, 1109-1115.

Lacefield, S., Lau, D.T. and Murray, A.W. (2009) Recruiting a microtubule-binding complex to DNA directs chromosome segregation in budding yeast. *Nat Cell Biol*, **11**, 1116-1120.

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

Lermontova, I., Schubert, V., Fuchs, J., Klatte, S., Macas, J. and Schubert, I. (2006) Loading of Arabidopsis centromeric histone CENH3 occurs mainly during G2 and requires the presence of the histone fold domain. *Plant Cell*, **18**, 2443-2451.

Li, X. and Dawe, R.K. (2009) Fused sister kinetochores initiate the reductional division in meiosis I. *Nat Cell Biol*, **11**, 1103-1108.

Lohe, A.R. and Brutlag, D.L. (1987) Identical satellite DNA sequences in sibling species of Drosophila. *J Mol Biol*, **194**, 161-170.

Lomiento, M., Jiang, Z., D'Addabbo, P., Eichler, E.E. and Rocchi, M. (2008) Evolutionary-new centromeres preferentially emerge within gene deserts. *Genome Biol*, **9**, R173.

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

Maiato, H., DeLuca, J., Salmon, E.D. and Earnshaw, W.C. (2004) The dynamic kinetochore-microtubule interface. *J Cell Sci*, **117**, 5461-5477.

Malik, H.S. and Henikoff, S. (2003) Phylogenomics of the nucleosome. *Nat Struct Biol*, **10**, 882-891.

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

Maskell, D.P., Hu, X.W. and Singleton, M.R. (2010) Molecular architecture and assembly of the yeast kinetochore MIND complex. *J Cell Biol*, **190**, 823-834.

Matzke, A.J.M., van der Winden, J. and Matzke, M. (2003) Tetracycline operator/repressor system to visualize fluorescence-tagged T-DNAs in interphase nuclei of Arabidopsis. *Plant Mol Biol Rep*, **21**, 9-19.

McCleland, M.L., Gardner, R.D., Kallio, M.J., Daum, J.R., Gorbsky, G.J., Burke, D.J. and Stukenberg, P.T. (2003) The highly conserved Ndc80 complex is required for

kinetochore assembly, chromosome congression, and spindle checkpoint activity. *Genes Dev*, **17**, 101-114.

Mellone, B.G., Grive, K.J., Shteyn, V., Bowers, S.R., Oderberg, I. and Karpen, G.H. (2011) Assembly of Drosophila centromeric chromatin proteins during mitosis. *PLoS Genet*, **7**, e1002068.

Meluh, P.B., Yang, P., Glowczewski, L., Koshland, D. and Smith, M.M. (1998) Cse4p is a component of the core centromere of Saccharomyces cerevisiae. *Cell*, **94**, 607-613.

Meraldi, P., McAinsh, A.D., Rheinbay, E. and Sorger, P.K. (2006) Phylogenetic and structural analysis of centromeric DNA and kinetochore proteins. *Genome Biol*, **7**, R23.

Milks, K.J., Moree, B. and Straight, A.F. (2009) Dissection of CENP-C-directed centromere and kinetochore assembly. *Mol Biol Cell*, **20**, 4246-4255.

Mizuguchi, G., Xiao, H., Wisniewski, J., Smith, M.M. and Wu, C. (2007) Nonhistone Scm3 and histones CenH3-H4 assemble the core of centromere-specific nucleosomes. *Cell*, **129**, 1153-1164.

Moore, I., Samalova, M. and Kurup, S. (2006) Transactivated and chemically inducible gene expression in plants. *Plant J*, **45**, 651-683.

Moree, B., Meyer, C.B., Fuller, C.J. and Straight, A.F. (2011) CENP-C recruits M18BP1 to centromeres to promote CENP-A chromatin assembly. *J Cell Biol*, **194**, 855-871.

Murray, A.W. and Szostak, J.W. (1983) Construction of artificial chromosomes in yeast. *Nature*, **305**, 189-193.

Nakano, M., Cardinale, S., Noskov, V.N., Gassmann, R., Vagnarelli, P., Kandels-Lewis, S., . . . Masumoto, H. (2008) Inactivation of a human kinetochore by specific targeting of chromatin modifiers. *Dev Cell*, **14**, 507-522.

Nasuda, S., Hudakova, S., Schubert, I., Houben, A. and Endo, T.R. (2005) Stable barley chromosomes without centromeric repeats. *Proc Natl Acad Sci U S A*, **102**, 9842-9847.

Ogura, Y., Shibata, F., Sato, H. and Murata, M. (2004) Characterization of a CENP-C homolog in Arabidopsis thaliana. *Genes Genet Syst*, **79**, 139-144.

Olszak, A.M., van Essen, D., Pereira, A.J., Diehl, S., Manke, T., Maiato, H., . . . Heun, P. (2011) Heterochromatin boundaries are hotspots for de novo kinetochore formation. *Nat Cell Biol*, **13**, 799-808.

Page, S.L., Earnshaw, W.C., Choo, K.H. and Shaffer, L.G. (1995) Further evidence that CENP-C is a necessary component of active centromeres: studies of a dic(X; 15) with simultaneous immunofluorescence and FISH. *Hum Mol Genet*, **4**, 289-294.

Petrovic, A., Pasqualato, S., Dube, P., Krenn, V., Santaguida, S., Cittaro, D., . . . Musacchio, A. (2010) The MIS12 complex is a protein interaction hub for outer kinetochore assembly. *J Cell Biol*, **190**, 835-852.

Phan, B.H., Jin, W., Topp, C.N., Zhong, C.X., Jiang, J., Dawe, R.K. and Parrott, W.A. (2007) Transformation of rice with long DNA-segments consisting of random genomic DNA or centromere-specific DNA. *Transgenic Res*, **16**, 341-351.

Przewloka, M.R., Venkei, Z., Bolanos-Garcia, V.M., Debski, J., Dadlez, M. and Glover, D.M. (2011) CENP-C is a structural platform for kinetochore assembly. *Curr Biol*, **21**, 399-405.

Ranjitkar, P., Press, M.O., Yi, X., Baker, R., MacCoss, M.J. and Biggins, S. (2010) An E3 ubiquitin ligase prevents ectopic localization of the centromeric histone H3 variant via the centromere targeting domain. *Mol Cell*, **40**, 455-464.

Ravi, M., Shibata, F., Ramahi, J.S., Nagaki, K., Chen, C., Murata, M. and Chan, S.W. (2011) Meiosis-Specific Loading of the Centromere-Specific Histone CENH3 in Arabidopsis thaliana. *PLoS Genet*, **7**, e1002121.

Riera-Lizarazu, O., Vales, M.I., Ananiev, E.V., Rines, H.W. and Phillips, R.L. (2000) Production and characterization of maize chromosome 9 radiation hybrids derived from an oat-maize addition line. *Genetics*, **156**, 327-339.

Robinett, C.C., Straight, A., Li, G., Willhelm, C., Sudlow, G., Murray, A. and Belmont, A.S. (1996) In vivo localization of DNA sequences and visualization of large-scale chromatin organization using lac operator/repressor recognition. *J Cell Biol*, **135**, 1685-1700.

Saitoh, H., Tomkiel, J., Cooke, C.A., Ratrie, H., 3rd, Maurer, M., Rothfield, N.F. and Earnshaw, W.C. (1992) CENP-C, an autoantigen in scleroderma, is a component of the human inner kinetochore plate. *Cell*, **70**, 115-125.

Sanchez-Pulido, L., Pidoux, A.L., Ponting, C.P. and Allshire, R.C. (2009) Common ancestry of the CENP-A chaperones Scm3 and HJURP. *Cell*, **137**, 1173-1174.

Sato, H., Shibata, F. and Murata, M. (2005) Characterization of a Mis12 homologue in Arabidopsis thaliana. *Chromosome Res*, **13**, 827-834.

Schnable, P.S., Ware, D., Fulton, R.S., Stein, J.C., Wei, F., Pasternak, S., . . . Wilson, R.K. (2009) The B73 maize genome: complexity, diversity, and dynamics. *Science*, **326**, 1112-1115.

Screpanti, E., De Antoni, A., Alushin, G.M., Petrovic, A., Melis, T., Nogales, E. and Musacchio, A. (2011) Direct binding of Cenp-C to the Mis12 complex joins the inner and outer kinetochore. *Curr Biol*, **21**, 391-398.

Sekulic, N., Bassett, E.A., Rogers, D.J. and Black, B.E. (2010) The structure of (CENP-A-H4)(2) reveals physical features that mark centromeres. *Nature*, **467**, 347-351.

Shelby, R.D., Monier, K. and Sullivan, K.F. (2000) Chromatin assembly at kinetochores is uncoupled from DNA replication. *J. Cell Biol.*, **151**, 1113-1118.

Shi, J. and Dawe, R.K. (2006) Partitioning of the maize epigenome by the number of methyl groups on histone H3 lysines 9 and 27. *Genetics*, **173**, 1571-1583.

Song, K., Gronemeyer, B., Lu, W., Eugster, E. and Tomkiel, J.E. (2002) Mutational analysis of the central centromere targeting domain of human centromere protein C, (CENP-C). *Exp Cell Res*, **275**, 81-91.

Straight, A.F., Belmont, A.S., Robinett, C.C. and Murray, A.W. (1996) GFP tagging of budding yeast chromosomes reveals that protein-protein interactions can mediate sister chromatid cohesion. *Curr Biol*, **6**, 1599-1608.

Sullivan, B.A. and Karpen, G.H. (2004) Centromeric chromatin exhibits a histone modification pattern that is distinct from both euchromatin and heterochromatin. *Nat Struct Mol Biol*, **11**, 1076-1083.

Sullivan, B.A. and Schwartz, S. (1995) Identification of centromeric antigens in dicentric Robertsonian translocations: CENP-C and CENP-E are necessary componenets of functional centromeres. *Human Molecular Genetics*, **4**, 2189-2197.

Tachiwana, H., Kagawa, W., Shiga, T., Osakabe, A., Miya, Y., Saito, K., . . . Kurumizaka, H. (2011) Crystal structure of the human centromeric nucleosome containing CENP-A. *Nature*.

Talbert, P.B., Bryson, T.D. and Henikoff, S. (2004) Adaptive evolution of centromere proteins in plants and animals. *J Biol*, **3**, 18.

Talbert, P.B., Masuelli, R., Tyagi, A.P., Comai, L. and Henikoff, S. (2002) Centromeric localization and adaptive evolution of an Arabidopsis histone H3 variant. *Plant Cell*, **14**, 1053-1066.

Tanaka, K., Chang, H.L., Kagami, A. and Watanabe, Y. (2009) CENP-C functions as a scaffold for effectors with essential kinetochore functions in mitosis and meiosis. *Dev Cell*, **17**, 334-343.

Topp, C.N., Okagaki, R.J., Melo, J.R., Kynast, R.G., Phillips, R.L. and Dawe, R.K. (2009) Identification of a maize neocentromere in an oat-maize addition line. *Cytogenet Genome Res*, **124**, 228-238.

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

Trazzi, S., Bernardoni, R., Diolaiti, D., Politi, V., Earnshaw, W.C., Perini, G. and Della Valle, G. (2002) In vivo functional dissection of human inner kinetochore protein CENP-C. *J Struct Biol*, **140**, 39-48.

Van Hooser, A.A., Ouspenski, II, Gregson, H.C., Starr, D.A., Yen, T.J., Goldberg, M.L., . . . Brinkley, B.R. (2001) Specification of kinetochore-forming chromatin by the histone H3 variant CENP-A. *J Cell Sci*, **114**, 3529-3542.

Vazquez, J., Belmont, A.S. and Sedat, J.W. (2001) Multiple regimes of constrained chromosome motion are regulated in the interphase Drosophila nucleus. *Curr Biol*, **11**, 1227-1239.

Vermaak, D., Hayden, H.S. and Henikoff, S. (2002) Centromere targeting element within the histone fold domain of Cid. *Mol Cell Biol*, **22**, 7553-7561.

Voullaire, L.E., Slater, H.R., Petrovic, V. and Choo, K.H. (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**, 1153-1163.

Warburton, P.E., Cooke, C.A., Bourassa, S., Vafa, O., Sullivan, B.A., Stetten, G., . . . Earnshaw, W.C. (1997) Immunolocalization of CENP-A suggests a distinct nucleosome structure at the inner kinetochore plate of active centromeres. *Curr Biol*, **7**, 901-904.

Webb, C.D., Teleman, A., Gordon, S., Straight, A., Belmont, A., Lin, D.C., . . . Losick, R. (1997) Bipolar localization of the replication origin regions of chromosomes in vegetative and sporulating cells of B. subtilis. *Cell*, **88**, 667-674.

Wei, R.R., Al-Bassam, J. and Harrison, S.C. (2007) The Ndc80/HEC1 complex is a contact point for kinetochore-microtubule attachment. *Nat Struct Mol Biol*, **14**, 54-59.

Wei, R.R., Schnell, J.R., Larsen, N.A., Sorger, P.K., Chou, J.J. and Harrison, S.C. (2006) Structure of a central component of the yeast kinetochore: the Spc24p/Spc25p globular domain. *Structure*, **14**, 1003-1009.

Wei, R.R., Sorger, P.K. and Harrison, S.C. (2005) Molecular organization of the Ndc80 complex, an essential kinetochore component. *Proc Natl Acad Sci U S A*, **102**, 5363-5367.

Wieland, G., Orthaus, S., Ohndorf, S., Diekmann, S. and Hemmerich, P. (2004) Functional complementation of human centromere protein A (CENP-A) by Cse4p from Saccharomyces cerevisiae. *Mol Cell Biol*, **24**, 6620-6630.

Wigge, P.A. and Kilmartin, J.V. (2001) The Ndc80p complex from Saccharomyces cerevisiae contains conserved centromere components and has a function in chromosome segregation. *J Cell Biol*, **152**, 349-360.

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

Wolfgruber, T.K., Sharma, A., Schneider, K.L., Albert, P.S., Koo, D.H., Shi, J., . . . Presting, G.G. (2009) Maize centromere structure and evolution: sequence analysis of centromeres 2 and 5 reveals dynamic Loci shaped primarily by retrotransposons. *PLoS Genet*, **5**, e1000743.

Yang, C.H., Tomkiel, J., Saitoh, H., Johnson, D.H. and Earnshaw, W.C. (1996) Identification of overlapping DNA-binding and centromere-targeting domains in the human kinetochore protein CENP-C. *Mol Cell Biol*, **16**, 3576-3586.

Yu, H.G., Muszynski, M.G. and Kelly Dawe, R. (1999) The maize homologue of the cell cycle checkpoint protein MAD2 reveals kinetochore substructure and contrasting mitotic and meiotic localization patterns. *J Cell Biol*, **145**, 425-435.

Yu, W., Han, F., Gao, Z., Vega, J.M. and Birchler, J.A. (2007) Construction and behavior of engineered minichromosomes in maize. *Proc Natl Acad Sci U S A*, **104**, 8924-8929.

Yu, W., Lamb, J.C., Han, F. and Birchler, J.A. (2006) Telomere-mediated chromosomal truncation in maize. *Proc Natl Acad Sci U S A*, **103**, 17331-17336.

Zeitlin, S.G., Baker, N.M., Chapados, B.R., Soutoglou, E., Wang, J.Y., Berns, M.W. and Cleveland, D.W. (2009) Double-strand DNA breaks recruit the centromeric histone CENP-A. *Proc Natl Acad Sci U S A*, **106**, 15762-15767.

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

Zhou, Z., Feng, H., Zhou, B.R., Ghirlando, R., Hu, K., Zwolak, A., . . . Bai, Y. (2011) Structural basis for recognition of centromere histone variant CenH3 by the chaperone Scm3. *Nature*, **472**, 234-237.

Figure 1.1 Diagram of the CENH3 secondary structure.

CENH3 is mainly composed of two domains: a N-terminal tail and a histone fold domain. The CENPA centromere-targeting domain (CATD) is localized within the histone fold and contains the small loop1 (L1) and helix 2 (α2). The extreme C-terminus of CENH3, consisting of loop 2 (L2) and helix 3 (α3) interacts with inner kinetochore protein CENPC.

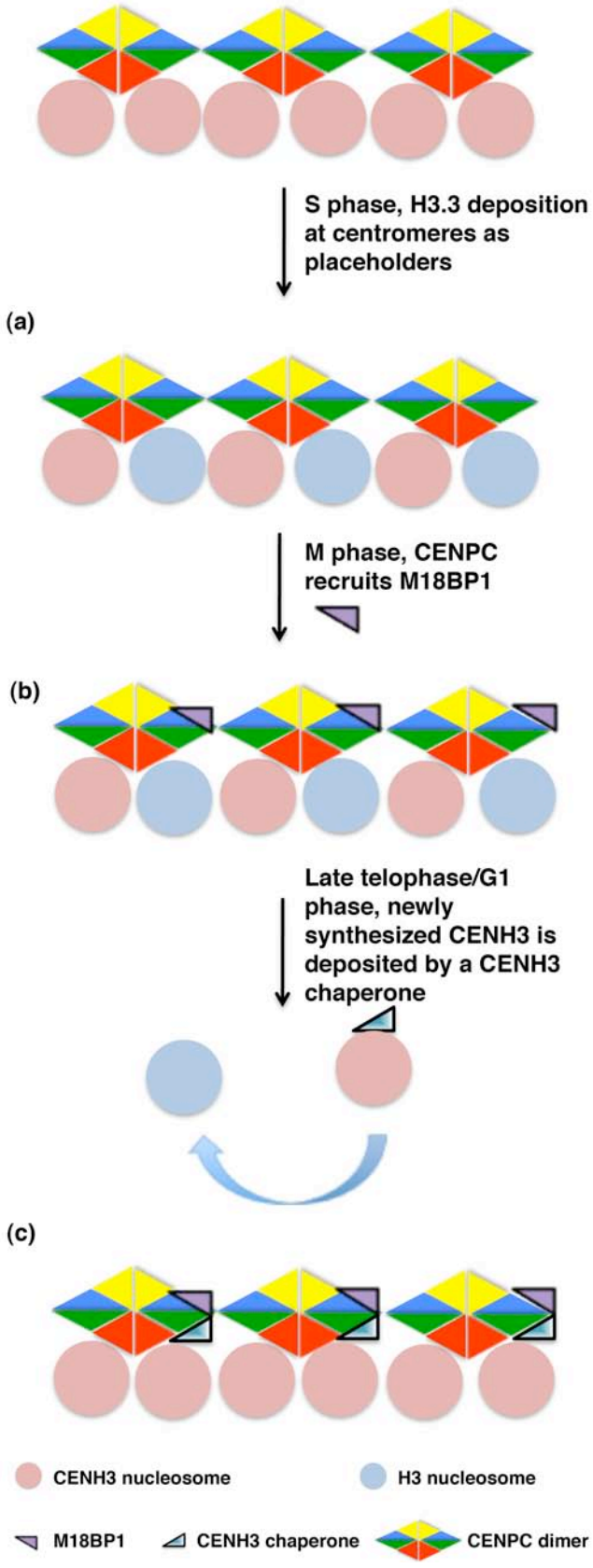

S phase, H3.3 deposition at centromeres as placeholders
(a)
M phase, CENPC recruits M18BP1
(b)
Late telophase/G1 phase, newly synthesized CENH3 is deposited by a CENH3 chaperone
(c)
CENH3 nucleosome
H3 nucleosome
M18BP1
CENH3 chaperone
CENPC dimer

Figure 1.2 Cartoon illustration of CENH3 maintenance and assembly at centromeres.

(a) After DNA replication, CENH3 is diluted to half of its original concentration. CENH3 deposition occurs independent of DNA replication. Instead,

H3-containing nucleosomes (in particular, H3.3 nucleosomes in *Drosophila*) are replicated during S phase and deposited at centromeres as placeholders.

(b) Before CENH3 deposition occurs, CENPC recruits M18BP1 and marks the positions of centromeres.

(c) Newly synthesized CENH3 are bound to CENH3 chaperones, which directly binds CENPC or indirectly binds CENPC through M18BP1. In this way, CENH3 is deposited by the chaperones at centromeres.

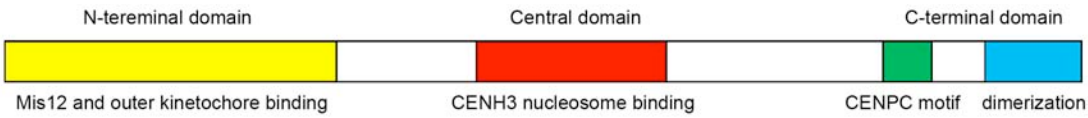


Figure 1.3 Diagram of the CENPC domains.

CENPC is composed of three domains: an N-terminal domain that binds the Mis12 complex and outer kinetochores, a central domain which is responsible for CENH3 nucleosome binding and a C-terminal domain. The C-terminal domain is further divided into two domains, a highly conserved CENPC motif of 23 amino acids, and a dimerization domain.

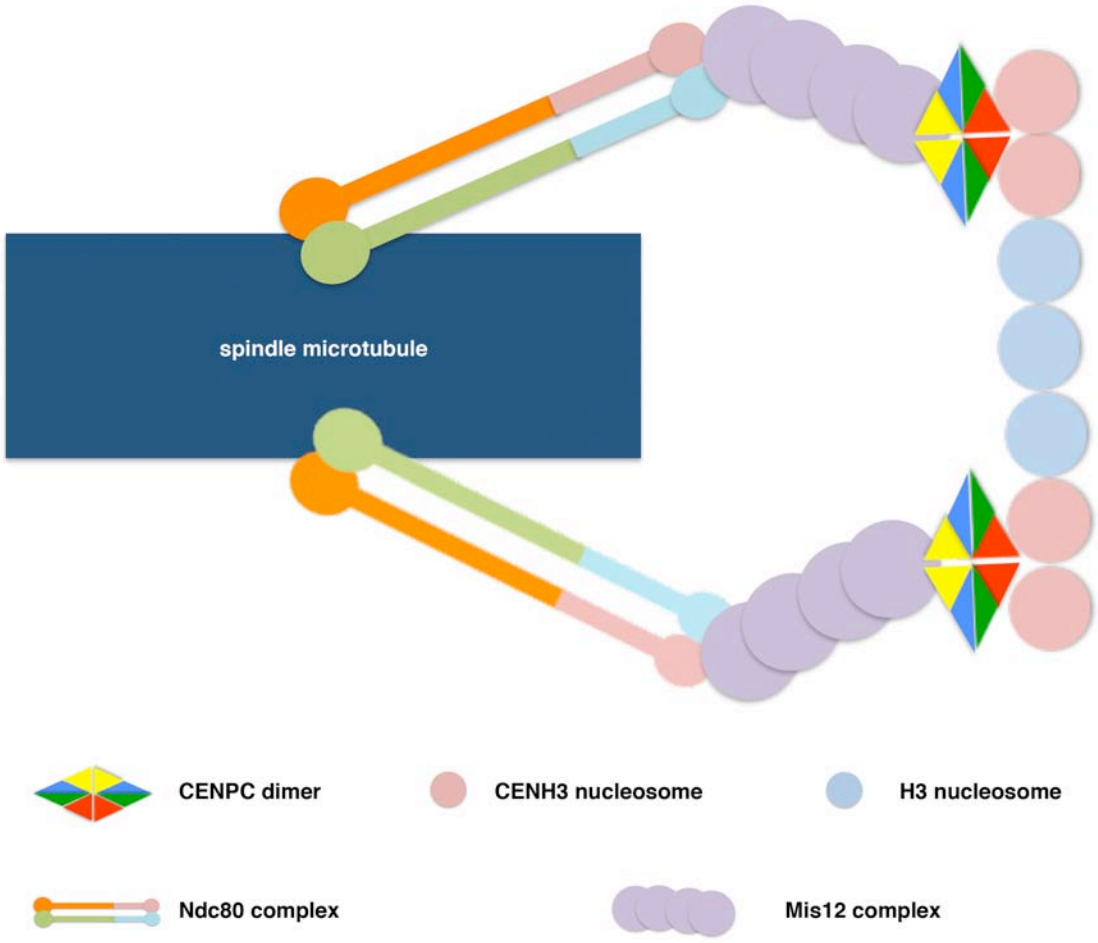


Figure 1.4 Cartoon illustration of the complete kinetochore complex.

CENH3-containing nucleosomes and H3-containing nucleosomes are interspersed in the centromere core. CENPC binds CENH3 nucleosomes with its central domain (red) and interacts with the Mis12 complex at its N-terminal domain (yellow). Both the Mis12 complex and the Ndc80 complex are composed of four proteins and provide a protein linkage between the centromeric chromatin and the spindle microtubules.

CHAPTER 2

STABLE INTEGRATION OF AN ENGINEERED MEGABASE REPEAT ARRAY INTO THE MAIZE GENOME[1]

[1] Han Zhang, Bao Phan, Kai Wang, Barbara A. Artelt, Jiming Jiang, Wayne A. Parrott and R. Kelly Dawe. Manuscript accepted in *The Plant Journal*, 2011.

Abstract

Plant genome engineering as a practical matter will require stable introduction of long and complex segments of DNA sequence into plant genomes. Here we show that it is possible to synthetically engineer and introduce centromere-sized satellite repeat arrays into maize. We designed a synthetic repeat monomer of 156 bp that contains four common DNA binding motifs (LacO, TetO, Gal4, LexA), and extended it into tandem arrays using an overlapping PCR method similar to what is commonly used in gene synthesis. The PCR products were then directly transformed into maize using biolistic transformation. We describe three resulting insertion sites (Arrayed Binding Sites, ABS), the longest of which is at least 1100 kb. The LacI DNA binding module is sufficient to efficiently tether YFP to ABS arrays. We conclude that synthetic repeats can be delivered into plant cells by skipping passage through *E. coli*, that they generally insert into one locus, and may achieve great lengths. These experimental approaches should be useful for future applications in artificial chromosome design.

Introduction

In recent years the field of plant improvement has shifted from the widespread use of single transgenes, into discussions of multi-trait stacking (Halpin 2005) and plant chromosome engineering (Purnick and Weiss 2009). This interest has been paralleled by exciting new methods for gene synthesis, for precisely targeting transgenes to specific loci (Urnov *et al.* 2010) and methods for manipulating plant ploidy (Ravi and Chan 2010). In *E. coli* and other smaller genome species, the horizon has further expanded as multiple authors consider the design or substantial redesign of entire genomes (Purnick and Weiss 2009). Similar ideas may one day be feasible in plants, but at present we are limited by severe technical bottlenecks. Two of these are the inability to clone and transfer large DNA molecules into plants, particularly those with repetitive sequences, and the lack of an empirical framework to design artificial centromeres.

One of the first examples of genome engineering was the creation of artificial chromosomes in *Saccharomyces cerevisiae*. Soon after the small (125 bp) yeast centromeres were identified, they were combined with origins of replication and telomeric repeats and shown to be fully transmissible with large foreign DNA cargoes (Murray and Szostak 1983). Similar approaches (referred to as "bottom-up" strategies) have been used to create synthetic human artificial chromosomes (Harrington *et al.* 1997, Ikeno *et al.* 1998, Henning *et al.* 1999). These small functional chromosomes may be linear or circular and include centromeric alpha-satellite arrays of natural or synthetic origin (Ebersole *et al.* 2000, Basu *et al.* 2005, Nakano *et al.* 2008). In some cases the resulting mini-chromosomes were found to be substantially larger than the original input DNA, suggesting that the introduced sequences had undergone duplication and expansion as a part of chromosome formation (Ikeno *et al.* 1998).

Plant and animal centromeres are similar in size and content, being ~300-2000 kb in size and composed primarily of tandem repeat arrays and interspersed transposons (Birchler and Han 2009). Given these similarities, several laboratories have attempted to create plant artificial chromosomes. The first published experiments demonstrated that a simple strategy of transforming rice centromeric BACs failed to result in artificial chromosomes (Phan *et al.* 2007). In a similar but more informative study, a second group identified maize centromeric BACs, retrofitted them with origins of replication and telomeric repeats, and then transformed them into maize (Ananiev *et al.* 2009). The results suggested that introduced centromeric BACs can initiate centromere assembly, although most of the mini-chromosomes appeared to have resulted from genome rearrangements. In a third study, centromeric BACs as small as 19 kb were assembled into a circular vector and shown to be both mitotic and meiotic inheritable, with evidence that the chromosomes were transmitted as independent ring chromosomes (Carlson *et al.* 2007). The latter study appeared to be most efficient, but doubts remain about the size of constructs and the mechanism of inheritance (Houben *et al.* 2008).

The factors limiting the success of de novo centromere assembly in plants are both technical and biological. Major technical limitations lie in the methods of plant transformation and the cloning vehicles that can be used. Prior authors have relied on existing centromere sequences that were cloned into BACs and introduced into plant genomes using biolistic transformation (Carlson *et al.* 2007, Phan *et al.* 2007, Ananiev *et al.* 2009). It should also be possible to introduce large (~150 kb) centromeric BACs into plant genomes using *Agrobacterium* and specialized T-DNA vectors (Hamilton *et al.* 1996, Liu *et al.* 1999). However, given the instability of repetitive repeat arrays in *E. coli* and *Agrobacterium*, there are inherent limitations to what can be accomplished using BAC clones (Song *et al.* 2001, Song *et*

al. 2003, Nakano *et al.* 2005, Chang *et al.* 2011). In addition, a major unaddressed biological limitation is the fact that centromeres are specified in a sequence-independent (epigenetic) fashion (Birchler and Han 2009). Even if we could easily clone and manipulate centromere-sized sequences, we do not yet understand which sequences might be most useful, or whether any existing sequences can be used reliably.

Recently, in budding yeast and animal cell cultures, it has been reported that de novo centromere activity can be achieved by tethering kinetochore proteins to synthetic sequences (Kiermaier *et al.* 2009, Lacefield *et al.* 2009, Barnhart *et al.* 2011, Gascoigne *et al.* 2011, Mendiburo *et al.* 2011). This approach takes advantage of the high affinity interaction between DNA binding proteins and their target sequences. A DNA binding protein such as the Lactose Repressor protein (LacI) will recognize and bind to lac operator sequences (LacO) integrated into the genome. Kinetochore proteins fused with LacI can be tethered to the LacO arrays to initiate the assembly of additional kinetochore proteins (Kiermaier *et al.* 2009, Lacefield *et al.* 2009, Barnhart *et al.* 2011, Gascoigne *et al.* 2011, Mendiburo *et al.* 2011). These studies have shown that it is possible to artificially assemble nearly complete kinetochores that display most if not all of the segregation properties of natural kinetochores. In principle, the tethering approach combines the predictability of a sequence-based (genetic) specification with the capacity to carefully control the process of kinetochore assembly.

In this study we report results that address one of the key technical barriers to creating engineered centromeres in plants. Very long synthetic tandem arrays containing DNA binding modules were generated by overlap extension PCR, and directly transformed into maize where they incorporated as single intact loci. The utility of the arrays was demonstrated by showing that

the arrays efficiently recruit LacI-YFP fusion proteins to form large fluorescent spots that are easily visible by standard microscopy.

Results

Long tandem repeats can be transformed into maize and remain intact

We designed a repeat monomer that contains the DNA recognition motifs for five DNA binding proteins (Figure 1): LacO, TetO, LexA, Gal4, and the CENPB box, which binds to the CENPB DNA binding protein found in human centromeres (See Methods for explanations and citations). Each monomer of the array was designed to be 156 bp, the size of the endogenous maize centromeric satellite CentC (Ananiev *et al.* 1998). Arrays were synthesized by PCR using two long oligonucleotides (98 and 99 bp) that overlap on both ends. In a standard PCR reaction, the complementary oligos anneal and extend the corresponding strands, and then extend the repeating monomers indefinitely (Figure 2). The engineered repeat arrays made in this way are referred to as Arrayed Binding Sites (ABS).

