ESTABLISHMENT AND MAINTENANCE OF PLANT CENTROMERE IDENTITY

by

CHRISTOPHER NEAL TOPP

(Under the Direction of R. Kelly Dawe)

ABSTRACT

Centromeres pose a special challenge to our understanding of the genome. Many basic functions of the cell are maintained by deoxyribonucleic acid (DNA) sequences that change very little over time. However centromere regions evolve rapidly; in fact no common sequence identifies centromeres of various species. Instead, repetitive DNA somehow propagates centromeres in one place over millions of years. The key to centromere stability is Centromere Histone H3 (CenH3). CenH3 is a variant type of histone that packages repetitive DNA into a functional centromere structure. How CenH3 is targeted to, and maintained at, ever-changing DNA sequences is a fundamental question of genome biology. Several approaches were used to understand how CenH3 is regulated in plant cells. We discovered that both strands of Zea mays (maize) centromeres are transcribed into ribonucleic acid (RNA), and that the transcripts are associated with CenH3. Usually, transcripts from regions of the genome that are transcribed from both strands hybridize and are cleaved into small interfering RNAs. In contrast, centromere RNAs remain single-stranded, suggesting a unique role in centromere maintenance and function. CenH3 maintenance appears to be very dynamic at newly initiated centromeres. We identified and characterized a ‘neocentromere’ that formed on a broken maize chromosome maintained in an oat genome. The maize chromosome was initially unstable during mitosis, suggesting a defect in centromere function. Although mitotic stability returned to the maize chromosome in progeny, its CenH3 domain was rapidly changing size among cells. These results point to CenH3-mediated instability in a newly established centromere. We suggest that the ultimate survival of a new centromere depends on its position and its mode of formation. CenH3 instability can conceivably be suppressed through experimental means.
We overexpressed the *Arabidopsis thaliana* CenH3 gene, *HTR12*, in an effort to induce centromere formation. Transgenic plants developed abnormally, and we traced these defects to chromosome mis-segregation and cell death, consistent with the formation of multiple centromeres. The evidence presented here further defines the identity of centromeres, and processes that affect their establishment.

**INDEX WORDS:** CenH3, CenH3 maintenance, CenH3 establishment, HTR12, centromere, neocentromere, epigenetic, histone variant, maize, *Zea mays*, *Arabidopsis thaliana*, centromere RNA, mis-segregation
ESTABLISHMENT AND MAINTENANCE OF PLANT CENTROMERE IDENTITY

by

CHRISTOPHER NEAL TOPP

Bachelor of Science, University of Georgia, 2000

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

2009
ESTABLISHMENT AND MAINTENANCE OF PLANT CENTROMERE IDENTITY

by

CHRISTOPHER NEAL TOPP

Major Professor: R. Kelly Dawe
Committee: Claiborne Glover
Richard Meagher
Michael Terns
Susan Wessler

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
May 2009
ACKNOWLEDGEMENTS

Of the very many that helped me realize my potential as a graduate student, none was more important than my advisor Kelly Dawe. He continues to influence me greatly as a scientist and as a person – Kelly is a true mentor in every sense of the term. I further attribute much of my scientific growth to my committee members Claiborne Glover, Rich Meagher, Michael Terns, and Sue Wessler, who continuously challenged me as a group, but provided constant support as individuals. In particular Rich Meagher and his lab were exceedingly generous with their time and resources, and set a model example for my own lab someday. The resources of the National Science Foundation, the University of Georgia Graduate School, and the UGA Plant Biology Department were also instrumental in every aspect of my research and training. Finally, I acknowledge my friends and family, who provided unending support throughout this journey. To my parents Carla and Greg, and my sister Gabriele, I would not have achieved this goal alone, thank you.
# TABLE OF CONTENTS

ACKNOWLEDGEMENTS ........................................................................................................ iv

LIST OF TABLES .................................................................................................................. vi

LIST OF FIGURES ............................................................................................................... vii

CHAPTER

I. Introduction and Literature Review ......................................................................................... 1
   Centromere-Kinetochoore Structure and Dynamics ............................................................. 1
   Centromere Cohesion ........................................................................................................... 3
   Centromere Identity ............................................................................................................. 5
   References ......................................................................................................................... 10

II. Centromere-Encoded RNAs are Integral Components of the Maize Kinetochore ............. 18
   References ......................................................................................................................... 31

III. Identification of a Maize Neocentromere in an Oat-Maize Addition Line ......................... 42
   References ......................................................................................................................... 54

IV. Overexpression of HTR12, the *Arabidopsis thaliana* CenH3, is Differentially Regulated on a
   Cell-to-Cell Basis, and Results in Chromosome Mis-Segregation and Cell Death ............. 70
   References ......................................................................................................................... 80

V. Summary and Conclusions .................................................................................................... 93
   References ......................................................................................................................... 101
## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 2.1: Association of RNA with CENH3</td>
<td>.................................................................</td>
<td>41</td>
</tr>
<tr>
<td>Table 3.1: Stability of Neocentromere Lines</td>
<td>........................................................................</td>
<td>68</td>
</tr>
<tr>
<td>Table 3.2: Relative Neocentromere Intensities</td>
<td>........................................................................</td>
<td>69</td>
</tr>
<tr>
<td>Table 4.1: Frequencies of Chromosome Mis-Division and Mininuclei</td>
<td>..........................................................</td>
<td>92</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

Page

Figure 2.1: CRM is Actively Transcribed................................................................. 36
Figure 2.2: Centromere-Encoded RNAs are Co-Immunoprecipitated with CENH3.................. 37
Figure 2.3: Centromeric siRNAs are not Detected Following CENH3-mediated ChIP .................. 39
Figure 3.1: Neocentromere Material Does Not Contain Maize Centromere DNA or SSR Marker DNA from the Maize Chromosome 3 Long Arm................................................................. 60
Figure 3.2: Maize Neocentromere Lines Lack Centromere DNA but are Stably Inherited............... 61
Figure 3.3: Neocentromere-Containing Chromosomes are Distinguished from Oat Chromosomes ....... 62
Figure 3.4: Identification of the Oat CENH3 Protein ...................................................... 64
Figure 3.5: The Maize Neocentromere Formed Submetacentrically ......................................... 65
Figure 3.6: CENH3 Stainings of Maize Neocentromeres are More Variable than Those of Established Oat Centromeres ........................................................................................................ 66
Figure 3.7: A Speculative Model of Neocentromere Establishment............................................ 67
Figure 4.1: HTR12 Overexpression Causes Multiple Developmental Abnormalities .................. 84
Figure 4.2: The Level of HTR12 Overexpression Correlates Weakly with the Phenotypic Severity of Transgenic Plants ........................................................................................................ 86
Figure 4.3: HTR12-YFP Protein Accumulates Differently on a Cell-to-Cell Basis ....................... 87
Figure 4.4: HTR12 Overexpression Causes Cell Division Defects Resulting in Mininuclei ............... 89
Figure 4.5: Plants Overexpressing HTR12 Exhibit Cell Death in Developing Floral Organs .......... 91
CHAPTER I:  
INTRODUCTION AND LITERATURE REVIEW

CENTROMERE-KINETOCHORE STRUCTURE AND DYNAMICS

A centromere is a region on each chromosome that ensures the precise segregation of replicated chromosomes during cell division. The primary role of centromeres is to provide the foundation for kinetochores, which are large multi-protein structures. Kinetochores link chromosomes to the microtubule bundles that pull each replicated chromosome to its respective pole during anaphase. Centromeres also prevent premature separation of replicated chromosomes by maintaining cohesion until the regulatory signals that promote anaphase are received. A slight variation of centromere-mediated cohesion directs the specialized meiotic cell division, during which homologous chromosomes recombine and segregate to form novel genotypes. Thus, centromeres underlie processes ranging from basic propagation of genetic material to promotion of genetic diversity.

The centromere-kinetochore complex is best understood in the budding yeast *Saccharomyces cerevisiae* (McAinsh et al., 2003; Guenatri et al., 2004; Pal-Bhadra et al., 2004). Genetic, molecular, and cytological analyses categorize this structure into three hierarchical levels (De Wulf et al., 2003; Westermann et al., 2003; Wei et al., 2005; Westermann et al., 2005; Meraldi et al., 2006). The first level consists of DNA sequences bound to a set of constitutive proteins (Westermann et al., 2003). Collectively these elements fashion a unique type of centromere chromatin that supplies the foundation for a larger set of constitutive proteins extending away from the face of the chromosome (Cleveland et al., 2003). The second level links chromosomes to microtubules directly through the NDC80 complex, or through interactions with a third level that assembles during cell division (De Wulf et al., 2003; Miranda et al., 2005). A mature budding yeast centromere-kinetochore complex is comprised of at least 60 proteins, with
the inner side attached to the chromosome and the outer side attached to microtubules (De Wulf et al., 2003). Regulatory crosstalk through the centromere-kinetochore coordinates microtubule binding with sister chromatid cohesion (McAinsh et al., 2003). When the proper conditions for chromosome separation are met, the centromere-kinetochore complex couples motor proteins with forces generated by shrinking microtubules to drive chromosomes poleward (Cleveland et al., 2003; McIntosh et al., 2008).

Centromeres of the budding yeast lineage bind only one microtubule per chromosome. Such ‘point’ centromeres can be considered archetypes of the regional centromeres in plants and animals that bind dozens of microtubules per chromosome (Zinkowski et al., 1991; Joglekar et al., 2008). For instance, induced fragmentation of regional centromeres results in multiple kinetochores (Brinkley et al., 1988). The suggestion that regional centromeres are functionally redundant point centromeres is also supported by the wide conservation of linker layer proteins. The linker layer of S. cerevisiae is arranged into four subgroups, MIND, COMA, NDC80, and SPC105 that direct particular aspects of kinetochore structure, and functional homologs have been found in many species with regional centromeres (De Wulf et al., 2003; McAinsh et al., 2003; Meraldi et al., 2006). Although many of the S. cerevisiae genes critical to cell cycle control such as MAD2, BUB1, and Aurora kinase are also conserved, the essential microtubule binding complex DASH and several kinesin motor proteins are not (Cleveland et al., 2003; Meraldi et al., 2006). In their place stand a poorly understood network of proteins that unify regional kinetochores into a coherent ‘binding-face’ (Pidoux and Allshire, 2003). Apparently new structures were needed to harness multiple microtubule interactions as centromeres grew in size. The foundations of regional centromeres are similarly reorganized (Blower et al., 2002; Cheeseman and Desai, 2008). The CBF complex is required in S. cerevisiae to bind a short, stringently conserved DNA motif that directs the localization of a nucleosome with a centromere-specific histone. Taxa with regional centromeres have neither the CBF complex, nor any common DNA elements. Such large centromeres have instead substituted strict binding motifs for arrays of repetitive elements whose sequences are not essential for centromere function (Allshire and Karpen, 2008). The only conserved elements of the budding yeast centromere foundation are the centromere-specific histone Cse4p and its structural counterpart Mif2p (Westermann et al., 2003;
Meraldi et al., 2006). Known as CenH3s and CENPCs, respectively, these protein classes are conserved across eukaryotes, where they function to coordinate the assembly of all other centromere and kinetochore proteins.

Remarkably, the DNA sequences associated with CenH3 and CENPC in regional centromeres are the most rapidly evolving domains of eukaryotic genomes (Henikoff et al., 2001). The sequence constraints on CenH3 and CENPC binding are so weak that these proteins can spontaneously establish neocentromeres at previously non-centromeric regions of the genome (Marshall et al., 2008). However, this occurrence is rare, and inherited neocentromeres are eventually restructured to include long repetitive arrays, suggesting the function of centromere DNA may simply be to provide a high concentration of stable binding sites for CenH3 and CENPC (Malik and Henikoff, 2001; Malik and Henikoff, 2002; Lomiento et al., 2008). In this view, epigenetic processes govern centromeres by controlling the position of centromere chromatin.

CENTROMERE COHESION

Centromeres are usually under strong pressure to stay in one place because their position is important to the proper pairing and separation of chromosomes (Kemp et al., 2004). Both of these processes depend on cohesion between two aligned centromeres. During meiosis, sister centromeres initially pair to present a unified face for microtubule binding (Petronczki et al., 2006; Riedel et al., 2006). This structure is necessary to create haploid cells in meiosis I, because it ensures homologous, not sister, chromosome segregation. In contrast, mitotic sister centromeres are paired back-to-back (Watanabe, 2005). This orientation ensures that sister kinetochores attach to microtubules from opposite ends of the spindle. Once these bipolar attachments are made, centromeric cohesion prevents the spindle from prematurely pulling sisters apart. The tension formed between the spindle and cohesion of sister centromeres relays a phosphorylation signal that triggers chromosome separation by the action of the anaphase promoting complex (APC). The APC in turn sets in motion a series of events that lead to the destruction of cohesion and the initiation of anaphase (Cleveland et al., 2003).
Cells use ring-shaped cohesin proteins to regulate chromosome cohesion. These four-molecule complexes contain two members of the highly conserved SMC family of proteins, SMC1 and SMC3, which heterodimerize to form a V-shaped backbone. There are several forms of the two non-SMC members, which allow cohesins to have distinct functions (Losada and Hirano, 2005). For example, Rec8 replaces the mitotic functioning subunit Rcc3 during meiosis to ensure sister centromeres do not prematurely separate (Watanabe and Nurse, 1999). However, cohesins also function at loci other than centromeres and are initially installed along the entire length of newly replicated chromosomes (Watanabe, 2005). In a poorly understood process, cohesin molecules are removed from chromosome arms and accumulate at centromeres in prophase.

Until recently, the centromeric reorganization of cohesins was attributed to the deeply-staining ‘heterochromatin’ that surrounds centromeres in cytological preparations. Pericentric heterochromatin is commonly methylated on lysine 9 of histone H3, either twice (H3K9me2) or three times (H3K9me3). Either of these modifications creates a binding site for the widely conserved heterochromatin protein 1 (HP1), which induces a compact chromatin structure (Maison and Almouzni, 2004). Chromosome segregation of the fission yeast *Schizosaccharomyces pombe*, in which heterochromatin is found predominantly at pericentromeres is severely disrupted by mutations that diminish heterochromatin (Bernard et al., 2001; Nonaka et al., 2002). However, it is not clear how plants and animals, with large islands of heterochromatin on chromosome arms, could regulate cohesion in the same way. Several studies of human, mouse, and insect cells deficient in pericentric heterochromatin reported normal cell division (Pal-Bhadra et al., 2004; Kanellopoulou et al., 2005; Koch et al., 2008), whereas other studies using similarly mutated chicken and human cells reported chromosome dysfunction (Fukagawa et al., 2004; Yamagishi et al., 2008). The plant homolog of HP1 does not localize to pericentromeres, and no mutation that disrupts plant pericentromere structure has been reported to affect cell division (May et al., 2005; Onodera et al., 2005; Fransz et al., 2006). At most, we can conclude that heterochromatin may support, but is not essential for, centromere cohesion in plants and animals (Topp and Dawe, 2006).
This conclusion is supported by the existence of stably inherited new centromeres, which can apparently initiate cohesion de novo. In fact, several proteins that regulate cohesins are recruited directly to centromeres, most notably Aurora B kinase (Meluh and Strunnikov, 2002). Aurora B is a key cell-cycle regulator important to cohesion and tension sensing across sister kinetochores (Liu et al., 2009). Its localization to the inner centromere of humans during prophase requires the phosphorylation of CenH3, which is itself a target (Kunitoku et al., 2003). Aurora B also phosphorylates histone H3 on serine 10 (H3S10ph) in a pattern that is initially restricted to pericentric regions, presumably by phosphorylation of the centromere core (Topp and Dawe, 2006). In many plant species, H3S10ph and a similar mark, H3S28ph, are found precisely between kinetochores under tension and appear to delimit the cohesive domain of pericentromeres (Kaszas and Cande, 2000; Gernand et al., 2003; Shibata and Murata, 2004; Zhang et al., 2005). Both H3S10ph and H3S28ph are present on the flanks of a stable barley neocentromere, suggesting that phosphorylation of the surrounding chromatin is important to establishing a functional centromere (Nasuda et al., 2005). Accordingly, maize CenH3 is also phosphorylated during prophase and has been implicated to begin a wave of H3S28ph that defines cohesion (Zhang et al., 2005).

Centromeres operate in separate domains with interdependent functions. Kinetochores require tension from cohesion for microtubule attachment, and pericentromeric cohesion is regulated by signals from kinetochores. If centromere chromatin can organize both domains a priori, then these functions are necessarily maintained together. Centromere integrity at established and newly initiated loci thus rests completely on the regulation of CenH3, the foundation of centromere chromatin.

CENTROMERE IDENTITY

The first CenH3 to be discovered was CENP-A, one of three centromere antigens recognized by the serum of a patient who suffered from the human autoimmune disease CREST scleroderma (Earnshaw and Rothfield, 1985). Medical researchers soon identified the DNA sequences associated with centromeric nucleosomes through their biochemical association with CREST antigens (Masumoto et al., 1989b; Masumoto et al., 1989a). Human centromere DNA is organized into short repetitive tandem arrays
named alpha satellites that stretch megabases (Vafa and Sullivan, 1997; Amor et al., 2004a). Each repeat is comprised of 160 nucleotides, roughly the length of DNA that wraps around one nucleosome to form the basic level of chromatin. Justification of the alpha-satellite as a functional element was made by experiments in which CENP-A was recruited to transfected alpha repeats, which form stably inherited centromeres de novo (Larin et al., 1994; Masumoto et al., 1998). Centromere function seemed to be specified by the interaction of a conserved DNA element and a structural protein, akin to other essential processes such as telomere maintenance. However, this view was questioned by reports of ‘dicentric’ chromosomes containing two alpha-satellite loci separated by non-centromere DNA (Fisher et al., 1997). On these chromosomes, one DNA locus was rendered functional by recruiting CENP-A, whereas the other locus was nonfunctional and did not attract CENP-A. Later research would demonstrate the capability of some dicentric chromosomes to switch active and repressed loci multiple times in clonal cell lineages, reinforcing the view that centromere chromatin can be rapidly established and removed (Higgins et al., 2005). Furthermore, the discovery of ‘neocentromeres’, in which stably inherited chromosomes recruit CenH3 upon loss of alpha satellite DNA, further undermined a genetic requirement for centromeres (Vouillaire et al., 1993; Ohashi et al., 1994). Research to date has identified CenH3 as present at every functional centromere, and as absent from every nonfunctional centromere. Therefore CenH3 is the biochemical keystone of centromere identity. How dynamic epigenetic mechanisms propagate stable centromere location over millions of years remains a fundamental question.

At the heart of this question lies the regulation of the CenH3 histone. Centromeres face a special problem because, unlike standard histones, CenH3 deposition occurs independently of DNA replication (Shelby et al., 2000; Smith, 2002). This feature effectively dilutes centromere chromatin in S-phase and would eventually erase the epigenetic identity of centromeres. The process of CenH3 replenishment remains obscure. Most genomes encode several other histone variants, such as the widely conserved H3.3 and H2AZ, which are incorporated independently of replication (Deal et al., 2007). However these variants depend on conserved nucleosome remodeling complexes to mediate their exchange with standard histones. Although general histone binding proteins are known to participate in centromere histone
assembly (Henikoff and Ahmad, 2005; Furuyama et al., 2006), the lack of similarity among CenH3s makes a conserved targeting mechanism difficult to envision. Instead, physical attributes of CenH3s may be the key to their centromere localization.

Phylogenetic analyses have identified CenH3 genes in a number of organisms using two criteria: an abnormally long histone fold domain (HFD) lacking a specific glutamine conserved in H3s, and a unique N-terminus. The N-terminal tails of CenH3 members range from 20 to 200 amino acids and share little identity among themselves or with H3 (H3.3 has one change in this region) (Malik and Henikoff, 2003). This commonality precludes CenH3s from the epigenetic regulation that occurs on bulk chromatin. Although N-termini appear to be dispensable to CenH3 localization, the changes to the HFD are vital. Chimeric H3 molecules containing human CenH3 HFDs are centromere localized, but not vice-versa (Shelby et al., 1997; Henikoff and Ahmad, 2005). Biophysical evidence suggests CenH3 HFDs are especially compact and bend DNA around the nucleosome differently (Black et al., 2004; Bloom et al., 2006). In support of an unusual nucleosome structure, biochemical data from *Drosophila* cells lead to the argument that fly CenH3 nucleosomes assemble with only four histones, as opposed to the octameric form conserved in standard chromatin (Dalal et al., 2007). A hexameric CenH3 nucleosome has been argued for budding and fission yeast (Pidoux et al., 2009). As a group centromeric nucleosomes are united by a unique structure that affects their association with DNA, and by their divergent N-termini, which prevent typical epigenetic regulation.

Centromere identity factors support the maintenance of centromeric chromatin. One explanation for the lack of a strong genetic component to centromere identity is that CenH3s can be widely incorporated into chromatin, but are retained only at centromere loci. Consistent with this view, vast overexpression of CenH3s result in their eventual spread outside of centromeres (Van Hooser et al., 2001; Heun et al., 2006; Collins et al., 2007). As demonstrated in budding yeast and *Drosophila*, directed proteolysis normally suppresses ectopic CenH3 (Collins et al., 2004; Moreno-Moreno et al., 2006), but why aren’t established centromeres similarly suppressed? Since centromere position is not genetically marked, then centromere maintenance must depend on epigenetic marks set by centromere chromatin. In chapter two of this
manuscript, we investigate the role that RNA transcripts from the centromere play in marking established centromere regions.

Counterintuitively, heterochromatin may also be a centromere identity factor. Although pericentric heterochromatin forms as a consequence of the repression of transposons and other repetitive elements that reside there (Topp and Dawe, 2006), evidence suggests that centromeres need heterochromatin to stably maintain CenH3. Under normal circumstances, centromeres will form on naked centromere DNA transfected into animal cells. These constructs are then stably transmitted as human or mammalian artificial chromosomes (HACs or MACs). The formation rates and stability of HACs are reduced by mutations that promote or ameliorate heterochromatin formation, suggesting that a natural balance between CenH3 and heterochromatin is important (Nakano et al., 2008). More specifically, human artificial chromosomes form efficiently only when flanked by heterochromatic marks on only one side of the centromere array (Nakashima et al., 2005). Native centromeres appear to be under similar constraints. In *S. pombe*, transcription of a tRNA gene on the boundary of the natural centromere/pericentromere is essential to prevent pericentric heterochromatin from encroaching on CenH3 chromatin and compromising the centromere (Scott et al., 2006). Clearly the antagonistic boundary established between CenH3 chromatin and modified histone H3 is critical to centromere identity, and may be of special importance to the formation of new centromeres. A recent genome–wide study in *S. pombe* found that upon deletion of native centromeres, neocentromeres were recovered almost exclusively at heterochromatic regions, but rarely in RNAi mutants deficient in heterochromatin (Ishii et al., 2008). Since CenH3 can theoretically bind anywhere on the chromosome, these results suggest that heterochromatin can promote centromere identity through stabilization of CenH3.

The processes that govern neocentromere formation are keys to understanding centromere evolution. Nearly all of the ~100 known instances of neocentromere formation in humans are associated with diseases stemming from partial or full aneuploidies that occur after chromosome breakages eliminate centromere DNA (Marshall et al., 2008). In these cases it is difficult to see how a neocentromere would be beneficial in an evolutionary context. However, in one recent patient with only slight abnormalities, a
neocentromere was inherited despite full retention of the native centromere DNA and no identifiable chromosome rearrangements (Amor et al., 2004b). After millions of years of epigenetic stability, centromeres can apparently move to an unrelated location and begin anew. This striking example linked neocentromere formation to the phenomenon known as ‘centromere repositioning’ (Marshall et al., 2008). Centromere repositioning refers to the observation that centromere positions have changed and become fixed over time in closely related species. Simple translocation of the original centromere can be ruled out in many cases using the synteny of flanking markers (Capozzi et al., 2008). Collective evidence suggests that centromere repositioning begins with a neocentromere seed, which millions of years later gains full genetic identity by acquisition of tandem repeat elements that favor CenH3 binding (Capozzi et al., 2008; Lomiento et al., 2008). Perhaps the evolutionary importance of centromere movement accounts for their odd epigenetic flexibility.

The subsequent content of this dissertation details our efforts to understand the initiation and maintenance of centromere identity in plants. Plants are an ideal group in which to study epigenetic features of centromere function for several reasons. First, plant centromeres encompass an unparalleled range of diversity in sequence and organization of centromere DNA. For example, the satellite DNA content of Zea mays (maize, or corn) centromeres has undergone dramatic fluctuations during their brief period of domestication and cultivation (Kato et al., 2004). However, the major centromere proteins have changed very little, pointing to as yet unidentified factors that maintain centromere stability. Plants also retain a high degree of developmental plasticity, allowing experimental methods that would likely cause lethality in animals. Finally, from a biotechnology standpoint, engineering de novo centromere function is the key element in creating artificial chromosomes that can be used as independently propagating platforms for gene introduction.

Chapter Two is a published manuscript reporting the novel finding that the centromere core DNA of Zea mays is transcribed and that the RNA molecules are physically associated with centromere chromatin. These data have implications for the maintenance of CenH3 at established centromeres. Chapter Three is another published manuscript, characterizing a newly initiated centromere in a hybrid of Avena sativa
(oat) and maize. The behavior of this neocentromere suggests that centromere identity can be fragile at its onset. Chapter Four details research aimed at uncovering the minimum requirements for centromere initiation and creating a robust system to study early initiation events, using the model plant *Arabidopsis thaliana*. We conclude with a summary of the implications of our data in the larger field of centromere biology.

REFERENCES


CHAPTER II:

CENTROMERE-ENCODED RNAS ARE INTEGRAL COMPONENTS OF THE MAIZE KINETOCHORE

---

Reprinted here with permission of the publisher. Copyright 2004 National Academy of Sciences U.S.A.
ABSTRACT

RNA is involved in a variety of chromatin modification events, ranging from large-scale structural rearrangements to subtle local affects. Here we extend the evidence for RNA-chromatin interactions to the centromere core. The data indicate that maize centromeric retrotransposons (CRM) and satellite repeats (CentC) are not only transcribed but that nearly half of the CRM and CentC RNA is tightly bound to Centromeric Histone H3 (CENH3), a key inner kinetochore protein. RNAs from another tandem repeat (180 bp knob sequence) or an abundant euchromatic retroelement (Opie) are undetectable within the same anti-CENH3 immune complexes. Both sense and antisense strands of CRM and CentC, but not siRNAs homologous to either repeat, were found to co-immunoprecipitate with CENH3. The bulk of the immunoprecipitated RNA ranged in size from 40-200 nt. These data provide evidence for a pool of protected, single stranded centromeric RNA within the centromere/kinetochore complex.
INTRODUCTION

One of the most highly specialized yet poorly understood regions of the chromosome is the centromere (Henikoff et al., 2001). Tandem repeat arrays are the only identifying feature of most higher eukaryotic centromeres, and even these diverge at astonishing rates. In several well-characterized centromeres tandem repeats are completely absent (e.g., Amor et al., 2004). The rapid sequence evolution, combined with the fact that centromeres from one species can sometimes function in related species (Jin et al., 2004) suggests that epigenetic determinants are involved in establishing the centromeric state. Although the mechanisms that establish and maintain centromeres are not known, one of the first steps in the process must involve the deposition of the histone H3 variant, Centromeric Histone H3 (CENH3). CENH3 is generally viewed as the core of the centromere (Choo, 2000; Henikoff et al., 2001), and is sufficient to recruit the major components of the inner kinetochore (Van Hooser et al., 2001). How a cell targets CENH3 to the proper chromosomal location is a key question in centromere biology.

DNA replication occurs during S-phase and is coincident with the deposition of the core histones. CENH3, however, is left out of this replication-coupled process and is deposited later in a replication-independent (RI) fashion (Shelby et al., 2000; Ahmad and Henikoff, 2002b). CENH3 RI deposition requires specialized centromeric targeting as well as a poorly understood histone exchange reaction (Sullivan, 2001). An early event in histone replacement is likely to be transcription, which can disrupt nucleosomes (Boeger et al., 2003) and in principle facilitate the incorporation of replacement histones (Jiang et al., 2003). Indeed, an analysis of a human neocentromere (that had recently formed over a gene-containing region) demonstrated that 15 of the 51 genes present produced apparently normal transcripts (Saffery et al., 2003). Similarly, rice centromere 8 contains fourteen genes, four of which are actively expressed (Nagaki et al., 2004). Two other papers provide direct links between transcription and centromere activation: in humans, selection for transcription flanking an ectopic centromere was shown to promote CENH3 (CENP-A) recruitment (Nakano et al., 2003), and in S. pombe the GATA-like
transcription factor Ams2 is required to initiate CENH3 (SpCENP-A) deposition and centromere formation (Chen et al., 2003).

There is also a considerable amount of evidence suggesting that RNA facilitates the targeting of chromatin modifying complexes to specific regions of the genome (Grewal and Moazed, 2003). A major breakthrough was the demonstration by Volpe et al. (2002) that RNAi is required to establish a heterochromatic state in the pericentromeric domains that flank the (CENH3-binding) centromere core of S. pombe. Mutations in the RNAi pathway release pericentromeric repeats from transcriptional repression and perturb normal centromere function (Volpe et al., 2002; Hall et al., 2003; Volpe et al., 2003; Fukagawa et al., 2004). Similarly, treatment of mouse cells with RNase causes the release Heterochromatin Protein 1 (HP1) and changes the spatial organization of histones in pericentromeric regions (Maison et al., 2002). These data, and the discovery and characterization of the RITS (RNA-induced Initiation of Transcriptional Silencing) effector complex (Verdel et al., 2004) support a model whereby siRNAs directly target homologous DNA sequences for chromatin modification (Volpe et al., 2002; Verdel et al., 2004). Notably however, RNAi mutations do not appear to affect the transcriptional status or protein composition of the functionally distinct CENH3-containing centromere core domains in either fission yeast or chicken cell lines (Partridge et al., 2000; Volpe et al., 2002; Fukagawa et al., 2004).

Although kinetochores differ in morphology from species to species, recent data have established that an important group of kinetochore proteins are conserved from S. cerevisiae to humans (Westermann et al., 2003). The fact that budding yeast has many of the same kinetochore proteins found in more complex eukaryotes suggests that the large plant and animal centromeres represent multiple iterations of the simple S. cerevisiae point centromere (Blower et al., 2002). A clue to higher order structure in humans came from chromatin immunoprecipitation (ChIP) experiments demonstrating that progressive micrococcal nuclease (MNase) digestion releases large complexes of constitutive centromere proteins before releasing the individual components (Ando et al., 2002). The molecules involved in the formation and stabilization of large-scale centromeric chromatin structure are not known. However, in several other chromatin protein complexes, RNA is an integral component. For instance, RNA is a known component
of sex chromosome dosage compensation complexes in mammals and *Drosophila* (Amrein, 2000; Wutz, 2003), of human pericentromeric heterochromatin (Maison et al., 2002), and of the yeast telomerase complex (Zappulla and Cech, 2004).