Agarose gel electrophoresis of ABS amplification reactions showed a negative relationship between primer concentration and size of the resulting PCR products: as primer concentration decreased, PCR product sizes increased (Figure 3a). PCR products rapidly reached sizes large enough to be excluded from the agarose matrix (Figure 3a, lane 6). We also tried contour-clamped homogeneous electric field (CHEF), which enables the resolution of macromolecules as large as two megabases (Chu *et al.* 1986), and again observed that the PCR products failed to enter the gel matrix (data not shown). These PCR products are undoubtedly very large, but we do not know their actual size or structure; they may be partially circular or

contain irregular multimers that prevent entry into agarose gels. Nevertheless after digestion of the PCR products with NdeI, which cleaves once in each monomer, the PCR products were reduced to 156 bp monomers and 312 bp dimers, showing that the vast majority of sequences are tandemly arrayed as predicted (Figure 3b).

For simplicity, we opted to directly transform the PCR products into maize embryogenic cultures. Prior work had shown that smaller PCR products can be transformed into rice by co-bombarding with a plasmid that supplies a selective marker (Kumar *et al.* 2006). The amplified ABS products were mixed with plasmid pAHC25 (Christensen and Quail 1996), which contains the BASTA resistance gene and the GUS reporter gene, and transformed into maize calli by microprojectile bombardment. A total of 59 individual transformation events were selected as BASTA resistant, and 22 (37.3%) contained detectable quantities of the ABS array by PCR. From this set of 22 calli, 17 mature plants (28.8%) were recovered.

We initially used Southern blots to assess the size of the integrated ABS loci. Genomic DNA was digested with multiple restriction enzymes that do not cleave in the array but do cleave within the co-bombarded plasmid (EcoRI, HindIII, BamHI, Kpn I). The blots were then probed with the ABS monomer. Of the 17 recovered plants, five showed bright bands at molecular weights greater than ~40 kb, the resolution of the agarose gel used (Figure 4a). One interesting exception is ABS3 (lane 8), which showed a large ~40 kb band as well as a second distinct band over 20 kb. We also observed faint smaller bands in all lanes, which may represent either short independent arrays or subdomains of the larger arrays. Importantly, when these DNAs were digested with NdeI, all of the ABS-hybridizing sequences were cleaved into short pieces representing 156 bp monomers (Figure 4b).

We found that we were able to identify progeny carrying the three longest arrays (ABS3, ABS4, and ABS7) using primers homologous to the pAHC25 marker plasmid, suggesting that at least one copy of the marker plasmid co-segregates with the arrays. Re-probing the Southern blot in Figure 4a with the GUS gene (not shown) demonstrated clear hybridization in the >40 kb molecular weight range. These results indicate that the pACH25 plasmid occasionally co-integrates with the PCR products, however this must be relatively rare, generally occurring at intervals greater than 40 kb in the longest arrays.

Three independent lines with ABS in different positions relative to centromeres

In order to determine the location of the ABS loci, fluorescence in situ hybridization (FISH) was performed on three transgenic lines with the most abundant ABS signal. Chromosomes can be identified based on size, arm ratio and staining intensity of the centromere repeat CentC (Kato *et al.* 2004). Examination of mitotic chromosome spreads revealed that the arrays are integrated into single sites on long arms of chromosomes 3, 4, and 7 (Figure 5), and are referred to as ABS3, ABS4 and ABS7 accordingly. In ABS3 and ABS4, the tandem arrays are located in central positions of the long arms, while in ABS7, the array is located near the endogenous centromere. The cytological position of ABS7 was further refined by analyzing meiotic pachytene chromosomes, which provide higher resolution (Wang *et al.* 2006). The results show that ABS7 is located in pericentromeric heterochromatin at position 7L.1 on the cytological map (roughly 10% of the distance to the telomere; see Figure 6).

To determine the physical lengths of the integrated ABS arrays, we used fiber-FISH technology, which can accurately estimate size from stretched DNA molecules (Figure 7, Table 1). More than 20 fiber-FISH images were collected and measured for each ABS line. The

lengths of the fiber-FISH signals (in micrometers) were converted to kilobases using a 3.21 kb/μm conversion rate derived from rice fibers (Cheng *et al.* 2002). The data reveal that ABS3 is composed of two satellite blocks of 28 kb and 53 kb that are separated by a 109 kb gap. The fact that ABS3 is divided into two separate regions, one of which is ~30 kb, may explain the existence of a second smaller band in the ABS3 lane of the Southern blot (Figure 4a). In contrast, the fiber-FISH signals from ABS4 and ABS7 lacked obvious intervening sequences and measured 1146 kb and 315 kb, respectively. We note that the dotted hybridization pattern is typical for fiber-FISH images and does not imply that the ABS arrays are interrupted at regular intervals. If we assume that the arrays are nearly continuous, ABS4 contains over 7000 tandemly arrayed copies of synthetic 156bp monomer.

All three ABS loci were reliably inherited over multiple generations. The fiber FISH images in Figure 7 were taken from plants that were at least four generations removed from the original transformants. These data suggest that synthetic repeat arrays, like natural centromeric repeats and the longer maize knob repeat arrays (Adawy *et al.* 2004), are stably replicated and transmitted.

Segregation and inheritance of tethered LacI-YFP in transgenic plants

To test whether the arrays function as tethering sites, we separately transformed maize with a LacI-YFP fusion construct driven by the constitutive 35S promoter. We anticipated that if plants expressing LacI-YFP were crossed to ABS lines, the YFP signal would aggregate to produce strong and focused fluorescent signals. As expected, in plants hemizygous for LacI-YFP and ABS4, we observed either a single bright dot in nuclei (Figure 8c, top row) or two closely paired signals representing the G2 phase of the cell cycle (Figure 8c, bottom row).

Sibling plants that contained LacI-YFP but lacked ABS showed no fluorescent foci (Figure 8a). We also crossed hemizygous plants to produce a population segregating for LacI-YFP and zero, one or two copies of ABS4 (Figure 8b, 8d and 8f). Using FISH to determine ABS genotype, we showed that in a sample of 13 plants, the number of visible YFP foci corresponded perfectly to the number of ABS arrays present in the cell (Figure 8).

ABS4 is the longest of the arrays, being almost four times the size of ABS7 and ten times the size of the combined ABS3 blocks. We found that fluorescent foci were always observed in LacI-YFP lines containing ABS4 (of 19 plants tested) and ABS7 (of 12 plants tested). However, visible YFP foci were observed in only one out of nine plants containing the ABS3 array. We were able to directly compare ABS3 to ABS4 by crossing the two lines together to create plants with both arrays as well as LacI-YFP. In these lines one bright dot (presumably ABS4) was observed in all cells, and occasionally a second much weaker dot barely at the limit of detection was also visible (not shown). It is possible that more LacO repeats are necessary to visualize YFP in maize than in Arabidopsis, where 120 copies of the lacO sequence were sufficient to visualize LacI fused to GFP (Kato and Lam 2001). However, other factors are also probably involved, such as the expression level of the 35S promoter in maize, the relative brightness of the fluorescent protein used in this study, and the transmission of fluorescence through intact maize roots.

Discussion

Our goal was to produce synthetic repeat arrays resembling natural centromeres, and introduce the arrays into the maize genome. At the time the study was initiated, LacO and TetO

repeat arrays were available as plasmid clones (Robinett *et al.* 1996, Michaelis *et al.* 1997), however these arrays (less than 6 kb) were far shorter than natural centromeres, and were known to be unstable in *E. coli*, and lacked the nucleosome-sized repeat structure typical of satellite repeats. We used a simple PCR method involving double-overlapping primers that extend a tandem repeat structure indefinitely within the confines of a PCR reaction. The large PCR products were then coated onto metal spheres along with a marker plasmid and shot into maize embryogenic cultures. The ABS4 array described here (Figure 7) is the longest segment of foreign DNA yet incorporated into a higher plant, and at 1.1 Mb, is the size of many functional centromere regions (Murata *et al.* 2008, Yan *et al.* 2008, Wolfgruber *et al.* 2009).

Previous studies have shown that biolistic transformation can be used to introduce nearly any type of DNA into plant genomes, but it often results in rearranged or duplicated insertions that sometimes include interspersed genomic sequences (Sautter *et al.* 1999, Friebe *et al.* 2001). In our case it is difficult to interpret the events that occurred upon transformation; however, consistent with prior work we found that the PCR products co-integrated with marker plasmids, and that there is a low level of host genome rearrangement upon insertion (see ABS3, Figure 7). More importantly, our data reveal that the great majority of integrated sequences are derived from ABS arrays, and that these long synthetic molecules survived biolistic transformation to become heritable features of the genome. While basic considerations of size and repeat character would have made this study unapproachable using *Agrobacterium*, at present there appears to be no inherent limit to the size or type of DNA molecule that can be transformed into plants using biolistic methods.

The value of long arrays of centromere-like repeats lies in their application to artificial chromosome technology. One might question the need for synthetic repeats, when in human

cells the natural centromere repeats (alpha-satellites) are sufficient to form functional artificial chromosomes (Harrington *et al.* 1997, Ebersole *et al.* 2000), and similar strategies in plants have met with at least partial success (Carlson *et al.* 2007, Ananiev *et al.* 2009). However, an artificial chromosome strategy based on success in cultured human cells has inherent weaknesses for higher plants. Unlike in plants, human alpha satellites have evolved in concert with a specialized binding protein called CENPB. Human CENPB appears to function primarily to establish or maintain centromere identity by binding to a CENPB-binding motif in functional centromeric domains (Masumoto *et al.* 1989, Muro *et al.* 1992). If the motif is removed, artificial centromeres do not form (Ohzeki *et al.* 2002). Plants, lacking any evidence of a similar DNA-to-protein specification mechanism, seem to rely heavily on epigenetic processes to ensure centromere assembly (Birchler and Han 2009).

In recent months a series of papers have illustrated the utility of using protein tethering to confer a genetic specification mechanism to centromeric regions. In yeast, a microtubule binding protein (Dam1) was tethered to a DNA binding array and shown to promote segregation of an otherwise acentric plasmid (Kiermaier *et al.* 2009, Lacefield *et al.* 2009). Similarly, tethering of human kinetochore proteins CENPC or CENPN was sufficient to assemble outer kinetochore proteins to an integrated repeat array (Gascoigne *et al.* 2011). Recent studies went a step further by directly tethering Centromeric Histone H3 (CENH3) or its deposition factor HJURP, and demonstrating the ectopic assembly of centromeric nucleosomes and fully functional kinetochores over the introduced repeat arrays (Barnhart *et al.* 2011, Mendiburo *et al.* 2011).

With the technical advances described here, it is now possible to conduct similar tests in plants. The strategy we describe for producing repeat arrays is simple and should make it

possible to rapidly redesign the monomer, for instance to better match known nucleosome positioning motifs in plant centromeres (Gent *et al.* 2011). In the near term, such tethering arrays should help us better understand the process of plant kinetochore assembly, which at present is almost entirely unknown. In the long term, our goal is to create self-sustaining synthetic centromeres that can be used in constructing plant artificial chromosomes, an area of active research with great promise (Purnick and Weiss 2009).

Materials and Methods

ABS preparation, transformation and detection

The sequence of the ABS array was designed to match the GC content of the maize CentC array. The sequence contains an NdeI site, the LacO operator (AATTGTGAGCGGCTCACAATT; (Simons *et al.* 1984, Lanzer and Bujard 1988)), the TetO operator (TCCCTATCAGTGATAGAGA; (Lutz and Bujard 1997)), the LexA UAS (TACTGTATATATATACAGTA; (Berg 1988)), the Gal4 UAS (CGGAGGACTGTCCTCCG; (Giniger *et al.* 1985, Reece and Ptashne 1993)) and the CENPB box (TTTCGTTGGAAACGGGA; (Masumoto *et al.* 1989)) which binds to a known DNA binding domain (Tanaka *et al.* 2001). Outside of the six constrained sequences, bases were chosen at random. Overlap extension PCR was performed using the following primers: syn3 for (5'-GTACCGCTAGTCCCTATCAGTGATAGAGATGCAGTCGCCGAATACTGTATATATATA CAGTAAGCCTTCTACTGGAATTGTGAGCGGCTCACAATTAGT-3') and syn3 rev (5'-CTGATAGGGACTAGCGGTACTCCGGAGGACAGTCCTCCGAGAAGGCATATGTCCCG TTTCCAACGAAAATTTCATCCTACTAATTGTGAGCCGCTCAC-3'). PCR was performed

using the BD Advantage™ 2 PCR Kit (Clonetech, Mountain View, CA) in a reaction containing 0.01 μM primers, 100 μM dNTPs, 1 μL BD Advantage™ 2 PCR buffer and 0.1 μL BD Advantage™ 2 PCR polymerase mix in a 10 μL reaction. PCR reactions using other kits and Taq polymerases were not as effective. PCR was performed using the following parameters: initial denaturing at 95°C for 2 min, 35 amplification cycles of 20 sec at 94°C, 1 min at 52°C and 1 min at 72°C, followed with a final incubation at 72°C for 8 min. The resulting PCR products are expected to have gaps, nicks and A overhangs. Most of the products are probably hybrid molecules with parts of multiple PCR products entwined. To correct these presumed errors, PCR products were treated with PfuI polymerase, which should fill gaps, correct errors (by its 3'-5' exonuclease activity) and remove the A overhangs. Finally, we treated the sample with DNA ligase to seal the nicks and join the molecules together.

A plasmid pAHC25 (Christensen and Quail 1996) which contains the reporter gene β-glucuronidase (GUS) and the selectable BAR gene which confers resistance to the herbicide BASTA was mixed with the PCR products at a molar ratio of 1:6, and was directly shot into maize by microprojectile bombardment as described previously (Phan *et al.* 2007). Target maize calli were subsequently selected for BASTA resistance and then screened for ABS by PCR using the following primers: syn3for_short (5'-GTACCGCTAGTCCCTATCAGTGAT-3') and syn3rev_short (5'-TCCGGAGGACAGTCCTCC-3'). Southern blot was performed according to (Phan *et al.* 2007) with probes specific for a single monomer of ABS (156bp).

FISH and fiber-FISH

FISH was performed on root tips as described previously (Kato *et al.* 2004), except that fluorescently labeled oligonucleotides were used to detect the ABS arrays. Immature tassels

were harvested and fixed in 1 X buffer A (80 mM KCl, 20 mM NaCl, 2 mM EDTA, 0.5 mM EGTA, 15 mM Pipes buffer, pH 7.0), 4% paraformaldehyde and 0.1% Triton X-100 for 2 ~3 hours before dissecting. Pachytene chromosomes were prepared as described previously (Shi and Dawe 2006) and FISH was performed as for root tips. Fiber-FISH was performed according to Jackson et al. (Jackson *et al.* 1998).

LacI-YFP construct preparation, transformation and detection

The LacI gene without stop codon was cloned into pENTR/D vector (Invitrogen) from *E.coli* genomic DNA (kindly provided by Sidney Kushner) using the following primers: LacINLS-F2 (5'-CACCATGAAGAAGAAGAGAAAGGTGGTGAAACCAGTAACGTTATACG-3') and LacI-R2 (5'-CTGCCCGCTTTCCAGTCGGGAA-3'). The forward primer introduces the nuclear localization sequence KKKRKV (from the SV40 large T antigen) as used in plants previously (Raikhel 1992). The sequence-confirmed LacI sequence was then recombined into pEarleygate vector pEG101 (an Invitrogen Gateway vector derivative; (Earley *et al.* 2006)) to fuse the LacI protein to YFP. The final 35S promoter-LacI-YFP fusion construct was introduced into *Agrobacterium tumefaciens* strain C58C1 (a gift from Richard Meagher and lab), and then transformed into maize HiII lines by the Iowa State University Plant Transformation Facility (http://www.agron.iastate.edu/ptf/). The transgenes were scored in the progeny using the following primers: Gus2F (5'-AGTGAAGGGCCAACAGTTCCTGAT-3') and Gus2R (5'-AAATATTCCCGTGCACTTGCGGAC-3') for ABS detection and 35S1F (5'-GGCCCTAACAAGCCCACCAA-3') and 35S1R (5'-GGGCAATGGAATCCGAGGAG-3') for LacI-YFP detection.

Root tips about 1 cm in length were cut from 3-day-old seedlings and fixed on the slides by taping the edges of the coverslips onto the slides. Phosphate Buffered Saline (PBS) was added into the space between coverslips and slides. The slides were then observed from the surface using a Zeiss Axio Imager microscope and a 63X objective oil lens. Images were taken with Slidebook software (Intelligent Imaging Innovations).

Acknowledgments

We thank Amy Luce for helping to prepare the LacI-YFP fusion constructs, Sarah Covert for providing access to a CHEF gel apparatus, Sidney Kushner for supplying *E. coli* genomic DNA and Richard Meagher for providing the *Agrobacterium tumefaciens* strain. This work was supported by grants 0421671 and 092270 from the National Science Foundation.

References

Adawy, S.S., Stupar, R.M. and Jiang, J. (2004) Fluorescence in situ hybridization analysis reveals multiple loci of knob-associated DNA elements in one-knob and knobless maize lines. *J Histochem Cytochem*, **52**, 1113-1116.

Ananiev, E.V., Phillips, R.L. and Rines, H.W. (1998) Chromosome-specific molecular organization of maize (Zea mays L.) centromeric regions. *Proc Natl Acad Sci U S A*, **95**, 13073-13078.

Ananiev, E.V., Wu, C., Chamberlin, M.A., Svitashev, S., Schwartz, C., Gordon-Kamm, W. and Tingey, S. (2009) Artificial chromosome formation in maize (Zea mays L.). *Chromosoma*, **118**, 157-177.

Barnhart, M.C., Kuich, P.H., Stellfox, M.E., Ward, J.A., Bassett, E.A., Black, B.E. and Foltz, D.R. (2011) HJURP is a CENP-A chromatin assembly factor sufficient to form a functional de novo kinetochore. *J Cell Biol*, **194**, 229-243.

Basu, J., Stromberg, G., Compitello, G., Willard, H.F. and Van Bokkelen, G. (2005) Rapid creation of BAC-based human artificial chromosome vectors by transposition with synthetic alpha-satellite arrays. *Nucleic Acids Res*, **33**, 587-596.

Berg, O.G. (1988) Selection of DNA binding sites by regulatory proteins: the LexA protein and the arginine repressor use different strategies for functional specificity. *Nucleic Acids Res*, **16**, 5089-5105.

Birchler, J.A. and Han, F. (2009) Maize centromeres: structure, function, epigenetics. *Annu Rev Genet*, **43**, 287-303.

Carlson, S.R., Rudgers, G.W., Zieler, H., Mach, J.M., Luo, S., Grunden, E., . . . Preuss, D. (2007) Meiotic transmission of an in vitro-assembled autonomous maize minichromosome. *PLoS Genet*, **3**, 1965-1974.

Chang, Y.L., Chuang, H.W., Meksem, K., Wu, F.C., Chang, C.Y., Zhang, M. and Zhang, H.B. (2011) Characterization of a plant-transformation-ready large-insert BIBAC library of Arabidopsis and bombardment transformation of a large-insert BIBAC of the library into tobacco. *Genome*, **54**, 437-447.

Cheng, Z., Buell, C.R., Wing, R.A. and Jiang, J. (2002) Resolution of fluorescence in-situ hybridization mapping on rice mitotic prometaphase chromosomes, meiotic pachytene chromosomes and extended DNA fibers. *Chromosome Res*, **10**, 379-387.

Christensen, A.H. and Quail, P.H. (1996) Ubiquitin promoter-based vectors for high-level expression of selectable and/or screenable marker genes in monocotyledonous plants. *Transgenic Res*, **5**, 213-218.

Chu, G., Vollrath, D. and Davis, R.W. (1986) Separation of large DNA molecules by contour-clamped homogeneous electric fields. *Science*, **234**, 1582-1585.

Earley, K.W., Haag, J.R., Pontes, O., Opper, K., Juehne, T., Song, K. and Pikaard, C.S. (2006) Gateway-compatible vectors for plant functional genomics and proteomics. *Plant J*, **45**, 616-629.

Ebersole, T.A., Ross, A., Clark, E., McGill, N., Schindelhauer, D., Cooke, H. and Grimes, B. (2000) Mammalian artificial chromosome formation from circular alphoid input DNA does not require telomere repeats. *Hum Mol Genet*, **9**, 1623-1631.

Friebe, B., Jackson, S.A., Zhang, P., Chen, W.P., Phillips, R.L., Muthukrishnan, S. and Gill, B.S. (2001) High-resolution structural analysis of biolistic transgene integration into the genome of wheat. *Theoretical and Applied Genetics*, **103**, 56-62.

Gascoigne, K.E., Takeuchi, K., Suzuki, A., Hori, T., Fukagawa, T. and Cheeseman, I.M. (2011) Induced Ectopic Kinetochore Assembly Bypasses the Requirement for CENP-A Nucleosomes. *Cell*, **145**, 410-422.

Gent, J.I., Schneider, K.L., Topp, C.N., Rodriguez, C., Presting, G.G. and Dawe, R.K. (2011) Distinct influences of tandem repeats and retrotransposons on CENH3 nucleosome positioning. *Epigenetics Chromatin*, **4**, 3.

Giniger, E., Varnum, S.M. and Ptashne, M. (1985) Specific DNA binding of GAL4, a positive regulatory protein of yeast. *Cell*, **40**, 767-774.

Halpin, C. (2005) Gene stacking in transgenic plants--the challenge for 21st century plant biotechnology. *Plant Biotechnol J*, **3**, 141-155.

Hamilton, C.M., Frary, A., Lewis, C. and Tanksley, S.D. (1996) Stable transfer of intact high molecular weight DNA into plant chromosomes. *Proc Natl Acad Sci U S A*, **93**, 9975-9979.

Harrington, J.J., Bokkelen, G.V., Mays, R.W., Gustashaw, K. and Willard, H.F. (1997) Formation of de novo centromeres and construction of first-generation human artificial chromosomes. *Nature Genetics*, **15**, 345-355.

Henning, K.A., Novotny, E.A., Compton, S.T., Guan, X.Y., Liu, P.P. and Ashlock, M.A. (1999) Human artificial chromosomes generated by modification of a yeast artificial chromosome containing both human alpha satellite and single-copy DNA sequences. *Proc Natl Acad Sci U S A*, **96**, 592-597.

Houben, A., Dawe, R.K., Jiang, J. and Schubert, I. (2008) Engineered plant minichromosomes: a bottom-up success? *Plant Cell*, **20**, 8-10.

Ikeno, M., Grimes, B., Okazaki, T., Nakano, M., Saitoh, K., Hoshino, H., . . . Masumoto, H. (1998) Construction of YAC-based mammalian artificial chromosomes. *Nat. Biotech.*, **16**, 431-439.

Jackson, S.A., Wang, M.L., Goodman, H.M. and Jiang, J. (1998) Application of fiber-FISH in physical mapping of Arabidopsis thaliana. *Genome*, **41**, 566-572.

Kato, A., Lamb, J.C. and Birchler, J.A. (2004) Chromosome painting using repetitive DNA sequences as probes for somatic chromosome identification in maize. *Proc Natl Acad Sci U S A*, **101**, 13554-13559.

Kato, N. and Lam, E. (2001) Detection of chromosomes tagged with green fluorescent protein in live Arabidopsis thaliana plants. *Genome Biol*, **2**, RESEARCH0045.

Kiermaier, E., Woehrer, S., Peng, Y., Mechtler, K. and Westermann, S. (2009) A Dam1-based artificial kinetochore is sufficient to promote chromosome segregation in budding yeast. *Nat Cell Biol*, **11**, 1109-1115.

Kumar, S., Arul, L., Talwar, D. and Raina, S.K. (2006) PCR amplification of minimal gene expression cassette: an alternative, low cost and easy approach to 'clean DNA' transformation. *Curr Sci India*, **91**, 930-934.

Lacefield, S., Lau, D.T. and Murray, A.W. (2009) Recruiting a microtubule-binding complex to DNA directs chromosome segregation in budding yeast. *Nat Cell Biol*, **11**, 1116-1120.

Lanzer, M. and Bujard, H. (1988) Promoters largely determine the efficiency of repressor action. *Proc Natl Acad Sci U S A*, **85**, 8973-8977.

Liu, Y.G., Shirano, Y., Fukaki, H., Yanai, Y., Tasaka, M., Tabata, S. and Shibata, D. (1999) Complementation of plant mutants with large genomic DNA fragments by a transformation-competent artificial chromosome vector accelerates positional cloning. *Proc Natl Acad Sci U S A*, **96**, 6535-6540.

Lutz, R. and Bujard, H. (1997) Independent and tight regulation of transcriptional units in Escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Res*, **25**, 1203-1210.

Masumoto, H., Masukata, H., Muro, Y., Nozaki, N. and Okazaki, T. (1989) A human centromere antigen (CENP-B) interacts with a short specific sequence in alphoid DNA, a human centromeric satellite. *J Cell Biol*, **109**, 1963-1973.

Mendiburo, M.J., Padeken, J., Fulop, S., Schepers, A. and Heun, P. (2011) Drosophila CENH3 is sufficient for centromere formation. *Science*, **334**, 686-690.

Michaelis, C., Ciosk, R. and Nasmyth, K. (1997) Cohesins: chromosomal proteins that prevent premature separation of sister chromatids. *Cell*, **91**, 35-45.

Murata, M., Yokota, E., Shibata, F. and Kashihara, K. (2008) Functional analysis of the Arabidopsis centromere by T-DNA insertion-induced centromere breakage. *Proc Natl Acad Sci U S A*, **105**, 7511-7516.

Muro, Y., Masumoto, H., Yoda, K., Nozaki, N., Ohashi, M. and Okazaki, T. (1992) Centromere protein B assembles human centromeric alpha-satellite DNA at the 17-bp sequence, CENP-B box. *J Cell Biol*, **116**, 585-596.

Murray, A.W. and Szostak, J.W. (1983) Construction of artificial chromosomes in yeast. *Nature*, **305**, 189-193.

Nakano, A., Suzuki, G., Yamamoto, M., Turnbull, K., Rahman, S. and Mukai, Y. (2005) Rearrangements of large-insert T-DNAs in transgenic rice. *Mol Genet Genomics*, **273**, 123-129.

Nakano, M., Cardinale, S., Noskov, V.N., Gassmann, R., Vagnarelli, P., Kandels-Lewis, S., . . . Masumoto, H. (2008) Inactivation of a human kinetochore by specific targeting of chromatin modifiers. *Dev Cell*, **14**, 507-522.

Ohzeki, J., Nakano, M., Okada, T. and Masumoto, H. (2002) CENP-B box is required for de novo centromere chromatin assembly on human alphoid DNA. *J Cell Biol*, **159**, 765-775.

Phan, B.H., Jin, W., Topp, C.N., Zhong, C.X., Jiang, J., Dawe, R.K. and Parrott, W.A. (2007) Transformation of rice with long DNA-segments consisting of random genomic DNA or centromere-specific DNA. *Transgenic Res*, **16**, 341-351.

Purnick, P.E. and Weiss, R. (2009) The second wave of synthetic biology: from modules to systems. *Nat Rev Mol Cell Biol*, **10**, 410-422.

Raikhel, N. (1992) Nuclear targeting in plants. *Plant Physiol*, **100**, 1627-1632.

Ravi, M. and Chan, S.W. (2010) Haploid plants produced by centromere-mediated genome elimination. *Nature*, **464**, 615-618.

Reece, R.J. and Ptashne, M. (1993) Determinants of binding-site specificity among yeast C6 zinc cluster proteins. *Science*, **261**, 909-911.

Robinett, C.C., Straight, A., Li, G., Willhelm, C., Sudlow, G., Murray, A. and Belmont, A.S. (1996) In vivo localization of DNA sequences and visualization of large-scale chromatin organization using lac operator/repressor recognition. *J Cell Biol*, **135**, 1685-1700.

Sautter, C., Uze, M. and Potrykus, I. (1999) Single-stranded DNA in the genetic transformation of wheat (Triticum aestivum L.): transformation frequency and integration pattern. *Theoretical and Applied Genetics*, **99**, 487-495.

Shi, J. and Dawe, R.K. (2006) Partitioning of the maize epigenome by the number of methyl groups on histone H3 lysines 9 and 27. *Genetics*, **173**, 1571-1583.

Simons, A., Tils, D., von Wilcken-Bergmann, B. and Muller-Hill, B. (1984) Possible ideal lac operator: Escherichia coli lac operator-like sequences from eukaryotic genomes lack the central G X C pair. *Proc Natl Acad Sci U S A*, **81**, 1624-1628.

Song, J., Bradeen, J.M., Naess, S.K., Helgeson, J.P. and Jiang, J. (2003) BIBAC and TAC clones containing potato genomic DNA fragments larger than 100 kb are not stable in Agrobacterium. *Theor Appl Genet*, **107**, 958-964.

Song, J., Dong, F., Lilly, J.W., Stupar, R.M. and Jiang, J. (2001) Instability of bacterial artificial chromosome (BAC) clones containing tandemly repeated DNA sequences. *Genome*, **44**, 463-469.

Tanaka, Y., Nureki, O., Kurumizaka, H., Fukai, S., Kawaguchi, S., Ikuta, M., . . . Yokoyama, S. (2001) Crystal structure of the CENP-B protein-DNA complex: the DNA-binding domains of CENP-B induce kinks in the CENP-B box DNA. *EMBO J*, **20**, 6612-6618.

Urnov, F.D., Rebar, E.J., Holmes, M.C., Zhang, H.S. and Gregory, P.D. (2010) Genome editing with engineered zinc finger nucleases. *Nat Rev Genet*, **11**, 636-646.

Wang, C.J., Harper, L. and Cande, W.Z. (2006) High-resolution single-copy gene fluorescence in situ hybridization and its use in the construction of a cytogenetic map of maize chromosome 9. *Plant Cell*, **18**, 529-544.

Wolfgruber, T.K., Sharma, A., Schneider, K.L., Albert, P.S., Koo, D.H., Shi, J., . . . Presting, G.G. (2009) Maize centromere structure and evolution: sequence analysis of centromeres 2 and 5 reveals dynamic Loci shaped primarily by retrotransposons. *PLoS Genet*, **5**, e1000743.

Yan, H., Talbert, P.B., Lee, H.R., Jett, J., Henikoff, S., Chen, F. and Jiang, J. (2008) Intergenic locations of rice centromeric chromatin. *PLoS Biol*, **6**, e286.

Table 2.1 Size estimates of ABS arrays by fiber-FISH.

Fiber-FISH images were measured in micrometers and converted to kilobases using a 3.21 kb/μm conversion rate.

		Average Length (μm)	SD	n	Estimated Length (kb)
ABS4		357.04	55.47	22	1146.1
ABS7		98.11	9.17	24	314.93
ABS3	Insert 1	8.76	1.01	31	28.12
	Gap	34.1	2.31	31	109.46
	Insert 2	16.67	2.09	31	53.51
	Total	59.53	3.66	31	191.09

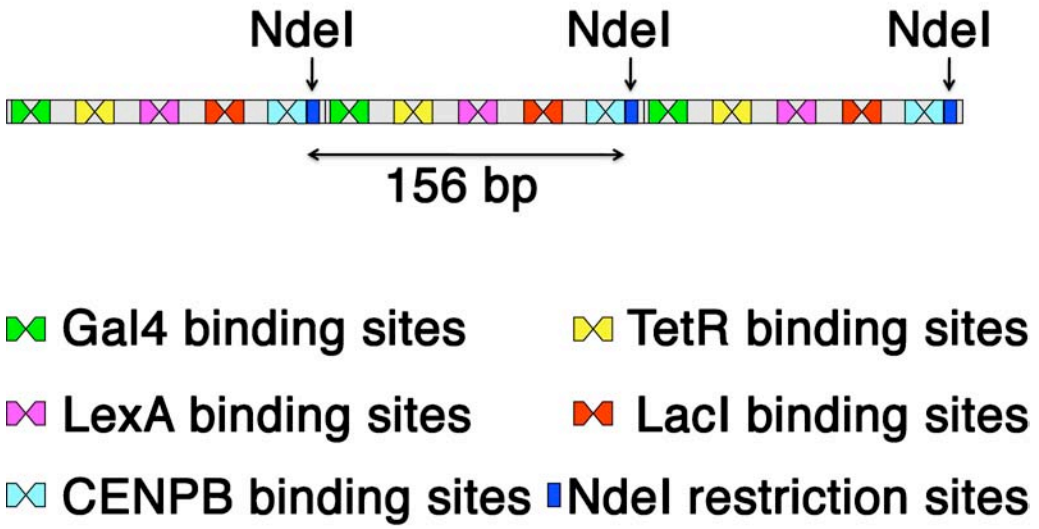


Figure 2.1 Design of the Arrayed Binding Sites (ABS).