The experiments described here were designed to test the hypothesis that the repeats of the maize centromere core are transcribed and that the resulting transcripts are bound in some fashion to centromeric chromatin (Dawe, 2003; Jiang et al., 2003). In maize, CENH3 binds to 156 bp CentC repeat arrays and Centromeric Retrotransposable (CR) elements that are arranged in nearly continuous, intermingled arrays and clusters (Jin et al., 2004). Both CR elements and the centromeric satellite repeat CentC co-immunoprecipitate with maize CENH3, supporting the view that the retroelements cooperate with tandem repeat arrays to assemble a functional kinetochore (Zhong et al., 2002). Here we used a variation of the sensitive native chromatin immunoprecipitation (ChIP) technique to show that CR and CentC RNA are tightly associated with the maize kinetochore. Subsequent analysis of the RNA revealed significant quantities of both strands of each repeat, ranging in size from 40-900 bp. The data show that, as within the *S. pombe* pericentromere (Volpe et al., 2002), some level of transcription is a native feature of the centromere core and suggest a potential role for non-coding RNA in the specification of centromeric chromatin.

**METHODS**

**EST analysis**

Fourteen ESTs were characterized and listed below with respect to where they align on the consensus CRM2 element (GenBank AY129008). The EST CB278268 aligns to bases 800-1093 of CRM2, CB179846 to bases 508-1062, CB278262 to bases 797-1075, CB278268 to bases 800-1093, AW076314 to bases 646-1198, AW065493 to bases 732-1204, AW076306 to bases 657-1264, CB179288 to bases 900-1376, CB278333 to bases 760-1456, and BM660209 to bases 7039-7391; each of these are homologous to the sense strand of CRM2 (Nagaki et al., 2003). The EST AW017992 aligned to bases 1458-847 and AW017999 to bases 1458-890; these are homologous to the antisense strand of CRM2.
BM335652 aligns to bases 6825-7300 and BM349031 to bases 6825-7292; these sequences were derived from the sense strand and are terminated by poly A tracts within the 3’ LTR.

**Immunoprecipitation and slot blot analysis**

In our previously published ChIP protocol (Zhong et al., 2002), we used micrococcal nuclease (MNase) to digest the chromatin prior to immunoprecipitation. Although MNase is an excellent reagent for digesting maize chromatin, it is a known RNase (Telford and Stewart, 1989). Preliminary studies established that when MNase-prepared samples were blotted for RNA and probed with centromere repeats, the S/N ratios were unacceptably low (less than 2 in three different experiments). Therefore, in all studies reported here, nuclei from the W23 inbred were isolated (Zhong et al., 2002) and for each ChIP roughly 50 OD units (measured at 260 nm) were treated with RNase-free DNaseI (Promega, WI) for 10 min at 37°C, using a concentration of 4 units/mg DNA. Fragments of roughly 300-800 nt gave us the highest recovery (%IP) of centromeric DNA.

To effectively separate DNA and RNA in our ChIP samples, we took advantage of the fact that nucleic acid hybridization occurs efficiently only on single stranded molecules. When a sample is treated with formamide and heated slightly prior to blotting (68°C in 1X SSC, 7% formamide and 50% formaldehyde for 15 minutes), RNA is preserved but the DNA remains double stranded. Although both DNA and RNA will bind to a nylon membrane under these conditions (Khandjian and Meric, 1986; Shihara and Shikita, 1990) only the RNA is freely available for hybridization (we used N+ Hybond membranes, Amersham Biosciences). Conversely, when samples are treated with a high pH denaturant (0.4 M NaOH) prior to blotting, DNA is well preserved and readily detected following hybridization, while RNA (which is unstable at high pH) is barely visible.

The %IP was defined as \( P/(P+S) \) from CENH3 Ab - \( P/(P+S) \) from the preimmune serum. Signal/Noise (S/N) ratios were calculated under the assumption that noise is the fraction of nuclear RNA immunoprecipitated by preimmune serum, and signal is the fraction immunoprecipitated by anti-CENH3 antibodies:

\[
S/N = \frac{P/(P+S)}{P/(P+S)} \text{ from anti-CENH3 treatment divided by the } \frac{P/(P+S)}{P/(P+S)} \text{ from preimmune control.}
\]

For S/N calculations, preimmune %IP values were rounded up to 1.0 when the actual numbers
were less. For the data shown in Figures 2.2A and D, IP samples were treated with RNaseA (0.4 µg/µl) at 37°C for 10 min after the sample had been treated with phenol/chloroform to remove chromatin proteins, antibodies and other protein reagents.

A plasmid containing a 1.7 kb fragment of the Opie GAG domain was obtained from Chris Della Vedova via Jon Lamb of the James Birchler lab (University of Missouri). The clone is homologous to Opie B, GenBank AF466932, bp 41014-42769 (Ramakrishna et al., 2002). All other DNA probes were described previously (Zhong et al., 2002). Strand-specific CRM and CentC probes were prepared by cloning the GAG.90 and CentC inserts (Zhong et al., 2002) in both orientations into pBluescript (Stratagene), and transcribing the sequence with T7 polymerase (Riboprobe kit, Promega). The sense GAG.90 construct expresses the sense strand of the CRM element (AY129008) while the ‘sense’ construct of CentC expresses the strand reported in GenBank record AF078923. Hybridization experiments involving single stranded (M13-generated) DNA molecules (not shown) and single stranded RNA oligonucleotides (Fig. 3A), demonstrated that the CRM riboprobes were essentially 100% strand-specific.

**Polymerase Chain Reaction**

For the data shown in Figure 2.1, cDNA was prepared using the Clontech smart cDNA library construction kit. The library was verified as free of detectable DNA by amplifying a portion of the maize CenH3 known to contain an intron. PCR was carried out on the Clontech-prepared cDNA using the following primers: GAG.65(F) 5’-agggaaatcgacactccttgetta-3’ and(R) 5’-gatcggcaagatgacccaggaa-3’; GAG.90(F) 5’-ctgttgggtgatatgagtcgga -3’ and (R) 5’-gatcggcaagatgcaccaggaa-3’; RT.42(F) 5’-atgtcagcatctttgctctgtta-3 and (R) 5’-tcgtatgaaaattgggaagatgaa-3’; INT.56 (F) 5’-ttgaatgtgatgctagtggaattgga-3 and (R) 5’-cctccatgcgcctcctgtaacaacaaaag-3’; LTR.32 (F) 5’-ttgaatgtgatgctagtggaattgga-3’ and (R) 5’-cctccatgcgcctcctgtaacaacaaaag-3’; LTR.32 (F) 5’-ttgaatgtgatgctagtggaattgga-3’ and (R) 5’-cctccatgcgcctcctgtaacaacaaaag-3’. Strand-specific RTPCR (Fig. 3B) was carried out on IP fractions that were digested twice with DNase (at 0.4 units/µl), using primers for the CRM GAG.90.
RNA detection by polyacrylamide gel electrophoresis

For the data shown in Figure 2.3A, samples were incubated for 10 min at 37°C with an excess of RNase-free DNaseI (0.4 units/µl; >1000X times the concentration used to digest chromatin for immunoprecipitation). It is difficult to estimate the quantity of RNA loaded, though each lane contains roughly 20% of a ChIP experiment. Samples were added to RNA loading buffer containing formamide, heated for 5 min at 95°C, and electrophoresed on 15% denaturing polyacrylamide (7M urea) gels. The marker lane contained 0.5 nmol of a 28 nt single-stranded RNA identical to the CRM GAG sense strand (5’-CCAAAUCUGCCCAGAAACCA-GCAGGUA-3’). Gels were transferred to either N⁺ Hybond or Ambion Nylon 66 membranes and hybridized with strand specific RNA probes. All data from ChIP samples were exposed to a phosphorimager for 18-21 hours. The exposure time for the 28 nt small RNA marker was 30 minutes.

RESULTS

CRM is an expressed retroelement

During our initial studies of CRM (Centromeric Retrotransposable element from Maize; 30) we noticed that there were at least 18 CRM-homologous ESTs in GenBank (listed in Methods). The majority (14/18) are non-chimeric, suggesting that they were initiated from within the retroelement, and 12 of the non-chimeric transcripts are derived from the sense strand. Further, at least two terminate in polyadenine tracts within the 3’ LTR, suggesting that a subset of the ESTs represent legitimate CRM-initiated cDNAs. Although we were not able to recover full-length cDNAs, at least five different regions of CRM were easily amplified by RTPCR from poly-A selected mRNA (Fig. 2.1B).

Centromeric RNA is immunoprecipitated with CENH3 Ab

To test for the presence of RNA at maize centromeres, we developed and employed a (RNase-free) DNaseI-based chromatin preparation method. Chromatin was immunoprecipitated using preimmune and anti-CENH3 antisera, and RNA immobilized on nylon membranes using an RNA slot blotting protocol. Blots containing the Supernatant (S) and Pellet (P) fractions from both treatments were sequentially
probed with the known centromeric sequences CentC and CRM, and two negative controls: the 180 bp knob repeat (Peacock et al., 1981), a tandemly arrayed sequence located exclusively on chromosome arms (Dawe and Hiatt, 2004), and Opie B, a retrotransposon with a uniform euchromatic (generally non-centromeric) distribution (SanMiguel et al., 1996; Mroczek and Dawe, 2003).

When RNA slot blots were probed with centromeric sequences, strong hybridization was observed in the pellet fractions. CRM and CentC signal intensities were far above background in all experiments (signal to noise (S/N) ratios significantly exceeded 10; Table 2.1) and were nearly or completely RNaseA-sensitive (Fig 2.2A, 2.2D). In addition, the RNA was tightly bound within the CENH3 immune complex. Ionic strengths as high as 1M, high enough to dissociate the majority of H2a-H2b dimers from canonical nucleosomes (Burton et al., 1978), had a surprisingly small effect on the recovery of RNA following chromatin immunoprecipitation (Fig. 2.2C).

The percentage of nuclear RNA associated with CENH3 was estimated using background subtraction. An average of five independent experiments demonstrated that close to half of the CRM (44%) and CentC (48%) RNA in purified nuclei is associated with CENH3 (Table 2.1, Fig. 2.2B). In one experiment, 86% of the CRM RNA and 69% of CentC RNA was recovered in the anti-CENH3 immune complex (Table 2.1). Conversely, the %IPs for knob and Opie RNA hovered around zero (Table 2.1, Fig. 2.2A,B), with S/N ratios significantly below two (Table 2.1). Centromeric DNA was also measured and quantified following CENH3-mediated immunoprecipitation. As shown in Figure 2.2B, the data indicate that the relative quantities of centromeric DNA and RNA mirror each other with remarkable accuracy.

**Both strands of CRM and CentC are associated with the CENH3 immune complex**

In addition to DNA probes for the sequences shown in Figure 2.2A, ChIP samples were hybridized with RNA probes specific for the forward and reverse strands of CRM (GAG) and CentC (Fig. 2.2D). Since the signal was almost entirely abolished by RNaseA treatment, these data confirm that single-stranded RNA homologous to both strands of CRM and CentC are co-immunoprecipitated with maize CENH3 antibodies. The reverse CRM transcripts may represent examples of transcription through nested CRM elements, since retroelements often insert into each other in reverse orientations (SanMiguel et al.,
1996). Using two forward CRM primers we were able to recover several different PCR products from genomic DNA; which when sequenced, proved to represent reverse CRM-CRM insertions (not shown). Similarly, CentC transcripts are likely derived from read-through transcription by such CRM elements (see Discussion).

The majority of CENH3-associated RNA is greater than 40 nt in length

The simultaneous presence of forward and reverse transcripts is expected to activate the RNAi pathway and produce siRNAs (Hannon, 2002). However, on polyacrylamide gels we were not able to detect centromeric RNA in the size range expected for siRNAs (22-30 nt). Rather, the bulk of the CentC and CRM RNA was between 40 and 250 nt in length (Fig. 2.3A). The banding patterns in supernatant and pellet fractions were very similar, with both forward and reverse probes identifying several distinct bands (in four independent experiments). It is possible that the ~40 and 60 nt bands homologous to CentC are dicer-like products, but to our knowledge no bona fide siRNAs in this size range have been reported.

The apparent upper limit on the size of the RNA may be an artifact of the procedure, since much larger RNAs could be detected using more sensitive methods. In two experiments we were able to detect 900 nt RNAs from both strands of the CRM GAG domain by RTPCR (Fig. 2.3B). Taken together, these data suggest that the centromere-associated RNAs are variable in size, but are rarely as small as would be expected if the RNAs were the product of RNA interference.

DISCUSSION

The centromere core is often viewed as a genetically inert domain of the chromosome (Ahmad and Henikoff, 2002b; Pidoux et al., 2003). However, recent results clearly indicate that genes within centromeres can be transcribed (Saffery et al., 2003; Nagaki et al., 2004) and suggest that transcription may contribute to centromere formation (Chen et al., 2003; Nakano et al., 2003). Here we add the observations that centromere repeats are actively transcribed, and that a significant fraction of the RNA is bound, directly or indirectly, to CENH3. The fact that CentC and CRM transcripts co-immunoprecipitate with native (not chemically cross linked) CENH3 complexes, at ionic strengths sufficient to partially
disrupt nucleosomes, indicates that RNA is an integral component of centromeric chromatin (Table 2.1, Fig. 2.2, 2.3).

The formation of centromeric RNAs

To explain the origin of CentC transcripts, we refer to a convincing body of evidence indicating that retroelements can initiate aberrant read-through transcription of flanking DNA (Whitelaw and Martin, 2001; Kashkush et al., 2003). Although most retrotransposons are rare in centromeres (Sun et al., 1997; Schueler et al., 2001; Mroczek and Dawe, 2003), CRM elements are abundant and actively transcribed. CR elements can occupy as much 60% of the centromeric DNA in both maize and rice (Jin et al., 2004; Nagaki et al., 2004), providing ample opportunity for transcripts to be initiated within the centromere. If the maize centromeric BAC 16H10 is used as a guide (Nagaki et al., 2003), roughly seven CRM/CentC junctions can be expected in a 90-Kb CRM-rich region of centromeric DNA. We also found evidence of reverse CRM transcripts in the EST database, on blots (Fig. 2.2D, 2.3A), and by RTPCR (Fig. 2.3B). These RNAs likely represent examples of the CRM promoter driving transcription through reverse-nested insertions. Such nested insertions, truncations, and other rearrangements known to be associated with retroelements have the potential to provide an array of forward and reverse templates for CRM and CentC. In principle, any promoter could initiate similar centromeric transcripts, whether they are from genes (Saffery et al., 2003; Nagaki et al., 2004) or other forms of transposable elements.

Size of the RNAs

Despite the presence of both strands of CentC and CRM, we did not detect canonical siRNA-sized molecules within the nucleus or in association with CENH3. It remains possible that low concentrations of centromeric siRNAs were present but not detected in our assays. We also note that since our experiments were designed to study nuclear RNA, the possibility that centromeric siRNAs might exist within the cytoplasm was not addressed. In nuclei, however, the fact that larger RNAs from both strands of centromeric repeats were readily detected suggests that the RNAi machinery did not process these RNAs. A portion of the centromeric RNA may be kept in a single stranded state within the kinetochore, or otherwise protected from the RNase-III like (dicer) enzymes that initiate RNA interference.
RNAs ranging in size from 40-900 nt are recovered from immunoprecipitates (Fig. 2.3), but it is difficult to know how long the centromere-associated RNAs are in vivo. An analogy can be drawn to the Xist RNA that associates directly with chromatin during human X chromosome inactivation (Wutz, 2003). Xist is 17 kb when transcribed, but only small fragments of the RNA (e.g. 242 bp) have been detected after immunoprecipitation with antibodies to macroH2A, a histone variant that is deposited on the inactive X chromosome subsequent to coating by Xist (Gilbert et al., 2000).

**RNA in the initiation and maintenance of the centromeric state**

CENH3 is one of many replacement histones that are incorporated into chromatin after DNA replication (unlike core histones, which are assembled during DNA replication; (Shelby et al., 2000; Ahmad and Henikoff, 2002b). Although the mechanisms of histone replacement are not well understood, it is likely that transcription, apparently a common feature of centromeres (11, 12, this report), facilitates the process. Transcription partially disassembles nucleosomes (Boeger et al., 2003) and is correlated with the replacement of histone H3 with a variant known as histone H3.3 (McKittrick et al., 2004; Tagami et al., 2004). The fact that the *S. pombe* Ams2 transcription factor mediates CENH3 (SpCENP-A) localization (Chen et al., 2003) provides strong support for the idea that transcription is involved in establishing the centromeric state.

Our data also establish that centromeric RNA can remain bound to the centromere/kinetochore complex following transcription. Interestingly, early ultrastructural studies strongly support the idea that RNA is present at plant and animal kinetochores (Braselton, 1975; Rieder, 1979). In its chromatin-bound capacity, centromeric RNA may have a targeting and/or stabilizing role. Telomerase RNA, for instance, not only targets the telomerase complex to the ends of chromosomes by base pairing, but also serves as a flexible scaffold for associated proteins (Zappulla and Cech, 2004). Zappulla et al (2004), working with telomerase RNA, suggested that: “the overall structure of telomerase is at least somewhat flexible” and that “telomerase RNA tethers [proteins] to the RNP rather than positioning them precisely within a highly structured complex” (Zappulla and Cech, 2004). In a broad sense telomerase can be compared to RITS, which is thought to use RNA to target (by base pairing) a three-protein ribonucleoprotein complex on
specific chromatin domains (Volpe et al., 2002; Verdel et al., 2004). Similarly, we show that the relative quantity of centromeric and non-centromeric RNA mirrors the DNA within CENH3 immune complexes (Fig. 2.2B). Although our approach does not allow us to measure the stoichiometry or specificity of the RNA-DNA associations, the data are consistent with the idea that centromeric RNA interacts with cognate DNA sequences within CENH3-containing nucleosomes.

There are also strong parallels between our data, telomerase, and the RNAs that regulate dosage compensation in humans and Drosophila. In both species, long RNAs interact with regulatory proteins to form a complex known in Drosophila as the compensasome (Wutz, 2003). RNA-containing compensasomes spread along entire chromosomes to either shut down (humans) or double (Drosophila) gene expression (Wutz, 2003). One working model is that compensasome RNA provides low affinity contacts that facilitate higher-order interactions among chromatin proteins (Amrein, 2000; Wutz, 2003). The authors again point to the fact that RNA is more flexible than protein, and can tolerate rapid sequence divergence while still maintaining function (Amrein, 2000; Wutz, 2003). Unlike telomerase, however, dosage compensation complexes function exclusively in cis, spreading to linked sites but never from one chromosome to another.

The available data suggest that centromere transcription may contribute to both the deposition (initiation) and stabilization (maintenance) of kinetochore chromatin structure. Transcription could initiate the process by opening chromatin to allow the replacement of histone H3 with CENH3, a key early event in centromere specification (Choo, 2001; Sullivan, 2001). Secondly, centromere-associated RNA could provide a flexible scaffold that brings together and stabilizes the proteins of the inner kinetochore. Centromeric targeting may be conferred by base pairing (consistent with the roles of siRNA or telomerase RNA; (Volpe et al., 2002; Verdel et al., 2004) or by a cis-acting/spreading mechanism (as exemplified by the compensasome; (Amrein, 2000; Wutz, 2003). Importantly, a RNA scaffold could also help to explain how centromeric repeats evolve so rapidly (Henikoff et al., 2001) yet maintain their function. If RNA has an important functional role in maintaining the centromeric state, then DNA sequence may be less important than the secondary structure of the folded RNA derived from it.
ACKNOWLEDGEMENTS

We are grateful to Roger Deal for helping to establish the conditions used for DNaseI ChIP and to Amy Bouck for useful statistics discussions. This work was supported by a grant from the National Science Foundation (9975827) to R.K.D.

REFERENCES


Figure 2.1 CRM is Actively Transcribed

At top is a map of CRM element (30, later called CRM2; (Nagaki et al., 2003). The approximate locations of the GAG, reverse transcriptase (RT), and integrase (INT) domains are shown. Below is an agarose gel showing that several internal regions of CRM2 are readily amplified by RTPCR. The locations of the amplified regions are indicated on the map.
Figure 2.2 Centromere-Encoded RNAs are Co-Immunoprecipitated with CENH3

A) Chromatin samples were immunoprecipitated with anti-CENH3 antibodies, blotted for RNA, and probed with DNA probes for the sequences indicated. Supernatant (S) and pellet (P) fractions for the preimmune and anti-CENH3 treatments are shown. As shown in the bottom lane, RNaseA treatment removed the majority of RNA hybridization. These slot blot images were used to acquire the numerical data in Table 1, Experiment 1. 

B) The RNA and DNA %IPs from five different experiments. The RNA %IP data (from Table 1) and associated DNA %IP values from each of five experiments are shown as
mean ± SE. C) High ionic strengths have a minimal impact on the immunoprecipitation of CentC RNA. The preimmune (S and P) fractions are shown in the top two lanes, standard ChIP (using a 0.3 M NaCl wash) in the second and third lanes, and standard ChIP following a 1M NaCl wash in the final two lanes. This experiment was not repeated in kind, but a second experiment with 0.7M NaCl gave similar results. 

D) Both strands of CRM and CentC are present after anti-CENH3-mediated ChIP. P fractions were blotted for the presence of DNA and RNA, either with or without prior RNaseA treatment. The sense and antisense strands are defined in Methods. This experiment was repeated three times with essentially identical results.
Figure 2.3 Centromeric siRNAs are not Detected Following CENH3-Mediated ChIP

A) The bulk of immunoprecipitated centromeric RNA ranges in size from 40-250 nt in length. The supernatant (S) and pellet (P) fractions are shown for both CRM and CentC. For the P fraction, hybridization with the opposite strand is shown. Hybridization to a 28 nt single-stranded synthetic RNA (SSRNA) homologous to the sense strand of the CRM GAG domain is shown at left. The 28 nt marker is underexposed relative to the other lanes. DNA markers (not shown, but indicated in nt) were used for higher molecular weight estimates.

B) ~900 bp fragments of the CRM GAG domain were detected by strand-specific RTPCR. This technique is
not quantitative. The absence of product when no reverse transcriptase (RT -) was added indicates the bands were derived from immunoprecipitated RNA, not DNA.
### Table 2.1 Association of RNA with CENH3

<table>
<thead>
<tr>
<th>Exp.</th>
<th>CRM</th>
<th>Opie</th>
<th>CentC</th>
<th>Knob</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>S/N</td>
<td>%IP</td>
<td>S/N</td>
<td>%IP</td>
</tr>
<tr>
<td>1</td>
<td>26.3</td>
<td>26.0</td>
<td>1.0</td>
<td>-0.3</td>
</tr>
<tr>
<td>2</td>
<td>12.9</td>
<td>12.6</td>
<td>1.0</td>
<td>-0.9</td>
</tr>
<tr>
<td>3</td>
<td>41.5</td>
<td>46.2</td>
<td>1.0</td>
<td>0.0</td>
</tr>
<tr>
<td>4</td>
<td>42.1</td>
<td>46.9</td>
<td>1.0</td>
<td>0.0</td>
</tr>
<tr>
<td>5</td>
<td>86.4</td>
<td>86.4</td>
<td>0.8</td>
<td>-1.0</td>
</tr>
</tbody>
</table>

1 Data are expressed as Signal/Noise ratio [S/N = (P/(P+S)) from anti-CENH3 treatment divided by (P/(P+S)) from preimmune control] and %IP [%IP = (P/(P+S)) from anti-CENH3 treatment minus (P/(P+S)) from preimmune control]. S/N ratio is an indication of the overall quantity of RNA, specificity of the antibody, the strength of interactions within the immune complex, and the overall efficiency of the procedure, while %IP is the percentage of nuclear RNA associated with CENH3 under the experimental conditions used. The S/N ratios were analyzed for statistical significance using one-tailed t-tests (P ≤ 0.025). For CRM and CentC, the S/N ratios were significantly greater than 10, and for Opie and Knob the S/N ratios were significantly less than 2.
CHAPTER III:

IDENTIFICATION OF A MAIZE NEOCENTROMERE IN AN OAT-MAIZE ADDITION LINE

ABSTRACT

We report a neocentromere event on maize chromosome 3 that occurred due to chromosome breakage. The neocentromere lies on a fragment of the short arm that lacks the primary centromere DNA elements, CentC and CRM. It transmitted in the genomic background of oat via a new centromere (and kinetochore), as shown by immunolocalization of the oat CENH3 protein. Despite normal transmission of the maize fragment in most progeny, neocentromeres appear to vary in size within the same tissue, as shown by fluorescent measurements. A secondary truncation in one line lowered mitotic transmission to 3% and precipitously reduced the size of the chromosome. The results support the view that neocentromere formation is generally associated with major genomic disturbances such as wide species crosses or deletion of an existing centromere. The data further suggest that new centromeres may undergo a period of instability that is corrected over a period of several generations.
INTRODUCTION

Centromeres must be positionally stable – otherwise homologous chromosomes will not pair correctly and can be lost at high frequencies during meiosis. The stability is perplexing because centromeres are not determined in a strictly genetic sense (Dawe and Henikoff, 2006). A special histone called centromeric histone H3 (CENH3) replaces standard histone H3 and creates an environment that demarcates centromeres from surrounding chromatin. Several other proteins are required to build the kinetochore over a centromere, but CENH3 is widely considered the ‘centromere identifier’ (Warburton et al., 1997; Malik and Henikoff, 2001; Van Hooser et al., 2001). Although centromere chromatin typically forms on tandem arrayed satellite repeats or special retrotransposon families, there are no common constraints on the DNA sequences CENH3 may interact with (Jiang et al., 2003). Other data support and extend the idea that centromeres are propagated by epigenetic means. Some of the best evidence comes from chromosome addition lines where alien centromeres are maintained by centromere proteins of the host species (Jin et al., 2004), and rare cases in which the centromere chromatin forms at a new locus, creating a neocentromere (Maggert and Karpen, 2001; Nasuda et al., 2005; Alonso et al., 2007; Marshall et al., 2008).

Neocentromerization has been proposed to underlie key evolutionary events such as rapid centromere evolution, hybrid incompatibility, and speciation (Ventura et al., 2001; O'Neill et al., 2004; Marshall et al., 2008). Extensive work in human cell lines suggests that neocentromeres have different DNA sequences but structurally similar kinetochores (Saffery et al., 2000; Alonso et al., 2003; Alonso et al., 2007). In contrast, ‘classical’ neocentromeres form at terminal heterochromatic knob repeats of some plants and do not share any protein similarity to the normal centromere/kinetochore (Dawe and Hiatt, 2004). Despite its presumed importance, the process of new centromere formation is not well understood.

One mode of neocentromere formation appears to involve the direct transfer of kinetochore proteins to a new locus. Such a kinetochore transfer mechanism was first suggested by a study in Drosophila, where chromosomes were broken next to a centromere and the centromere/kinetochore moved laterally to neighboring DNA sequences (Maggert and Karpen, 2001). In wheat, a similar neocentromere event
occurred on a barley chromosome that had been introduced by a wide cross. A spontaneous breakage produced an isochromosome with no known centromere repeats and two mirror images of the chromosome 7 short arm. The newly formed centromere was transmitted normally, although some chromosome variants were produced; for example some had one or part of one of the two original arms. Immunolocalization of CENH3 and other key proteins confirmed that a neocentromere had formed close to the position of the original barley centromere (Nasuda et al., 2005). A common feature of most neocentromeres is their association with genome rearrangements that cause a loss of the original centromere. It seems likely that neocentromeres arise as a natural means of stabilizing broken genomes.

We report the characterization of a second plant neocentromere. An unstable chromosome lacking a visible centromere was reported by Muehlbauer et al. as a derivative of maize chromosome 3 in an oat genomic background (Muehlbauer et al., 1999). We analyzed progeny and show that the unusual chromosome is a fragment of the chromosome 3 short arm (ch3S) and lacks known maize centromere DNA sequences. In contrast to the highly unstable derivative originally reported, this neocentromere now reliably transmits the maize chromosome in somatic tissue. We cloned the oat CENH3 gene and raised an antibody to the predicted protein in order to demonstrate the location of the new centromere. In apparent contrast to the neocentromeres formed in Drosophila and barley, the maize neocentromere formed at a site distant from the original, making a templating event in cis unlikely. We also observed wide variance in the size of stably transmitting neocentromeres, suggesting to us that newly formed centromeres undergo a period of epigenetic flux before (presumably) stabilizing.

RESULTS

A neocentromere event on maize chromosome arm 3S

GISH analysis of progeny resulting from self mating a monosomic oat-maize chromosome 3 addition line detected a small chromosome comprised solely of maize DNA (Muehlbauer et al., 1999). The lack of a primary constriction indicates it formed by a breakage event within or near the centromere (Figure 5D from Muehlbauer et al). We became interested in this line, BC1F2-4/9, when a Southern blot of a sibling
line suggested that it had lost its entire native centromere (summarized in Figure 3.1). Further mapping with SSR markers ordered along the length of maize chromosome 3 suggest that only DNA from the short arm (ch3S) remains. We hypothesize that the full maize chromosome underwent a breakage event on the short arm of ch3, distal to the centromere but proximal to marker bnlg1957, placing it between bnlg1957 and AY110151 (between contig 120 and 121, IBM2 map) (Figure 3.1). A new centromere stabilizing the acentric fragment is the most plausible explanation for the origin of the truncated chromosome. We wondered if newly formed centromeres have any unusual properties compared to those long established.

**Stability of the neocentromere**

We were especially intrigued by Muehlbauer and colleagues’ evidence suggesting that the truncated fragment was completely absent in older tissue, such as tillers (Muehlbauer et al., 1999). Somatic loss could result from defects in the neocentromere. As a test we screened young leaves of T2 progeny selfed from the plant containing the original neocentromere event, and compared them to a full disomic ch3 addition line, OMAd3.1. A primer pair that amplifies maize, but not oat, members of the abundant OPIE family of retroelements identified 6 of 40 positive progeny. Quantitative PCR analysis confirmed that the truncated fragment had lost its native centromere. Maize centromere DNA is comprised of a tandem repeated satellite named CentC, interspersed with a centromere-specific retroelement family named CRM (Zhong et al., 2002a; Jin et al., 2004). CentC did not amplify in any line and CRM retroelements amplified at very low levels (similar to the oat parental background line Sun II, not shown). In contrast both elements robustly amplified from a full ch3 addition line (Figure 3.2). OPIE elements were similarly amplified in 5 of 6 neocentromere progeny (Figure 3.2) as well as in another leaf from the same tiller (not shown). Consistent OPIE amplification suggested that the original neocentromere reported by Muehlbauer and colleagues had stabilized, and is now transmitting at a much more stable frequency. The sole exception was line CTo1_3, which showed dramatically reduced levels of OPIE and CRM elements (Figure 3.2). It appeared that the neocentromere in CTo1_3 had spontaneously destabilized.