A trimer of the 156 bp monomer is shown. Binding sites for Gal4, TetR, LexA, LacI and CENPB, as well as restriction sites of NdeI were designed into each monomer of the array. Outside of the six constrained sequences, bases were chosen at random.

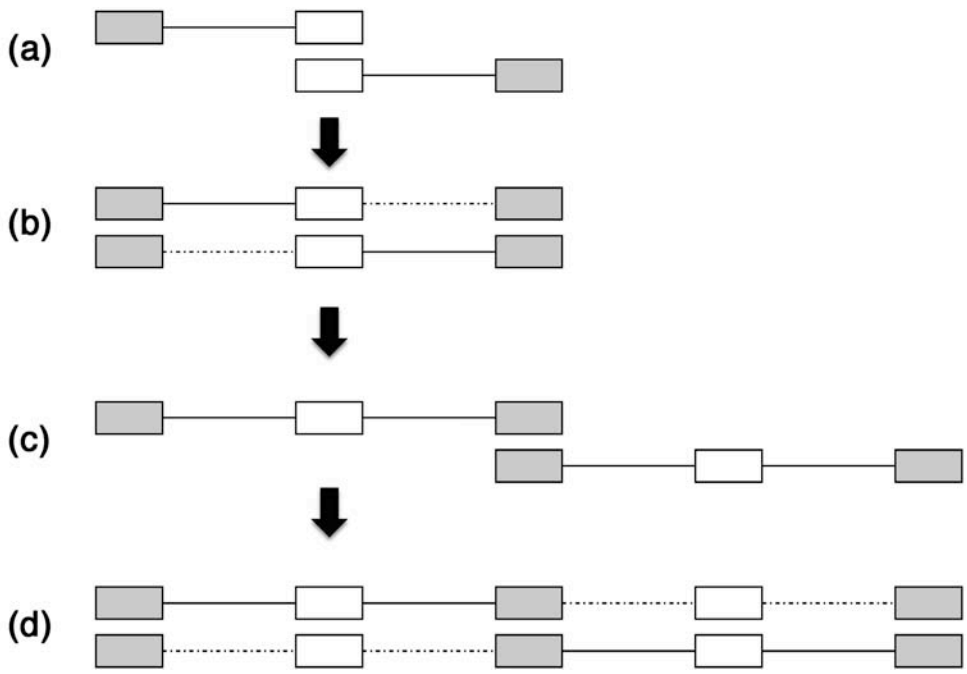


Figure 2.2 Scheme of overlap extension PCR.

Two long oligonucleotides (98 and 99 bp) overlapping at both 5' (grey boxes) and 3' (white boxes) ends were used in the PCR reaction to generate long repeat arrays. Upon denaturing, complementary oligos attach to each other at homologous regions (a and c) and extend their corresponding strands (b and d). Subsequent cycles further extend the array.

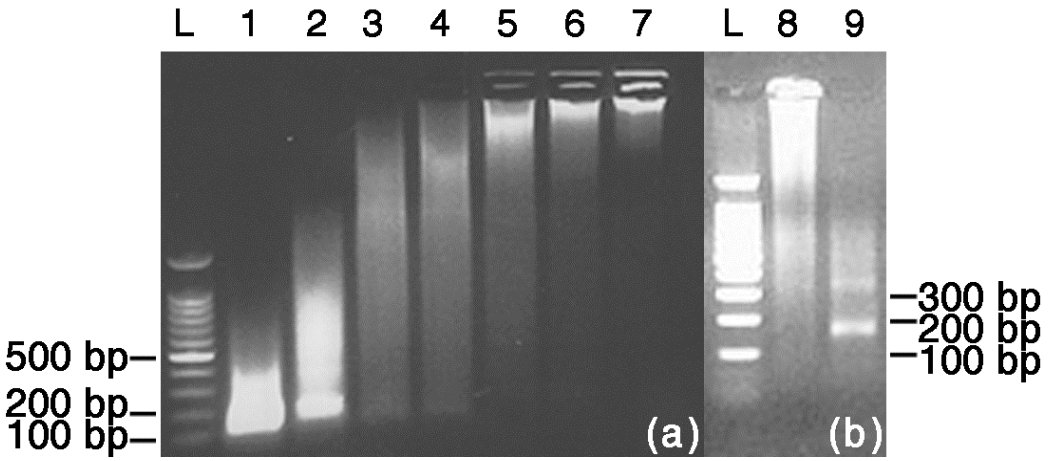


Figure 2.3 Synthesis of long repeat arrays with expected tandem structure.

(a) Gel electrophoresis of the overlap extension PCR products using a series dilution of primers. L, DNA ladder; lane 1-6, PCR products using primer concentrations of: lane 1, 0.4 μM; lane 2, 0.2 μM; lane 3: 0.1 μM; lane 4, 0.04 μM; lane 5, 0.02 μM; lane 6, 0.01 μM; lane 7, positive control using products of lane 6 as templates for a second round of PCR.

(b) Gel electrophoresis of the PCR products before (lane 8) and after (lane 9) restriction digest with NdeI. L, DNA ladder. The PCR products were cleaved into 156 bp monomer and 312 bp dimers.

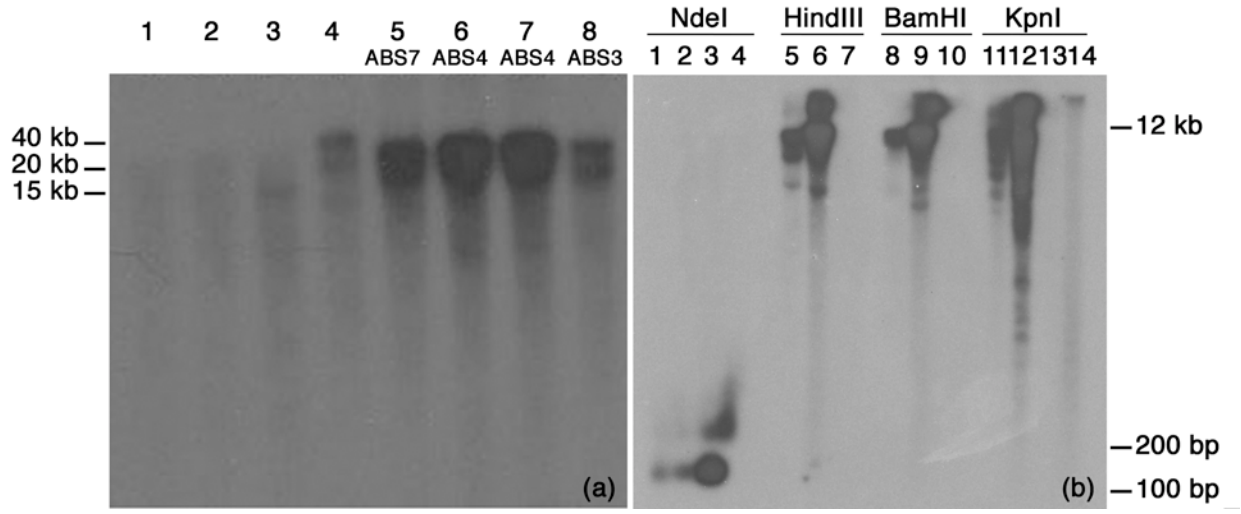


Figure 2.4 Southern blots analysis showing the size of the integrated ABS repeat arrays.

(a) Comparison of different transformed lines. Genomic DNAs from wild type (lane 1) and seven transgenic lines (lane 2-8) were digested with EcoRI and probed with the ABS monomer. Five lines (lane 4-8) showed bright bands at a molecular weight greater than 40 kb. Three of these were further pursued and renamed ABS3, ABS4, and ABS7 as noted (lanes 6 and 7 were later shown to be clones of the same event).

(b) Demonstration that ABS arrays are composed of tandem repeats. Genomic DNAs from wild type (lanes 4, 7, 10, 13), ABS3 (lanes 2, 5, 8, 11) and ABS4 (lanes 3, 6, 9, 12) were digested with NdeI, HindIII, BamHI and KpnI and probed with ABS monomers. PCR products digested with NdeI (lane 1) and KpnI (lane 14) were loaded as controls to indicate the size.

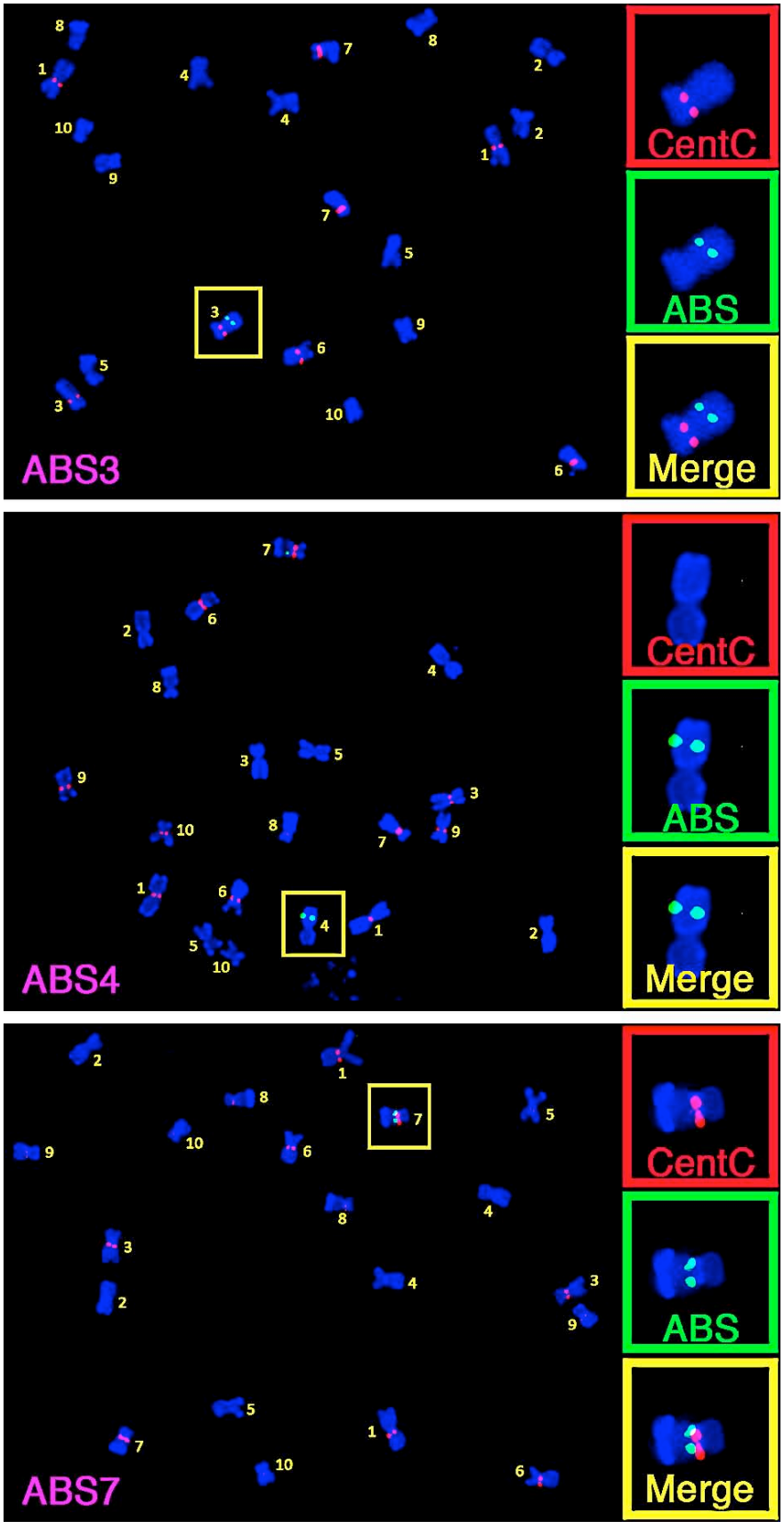

ABS3
CentC
ABS
Merge
ABS4
CentC
ABS
Merge
ABS7
CentC
ABS
Merge

Figure 2.5 Localization of ABS arrays by fluorescence in situ hybridization.

Maize mitotic chromosome spreads were hybridized with endogenous centromeric satellite CentC (red) and synthetic ABS repeats (green). Each chromosome is identified by its number. Chromosomes with ABS arrays are enlarged as insets to show their positions on the long arms of chromosome 3, 4 and 7. CentC quantity varies among chromosomes and lines, and the near absence of CentC signal on chromosome 4 has been documented previously (Kato *et al.* 2004).

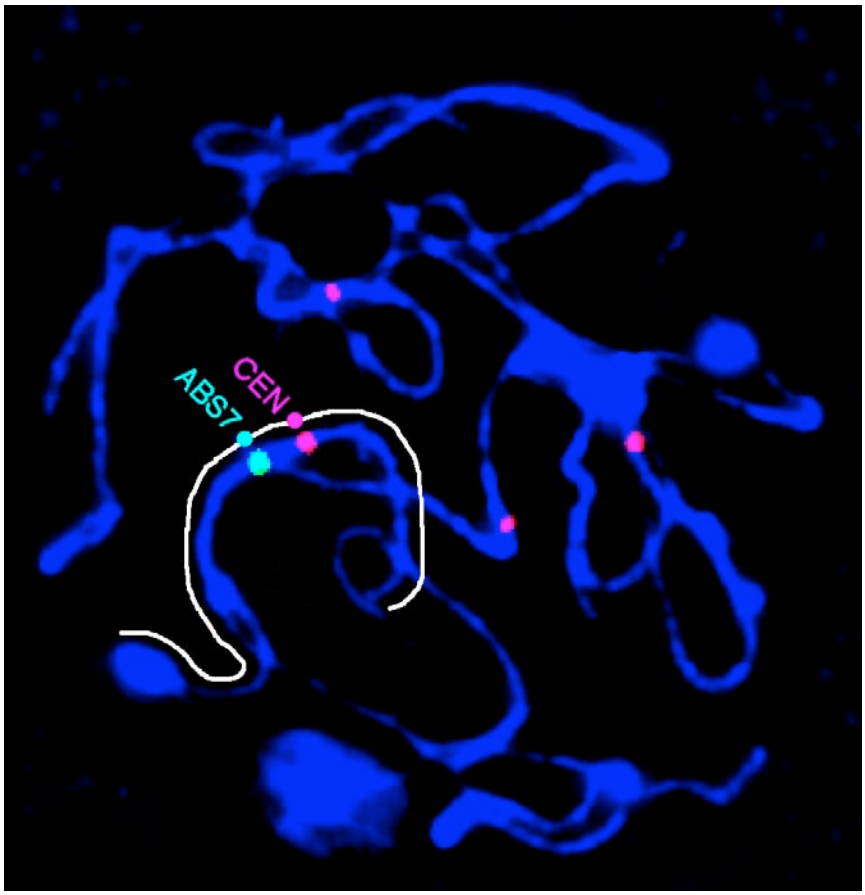


Figure 2.6 Localization of ABS7 in the pericentromeric heterochromatin of chromosome 7. Maize pachytene chromosomes were hybridized with endogenous centromeric satellite CentC (red) and synthetic ABS repeats (green). Chromosome 7 is identified based on the large knob at the end of the long arm and is depicted by a white curve.

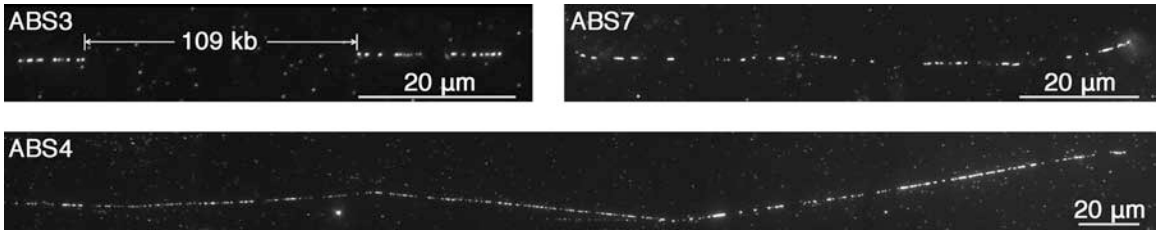


Figure 2.7 Size estimation of ABS arrays by fiber-FISH.

Stretched DNA molecules from the ABS lines were hybridized with synthetic ABS repeat probes (white). ABS4 and ABS7 are continuous or nearly continuous, while ABS3 is composed of two arrays separated by a 109 kb gap. Scale bar is 20µm.

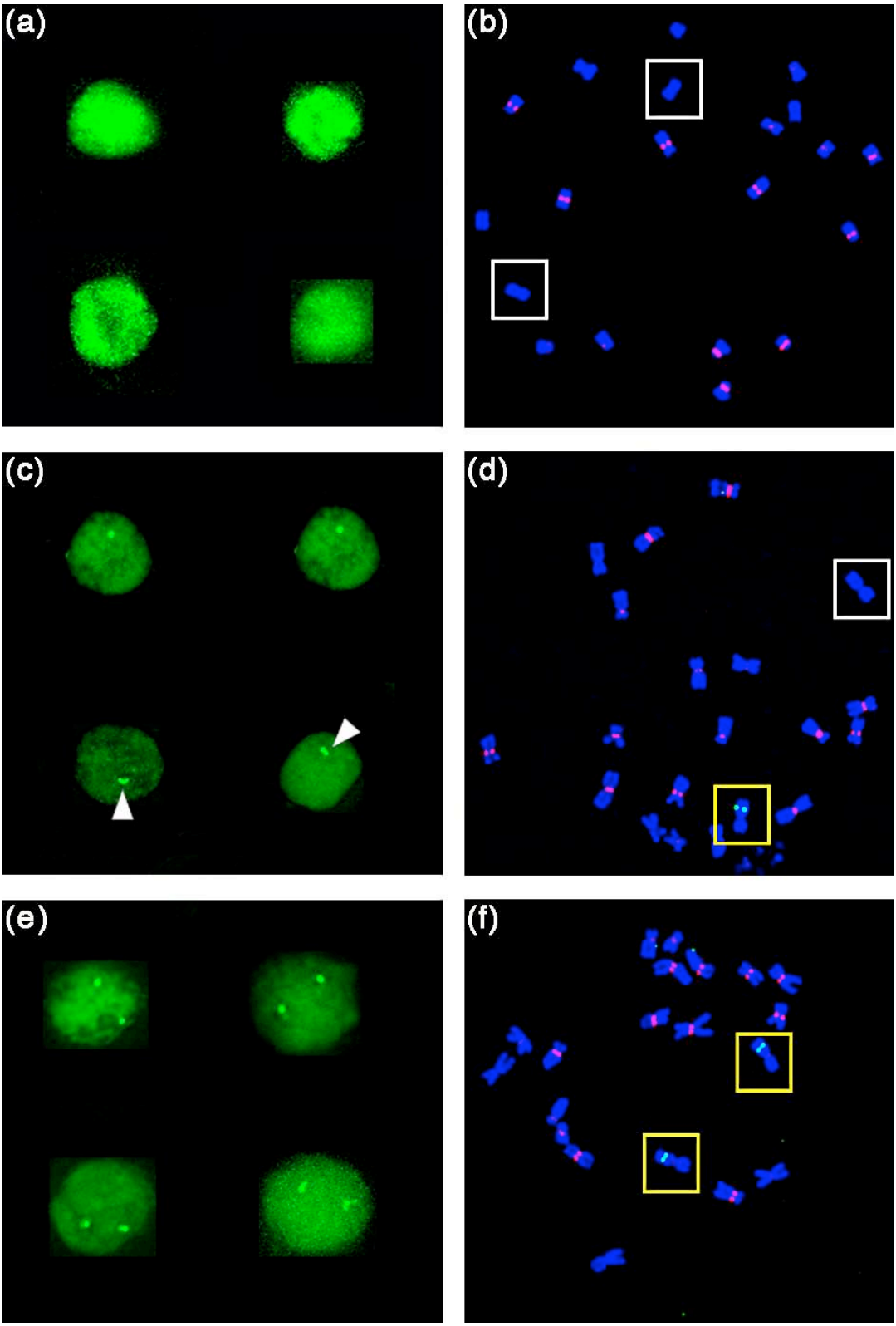

(a)
(b)
(c)
(d)
(e)
(f)

Figure 2.8 Tethered LacI-YFP in ABS4 segregating lines.

(a, c, e) YFP fluorescence in ABS4 segregating lines. The green spots represent LacI-YFP proteins localization at ABS4 arrays, as confirmed by genotyping and FISH. The two closely paired dots in the bottom row of c (arrowheads) represent sister-chromatids at the G2 phase of the cell cycle.

(b, d, f) FISH on ABS4 segregating lines. Maize mitotic spreads were hybridized with endogenous centromeric satellite CentC (red) and synthetic ABS repeats (green). Each chromosome 4 is encircled in a white box (without ABS) or yellow box (with ABS). The number of YFP spots matches the number of ABS loci in the line: No ABS (a-b), ABS heterozygous (c-d, one spot) and ABS hemizygous (e-f, two spots).

CHAPTER 3

TETHERING KINETOCHORE PROTEINS CENH3 AND CENPC INDUCES FAILURES IN CHROMOSOME SEGREGATION BUT NOT NEOCENTROMERE FORMATION IN MAIZE[2]

[2] Han Zhang, Chris Topp, Sarah Wolf, R. Kelly Dawe. Manuscript to be submitted to *Genetics*.

Abstract

Centromeres are critical to maintain genome stability and ensure chromosome inheritance. Although centromeric DNAs are highly divergent across species, proteins necessary to form kinetochores are quite conserved. Among them, CENH3 and CENPC are considered to be the foundation proteins that specify centromeric chromatin and establish a scaffold for kinetochore assembly. It has been shown that under artificial conditions, CENH3 or CENPC can be sufficient to induce neocentromere formation in animal cell cultures. Here we described similar tests in plants. We showed that in maize overexpressing CENH3 and CENPC fused to varied DNA binding domains resulted in high levels of centromere misdivisions and chromosome segregation defects. Although these experiments were designed to induce neocentromeres on chromosome arms we found no evidence of new centromere formation. Rather, out data suggest that fusion tags causes strong dominant-negative affects on CENH3 and CENPC during mitosis and meiosis.

Introduction

Centromeres are the chromosomal domains that are responsible for kinetochore assembly and microtubule attachment. In eukaryotes, centromeres are essential for proper segregation of chromosomes and inheritance of their genetic contents. Despite these conserved functions, centromeric DNA across species diverges so rapidly that no universal motif has been identified. Centromeres can also form spontaneously in chromosomal regions lacking centromere-specific sequences (referred to as "neocentromeres") and yet display all the properties of regular centromeres (Voullaire *et al.* 1993, Karpen and Allshire 1997, Nasuda *et al.* 2005, Topp *et al.* 2009).

It is now well established that centromeres are specified by a histone H3 variant, named Centromere Histone H3 (CENH3). CENH3 is present in the functional centromeres of all eukaryotes studied, including neocentromeres (WARBURTON *et al.* 1997). The incorporation of CENH3 marks a specialized chromatin state that is distinct from euchromatin or heterochromatin. The unique properties of CENH3-containing nucleosomes have been proposed to provide the foundation for centromere identity and maintenance (BLACK and CLEVELAND 2011).

Like conventional histone H3, CENH3 is mainly composed of two domains: an amino-terminal tail that is dramatically divergent from H3, and a histone fold domain that is relatively well conserved (HENIKOFF *et al.* 2001). A small region containing loop1 and helix 2 within the histone core of CENH3, known as the CENPA centromere-targeting domain (CATD), has been shown to be both required and sufficient for centromere targeting (Vermaak *et al.* 2002, Black *et al.* 2004, Black *et al.* 2007). On the contrary, the function of the N-terminal tail is still unclear

and seems to vary among species (Morey *et al.* 2004, Wieland *et al.* 2004, Lermontova *et al.* 2006, Ravi *et al.* 2010, Ravi *et al.* 2011).

Although much is known about CENH3 itself, this has not led to a full understanding of centromere establishment. Since endogenous centromeres are always maintained at the same positions, researchers have mainly focused on dissecting neocentromeres to address the question how centromeres are established. Naturally occurring neocentromeres are rare and mostly result from genome rearrangements that cause losses of original centromeres (Nasuda *et al.* 2005, Marshall *et al.* 2008, Topp *et al.* 2009). It is likely that neocentromeres arise as a natural means to maintain genome stability and prevent chromosome loss. As individuals carrying neocentromeres are not necessarily phenotypically different from wild types, identification of these neocentromeres requires extensive cytological examination.

Neocentromeres can also be induced experimentally by targeting CENH3 to non-centromeric loci. It has been reported in multiple organisms that as the expression level of CENH3 increases, the protein spreads to chromosome arms (Van Hooser *et al.* 2001, Heun *et al.* 2006, Collins *et al.* 2007, Gascoigne *et al.* 2011). Ectopically localized CENH3 was able to recruit a subset of kinetochore components, yet insufficient to direct microtubule attachment or chromosome segregation (Van Hooser *et al.* 2001, Gascoigne *et al.* 2011). In some rare cases, CENH3 localized on the chromosome arms is sufficient to assemble nearly complete kinetochores and mediate poleward movement of otherwise acentric fragments (Heun *et al.* 2006, Olszak *et al.* 2011). Although CENH3 was broadly localized throughout euchromatin in these experiments, only a few loci supported microtubule attachment.

The fact that incorporation of CENH3 may or may not cause neocentromere formation leads to the speculation that the assembly of neocentromeres may be limited by the availability

of other kinetochore proteins. Indeed, it has been shown in *Drosophila* that the incorporation of excess CENH3 is prevented by limiting the levels of CAL1, a CENH3 chaperone (SCHITTENHELM *et al.* 2010). CENH3 chaperones complex with soluble CENH3 and histone H4, transiently localize at centromeres, and load pre-nucleosomal CENH3 onto newly replicated DNA (Mizuguchi *et al.* 2007, Dunleavy *et al.* 2009, Foltz *et al.* 2009, Cho and Harrison 2011, Mellone *et al.* 2011, Zhou *et al.* 2011). Excess CENH3 that is not bound to chaperones is unlikely to be deposited. In contrast, when CENH3 chaperones were provided in excess, CENH3 may be targeted ectopically and induce neocentromere formation on the chromosome arms (BARNHART *et al.* 2011).

Neocentromere formation may also require high local density of CENH3. In the budding yeast, excess CENH3 is regulated by ubiquitin-mediated proteolysis (Collins *et al.* 2004, Hewawasam *et al.* 2010, Ranjitkar *et al.* 2010). Ectopically localized CENH3 is ubiquitylated and targeted for proteolysis while centromerically localized CENH3 is protected by their chaperones from degradation. In *S. cerevisiae*, overexpressed CENH3 localized to the euchromatic regions only when the proteolysis pathway is disrupted (Hewawasam *et al.* 2010, Ranjitkar *et al.* 2010).

Another way to induce neocentromere formation is through the activity of Centromere protein C (CENPC). CENPC is a highly conserved protein that has been found in plants, animals, and yeasts. CENPC lies at the innermost face of the kinetochore, and in maize and other species interacts directly with centromeric DNA (DU *et al.* 2010). In addition, CENPC interacts with CENH3 on its C-terminal end (TRAZZI *et al.* 2009) and outer kinetochore proteins via N-terminal sequences (Przewloka *et al.* 2011, Screpanti *et al.* 2011). Therefore, CENPC

functions as the principal linkage between inner centromeric chromatin and outer kinetochore proteins (Przewloka *et al.* 2011).

Given its importance, several laboratories have questioned whether CENPC is sufficient for kinetochore assembly. General overexpression of CENPC in chicken cells caused excess CENPC localization on chromosome arms but neocentromere formation was not observed (FUKAGAWA *et al.* 1999). In contrast, CENPC targeted to a specific locus outside centromeres is sufficient to assemble de novo kinetochores (Gascoigne *et al.* 2011, Przewloka *et al.* 2011). In one case, the N-terminal protein binding domain of CENPC was fused with a centrosome targeting protein (Przewloka *et al.* 2011). It was found that core components of the kinetochores but not CENH3 were recruited to the centrosomes (Przewloka *et al.* 2011). In another study, CENPC was targeted to synthetic binding motifs integrated on chromosome arms and was shown to interact with microtubules and function in chromosome segregation (Gascoigne *et al.* 2011).

In the cases described above, although partially or fully functional neocentromeres were assembled transiently, the mitotic (or meiotic) persistence of neocentromeres remained unexamined because of the limitations of the in vitro culture system. In contrast, plants are known to tolerate high levels of chromosome missegregation while still producing flowers and seeds (LI and DAWE 2009). Both CENH3 and CENPC have been characterized in maize (Dawe *et al.* 1999, Zhong *et al.* 2002). Therefore, plants provide a unique model system for neocentromere study.

Here we described studies designed to induce neocentromere formation in plants. We fused CENH3 and CENPC with varied DNA binding domains and expressed them in plants containing an array of DNA binding sites. We observed high levels of centromere misdivisions, lagging chromosomes and aneuploidies. Careful cytological analysis of these transgenic plants

revealed that although CENH3 or CENPC may be transiently targeted to non-centromeric loci, they rarely resulted in neocentromeres. Tethered CENH3 or CENPC are more likely to destabilize the function of existing kinetochores than to induce the formation of new ones. Our data suggest that plant kinetochores are designed to be exceptionally stable features that rarely move even under strong selection.

Results

Attempts to mis-target CENH3 to an ectopic site

Considering that the formation of de novo kinetochores may require high local density of CENH3 (Heun *et al.* 2006, Olszak *et al.* 2011), we attempted to concentrate CENH3 proteins specifically on a particular locus on the chromosome arm. We previously constructed a maize transgenic line containing a 1.1 Mb tandem repeat array (ABS4) made of 156 bp repeat monomers, each consisting of a Gal4 binding motif, a LexA binding motif, a LacI binding domain and a CENPB-box (Chapter 2). ABS4 is estimated to contain over 7000 copies of each binding module localized to a single locus in the middle of chromosome 4L. This line was designed to be used as a tethering site, similar to the approach used by several recent authors (Barnhart *et al.* 2011, Gascoigne *et al.* 2011) to show that CENPC and CENH3 chaperones are sufficient to organize kinetochore regions.

To test the idea that CENH3 can be targeted to a binding array, the maize CENH3 gene was fused with the Gal4 binding protein, the LexA binding domain, or the CENPB binding protein at its N-terminal end (Figure 3.1a). Only the minimal DNA binding domains (74 amino acids for Gal4, 72 amino acids for LexA and 125 amino acids for CENPB) required for binding

capacity (Fogh *et al.* 1994, Aoyama and Chua 1997, Tanaka *et al.* 2001) were fused with CENH3. These domains are similar in size as the native CENH3 tails, which vary from 20 to 200 amino acids and are highly divergent across lineages (Henikoff *et al.* 2001). To differentiate transgenic CENH3 from endogenous protein, we replaced the N-terminal tail of maize CENH3 with its corresponding sequence from oat (60 amino acids of the oat CENH3, Figure 3.1a), which can be recognized specifically by an oat peptide antibody (Topp *et al.* 2009). Fluorescent protein tags were not used for detection because of their large sizes (for example, GFP is of 238 amino acids). These chimeric CENH3 proteins were expressed under the control of either the 35S promoter for constitutive overexpression or 1 kb of the endogenous promoter from CENH3.

We first confirmed the expression of the chimeric CENH3 genes by either RT-PCR (Figure 3.1b, Table 3.1) or western blot using the oat peptide antibody, which recognizes the transgenic proteins but not the endogenous maize CENH3 (Figure 3.1c, Table 3.1). Immuno-fluorescence was also performed to detect the localization of the chimeric CENH3 proteins using the same peptide antibody (Table 3.1). Transgenic CENH3 proteins co-localized with endogenous CENPC proteins, suggesting that they were targeted to the native kinetochores (Figure 3.2). The localization pattern indicates that the replacement of part of the endogenous CENH3 tail with its oat counterpart, as well as the addition of a larger tethering protein at the amino terminus, does not affect CENH3 deposition. Our data are consistent with previous reports showing that the histone fold domain is sufficient for CENH3 targeting (Morey *et al.* 2004, Lermontova *et al.* 2006).

We then tested whether the chimeric CENH3 proteins can be targeted to the binding array localized on the chromosome 4L. Plants expressing transgenic CENH3 were crossed to plants homozygous for ABS4. Progenies of the cross were examined by immunoFISH to detect the

localization of the transgenic proteins and the ABS4 arrays simultaneously. Unfortunately, the oat peptide antibody we used to detect the localization of the transgenic CENH3 proteins did not provide fluorescent signals strong enough to be combined with FISH procedures. Other CENH3 antibodies that have been shown to recognize maize CENH3 (Zhong *et al.* 2002, Jin *et al.* 2004) do not recognize the transgenic CENH3 proteins (Figure 3.1a). For this reason, we decided to examine other kinetochore proteins instead of CENH3 to see if these proteins are recruited to the binding arrays by tethered CENH3.

An anti-CENPC antibody (Dawe *et al.* 1999) and an anti-MIS12 antibody (LI and DAWE 2009) have been shown to recognize maize CENPC and MIS12, which are key kinetochore proteins conserved in eukaryotes. We reasoned that tethered CENH3 may recruit one or both of these proteins to the ABS4 site. Of 26 different transgenic lines examined, only one showed co-localization of CENPC with ABS4 (Figure 3.3). This form of co-recruitment event is rare, as co-localization was only detected in 3 out 17 cells from one transgenic line. The vast majority of the CENPC proteins were targeted at endogenous kinetochores and not co-localized with the ABS4 loci on the chromosome arms (Figure 3.4). It also appeared that the recruitment occurred in a transient fashion, since in all three cases, CENPC and ABS4 were co-localized on early pachytene/diplotene chromosomes. We did not observe any co-localization of MIS12 with ABS4 (data not shown). This result is consistent with previous studies showing that ectopically localized CENH3 was able to recruit a subset of kinetochore components, including CENPC, yet insufficient to assemble complete kinetochores or function autonomously (Van Hooser *et al.* 2001, Gascoigne *et al.* 2011).

Attempts to mis-target CENPC to an ectopic site

We also tested whether CENPC can be tethered to the chromosome arms and induce neocentromere formation. Two constructs of CENPC were made for this purpose. In one construct, CENPC was fused with LacI binding protein on the N-terminus (C-terminal fusions do not localize to centromeres of plants (Du *et al.* 2010), and driven by the 35S promoter (Figure 3.5a). In the second construct, we made use of the fact that the DNA binding domain on maize CENPC is limited to exons 9-12 (Du *et al.* 2010). CENPC localization is only slightly reduced when the DNA exon 9-12 binding domain is removed or replaced (Du *et al.* 2010). We replaced this DNA binding domain with the Gal4 binding domain (Figure 3.5b), and drove the construct using 1 kb of the endogenous CenpcA promoter (Dawe *et al.* 1999). The expression of the transgenic CENPC gene was confirmed using RT-PCR (Figure 3.5c).

As antibodies against LacI or Gal4 are not available for immunofluorescence, we cannot detect the localization of the fusion proteins specifically. Instead, we used an anti-CENPC antibody to detect both endogenous and transgenic CENPC. In the plants containing both transgenic CENPC and ABS4, no CENPC signals were observed on the ABS4 loci in interphase or meiotic cells (Figure 3.6). Antibodies against MIS12 were also applied, but no co-localization of MIS12 proteins and ABS4 was detected (data not shown). Although these data are far from definitive, they suggest that neither CENH3 nor CENPC are stable on chromosome arm positions.

Tethered CENH3 and CENPC causes severe meiotic defects

It has been reported that replacement of endogenous CENH3 with transgenic CENH3 with fusion tags on the N-terminus causes sterile but viable *Arabidopsis* plants, suggesting that

meiosis is affected by the fusion tags of the CENH3 proteins (Ravi and Chan 2010, Lermontova *et al.* 2011, Ravi *et al.* 2011). We wondered whether similar defects occur in maize plants expressing tethered CENH3 or CENPC. Most maize plants expressing chimeric CENH3 or CENPC proteins did not display obvious whole plant phenotypes. However, when they were crossed to different genetic backgrounds in large segregating populations, dwarf and sterile plants were observed sporadically (data not shown), suggesting that aneuploid progeny were produced (similar to (LI and DAWE 2009)). Some plants also displayed cracked leaves, which are indicative of cell death (Figure 3.7). The nature of these phenotypes suggests that both meiosis and mitosis are affected in the transgenic CENH3/CENPC lines.

We first examined maize meiocytes from plants expressing tethered CENH3 or CENPC to see whether kinetochore defects can be observed in meiosis (Table 3.1). On a single branch of maize tassels, all stages of meiosis occur in a linear way so that the entire cell division process, from pre-meiosis to the formation of microspores, can be easily observed. No cytological defects were observed before or during prometaphase when outer kinetochore assembly begins (Figure 3.8, a and b). In contrast, chromosome alignment (metaphase) and segregation (anaphase) were severely affected (Figure 3.8, c-f), and resulted in chromosome loss and mini-nuclei (Figure 3.8, g and h).

Cell division is a precise cellular process with each step accurately monitored. At metaphase, the chromosomes align because of stable interactions between kinetochores and microtubules. In our transgenic CENH3/CENPC lines, however, some chromosomes failed to align on the metaphase plates, indicating that the interaction between kinetochores and microtubules were disrupted (Figure 3.8, c and d). Once chromosomes are aligned properly, kinetochores will be pulled towards opposite poles so that either homologs (anaphase I) or sister

chromatids (anaphase II) can be separated. In maize plants expressing fusion CENH3/CENPC, we observed many lagging chromosomes, which were stalled in the middle of the division plane at both anaphase I and anaphase II (Figure 3.8, e and f). Evidence of sister chromatid pre-separation at anaphase I was also observed (data not shown). Taken together, these observations suggest that meiosis is severely affected when CENH3 or CENPC fusion tags are expressed.

Tethered CENH3 causes severe centromere misdivisions

Although CENH3 fusions proteins do not appear to have any negative affects on mitosis in *Arabidopsis* (Ravi and Chan 2010, Lermontova *et al.* 2011, Ravi *et al.* 2011), it has been reported that GFP-tagged CENH3 causes chromosome missegregation, aneuploidy and apoptosis in mice (KALITSIS *et al.* 2003). We examined whether mitosis is affected by tethered CENH3 or CENPC in maize. Root tip cells from plants expressing chimeric CENH3 or CENPC proteins were examined by FISH using a centromere-specific probe. One outcome of kinetochore failure at mitosis is centromere misdivision, a phenomenon where chromosomes break into two parts near or within centromere (Sears 1952). We observed numerous centromere misdivision events mostly in plants expressing fusions with CENH3 and occasionally with CENPC (Figure 3.9, Table 3.1). The frequency of misdivision increased when the transgenic CENH3 constructs were driven by the 35S promoter, and varied between different transgenic lines. In one particular line driven by the 35S promoter, misdivisions were observed in more than half of the plants examined (Table 3.1).

In all cases the breakage events appeared to have occurred within centromere cores to produce telocentric chromosomes (Figure 3.9, a and b). Some maize chromosomes contain low quantities of the CentC satellite repeats (KATO *et al.* 2004) and in these cases we could not detect

CentC on the misdivision derivatives (Figure 3.9, b and c). Breakage appeared to have occurred within the primary constrictions, as judged by the size and shapes of the chromosome arms, although we cannot rule out the possibility that misdivisions occurred at loci close to the native centromeres. In some plants, partially separated centromeres were also observed (Figure 3.9d). The nature and frequency of misdivision events strongly suggests that the N-terminal region of CENH3 has functional roles in kinetochore assembly, and that fusion tags cause strong dominant negative affects that interfere with chromosome alignment and segregation.

Discussion

The primary intent of this study was to replicate and extend previous research in animals showing that overexpression or tethering of CENH3 and CENPC can induce neocentromeres at ectopic loci. To this end, both CENH3 and CENPC were overexpressed as fusion constructs that were designed to tether the proteins to an engineered repeat array on chromosome 4 (ABS4). Our observations revealed that the CENH3 fusion proteins rarely localized to the corresponding binding motifs on the chromosome arm (Figure 3.4), but rather localized to endogenous kinetochores (Figure 3.2). In one plant expressing a CENPB-CENH3 fusion protein we observed that CENPC was recruited to ABS4 (Figure 3.3), but the frequency of the targeting/recruitment events was extremely low. Likewise, tethering CENPC with the LacI protein or the Gal4 binding domain failed to target CENPC to non-centromeric loci (Figure 3.6).

The fact that neocentromeres were not formed even under strong selection can be attributed to several possibilities. First, in other species CENH3 must be deposited by a chaperone, and the availability of the chaperone restricts CENH3 localization. CENH3

deposition factors have been identified in yeast, *Drosophila* and human (Mizuguchi *et al.* 2007, Dunleavy *et al.* 2009, Foltz *et al.* 2009, Mellone *et al.* 2011). Although functionally conserved, little homology can be detected between known CENH3 chaperones (Sanchez-Pulido *et al.* 2009, Mellone *et al.* 2011). Plant CENH3 chaperones have not yet been identified by sequence homology, but it is likely that functional homologs exist. We may be more successful if we design experiments to overexpress or tether the CENH3 chaperone.

Another potential explanation is that the tethered CENH3 was targeted for proteolysis. It has been reported that in budding yeast, an E3 ubiquitin ligase called Psh1 regulates excess CENH3 and restricts CENH3 localization. Ectopically localized CENH3 is ubiquitylated by Psh1 and degraded by proteolysis. In *psh1* mutants, however, CENH3 localized to the euchromatic regions when overexpressed (Hewawasam *et al.* 2010, Ranjitkar *et al.* 2010). Similar proteolysis pathway may be involved in CENH3 regulation in plants to prevent neocentromere formation.

It is also important to consider the fact that dicentric chromosomes are naturally unstable. Dicentric chromosomes are rare in both plants and animals, and even when they form, one of the two is usually inactivated shortly thereafter (Therman *et al.* 1986, Han *et al.* 2006). By the time we assessed the formation of neocentromeres, the plants had undergone many cycles of cell divisions. It is possible that neocentromeres were formed in the fertilized eggs, but were inactivated during development to prevent chromosome missegregation. Neocentromeres may be more likely to form and transmit under conditions where chromosomes have otherwise lost their true kinetochores (Nasuda *et al.* 2005, Topp *et al.* 2009).

A final, important consideration is that the fusion proteins we used have dominant negative affects on the natural kinetochores. We presume that the failures we observed during

meiotic chromosome segregation, and frequent misdivision defects induced during mitosis (Figure 3.7 and Figure 3.8) would have also diminished the likelihood of inducing neocentromeres at an ectopic site. The effects of fusion tags on kinetochore proteins have been reported previously (Kalitsis *et al.* 2003) but not widely recognized. Recent studies in *Arabidopsis* re-emphasized the importance of an intact N-terminus of CENH3 during meiosis (Lermontova *et al.* 2011, Ravi *et al.* 2011). The defects are unlikely to be caused by specific tags, since all CENH3 constructs involving different tags displayed similar results. The fact that similar cytological defects were observed in plants expressing CENPC with tethered proteins (Table 3.1) further supports the idea that kinetochore functions were disrupted in the transgenic plants.

Taken together our data suggest that CENH3 or CENPC may be targeted to non-centromeric loci in plants by tethering, but these events tend to be transient and not sufficient to form stable neocentromeres. Plants do not seem to support the same level of CENH3/CENPC-induced de novo kinetochore formation that was originally observed in animal cell cultures. Rather, in plants, overexpressed CENH3 and CENPC fusion proteins are most likely to interfere the function of existing kinetochores, a negative trait that would presumably reduce the activation of new kinetochores.

Materials and Methods

Construct preparation and transformation

Maize CENH3 constructs tethered with Gal4 and LexA (see supplemental information for complete sequences) were synthesized by IDT (http://www.idtdna.com/) in a pUC plasmid

and cloned into pENTR/D vector (Invitrogen) by restriction digest with NotI and AscI. The sequence-confirmed constructs were then recombined into pEarleygate vector pEG100 (an Invitrogen Gateway vector derivative; (Earley *et al.* 2006)) for overexpression driven by the 35S promoter. Maize CENH3 endogenous promoter ~1.7 kb upstream of the coding sequence in maize HiII plants was cloned into TOPO vector (Invitrogen) using the following primers: CENH3 promoter 2F (5'-TTCGAAGGCAATTGCAGTAGTGCCT-3') and CENH3 promoter R (5'-CTCGAGCGCGGTGGGCGCCTCGCA-3'). The promoter sequence was then digested with BstBI and XhoI to replace the 35S promoter in the 35S promoter-driven CENH3 constructs for gene-specific expression with endogenous CENH3 promoter.

The CENPB-CENH3 construct and a CENPC construct with exon 9-12 replaced by the Gal4 binding domain was synthesized by DNA2.0 (www.dna20.com/) in pDONR vectors and recombined into pEarleygate vector pEG301 for gene-specific expression. Both constructs include the endogenous CENH3 promoter extending 1 kb upstream of the start codon (see supplemental information for complete sequences)

The maize LacI-CENPC construct (see supplemental information for complete sequence) was constructed in two steps. First the LacI gene was cloned into TOPO vector (Invitrogen) from a previously constructed plasmid (Chapter 2) using the following primers: LacI_BamHI1F (5'-GGATCCCACCATGAAGAAGAAGAGAA-3') and LacI_EcoRI1R (5'-GAATTCCTGCCCGCTTTCCAGTCGGG-3). The resulting plasmid was then digested with BamHI and EcoRI and ligated into the pENTR1A vector (Invitrogen). Second, the CENPC gene cloned in pENTR/D vector (Du *et al.* 2010) was digested with NotI and PvuI and ligated into the pENTR1A vector containing LacI. The resulting clone was recombined into the pEarleygate vector pEG100 for overexpression driven by the 35S promoter.

All the final sequence-confirmed constructs in pEarleygate binary vectors were introduced into *Agrobacterium tumefaciens* strain C58C1 (a gift from Richard Meagher and lab), and then transformed into maize line HiII by the Iowa State University Plant Transformation Facility (http://www.agron.iastate.edu/ptf/). Progenies segregating for transgenic proteins were screened by PCR with the following primers: 35S1F (5'-GGCCCTAACAAGCCCACCAA-3') and 35S1R (5'-GGGCAATGGAATCCGAGGAG-3') for constructs driven by 35S promoters, Gal4-CENH3 1F (5'-GAAGCTGAAGTGCTCCAA-3') and Gal4-CENH3 1R (5'-GGAAGATGAGCAGGAACAGC-3') for Gal4-CENH3, LexA 1F(5'-AGCAAGAGGTGTTCGACCTG-3') and LexA 1R (5'-CCGAACTTGAGCTGCTTCTT-3') for LexA-CENH3, CENPB 1F (5'-CGCGAGCGAGCGCAAGTACG-3') and CENPB 1R (5'-GCCAGCCGTTTGAAGCCGTGA -3') for CENPB-CENH3, LacINLS-F2 (5'-CACCATGAAGAAGAAGAGAAAGGTG-3') and LacI-R2 (5'-CTGCCCGCTTTCCAGTCGGGAA-3') for LacI-CENPC, and Gal4-CENPC 1F (5'-TGAGTTCCATCGAGCAAGCGTGTG-3') and Gal4-CENPC 1R (5'-CGAGATCCTCGCGGGGGAAG-3') for Gal4-CENPC.

RT-PCR and protein blots

RNA from young leaves of transgenic plants were extracted using RNeasy Mini Kit (Qiagen) and reverse transcribed using gene specific primers Gal4-CENH3 1R (5'-GGAAGATGAGCAGGAACAGC-3') for Gal4-CENH3, CENPB 1R (5'-GCCAGCCGTTTGAAGCCGTGA -3') for CENPB-CENH3, LacI-R2 (5'-CTGCCCGCTTTCCAGTCGGGAA-3') for LacI-CENPC, Gal4-CENPC 1R (5'-CGAGATCCTCGCGGGGGAAG-3') for Gal4-CENPC and Superscript III first strand

(Invitrogen). Western blots were performed as previously described (JACKSON *et al.* 2004) using the oat peptide antibody (1:1000).

Cytological preparation and observation

Transgenic and wild type maize plants were grown for two months until they formed about eight fully developed leaf sheaths. Meiocytes were prepared as previously described (SHI and DAWE 2006). Immuno-fluorescence was performed using the following antibodies: oat CENH3 antibodies (Topp *et al.* 2009), maize CENPC antibodies (Zhong *et al.* 2002), or maize Mis12 antibodies (Li and Dawe 2009). ImmunoFISH was performed following the immuno-fluorescence procedure. Slides were observed under fluorescent microscope to validate the quality. After observation, coverslips were removed gently to prevent oil from getting into the slides and rinsed in PBS buffer three times to remove Vectashield. Slides were fixed again in 4% paraformaldehyde diluted in PBS buffer for 10 minutes and rinsed briefly in PBS before performing FISH.

Maize mitotic chromosome spreads were prepared and hybridized with CentC probes as previously described (Kato *et al.* 2004). All fluorescent images were captured using a Zeiss Axio Imager microscope and processed using SlideBook 5.0 software (Intelligent Imaging Innovations, Denver, CO, USA).

Acknowledgments

We thank Richard Meagher for providing the *Agrobacterium tumefaciens* strain, and Harold Garner for technical support. This work was supported by grants 0421671 and 092270 from the National Science Foundation.

References

Aoyama, T. and Chua, N.H. (1997) A glucocorticoid-mediated transcriptional induction system in transgenic plants. *Plant J*, **11**, 605-612.

Barnhart, M.C., Kuich, P.H., Stellfox, M.E., Ward, J.A., Bassett, E.A., Black, B.E. and Foltz, D.R. (2011) HJURP is a CENP-A chromatin assembly factor sufficient to form a functional de novo kinetochore. *J Cell Biol*, **194**, 229-243.

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

Black, B.E., Foltz, D.R., Chakravarthy, S., Luger, K., Woods, V.L., Jr. and Cleveland, D.W. (2004) Structural determinants for generating centromeric chromatin. *Nature*, **430**, 578-582.

Black, B.E., Jansen, L.E., Maddox, P.S., Foltz, D.R., Desai, A.B., Shah, J.V. and Cleveland, D.W. (2007) Centromere identity maintained by nucleosomes assembled with histone H3 containing the CENP-A targeting domain. *Mol Cell*, **25**, 309-322.

Cho, U.S. and Harrison, S.C. (2011) Recognition of the centromere-specific histone Cse4 by the chaperone Scm3. *Proc Natl Acad Sci U S A*, **108**, 9367-9371.

Collins, K.A., Camahort, R., Seidel, C., Gerton, J.L. and Biggins, S. (2007) The overexpression of a Saccharomyces cerevisiae centromeric histone H3 variant mutant protein leads to a defect in kinetochore biorientation. *Genetics*, **175**, 513-525.

Collins, K.A., Furuyama, S. and Biggins, S. (2004) Proteolysis contributes to the exclusive centromere localization of the yeast Cse4/CENP-A histone H3 variant. *Curr Biol*, **14**, 1968-1972.

Dawe, R.K., Reed, L.M., Yu, H.G., Muszynski, M.G. and Hiatt, E.N. (1999) A maize homolog of mammalian CENPC is a constitutive component of the inner kinetochore. *Plant Cell*, **11**, 1227-1238.

Du, Y., Topp, C.N. and Dawe, R.K. (2010) DNA binding of centromere protein C (CENPC) is stabilized by single-stranded RNA. *PLoS Genet*, **6**, e1000835.

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

Earley, K.W., Haag, J.R., Pontes, O., Opper, K., Juehne, T., Song, K. and Pikaard, C.S. (2006) Gateway-compatible vectors for plant functional genomics and proteomics. *Plant J*, **45**, 616-629.

Fogh, R.H., Ottleben, G., Ruterjans, H., Schnarr, M., Boelens, R. and Kaptein, R. (1994) Solution structure of the LexA repressor DNA binding domain determined by 1H NMR spectroscopy. *EMBO J*, **13**, 3936-3944.

Foltz, D.R., Jansen, L.E., Bailey, A.O., Yates, J.R., 3rd, Bassett, E.A., Wood, S., . . . Cleveland, D.W. (2009) Centromere-specific assembly of CENP-a nucleosomes is mediated by HJURP. *Cell*, **137**, 472-484.

Fukagawa, T., Pendon, C., Morris, J. and Brown, W. (1999) CENP-C is necessary but not sufficient to induce formation of a functional centromere. *EMBO J*, **18**, 4196-4209.

Gascoigne, K.E., Takeuchi, K., Suzuki, A., Hori, T., Fukagawa, T. and Cheeseman, I.M. (2011) Induced Ectopic Kinetochore Assembly Bypasses the Requirement for CENP-A Nucleosomes. *Cell*, **145**, 410-422.

Han, F., Lamb, J.C. and Birchler, J.A. (2006) High frequency of centromere inactivation resulting in stable dicentric chromosomes of maize. *Proc Natl Acad Sci U S A*, **103**, 3238-3243.

Henikoff, S., Ahmad, K. and Malik, H.S. (2001) The centromere paradox: stable inheritance with rapidly evolving DNA. *Science*, **293**, 1098-1102.

Heun, P., Erhardt, S., Blower, M.D., Weiss, S., Skora, A.D. and Karpen, G.H. (2006) Mislocalization of the Drosophila centromere-specific histone CID promotes formation of functional ectopic kinetochores. *Dev Cell*, **10**, 303-315.

Hewawasam, G., Shivaraju, M., Mattingly, M., Venkatesh, S., Martin-Brown, S., Florens, L., . . . Gerton, J.L. (2010) Psh1 is an E3 ubiquitin ligase that targets the centromeric histone variant Cse4. *Mol Cell*, **40**, 444-454.

Jackson, J.P., Johnson, L., Jasencakova, Z., Zhang, X., PerezBurgos, L., Singh, P.B., . . . Jacobsen, S.E. (2004) Dimethylation of histone H3 lysine 9 is a critical mark for DNA methylation and gene silencing in Arabidopsis thaliana. *Chromosoma*, **112**, 308-315.

Jin, W., Melo, J.R., Nagaki, K., Talbert, P.B., Henikoff, S., Dawe, R.K. and Jiang, J. (2004) Maize centromeres: organization and functional adaptation in the genetic background of oat. *Plant Cell*, **16**, 571-581.

Kalitsis, P., Fowler, K.J., Earle, E., Griffiths, B., Howman, E., Newson, A.J. and Choo, K.H. (2003) Partially functional Cenpa-GFP fusion protein causes increased chromosome missegregation and apoptosis during mouse embryogenesis. *Chromosome Res*, **11**, 345-357.

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

Kato, A., Lamb, J.C. and Birchler, J.A. (2004) Chromosome painting using repetitive DNA sequences as probes for somatic chromosome identification in maize. *Proc Natl Acad Sci U S A*, **101**, 13554-13559.

Lermontova, I., Koroleva, O., Rutten, T., Fuchs, J., Schubert, V., Moraes, I., . . . Schubert, I. (2011) Knockdown of CENH3 in Arabidopsis reduces mitotic divisions and causes sterility by disturbed meiotic chromosome segregation. *Plant J*, **68**, 40-50.

Lermontova, I., Schubert, V., Fuchs, J., Klatte, S., Macas, J. and Schubert, I. (2006) Loading of Arabidopsis centromeric histone CENH3 occurs mainly during G2 and requires the presence of the histone fold domain. *Plant Cell*, **18**, 2443-2451.

Li, X. and Dawe, R.K. (2009) Fused sister kinetochores initiate the reductional division in meiosis I. *Nat Cell Biol*, **11**, 1103-1108.

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

Mellone, B.G., Grive, K.J., Shteyn, V., Bowers, S.R., Oderberg, I. and Karpen, G.H. (2011) Assembly of Drosophila centromeric chromatin proteins during mitosis. *PLoS Genet*, **7**, e1002068.

Mizuguchi, G., Xiao, H., Wisniewski, J., Smith, M.M. and Wu, C. (2007) Nonhistone Scm3 and histones CenH3-H4 assemble the core of centromere-specific nucleosomes. *Cell*, **129**, 1153-1164.

Morey, L., Barnes, K., Chen, Y., Fitzgerald-Hayes, M. and Baker, R.E. (2004) The histone fold domain of Cse4 is sufficient for CEN targeting and propagation of active centromeres in budding yeast. *Eukaryot Cell*, **3**, 1533-1543.

Nasuda, S., Hudakova, S., Schubert, I., Houben, A. and Endo, T.R. (2005) Stable barley chromosomes without centromeric repeats. *Proc Natl Acad Sci U S A*, **102**, 9842-9847.

Olszak, A.M., van Essen, D., Pereira, A.J., Diehl, S., Manke, T., Maiato, H., . . . Heun, P. (2011) Heterochromatin boundaries are hotspots for de novo kinetochore formation. *Nat Cell Biol*, **13**, 799-808.

Przewloka, M.R., Venkei, Z., Bolanos-Garcia, V.M., Debski, J., Dadlez, M. and Glover, D.M. (2011) CENP-C is a structural platform for kinetochore assembly. *Curr Biol*, **21**, 399-405.

Ranjitkar, P., Press, M.O., Yi, X., Baker, R., MacCoss, M.J. and Biggins, S. (2010) An E3 ubiquitin ligase prevents ectopic localization of the centromeric histone H3 variant via the centromere targeting domain. *Mol Cell*, **40**, 455-464.

Ravi, M. and Chan, S.W. (2010) Haploid plants produced by centromere-mediated genome elimination. *Nature*, **464**, 615-618.

Ravi, M., Kwong, P.N., Menorca, R.M., Valencia, J.T., Ramahi, J.S., Stewart, J.L., . . . Chan, S.W. (2010) The rapidly evolving centromere-specific histone has stringent functional requirements in Arabidopsis thaliana. *Genetics*, **186**, 461-471.

Ravi, M., Shibata, F., Ramahi, J.S., Nagaki, K., Chen, C., Murata, M. and Chan, S.W. (2011) Meiosis-Specific Loading of the Centromere-Specific Histone CENH3 in Arabidopsis thaliana. *PLoS Genet*, **7**, e1002121.

Sanchez-Pulido, L., Pidoux, A.L., Ponting, C.P. and Allshire, R.C. (2009) Common ancestry of the CENP-A chaperones Scm3 and HJURP. *Cell*, **137**, 1173-1174.

Screpanti, E., De Antoni, A., Alushin, G.M., Petrovic, A., Melis, T., Nogales, E. and Musacchio, A. (2011) Direct binding of Cenp-C to the Mis12 complex joins the inner and outer kinetochore. *Curr Biol*, **21**, 391-398.

Sears, E.R. (1952) Misdivision of univalents in common wheat. *Chromosoma*, **4**, 535-550.

Shi, J. and Dawe, R.K. (2006) Partitioning of the maize epigenome by the number of methyl groups on histone H3 lysines 9 and 27. *Genetics*, **173**, 1571-1583.

Tanaka, Y., Nureki, O., Kurumizaka, H., Fukai, S., Kawaguchi, S., Ikuta, M., . . . Yokoyama, S. (2001) Crystal structure of the CENP-B protein-DNA complex: the DNA-binding domains of CENP-B induce kinks in the CENP-B box DNA. *EMBO J*, **20**, 6612-6618.

Therman, E., Trunca, C., Kuhn, E.M. and Sarto, G.E. (1986) Dicentric chromosomes and the inactivation of the centromere. *Hum Genet*, **72**, 191-195.

Topp, C.N., Okagaki, R.J., Melo, J.R., Kynast, R.G., Phillips, R.L. and Dawe, R.K. (2009) Identification of a maize neocentromere in an oat-maize addition line. *Cytogenet Genome Res*, **124**, 228-238.

Trazzi, S., Perini, G., Bernardoni, R., Zoli, M., Reese, J.C., Musacchio, A. and Della Valle, G. (2009) The C-terminal domain of CENP-C displays multiple and critical functions for mammalian centromere formation. *PLoS One*, **4**, e5832.

Van Hooser, A.A., Ouspenski, II, Gregson, H.C., Starr, D.A., Yen, T.J., Goldberg, M.L., . . . Brinkley, B.R. (2001) Specification of kinetochore-forming chromatin by the histone H3 variant CENP-A. *J Cell Sci*, **114**, 3529-3542.

Vermaak, D., Hayden, H.S. and Henikoff, S. (2002) Centromere targeting element within the histone fold domain of Cid. *Mol Cell Biol*, **22**, 7553-7561.

Voullaire, L.E., Slater, H.R., Petrovic, V. and Choo, K.H. (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**, 1153-1163.

Warburton, P.E., Cooke, C.A., Bourassa, S., Vafa, O., Sullivan, B.A., Stetten, G., . . . Earnshaw, W.C. (1997) Immunolocalization of CENP-A suggests a distinct nucleosome structure at the inner kinetochore plate of active centromeres. *Curr Biol*, **7**, 901-904.

Wieland, G., Orthaus, S., Ohndorf, S., Diekmann, S. and Hemmerich, P. (2004) Functional complementation of human centromere protein A (CENP-A) by Cse4p from Saccharomyces cerevisiae. *Mol Cell Biol*, **24**, 6620-6630.

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

Zhou, Z., Feng, H., Zhou, B.R., Ghirlando, R., Hu, K., Zwolak, A., . . . Bai, Y. (2011) Structural basis for recognition of centromere histone variant CenH3 by the chaperone Scm3. *Nature*, **472**, 234-237.

Table 3.1 Mitotic and meiotic defects of different transgenic lines.