To directly analyze the stable and putative unstable forms of the neocentromere chromosome, we performed FISH on root tip spreads of the five stable lines, CTo1_3, and the OMAd3.1 control. Neither
CentC nor CRM elements were detected in any neocentromere line, stable or putative unstable, as opposed to the clear signals detected in OMAd3.1 (Figure 3.3a, b; CRM not shown). The neocentromere chromosome is consistently seen in all five stable lines by its reduced size relative to oat chromosomes and its OPIE signal (Figure 3.3a). Because the OPIE signal appears symmetrical, we wondered if the ch3 short arm was duplicated around the center, despite the lack of a primary constriction typical of isochromosomes. In a second assay, we hybridized the chromosomes with a FISH probe for CentA. CentA is a family of retrotransposons that are closely related to the centromeric CRM elements, and generally localize in the centromere/pericentromere regions of grasses, including oat, but also speckle chromosome arms in maize (Figure 3.3c). The speckled, unsymmetrical CentA signal obtained from stable neocentromere lines confirms its origin as a simple truncation of maize chromosome 3 (Figure 3.3d). Upon its detection, the putative unstable chromosome of CTo1_3 was seen as two tiny symmetrical arms flanking a primary constriction (Figure 3.3e). Using OPIE signals of interphase cells, we calculated somatic transmission rates of nearly 100% in stable lines CTo1_12 and CTo1_33, as opposed to 3% in CTo1_3 (Table 3.1). Apparently a secondary breakage event occurred in Cto1_3 that further truncated the chromosome and either destabilized the original neocentromere or formed a new and unstable centromere. The dramatic increase in stability of the neocentromere from the T_0 (original Muehlbauer line) to the T_2 generation, combined with the rapid loss of stability in Cto1_3, raise the possibility that some neocentromere events may be initially unstable and subject to a period of flux, during which time they may be lost.
Position of the neocentromere

To investigate the functional aspects of neocentromeres, we developed polyclonal antibodies against a poorly conserved region of the oat CENH3 protein (here designated *Avena sativa* CENH3, or oat CENH3; Figure 3.4a). Oat CENH3 antibodies stain oat, but not maize cells (Figure 3.4b). Oat CENH3 antibodies also show the constitutive localization pattern that is typical for CENH3 (Talbert et al., 2002; Zhong et al., 2002b), clearly labeling the primary constrictions of all chromosomes (Figure 3.4c, d).

In *Drosophila* and barley the neocentromere formed close to the site of the original centromere. Based on our SSR work, we assumed the oat CENH3 signal would be on one end of the maize fragment. However, our cytological analysis of stable lines CTo1_12, CTo1_33, and CTo1_40, revealed a submetacentric location for oat CENH3 (Figure 3.5a-c). Thus the maize neocentromere appears to have formed *de novo* at a site unlinked to the original (lost) centromere.

We found only one example of the highly unstable CTo1_3 isochromosome in a meiotic preparation. Despite strong staining of the oat kinetochores, oat CENH3 staining of the isochromosome was below the detectable limit (Figure 3.5d). Although plant spindles are capable of exerting polar forces on acentric fragments (Khodjakov et al., 1996), it is doubtful whether such a mechanism is capable of sustained transmission. We think it more likely that the unstable kinetochore has been reduced to a barely functional size.

Kinetochore size within stable neocentromere lines vary

Classical plant cytogenetic research recognized a positive correlation between centromere size and somatic stability, but could only speculate as to why (Rhoades, 1940; Steinitz.Lm, 1966). A potential mechanism for small centromere instability is suggested by later work that supports a minimum size requirement for kinetochore and CENH3 domains of stable chromosomes (Cherry et al., 1989; McEwen et al., 1998; Okamoto et al., 2007). We could not analyze the kinetochore size of the original unstable neocentromere, but reasoned that analyzing the kinetochores of stable progeny would help us understand post-formation events.
The fluorescence intensity of immunodetected CENH3 is a good estimate of kinetochore size because CENH3 chromatin is the structural basis of kinetochores. The primary function of kinetochores are to bind microtubules, and microtubule binding capacity is dependent in large part on the structural densities of kinetochores (Mitchison and Kirschner, 1985; McEwen et al., 1998). We compared CENH3 intensities of neocentromeres from stable lines and their oat counterparts to approximate their functional sizes (Materials and Methods). The CENH3 staining intensity of most neocentromeres was two or more standard deviations smaller than the corresponding mean oat centromere value from the same cells, but some neocentromeres were much smaller than others (expressed as % of mean oat; Table 3.2; Figure 3.6). The apparent size variation among neocentromeres is striking considering that previously measured human neocentromeres were much less variable (Irvine et al., 2004). We used an F-test to compare the CENH3 staining of neocentromeres to the mean staining of oat centromeres, with a null hypothesis that these values were equal. The variation of maize neocentromere size is greater than that of native oat centromeres at a 95% significance level (Materials and Methods). Considering the apparent size plasticity of neocentromeres, we speculate that a sub-minimal kinetochore caused instability of the maize chromosome in the T₀ generation, and that an increase in kinetochore size is responsible for its stability in the T₁ and T₂ generations.

In anther tissue the neocentromere chromosome is often marginalized from the main nuclear body as the spindle forms in early prophase. We noticed that neocentromeres attached to the spindle were larger than neocentromeres isolated from the spindle (Figure 3.6). Though it seems likely that larger neocentromeres are simply better at binding microtubules, there is the additional possibility that microtubule attachments somehow enhance kinetochore size. Regardless, this observation supports the suggestion that neocentromere stability depends on its ability to interact with the spindle.
DISCUSSION

The stability of centromeres provides the foundation for chromosomal, and hence genetic, inheritance. Yet because centromeres have an epigenetic basis, they are capable of rapid adaptation to genomic instabilities such as species hybridization and chromosome breakage. While neocentromere adaptation may be detrimental in some cases, cancers for example, they may also allow centromeres to affect evolutionary change. The current paradigm of centromere formation involves 1) establishing a critical mass of CENH3 chromatin sufficient to form a functional centromere/kinetochore, and 2) reinforcement of the centromere position over evolutionary timescales by the accumulation of specific repeats within the centromere and the expansion of flanking heterochromatin domains by genomic rearrangements and transposons (Nagaki et al., 2004; Topp and Dawe, 2006; Marshall et al., 2008). It has been shown that the establishment phase may be accomplished by overwhelming a locus with CENH3 proteins (Heun et al., 2006). Our results suggest that CENH3 accumulates over time and that the process is very dynamic, at least in the first two generations.

The idea of a flux period might have been expected given what is known about CENH3 incorporation into chromatin. The CENH3 nucleosome density of centromeres is diluted during chromosome replication because CENH3 deposition is not coupled to the replication machinery like canonical histones (Shelby et al., 2000; Sullivan and Karpen, 2001; Lermontova et al., 2007). Various scenarios are proposed for centromere replenishment (Sullivan et al., 2001; Smith, 2002; Mellone and Allshire, 2003), but the weight of evidence points to a mechanism in which general chromatin factors deposit CENH3 histones at accessible locations (Henikoff and Ahmad, 2005; Furuyama et al., 2006). Established centromeres are a highly favorable environment for new CENH3 incorporation, reinforcing centromere position. Cells have robust mechanisms for removing CENH3 nucleosomes as well, as evidenced by the rapid inactivation of entire centromeres in plants and animals (Amor et al., 2004; Higgins et al., 2005; Han et al., 2006), and CENH3 proteolytic degradation in yeast and flies (Collins et al., 2004; Moreno-Moreno et al., 2006).
One interpretation of the stability of established centromeres is that they are protected from CENH3 removal by some inherent property. A newly formed centromere would therefore be especially susceptible to loss until it gained enough centromere ‘identity’ to resist removal forces. An obvious feature of established centromeres is flanking heterochromatin, and several lines of evidence have led researchers to suggest that heterochromatin is integral to both centromere stability and de novo formation (Henikoff et al., 2000; Maggert and Karpen, 2000; Nakashima et al., 2005; Okada et al., 2007; Ishii et al., 2008; Nakano et al., 2008). While general heterochromatin is unlikely to be a centromere identity factor per se, it may help to protect fledgling neocentromeres from removal.

The interplay between centromere identity factors and CENH3 removal forces may help explain the mosaic inheritance patterns reported for some human neocentromeres (Amor and Choo, 2002), as well as the rapid stabilization and variance in CENH3 domain size of the maize neocentromere reported here. For example, neocentromeres that were apparently cis-templated from nearby centromeres are stable, perhaps due to their formation near pericentric heterochromatin (Maggert and Karpen, 2001; Nasuda et al., 2005). We might expect de novo neocentromeres to have less centromere identity and less protection from CENH3 removal. Perhaps the initial instability of the maize neocentromere resulted from its local susceptibility to CENH3 removal, which would reduce the kinetochore to a sub-minimal size.

Based on our observations and the current understanding of CENH3 dynamics, we speculate on the processes that may occur during new centromere establishment (Figure 3.7). New centromeres undergo an intermediate period of CENH3 fluctuation, where reinforcement factors such as CENH3 from the last cell cycle and flanking heterochromatin counteract CENH3 removal. The period and outcome of the establishment process depends on the relative strengths of these forces and the size of the minimal CENH3 domain. If the neocentromere formed in a favorable environment, the initial reinforcement factors are strong and will rapidly push CENH3 density past the threshold. If the neocentromere formed in a less than favorable environment, the initial reinforcement factors are weak and the neocentromere will be unstable and may be lost. Alternatively, an unstable neocentromere may gain enough centromere
identity through repeated cell divisions to reinforce itself (Mellone and Allshire, 2003). Under this view, the location and initial events of neocentromere formation are critical to its ultimate stability.

**MATERIALS AND METHODS**

**Mapping of the maize chromosome 3 breakage site**

The following conditions were used to PCR amplify markers from each material: 50 ng genomic DNA, 0.2 units HotStarTaq (Qiagen), 0.5 µM each primer forward and reverse, 2.67 µM dNTPs, and 1X supplied buffer in a 15 µL total volume. The cycling program began with a 15 minute incubation at 95°C to activate the enzyme followed by 36 cycles of 94°C for 30 seconds, 57°C or 52°C for 30 seconds and 72°C for 75 seconds, and finally two minutes at 72°C. PCR primers were either downloaded from www.maizegdb.org or designed using the Primer3 program with an optimal annealing temperature of 63°C (Rozen and Skaletsky, 2000). Southern blots were performed using standard conditions.

**Quantitative PCR of neocentromere material**

qPCR was performed on an Eppendorf Realplex machine using the following conditions: ~100 ng genomic DNA, 0.375 units AmpliTaq Gold (Applied Biosystems), 3 mM MgCl₂, 0.3 µM each primer forward and reverse, 200 µM dNTPs, 5% DMSO, 1X SYBR Green (Invitrogen), and 1X supplied buffer in a 15 µL total volume. Primers pairs are as follows: CentC F-(GATTGGGGCATGTTCGTGGTGTG), R- (CACTACTTATTAGGTCGAAAC); CRM F-(CTCGTGCTCGTCAACTCAA), R– (ACCGTCACAAAGTTGGTTGT); OPIE F-(GATTCCTCGCAAACGGGAW), R- (CTTGCTACTCCACGTGGT); 5S rDNA F-(GATGCGATCATACACCAGCACA), R- (GAATGCAACACGGGACTT). The cycling program began with a 5 minute incubation at 95°C to activate the enzyme followed by 35 cycles of 95°C for 15 seconds, 53°C for 15 seconds and 72°C for 30 seconds, and finally 1 minute at 72°C. Fold depletion was calculated by the $2^{-\Delta\Delta CT}$ method (Livak and Schmittgen, 2001) using 5S rDNA as the reference standard. Each reaction was performed in triplicate wells and averaged, excluding replicates in which the variability exceeded one cycle.
Fluorescence in situ hybridization (FISH)

Root tip preparation and FISH were performed essentially as described, except the digestion was 80 minutes (Lamb et al., 2007). PCR products were amplified from maize cultivar Seneca 60, using the aforementioned primers for CentC, OPIE, CRM, or as reported for CentA (Mroczek and Dawe, 2003)

Identification of oat CENH3 (\textit{Avena sativa} CENH3)

To generate genomic clones of the 5’ region of the oatCENH3 gene, primers F2 (GCACMGGCSTGAGGA) and R1 (TTCTCTCTCTCCGCAGYGC) were used. From the genomic DNA sequence (Genbank accession FJ155067), new primers were used to screen a cDNA preparation from ‘Carolina Oat’. Primers were oatCH3F2 (ACGCCCAGAAGCAGCTCAAG) and octrace2 (GCAGCCGTGCCAGGCTTAAC). The RTPCR products were cloned, sequenced and the sequence was deposited into GenBank (accession FJ155068). The translated sequence was used to identify a unique peptide, SKPTPKKQLKGRSPGQTAE, which was subsequently used to generate a polyclonal antibody (Biosource, now Invitrogen).

Immunolocalization

Immunolocalization on young anther tissue was performed essentially as described (Shi and Dawe, 2006), except that dissected, unfixed anthers were treated with 2% cellulase R10, 1% pectolyase Y23 (Karlan) in 1X PHEM buffer for 15-30 minutes to degrade cell walls prior to fixation.

Measurements and statistical analysis of kinetochores

Fluorescent signal intensities were measured using SlideBook 4.0 software (Intelligent Imaging Innovations, California), essentially as described previously (Du and Dawe, 2007). In brief, a projection image of relevant z-stacks from a raw 3D-image was created for each cell preparation. Total CY3 intensity values were measured for non-overlapping kinetochores using a brush tool to paint each signal. Kinetochores intensities were each divided by the area and then averaged to calculate densities. When the signal was split (in anaphase I lagging chromosomes), the values were averaged and taken as one data point. Excel software was used to calculate the variance of the neocentromere and averaged oat kinetochores intensities among cells. An F-test of equality of the kinetochores variances was conducted as
follows, where $A$ is the neocentromere value, and $B$ is the average oat value: $H_0: \frac{\delta_A^2}{\delta_B^2} = 1$ versus $H_A: \frac{\delta_A^2}{\delta_B^2} > 1; F = \frac{s_A^2}{s_B^2} = 4.5019$, d.f. = (6,6); $F_{0.05}(6,6) = 4.2839$ (F-table: Bhattacharyya & Johnson, *Statistical Concepts and Methods*; p604); $4.5019 > 4.2839$, therefore $H_0$ is rejected at $\alpha = .05$.

**FUNDING AND ACKNOWLEDGEMENTS**

This work was supported by a grant from the National Science Foundation (0421671) to R.K.D., and by a University of Georgia Graduate School Dissertation Completion Award to C.N.T. We sincerely thank Beau Brouillette and Eleanor Kuntz for helpful statistics discussions.