Constructs	Expression confirmation methods	Meiotic defects	Centromere misdivisions (# containing misdivision/# scored)
35S-Gal4-CENH3	RT-PCR, Western blot, Immuno-fluorescence	not expressed during meiosis	strong, 8/16, second generation 31/52
35S-LexA-CENH3	N/A	not expressed during meiosis	not observed, 0/1
cenh3 pro-Gal4-CENH3	RT-PCR	mis-alignment, lagging chromosome, mini-nuclei	mild, 1/16, second generation 1/17
cenh3 pro-LexA-CENH3	N/A	N/A	not observed, 0/2
cenh3 pro-cenpb-CENH3	*RT-PCR*	mis-alignment, lagging chromosome, mini-nuclei	mild, 1/9
35S-LacI-CENPC	RT-PCR	not expressed during meiosis	mild, 1/21
cenpc pro-Gal4-CENPC	*RT-PCR*	mis-alignment, lagging chromosome, mini-nuclei	not observed, 0/8

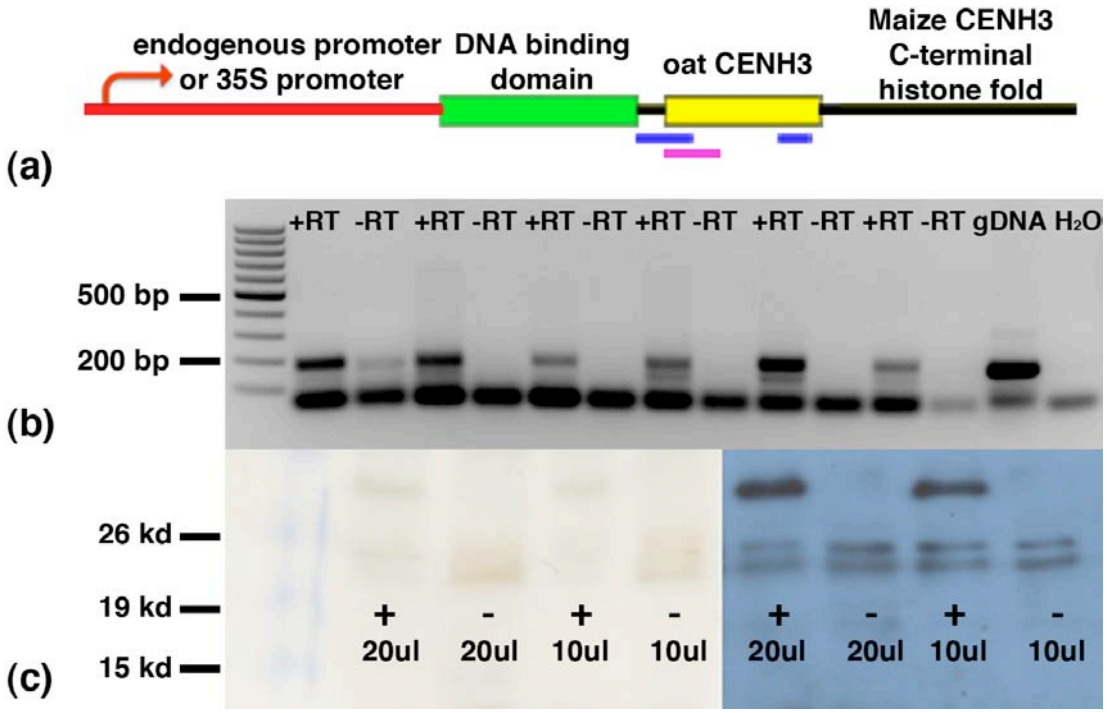


Figure 3.1 Expression of transgenic CENH3 constructs.

(a) Schematic illustration of transgenic CENH3 constructs. Minimal DNA binding domain of Gal4 or LexA or CENPB is shown in green and the segment of oat CENH3 sequence is shown in yellow. The native maize CENH3 sequences are shown in black, consisting of a small N-terminal sequence (11 amino acids) and the histone fold domain. Sequences recognized by the oat peptide antibody are underlined in magenta. Sequences recognized by other CENH3 antibodies (which do not recognize these fusion proteins) are indicated in blue.

(b) RT-PCR results showing that the Gal4-CENH3 gene is expressed as expected. Reverse transcriptase positive (+RT) and negative (-RT) are indicated (167 bp).

(c) Western blots showing that the 35S-Gal4-CENH3 protein (expected to be 27.44 KD) is expressed in transgenic plants (+) but not wild type plants (-).

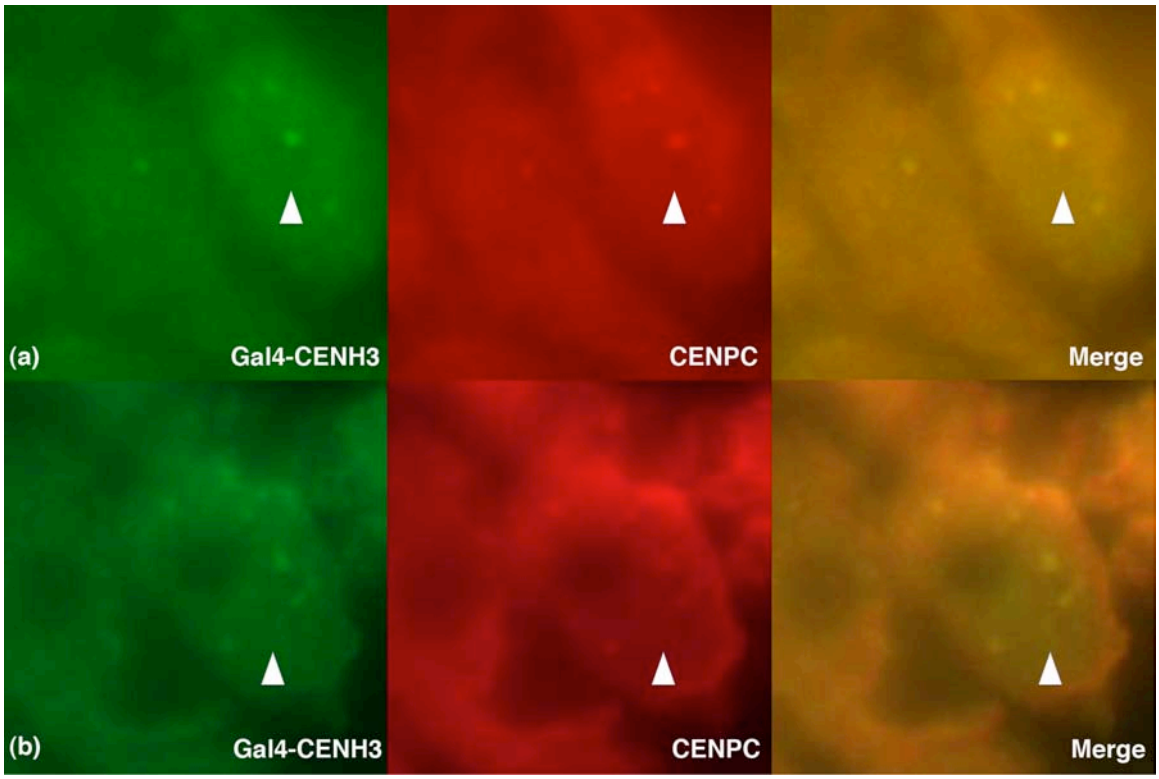


Figure 3.2 Immunofluorescence images from a Gal4-CENH3 overexpression line showing that transgenic Gal4-CENH3 co-localized with endogenous CENPC proteins (indicated by white arrowheads). Images from two different cells (a and b) were stained with oat peptide antibodies (green) to represent transgenic Gal4-CENH3 proteins and maize CENPC antibodies (red) to indicate endogenous kinetochores.

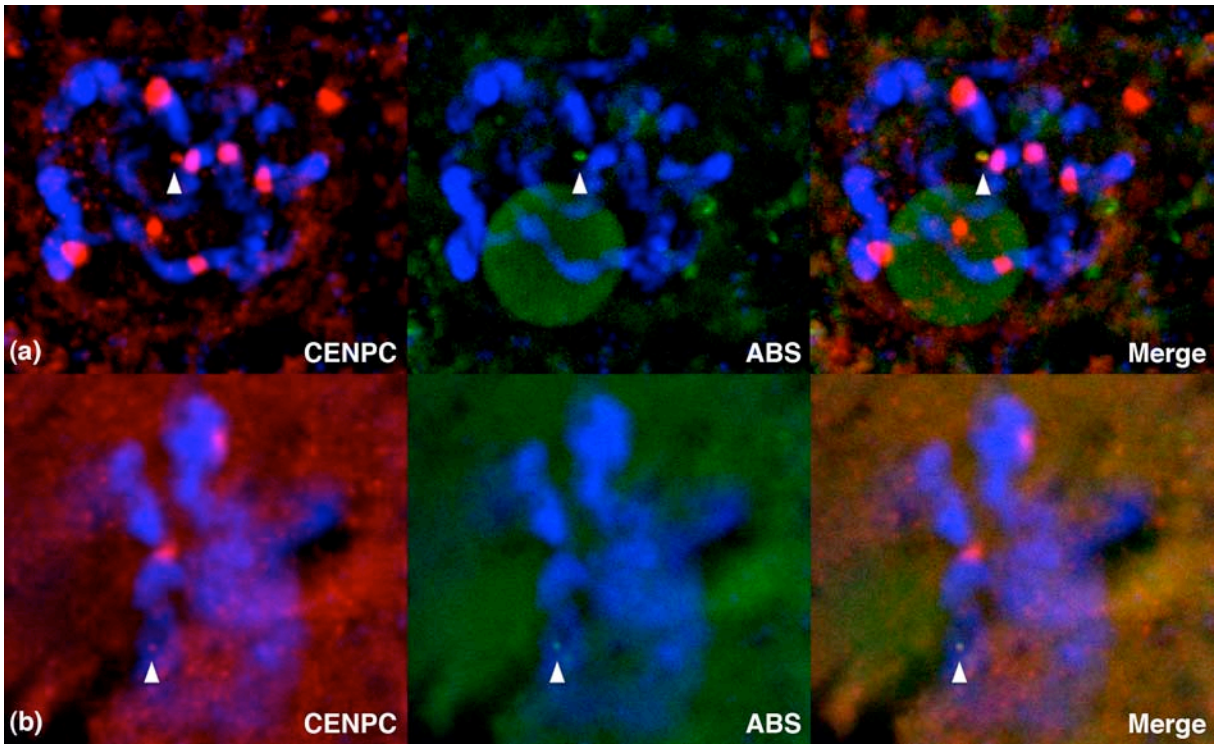


Figure 3.3 Transient localization of CENPC proteins on chromosome arms.

Immuno-FISH images from a CENPB-CENH3 line showing that CENPC can be recruited to the ABS4 array (indicated by white arrowheads). Since the ectopically expressed proteins were CENPB-CENH3, we presume CENPC was recruited to ABS4 indirectly, by interacting with tethered CENH3. Maize pachytene (a) and diplotene (b) chromosomes were stained with CENPC antibodies (red) and hybridized with ABS-specific probes (green).

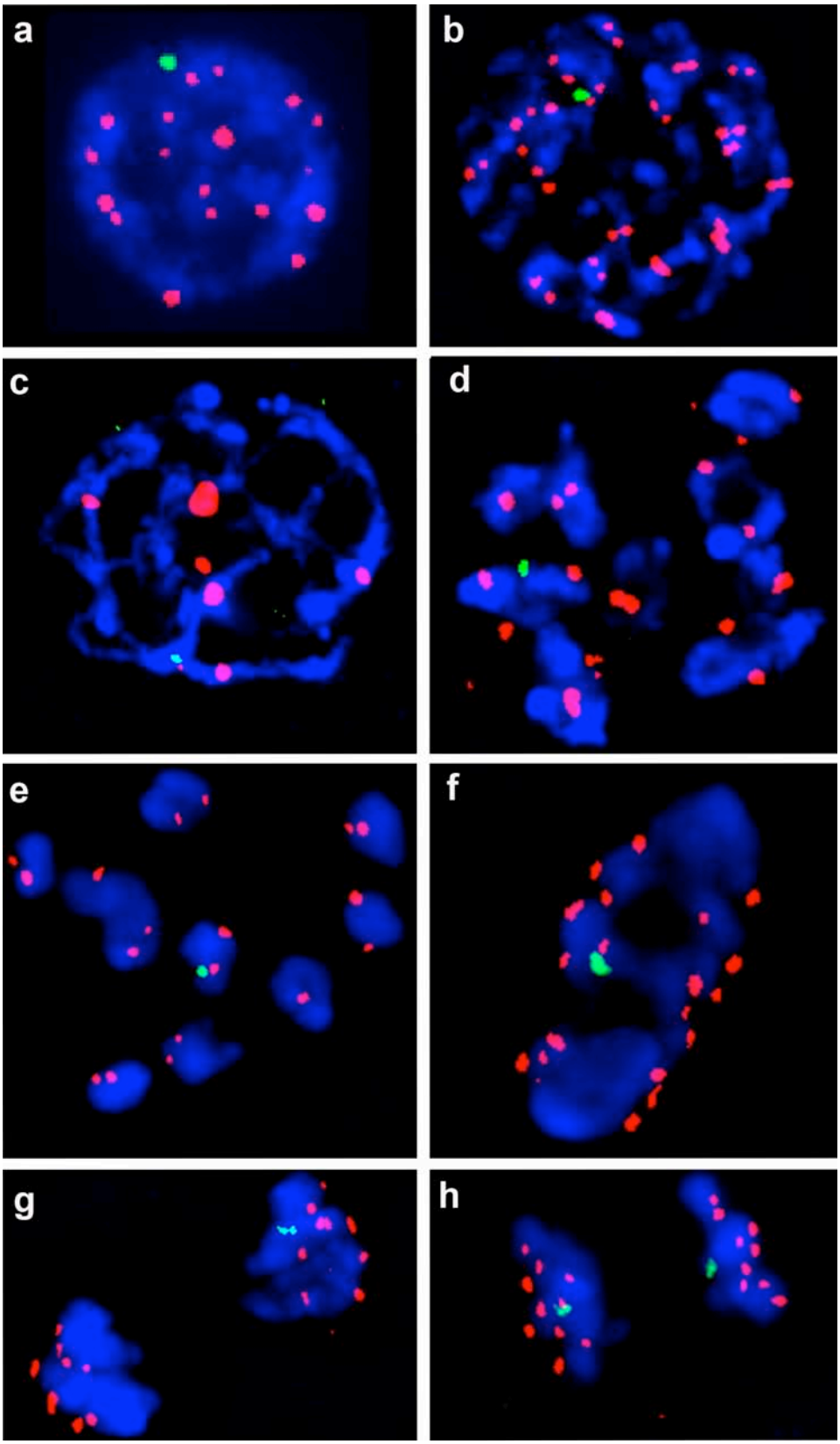

a
b
c
d
e
f
g
h

Figure 3.4 Immuno-FISH images from Gal4-CENH3 lines showing that CENPC proteins were normally targeted to the endogenous centromeres but not the ABS4 array. Cells were stained with CENPC antibodies (red) and DAPI (blue) and hybridized with ABS-specific probes (green). Transgenic CENH3 proteins were not localized at the ABS loci in mitotic interphase (a), mitosis (b) and all stages of meiosis, including pachytene (c), diakinesis (d), prometaphase I (e), metaphase I (f), anaphase I (g) and anaphase II (h).

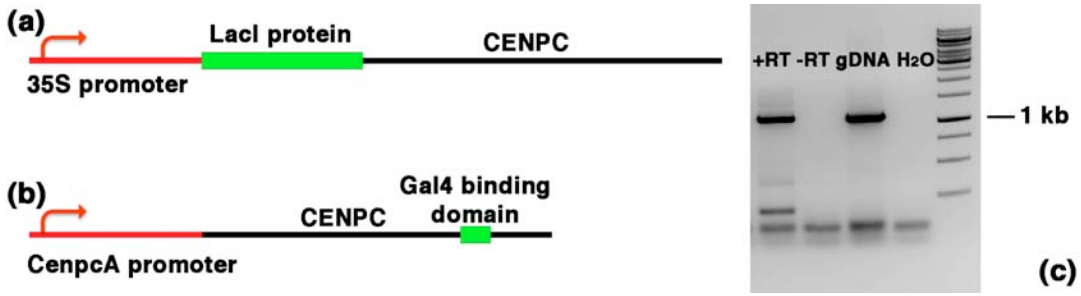


Figure 3.5 Design and expression of transgenic CENPC proteins.

(a) Schematic illustration of the LacI-CENPC construct. The 35S promoter is shown in red; LacI protein is shown in green and CENPC sequence is shown in black.

(b) Schematic illustration of transgenic Gal4-CENPC construct. Maize endogenous CenpcA promoter is shown in red; Gal4 binding domain is shown in green and CENPC sequence is shown in black. In this construct the native DNA binding domain is replaced with the Gal4 DNA binding module.

(c) RT-PCR results showing transgenic 35S-LacI-CENPC is expressed. The expected band is seen when reverse transcriptase is added (+RT) but not in the negative control (-RT).

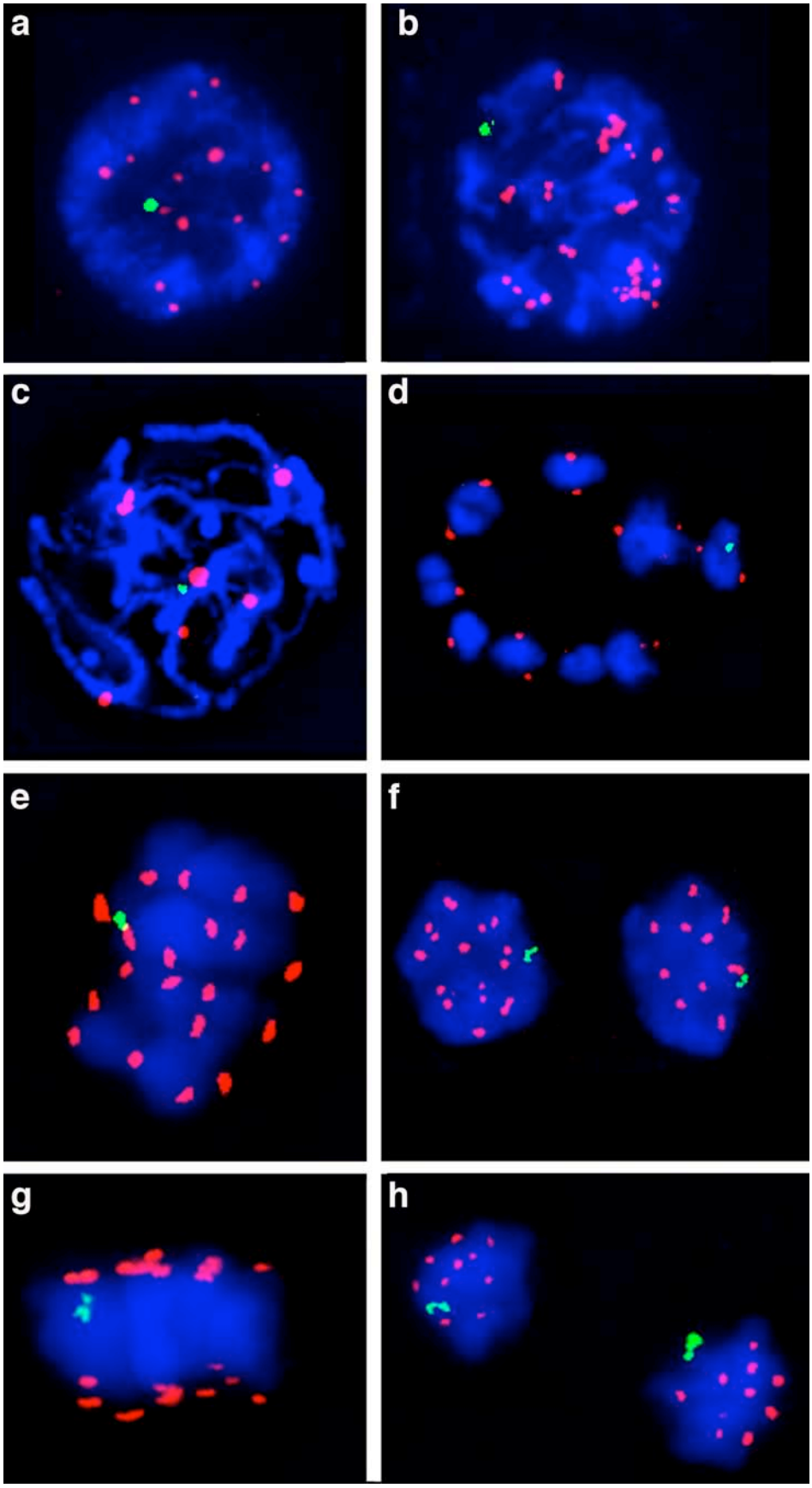

a
b
c
d
e
f
g
h

Figure 3.6 ImmunoFISH images of Gal4-CENPC lines showing that CENPC proteins were targeted to endogenous centromeres but not ABS arrays on the chromosome arms. Cells were stained with CENPC antibodies (red) and DAPI (blue) and hybridized with ABS-specific probes (green). Transgenic CENPC proteins were not localized at the ABS loci in interphase (a), mitosis (b) and all stages of meiosis, including pachytene (c), diakinesis (d), metaphase I (e), prophase II (f), metaphase II (g) and anaphase II (h).

Figure 3.7 Morphological defects in maize plants overexpressing transgenic Gal4-CENH3. Cracks and gaps in the leaves are indicated by white arrowheads.

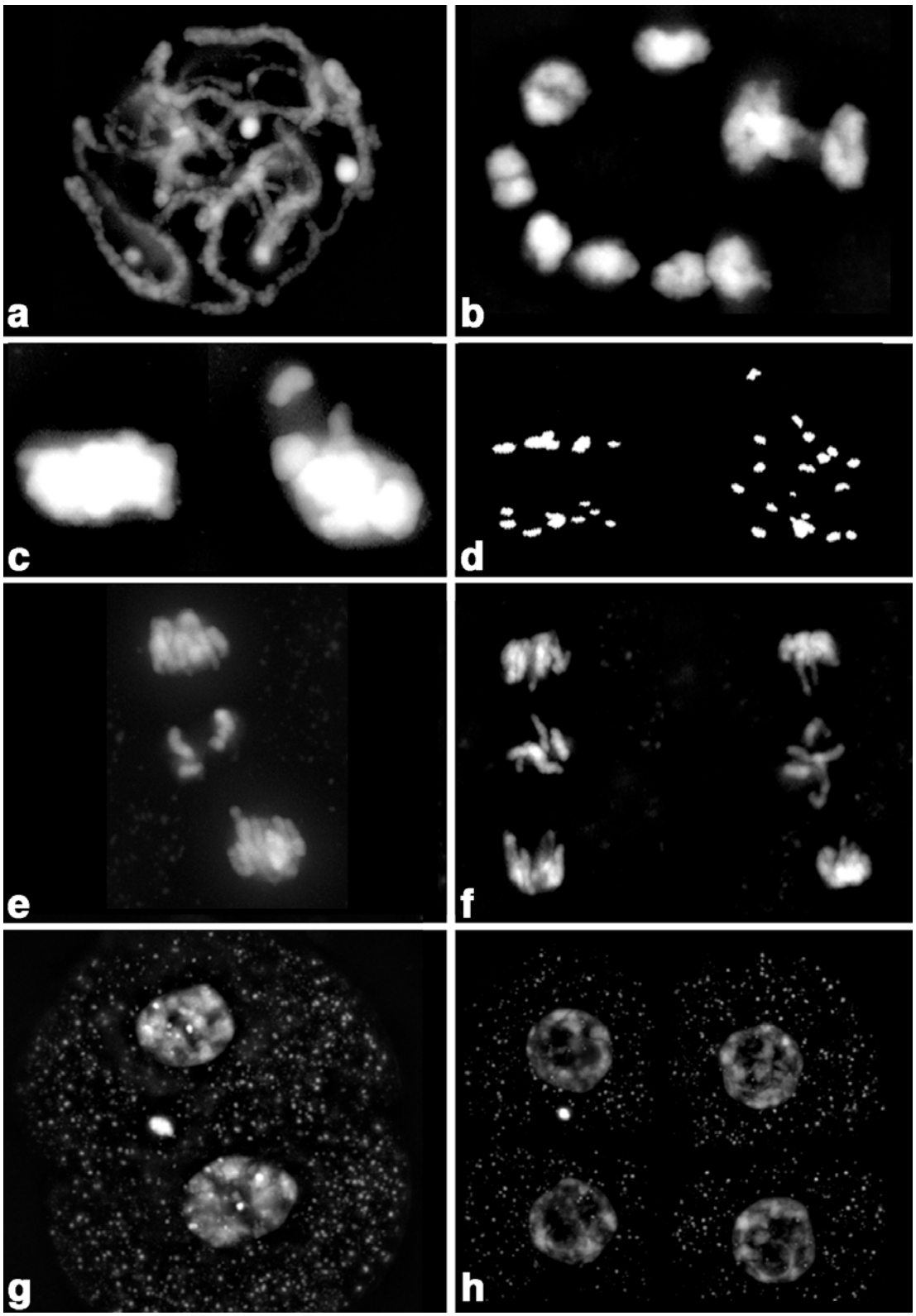

a
b
c
d
e
f
g
h

Figure 3.8 Meiotic defects in plants expressing the Gal4-CENPC construct. No defects were observed in pachytene (a) or diakinesis (b). (c) Evidence of misalignment at metaphase I as viewed with DAPI (c) and CENH3 staining (d)**.** Lagging chromosomes were observed in anaphase I (e) and anaphase II (f). Minichromosomes are one common result chromosome segregation failures, and these were observed at telophase I (g) and the tetrad stage (h).

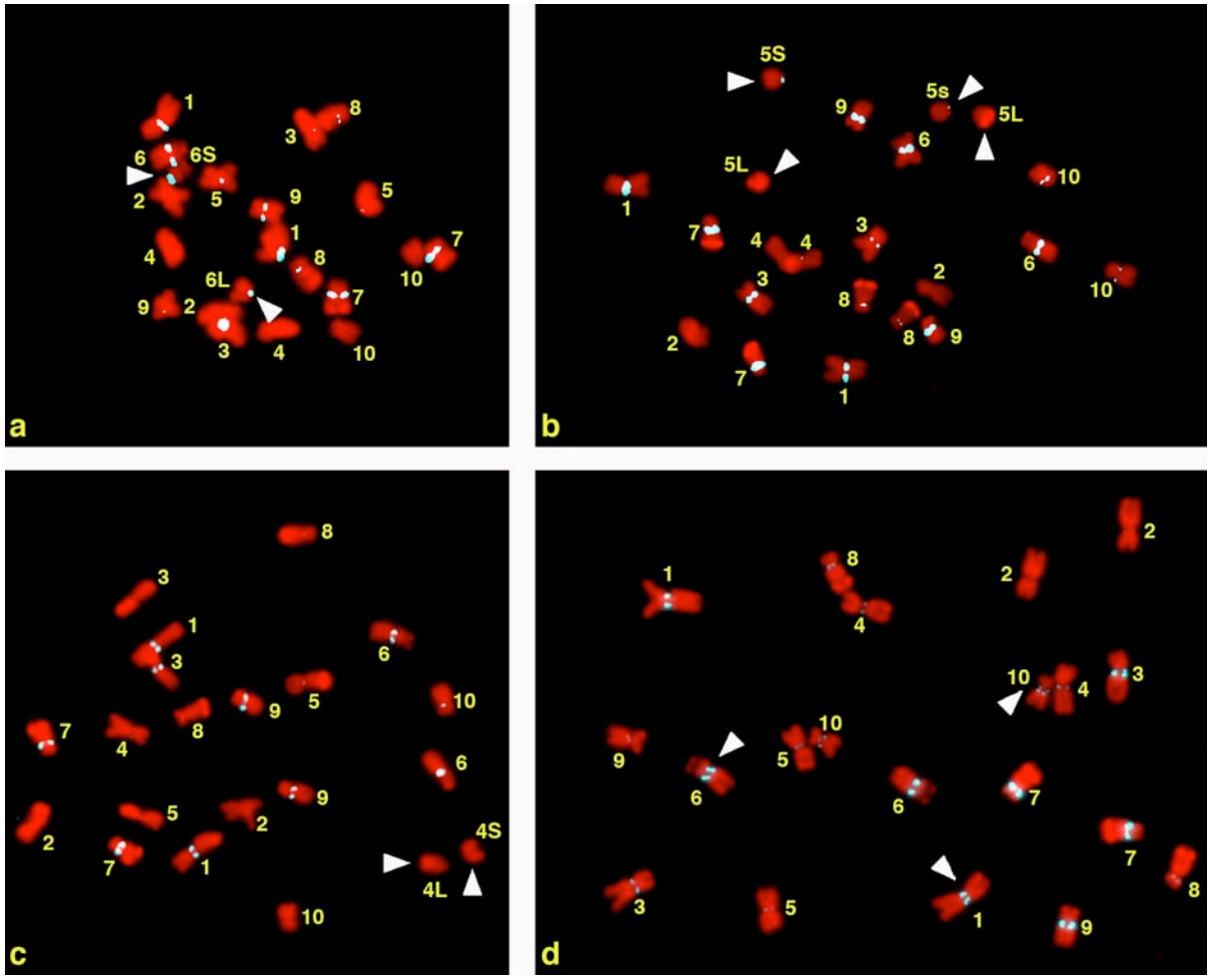


Figure 3.9 Centromere misdivisions in Gal4-CENH3 overexpression lines. Chromosomes were hybridized with a maize centromeric DNA probe CentC (white) and stained with DAPI (red). Chromosomes are identified by their numbers. Broken chromosomes with CentC staining (a and b) or without staining (b and c) are indicated with white arrowheads. Examples of partially separated centromeres were also observed (d).

Supplemental data

1. Sequence of Gal4-CENH3-oat epitope-CENH3

ATGAAGCTGCTCTCCTCTATCGAGCAGGCCTGCGACATCTGCAGGCTGAAGAAGCTGAAGTGCTCCAAGGAGAAGCCGAAGTGC
GCCAAGTGCCTGAAGAACAACTGGGAGTGCAGGTACTCCCCGAAGACCAAGAGGTCCCCGCTGACCAGGGCCCACCTGACCGAG
GTGGAGTCCAGGCTGGAGAGGCTGGAGCAGCTGTTCCTGCTCATCTTCCCGAGGATGGCTCGAACCAAGCACCAGGCCGTGAGGA
AGTCGAAGCCGACGCCCAAGAAGCAGCTCAAGTTCGGCCGCTCCCCGGGCCAGACGGCGGAGCAGGAGACAGGCGGCCCGAGC
ACGTCGGCGGCACCAAGGCGAGGCGCGCGGAGGCCAGCAGCAACGCCGGCTCCAGGGGCACCTGCGCAACAGAGGGCGAGGAA
GCCGCATCGGTTCAAGCCAGGGACTGTAGCGCTGCGGGAGATCAGGAAGTACCAGAAGTCCACTGAACCGCTCATCCCCTTTGCG
CCTTTCGTCCGTGTGGTGAGGGAGTTAACCAATTTCGTAACAAACGGGAAAGTAGAGCGCTATACCGCAGAAGCCCTCCTTGCGC
TGCAAGAGGCAGCAGAATTCCACTTGATAGAACTGTTTGAAATGGCGAATCTGTGTGCCATCCATGCCAAGCGTGTCACAATCAT
GCAAAAGGACATACAACTTGCAAGGCGTATCGGAGGAAGGCGTTGGGCATGA

2. Sequence of LexA-CENH3-oat epitope-CENH3

ATGAAGGCCCTGACCGCCAGGCAGCAAGAGGTGTTCGACCTGATCAGGGACCACATCTCCCAGACCGGCATGCCGCCGACCAGGGCCG
AGATCGCCCAGAGGCTGGGCTTCAGGTCCCCGAACGCCGCGGAGGAGCACCTGAAGGCCCTGGCCAGGAAGGGCGTGATCGAGATCGT
GTCCGGCGCCTCCAGGGGCATCAGGCTGCTGCAGGAGGAGATGGCTCGAACCAAGCACCAGGCCGTGAGGAAGTCGAAGCCGACGCCC
AAGAAGCAGCTCAAGTTCGGCCGCTCCCCGGGCCAGACGGCGGAGCAGGAGACAGGCGGCCCGAGCACGTCGGCGGCACCAAGGCGA
GGCGCGCGGAGGCCAGCAGCAACGCCGGCTCCAGGGGCACCTGCGCAACAGAGGGCGAGGAAGCCGCATCGGTTCAAGCCAGGGACT
GTAGCGCTGCGGGAGATCAGGAAGTACCAGAAGTCCACTGAACCGCTCATCCCCTTTGCGCCTTTCGTCCGTGTGGTGAGGGAGTTAAC
CAATTTCGTAACAAACGGGAAAGTAGAGCGCTATACCGCAGAAGCCCTCCTTGCGCTGCAAGAGGCAGCAGAATTCCACTTGATAGAAC
TGTTTGAAATGGCGAATCTGTGTGCCATCCATGCCAAGCGTGTCACAATCATGCAAAAGGACATACAACTTGCAAGGCGTATCGGAGGA
AGGCGTTGGGCATGA

3. Sequence of 1kb cenh3 promoter-CENPB(2-125)-oat epitope-CENH3

TACAATTGTACCTTTCACATGTCTTCTAGATGGTTCCCAACCCTTTGGCCAAGATCGTACAGATAATATTGCGAGGAGCCCAAATC
AACGGTGTCCATATGTTATGTTGATGTGGATGGTTTACCTAGGCGCAAAAGTGCGCTGGTTTCGTCCGTACAAATATACTTTAAGT
ATGGTTTTGATTTTTTTCTATTTTTCATTTTTTAAATAAAACGAGACAATCAAATCTGATATAAAAATCAAATGAATTATAAATAGA
GACGGAAAGAGTATATATATTTGTTTTGCTATTATTTAAAGTATTAAAAGATAGTGGACGAATGAACGTCCTCTATGTTTAAAAGA
ACGTTTTAGAGGACGTTGTGTTGTTGAAGGAAATATGAAAAAAAAAATCTTCTGCATATTTAGAAGGGAGGAGCGTTTACACATTA
CTTTCGGGACTTCAACCCAAATATGTCAAGGTTTGTGAGTGGCTCAGTGCGGAAAAAAAATCCTATATATACCAGATGTAAACAC
TATCTTTTACAGCCTATCACATTCACATTTAGAGGTTCACAAAGATAGATCAAAATTTATAAAATAATCATTTAATATTTTTTTTAT
TTTATTTATATGGATAAGCAGCTGGTGTATGTGAGGAGCTGTAAAAGATATTTTTTACATCCGAGATGTAAAGATTTTTTTTAACTC
AATGCTGGTTACCGGfCTGGGAGGACGATGATAAAGAAAGCATCTCTCACTGCATTCCGGGCCCACTACTCAAACGTTCGGCACG
CCAGGTTGGCAGGTAGCCGTTACATCGATAGGCACTCGGCCACTCGCACGCAGACACCACACCAGTGTGCTCAGTGCTCACTGCT
CACCATAATAACGCTGCACCTCTTTTCATTTCACCATCTCCTGCCCCCTTAAAAAAAAGACTCACCGTCGACACGCCCTCCCGTCC
CGAGAGTTCTGAATCGAAACCGTCGGCCACGAGAGCAGTGCGAGGCGCCCACCGCGATGGGGCCGAAACGGAGACAGCTCACTT
TTAGGGAGAAGTCGCGGATTATCCAAGAAGTGGAGGAGAACCCAGACTTGAGGAAAGGGGAAATCGCGCGCAGATTCAATATCC
CTCCGAGTACACTGTCCACGATCTTGAAGAACAAGCGCGCCATTCTCGCGAGCGAGCGCAAGTACGGTGTCGCATCTACCTGCCG
CAAGACCAATAAGCTCTCCCCATATGATAAGCTGGAGGGCCTTCTTATAGCATGGTTTCAGCAAATCCGTGCCGCTGGTCTGCCCG
TGAAGGGAATCATTCTCAAAGAAAAGGCTCTGAGGATAGCGGAGGAGCTTGGCATGGACGACTTCACGGCTTCAAACGGCTGGC
TGGATAGGACGCCCAAGAAGCAGCTCAAGTTCGGCCGCTCCCCGGGCCAGACGGCGGAGCAGGAGCGCTGGCGGCCAGGGACTG
TAGCGCTGCGGGAGATCAGGAAGTACCAGAAGTCCACTGAACCGCTCATCCCCTTTGCGCCTTTCGTCCGTGTGGTGAGGGAGTT
AACCAATTTCGTAACAAACGGGAAAGTAGAGCGCTATACCGCAGAAGCCCTCCTTGCGCTGCAAGAGGCAGCAGAATTCCACTTG
ATAGAACTGTTTGAAATGGCGAATCTGTGTGCCATCCATGCCAAGCGTGTCACAATCATGCAAAAGGACATACAACTTGCAAGGC
GTATCGGAGGAAGGCGTTGGGCATGA

4. Sequence of 1kb CenpcA promoter-Gal4-CENPC

TAGTTAATATTGTTCTTTGTTGAGTGTGGAATATTAAGTTGCAAACTCTATTCTTTTTTTCTGCATAGATCTCCTTTTTTTCTTTTGCT
ACATGTTTTTATATACGTTTGTAGACATTAGACATGATATTATCGAACTAGCTCGCGAGCTAAACGAGCCAGCTCGAGTTGTCAAA
CGAGTCGAGCCGAGCCAGCTCGTTATCTTAACGAGCCAGTTTGAGTCGAGTCGAGCCGAGCCAGCTCGATATCCACCCCTAGTGG
AGTCTCACGCTATGGGAGAAATTTGTGATTTTGCCACTCTTAATTTTAGTTGTCTGTTATTATGACATCCCGTGTGTATGACTGATA

AGACCATACGTGTCTATGATTTGTGGACTTGATGACAATATAACGATGCTCATAAATAATAATGATAAAATTGACAATGGCTACA
GTGGTAGGGGAGGGTGATGCTCGCGGGGAGCGACGGCCTTAGGGTGTGCGCAGCGGTAAACACTAGCCATCCCCCTTCCCTTGCA
TGCCCGCTGTAGGGGGCACCCCACGGTCGCAGCGGTGCCCTCTACGCGTCGGTCCCCTTCTCTCCCCCCAAGATCTAGCGTGTCAC
TGTTCATCACCCTAAACACTGTGCTGTGGGTCCACGAGTAATTGTTAGTAGATAAAAAATTAATAATTGATATAGAAAATGATATT
TTATAGTTGTAGTGGGGTATATGGGAGTATTTAGGGGGAACCGCTGCGGGAGATGAAAAAATATGGAAAAACATAGGGGAGAAA
AGTGATATAGGGGGAAAAATTTAGGGGTAACGGTTGCGGATAGCCTTACCCTTCCCCATCCTCACTCTCCCGTCATCCAGCCTAGT
AGCCTACCGTTTCAATTTTTGAACTCCCGGTCCGGCAACCCTCCCCAGCCAGCGTTCCACTCACTCCGGCATCGCTTCGAATCTCC
GCGCGGCGGCGGGACCGCGTGACCGACGCGCGGCGCTGCCCCTTCAGAAAGCCGCATGTACCCCTACGACGTTCCCGATTATGCA
TACCCCTACGACGTTCCCGATTATGCATACCCCTACGACGTTCCCGATTATGCAGACGCCGCCGACCCCCTCTGCGCCATCTCCTC
GACTGCGCGACTCCTTCCGCGGACCCTTGGCCCCGCCATTGGGCCTTCCCCCTCGAACCCCCGCGACGCGCTCCTCGAGGCCATCG
CACTCGCCCGGTCTCTGAAAGGGTCTGAGGAGCTGGTTAAGCAGGCTACAATGGTGCCGAAGGAGCACGGGGACATCCAGGCTC
TCTACCATGATGATGGAGTGAAAGGCTGGCCTCCTGCAAATGGCAGCAAAGAGCAACAGGGAAGAAGGCCAGCACTGGATCGTA
AACGGGCTCGGTTTGCTATGAAGGACACTGGAAGCAAACCGGTGCCTGTTGTGGACCAATCTAAGTTATCAAATATTTCAGATCC
AATCACATTCTTCATGACCTTGGATCGGCTTGAAGAAGCCGAGGAGGAGATCAAACGGCTAAATGGAGAGGCAGAAAAACGCAC
ATTGAATTTTGATCCTGTAGATGAACCCATTAGACAGCCTGGTTTACGTGGGAGAAAATCAGTCCGCAGTTTCAAGGTAATTGAG
GATGTTGGTACTCAAGATCCCAATGAGGCACCGGCTTCCCAAACAGCAACTATGACAGGATCTCAGTTGTCACAGGATGTTATGC
ATGCTGTTGCGGGCAAAAATGGACGATCTGTTTCCTCAAGATCTAGTGAAGCTATTTCAGAGAAAGAAGTTTCATTAGCTGAGAA
GGATGGCCGTGACGATTTGACTTATATACTGACCTCAATCCAAGATTTAGATGAATCTGAAGAGGAAGAGTTTATCCGGAAGACC
TTAGGGATTAAAGAAACAAGGAAAGAAAGAGTTAGTCTTCGTAACTCCATTCTTGGAGTTAGGTCCCTAAGAAGCAATACTGAAC
AAAAAGGTTCAATGAAGGTCCACCCTCCAGAAAGTCATTTGCCTCAGACTCGCCAAGATCGAATTTCAGAGTTAGAGAAACATTT
GTTTCCTGGAGGTGCAGCAAATGCTAAATGCACAGATGATGAATCTGAAGGGTCACCAGATATTGTGATGGGTGAACCATCATTG
GTGCATTATTCTTCTGATGTTCTGATGACTGATGAGAACTTTACTGCAAGTGACGTTGACAGGGAGACCCCAAATCTGGGTGCCAG
AGCAGCAGACCCTGTTCTTGATCCTGAACCAAACTTGCCTGATCATGCATATGAAAGGCAACCTGGAGGTTCCTCACTTGGTTTGT
ACAGAGACACAGAAGTTGGAGTTGCCAAAGAAAATGAGACATGTAACAGGAGCAATATATCTGTGGAGGAAGATGATGTGCCAA
TAGACTATATAACTATTGGCAGGTCTACTAATGAAACAGAAGTTTCTTCTTCTCATCCTTTGGAGGGCAGCTCGACAGAAGAATTG
GTCACTAAACCAGGTAGACATGCGGCTCCAGATGGTATCTGTAGGACTTCACATGCAGCTGAGGATAGCATTCAACATCTAGAAG
CGGTCAAAGAGGGTGGTGTCCTACAAAAACTTCTGAGTTCCATCGAGCAAGCGTGTGATATATGTCGCCTGAAGAAGCTCAAGTG
CAGCAAAGAGAAGCCAAAGTGCGCGAAGTGCCTCAAAAACAATTGGGAGTGCAGATACTCCCCTAAGACAAAGCGCTCCCCTCT
CACGAGGGCGCACTTGACGGAAGTCGAAAGTAGATTGGAACGCCTGGAGCAACTGTTTCTTCTCATCTTCCCCCGCGAGGATCTC
GACATGATTCTGAAAATGGACTCCCTTCAGGACATTAAGGCCCTGCTCACTGATGCTGACCTTGCATGCCAGCCTGGAGTGAGAA
AAAGCTCAAGAACACGCTCAAGGCCTTTGGAGTATTGGCTTGGTGAAAGATTACTATATGGGCCTATACATGACAATTTGCATGG
AGCTATTGGCATCAAAGCATACTCTCCTGGCCAAGATGGCAAGAGATCATTGAAAGTGAAATCTTTCGTGCCTGAACAGTATTCA
GATCTTGTTGCTAAATCTGCAAGGTACTGA

5. Sequence of LacI-CENPC

ATGAAGAAGAAGAGAAAGGTGAAACCAGTAACGTTATACGATGTCGCAGAGTATGCCGGTGTCTCTTATCAGACCGTTTCCCGCG
TGGTGAACCAGGCCAGCCACGTTTCTGCGAAAACGCGGGAAAAAGTGGAAGCGGCGATGGCGGAGCTGAATTACATTCCCAACC
GCGTGGCACAACAACTGGCGGGCAAACAGTCGTTGCTGATTGGCGTTGCCACCTCCAGTCTGGCCCTGCACGCGCCGTCGCAAAT
TGTCGCGGCGATTAAATCTCGCGCCGATCAACTGGGTGCCAGCGTGGTGGTGTCGATGGTAGAACGAAGCGGCGTCGAAGCCTGT
AAAGCGGCGGTGCACAATCTTCTCGCGCAACGCGTCAGTGGGCTGATCATTAACTATCCGCTGGATGACCAGGATGCCATTGCTG
TGGAAGCTGCCTGCACTAATGTTCCGGCGTTATTTCTTGATGTCTCTGACCAGACACCCATCAACAGTATTATTTTCTCCCATGAAG
ACGGTACGCGACTGGGCGTGGAGCATCTGGTCGCATTGGGTCACCAGCAAATCGCGCTGTTAGCGGGCCCATTAAGTTCTGTCTC
GGCGCGTCTGCGTCTGGCTGGCTGGCATAAATATCTCACTCGCAATCAAATTCAGCCGATAGCGGAACGGGAAGGCGACTGGAGT
GCCATGTCCGGTTTTCAACAAACCATGCAAATGCTGAATGAGGGCATCGTTCCCACTGCGATGCTGGTTGCCAACGATCAGATGG
CGCTGGGCGCAATGCGCGCCATTACCGAGTCCGGGCTGCGCGTTGGTGCGGATATCTCGGTAGTGGGATACGACGATACCGAAGA
CAGCTCATGTTATATCCCGCCGTCAACCACCATCAAACAGGATTTTCGCCTGCTGGGGCAAACCAGCGTGGACCGCTTGCTGCAA
CTCTCTCAGGGCCAGGCGGTGAAGGGCAATCAGCTGTTGCCCGTCTCACTGGTGAAAAGAAAAACCACCCTGGCGCCCAATACGC
AAACCGCCTCTCCCCGCGCGTTGGCCGATTCATTAATGCAGCTGGCACGACAGGTTTCCCGACTGGAAAGCGGGCAGAAGGGCGA
ATTCGCGGCCGCCCCCTTCACCATGGACGCCGCCGACCCCCTCTGCGCCATCTCCTCGACTGCGCGACTCCTTCCGCGGACCCTTG
GCCCCGCCATTGGGCCTTCCCCCTCGAACCCCCGCGACGCGCTCCTCGAGGCCATCGCACTCGCCCGGTCTCTGAAAGGGTCTGAG
GAGCTGGTTAAGCAGGCTACAATGGTGCCGAAGGAGCACGGGGACATCCAGGCTCTCTACCATGATGATGGAGTGAAAGGCTGG
CCTCCTGCAAATGGCAGCAAAGAGCAACAGGGAAGAAGGCCAGCACTGGATCGTAAACGGGCTCGGTTTGCTATGAAGGACACT
GGAAGCAAACCGGTGCCTGTTGTGGACCAATCTAAGTTATCAAATATTTCAGATCCAATCACATTCTTCATGACCTTGGATCGGCT
TGAAGAAGCCGAGGAGGAGATCAAACGGCTAAATGGAGAGGCAGAAAAACGCACATTGAATTTTGATCCTGTAGATGAACCCAT
TAGACAGCCTGGTTTACGTGGGAGAAAATCAGTCCGCAGTTTCAAGGTAATTGAGGATGTTGGTACTCAAGATCCCAATGAACCG
GCTTCCCAAACAGCAACTATGACAGGATCTCAGTTGTCACAGGATGTTATGCATGCTGTTGCGGGCAAAAATGGACGATCTGTTTC
CTCAAGATCTAGTGAAGCTATTTCAGAGAAAGAAGTTTCATTAGCTGAGAAGGATGGCCGTGACGATTTGACTTATATACTGACC
TCAATCCAAGATTTAGATGAATCTGAAGAGGAAGAGTTTATCCGGAAGACCTTAGGGATTAAAGAAACAAGGAAAGAAAGAGTT
AGTCTTCGTAACTCCATTCTTGGAGTTAGGTCCCTAAGAAGCAATACTGAACAAAAAGGTTCAATGAAGGTCCACCCTCCAGAAA
GTCATTTGCCTCAGACTCGCCAAGATCGAATTTCAGAGTTAGAGAAACATTTGTTTCCTGGAGGTGCAGCAAATGCTAAATGCAC
AGATGATGAATCTGAAGGGTCACCAGATATTGTGATGGGTGAACCATCATTGGTGCATTATTCTTCTGATGTTCTGATGACTGATG
AGAACTTTACTGCAAGTGACGTTGACAGGGAGACCCCAAATCTGGGTGCCAGAGCAGCAGACCCTGTTCTTGATCCTGAACCAAA
CTTGCCTGATCATGCATATGAAAGGCAACCTGGAGGTTCCTCACTTGGTTTGTACAGAGACACAGAAGTTGGAGTTGCCAAAGAA
AATGAGACATGTAACAGGAGCAATATATCTGTGGAGGAAGATGATGTGCCAATAGACTATATAACTATTGGCAGGTCTACTAATG

AAACAGAAGTTTCTTCTTCTCATCCTTTGGAGGGCAGCTCGACAGAAGAATTGGTCACTAAACCAGGTAGACATGCGGCTCCAGA
TGGTATCTGTAGGACTTCACATGCAGCTGAGGATAGCATTCAACATCTAGAAGCGGTCAAAGAGGGTGGTGTCCTACAAGATAAG
TCAAGCCAGTTGTTGGAGATGCCTCAGGAAGATATTAATCCACTGAATCAGGCTCAGATGCATGGTGGAAGCACTAAGAAATCGG
CACCTGATCTATCCCCGACCAAACAGAAGAAGCAACAGGCAGTTCAAGAAAGGAAAAGGAAGCAACAGTCAAAGAGGGGCAAG
AAAGTGGCTGGTGAATCCCTGGAAATACCCCAAACAAATTTTGAATCGGAGAATCAACCTCATAATGACGATGTCAACATTGAGA
AACAGACAATTACGAGTAGCACACTCTCACCAAATGGTGCCAAGGGGCAAAAGGGAGCTCAAAGGAGAAACCAGACCAAAAAA
AGGAATCAAAGAAAAATCCTTGGAGATGCTGACCTTGCATGCCAGCCTGGAGTGAGAAAAAGCTCAAGAACACGCTCAAGGCCT
TTGGAGTATTGGCTTGGTGAAAGATTACTATATGGGCCTATACATGACAATTTGCATGGAGCTATTGGCATCAAAGCATACTCTCC
TGGCCAAGATGGCAAGAGATCATTGAAAGTGAAATCTTTCGTGCCTGAACAGTATTCAGATCTTGTTGCTAAATCTGCAAGGTACT
GA

CHAPTER 4

EPIGENETIC REGULATION OF CENTROMERE SIZE[3]

[3] Han Zhang and R. Kelly Dawe, Manuscript to be submitted to *Chromosoma*.

Abstract

It has been known for decades that centromere size varies across species, but the factors involved in regulating this process are unknown. As a means to address this question we measured centromere size in ten species of the grass family including rice, maize, and wheat, which diverged 60~80 million years ago and vary by 40-fold in genome size. These measurements revealed that centromere size does not correlate well with genome size, or with chromosome number, but can be correctly expressed as a function of genome size divided by chromosome number. The data suggest that simple genome characteristics can be used to predict average centromere size. We then studied the outcome when two species with different sized centromeres are crossed together. We measured centromere sizes on the maize chromosomes in three oat-maize additional lines and found that centromere sizes in the hybrid differed from the original sizes inherited from the progenitors. Our data suggest that centromere size is a regulated feature that is at least partially responsive to cell size.

Introduction

Centromeres are the chromosomal domains responsible for accurate chromosome segregation during mitosis and meiosis. In higher eukaryotes, centromeres are characterized by long segments of tandem repeats. Although these arrays extend several megabases in plants and animals, only a fraction of the repetitive sequences are incorporated into centromeric chromatin, which is specified by the specialized histone H3 variant CENH3 (Blower *et al.* 2002, Zhong *et al.* 2002). It has been estimated that CENH3-containing nucleosomes occupy 2,500 kb of satellite repeats on human centromeres, compared to ~750 kb on rice centromeres and ~10 kb in fission yeast *S. pombe* (Partridge *et al.* 2000, Nagaki *et al.* 2004, Black *et al.* 2007, Yan *et al.* 2008). These observations suggest that each species is typified by an average centromere size.

CENH3 and its distinct chromatin environment are directly responsible for recruiting the overlying kinetochore proteins that ultimately interact with microtubules (Cheeseman and Desai 2008). Similar to what have been observed for kinetochores, each species has a characteristic number of kinetochore microtubules (Peterson and Ris 1976, Lin *et al.* 1981, Rieder 1982, Ding *et al.* 1993). In the budding yeast *S. cerevisiae*, which has the smallest known centromere at 125 bp, a single CENH3 nucleosome is linked to a single spindle microtubule through a protein linkage consisting of 1-2 copies of CENPC, 6-7 copies of the MIS12 complex and 8 copies of the NDC80 complex (Joglekar *et al.* 2006). Fission yeast has a similar molecular architecture, building upon 2-3 CENH3 nucleosomes and interacting with 2-4 microtubules (Joglekar *et al.* 2008). In higher eukaryotes, it has been reported that there is a positive correlation between kinetochore size (as determined by electron microscopy) and number of kinetochore

microtubules, although small kinetochores tend to interact with spindle microtubules at high density (McEwen *et al.* 1998).

Among chromosomes within a species, the size of the CENH3-containing domain appears to be relatively constant. The size of the CENH3 domain has been measured by means of indirect immuno-fluorescence in human and mammalian cells (Cherry and Johnston 1987, Cherry *et al.* 1989, Fantes *et al.* 1989, Schmitz *et al.* 1992). These authors showed that differences in kinetochore fluorescence between chromosomes are minor and appear to be independent of chromosome length. In a more recent study, Irvine and colleagues measured direct fluorescence of CENH3 by fusing it with a fluorescent proteins EGFP and expressing the fusion proteins in human cell culture (Irvine *et al.* 2004). This study revealed a weak (40%) correlation between chromosome size and CENH3 incorporation (Irvine *et al.* 2004).

Beyond the idea that longer chromosomes should require larger kinetochores, it is not clear what factors might be involved in regulating overall kinetochore size. It has been observed that chromosome number is inversely correlated with average kinetochore size in mammalian species (Cherry *et al.* 1989). Yet the authors showed that a four-fold increase in chromosome number resulted in only a 50% decrease in average kinetochore size, suggesting that additional factors are involved. It has also been found that the size of the CENH3-specified domains is in proportion to total centromeric satellite repeats in humans (Sullivan *et al.* 2011). However, this correlation may not be applicable to other species. For example, although the number of centromere-specific satellite repeats varies by a factor of five between rice centromeres 4 and 8, CENH3-containing nucleosomes occupy a domain of similar size on these two chromosomes (Feng *et al.* 2002, Nagaki *et al.* 2004).

CENH3 domains can also change in size without extending their boundaries. Normally in the centromere core, blocks of CENH3-containing nucleosomes are interspersed with blocks of H3-containing nucleosomes (Blower *et al.* 2002). When CENH3 is overexpressed in yeast and humans, H3-containing nucleosomes are replaced by CENH3 nucleosomes (Lam *et al.* 2006, Burrack *et al.* 2011). As a consequence, more kinetochore proteins and spindle microtubules are recruited to the centromeres but CENH3 incorporation does not spread to the DNA sequences flanking centromeres (Burrack *et al.* 2011). This observation is different from neocentromere formation, where CENH3 incorporation spreads to the chromosome arms when CENH3 expression level is increased (Heun *et al.* 2006, Olszak *et al.* 2011).

Size plasticity of the CENH3 domains has also been reported in plants. It has been shown that the amount of CENH3 at maize B centromeres is often less than that in the A centromeres (Jin *et al.* 2005). Size variation is more striking in the situations when two species are crossed. In the oat-maize additional lines, neocentromeres can be assembled on maize chromosome fragments that have lost their original kinetochores (Topp *et al.* 2009). The size of the CENH3 domains on these fragments ranges from 30% to 90% of the oat centromeres (Topp *et al.* 2009). The fact that CENH3 domains on the neocentromeres and B centromeres are substantially smaller than normal centromeres indicates that kinetochores can reduce in size without interference on their functions.

Here we took advantage of an anti-rice CENH3 antibody that can be used in multiple grass species for immunofluorescence (Jin *et al.* 2004, Nagaki *et al.* 2004, Liu *et al.* 2008, Sanei *et al.* 2011). By comparing species that diverged 60~80 million years ago (Bennetzen and Freeling 1997) and which vary by 40-fold in genome size, we showed that the size of the CENH3 domains correlates with genome size and chromosome number in a predictable way. In

addition we analyzed plants derived from inter-species crosses, and found that centromere size is flexible. Our results suggest that average centromere size is strongly influenced by overall cell size and genome architecture.

Results

CENH3 domains vary in size across species

In order to examine size variation of the CENH3-containing nucleosomes, we performed immunofluorescence in ten species of the grass family using an antibody that recognizes the extreme amino-terminus of the rice CENH3 protein (Nagaki *et al.* 2004). Although this region is dramatically divergent among species, it is relatively well conserved in the cereal grasses and close relatives (Figure 4.1a). Previous studies have shown that this antibody is able to recognize CENH3 in rice, maize, oat, barley and wheat (Jin *et al.* 2004, Nagaki *et al.* 2004, Liu *et al.* 2008, Sanei *et al.* 2011). We found that this antibody also stains rye, *Zea luxurians*, pearl millet, foxtail millet and sorghum (Figure 4.1b).

We then compared the size of the CENH3 domains across all ten taxa. We were concerned that signal intensity of CENH3 staining may vary naturally due to different affinities between the antibodies and the different CENH3 proteins. Our method is based on staining area measurement instead of staining intensity and it only requires that the staining be strong and above background so as to establish a clear threshold. As shown in Figure 4.2, the threshold was set as one unit (based on fluorescence intensity) above the brightest background staining so that only kinetochore stainings were measured. This approach is independent of signal exposure time, as overexposing images will increase background staining as well as kinetochore staining

at the same magnitude. We also performed immunofluorescence using antibodies specific to maize and oat (Zhong *et al.* 2002, Topp *et al.* 2009) and compared the results with measurements using the rice CENH3 antibody. Our results suggested that kinetochore staining area measurements reflect the actual size of the CENH3 domains and are not affected by antibody binding affinities (data not shown).

For each species, CENH3 staining area per chromosome was measured in at least 20 cells from two-dimensional projection images (see materials and methods). The areas were then transformed into volume measurements presuming kinetochores are spherical structures (Dawe *et al.* 2005). Each species has a statistically unique kinetochore volume except between oat and pearl millet and between sorghum and foxtail millet (Figure 4.1c).

CENH3 domain is correlated with genome size divided by chromosome number

It has been documented previously that there is an inverse relationship between average kinetochore size and chromosome number in mammalian cells (Cherry *et al.* 1989). To test whether chromosome number is a factor in determining size of the CENH3 domains in plants, we examined ten species of the grass family, which vary from 14 chromosomes to 42 chromosomes. Comparisons of CENH3 staining to chromosome number did not reveal any apparent correlations (Figure 4.3). We also considered that total genome size might be a better predictor of chromosomal CENH3. These comparisons revealed a weak positive correlation, suggesting that genome size can explain about 26% of the variation in CENH3 staining (Figure 4.4). Notably, data points representing rye and barley deviate from the linear regression most remarkably (Figure 4.4), both of which have large genomes with only a few chromosomes.

We next investigated whether CENH3 staining is better correlated with average

chromosome size. In most members of the grass family, chromosome size within a species varies by about two fold (Takashi Matsumoto1 *et al.* 2005, Paterson *et al.* 2009, Schnable *et al.* 2009) and CENH3 staining on the chromosomes of a species is generally constant (Topp *et al.* 2009). In contrast, comparison of chromosomes across species can reveal very large differences. For example, rye chromosomes are about 30 times bigger than rice chromosomes on average. To assess CENH3 staining on a chromosome size basis, we compared CENH3 staining to genome size divided by chromosome number. These data revealed a strong 91.1% correlation between these two variables (Figure 4.5). Correlations such as those in Figure 4.5 can be strongly affected by evolutionary context; for instance, in cases where the species cluster into closely related subgroups the correlations may be artificially elevated (Felsenstein 1985). To test for such artifacts, the same regression analysis was performed under conditions weighted by evolutionary context (Martins 2004) using a phylogeny based on the trnL-trnF intergenic spacer of the chloroplast genome (Drabkova *et al.* 2004). This modification of the statistics mildly loosened the correlation, indicating that 90% of total variation in CENH3 staining can be explained as a function of genome size divided by chromosome number (data not shown).

We also investigated whether there might be a correlation between the size of the CENH3 domains and the number of microtubule attachments on the kinetochores. The tubulin staining assay is not as definitive as a kinetochore staining assay, since microtubule mass can be affected by spindle length and microtubule number. We were unable to differentiate these two factors. It is clear, however, that the density of microtubules is much higher in species with larger CENH3 domains (e.g. barley) and vice versa (e.g. sorghum, Figure 4.6).

CENH3 domains change in size rapidly in new environment

Previous studies have shown that CENH3 domains change in size on maize neocentromeres in the oat background (Topp *et al.* 2009). To investigate whether the amount of CENH3 changes on regular centromeres when two species are crossed, we studied oat-maize additional lines retaining intact maize chromosomes (Ananiev *et al.* 1997, Kynast *et al.* 2004). Immunofluorescence was performed on root tip cells from three oat-maize additional lines, which contain maize chromosomes 2, 6, and 9. A maize-specific FISH probe was used to differentiate the maize chromosomes from the oat chromosomes (Figure 4.7a). In all three lines, the sizes of the CENH3 domains on the introduced maize chromosomes were more variable and significantly larger than those from an average maize chromosome in the progenitors (Figure 4.7b). These data suggest that when oat and maize are crossed, the size of their CENH3 domains tends to become homogenized; CENH3 domains on the oat chromosomes remain unchanged in size, while maize chromosomes have recruited substantially more CENH3 to their centromeres. Our results support the assertion that centromeres may change in size when introduced into a new genetic background.

Discussion

In eukaryotes, each species has established a specialized CENH3 domain for kinetochore assembly and microtubule attachment. For example, centromeres in the budding yeast *S. cerevisiae* have a single CENH3 nucleosome and bind one spindle microtubule (Joglekar *et al.* 2006). Centromeres in fission yeast and higher eukaryotes have much longer CENH3 domains and multiple microtubule attachment sites (Cheeseman and Desai 2008, Joglekar *et al.* 2008).

Due to the highly repetitive characteristics of centromeric sequences, an accurate assessment of centromere sizes is only possible in species with minimal or frequently interrupted satellite repeats (Yan *et al.* 2008, Wolfgruber *et al.* 2009). In contrast, measuring CENH3 domains by means of immunofluoresence is more technically feasible and has been performed in several previous studies (Cherry and Johnston 1987, Cherry *et al.* 1989, Fantes *et al.* 1989, Schmitz *et al.* 1992). The limitations of this method for comparative analysis, however, lie in the availability of a common antibody that will recognize kinetochores across multiple related species. Such antibodies are very rare because the defining N-terminal region of CENH3 is variable by nature (Henikoff *et al.* 2001). Grasses are unusual in having a conserved motif at the far N-terminus (Talbert *et al.* 2004) and we are fortunate to have access to an antibody raised against rice CENH3 with broad reactivity (Nagaki *et al.* 2004). Ten species were selected for this study that together represent much of the total variability in the true grasses (Soreng and Davis 1998).

Our measurements showed that each species in the grass family has a characteristic average centromere size, as represented by CENH3 (Figure 4.1C). Further, there is a remarkably strong correlation between average CENH3 volume per chromosome and the quotient of genome size divided by chromosome number (Figure 4.5). This correlation enables us to estimate centromere sizes in the grass species given the value of rice centromeres (Table 1). We estimated that maize centromeres occupy ~ 4800 kb domains, similar to what have been documented for maize chromosome 5 (4200 kb) based on chromatin immuno-precipitation (Wolfgruber *et al.* 2009). Compare to the conventional centromere mapping method, our approach is much faster and can be achieved in species with highly repetitive sequences.

One interpretation of the correlation between centromere size and average chromosome size would be that larger chromosomes require larger kinetochores; however, we do not think

this is the best explanation. In humans, except for Y chromosome, which has a substantially small kinetochore, interchromosomal kinetochore variation is less than two fold, when chromosome size varies by more than five fold (Cherry and Johnston 1987, Irvine *et al.* 2004). In the grasses, chromosome size varies within a species by two-three fold, but kinetochores sizes do not appear to follow this trend (Jin *et al.* 2005, Topp *et al.* 2009). Our own results support the interpretation of a roughly uniform kinetochore size in any given species (Figure 4.1b). Considering that DNA content positively correlates with cell and nuclear volume (Price *et al.* 1973, Szarski 1976), we speculate that the size of the CENH3 domains (and overlying kinetochore complexes) are more strongly influenced by overall cell size.