**REFERENCES**


Figure 3.1 Neocentromere Material Does Not Contain Maize Centromere DNA or SSR Marker DNA from the Maize Chromosome 3 Long Arm

A series of SSR markers spanning maize chromosome 3 were used to probe a panel of chromosome three truncation lines, BC$_1$F$_2$-4/9 ‘neocentromere’ progeny, a full maize chromosome 3 disomic addition line (OMAd3.1), and the oat parent of the addition cross (Sun II). The grey bars represent BAC contigs (ctg) that correlate the SSR marker position to the IBM2 physical map. The putative ch3 breakage site maps to the gap between ctg120 and ctg121 on the short arm. Each line was also Southern blotted and probed with the major maize centromere DNA element, CentC, demonstrating the loss of the original centromere in neocentromere lines.
Figure 3.2 Maize Neocentromere Lines Lack Centromere DNA but are Stably Inherited

Quantitative PCR analyses of leaf tissue from six neocentromere siblings (CTo1_3, 8, 12, 33, and 40) show a sharp decrease in OPIE and CRM retroelements, and did not detect the CentC centromere repeat (Ct = 35 was used arbitrarily for illustration, see Materials and Methods). The full disomic addition line, OMAAd3.1 is used as a control. Quantitative measurements presented below the graph are expressed as depletions relative to OMAAd3.1 after normalization to the 5S rDNA content of each sample. The maize chromosome of Cto1_3 seems to be inherited at lower frequency than the other neocentromere chromosomes, all of which share the same parent.
Figure 3.3 Neocentromere-Containing Chromosomes are Distinguished from Oat Chromosomes

(a) DAPI (blue) staining and FISH using an OPIE probe (green) identifies stable neocentromere chromosomes in a metaphase spread, whereas a CentC FISH probe (red) does not stain. (b) Full addition line chromosomes are larger than their neocentromere counterparts and their native centromeres are identified by CentC staining. (c) FISH with a CentA probe (green) stains centromeres, pericentromeres, and speckles chromosome arms of maize B73 chromosomes. (d) The same CentA probe stains oat centromeres/pericentromeres, and stains the maize ch3S neocentromere chromosome asymmetrically (circled in red). (e) The unstable neocentromere chromosome CTo1_3 (red arrowhead) is a fraction of the stable size (a), and is symmetrical around a primary constriction.
Figure 3.4 Identification of the Oat CENH3 Protein

(a) A CLUSTAL (Larkin et al., 2007) alignment using the N-termini of the rice, maize, and oat forms of CENH3 proteins. (b) The underlined sequence was used to generate a peptide antibody that detects oat but not maize CENH3 in mixed cell immunolocalization-FISH experiments. Oat cells are larger and stain for CENH3 (red), whereas maize cells are smaller and stain for CentC (green) but not oat CENH3. (c) Oat pachytene and (d) metaphase I immunolocalizations using the oat CENH3 antibody (red or green), stains oat centromeres.
Figure 3.5 The Maize Neocentromere Formed Submetacentrically

(a, b, c) CENH3 immunolocalization on meiocytes of stable lines reveal that the neocentromere is located at a submetacentric position on the chromosome arm. CENH3 is shown in red and tubulin is shown in green. (d) No CENH3 was detected on the unstable CTo1_3 neocentromere from a pachytene cell, despite clear staining on corresponding oat chromosomes (signal intensity of the maize chromosome has been enhanced for viewing - inset).
Figure 3.6 CENH3 Stainings of Maize Neocentromeres are More Variable than Those of Established Oat Centromeres

Cell spreads were immunostained with oat CENH3 (red), tubulin (green, c-f only), and counterstained with DAPI (blue). The fluorescence intensity was measured to estimate kinetochore size. Numbers indicate the relative intensity of the CENH3 of neocentromeres (arrowheads) relative to the mean intensity of oat kinetochores. (a, b) Two spreads from the same slide illustrating variation in CENH3 intensity. (c, d & e, f) Two similarly staged cells from the same neocentromere line illustrating a correlation between kinetochore size and attachment to the spindle. The smaller neocentromere (60% of mean oat; c, d) appears to be detached from the spindle, whereas a proportionally much larger neocentromere (90% of mean oat; e, f) is attached to a kinetochore fiber.
The conditions under which a centromere forms may contribute to its stability. Upon formation, a neocentromere (blue oval) is subject to removal by endogenous forces that normally suppress ectopic centromeres (green arrows). Under certain conditions, reinforcement factors (red and yellow arrows) counteract CENH3 removal and stabilize the neocentromere. If the reinforcement factors are relatively strong (size reflects its relative strength; right panel), establishment is favored and the centromere stabilizes rapidly. If reinforcement factors are weak to begin with (left panel), the centromere may be unstable and fall below the minimal threshold (dotted line). Alternatively, the reinforcement factors may gain strength over time, eventually resulting in a stable centromere.

**Figure 3.7 A Speculative Model of Neocentromere Establishment**
Table 3.1 Stability of Neocentromere Lines

<table>
<thead>
<tr>
<th>Plant Line</th>
<th>Number of cells with OPIE signal</th>
<th>Number of cells without OPIE signal</th>
<th>Percent transmission</th>
</tr>
</thead>
<tbody>
<tr>
<td>OMAd3.1 – control</td>
<td>331</td>
<td>0</td>
<td>100%</td>
</tr>
<tr>
<td>Cto1_3 – unstable</td>
<td>4</td>
<td>131</td>
<td>3.0%</td>
</tr>
<tr>
<td>Cto1_12 – stable</td>
<td>408</td>
<td>2</td>
<td>99.5%</td>
</tr>
<tr>
<td>Cto1_33 - stable</td>
<td>318</td>
<td>0</td>
<td>100%</td>
</tr>
</tbody>
</table>
Table 3.2 Relative Neocentromere Intensities

<table>
<thead>
<tr>
<th>cell</th>
<th>percent of mean oat intensity</th>
<th>Number of standard deviations from mean oat</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>75.2%</td>
<td>3.1</td>
</tr>
<tr>
<td>2</td>
<td>59.6%</td>
<td>3.2</td>
</tr>
<tr>
<td>3</td>
<td>87.0%</td>
<td>2.0</td>
</tr>
<tr>
<td>4</td>
<td>32.5%</td>
<td>2.8</td>
</tr>
<tr>
<td>5</td>
<td>54.3%</td>
<td>6.8</td>
</tr>
<tr>
<td>6</td>
<td>79.4%</td>
<td>3.4</td>
</tr>
<tr>
<td>7</td>
<td>80.7%</td>
<td>2.3</td>
</tr>
<tr>
<td>8</td>
<td>90.3%</td>
<td>1.1</td>
</tr>
</tbody>
</table>

The number of standard deviations (σ) of each neocentromere kinetochore from the mean oat intensity was calculated from the following formula: \( x = \frac{\text{mean oat intensity} - \text{neocentromere intensity}}{\sigma_{\text{oat}}} \).
CHAPTER IV:

OVEREXPRESSION OF HTR12, THE ARABIDOPSIS THALIANA CENH3, IS DIFFERENTIALLY REGULATED ON A CELL-TO-CELL BASIS, AND RESULTS IN CHROMOSOME MIS-SEGREGATION AND CELL DEATH.

INTRODUCTION

The centromere locus ensures the inheritance of genetic information, yet has little genetic identity of its own. Centromere DNA changes so rapidly that there are no unifying sequence elements among taxa (Malik and Henikoff, 2002). In stark contrast, centromeres share dozens of conserved proteins necessary for centromere function in cell division. Many of these proteins assemble into various subgroups during M-phase to form the kinetochore, which is the large proteinaceous structure that links chromosomes to the spindle (Meraldi et al., 2006). Proper localization and regulation of the kinetochore is directed by a set of constitutive centromere proteins called ‘foundation’ proteins (Amor et al., 2004). This core group of proteins, including CenH3, CENP-C, and MIS12, interact with centromere DNA to create a unique chromatin structure that specifies the site of kinetochore assembly. Centromere chromatin must be stably propagated in the same position otherwise chromosomes will align poorly during meiosis, causing genetic defects (Kemp et al., 2004). Despite their ever-shifting genetic components, centromeres maintain their vital function.

At the heart of this seeming paradox is a group of histone variants, collectively known as Centromere Histone 3s, or CenH3s. At centromeres, CenH3s replaces the canonical histone H3 (H3) in the nucleosome structure that winds DNA into the most basic level of chromatin. Normally the protruding N-termini of H3 molecules are subject to a suite of conserved covalent modifications that direct higher-order structuring of chromatin fibers. CenH3s are refractory to such regulation, because their N-termini have different amino acid sequences (Malik and Henikoff, 2003). CenH3 nucleosomes also wind DNA with a different geometry than that of usual histones (Black et al., 2004; Bloom et al., 2006). Thus, chromatin with distinct physical properties provides the foundation for the centromere (Dalai et al., 2007).

How CenH3s are regulated is then a central question of centromere biology. Major insight came from experiments which demonstrated that in contrast to canonical H3, CenH3 deposition is uncoupled from and occurs after DNA replication (Shelby et al., 2000; Ahmad and Henikoff, 2001; Lermontova et al., 2007). Centromere chromatin must therefore exist in an intermediate state for part of the cell cycle, which would seemingly endanger its stable propagation. Either CenH3s are targeted to centromeres
containing H3 nucleosomes deposited during replication, or to centromeres containing nucleosome gaps (Smith, 2002). Further confusing the issue are differences among plants and animals as to when CenH3 deposition actually occurs. For example, Arabidopsis thaliana centromeres are replenished during late G2 (Lermontova et al., 2007), whereas the most current animal data provides evidence for anaphase replenishment, implying that centromeres function with half the pre-replication complement of nucleosomes (Allshire and Karpen, 2008). No common set of proteins have been found that target CenH3s to centromeres. In yeasts and animals dozens of networked proteins have been implicated in CenH3 localization, including general histone binding proteins (Foltz et al., 2006; Allshire and Karpen, 2008). Yet the larger question of how these proteins are themselves localized to the centromere remains unanswered.

Centromeres occasionally form at new chromosomal loci, and these events are thought to have profound effects on karyotype and even species evolution (Ventura et al., 2001; Capozzi et al., 2008; Marshall et al., 2008). New centromeres are rare in nature but can be induced by disrupting centromere replication, maintenance, or surveillance systems. In budding yeast and Drosophila melanogaster, interference with proteolysis results in non-centromeric CenH3 deposition (Collins et al., 2004; Moreno-Moreno et al., 2006) raising the possibility that CenH3 is incorporated broadly but retained only at centromeres. Heterochromatin surrounds centromeres in most organisms, and several lines of evidence suggest that it is also important for new centromere formation. Studies of engineered mammalian chromosomes and yeast neocentromeres, show that efficient centromere formation requires a proper balance between heterochromatin and centromeric chromatin (Nakashima et al., 2005; Okamoto et al., 2007; Ishii et al., 2008; Nakano et al., 2008).

Kinetochore stability is also affected by the abundance of CenH3. Studies in Saccharomyces cerevisiae, Drosophila, and human cells found that increasing the expression of transgenic CenH3 causes the protein to spread into chromosome arms (Van Hooser et al., 2001; Heun et al., 2006; Collins et al., 2007). Other data from humans showed that CenH3 overexpression caused a partial recruitment of kinetochores that, while lacking the capacity to move chromosomes, included the key inner kinetochore
protein CENP-C (Van Hooser et al., 2001). A more recent study in *Drosophila* re-examined this issue using inducible promoters to control the expression of transgenic CenH3 fused to GFP (Heun et al., 2006). CenH3 protein levels of 10, 20, and 30 fold greater than controls resulted in sharply increasing mitotic dysfunction and cell death. Careful examination of aberrant mitoses revealed multiple previously non-centromeric microtubule attachments. Ectopic attachment sites co-localized with inner and outer kinetochore proteins, demonstrating bona fide neocentromere function. Although CenH3 was broadly localized throughout euchromatin, only a few loci were attached to the spindle, leading to an interesting speculation that CenH3 density must meet a threshold to recruit kinetochores (Heun et al., 2006).

We interpret these data as implying that centromere formation is typically suppressed by factors such as proteolysis of CenH3. However, when CenH3 is protected by kinetochore proteins, embedded in heterochromatin-rich region, or overexpressed, centromere formation is favored. The establishment and continued stability of a neocentromere may depend on the confluence of centromere identity factors including heterochromatin, RNA, sequence elements that favorably bind CenH3, and other centromere proteins such as CENP-C.

Here we show that constitutive overexpression of the *Arabidopsis thaliana* CenH3 gene, *HTR12* (Talbert et al., 2002), can cause chromosome mis-segregation and cell death in the tissues of transgenic plants. Families overexpressing *HTR12* exhibited a wide range of organ abnormalities, but the severity of these phenotypes was not tightly linked to overall RNA levels. Cytological inspection of individual cells revealed dramatic cell-to-cell differences in HTR12 protein levels, providing evidence that some cells could not suppress ectopic chromatin. Cells that accumulate high levels of HTR12 may be prone to form neocentromeres, explaining the observed cytological and morphological defects.
RESULTS

HTR12 overexpression causes developmental defects

In order to test our hypothesis that plant CenH3 is sufficient to promote neocentromere activity, we overexpressed the *Arabidopsis thaliana* gene HTR12 in transgenic plants (HTR12ox plants). We used the 35S promoter to drive constitutive expression of the either the native HTR12 gene, HTR12 fused to the epitope tag hemagglutinin (HTR12-HA), or HTR12 fused to yellow fluorescent protein (HTR12-YFP) (see Materials and Methods). A wide range of aberrant phenotypes were noted, including extremely curled leaves with irregular edges, a determinate primary inflorescence that withers off, misplaced, misshapen, and fused organs, as well as various degrees of male and female sterility (Figure 4.1). Developmental defects of transgenic plants were noted in families from each construct-type, suggesting that the fusion tags made little difference to protein function. Furthermore, analyses of several families for up to three generations yielded no discernable pattern of inheritance for the abnormal phenotypes – selfed progeny of some severely affected plants gave rise to phenotypically normal progeny and vice-versa. Surprisingly, the cumulative defects of any given plant were only loosely related to its level of HTR12 overexpression. We developed an index to categorize the relative phenotypic severity of several plants on a 0-6 scale and compared this number to the amount of HTR12 RNA as measured by qRT-PCR (Figure 4.2a) (see Materials and Methods). Only a weak positive relationship could be inferred between these two variables, suggesting the involvement of additional factors. In fact, many siblings with the same phenotypic severity differed greatly in HTR12 levels (Figure 4.2b, c), and some plants that looked normal expressed abundant HTR12 RNA (Figure 4.2d). Given what is known about the regulation of CenH3 in general, we wondered if protein abundance affected the observed phenotypes.
**HTR12 is regulated on a cell-to-cell basis**

To assess cellular levels of transgenic HTR12 protein, we used wide field and confocal microscopy to view fixed and live tissue of HTR12-YFP plants (see Materials and Methods). Interphase nuclei prepared from young flower buds revealed striking differences in the patterns and amounts of YFP fluorescence among cells (Figure 4.3a-c). In some cells the signal was nearly undetectable. Generally some signal was detected in punctate spots corresponding to the ten centromeres of diploid Arabidopsis cells, but the brightness of these spots varied. Nuclei with the brightest spots were usually accompanied by diffuse nuclear staining of varying intensities, sometimes nearly as strong as centromeric signals. These data are in line with the tendency of mis-regulated CenH3s to spread from established centromeres onto chromosome arms. In one exceptional case, eleven clear spots were observed (Figure 4.3a-c inset) indicating that a localized concentration of HTR12 had formed in a region of condensed heterochromatin. This observation supports a model of heterochromatin-mediated CenH3 stabilization, especially since there is otherwise sparse non-centromeric YFP signal in this cell. Three-dimensional imaging of live roots further suggests that HTR12 is differentially regulated among cells (Figure 4.3d). Native YFP in some regions was not detected despite extremely abundant signal in others. Regional differences may be explained in part by the stereotypical division patterns of root cells. For instance, cells of a particular lineage may inherit different amounts of HTR12-YFP protein. Regardless, we expect constitutive expression of the HTR12-YFP transgene to result in some signal, unless the protein is efficiently removed.

**HTR12 overexpression causes cell division defects resulting in mininuclei**

The abundance of ectopic HTR12 protein in some cells provided the possibility of neocentromere activity during cell division. Cell division is an extremely precise process; particularly anaphase when sister chromatids separate (Figure 4.4a). HTR12 overexpression caused a highly significant increase of lagging chromosomes during anaphase (p-value = .001; Table 4.1). The small size of Arabidopsis mitotic chromosomes and the disorder of most spindles did not allow us to unambiguously determine where microtubules attached (Figure 4.4b). However, chromosomes were often stretched horizontally across the
division plane (Figure 4.4b-d). This unusual geometry is often an outcome of microtubules pulling opposite arms of a chromatid apart due to neocentromere function (Heun et al., 2006). Vertically oriented chromosomes were observed in some cases, but it was not clear whether they were a result of cohesion defects or bipolar attachments to a single kinetochore (example Figure 4.4b).

Chromosomes slow to segregate are often broken or mis-segregate and form ‘mininuclei’ in daughter cells. We observed a highly significant percentage of mininuclei in flower buds from Htr12ox plants with severe floral phenotypes (p-value = .001; Table 4.1). By examining dissected flower buds (see Materials and Methods), we observed that some groups of cells had a high percentage of mininuclei and were clearly associated with lagging chromosomes (Figure 4.4e). Some mininuclei had two distinct DAPI bodies (Figure 4.4f), consistent with chromosome breakage or multiply mis-segregated chromosomes. However, mininuclei and cell division defects were rare in some clumps of cells (image not shown; data contributed to table 4.1), consistent with the regional differences in HTR12 abundance (Figure 4.3d).

**HTR12 overexpression causes cell death**

Repeated dysfunction resulting in chromosome loss or breakage will eventually cause lethality in some cells. We stained living plant organs with Propidium Iodide (PI), a fluorescent molecule that can stain nuclei of dead or dying cells, but is excluded from living cells (see Materials and Methods). Cells in the tissue of wild type controls were refractory to PI stain (Figure 4.5a), except for those damaged at the point of dissection (not shown). Clusters of nuclei from phenotypically severe Htr12ox floral organs were commonly stained (Figure 4.5b, c). These results provide a physiological basis for many of the aberrant morphologies observed in Htr12ox plants (Figure 4.1).

**DISCUSSION**

The overexpression of the *HTR12* gene is clearly disruptive to *Arabidopsis* plant architecture. One explanation is that ectopic HTR12 mis-regulates important developmental genes. While this may occur in some cases, many persistent phenotypes such as curled leaves with irregular margins are better explained by unequal growth rates of cells within a tissue. For example, disruption of the normal cell division
patterning of *Antirrhinum* leaves caused buckling of the surface (Nath et al., 2003). Furthermore, cell death is clearly causative of some HTR12ox phenotypes (Figure 4.1b, d, h-j). Similar cell death and consequent morphologies have been described before. For instance, a mutation in an essential *A. thaliana* gene caused localized cell death and very similar phenotypes, including bent siliques, wrinkly leaves and premature senescence of the primary inflorescence (Mou et al., 2000). *Arabidopsis* plants deficient in telomerase undergo repeated chromosome breakages and retain little semblance of developmental patterning (Riha et al., 2001). Other defects such as organ fusions (Figure 4.1e, f, g) would be a likely result of cell death at the borders of developing organs in the meristem. We provide direct evidence of cell death in floral organs of HTR12ox plants (Figure 4.5). Since the meristem is actively dividing, the most likely cause of the observed developmental abnormalities is HTR12-induced chromosome mis-segregation and cell death.

We assume that as in *Drosophila*, cells with high levels of CenH3 protein are most prone to dysfunction. However, we also observed a condensed locus of HTR12-YFP in a cell with otherwise little fluorescent signal (Figure 4.3a and c, insets). This cell type indicates that HTR12 is preferentially stabilized at some loci, such as constitutive heterochromatin. Given the typical suppression of neocentromeres through the removal of CenH3 (Figure 4.3) (Collins et al., 2004; Moreno-Moreno et al., 2006), such stabilized loci may represent the functional ‘threshold’ postulated by Heun et al. For example, cell-to-cell differences in protein levels could result from self-reinforcing HTR12 stabilization in some cells, and efficient removal from others (Figure 4.3). In that case, cell-specific regulation of HTR12 could account for the noted regional differences in cell division abnormalities and cell death (Figures 4.4 and 4.5).

Nonetheless, a specific role for HTR12 in chromosome mis-segregation cannot be unequivocally assigned. Neocentromeres induced by CenH3 overexpression in *Drosophila* were defined by a full suite of kinetochore proteins. Heun et al further demonstrated that multiple points of spindle attachment resulted in chromosomes being stretched in unusual ways. Most lagging chromosomes in HTR12ox tissue were similarly stretched from pole-to-pole, although we could not verify where they were attached to the...
spindle. A dominant negative effect on native centromere function remains a possibility. However, no study using low levels of CENH3-YFP has reported dysfunction, despite reporting clear localization of the protein to native centromeres (Henikoff et al., 2000; Van Hooser et al., 2001; Fang and Spector, 2005). Based on this evidence, we think that ectopic HTR12 is sufficient to induce neocentromere function in Arabidopsis.

In summary, overexpression of CenH3 results in severe developmental abnormalities in plants. These defects are coincident with dramatic levels of chromosome mis-segregation and regional cell death. Constitutive expression of CenH3 resulted in non-centromeric accumulation in many cells, suggesting a failure to suppress ectopic centromere chromatin. Besides extending key data from Drosophila, the work here also represents the first study of CenH3 overexpression in a mature organism, because excess CenH3 was lethal to fly embryos. Future experiments could capitalize on the fertility of HTR12 overexpression plants, and follow putative neocentromeric chromosomes through meiosis into following generations.

MATERIALS AND METHODS

Construct Preparation and Transformation

Arabidopsis thaliana ecotype Columbia was used exclusively in this study and were grown in pots at 22°C with 16 hours of fluorescent light per day (long day conditions). The HTR12 gene (locus At1g01370) was cloned into pENTR/D (Invitrogen) from a floral cDNA library (kindly given by Richard Meagher and laboratory) using the following primers: HTR12-F1 (5’-CACCATGCGAGAACATCGC GTT-3’); and HTR12stop-R1 (5’-TCACTGGTTCTGCCTCCAA-3’) for N-terminal or untagged constructs, or HTR12nostop-R1 (5’-CCACTGTCTGGTTCTCCAAAG-3’) for C-terminal tagged constructs. ‘Stop’ or ‘nostop’ clones were transferred into the following pEARLEYGATE vectors (Earley et al., 2006) using the Clonase II enzyme (Invitrogen): pEG100 (no tag); pEG101 (C-terminal YFP-HA tag); pEG201 (C-terminal HA tag); or pEG104 (N-terminal YFP tag). All constructs constitutively express the HTR12 gene through an enhanced CaMV 35S promoter; N and C-terminal tags gave indistinguishable phenotypes. Sequences of pEG plasmids were verified prior to
their transfer to *Agrobacterium tumefaciens* strain C58C1 (also a gift of the Meagher lab). Plants were transformed using the standard floral dip method.

**Phenotype severity index**

The following scoring system was used to assign cumulative point values to commonly observed phenotypes of HTR12ox plants: small (0.5 point) or minute plant (1); majority of leaves curly (1); majority of leaves with severely misshapen leaf margins (2); misshapen siliques (1); no/abrogated petals (1) for an index of 0-6 points.

**Quantitative Reverse Transcriptase Mediated Polymerase Chain Reaction (qRT-PCR)**

DNase I treated RNA extracted from young rosette leaves (RNeasy kit -Qiagen) was reverse transcribed using random hexamers and Superscript III enzyme (+RT) as specified by the manufacturer. Roughly 400ng of the resulting cDNA (1µl) was subsequently used in a quantitative PCR reaction (Tm = 58°C, described in Chapter III, Materials and Methods) with the following primer sets: HTR12-F1 (5’-GAG AAC CAA GCA TCG CGT TA-3’) and HTR12-R1 (5’-GGC GGC ATC AGT TTG ATT C-3’); At18s-qPCR879-s (5’-GGG GGC ATT CGT ATT TCA TA-3’) and At18s-Qpcr944-a (5’-TTC GCA GTT CGT CTT TC-3’) (18S primers another gift from the Meagher lab). Relative fold enrichment (RFE) of *HTR12* was calculated by the 2-ΔΔCt method (Livak and Schmittgen, 2001). *HTR12* expression (as measured by C(t) value, the fluorescence intensity at an arbitrary threshold) in transgenic plants was first normalized against the mean endogenous *HTR12* levels in wild type controls, and subsequently against ubiquitous 18S rRNA transcript levels (to account for small differences in sample concentration). All reactions were performed in triplicate wells, and –RT controls were used to verify the absence of DNA contamination. Similar results were achieved with another *HTR12* primer set, as well as primer sets that amplify a portion of YFP (not shown).

**Microscopy and Staining Techniques**

For widefield microscopy of nuclei, HTR12-YFP or wild type flower buds were fixed in 4% paraformaldehyde diluted in Phosphate Buffered Saline (PBS) under a vacuum for 30 minutes. After rinsing, buds were chopped finely with a razor blade and passed through a 48µm nylon filter. ~10µl of
nuclei were mounted in Vectashield anti-fade with DAPI (Vector Laboratories). Images were taken with Slidebook software (Intelligent Imaging Innovations) using a 40X objective and FITC or DAPI filter sets. Chromosome spreads were prepared in a similar way, except after fixation, flower buds were digested in an enzyme mix of 0.33% Cellulase, 0.33% Pectolyase, and 0.33% Cytohelicase in 10mM Citric Buffer, pH 4.5 for up to two hours. In some cases, mouse anti-tubulin antibodies were applied and detected by fluorophore-conjugated anti-mouse secondary antibodies as previously described (Dawe et al., 1999).

Confocal Scanning Laser Microscopy of live cells expressing HTR12-YFP was conducted at the Center for Ultrastructural Research, University of Georgia (http://www.uga.edu/caur). Roots were growing in chambered coverslips containing 1/2MS media as described (Fang and Spector, 2005). Propidium Iodide staining was as follows: organs from middle-staged flower buds were carefully dissected and placed in 10µg/ml PI diluted in water for 10 minutes, before rinsing and mounting under a coverslip. Images were taken with a Cy3 filter set on a 40x objective.

REFERENCES


Figure 4.1 *HTR12* Overexpression Causes Multiple Developmental Abnormalities

(a) Wild type *Arabidopsis thaliana* ecotype Columbia-0 grown under standard conditions (see Materials and Methods). (b) Example of an *HTR12*overexpression plant at a similar developmental stage exhibiting numerous defects. The following sub-figures are not necessarily from this plant and are meant to encompass part of the pleiotropic effects caused by overexpressing *HTR12*. (c) WT (top) and mutant (bottom) leaves of similar developmental stages. The margins of mutant leaves are often twisted and malformed. (d) Close-up of a terminal flower bud (from (b)). Such buds initially appear to develop normally before turning yellow with purple tips, then completely browning and withering off the plant. Only the first floral cluster of some mutant primary inflorescences developed in this way, and secondary branches were fertile. (e) Fused inflorescences. Normally flower buds branch off the continuously
elongating inflorescence stem at fairly regular alternating intervals (a). Many mutant inflorescences initiated sporadically (b) or in a large determinate cluster. (f) Bizarre mutant inflorescence with multiple abnormalities including random organ initiation and misshapen flower buds and siliques. (g) Three fused mutant siliques. (h) A moderately bent mutant silique indicating embryo death. More severely bent siliques were often completely sterile. (i and j) Pedicel (i) and stem (j) rot indicating localized cell death in these tissues.
Figure 4.2 The Level of *HTR12* Overexpression Correlates Weakly with the Phenotypic Severity of Transgenic Plants

(a) Graph plots the phenotypic severity of each transgenic plant versus the relative fold enrichment (RFE) of the *HTR12* gene as calculated by qRT-PCR (see Materials and Methods). Several different *HTR12* overexpression families, as well as two WT plants are included. These data show a very weak linear correlation between the amount of *HTR12* RNA and the phenotype of the plant (Pearson’s correlation coefficient, r = 0.149). (b and c) Two mutant siblings with identical phenotypes but an almost 20x difference in *HTR12* expression levels. (d) Strong *HTR12* overexpressing plant with no mutant phenotype. Inflorescences were removed in (b,c,d) for imaging purposes.
Figure 4.3 HTR12-YFP Protein Accumulates Differently on a Cell-to-Cell Basis

(a) Native YFP fluorescence from fixed nuclei isolated from an HTR12-YFP expressing flower bud is shown in green, and merged with 4’, 6-diamidino-2-phenylindole (DAPI) staining of DNA in blue. (b) DAPI channel alone. (c) YFP channel alone. (a and c). Cells from the same tissue are highly variable for HTR12-YFP protein levels. White arrow marks the cell enlarged in the inset of (a,b,c). This cell clearly shows eleven punctate YFP spots, corresponding to ten diploid A. thaliana centromeres, and one smaller
locus of ectopic HTR12 protein, indicated by the red arrow. (d) Native YFP fluorescence from a live lateral root tip of an HTR12-YFP plant shown in yellow. This image represents a z-stack of two-dimensional images that pass through the root. HTR12-YFP is variable among neighboring cells, with some regions of high concentration and some regions with very little protein.
**Figure 4.4 HTR12 Overexpression Causes Cell Division Defects Resulting in Mininuclei**

Cells were isolated from fixed flower buds and prepared according to Materials and Methods. DNA is stained with 4’, 6-diamidino-2-phenylindole (DAPI), shown in blue, and microtubules are shown in red using an anti-tubulin antibody. (a) Late anaphase of a typical wild type *A. thaliana* cell. (b) Late anaphase of an *HTR12* overexpressing cell is massively disordered. Green arrow indicates one vertically-oriented chromosome, or chromosome pair, attached to microtubules from both sides of the spindle. (c and d) Two examples of the typical anaphase defects observed in mutant cells. Lagging chromosomes are usually stretched horizontally, indicating non-centromeric sites of microtubule attachment. (e) A field of
adjacent cells from mutant tissue. White arrows indicate mininuclei that form as a result of chromosome mis-division, such as indicated by the yellow arrow. (f) Close-up of a mininucleus suggests the presence of multiple chromosomes or chromosome fragments.
Figure 4.5 Plants Overexpressing *HTR12* Exhibit Cell Death in Developing Floral Organs

Propidium Iodide (PI) stains the nuclei of dead or dying cells (shown in bright red), but is excluded by the plasma membrane of living cells. (a) No detectable signal from a wild type gynoecium indicates the tissue is healthy. (b and c) Staining of gynoecium (b) and anther filament (c) from an *HTR12* overexpressing plant indicates regional cell death. Gynoecia of mutant plants were commonly malformed, partially fused to other organs, and sterile. Anthers of mutant plants were commonly truncated, fused, sterile, or absent.
Table 4.1 Frequencies of Chromosome Mis-Division and Mininuclei

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Total Anaphase Figures</th>
<th>Frequency of Lagging Anaphase Chromosomes$^1$</th>
<th>Total Interphase Cells</th>
<th>Frequency of Mininuclei$^1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>HTR12ox</td>
<td>32</td>
<td>$46.9 \pm 23.4%$***</td>
<td>332</td>
<td>$8.4 \pm 3.9%$***</td>
</tr>
<tr>
<td>Wild Type</td>
<td>67</td>
<td>3%</td>
<td>300</td>
<td>0%</td>
</tr>
</tbody>
</table>

$^1$ HTR12ox treatment resulted in significant increase over wild type; p-value = .001
CHAPTER V:
SUMMARY AND CONCLUSIONS

The maintenance and function of eukaryotic centromeres is an unparalleled epigenetic phenomenon. Often megabases in size, centromeres are regulated by stringently conserved genetic pathways to perform a most fundamental task, yet they are themselves not conserved at the sequence level. Regardless, centromeres somehow occupy stable positions on each chromosome for millions of years. Understanding the relationship between centromere proteins and their DNA binding partners is crucial to illuminate not only the basic features of centromere biology, but also new modes of epigenetic stability.