The numbers of kinetochore microtubules seem to follow the same trend as the kinetochores: while each species has a distinct number of microtubules per kinetochore, within a species chromosome volume can vary up to ten fold with little or no increase in microtubule numbers (Moens 1979, Lin *et al.* 1981, Rieder 1982)(McEwen *et al.* 1998). To analyze the relationship between microtubule number and average chromosome size, we used previous measurements in varied species based on electron microscopy (Table 2) and found a positive relationship between the microtubule number and average chromosome size (R=0.92). This correlation indicates that there is an association between centromere size and kinetochore microtubules. Indeed, our immunofluorescent images revealed that kinetochore microtubules are more abundant on larger kinetochores (Figure 4.6).

To accommodate the relationship between genome size, chromosome number, kinetochore size and microtubule numbers, we proposed a model that incorporates cell volume, microtubule distribution and CENH3 domains (Figure 4.8). We suggest that the size of the CENH3 domains is a reflection of how much kinetochore area is required to stabilize spindle

microtubules: Cells with few chromosomes will have relatively large kinetochores that are similar in total volume to a cell with many chromosomes each with small kinetochores (Figure 4.8). This model does not imply that every kinetochore must be the same size as all others in a cell, but rather suggests that uniform kinetochore size is a natural stable state.

Our model suggests that most centromeres, even those on small chromosomes, will stabilize at a size that matches a genome average. We tested this hypothesis by studying the result when two species with different sized CENH3 domains are crossed. Such a cross was made previously between maize and oat (Rines *et al.* 1995). By analyzing oat lines with single maize chromosomes, we showed the oat genome does appear to have the affect of increasing the size of an introduced maize centromere (Figure 4.7). It is interesting to consider the questions of how and where on the maize chromosomes the CENH3 domains expanded. Although we did not address these issues directly, we favor a "filling in" mechanism, where the centromeres do not expand into the flanking pericentromeric heterochromatin but instead increasing the density of CENH3 nucleosomes within the confines of the existing centromere, presumably by replacing H3-containing nucleosomes. Similar replacement of H3 nucleosomes has been documented in yeast and human, where overexpression of CENH3 resulted in the CENH3 domains expanding internally (Lam *et al.* 2006, Burrack *et al.* 2011).

CENH3 domains not only expand but also shrink. It has been observed in *Neurospora crassa* when heterochromatin is negatively affected (Smith *et al.* 2011). The published works together with our observations suggest that the exchange between H3 nucleosomes and CENH3 nucleosomes is dynamic and may be regulated by epigenetic factors. The flexibility of the CENH3 domains, including de novo formation on ectopic loci, spreading to proximate regions, expand and shrinkage, facilitate cells to better adjust to new environment. Future studies will

need to address whether such changes are a cumulative result of multiple rounds of adjustment during mitosis or occur within a single cell cycle.

Materials and methods

Materials

Seeds of oat (Southern States), barley (Fox), rye (Athens Abruzzi) and *sorghum bicolor* were ordered from USDA-ARS. Seeds of wheat (Chinese Spring), foxtail millet (Yugul), pearl millet (Tift23db) and rice were generously provided by Katrien M. Devos. Maize (B73) and *Z. luxurians* were from our own lab.

Cytological preparation and observation

Seeds from different species were germinated on wet paper towels for three days to a week depending on the species. Root tips were fixed in 4% paraformaldehyde diluted in Phosphate Buffered Saline (PBS) buffer for 30 min and stored in methanol if not directly used for immunolocalization. Followed fixation, root tips were digested in an enzyme mix containing 4% cellulase, 2% pectolyase in 10mM Citric Buffer, pH 5.5, for 20 min~1 hours at 37°C. Digested root tip cells were rinsed in PBS buffer before dropping on the slides and centrifuged at 100g for 1min to flatten the cells. Slides were incubated with an anti-rice CENH3 antibody (1:200; generously provided by Jiming Jiang) overnight at 4°C, followed by 1h blocking with goat serum (1:10) and 2h incubation with secondary FITC-conjugated goat anti-rabbit antibodies (1:200) at room temperature. Vectashield containing DAPI was mounted on the slides. All images were

captured and processed using a Zeiss Axio Imager microscope and SlideBook 5.0 software (Intelligent Imaging Innovations, Denver, CO, USA).

ImmunoFISH was performed the same way as immunofluorescence, followed by ten minutes wash in PBS buffer, ten minutes post-fixation in 4% paraformaldehyde diluted in PBS buffer, and brief rinse in PBS before performing FISH. FISH was performed as previously described (Kato *et al.* 2004), except that probes and slides were denatured at the same time in the boiling water.

Kinetochore measurement and statistical analysis

Fluorescent signals were captured as 3D images using the Digital Microscope Workstation as previously described (Du and Dawe 2007). A projection image was made for each capture and deconvolved using SlideBook 5.0 software to reduce background interference. Fluorescent signals were measured using the mask tool to select regions of interest. A threshold of signal intensity was set so that none of the regions selected were outside the nuclei. Areas covering the selected regions were measured and transformed into volume assuming the kinetochores are spherical structures. Excel software was used to calculate the average size of the CENH3 domains and standard deviation of each species. Linear regression and correlation coefficient was also calculated using the excel software.

Standard T-test was performed to test the statistical significance of differences between average sizes of the CENH3 domains in different species (Figure 4.1c) and between CENH3 domains on oat chromosomes and maize chromosomes in the oat-maize additional lines (Figure 4.7). Independence contrast was applied using the COMPARE 4.6b software (http://www.indiana.edu/~martinsl/compare/) with the following newick trees: without branch

length (((Sor,(Mai,Lux)),(Fox,Pea)),((Oat,(Bar,(Rye,Whe))),Ric)) and with branch length (((Sor:0.00950939280685307413,(Mai:0.00231520485691936832,Lux:0.010523325331448282 06):0.00367176807056998015):0.0104171180333998275 6,(Fox:0.01363691334023187443,Pea: 0.01520671944747864317):0.01592780163032410726):0.02586075002500199532,((Oat:0.0307 2769607026095903,(Bar:0.01983366653899813412,(Rye:0.01449881409177665101,Whe:0.003 50123323061236674):0.00909150693055843159):0.02146457938246661493):0.018074214022 16232917,Ric:0.04736460143268115403):0.00833952027397439483).

Acknowledgements

We thank Katrien M. Devos for providing the seeds stocks and Michael McKain for generating the newick trees for phylogenic study. This work was supported by grants 0421671 and 092270 from the National Science Foundation.

References

Ananiev, E.V., Riera-Lizarazu, O., Rines, H.W. and Phillips, R.L. (1997) Oat-maize chromosome addition lines: a new system for mapping the maize genome. *Proc Natl Acad Sci U S A*, **94**, 3524-3529.

Bennetzen, J.L. and Freeling, M. (1997) The unified grass genome: synergy in synteny. *Genome Res*, **7**, 301-306.

Black, B.E., Jansen, L.E., Maddox, P.S., Foltz, D.R., Desai, A.B., Shah, J.V. and Cleveland, D.W. (2007) Centromere identity maintained by nucleosomes assembled with histone H3 containing the CENP-A targeting domain. *Mol Cell*, **25**, 309-322.

Blower, M.D., Sullivan, B.A. and Karpen, G.H. (2002) Conserved organization of centromeric chromatin in flies and humans. *Dev Cell*, **2**, 319-330.

Burrack, L.S., Applen, S.E. and Berman, J. (2011) The requirement for the Dam1 complex is dependent upon the number of kinetochore proteins and microtubules. *Curr Biol*, **21**, 889-896.

Cheeseman, I.M. and Desai, A. (2008) Molecular architecture of the kinetochore-microtubule interface. *Nat Rev Mol Cell Biol*, **9**, 33-46.

Cherry, L.M., Faulkner, A.J., Grossberg, L.A. and Balczon, R. (1989) Kinetochore size variation in mammalian chromosomes: an image analysis study with evolutionary implications. *J Cell Sci*, **92 (Pt 2)**, 281-289.

Cherry, L.M. and Johnston, D.A. (1987) Size variation in kinetochores of human chromosomes. *Hum Genet*, **75**, 155-158.

Dawe, R.K., Richardson, E.A. and Zhang, X. (2005) The simple ultrastructure of the maize kinetochore fits a two-domain model. *Cytogenet Genome Res*, **109**, 128-133.

Ding, R., McDonald, K.L. and McIntosh, J.R. (1993) Three-dimensional reconstruction and analysis of mitotic spindles from the yeast, Schizosaccharomyces pombe. *J Cell Biol*, **120**, 141-151.

Drabkova, L., Kirschner, J., Vlcek, C. and Paces, V. (2004) TrnL- trnF intergenic spacer and trnL intron define major clades within Luzula and Juncus (Juncaceae): importance of structural mutations. *J Mol Evol*, **59**, 1-10.

Fantes, J.A., Green, D.K., Malloy, P. and Sumner, A.T. (1989) Flow cytometry measurements of human chromosome kinetochore labeling. *Cytometry*, **10**, 134-142.

Felsenstein, J. (1985) Phylogenies and the Comparative Method. *Am Nat*, **125**, 1-15.

Feng, Q., Zhang, Y., Hao, P., Wang, S., Fu, G., Huang, Y., . . . Han, B. (2002) Sequence and analysis of rice chromosome 4. *Nature*, **420**, 316-320.

Henikoff, S., Ahmad, K. and Malik, H.S. (2001) The centromere paradox: stable inheritance with rapidly evolving DNA. *Science*, **293**, 1098-1102.

Heun, P., Erhardt, S., Blower, M.D., Weiss, S., Skora, A.D. and Karpen, G.H. (2006) Mislocalization of the Drosophila centromere-specific histone CID promotes formation of functional ectopic kinetochores. *Dev Cell*, **10**, 303-315.

Irvine, D.V., Amor, D.J., Perry, J., Sirvent, N., Pedeutour, F., Choo, K.H. and Saffery, R. (2004) Chromosome size and origin as determinants of the level of CENP-A incorporation into human centromeres. *Chromosome Res*, **12**, 805-815.

Jin, W., Lamb, J.C., Vega, J.M., Dawe, R.K., Birchler, J.A. and Jiang, J. (2005) Molecular and functional dissection of the maize B chromosome centromere. *Plant Cell*, **17**, 1412-1423.

Jin, W., Melo, J.R., Nagaki, K., Talbert, P.B., Henikoff, S., Dawe, R.K. and Jiang, J. (2004) Maize centromeres: organization and functional adaptation in the genetic background of oat. *Plant Cell*, **16**, 571-581.

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

Joglekar, A.P., Bouck, D.C., Molk, J.N., Bloom, K.S. and Salmon, E.D. (2006) Molecular architecture of a kinetochore-microtubule attachment site. *Nat Cell Biol*, **8**, 581-585.

Kato, A., Lamb, J.C. and Birchler, J.A. (2004) Chromosome painting using repetitive DNA sequences as probes for somatic chromosome identification in maize. *Proc Natl Acad Sci U S A*, **101**, 13554-13559.

Kynast, R.G., Okagaki, R.J., Galatowitsch, M.W., Granath, S.R., Jacobs, M.S., Stec, A.O., . . . Phillips, R.L. (2004) Dissecting the maize genome by using chromosome addition and radiation hybrid lines. *Proc Natl Acad Sci U S A*, **101**, 9921-9926.

Lam, A.L., Boivin, C.D., Bonney, C.F., Rudd, M.K. and Sullivan, B.A. (2006) Human centromeric chromatin is a dynamic chromosomal domain that can spread over noncentromeric DNA. *Proc Natl Acad Sci U S A*, **103**, 4186-4191.

Lin, H.P., Ault, J.G. and Church, K. (1981) Meiosis in Drosophila melanogaster. I. Chromosome identification and kinetochore microtubule numbers during the first and second meiotic divisions in males. *Chromosoma*, **83**, 507-521.

Liu, Z., Yue, W., Li, D., Wang, R.R., Kong, X., Lu, K., . . . Zhang, X. (2008) Structure and dynamics of retrotransposons at wheat centromeres and pericentromeres. *Chromosoma*, **117**, 445-456.

Martins, E. P. (2004) COMPARE, version 4.6b. Computer programs for the statistical analysis of comparative data. Distributed by the author at http://compare.bio.indiana.edu/. Department of Biology, Indiana University, Bloomington IN.

Matsumoto, T., Wu, J., Kanamori, H., Katayose, Y., Fujisawa, M., Namiki, N., . . . Burr, B. (2005) The map-based sequence of the rice genome. *Nature*, **436**, 793-800.

McEwen, B.F., Ding, Y. and Heagle, A.B. (1998) Relevance of kinetochore size and microtubule-binding capacity for stable chromosome attachment during mitosis in PtK1 cells. *Chromosome Res*, **6**, 123-132.

Moens, P.B. (1979) Kinetochore microtubule numbers of different sized chromosomes. *J Cell Biol*, **83**, 556-561.

Nagaki, K., Cheng, Z., Ouyang, S., Talbert, P.B., Kim, M., Jones, K.M., . . . Jiang, J. (2004) Sequencing of a rice centromere uncovers active genes. *Nat Genet*, **36**, 138-145.

Olszak, A.M., van Essen, D., Pereira, A.J., Diehl, S., Manke, T., Maiato, H., . . . Heun, P. (2011) Heterochromatin boundaries are hotspots for de novo kinetochore formation. *Nat Cell Biol*, **13**, 799-808.

Partridge, J.F., Borgstrom, B. and Allshire, R.C. (2000) Distinct protein interaction domains and protein spreading in a complex centromere. *Genes Dev*, **14**, 783-791.

Paterson, A.H., Bowers, J.E., Bruggmann, R., Dubchak, I., Grimwood, J., Gundlach, H., . . . Rokhsar, D.S. (2009) The Sorghum bicolor genome and the diversification of grasses. *Nature*, **457**, 551-556.

Peterson, J.B. and Ris, H. (1976) Electron-Microscopic Study of Spindle and Chromosome Movement in Yeast Saccharomyces-Cerevisiae. *Journal of Cell Science*, **22**, 219-242.

Price, H.J., Sparrow, A.H. and Nauman, A.F. (1973) Correlations between Nuclear Volume, Cell Volume and DNA Content in Meristematic Cells of Herbaceous Angiosperms. *Experientia*, **29**, 1028-1029.

Rieder, C.L. (1982) The formation, structure, and composition of the mammalian kinetochore and kinetochore fiber. *Int Rev Cytol*, **79**, 1-58.

Rines, H. W., Riera-Lizarazu, O., Phillips, R. L. (1995) in *Modification of Gene Expression and Non-Mendelian Inheritance*, eds Oono K, Takiwa F(National Institute of Agrobiological Resources, Tsukuba, Japan), pp 235–251.

Sanei, M., Pickering, R., Kumke, K., Nasuda, S. and Houben, A. (2011) Loss of centromeric histone H3 (CENH3) from centromeres precedes uniparental chromosome elimination in interspecific barley hybrids. *Proc Natl Acad Sci U S A*, **108**, E498-505.

Schmitz, A., Guilly, M.N., Trask, B. and van den Engh, G. (1992) Immunofluorescent labeling of centromeres for flow cytometric analysis. *Cytometry*, **13**, 478-484.

Schnable, P.S., Ware, D., Fulton, R.S., Stein, J.C., Wei, F., . . . Wilson, R.K. (2009) The B73 maize genome: complexity, diversity, and dynamics. *Science*, **326**, 1112-1115.

Smith, K.M., Phatale, P.A., Sullivan, C.M., Pomraning, K.R. and Freitag, M. (2011) Heterochromatin is required for normal distribution of Neurospora crassa CenH3. *Mol Cell Biol*, **31**, 2528-2542.

Soreng, R.J. and Davis, J.I. (1998) Phylogenetics and character evolution in the grass family (Poaceae): Simultaneous analysis of morphological and chloroplast DNA restriction site character sets. *Bot Rev*, **64**, 1-85.

Sullivan, L.L., Boivin, C.D., Mravinac, B., Song, I.Y. and Sullivan, B.A. (2011) Genomic size of CENP-A domain is proportional to total alpha satellite array size at human centromeres and expands in cancer cells. *Chromosome Res*, **19**, 457-470.

Szarski, H. (1976) Cell size and nuclear DNA content in vertebrates. *Int Rev Cytol*, **44**, 93-111.

Talbert, P.B., Bryson, T.D. and Henikoff, S. (2004) Adaptive evolution of centromere proteins in plants and animals. *J Biol*, **3**, 18.

Tenaillon, M.I., Hufford, M.B., Gaut, B.S. and Ross-Ibarra, J. (2011) Genome size and transposable element content as determined by high-throughput sequencing in maize and Zea luxurians. *Genome Biol Evol*.

Topp, C.N., Okagaki, R.J., Melo, J.R., Kynast, R.G., Phillips, R.L. and Dawe, R.K. (2009) Identification of a maize neocentromere in an oat-maize addition line. *Cytogenet Genome Res*, **124**, 228-238.

Wolfgruber, T.K., Sharma, A., Schneider, K.L., Albert, P.S., Koo, D.H., Shi, J., . . . Presting, G.G. (2009) Maize centromere structure and evolution: sequence analysis of centromeres 2 and 5 reveals dynamic Loci shaped primarily by retrotransposons. *PLoS Genet*, **5**, e1000743.

Yan, H., Talbert, P.B., Lee, H.R., Jett, J., Henikoff, S., Chen, F. and Jiang, J. (2008) Intergenic locations of rice centromeric chromatin. *PLoS Biol*, **6**, e286.

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

Table 4.1 Estimation of centromere size in the grass species.

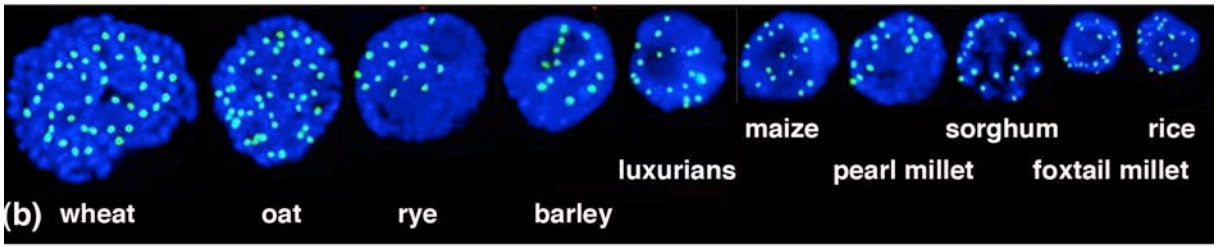
Species	Haploid genome size (Mb)	Haploid chromosome number	Average chromosome size (Mb)	Estimated centromere Size (kb)
Wheat	16937	21	806.5	16880.7
Oat	11300	21	538.1	11262.5
Rye	7917	7	1131.0	23672.1
Barley	5100	7	728.6	15249.2
Luxurians	3300	10	330.0	6907.0
Pearl millet	2350	7	335.7	7026.6
Maize	2300	10	230.0	4814.0
Sorghum	735	10	73.5	1538.4
Foxtail millet	490	9	54.4	1139.5
Rice	430	12	35.8	750.0

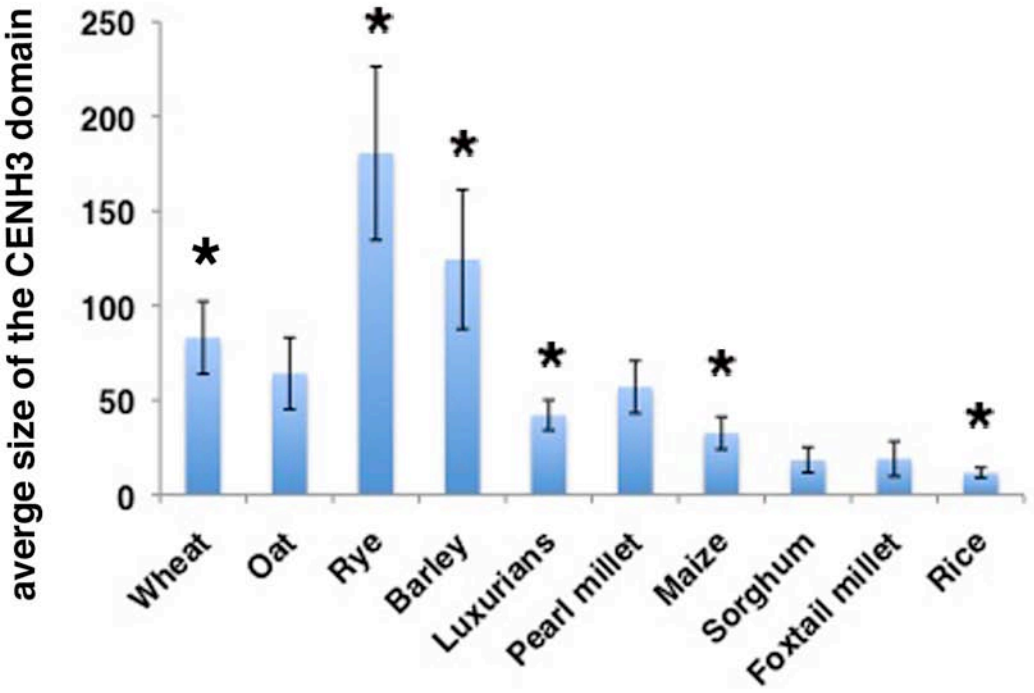
Table 4.2 Number of kinetochore microtubules in different species.

	Species	Haploid genome size (Mb)	Haploid Chromosome number	Average chromosome size (Mb)	Microtubule number per kinetochore
yeast	S.pombe	14.1	3	4.7	3
	S.cerevia	12	16	0.8	1
	Candida albicans	16	8	2.0	1
animals	drosophila	165	4	41.3	5
	human	3000	23	130.4	17
	fetal rats	2800	21	133.3	7
	CHO cells	3032	11	275.6	12
	PtK cells	3000	6	500.0	24
alga	Ostreococcus tauri	12	20	0.6	0.4

```
sorghum   MARTKQTARKSTGGKAPRKQ----------------LATKAARKSAPATGG-------- 35
wheat     MARTKQTARKSTGGKAPRKQ----------------LATKAARKSAPATGG-------- 35
maize     MARTKHQAVRKTAEKPKKKLQFERS----------GGASTSATPERAAGTGGR-AASGGD 49
rice      MARTKHPAVRKSKAEPKKKLQFERSPRPSKAQRAGGGTGTSATTRSAAGTSASGTPRQQT 60
barley    MARTKHPAVRKSKAPPRKKVGSARAPAAAQRRHETDGAGTSETPRRGPAPAADQGAPGEP 60
oat       ----------SKPTPKKQLK----------------FGRSPGQTAEQETGG-------- 25
                    :   . ::                    . .       ...
```

(a)





(c)

Figure 4.1 Size variation of CENH3 domains in different species in the grass family.

(a) ClustalW2 alignment of the N-terminal sequences of CENH3 genes in sorghum, wheat, maize, rice, barley and oat. The sequence from which the anti-rice CENH3 antibody was generated was underlined.

(b) Immuno-fluorescence images showing that the CENH3 domains differ in size in different species of the grass family. Interphase cells were stained with anti-rice CENH3 antibody (green) and DAPI (blue).

(c) Distinct sizes of the CENH3 domains in the grass family. Species whose CENH3 domains are significantly different from others were indicated with stars. CENH3 domains between oat and pearl millet and between sorghum and foxtail millets are not significantly different from each other but significantly different from any other species.

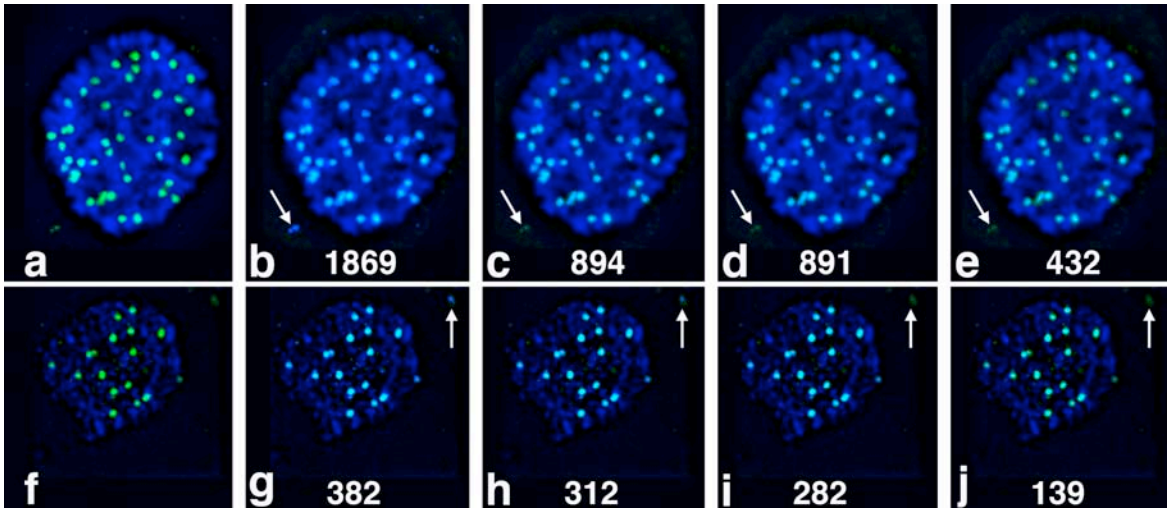


Figure 4.2 Measurement of kinetochore staining area based on different thresholds. Immunofluorescence were performed on oat (a~e) and maize (f~j) root tip cells and stained with DAPI (dark blue) and rice CENH3 antibodies (green). Fluorescent signals were measured using a mask tool to select regions where signal intensities were above the threshold. Kinetochore areas (pixels) covering the selected regions were listed below the corresponding images. Different thresholds setting resulted in different measurements: a and f, no threshold; b and g, excessive background staining; c and h, minor background staining; d and i, no background staining; e and j, reduced staining. In our measurements, the threshold was set as one unit above the brightest background staining (indicated with arrows) so that only kinetochore stainings were measured (d and i).

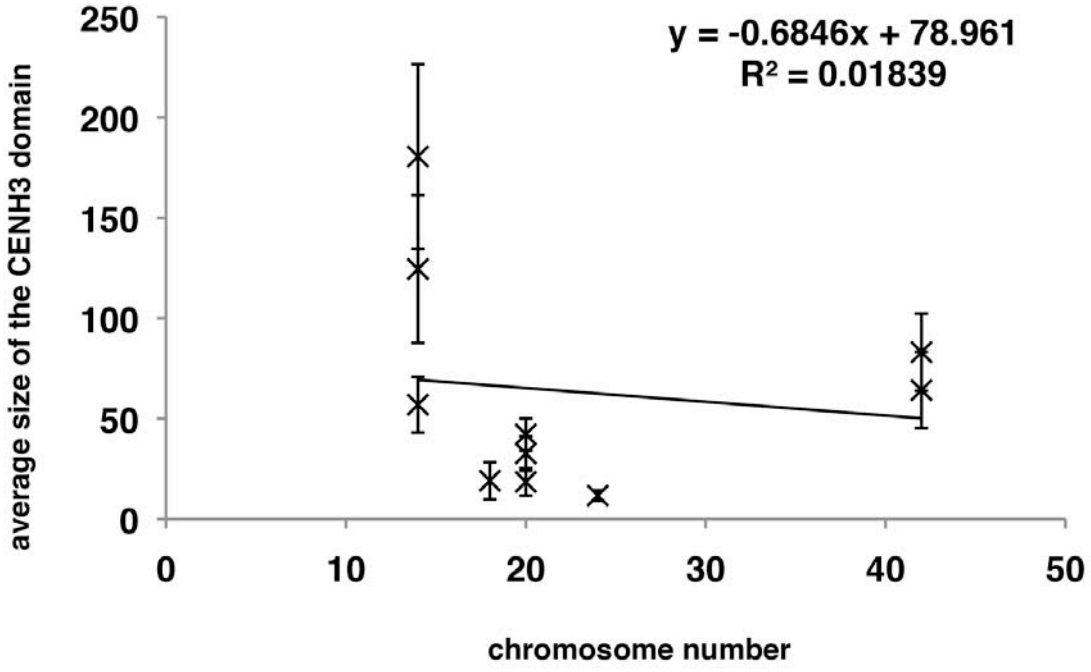


Figure 4.3 Correlation between the size of the CENH3 domains and chromosome number in the grass species.

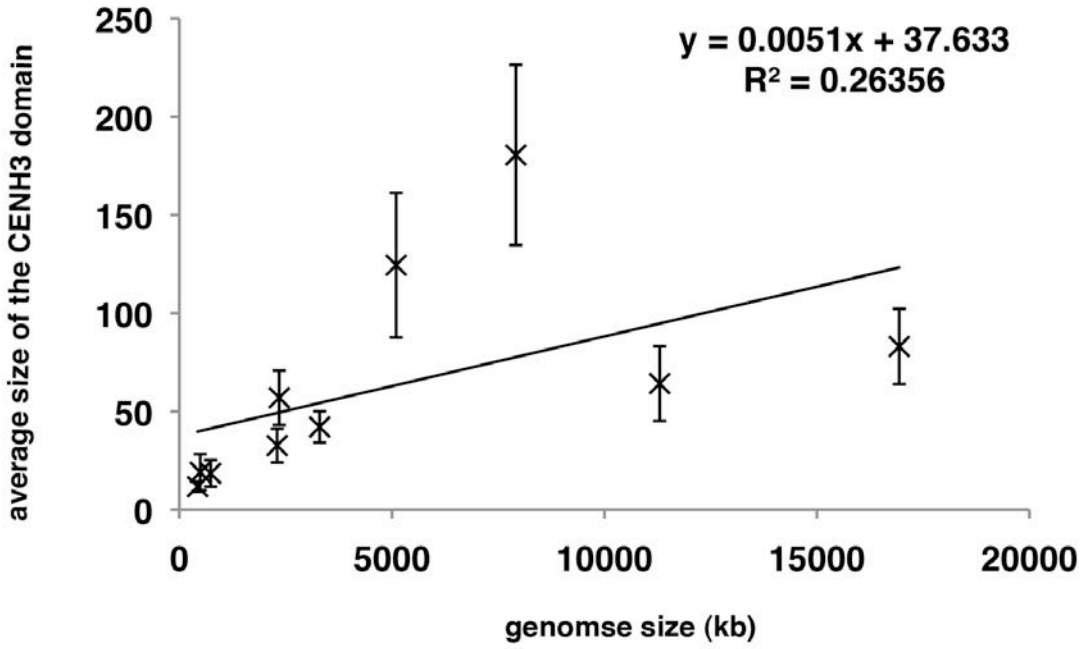


Figure 4.4 Correlation between the size of the CENH3 domains and genome size in the grass species.

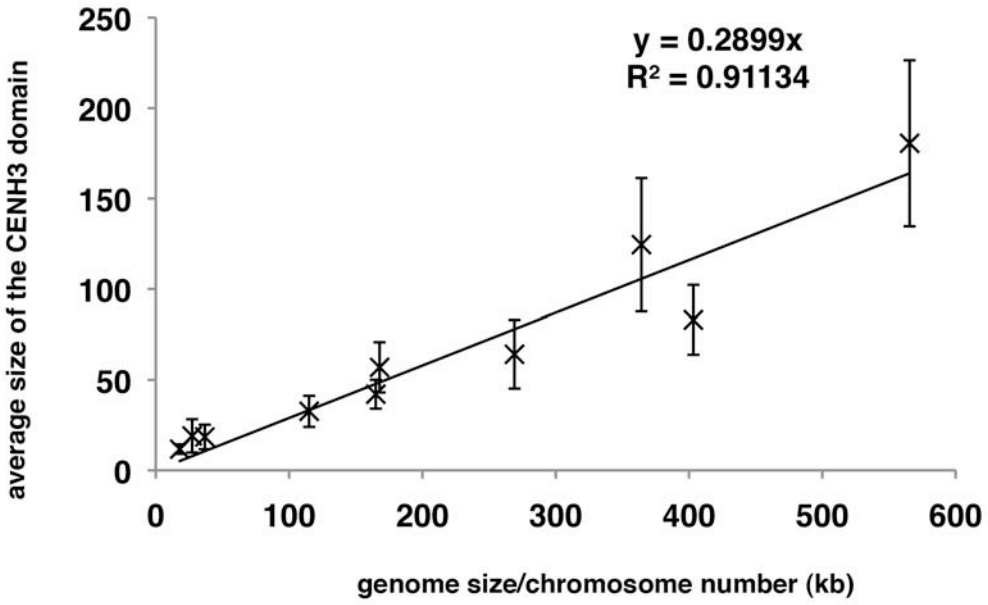


Figure 4.5 Correlation between the size of the CENH3 domain and average chromosome size (genome size/chromosome number) in the grass species.