Centromeres of cereal plants such as maize, rice, barley, and oats provide an interesting exception to the poor conservation of most centromere DNA. While each species has a unique tandem arrayed satellite repeat, all cereals share a group of well-conserved retrotransposons that are specific to centromere regions (Miller et al., 1998; Presting et al., 1998; Hudakova et al., 2001). The maize member of this group is known as CRM, or centromeric retrotransposon maize. CRMs are distributed widely across the centromere of each maize chromosome, interspersed with the maize satellite repeat CentC (Ananiev et al., 1998c). CRMs in maize retain a high level of sequence identity with one another, diagnostic of recent transposon activity (Zhong et al., 2002; Jiang et al., 2003). We were therefore intrigued at the prospect that CRMs may play a role in the rapid sequence evolution of maize and other cereal centromeres. In a set experiments using the chromatin immunoprecipitation method (ChIP), we had previously demonstrated that CRM and CentC were the major elements bound by nucleosomes containing the centromere specific histone variant CenH3 (Zhong et al., 2002). However, it remained unclear how a dynamic process such as retrotransposition could aid the function of a highly conserved and regulated set of centromere proteins.

Concurrent reports linking RNA to the structure and function of pericentromeric heterochromatin provided a major clue. Maison et al. published evidence that in human cells, RNA was necessary for the
structure of constitutive chromocenters formed by condensed heterochromatin during interphase (Maison et al., 2002). Volpe et al. used mutants of *Saccharomyces pombe* to elucidate how such a structure could be arranged by an RNAi-like system: low levels of transcription from one strand of pericentromeric tandem arrayed elements are amplified by an RNA-dependant polymerase, cleaved into siRNAs by Dicer that are then bound by Argonaute to direct the localization of histone methylases and the yeast HP-1 homolog, which function to repress transcription from the complementary strand of DNA (Volpe et al., 2002). Volpe et al. further interpreted their results as the mechanism of centromere function, because RNAi mutants exhibited severe chromosome segregation defects (Volpe et al., 2002; Volpe et al., 2003). However, we interpreted the Maison and Volpe findings differently, since chromosome mis-segregation could stem from pericentromeric cohesion defects, and not kinetochore function per se. Our view is supported by data from Volpe et al., 2002, showing that while RNAi mutants de-repressed transcription from complementary strands of pericentromeric repeats, the unique repeats at the kinetochore-forming centromere core remained silenced. By definition, CenH3 chromatin cannot participate in the described repressive loop, because CenH3s lack the required N-terminal residues.

We hypothesized that RNA transcribed from maize CRM elements may also contribute to centromere function, but in a novel way. Expressed sequence tag data deposited in GenBank and our own reverse transcriptase-mediated PCR (RT-PCR) experiments confirmed that CRM retroelements are transcribed in maize tissue. To test a structural role for CRM transcripts, we returned to the ChIP method with modifications that allowed us to recover and assay RNA as well as DNA. Notably, we used the ‘native’ ChIP variation that does not use a crosslinking chemical to fix protein-protein and protein-nucleic acid associations in vivo. We reasoned that avoiding a crosslinker would yield only the strongest covalent and ionic interactions in our biochemically purified sample. Native ChIP using our well characterized antibody against maize CenH3 (Zhong et al., 2002) revealed that RNA identical to the centromere-specific CRM family, but not the widely distributed OPIE retroelement, was physically associated with the CenH3-immune sample. To our surprise, we also discovered CenH3 associated with abundant
transcripts from the non-coding centromere repeat CentC, but not the abundant non-centromeric Knob repeat.

In order to verify our samples actually contained RNA, we treated nucleic acid purified from CenH3 immune samples with RNase A, a single-strand-specific exonuclease. Upon this treatment we found that nearly the entire RNA signal was abolished, leading us to two possible conclusions: either centromere RNAs were only transcribed from one strand, or both strands were transcribed but did not hybridize together. We investigated this question using both DNA and RNA probes specific to each strand of CRM and CentC in membrane blotting experiments. Unlike *S. pombe* pericentromeres, maize centromere cores are naturally transcribed from both strands of DNA in abundance and both strands are associated with CenH3 immune complexes. Further examination of these RNAs by PAGE demonstrated discretely-sized transcripts ranging from 40 to >900 nucleotide (nt) in distinct patterns associated with either strand. Despite the detection of minute quantities of a 28 nt control RNA, we found no evidence for siRNA-sized RNA molecules, suggesting to us that homologous centromere transcripts were specifically maintained in a single-stranded state and thus protected from processing by RNAi.

Our discoveries that maize centromeres are transcribed and that the transcripts are part of the centromere impacted the field in two important ways. First, the longstanding premise that centromeres are heterochromatic and transcriptionally inert, akin to pericentromeres, was directly challenged. Instead our evidence supported the newly emerging view that centromere chromatin has unique physical properties (Dalai et al., 2007; Houben et al., 2007). Studies in rice and humans have since demonstrated that the cores of functional centromeres are hypomethylated, in contrast to pericentromeres (Wong et al., 2006; Zhang et al., 2008). Centromere repeats from numerous plants and animals have been found to be transcriptionally active, pointing to the generality of our findings (Fukagawa et al., 2004; May et al., 2005; Lee et al., 2006). Additionally, numerous functional genes have been identified at a centromere core in rice, and in a human neocentromere region (Lo et al., 2001; Nagaki et al., 2004).

We also identified single-stranded RNA as a component of centromere chromatin and provided evidence that these transcripts are novelly regulated. In the current paradigm of RNAi, transcription from
homologous strands of DNA results in a dsDNA molecule that is cleaved into 21-28 nt small RNAs by a class of endonucleases typified by Dicer (Volpe et al., 2002). Maize CenH3-associated transcripts are somehow refractory to such regulation, but based on their discrete sizes we might speculate that they are processed by a different mechanism. Subsequent studies in mouse and marsupials also describe discretely-sized RNAs encoded from centromere satellites and retrotransposons, although their regulation remains an unanswered question (Bouzinba-Segard et al., 2006; Carone et al., 2009).

The precise function of centromere-encoded RNA also remains unknown. One hypothesis is that the act of transcription aids in the replication-independent replacement of H3 histones with CenH3s, a process necessary for centromere maintenance (Jiang et al., 2003). While this hypothesis has not been tested, it would not seem to explain the retention of transcripts at centromere chromatin. We favor a direct role for RNA in centromere structure, similar to the function of telomeric RNA (Zappulla and Cech, 2004). Additional data from our lab and human centromere research are highly consistent with this view. The maize homolog of the essential centromere foundation protein CENP-C was shown in vitro to bind RNA in a single strand specific manner (Du, Y. and Dawe, R.K. manuscript in preparation). Furthermore, this activity was mapped to a region of the C-terminus whose deletion reduced centromere localization of transgenic CENP-C in maize plants. Transcripts from human centromere repeats were also found to be necessary for the continued maintenance of CENP-C at human centromeres (Wong et al., 2007). Astonishingly, a recent report established that transcripts from a single retroelement at the precise core of the human neocentromere Mardel10 were essential to the maintenance of CenH3 and centromere function (Chueh et al., 2009). Collectively, these results provide convincing evidence that a conserved mechanism involving RNA is required to maintain the principle components of centromere chromatin, and hence the epigenetic identity of centromeres.

RNA may bridge the gap between the genetic and epigenetic identities of centromeres. Single stranded RNAs are well known to form secondary structures essential to their function. Conceivably, selective pressure on a secondary structure of centromere RNA would account for the rapid changes in DNA sequence common to most centromeres. Examination of the C-terminal ssRNA binding region of
maize CENPC reveals a history of adaptive evolution (Talbert et al., 2004), suggesting that constitutive centromere proteins may be ‘keeping pace’ with changes in DNA via RNA transcripts (Jiang et al., 2003). Without further knowledge of the physical function of centromere RNA, however, an evolutionary function will be difficult to study.

Many features of established centromeres have been identified, including a core set of constitutive proteins, transcriptionally active repetitive DNA, and pericentric heterochromatin (Allshire and Karpen, 2008). However, the individual contribution of such factors to the maintenance of centromere identity is difficult to determine because of their interdependency. The study of newly formed centromeres offers an alternative approach to understanding the most critical features of centromere function. For example, analysis of centromere movement in primates has demonstrated that new centromeres often form in gene-poor regions on chromosome arms, but rapidly gain a ‘normal’ centromere structure including satellite repeats (Lomiento et al., 2008; Marshall et al., 2008; Stanyon et al., 2008). We identified a neocentromere event in plants, and took this rare opportunity to investigate epigenetic processes that establish new centromeres.

In certain scenarios, widely diverged plant species can be crossed resulting in retention of one chromosome of the paternal donor by the maternal genome (Ananiev et al., 1997). Chromosomes present in one copy are transmitted normally through mitotic cell division. However, meiotic division requires chromosome pairs, and chromosomes without a partner are often broken by the spindle. We identified a fragment of maize chromosome 3 that had likely broken by such a scenario (Muehlbauer et al., 2000). Several lines of evidence verified that the truncated chromosome lost its native centromere and most of the long arm, yet was retained in the F1 generation (we designated it chromosome 3 short arm, ch3S). Broken fragments without a centromere are usually lost during embryogenesis unless a new centromere forms on them. Curious as to how ch3S was transmitted, we developed an antibody that would recognize the oat CenH3 molecule, since CenH3 is necessary for centromere function, and the maize CenH3 gene is not on chromosome 3. Immunolocalization demonstrated that a new centromere had formed, only the
second such verification in plants. The first neocentromere was similarly identified after a barley chromosome was truncated in a wheat genomic background (Nasuda et al., 2005).

These two cases draw some important comparisons, but also highlight apparently distinct mechanisms of neocentromere formation. Both scenarios were associated with a chromosome breakage event. The idea that centromeres form as a natural means of stabilizing broken genomes is supported by data from dozens of human neocentromeres, by historical evidence of centromere repositioning in animals, and by the efficiency of neocentromere formation after experimental breakage in yeasts (Ishii et al., 2008; Marshall et al., 2008; Ketel et al., 2009). In many of these cases, breakage often occurs near the site of the original centromere. This fact, coupled with the demonstration that CenH3 can laterally ‘spread’ (presumably a process similar to heterochromatin-mediated Position Effect Variegation) (Lam et al., 2006; Scott et al., 2006), have led to arguments that neocentromeres are templated from native centromeres (Maggert and Karpen, 2001). The barley neocentromere may have formed by such a mechanism because the CenH3 chromatin is localized at the very end of the chromosome next to the breakage point. However, the CenH3 chromatin of maize ch3S clearly formed distal to the breakpoint, suggesting de novo centromere formation.

Templat ing and de novo processes can be considered in terms of the stability of new centromeres. Given the normal suppression of ectopic centromeres, it stands to reason that a direct transfer of CenH3 and associated centromere identity factors from an established centromere could result in a more stable neocentromere than would de novo initiation. For instance, the barley neocentromere stably transmitted its chromosome in the first generation. In contrast, the maize ch3S fragment was initially highly unstable in mitosis. Nonetheless, it was inherited and by the F2 was transmitting efficiently. The extremely rapid stabilization of ch3S can most easily be accounted for by epigenetic processes, implicating CenH3. Prior evidence suggests centromere stability depends on sufficient amounts of CenH3 to faithfully recruit a kinetochore. Our observation that de novo-established CenH3 domains fluctuate in size more than those of established centromeres lends anecdotal, yet plausible, support to the idea that the stability of new
centromeres depends on the epigenetic maintenance of CenH3, which in turn depends on its mode of initiation.

Further experimentation in this system would provide an unparalleled opportunity to examine the molecular features of a neocentromere in the generations immediately following its initiation. For example, ChIP purification and deep-sequencing of oat CenH3-associated DNA will identify maize-specific sequences participating in the neocentromere chromatin. In this way the specific positions of the neocentromere could be mapped over successive generations of ch3S inheritance. Identification of neocentromeric DNA sequences would also facilitate the investigation of RNA molecules that may be involved in the epigenetic establishment of centromeres.

To further assess the importance of CenH3 in centromere initiation, we turned to the model plant species *Arabidopsis thaliana*. Ectopic localization of CenH3 in Drosophila was previously reported to induce neocentromere activity. The efficiency of this process increased with rising levels of CenH3, suggesting a relationship between the amount of CenH3 and its ability to recruit downstream components of the kinetochore. We constitutively overexpressed the Arabidopsis CenH3 gene with the expectation that a similar relationship exists in plants.

Neocentromere activity initiated by excess CenH3 should result in cell division defects due to multiple spindle attachments. Unfortunately, the size of Arabidopsis mitotic chromosomes is below the resolving power of a light microscope to unambiguously assign microtubule attachment points. Nonetheless, we detected numerous chromosome segregation defects in anaphase of CenH3 overexpressors, compared to wild type. We could not definitively prove these defects did not stem from a dominant negative effect on the native centromere, however, the horizontal orientation of most mis-segregating chromosomes provides a strong argument for multiple sites of attachment. In support, detailed cytological studies from Heun et al. clearly demonstrate that microtubule attachments on both arms of a chromosome will pull it horizontally ‘taut’ during anaphase (Heun et al., 2006).

Mis-segregating or broken chromosomes are often excluded from the main nucleus and instead form small satellite nuclei. Such mininuclei are diagnostic of the partial aneuploidies and polyploidies that
result from cell division defects. We observed a highly significant increase of mininuclei in the same mutant tissue also enriched for anaphase defects. The imbalances in gene dosage caused by aneuploidies are known to cause cell death, leading us to query the viability of cells in mutant tissue. The membrane exclusion dye propidium iodide was used to demonstrate small regions of cell death in what is normally healthy tissue. A similar localized-cell-death phenotype was previously reported for a mutation in an essential gene involved in fatty acid biosynthesis (Mou et al., 2000). These mutant plants had a plethora of developmental abnormalities tightly correlated with cell death.

Transgenic plants overexpressing CenH3 also exhibited a wide range of developmental defects, but there was surprisingly little overall correlation between the amount of CenH3 RNA and the severity of the plant phenotype. Cytological inspection of both live and fixed tissue showed that cells within an individual organ varied widely in the amount and nuclear distribution of CenH3. As demonstrated by Heun et al., CenH3-mediated neocentromeres and cell death occur most frequently when high levels of CenH3 are incorporated into chromosome arms. A live cell assay using one of our overexpression lines similarly associated mis-segregating chromosomes with high levels of ectopic CenH3 incorporation. Collectively, our data argue that some cells are more prone than others to CenH3-mediated cell division defects and cell death. This argument explains the poor correlation between phenotype and RNA levels, because stochastic death of key developmental cells (such as meristem cells) would cause severe morphological defects, whereas death in established organs could be compensated for by neighboring cells.

A unique form of epigenetic control confers positional stability to centromeres over millions of generations, yet is capable of repositioning them in just one. We have discovered features of plant centromere initiation and maintenance that converge on the regulation of the centromere-specific histone, CenH3. The dynamics of CenH3 are inherently difficult to study given the size, repetition, and essential function of centromeres. Therefore, the most promising avenues of research lie in systems that induce stable centromeres at identifiable or even predictable chromosomal sites. The details of specific processes
that direct centromere initiation and maturation promise to reveal new mechanisms of long-term epigenetic maintenance.

**REFERENCES**