The (0,0) point was included arbitrarily under the assumption that a null genome would have no CENH3. The correlation coefficient (R^2=0.911) was calculated using the (0,0) point. Without the (0,0) point, R^2=0.914.

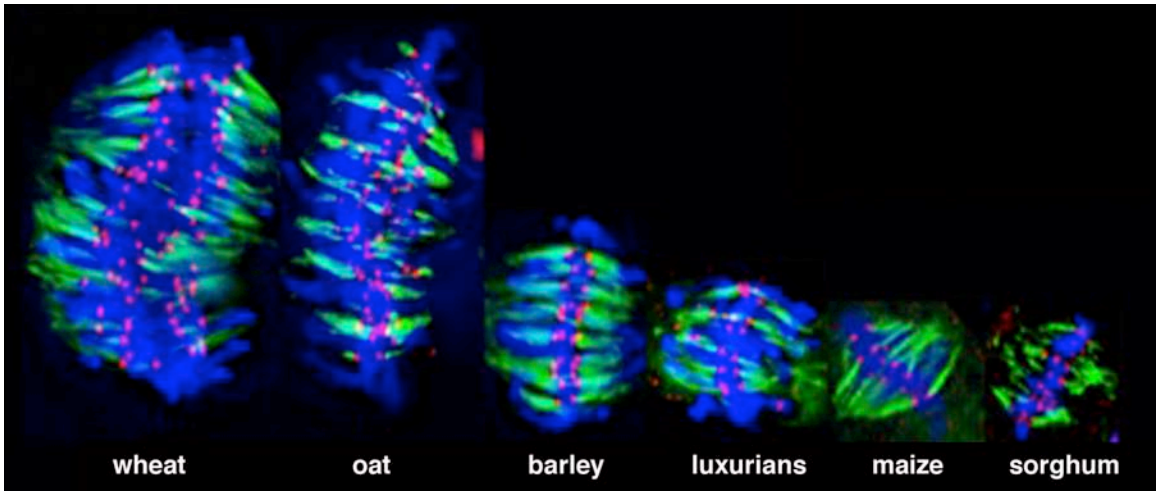


Figure 4.6 Relationship between microtubule density and CENH3 domains in different grass species.

Metaphase chromosomes were stained with anti-rice CENH3 antibody (red), anti-tubulin antibody and DAPI (blue). Notice that microtubules are more abundant in species with large CENH3 domains (e.g. barley) and less abundant in species with small CENH3 domains (e.g. sorghum).

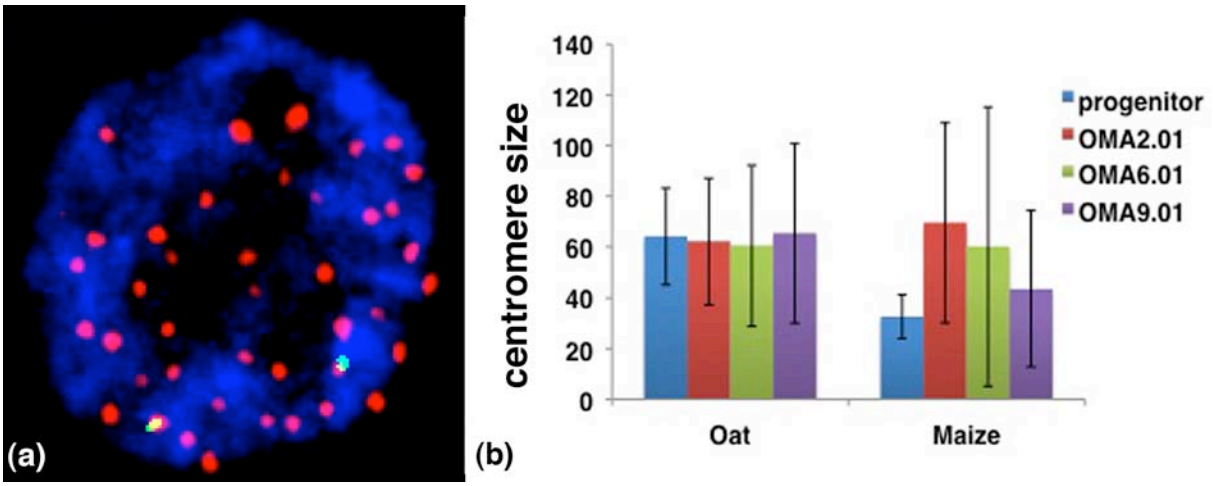


Figure 4.7 Comparison of CENH3 domains on the maize chromosomes and oat chromosomes in the oat-maize additional lines.

(a) An interphase cell from oat-maize additional lines was stained with anti-rice CENH3 antibody (red) and DAPI (blue) and hybridized with a maize centromere-specific probe (green).

(b) Size comparisons between CENH3 domains on the oat chromosomes and maize chromosomes in the oat-maize additional lines OMA2.01, OMA6.01, OMA9.01 as well as the progenitor oat and maize lines.

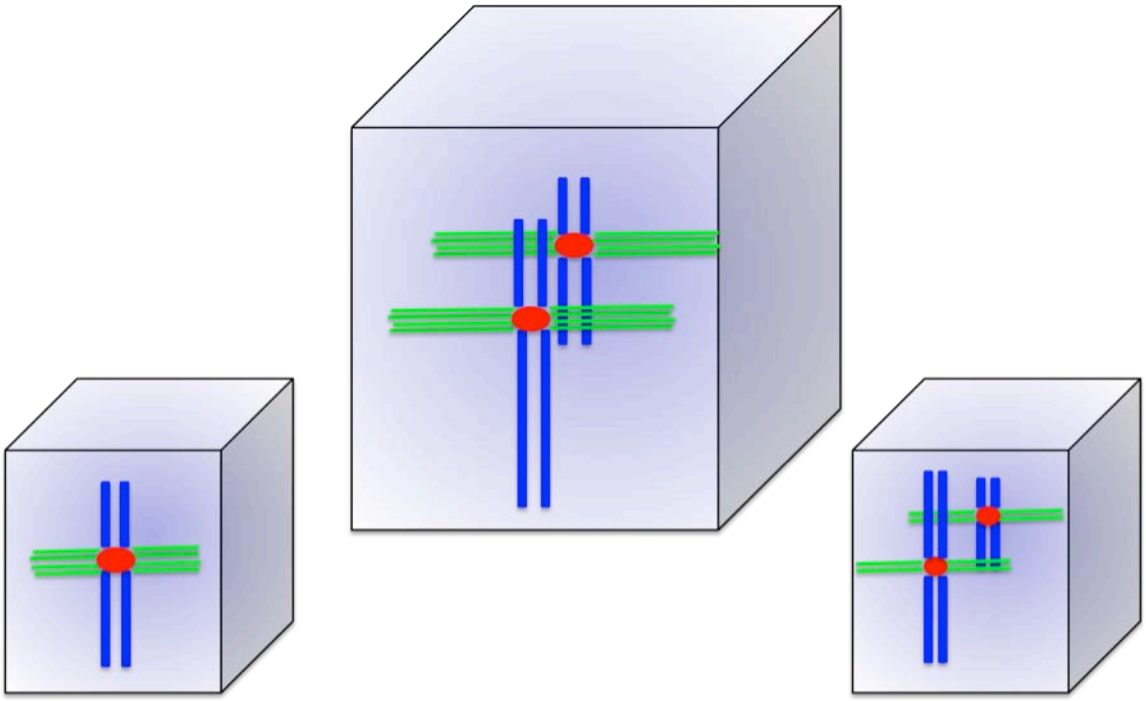


Figure 4.8 A model demonstrating the relationship between kinetochore size, cell size and microtubule abundance.

Within species, CENH3 domains are of similar size independent of chromosome length (middle and right). Sizes of the CENH3 domains in two species with the same genome size (left and right) depend on chromosome number. Less microtubules and smaller CENH3 domains are recruited in species with more chromosomes (right) compared to fewer chromosomes (left). Sizes of the CENH3 domains in two species with same chromosome number (middle and right) depend on genome size. More microtubules and larger CENH3 domains are recruited in species with larger genome (middle) compared to smaller genome (right).

CHAPTER 5

CONCLUSIONS AND FUTURE STUDY

The primary intent of this study was to produce a synthetic repeat array system and use this system to induce neocentromere formation in maize by kinetochore tethering. To this end, we have successfully engineered and introduced centromere-sized satellite repeat arrays into maize by particle bombardment. Three resulting insertion sites (Arrayed Binding Sites, ABS) were described and showed to be sufficient to efficiently tether YFP. The ABS arrays system was designed to be used as a tethering site, similar to the approach used by several recent authors (Barnhart *et al.* 2011, Gascoigne *et al.* 2011) to show that de novo kinetochores can be assembled by tethering kinetochore proteins to specific loci.

We fused kinetochore proteins CENH3 and CENPC with varied tethering proteins and tested whether transgenic fusion proteins can be targeted to the ABS loci on the chromosome arms. Our data showed that tethered CENH3 and CENPC were transiently localized at the binding arrays but did not form neocentromeres. Instead, we observed severe chromosome segregation defects in both mitosis and meiosis. We concluded that plant kinetochores are exceptionally stable and rarely move even under strong selection. Therefore tethered CENH3 or CENPC are more likely to destabilize the function of existing kinetochores than to induce the formation of new ones. Possible reasons of neocentromere suppression were discussed in Chapter 3. Here in this chapter I propose relevant studies that will extend our understanding of

plant kinetochore formation and neocentromere establishment. The chapter is worded in the manner of a grant-proposal, often using the "we" and "will" terminology typical of this format.

Characterization of Subdomains within Plant CENPC

CENPC is a key kinetochore protein that has been found in plants, animals, and yeasts. CENPC lies at the innermost face of the kinetochore, and functions as the principal linkage between inner centromeric chromatin and outer kinetochore proteins (Przewloka *et al.* 2011). It has been well studied that in animals CENPC is composed of three domains: an amino-terminal domain that binds to the MIS12 complex and faces outer kinetochores (Screpanti *et al.* 2011), a central domain which binds centromeric chromatin (Yang *et al.* 1996, Song *et al.* 2002, Trazzi *et al.* 2002, Milks *et al.* 2009, Carroll *et al.* 2010) and a conserved C-terminal end that is required for new CENH3 deposition (Moree *et al.* 2011). Within the C-terminal end, a small region of 23 amino acids (referred to as the CENPC motif) is highly conserved and has been found in all the eukaryotes studied (Brown 1995, Talbert *et al.* 2004).

No sequence similarity between plant and animal CENPC can be detected outside the CENPC motif, nevertheless, plant and animal CENPC appear to share an overall architecture (Talbert *et al.* 2004). The carboxyl-terminus of CENPC is highly conserved within angiosperm families (Talbert *et al.* 2004). It seems reasonable that this domain is responsible for new CENH3 assembly and centromere targeting, similar to what has been documented in animals (Trazzi *et al.* 2009, Moree *et al.* 2011). The central domain of plant CENPC is under adaptive selection (Talbert *et al.* 2004), suggesting that this domain is interacting with the highly divergent centromeric DNA. Indeed, it has been reported that in maize DNA binding capacity of

CENPC is limited to exon 9-12 (Du *et al.* 2010). Removal or replacement of the DNA binding domain reduces centromere localization slightly, suggesting that this is another centromere targeting domain within CENPC (Du *et al.* 2010). These observations are in consistent with the animal literature, which showed that two centromere targeting domains are present within CENPC, one in the central domain and the other in the C-terminal domain (Milks *et al.* 2009, Trazzi *et al.* 2009).

The amino-terminus of CENPC is poorly characterized in plants, but well studied in animals. It has been shown that in both drosophila and human cell culture the N-terminal domain is sufficient to assemble de novo kinetochores (Gascoigne *et al.* 2011, Przewloka *et al.* 2011), although no sequence homology is shared between these two species at this region (HEEGER *et al.* 2005). Considering that the overall structure of CENPC is well conserved in eukaryotes, we speculate that the N-terminal domain of CENPC in plants functions similarly as its homolog in animals. Within angiosperms, the amino-terminus of CENPC is conserved in the first five exons as well as the beginning of the sixth exon (Talbert *et al.* 2004). We propose this relatively conserved region is the protein-binding domain of plant CENPC that interacts with outer kinetochores.

A functional dissection of Arabidopsis CENPC protein will be performed using truncated constructs (Figure 5.1). Fluorescent proteins will be fused with each construct in frame to test their sufficiency for centromere targeting. Biotin tags will also be fused to these constructs to test whether they can co-immunoprecipitate with MI12 complex (FURUYAMA and HENIKOFF 2006). Yeast two-hybrid screening or bimolecular fluorescence complementation may also be performed to test the protein-protein interaction between truncated CENPC proteins and MIS12

complexes in *Arabidopsis*. By this means, we will be able to characterize each subdomain of the CENPC protein and compare them with animal CENPC.

After the minimal protein binding domain is identified, it can be used to pull down the complete CENPC-interacting complex. Components of the complex will be detected and analyzed by Mass Spectrometry. Full length CENPC proteins will also be used for pull down assay and Mass Spectrometry as a positive control. Using this method, we may identify additional components of the constitutive centromere-associated network in plants (Foltz *et al.* 2006, Hori *et al.* 2008).

In the meanwhile, a functional analysis of maize CENPC will be performed using a transient expression assay (Du *et al.* 2010). Truncated maize CENPC constructs fused with fluorescent proteins will be delivered into maize embryogenic cultures by particle bombardment. A minimal centromere targeting domain can be identified by comparing localization patterns of different truncations. Then we will design a new full length CENPC construct with its centromere targeting domain replaced with an alternate DNA binding domain (Gal4, LexA, LacI, or TetR). Fluorescent protein tags will be fused to the N-terminal end of this new construct and expressed in maize plants containing the ABS binding arrays. Immunofluorescence will be performed to detect whether ectopic kinetochores can be assembled on the chromosome arms by the tethering approach (Gascoigne *et al.* 2011).

Identification and Characterization of Plant CENH3 Chaperones

The fact that targeting human CENH3 chaperone to chromosome arms is sufficient to recruit CENH3 ectopically and assemble nearly complete kinetochores prompted us to identify

plant CENH3 chaperones. CENH3 chaperones have been found in yeast (Scm3), drosophila (CAL1) and human (HJURP) (Mizuguchi *et al.* 2007, Dunleavy *et al.* 2009, Foltz *et al.* 2009, Mellone *et al.* 2011). Although functionally conserved, little homology can be detected between these known CENH3 chaperones. For example, drosophila CENH3 chaperone CAL1 is only found within Dipterans (Mellone *et al.* 2011). Little is known about how such non-homologous proteins perform analogous functions in these organisms.

We propose to identify plant homologs of known CENH3 chaperones using bioinformatic tools such as HHpred (Soding *et al.* 2005) and HMMer (Finn *et al.* 2011), which are more sensitive in identifying remote homologs. Similar approaches have been used to show that CENH3 chaperones in human (HJURP) and yeast (Scm3) share a common ancestry (Sanchez-Pulido *et al.* 2009). Besides sequence-based homolog search, we will also identify functional homologs in the CENPC-interacting complex (or more likely, the CENH3-interacting complex) pulled down with biotin tags. Homologs identified by these two approaches will be verified by analyzing their localization pattern. CENH3 chaperones will be transiently localized at centromeres coinciding with CENH3 deposition (Mizuguchi *et al.* 2007, Dunleavy *et al.* 2009, Foltz *et al.* 2009, Mellone *et al.* 2011).

Once the localization and function of the CENH3 chaperone in plants is confirmed, it will be fused with a DNA binding domain (Gal4, LexA, LacI, or TetR) and a fluorescent protein and expressed in the maize plants containing the ABS binding arrays. We will test whether CENH3 can be recruited to the ABS loci and whether functional neocentromeres can be established ectopically (Barnhart *et al.* 2011). Plants with neocentromere activities will be crossed with wild types to remove the fusion proteins. Progenies from this cross will be assessed to see if neocentromere activities are meiotically maintained at the ABS loci.

Inducing Acentric Fragment Formation for Neocentromere Establishment

Dicentric chromosomes are rare in plants, and even when they form, one of the two centromeres is usually inactivated shortly thereafter (Han *et al.* 2006, Gao *et al.* 2011). New centromere formation is more likely to occur under conditions where chromosomes have otherwise lost their original kinetochores (Nasuda *et al.* 2005, Topp *et al.* 2009). To mimic the natural conditions of neocentromere formation, we artificially induced chromosome breakages in maize using plants that bear a paracentric inversion (This experiment was not included in the dissertation, due to a lack of data. Nonetheless the figure was shown in this chapter and related materials will be used for future studies.)

Two different inversion lines (Inv3c, inverted at 3L.09 to 3L.81, and Inv4h, inverted at 4L.16 to 4L.56, stocks from maizeGDB (http://www.maizegdb.org/)) were crossed with plants homozygous for ABS3 or ABS4, respectively. In the progenies of this cross, we found that anaphase bridges were formed at anaphase I. Hybridization of the anaphase chromosomes with an ABS specific probe revealed that the ABS binding arrays were located on the anaphase bridge (Figure 5.2). These observations indicate that acentric fragments containing ABS binding arrays can be released from their original loci. In the subsequent cell divisions, these acentric fragments cannot bind to the microtubules and therefore will not be included in the daughter nuclei.

The inversion lines will be used as a means to test neocentromere formation on acentric fragments. Transgenic plants expressing tethered CENPC protein-binding domain or tethered CENH3 chaperone in plants homozygous for ABS will be crossed to the inversion lines. Meiocytes of the progenies from the cross will be examined. If de novo kinetochores can be

assembled at the ABS loci, acentric fragments caused by inversion and the resulting anaphase bridge will be stabilized and will segregate to one of the spindle poles. Recovery of the otherwise acentric fragments containing ABS arrays in the subsequent mitotic segregation will provide evidence of neocentromere establishment and maintenance.

In conclusion, the construction of plant artificial centromere will not be simple or straightforward, but with the biological and technical advances developed in this study, we believe that we are at least one step closer towards our ultimate goal. By combining gene discovery with the synthetic repeat arrays system, and additional maize resources such as inversion lines, it should be possible to assemble de novo kinetochores on defined sequences. With breakthroughs in creating functional artificial centromeres and parallel progress in gene synthesis, it is reasonable to believe that one day plant artificial chromosomes can be achieved.

References

Barnhart, M.C., Kuich, P.H., Stellfox, M.E., Ward, J.A., Bassett, E.A., Black, B.E. and Foltz, D.R. (2011) HJURP is a CENP-A chromatin assembly factor sufficient to form a functional de novo kinetochore. *J Cell Biol*, **194**, 229-243.

Brown, M.T. (1995) Sequence similarities between the yeast chromosome segregation protein Mif2 and the mammalian centromere protein CENP-C. *Gene*, **160**, 111-116.

Carroll, C.W., Milks, K.J. and Straight, A.F. (2010) Dual recognition of CENP-A nucleosomes is required for centromere assembly. *J Cell Biol*, **189**, 1143-1155.

Du, Y., Topp, C.N. and Dawe, R.K. (2010) DNA binding of centromere protein C (CENPC) is stabilized by single-stranded RNA. *PLoS Genet*, **6**, e1000835.

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

Finn, R.D., Clements, J. and Eddy, S.R. (2011) HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res*, **39**, W29-37.

Foltz, D.R., Jansen, L.E., Bailey, A.O., Yates, J.R., 3rd, Bassett, E.A., Wood, S., . . . Cleveland, D.W. (2009) Centromere-specific assembly of CENP-a nucleosomes is mediated by HJURP. *Cell*, **137**, 472-484.

Foltz, D.R., Jansen, L.E., Black, B.E., Bailey, A.O., Yates, J.R., 3rd and Cleveland, D.W. (2006) The human CENP-A centromeric nucleosome-associated complex. *Nat Cell Biol*, **8**, 458-469.

Furuyama, T. and Henikoff, S. (2006) Biotin-tag affinity purification of a centromeric nucleosome assembly complex. *Cell Cycle*, **5**, 1269-1274.

Gao, Z., Fu, S., Dong, Q., Han, F. and Birchler, J.A. (2011) Inactivation of a centromere during the formation of a translocation in maize. *Chromosome Res*, **19**, 755-761.

Gascoigne, K.E., Takeuchi, K., Suzuki, A., Hori, T., Fukagawa, T. and Cheeseman, I.M. (2011) Induced Ectopic Kinetochore Assembly Bypasses the Requirement for CENP-A Nucleosomes. *Cell*, **145**, 410-422.

Han, F., Lamb, J.C. and Birchler, J.A. (2006) High frequency of centromere inactivation resulting in stable dicentric chromosomes of maize. *Proc Natl Acad Sci U S A*, **103**, 3238-3243.

Heeger, S., Leismann, O., Schittenhelm, R., Schraidt, O., Heidmann, S. and Lehner, C.F. (2005) Genetic interactions of separase regulatory subunits reveal the diverged Drosophila Cenp-C homolog. *Genes Dev*, **19**, 2041-2053.

Hori, T., Amano, M., Suzuki, A., Backer, C.B., Welburn, J.P., Dong, Y., . . . Fukagawa, T. (2008) CCAN makes multiple contacts with centromeric DNA to provide distinct pathways to the outer kinetochore. *Cell*, **135**, 1039-1052.

Mellone, B.G., Grive, K.J., Shteyn, V., Bowers, S.R., Oderberg, I. and Karpen, G.H. (2011) Assembly of Drosophila centromeric chromatin proteins during mitosis. *PLoS Genet*, **7**, e1002068.

Milks, K.J., Moree, B. and Straight, A.F. (2009) Dissection of CENP-C-directed centromere and kinetochore assembly. *Mol Biol Cell*, **20**, 4246-4255.

Mizuguchi, G., Xiao, H., Wisniewski, J., Smith, M.M. and Wu, C. (2007) Nonhistone Scm3 and histones CenH3-H4 assemble the core of centromere-specific nucleosomes. *Cell*, **129**, 1153-1164.

Moree, B., Meyer, C.B., Fuller, C.J. and Straight, A.F. (2011) CENP-C recruits M18BP1 to centromeres to promote CENP-A chromatin assembly. *J Cell Biol*, **194**, 855-871.

Nasuda, S., Hudakova, S., Schubert, I., Houben, A. and Endo, T.R. (2005) Stable barley chromosomes without centromeric repeats. *Proc Natl Acad Sci U S A*, **102**, 9842-9847.

Przewloka, M.R., Venkei, Z., Bolanos-Garcia, V.M., Debski, J., Dadlez, M. and Glover, D.M. (2011) CENP-C is a structural platform for kinetochore assembly. *Curr Biol*, **21**, 399-405.

Sanchez-Pulido, L., Pidoux, A.L., Ponting, C.P. and Allshire, R.C. (2009) Common ancestry of the CENP-A chaperones Scm3 and HJURP. *Cell*, **137**, 1173-1174.

Screpanti, E., De Antoni, A., Alushin, G.M., Petrovic, A., Melis, T., Nogales, E. and Musacchio, A. (2011) Direct binding of Cenp-C to the Mis12 complex joins the inner and outer kinetochore. *Curr Biol*, **21**, 391-398.

Soding, J., Biegert, A. and Lupas, A.N. (2005) The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res*, **33**, W244-248.

Song, K., Gronemeyer, B., Lu, W., Eugster, E. and Tomkiel, J.E. (2002) Mutational analysis of the central centromere targeting domain of human centromere protein C, (CENP-C). *Exp Cell Res*, **275**, 81-91.

Talbert, P.B., Bryson, T.D. and Henikoff, S. (2004) Adaptive evolution of centromere proteins in plants and animals. *J Biol*, **3**, 18.

Topp, C.N., Okagaki, R.J., Melo, J.R., Kynast, R.G., Phillips, R.L. and Dawe, R.K. (2009) Identification of a maize neocentromere in an oat-maize addition line. *Cytogenet Genome Res*, **124**, 228-238.

Trazzi, S., Bernardoni, R., Diolaiti, D., Politi, V., Earnshaw, W.C., Perini, G. and Della Valle, G. (2002) In vivo functional dissection of human inner kinetochore protein CENP-C. *J Struct Biol*, **140**, 39-48.

Trazzi, S., Perini, G., Bernardoni, R., Zoli, M., Reese, J.C., Musacchio, A. and Della Valle, G. (2009) The C-terminal domain of CENP-C displays multiple and critical functions for mammalian centromere formation. *PLoS One*, **4**, e5832.

Yang, C.H., Tomkiel, J., Saitoh, H., Johnson, D.H. and Earnshaw, W.C. (1996) Identification of overlapping DNA-binding and centromere-targeting domains in the human kinetochore protein CENP-C. *Mol Cell Biol*, **16**, 3576-3586.

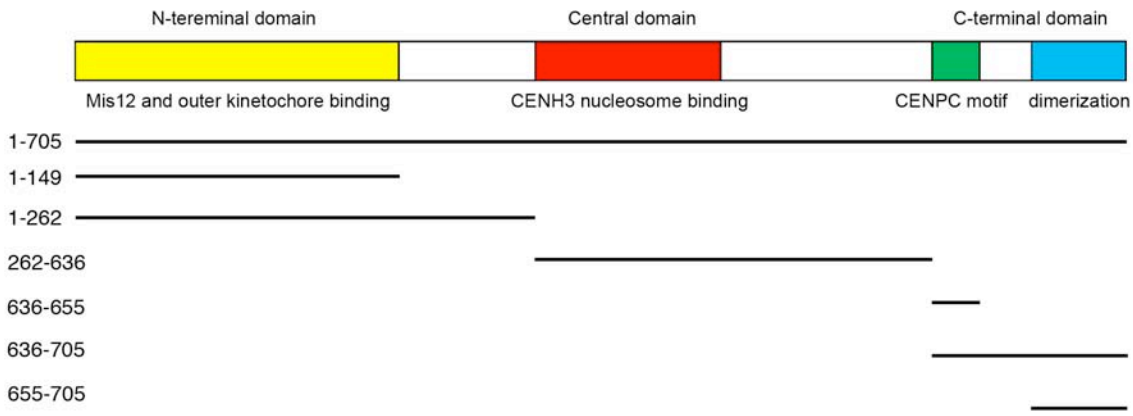


Figure 5.1 Diagram of different truncated constructs of CENPC.

In animals, CENPC is composed of three domains: an N-terminal domain, a central domain and a C-terminal domain. The C-terminal domain is further divided into two domains, a highly conserved CENPC motif of 23 amino acids, and a dimerization domain. The CENPC homolog in *Arabidopsis* has 705 amino acids and an overall structure very similar to the animal CENPC. Each domain are identified based the following criteria: Full length (1-705 aa), N-terminal domain (1-149 aa, exon 1-4, based on homology with maize CENH3 at the N-terminal end), N-terminus plus (1-262 aa, exon 1-6, based on homology with other plant CENH3, (Talbert *et al.* 2004)), Central domain (262-636 aa), CENPC motif (636-655 aa, based on homology with animal and yeast CENPC), C-terminal domain (636-705 aa), Dimerization domain (655-705 aa).

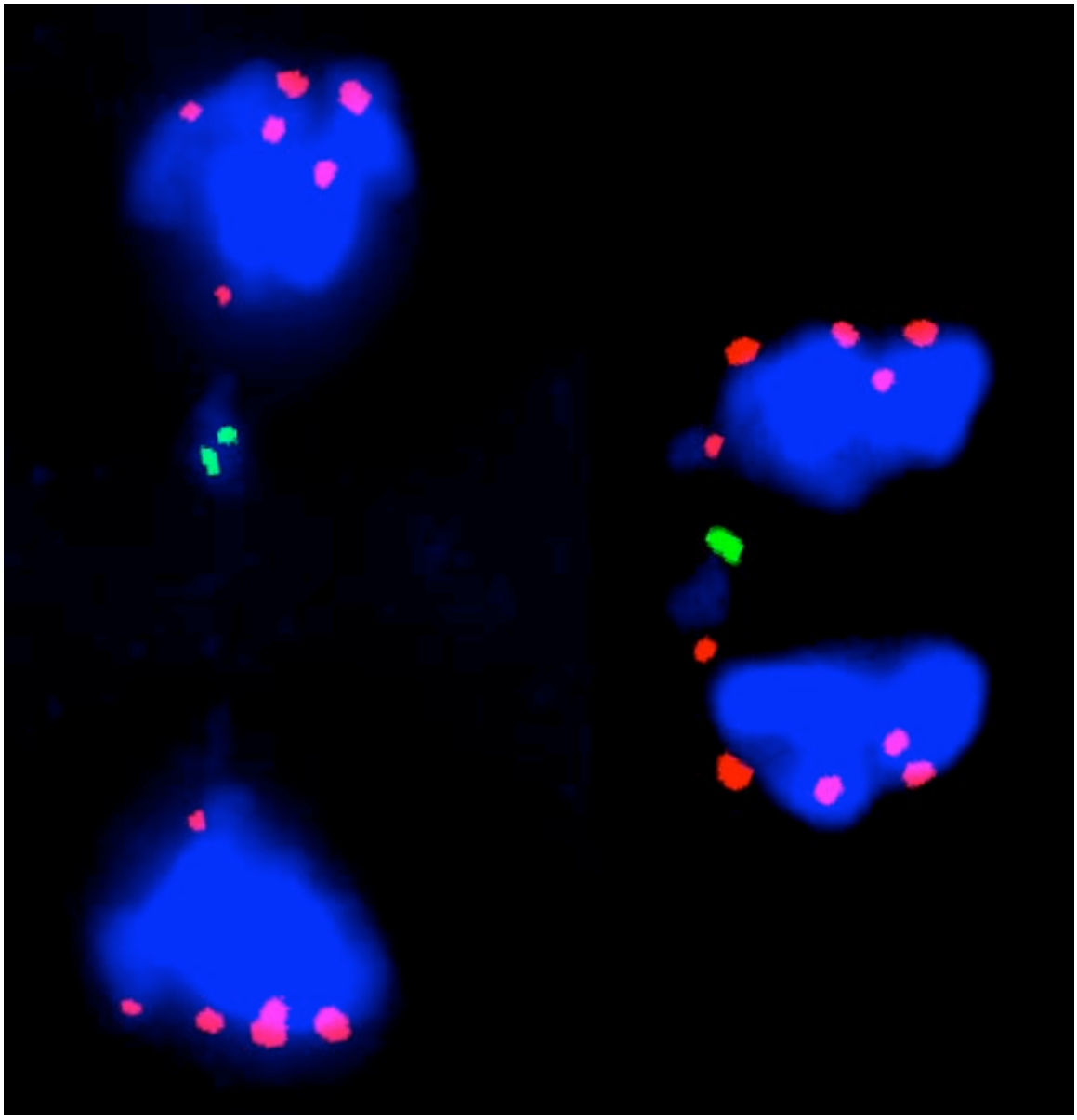


Figure 5.2 ABS binding arrays located on anaphase bridges.

Maize chromosomes at anaphase I were stained with CENPC antibodies (red) and hybridized with ABS-specific probes (green). Notice that ABS arrays were localized on the anaphase bridges.